Online Expert-Based Prediction for Cognitive Radio Secondary Markets

Juan Vanerio and Federico Larroca.
Facultad de Ingeniería, Universidad de la República.
Montevideo, Uruguay.
Email: jvanerio@fing.edu.uy, flarroca@fing.edu.uy

***Abstract*—The growing importance of wireless communications drives an increasing interest in dynamic access to spectrum resources. This requires efficient management policies that allow spectrum sharing between licensed primary users (PU) and unlicensed secondary users (SU). On such scenario, PUs shall preserve their usage priority right over any SU. Also, no SU shall interfere on any PU. Technical viability can be achieved through *Cognitive Radio* devices that adjust their operating parameters adaptively.**

After discussing several economic and technical models to achieve efficient spectrum sharing, we propose an on-demand secondary market model regulated by a *spectrum broker* who controls resource allocation. This model provides economic incentives for both kind of users to cooperate: SUs are charged by the *broker* on behalf of PUs for resource utilization but are indemnified if expelled to ensure PU priority. We describe the main characteristics of such a system and address the question of what allocation decisions should the *broker* take in order to achieve economic benefit regardless of users behavior. Several online expert-based no-regret algorithms are proposed to guide the decision taking process and evaluated under different user behavior patterns. Their results are compared with the ones achieved by dynamic programming to assess its convenience.

***Index Terms*—Access control, Dynamic spectrum access, Decision-making, Communication system economics.**

I. Introduction

In recent years there has been a dramatic increase in the demand (and thus cost) for radio spectrum [1], [2], [4], mainly due to the evolution and growth of wireless networks. This trend is only expected to increase, driven by technologies such as 5G or the "Internet of Things" (IoT) [3], [26].

Instead of relying on traditional policies, where a spectrum administrator issues long-term and exclusive licenses, in the context of Cognitive Radio (CR) and Dynamic Spectrum Access (DSA) the concept of *Spectrum Secondary Market* has emerged [4], [37]. Basically, it allows the exchange of usage rights between licensed primary users (PUs) and unlicensed secondary users (SUs), which should result (in the absence of market failures) in an efficient allocation of the scarce resources due to the information transmitted by prices [29], [31], [41].

After reviewing the possible implementations of the spectrum secondary market proposed so far in the literature, we argue why a centralized, real-time and on-demand scheme should be favored. In this implementation, allocation of radio resources to SUs is managed by a so-called *Spectrum Broker* (or simply "*broker*") [42], [46], [50]. Moreover, any authorized SU can request resources from the broker as they see fit, which in turn decides whether to accept the request or not. The broker is also responsible for enforcing PUs priority usage rights. This means that if a PU requires resources and there is not enough available capacity, it will remove a SU from the system (if any are present) and indemnify her for that situation.

However, for DSA to be viable both types of users must be willing to cooperate and share the spectrum. For SUs motivation comes from fast on-demand resource allocation and eventual compensation for service suspension. Earnings coming from the fees charged to SUs should be the economic incentive for the spectrum broker and the PUs to participate (as long as the latter's priority usage is asserted at all times).

Our main contribution is to study how this economic incentive may be achieved under a broad set of user behaviors and operating conditions. In particular, we are interested in providing the spectrum broker with simple and robust decision taking rules that maximize the benefits. We also discuss how, when benefits are maximized, the compensation due to a suspended service provides proper incentives (in terms of the resulting Quality of Service) for the SUs to participate.

The main difficulty of this maximization problem lies on how to model the user's behavior taking as few assumptions as possible. To handle this issue we resort to mathematical tools known as *no-regret algorithms* or *expert-based sequential prediction* [16]. In a nutshell, to decide whether to accept a SU or not, these algorithms compare suggestions from several arbitrarily complex decision rules called *experts*, and combine them to provide similar results to those of the best available expert in hindsight.

We discuss the particularities of the spectrum broker's problem in this context, and propose adaptations to the base algorithms to consider them. For instance, the fact that a SU has to be accepted in order to verify if it will eventually constitute an earning (i.e. not indemnified due to service suspension) constitutes a variation of the so-called Multi-Armed Bandit problem.

As we show through extensive simulations, our proposal is capable of managing essentially arbitrary behaviors for both kind of users and obtain nearly optimal results. As a practical benchmark, we compare our results with the ones achieved by *Dynamic Programming Algorithms* (DPA), which are known to be optimal for online *Markov Decision Processes* (MDPs), and extensively used for this type of sequential decision-making problems. Our results show that if the behavior does not comply to a markovian setting (an expected behavior), our algorithm largely outperforms DPA (and in addition is much simpler computationally).

The rest of this article is organized as follows. On section II, after reviewing key aspects of cognitive radio, we compare different technological and economic models present in the literature geared towards more efficient sharing of spectrum resources. Arguments in favor of a centralized, real-time and on-demand secondary market are discussed in that section. Section III provides a deeper description of *no-regret expert-based algorithms* for robust predictions. Then, it proposes a model of the *spectrum broker* problem and our solution. Why compensations are necessary in order to provide proper incentives to SUs is discussed at the end of this section. Afterwards section IV describes the methodology to evaluate the performance of each algorithm and then discusses the obtained results. Finally, section V presents the conclusions of the work.

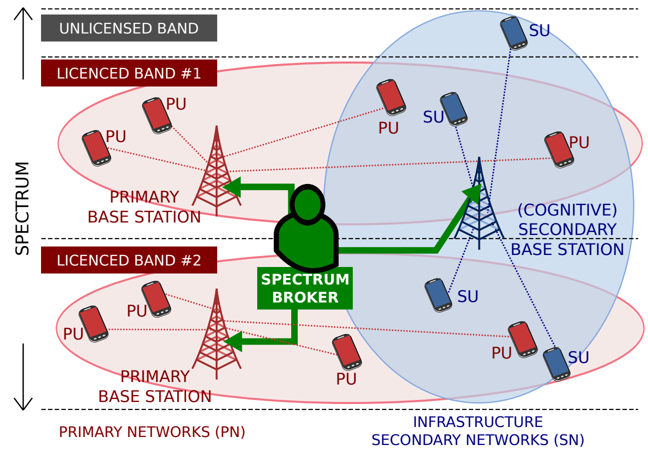

Fig. 1. CR Infrastructure Network Architecture. Red elements belong to primary networks, blue elements to secondary networks. Ovals represent network span.

II. Cognitive Radio

A. Technological Context

Cognitive Radio (CR) are wireless communications systems capable of adapting their operating parameters in real time to exploit vacant spectrum resources [39]. A spectrum resource is any portion of the spectrum which could potentially be used as a communication channel. As such, it is composed by several dimensions [33] like central frequency, channel bandwidth, transmission power, codification/modulation scheme, beam directionality, or polarity among others.

In the most general case, a CR node should implement the spectrum sensing, analysis, decision, mobility and sharing functionalities [6], [7], [50]. However, in *centralized* (and thus also *cooperative*) schemes a so-called spectrum broker exists, which decides on resource allocation and coordinates medium access procedures [42], [46], [50]. In infrastructure networks the central base station entity may play the role of the spectrum broker. In ad-hoc networks, the broker role could be performed by any SU, group of, or even some distributed protocol like in [9].

Several advantages stem from cooperation among both types of users, where the spectrum broker connects to primary and secondary networks. Firstly, the broker gets notified of all communication activities and available resources over the networks' areas. With this information at hand the broker can guarantee an operation without mutual interference simply allocating each user to a portion of the multidimensional spectrum that do not interfere with anyone else.

Secondly, in such architecture the SUs mechanism of performing distributed spectrum sensing and sending the results to the central entity becomes unnecessary. Relevant information can be directly provided by the broker through a control channel. This simplifies SUs cognitive functions and receiver implementations as well as their cost, while efficiently using the spectrum.

Several works agree that such an architecture (depicted in Fig. 1 for the case of an infrastructure secondary network), is indeed beneficial for SUs and can achieve better spectrum utilization. For instance, the authors of [9] propose a protocol over a virtual channel for coordination of simultaneous access between SUs and PUs, whereas a common signaling channel with the participation of explicit spectrum brokerage to coordinate access is proposed in [15]. Similarly, [14] uses a spectrum broker figure to centralize coordination, devices' management and access of cooperative users, storing and distributing allocation information on a dynamic database. More recently, [24] proposes a TV white spaces CR system in which a broker has access to a shared database specifying incumbent licensees (PUs), available frequency bands and maximum transmit powers at the broker geolocation. SUs can use resources marked as available in the database. A similar approach is also taken by both IEEE 802.22 [40] and IEEE802.11af [27] standards. A GPS-assisted geolocation is used by the base transceivers (BTS) to access a centralized database providing temporarily local available TV bands, location of primary stations and maximum allowable transmission power.

Because of its significant benefits, we will consider this kind of architecture for the rest of this work.

B. Economic Context

We now discuss the economic context under which CR may be implemented. The simplest form is one where SUs would attempt to access the spectrum resources opportunistically at zero-cost (the so-called *free-sharing* scheme) [38]. However, no incentives are provided to PUs to cooperate in the allocation, and it moreover leads to complex (and costly) technological SU devices as discussed in the previous section.

The alternative is a *paid-sharing* model [38] in which SUs pay for resource usage. That is to say, the implementation of a spectrum secondary market. This approach provides economic incentives for incumbent PUs to permit their underutilized licensed resources to be shared, as long as their usage priority right is honored at all times (which may be achieved by means of the spectrum broker as we also discussed in the previous section).

Taking into account the benefits of paid-sharing models, many works and government agencies propose different implementations of this spectrum secondary markets (SSM) [1], [4], [25], [37], which we now review.

Auctions are recommended by the literature as the method that provides greater potential benefits for all actors while increasing the spectrum utilization efficiency via reuse [17], [34], [37]. Several variants of auctions have been introduced specially for the *spectrum sharing* case [52]. For instance, the Federal Communications Commission (FCC) after trying several traditional approaches currently uses auctions for allocation of spectrum resources [20]. However, auctions have their problems. Existing efforts on dynamic spectrum auctions mostly strive at maximizing the auction's revenue. This cannot ensure that the bidder with the highest social value (reflecting the benefits of improved competition) wins as intended, but the one with the highest private value (the bid in the auction) [19], [20], [51].

This is not the case of on-demand allocation with a fixed price, where available resources are allocated to SUs that actually need them, thus being more likely to be the high social value bidders and to achieve an efficient allocation policy. Auctions can also be prone to manipulations, which might result in lower revenues for the auctioneer [22].

Moreover, auctions (as well as the more traditional contests or lotteries) require the gathering of interested candidates to work properly. This poses a limit on the flexibility of spectrum sharing, as SUs needing immediate access to resources might not be well served by this allocation scheme. For example, the closer the spectrum exchange is to being able to make spectrum trades in real time, the more attractive the system is for service providers acting as SUs needing to solve short term capacity shortages [10]. Such users or applications may be better served by an on-demand *real-time spectrum secondary market* (RTSSM), where SUs request immediate temporary access to spectrum as needed [43].

To compensate the drawbacks of real-time markets several measures could be taken. As an example, previous evaluations of SUs could be used in order to allow only properly qualified users to ask for service. Such a measure could also be useful to prevent ill-intentioned users to enter the system. Proper regulatory frameworks must prevent other undesirable economic effects such as the hoarding and intentional underutilization of the spectrum (i.e. *artificial spectrum scarcity*) [10].

Some works had studied this scenario, suggesting good results with fee-based pricing strategies in which the price paid by SUs is essentially fixed given the characteristics of resources [13], [23], [36], [43]. In particular, the authors of [13] argue that fixed-price policies tend to minimize spectrum fragmentation and claims that performance evaluation simulations verified the validity of the proposed architecture.

The discussion up to this point clearly justifies the choice of a centralized, real-time and on-demand secondary market. Regarding incentives for the SUs, a paid-sharing model and a dynamic market allows them to acquire spectrum when needed, and furthermore simplify their operation. However, an aspect not taken into account in all works cited so far, is that in order for the proposal to be attractive enough for SUs to pay for the resource, it is expected that they demand a compensation if their service is terminated due to the arrival of a PU. Actually, as we discuss at the end of the next section, this reimbursement implicitly considers the Quality of Service obtained by SUs.

The question remains then, and it is the subject of the rest of this article, on how may the broker decide whether to accept or not a new SU to the system in order to maximize its total payoff. To the best of our knowledge, only [48], [49] consider the same scenario as we do. However, they both assume a Markovian behavior of users. As we show in Sec. IV, designing the broker's decision algorithm assuming this to be true has an important negative effect when this is not the case.

Besides being robust to user's behaviors, the proposed decision algorithms are simple enough to be used online (real-time), and are scalable with respect to the system's capacity. These characteristics distinguish our work from that of [48], [49] that make use of dynamic programming which has computationally prohibitive costs. This aspect is further discussed in Sec. IV-D.

III. Online No-Regret Expert-Based Prediction

A. Basic Model

Having discussed the architectural and economical possibilities, and after justifying our choice, we now focus on the spectrum broker's problem: obtaining economical benefits from the system operation. To this end, let us now discuss the notation and main assumptions we will use in the rest of the article. Let x and y represent the number of PUs and SUs currently using the system. We will assume that the system's capacity is C, so that $x + y \leq C$. Moreover, accepted SUs will pay a fixed amount equal to R for the resource, and are compensated with an amount K in the event of a service interruption.

More complex situations, where for instance SUs and PUs are not equivalent in terms of the occupied resources, may be easily considered through simple modifications to the model. The tuple $(x, y) \in \mathcal{S} = \{(x, y) \in \mathbb{N}^2 : x + y \leq C\}$ represents the current state of the system.

Let us index with $\tau \in \mathbb{N}$ the moments when PUs or SUs either arrive of leave the system. At any given τ one of the following scenarios may happen:

1) A PU arrives to the system and there is enough remaining capacity (i.e. $x + y < C$), then $x \leftarrow x + 1$.
2) A PU arrives to the system when $x + y = C$ and $y > 0$, then a SU will be expelled due to the PUs priority. Thus, in this case $x \leftarrow x + 1$ and $y \leftarrow y - 1$, and the spectrum broker compensates the SU by paying her an amount equal to K.
3) A SU arrives to the system when $x+y < C$ and is granted access by the spectrum broker. Then, the SU pays R and $y \leftarrow y + 1$.
4) A SU arrives to the system when $x + y < C$ but the spectrum broker decides to refuse her access. Thus, the system state and total payoff remain unmodified.

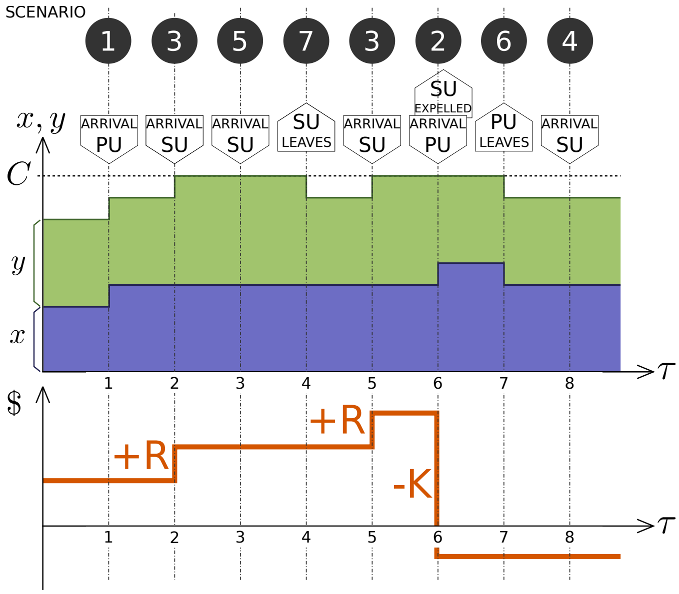

Fig. 2. Example sequence of basic events. At $\tau = 1$ a PU arrives to the system (scenario 1), and at $\tau = 6$ a PU arrives and causes a SU to be expelled (scenario 2). At $\tau = 2$ and $\tau = 5$ a SU arrives and is accepted and increases payoff (scenario 3), at $\tau = 8$ the SU is rejected (scenario 4).

5) A SU attempts to arrives when the system is full ($x + y = C$) so she cannot be accepted. No decision is necessary, and the system state remains unaltered.
6) A PU leaves, then $x \leftarrow x - 1$.
7) A SU leaves, then $y \leftarrow y - 1$.

Figure 2 provides an illustrative example of the evolution of the system. The upper graph shows the state of the system by stacking the number of PUs and SUs on the system, i.e. how many PUs (x, in blue at the bottom of the stack) and SUs (y, green at the top of the stack) are currently being served for any given time τ. The lower graph shows the evolution of the total payoff. On top of the upper graph there is an indication of which events are happening and which scenario number describes them (black circles with light numbers).

The figure shows that the spectrum broker's problem is an online decision problem. At each time τ which includes a SU arrival (which we will call *round* and denote as t), the broker has to decide whether to accept or reject (allocate resources or not) the incoming SU. The question becomes which sequence of decisions should be taken in order to maximize earnings (and thus incentivize the PUs' participation). This work's objective is to design a simple and robust decision-making algorithm that achieves this.

B. Online Non-Regret Expert-Based Prediction

The optimal decision for the broker will naturally depend on the system's current state, its parameters (i.e. capacity but also the values of K and R) and on the arrival and service processes for each kind of user. One may assume that those processes are stochastic with known distribution families. Nevertheless that is a strong hypothesis that might not be verified in practice. To account for this fact this work states the problem in the framework of Online Non-Regret Expert-Based Prediction (ONREBP) [16].

Instead of making any hypothesis regarding users processes, let us assume that on each round t the broker has access to advices from N external forecasters (called *experts*) and their past results. These experts are essentially arbitrary complex black boxes that at each round generate a recommendation to accept the SU or reject it. The broker then makes its own decision. If accepted, eventually the fate of the SU is revealed (she either completed her service or got expelled).

The objective of ONREBP algorithms is to obtain results similar to those of the best available expert (naturally, without previously knowing which this expert is). In other words, to minimize the regret of not having followed its advice.

Let us formalize this notion. Let $f_{i,t}$ ($i = 1, \ldots, N$) be the i-th expert's advice on round t. We will arbitrarily code $f_{i,t} = 0$ if the recommendation is to reject the SU, and $f_{i,t} = 1$ if it is to accept her. The decision by the broker on round t will be denoted by $\hat{p}_t$ and let u_t represent the outcome of the service provided to the SU; if it completed its service then $u_t = 1$, and $u_t = 0$ if the SU was expelled. Also, let $h(d_t, u_t)$ be the payoff generated by decision $d_t \in \{0, 1\}$ (d_t will be evaluated as $\hat{p}_t$ or as $f_{i,t}$ depending on the case) and SU outcome u_t.

The following cases are possible:

1) The SU is accepted ($\hat{p}_t = 1$), pays R, uses the allocated resource for an arbitrary period and then leaves the system. The broker receives a *payoff* $h(\hat{p}_t, u_t) = h(1, 1) = R$. Experts which recommended to accept will also receive payoff $h(f_{i,t}, u_t) = h(1, 1) = R$, otherwise they will receive payoff $h(f_{i,t}, u_t) = h(0, 1) = 0$ (since no gain nor loss would have been obtained if their advice was followed).
2) The SU is accepted and pays R. Eventually, the spectrum broker expels the SU from the system to ensure PU privilege, indemnifying her with an amount K. The broker receives a payoff $h(1, 0) = R - K$ as do experts which recommended to accept. As in the previous case, the other experts receive a payoff $h(0, 0) = 0$.
3) The SU is rejected ($\hat{p}_t = 0$). Since this decision does not generate loss nor gain (i.e. $h(0, 1) = h(0, 0) = 0$), the payoff of the broker and of the experts that recommended to reject is left unchanged. However, the outcome u_t of this particular SU, would it have been accepted, cannot be known. So neither can the instantaneous payoff of the experts that recommended to accept. We will handle this specific issue in following sections.

Note that the payoffs corresponding to a particular accepted SU are definitely fixed once she quits the system (either because she finished or she was expelled). This topic will also be further discussed in the following sections.

Let H_n and $H_{i,n}$ be the broker's and the i-th expert's accumulated payoff or simply payoff up to time n respectively. Then,

$$H_n = \sum_{t=1}^{n} h(\hat{p}_t, u_t), \tag{1}$$

$$H_{i,n} = \sum_{t=1}^{n} h(f_{i,t}, u_t). \tag{2}$$

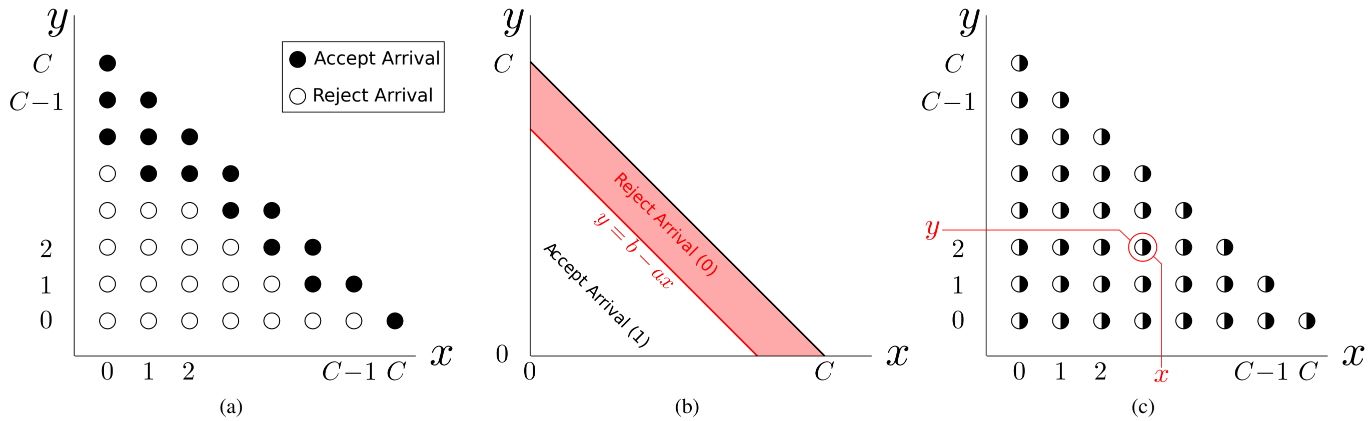

Fig. 3. Types of Experts: (a) Action Map Experts - AME, (b) Linear Boundary Experts - LBE, (c) Class Based Experts - CBE

Let the *regret* of the spectrum broker with respect to the i-th expert be the difference between their cumulative payoffs up to round n [16]:

$$R_{i,n} = H_{i,n} - H_n = \sum_{t=1}^{n} h(f_{i,t}, u_t) - h(\hat{p}_t, u_t). \tag{3}$$

Given any set of experts, the goal of an ONREBP algorithm is to minimize the cumulative regret R_n generated by not following the advice of the best available expert for any sequence of outcomes:

$$R_n = \max_{i=1,\ldots,N} H_{i,n} - H_n = \max_{i=1,\ldots,N} \sum_{t=1}^{n} h(f_{i,t}, u_t) - \sum_{t=1}^{n} h(\hat{p}_t, u_t). \tag{4}$$

Formally, an algorithm is said to be *Hannan Consistent* or to have *No Regret Property* with respect to the whole set of available experts if for any sequence $u_1, u_2, \ldots, u_n$ of n outcomes the following inequality holds:

$$\lim_{n\to\infty} \frac{R_n}{n} \leq 0 \text{ a.s.} \,\forall\, u_1, u_2, \ldots, u_n, \tag{5}$$

where the almost surely convergence is taken with respect to any possible randomization that the algorithm uses [30].

Then, the no-regret property of ONREBP algorithms can be interpreted as having a per round regret that vanishes with time, meaning the cumulative payoff difference between the best available expert in hindsight and the broker is sublinear with the number of rounds. The result will be that the broker's payoff will be close to that of the best expert in hindsight, a desirable property for a decision-making rule.

Before discussing how this may be achieved, let us present the experts we will consider in this work.

C. Experts

An expert could be any arbitrarily complex black box that at each round generates a recommendation to accept or to reject the SU. Experts might have access to information unavailable to the spectrum broker. They could also be known solutions for certain conditions (e.g. for certain stochastic cases).

We consider three types of simple decision rules as experts. The idea is to compare the performance of the different families of experts to decide on the one that consistently achieves the best results.

1) Action Map Experts (AME): Each expert is an explicit mapping of each possible system state (x, y) to a decision; i.e. $f_i(x, y) : \mathcal{S} \rightarrow \mathcal{D} = \{0, 1\}$. An example mapping is shown in Fig. 3a. The rationale is that SU's success may depend on the system's state at arrival, but also on the primary's behavior. A SU accepted when the PUs arrive in bursts might be likely to be expelled from the system before finishing her session, even if the system was largely underutilized when it was accepted. In this case the optimal policy would be to *reject all arrivals*. The opposite happens for situations where few PUs arrive, for which an *accept all arrivals* policy would be optimal. According to these observations, experts in this case are chosen to cover a wide range of mappings from all rejection ($f_i(x, y) = 0 \,\forall x, y$) to all acceptance ($f_i(x, y) = 1 \,\forall x, y$) of arrivals. To be coherent, experts must comply $f_i(x, y) \geq f_i(x + \delta_x, y + \delta_y) \,\forall x, y$ and $\delta_x, \delta_y \in \{0, 1\}$.

2) Linear Boundary Experts (LBE): This expert can be seen as a particular case of the previous one. In this case, expert i is represented by two real numbers (a_i, b_i) with $0 < a_i \leq C$ and $0 < b_i \leq C$ corresponding to the parameters of a line segment according to the following decision rule:

$$f_i(x, y) = \begin{cases} 1 & \text{if } b_i - a_i x \geq y, \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

According to [47], [49] under markovian assumptions on the user processes of arrival and sojourn, the broker's problem has an optimal decision boundary that is approximately linear. An LBE expert represents a simple possible solution for such a condition. Experts in this class are chosen to cover the whole state space similarly to AME experts as shown in figure 3b.

3) Class Based Experts (CBE): Instead of having a single set of experts providing a decision for each possible state of the system, the broker may choose a different set of experts for each state of the system. We will consider a simple case, where each set is composed by only two experts: *always accept* or *always reject*, as depicted by figure 3c. Regret comparisons has to be done on a per-state basis only.

D. Particularities of the Broker's Problem

This section considers significant and particular aspects of the spectrum broker's online decision problem. As stated, the system can be thought of as a game between the broker and the *market* (or the *environment*). CR secondary markets are composed of many disparate and legitimate users that only care about their own service requirements. As they act with *good faith* they do not intend to act against the spectrum broker. This condition might be met at an administrative level through screening of SU candidates. The sequence of *intended* SU's arrival and service times is then independent of the broker's actions.

Consequently, the model of a state-aware *oblivious opponent* [16] appears as the most appropriate for the market. Thus, the intended sequence of PUs and SUs arrivals and services could as well be computed in advance, at least for simulating the system.

Note that in our case the outcome (the market action) is known only after an arbitrary delay (when the SU either finishes or is expelled). This is a deviation from usual game theory hypothesis. Following theoretical work in [35], in this paper all tested algorithms decisions are taken considering only *revealed outcomes* up to that time. In practice the effect is similar to having taken the decision later in time (just before the outcome was known), yet possibly with slight less information.

Moreover, and as noted before, if the spectrum broker rejects a SU arrival, then it becomes impossible to know the payoff of the experts that recommended to accept it and regret cannot be calculated. This is known as a *Partial Information* [16] scenario, in which the decision-taking algorithm has to compensate for the missing information.

Nevertheless if the broker chooses to accept the SU the payoff of *all* experts will be known. It is then said that $\hat{p}_t = 1$ is a *Revealing Action*. Essentially, the *spectrum broker* will have to find a proper balance between taking the decision that provided the best payoff in the past (*exploitation*) and choosing the revealing action to get information about all experts (*exploration*). It can be shown that if a revealing action exists for a given problem, then a no-regret predictor also exists (see chapter 6 of [16]). Basically the decision-taking algorithm must ensure that the revealing action is taken at least a fraction of the times to ensure proper exploration of the environment. For examples of revealing-action algorithms for partial information settings refer to chapter 6 of [16].

E. Prediction Techniques

Let us now discuss how the broker may combine the experts' advice in order to obtain a total payoff close to the best expert. In order to limit the broker's regret, a simple algorithm would be to chose an expert at random at every round. We might start using the same probability of choosing each expert (maximum entropy), reflecting our ignorance. As decisions are taken, some experts might achieve greater payoffs than others, and it seems convenient to follow the former's advice with a higher probability. That is, to lower the distribution's entropy to reflect our increased knowledge [8].

Most no-regret algorithms roughly work as described above. Others combine the predictions with weights according to each expert's regret. Table I summarizes all tested prediction techniques. More details of the techniques we used follow.

1) FTPL - Follow the Perturbed Leader: An algorithm introduced by James Hannan in a seminal paper in 1956 [30]. It states that the spectrum broker must follow the advice of the expert whose regret plus a random signal is the greatest up to the last round. That is to say:

$$\hat{p}_t = f_{I_t,t}, \text{ where } I_t = \underset{i=1,..,N}{\arg max} \left(Z_{i,t}^{(\eta)} + \sum_{s=1}^{t-1} h(i, u_s) \right) \quad (7)$$

with random variable $Z_{i,t}^{(\eta)} \sim Laplace\left(0, \frac{1}{\eta}\right)$ acting as a perturbation. The distribution of $Z_{i,t}^{(\eta)}$ is chosen as discussed in [16].

The logic behind following the largest regret is deterministic, simple and intuitive. Nevertheless, if no perturbation is added there will always be a sequence of outcomes such that the broker's regret is not sublinear. The added noise introduces a random perturbation that allows the algorithm to escape sequences that otherwise would lead to suboptimal decisions. Then it can be proved that FTPL is a no regret algorithm [30].

2) EPRF - Exponential Potential Random Forecaster: A slightly different approach is to use explicit probabilities to choose the expert at random at every round. Intuitively the experts that achieve greater payoffs (greater regret) should have a higher probability and consequently be chosen more often than the rest. A simple way to do this is by using the *Hedge Algorithm* by Freund and Schapire [8], [16], [28]. It consists of updating the probability $p_{i,t}$ of choosing expert i ($\forall i$) at round t according to a weighted exponential function of its observed payoff up to round $t-1$. That is to say, at round t, the *broker* chooses expert i according to distribution:

$$p_{i,t} = \frac{e^{\eta H_{i,t-1}}}{\sum_{j=1}^{N} e^{\eta H_{j,t-1}}} \ \forall i \in \{1, \ldots, N\}, \eta > 0, \quad (8)$$

where η acts as a learning rate parameter. This algorithm has achieved success on several cases, see for example [18] [45] [12] [21].

The rationale behind this rule is that the larger an expert's payoff (or equivalently its regret) with respect to other experts is, the larger its probability of being chosen. Also all experts will always have a non-zero probability of being chosen, which can be useful for exploration. Thus, the balance between exploitation and exploration is tuned according to the past performance of each expert. Experts showing consistently good performance will be chosen most of the times but if their relative performance drops (i.e., some other experts starts to obtain greater payoffs) then the algorithm will adjust and become more exploratory.

Another interesting point about the *Hedge Algorithm* is that it makes no assumption about the sequence of events. This fact added to the algorithm's rationale suggests that this algorithm is a no-regret forecasting strategy. Proof is given in [16] chapters 2 (directly proving bounds on the algorithm's regret using potential functions) and 7 (under the framework of Blackwell's approachability [5], [11]) and also in [8].

3) EWAF - Exponential Weight Average Forecaster: With this algorithm, instead of choosing an expert for each round t and following its advice, the broker receives advice $\{f_{\cdot,t}\}$ from each expert and combines them to get a new advice. The recommendations of the experts are averaged using weights proportional to the exponential of their payoffs to get the broker's decision as follows:

$$\hat{p}_t = \frac{\sum_{i=1}^{N} e^{\eta H_{i,t-1}} f_{i,t}}{\sum_{j=1}^{N} e^{\eta H_{j,t-1}}}, \tag{9}$$

where η acts as a learning rate parameter. As expert recommendations are always in the $\{0,1\}$ domain, the prediction will belong to the continuous domain $[0,1]$ and can be interpreted as a score or confidence level on which the SU should be accepted. A decision is taken using a threshold at 0.5.

According to [32], online convex optimization methods can be used to prove that this algorithm obtains regret rates of order $O(\sqrt{n})$ for payoff functions that are linear on their first argument (the prediction) and convex domain.

4) Main Decision Algorithm: The general algorithm to be used by the spectrum broker is described in Algorithm 1.

Algorithm 1 Spectrum Broker Main Algorithm

Require:

- System Capacity C (determines set $\mathcal{S}$ of system states).
- Resource access price R.
- Expulsion compensation K.
- Set $\mathcal{E}$ of experts with $N = |\mathcal{E}|$ (LBE,AME) or a set of two experts $\mathcal{E}_{(x,y)} = \{\text{reject (0), accept (1)}\}$ for each state (x, y) (CBE).

For each round $t = 1, 2, \ldots$

(1) The *spectrum broker* gets to know her payoff H_t and the payoffs revealed for the experts $H_{i,t}$ up to time t.
(2) Current system state $(x(t), y(t))$ is revealed.
(3)
- If using CBE experts, the *broker* chooses an action according to the payoffs of the state experts.
- Else, each expert $i = \{1, \ldots, N\}$ makes its advice $f_{i,t} \in \{\text{reject(0) , accept(1)}\}$ and tells the *broker*.

(4) The *broker* takes its decision $\hat{p}_t \in \{0,1\}$ according to the chosen forecasting technique.
- If $\hat{p}_t = 0$ she observes $h(0, u_t) = 0$ immediately. The (estimated) payoff for the experts is also zero (payoffs remain unchanged).
- If $\hat{p}_t = 1$ (*revealing action*), both the *broker* and the experts will have to wait until the start of some later round to know their respective payoffs.

F. On the Quality of Service as perceived by SUs

We have so far focused on maximizing the total payoff of the broker. However, a pertinent question is how will this policy impact on the quality of the service (QoS) provided to the SUs. Before presenting the simulation results, we will briefly discuss this important aspect.

Given that each SU will use the spectrum as they see fit, the only two indicators that may be considered by the spectrum broker regarding the SUs service are the access and completion probability. That is to say, what proportion of all incoming SUs are accepted to the system, and of these, what proportion are not expelled (denoted as P_{accept} and $1-P_{\text{expel}}$ respectively). Let us then write the spectrum broker's problem (over a certain number of rounds n) in terms of these two probabilities:

$$\begin{aligned}
&\underset{\{\hat{p}_t\}_{t=1,\ldots,n}}{\arg\max}\; R\sum_{t=1}^{n}\mathbb{1}(\hat{p}_t = 1) - K\sum_{t=1}^{n}\mathbb{1}(\hat{p}_t = 1 \cap u_t = 0) = \\
&\underset{\{\hat{p}_t\}_{t=1,\ldots,n}}{\arg\max}\; \frac{1}{n}\Big(R\,\#\{\text{accepted SUs}\} - K\,\#\{\text{expelled SUs}\}\Big) = \\
&\underset{\{\hat{p}_t\}_{t=1,\ldots,n}}{\arg\max}\; RP_{\text{accept}} - \frac{K}{n}\frac{\#\{\text{accepted SUs}\}}{\#\{\text{accepted SUs}\}}\#\{\text{expelled SUs}\} = \\
&\underset{\{\hat{p}_t\}_{t=1,\ldots,n}}{\arg\max}\; RP_{\text{accept}} - KP_{\text{accept}}P_{\text{expel}} = \\
&\underset{\{\hat{p}_t\}_{t=1,\ldots,n}}{\arg\max}\; P_{\text{accept}}\left(1 - \frac{K}{R}P_{\text{expel}}\right).
\end{aligned}$$

The above equality means that both QoS indicators are taken into account implicitly on the broker's problem: the broker is actually maximizing the multiplication of both indicators, where the ratio between K and R dictates the relative importance of P_{expel}. For instance, these two pricing parameters may be thus adjusted by the broker depending on the SUs preference.

Further considerations on these aspects are left for future work. As we mention on the conclusions section, we intend to incorporate new pricing schemes and market dynamics on the next stages of research, allowing for a more direct measurement of the quality of the provided service.

IV. Simulation Results and Analysis

A. Methodology

The objective is to evaluate the performance achieved by the different combinations (*mechanisms*) of prediction techniques and expert families. In order to do this, the main measurement used is the *payoff per round*

$$m = \frac{H_n}{n}. \tag{10}$$

Where n is the total number of rounds and H_n is the payoff from eq. (1) up to round n. As the payoff is measured relative to the total number of rounds, the results from different processes and run times can be compared. It also has the nice property of being bounded between $R - K$ and R.

Notice that it would be interesting to directly measure the *regret per round*, $\frac{R_n}{n} = \max_{i=1,\ldots,N}\frac{H_{i,n}}{n} - m$ as we seek methods that make that value vanish with time. Unfortunately, that is impossible as the true accumulated payoff $H_{i,n}$ of each expert cannot be known if the revealing action (accept) is not taken by the broker. This fact leaves us with m as the best *observable* metric, and also a direct measure of the broker's earnings, i.e., the ultimate interest of the broker. Moreover, and as discussed on section III-F, m also serves as an indirect measure of SU's level of satisfaction.

TABLE I
IDS FOR COMBINATIONS OF PREDICTIVE ALGORITHM AND EXPERT TYPE

PRED ↓ ALG →	**Action Map (AME)**	**Line Segment (LBE)**	**State Based (CBE)**
Follow-the-Perturbed-Leader (FTPL)	AME-FTPL	LBE-FTPL	CBE-FTPL
Exponential Potential Random Forecaster (EPRF)	AME-EPRF	LBE-EPRF	CBE-EPRF
Exponential Weight Average Forecaster (EWAF)	AME-EWAF	LBE-EWAF	

As we intend to find simple an fast methods, another important metric for this work is the computational complexity of the technique. This is directly calculated on section IV-D.

B. System Dynamics

The system dynamic can be roughly split in three cases:

- **Saturated System**. When PUs arrive in large bursts or when most of the time most system's resources are allocated to PUs, accepted SU are likely to get expelled often. In this case a *reject all SUs* policy leads to a (likely) optimal result of $m \approx 0$. Every expert family includes an expert suggesting *reject all SUs*.
- **Idle System**. The opposite happens when most resources are generally available, making it unlikely for accepted SUs to be expelled. On this scenario an *accept all SUs* policy leads to maximal earnings, $m \rightarrow R$. Every expert family includes an expert suggesting an *accept all SUs*.
- **Loaded System**. Finally when most of the time $x + y \approx C$ and x and y are not very different (i.e., less than a factor of 5) then the SUs might (or not) have a successful and lasting operation. This is an interesting dynamic as there is no self-evident optimal policy so we will focus in this case.

C. General Tests

Unless stated otherwise, all tests were run according to:

- $C = 20$
- $K = 3$
- $R = 1$
- For each system dynamic, we generate several sets of the users' random processes parameters and instantiate ten different simulations per set.
- After every simulation a result $m(\omega)$ is obtained for each mechanism ω. Being interested in identifying methods that provide good performance under worst case scenarios, results of different simulations for the same process' set of parameters are summarized with its minimum value.
- For all usages of AME and LBE algorithms the experts were chosen from five different slopes and eight different intersection points (when seen as line segments).

As shown in equations 7, 8 and 9, the considered forecasting techniques have each a single parameter (η), although with a different meaning: perturbation intensity for FTPL, a learning rate parameter for ERPF and EWAF. Due to space constrains we cannot provide a complete discussion about the way in which η was determined in each case so we briefly mention the main lines.

The objective when choosing a value for η is to provide the predictive technique with a value that achieves a robust performance in a wide range of scenarios, but such that said performance does not depend greatly in the specific value –algorithm performance should be relatively insensitive to the value of η. To determine the best value for η for each forecasting technique we used a set of simulations with Poisson processes. In particular we focused on loaded systems, where the arrival rate of PUs and SUs are uniformly chosen from the interval $(C/10, 9C/10)$ and $(C/10, 1.5C)$ respectively, and the departure rates of PUs in the interval $(0.25, 4)$, whereas the departure rate of SUs is fixed at 1.

We used an exploratory approach by trying different values of η, uniformly spaced for mechanisms with $[0, 1]$ domains and logarithmically spaced for $[0, +\infty)$. Then we filtered the possible values for η for each technique to the ranges that exhibited low sensitivity to variation across the different tests. In the last step of estimating the best value for η we chose values that showed both high values for the median and minimum of m. This method allowed us to pick values for η in each forecasting technique and expect a robust performance of them.

Finally, to select the most promising mechanisms (i.e. forecasting technique and set of experts), we considered several aspects. On the one hand, we prefer those mechanisms that achieved the higher values for the minimum of m on the general tests. Nevertheless, to account for the (likely) bias introduced by using Poisson processes, we also considered a *diversity criteria* – the final set of candidates mechanisms must have at least one mechanism from each forecasting technique and one mechanism from each expert family.

This lead us to chose the following set of candidates: AME-EWAF, CBE-ERPF and LBE-FTPL (cf. Table I).

D. Performance against Dynamic Programming

The first test consisted in evaluating the performance of the selected mechanisms on a Poisson arrival and departures scenario. To benchmark the mechanisms their performances are compared against the static policy obtained by a particular DPA called *Modified Policy Iterator Algorithm (MPIA)* [44], which is known to provide an optimal policy for the considered scenario.

The simulations provided very close results between MPIA and the expert-based mechanisms for idle and saturated systems. Figure 4 shows the minimum m achieved for each mechanism and MPIA after 10 simulations for 50 different experiments (i.e. set of parameters for the arrival and departure processes) corresponding to loaded systems. In particular, we plot as a point each experiment's result corresponding to

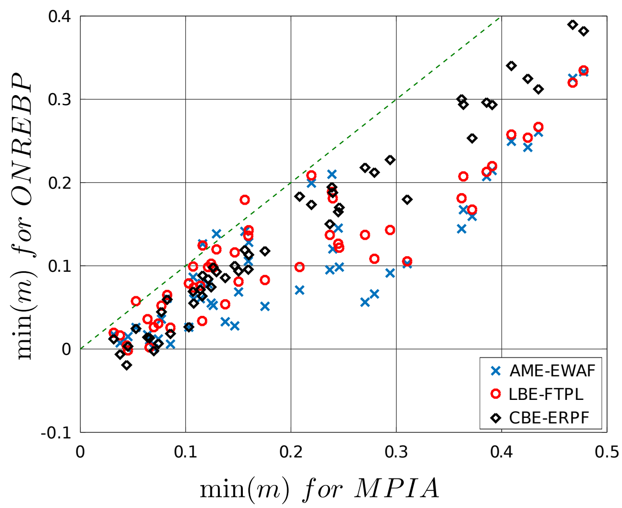

Fig. 4. ONREBP decision algorithms against DPA.

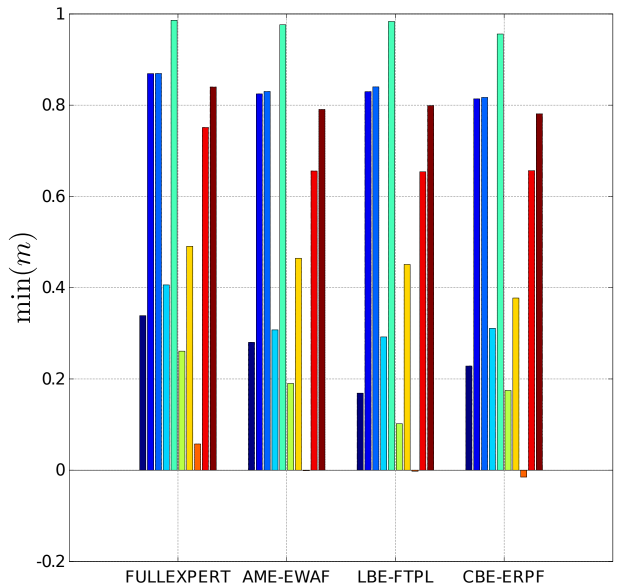

Fig. 5. Decision algorithms on large capacity systems.

MPIA (in the abscissa) against those obtained by the proposed mechanisms. Arrival and departure rates where chosen as in the previous subsection.

Note that in this case MPIA is optimal in the long-run, so almost all points are below the $y = x$ line (shown as a dashed line in the graph), and those that are not are very near it. However, it can be seen that expert-based mechanisms almost always achieved positive payoffs. Moreover, MPIA's optimal policy and expert-based mechanisms results are similar in the sense that when MPIA's policy got a high payoff so did the experts mechanisms. Finally, among the proposed algorithms, CBE-ERPF is the one that performs best, with results relatively similar to MPIA in all experiments.

Regarding complexity, we equate it to the number of operations the algorithm requires to operate. Table 8.7.2 of [44] shows that in order to compute the optimal policy, the operations per iteration of MPIA (ζ_{MPIA}) is proportional to $|\mathcal{S}|^2$, where $|\mathcal{S}|$ is the number of states (the policy is computed through several iterations). On the broker's problem $|\mathcal{S}| = \frac{(C+1)(C+2)}{2}$, so we can state that ζ_{MPIA} is $O(C^4)$.

On the other hand, expert-based mechanisms only perform operations during each decision round t. At this moments, the operations are over vectors of size N (the number of experts). Then the number of operations required by these mechanisms depends only on the number of experts N, and not on the number of states $|\mathcal{S}|$ nor the system capacity C. Even in the case of CBE, where $N = C$, the computations per round involve only two experts. This is a great scalability advantage, and proves that the expert-based mechanisms are conveniently simple algorithms even for cases were an optimal yet costly algorithm is known.

E. Large Capacities

According to the results of the previous section, one could argue that expert-based mechanisms provide a useful alternative to DPAs for Poisson arrivals and exponential service times when the capacity C of the system is large enough to make DPAs impractical or even unfeasible. To test this hypothesis, we consider a large system with $C = 50$. We tried MPIA on this scenario but it did not converge to a solution after running for several days, while the expert-based mechanisms run in matter of minutes. To have a reference against which to compare performance, we note as "FULL EXPERT" an AME decision algorithm considering all different possible experts from the same set of slopes as the algorithms under test.

Figure 5 shows the results achieved by each mechanism under test, for 10 different experiments of loaded systems (arrival and departure rates were chosen as before). The simulations are in the same order for each mechanism for ease of comparison. It can be seen that the expert-based algorithms achieved payoffs slightly lower than the "FULL EXPERT" case. On idle and saturated systems results were almost identical between all mechanisms, as expected from all of them having an expert suggesting the optimal decision. The tiny performance loss on those cases comes from the initial exploratory transient until the optimal policy is identified.

This shows that the expert-based approach proposed here can be an efficient alternative to computational costly DPAs.

F. Oblivious Opponents

Finally we drop Poisson-based users behavior and test the expert-based mechanisms under different processes complying with the *oblivious opponent* hypothesis. For instance, we focus on two different kinds of processes.

First, we considered a Cauchy heavy-tailed distribution for arrival and service intended times of both PUs and SUs. Then, we considered processes with two phases: one were the users arrive with Poisson distribution and serviced with exponential times (the *ON* phase) and one when users cease to arrive (the *OFF* phase). The switching times from one phase to another are randomly chosen as we discuss below. We call these

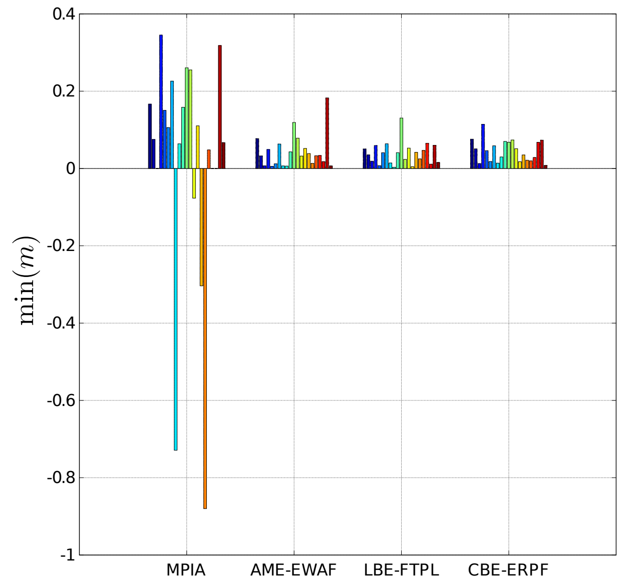

Fig. 6. Performance against DPAs with Heavy tailed Opponents.

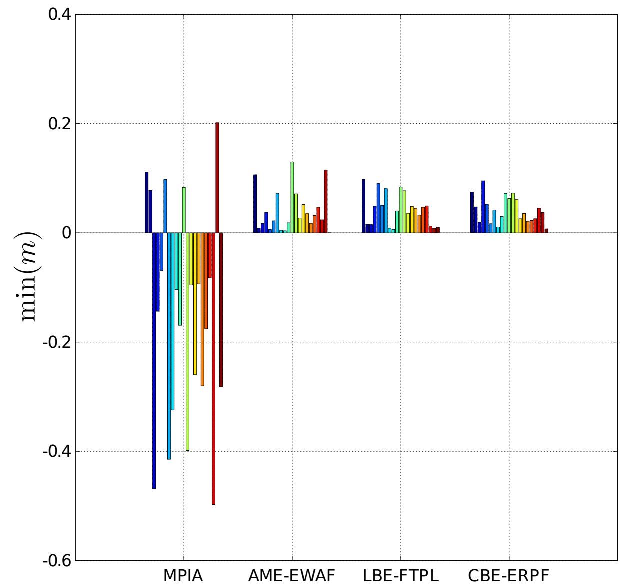

Fig. 7. Performance against DPAs with ON-OFF Opponents.

behavior an "ON-OFF Process" and use them independently for PUs and SUs.

Again the performance achieved by the algorithms under test is compared to the one achieved by MPIA. In order to apply a DPA algorithm, we proceed as the broker would: we estimate the involved parameters from the incoming traffic, using maximum-likelihood estimators. The DPA will then provide the optimal static policy that best adjusts to a Poisson/exponential approximation of the users' process.

1) Heavy Tail Processes: Figure 6 shows the results of the tests for heavy-tailed processes. We considered 21 different sets of parameters for simulation of users' processes, with 10 repetitions of each one and then the minimum achieved payoff is considered. On these set of simulations the system behaved between a loaded and a saturated system. In particular, the interarrival and service times were chosen as the absolute value of a random variable with a distribution $Cauchy(0, 1/\lambda)$ (where $1/\lambda$ is the so-called scale). Parameter λ is uniformly chosen in the intervals $(0.4C, 0.6C)$ and $(0.5C, C)$ for the arrival time of the PUs and SUs respectively, and the intervals $(1, 2)$ and $(0.1, 0.5)$ for the service times.

It is possible to see that while several times MPIA achieved greater payoffs than the expert-based mechanisms, it fails to be consistent. In at least 4 cases, the MPIA policy was off target and lead to losses, some of them considerable. In other cases, while not resulting in losses, MPIA obtained lower payoffs than the algorithms under test.

On the other hand, expert-based algorithms consistently achieved positive payoffs, as intended. This shows that all of them behave robustly to these kind of processes. Between them the higher median and the greatest minimum for m across simulations was achieved by CBE-ERPF.

For systems that behaved like saturated or idle systems, no clear difference could be found between MPIA and the algorithms under test.

2) ON-OFF Processes: The results are shown in Figure 7. Again 21 different simulation sets were considered, with 10 repetitions each one. The system behaved sometimes as a loaded system and sometimes as a saturated system. In particular, service and departure rates were uniformly chosen from the same intervals as in the Cauchy case. Moreover, in this case, the number of switches between on and off periods was uniformly chosen between 3 and 11. Switching times are then uniformly chosen in the simulation interval.

At first sight results are similar to the heavy-tailed scenario, except by the bad performance obtained by the MPIA algorithm most of the time. This shows that the MPIA is not a robust algorithm and suffered from the hypothesis deviation.

In contrast expert-based algorithms managed to obtain non-negative payoffs every time (with only one exception for AME-EWAF that incurred minimal loss on start), confirming their robustness. Although AME-EWAF obtained the highest values for $\min(m)$ occasionally, it was again CBE-ERPF that achieved the greatest minimum of minimums across simulations. This time the greatest median for minimum of m across simulations was for LBE-FTPL.

It can be concluded that all expert-based mechanisms are simple and accomplished robust performances against non-Poisson opponents. Between them, although having very close performances, CBE-ERPF showed perhaps the most consistent results by obtaining the highest minimums across simulations and competitive medians.

V. Conclusions

We studied the current problem of radioelectric spectrum usage and allocation. We proposed the usage of dynamic allocation policies under a secondary market system with fixed price and on demand-basis. A central entity called the *spectrum broker* is in charge of receiving arriving requests from secondary users and deciding whether or not to provide them access.

Such a system would be enabled by cognitive radio, a software controlled radio technology that allows dynamic coexistence between different users (for example with interference avoidance techniques) thus allowing the adoption of a dynamic market system.

We provided reasons and references showing that such a system allows the licensed users to sell or rent already allocated bands to secondary users with the highest possible dynamism and likely achieving high spectrum efficiency. We cited some works already making use of similar architectures.

The incentive for secondary users to actively cooperate with the broker when sharing the spectrum (as opposed to them using opportunistic access techniques) lies on simplified hardware and software (equating to lower costs) and also to the possibility of a compensation in case of service cancellation by the broker.

We propose several simple and robust practical mechanisms labeled as *ONREBP* to aid the broker decision-taking in an attempt to make a profit despite having to pay compensations. These mechanisms use the framework of expert-based prediction with different kinds of experts. They also use a variety of predictive techniques, from the fields of no-regret prediction and online-convex optimization.

After testing on simulations, we found the proposed tools to be effective independently of user's behavior, avoiding losses and also being able to make a profit most times. We compared their performance against a commonly used Dynamic Programming Algorithm (DPA) and found several advantages. First, ONREBP mechanisms have lower complexity and do not require previous knowledge nor estimations of the users arrival and services processes. This allow them to be used on scenarios were DPA becomes prohibitive. Secondly, the performance against Poisson/exponential processes were close to that of optimal DPA-derived rules, proving them as a lower cost alternative. Finally, they outperformed DPA on more general non-Poisson dynamics, in terms of achieving higher minimum cumulative payoffs and avoiding losses, thus providing the sought guarantees against worst case scenarios.

The results achieved by ONREBP mechanisms were very similar to each other. Nevertheless, algorithm CBE-ERPF exhibited the most consistent results by obtaining the highest minimums across simulations.

We believe that these results and techniques also apply to other economic scenarios involving prioritization privileges with reimbursement, admission control decisions and resource allocation. Finally, the next stage in our research line would be to incorporate to the problem different market dynamics such as dynamic pricing features and auctions. On the expert-based learning aspect of our research, we will study the use of dynamic experts that can themselves learn from the experience by implementing evolutionary algorithms or Q-Learning.

References

[1]

[2] The economics of spectrum management: A review. Technical report, Australian Communication and Media Authority (ACMA), 2007. Accesible at http://www.acma.gov.au/~/media/mediacomms/Research

[3] Broadband series: Exploring the value and economic valuation of spectrum. Technical report, ITU-D, Abril 2012.

[4] Telecommunications research and engineering at the communications technology laboratory: Meeting the nation's telecommunications needs. Technical report, National Academies of Sciences, Engineering, and Medicine., 2015.

[5] Report ITU-R SM.2012-5: Economic aspects of spectrum management. June 2016.

[6] Jacob Abernethy, Peter L Bartlett, and Elad Hazan. Blackwell approachability and no-regret learning are equivalent. In *COLT*, pages 27–46, 2011.

[7] Ian F Akyildiz, Won-Yeol Lee, Mehmet C Vuran, and Shantidev Mohanty. Next generation/dynamic spectrum access/cognitive radio wireless networks: a survey. *Computer networks*, 50(13):2127–2159, 2006.

[8] Ian F Akyildiz, Won-Yeol Lee, Mehmet C Vuran, and Shantidev Mohanty. A survey on spectrum management in cognitive radio networks. *IEEE Communications magazine*, 46(4):40–48, 2008.

[9] Sanjeev Arora, Elad Hazan, and Satyen Kale. The multiplicative weights update method: a meta-algorithm and applications. *Theory of Computing*, 8(1):121–164, 2012.

[10] J. Backens, C. Xin, and M. Song. A novel protocol for transparent and simultaneous spectrum access between the secondary user and the primary user in cognitive radio networks. *Computer Communications*, 69:98–106, 2015.

[11] Carlos E Caicedo Bastidas. Spectrum trading: Market-based architectures for dynamic radio frequency spectrum access. *Journal of Information Policy*, 3:485–500, 2013.

[12] David Blackwell. An analog of the minimax theorem for vector payoffs. *Pacific J. Math.*, 6(1):1–8, 1956. Accesible at http://projecteuclid.org/euclid.pjm/1103044235.

[13] Avrim Blum and Jason D Hartline. Near-optimal online auctions. In *Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 1156–1163. Society for Industrial and Applied Mathematics, 2005.

[14] Athina Bourdena, Evangelos Pallis, George Kormentzas, and George Mastorakis. A prototype cognitive radio architecture for tvws exploitation under the real time secondary spectrum market policy. *Physical Communication*, 10:159–168, 2014.

[15] Vladimir Brik, Eric Rozner, Suman Banerjee, and Paramvir Bahl. Dsap: a protocol for coordinated spectrum access. In *DySPAN 2005*, pages 611–614. IEEE, 2005.

[16] Nicolo Cesa-Bianchi and Gábor Lugosi. *Prediction, learning, and games*. Cambridge university press, 2006.

[17] Yanjiao Chen and Qian Zhang. *Dynamic spectrum auction in wireless communication*. Springer, 2015.

[18] William W Cohen, Robert E Schapire, and Yoram Singer. Learning to order things. In *Advances in Neural Information Processing Systems*, pages 451–457, 1998.

[19] Peter Cramton. Market design in energy and communications. 2015.

[20] Peter Cramton et al. Spectrum auctions. *Handbook of telecommunications economics*, 1:605–639, 2002.

[21] Neema Davis, Gaurav Raina, and Krishna Jagannathan. Taxi demand forecasting: A hedge based tessellation strategy for improved accuracy. *arXiv preprint arXiv:1805.06619*, 2018.

[22] Lara B Deek, Xia Zhou, Kevin C Almeroth, and Haitao Zheng. To preempt or not: Tackling bid and time-based cheating in online spectrum auctions. In *INFOCOM*, pages 2219–2227. Citeseer, 2011.

[23] Soumitra Dixit, Shalini Periyalwar, and Halim Yanikomeroglu. Secondary user access in lte architecture based on a base-station-centric framework with dynamic pricing. *IEEE Transactions on Vehicular Technology*, 62(1):284–296, 2013.

[24] D.S.A. Model rules and regulations for the use of television white spaces. Technical report, Dynamic Spectrum Alliance Limited, 21 St Thomas Street, Bristol BS1 6JS, United Kingdom, December 2017. Accesible at http://dynamicspectrumalliance.org/wp-content/uploads/2018/01/Model-Rules-and-Regulations-for-the-use-of-TVWS.pdf.

[25] FCC. Notice of proposed rule making and order. ET Docket 00-402, FCC, 2000.

[26] FCC. Report and order and second further notice of proposed rulemaking. Rulemaking 12-354, FCC, 2017.

[27] A. B. Flores, R. E. Guerra, E. W. Knightly, P. Ecclesine, and S. Pandey. Ieee 802.11af: a standard for tv white space spectrum sharing. *IEEE Communications Magazine*, 51(10):92–100, October 2013.

[28] Yoav Freund and Robert E Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 55(1):119–139, 1997.

[29] Sanford J Grossman and Joseph E Stiglitz. Information and competitive price systems. *The American Economic Review*, 66(2):246–253, 1976.
[30] James Hannan. Approximation to bayes risk in repeated play. *Contributions to the Theory of Games*, 3:97–139, 1957.
[31] Friedrich August Hayek. The use of knowledge in society. *The American economic review*, 35(4):519–530, 1945.
[32] Elad Hazan. The convex optimization approach to regret minimization. *Optimization for machine learning*, page 287, 2011.
[33] William D Horne. Adaptive spectrum access: Using the full spectrum space. In *Proc. Telecommunications Policy Research Conference (TPRC)*, 2003.
[34] Atif Jilani. Spectrum allocation methods: Studying allocation through auctions. *Journal of Economics, Business and Management*, 3(7), 2015.
[35] Pooria Joulani, András György, and Csaba Szepesvári. Online learning under delayed feedback. In *ICML (3)*, pages 1453–1461, 2013.
[36] Beatriz Lorenzo, Ivana Kovacevic, Francisco Javier González-Castaño, and Juan C Burguillo. Exploiting context-awareness for secure spectrum trading in multi-hop cognitive cellular networks. In *Globecom Workshops*, pages 1–7. IEEE, 2015.
[37] John W Mayo and Scott Wallsten. Secondary spectrum markets as complements to incentive auctions. *Georgetown Center for Business and Public Policy Economic Policy Vignette*, 2011.
[38] J. McMenamy, I. Macaluso, N. Marchetti, and L. Doyle. A methodology to help operators share the spectrum through an enhanced form of carrier aggregation. In *DYSPAN 2014*, pages 334–345, April 2014.
[39] Joseph Mitola III. *Cognitive radio*. PhD thesis, Royal Institute of Technology, 2000.
[40] Apurva N Mody and Gerald Chouinard. IEEE 802.22 wireless regional area networks. *Enabling Rural Broadband Wireless Access Using Cognitive Radio Technology*, June 2010. doc.: IEEE 802.22-10/0073r03.
[41] Dusit Niyato and Ekram Hossain. Spectrum trading in cognitive radio networks: A market-equilibrium-based approach. 15:71 – 80, 01 2009.
[42] OECD. Secondary markets for spectrum: Policy issues. *OECD Digital Economy Papers*, 95, 2005.
[43] Jon M Peha and Sooksan Panichpapiboon. Real-time secondary markets for spectrum. *Telecommunications Policy*, 28(7-8):603–618, 2004.
[44] Martin L Puterman. *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2005.
[45] Yuankai Qi, Shengping Zhang, Lei Qin, Hongxun Yao, Qingming Huang, Jongwoo Lim, and Ming-Hsuan Yang. Hedged deep tracking. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4303–4311, 2016.
[46] Chandrasekharan Raman, Roy D Yates, and Narayan B Mandayam. Scheduling variable rate links via a spectrum server. In *DySPAN 2005*, pages 110–118. IEEE, 2005.
[47] Claudina Rattaro and Pablo Belzarena. Cognitive radio networks: Analysis of a paid-sharing approach based on a fluid model. In *Proceedings of the 2016 Workshop on Fostering Latin-American Research in Data Communication Networks*, LANCOMM '16, pages 40–42. ACM, 2016. Accesible at http://doi.acm.org/2940116.2940120.
[48] Claudina Rattaro and Pablo Belzarena. Cognitive radio networks: Analysis of a paid-sharing approach based on admission control decisions. *Wireless Personal Communications*, 101(4):2053–2083, Aug 2018.
[49] A. Turhan, M. Alanyali, and D. Starobinski. Optimal admission control of secondary users in preemptive cognitive radio networks. In *WiOpt 2012*, pages 138–144. IEEE, May 2012.
[50] Beibei Wang and KJ Ray Liu. Advances in cognitive radio networks: A survey. *IEEE Journal of selected topics in signal processing*, 5(1):5–23, 2011.
[51] Guangen Wu, Pinyi Ren, and Qinghe Du. Dynamic spectrum auction with time optimization in cognitive radio networks. In *Vehicular Technology Conference (VTC Fall), 2012 IEEE*, pages 1–5. IEEE, 2012.
[52] Ning Zhang, Hao Liang, Nan Cheng, Yujie Tang, Jon W Mark, and Xuemin Sherman Shen. Dynamic spectrum access in multi-channel cognitive radio networks. *IEEE Journal on Selected Areas in Communications*, 32(11):2053–2064, 2014.

Juan Vanerio Juan Vanerio is a teaching assistant (grade 2/5) at the Engineering School of the Universidad de la República (Uruguay). was born in Montevideo, Uruguay in 1983. He received its degree in Telecommunication Engineering in 2007 and after some years of profesional activity at local companies and a ISP, he went to grad school at Universidad de la República and received his Masters degree in 2017 working on cognitive radio and no-regret algorithms. His current research interests are machine learning and no-regret algorithms, especially applied to networking security and anomaly detection.

Federico La Rocca Federico 'Larroca' La Rocca is an Assistant Professor at the Engineering School of the Universidad de la República (Uruguay). He was a research engineering (PostDoc) at Telecom ParisTech (ex ENST) during the first quarter of 2010, where he obtained his Ph.D. degree in Computer Science and Networking under the advisoring of Prof. Jean-Louis Rougier in December 2009. He received the degree in Telecommunication Engineering in 2006 from the Universidad de la República, with a thesis concerning the development of a propagation model and loss prediction software for cellular signals in suburban and urban environments. From 2004 to 2011 he held a teaching assistant position at the Universidad de la República.

His current research interests are related to the analysis and modeling of communication systems, and development on Software Defined Radio.