5

Dynamic Spectrum Access in Cognitive Radio: An MDP Approach

Juan J. Alcaraz, Mario Torrecillas-Rodríguez, Luis Pastor-González and Javier Vales-Alonso
Universidad Politécnica de Cartagena
Spain

1. Introduction

Cognitive radio refers to a set of technologies aiming to increase the efficiency in the use of the radio frequency (RF) spectrum. Wireless communication systems are offering increasing bandwidth to their users, therefore the spectrum demand is becoming higher. However, RF spectrum is scarce and operators gain access to it by a licensing scheme by which public administrations assign a frequency band to each operator. Currently, this allocation is static and inflexible in the sense that a licensed band can only be accessed by one operator and their clients (licensed users). However, it is a known fact that while some RF bands are heavily used at some locations and at particular times, many other bands remain largely underused FCC (2002). This is, in fact, a classical property of tele-traffic systems, *i.e.* traffic intensity is highly variable during a day. The consequence is a paradoxical situation: while the spectrum scarcity problem hinders the development of new wireless applications, there are large portions of unoccupied spectrum (*spectrum holes* or spectrum opportunities).

Cognitive radio provides the mechanisms allowing unlicensed (or secondary) users to access licensed RF bands by exploiting spectrum opportunities. Cognitive radio is based on software-defined radio, which refers to a wireless communication system that can dynamically adjust transmission parameters such as operating frequency, modulation scheme, protocol and so on. It is crucial that this opportunistic access is performed with the least possible impact on the service provided to licensed users. Therefore, cognitive users should implement algorithms to detect the spectrum use (*spectrum sensing*), identify the spectrum holes (*spectrum analysis*) and decide the best action based on this analysis (*decision making*). Once the decision is made, the cognitive user performs the *spectrum access* according to a medium access control (MAC) protocol facilitating the communication among unlicensed users with minimum collision with other licensed and unlicensed users.

Dynamic spectrum access (DSA) refers to the mechanism that manages the spectrum use in response to system changes (e.g. available channels, unlicensed user requests) according to certain objectives (e.g. maximize spectrum usage) and subject to some constraints (e.g. minimum blocking probability for licensed users). DSA can be implemented in a centralized or distributed fashion. In the former one, a central controller collects all the information required about current spectrum usage and the transmission requirements of secondary users in order to make the spectrum access decision, which is generally derived from the solution of some optimization problem. In distributed DSA unlicensed users make their own decisions autonomously, according to their local information. Compared to centralized DSA, this

scheme requires greater computational resources at the user terminal and generally does not achieve globally optimal solutions. On the other side, distributed schemes imply a smaller communication overhead.

MAC protocols for DSA can also include spectrum trading features. In situations of low spectrum usage, the licensed operator may decide to sell spectrum opportunities to unlicensed users. In order to do this in real-time, a protocol is required to support negotiations on access price, channel holding time, *etc*, between the spectrum owner and secondary users. There are several models for spectrum trading. In this work, we consider the bid-auction model, in which secondary users bid for the spectrum of a single spectrum owner.

This chapter addresses the design of DSA MAC protocols for centralized dynamic spectrum access. We explore the possibilities of a formal design based on a Markov decision process (MDP) formulation. We survey previous works on this issue and propose a design framework to balance the grade-of-service (e.g. blocking probability) of different user categories and the expected economic revenue. When two or more contrary objectives are balanced on an optimization problem, there is not an optimal solution, in the strict sense, but a Pareto front, defined as the set of values, for each individual objective, such that any objective can not be improved without worsening the others. In this work we study the Pareto front solutions for two possible access models. The first one consists of simply providing priority to the licensed users, and the second one is an auction-based model, where unlicensed users offer a bidding price for the spectrum opportunities. In the priority-based access, the centralized policy should balance the blocking probability of each class of users. In the auction-based access, the trade-off appears between the blocking probability of primary users and the expected revenue.

The content structure of the rest of this chapter is the following. Section 2 provides a brief introduction to Markov Decision Processes. Section 3 reviews previous works using the MDP approach in cognitive radio systems. Section 4 explains the system model and MDP formulation for both DSA procedures considered. Section 5 contains the performance analysis of each model based on numerical evaluations of practical examples. Section 6 summarizes the conclusions of this work.

2. Introduction to Markov Decision Processes

Markov Decision Processes (MDPs) are an application of a more general optimization technique known as dynamic programming (DP). The goal of DP is to find the optimal values of a variable when these values (decisions or actions) must be chosen in consecutive stages. The algorithms to solve DP problems rely on the principle of optimality, which states that in an optimal sequence of decisions, every subsequence must also be optimal. DP is generally applied in the framework of dynamical systems. Several basic concepts must be introduced to understand this framework:

- *State*: Is determined by the values of the variables that characterize the system.
- *Stage*: In a discrete-time dynamical system, a stage is a single step in the temporal advance of the process followed by the system. At each stage the system performs a transition from on state to an adjacent one. A process may consist of a finite or infinite number of stages.
- *Action*: At each state, there may be one or several variables whose value can be chosen in order to influence the transition performed at the present stage. The values selected constitute the action at this stage.

- *Cost*: Each pair state-action is associated to a return or outcome, which we will generally refer to as cost. Sometimes the outcome has a positive meaning and is considered a benefit. Additionally, we can compute the total outcome obtained in the whole process. Depending on how it is computed, this overall cost is referred to as total discounted cost or average cost, among others.
- *Policy*: A policy is a function that relates the states with the actions taken at each stage, for the whole duration of the process considered. An optimal policy is the one that attains the best overall cost for a given objective.

As can be anticipated from previous definitions, the goal of DP is to find the optimal policy for a given process. DP is, in fact, a decomposition strategy for complex optimization problems. In this case, the decomposition exploits the discrete-time structure of the policy.

Markov Decision Processes are the application of DP to systems described by controlled discrete-time Markov chains, that is, Markov chains whose transition probabilities are determined by a decision variable.

Let the integer k denote the k-th stage of an MDP. At a given stage, let i and u denote the state of the system and the action taken, respectively. The set of possible values of the state, the *state space*, is denoted by S, therefore $i \in S$. The control space U, is defined similarly. In general, at each state i only a subset of actions $U(i) \subseteq U$ is allowed. We restrict our attention to processes where both S, $U(i)$ and U are independent of k. In this case, the transition probability from state i to state j is denoted as $p_{ij}(u)$. A policy takes the form: $u = \mu(i)$, and because it does not depend on k it is said to be a *stationary* policy. It is said that a policy is admissible if $\mu(i) \in U(i)$ for $i \in S$. At each state i, the policy provides the probability distribution of next state as $p_{ij}(\mu(i))$, for $j \in S$.

The cost of each pair action-state is denoted by $g(i,u)$. Sometimes the costs are associated to transitions instead of states. Let $\tilde{g}(i,u,j)$ denote the transition cost from state i to state j. In this case, we use the *expected cost* per stage defined as:

$$g(i,u) = \textstyle\sum_{j \in S} \tilde{g}(i,u,j) p_{i,j}(u) \tag{1}$$

The objective of the MDP is to find the optimal stationary policy μ such that the total cost is minimized. The total cost may be defined in several ways. We will focus our attention on average cost problems. In this case, the cost to be optimized is given by the following equation

$$\lambda = \lim_{N \to \infty} \frac{1}{N} E \left\{ \sum_{k=1}^{N-1} g\left(x_k, \mu(x_k)\right) \right\} \tag{2}$$

where x_k represents the system's state at the k-th stage. Note that in the definition of the average cost λ we are implicitly assuming that its value is independent of the initial state of the system. This is generally not always true. However there are certain conditions under which this assumption holds. For example, in our scenario, the value of the per-stage cost is always bounded and both S and U are finite sets. Moreover, there is at least one state, n that is *recurrent* in every stationary policy. Given previous conditions, the limit in the right side of (2) exists and the average cost does not depend on the initial state.

Sometimes the system is modeled as a continuous-time Markov chain. In this case, as we shall see, the definition of the average cost is slightly different. In order to solve it by means of the known equations for average cost MDP problems, we have to construct an auxiliary discrete-time problem whose average cost equals the one of the continuous-time problem.

Given the conditions for the limit in 2 to exist, the *optimal* average cost can be obtained by solving the following Bellman's equation

$$h(i) = \min_{u \in U} \left[g(i,u) - \lambda + \sum_{j=1}^{N} p_{ij}(u) h(j) \right] \quad i \in S \tag{3}$$

with the condition $h(n) = 0$. It is known (see Bertsekas (2007)) that previous equations have a unique solution and the stationary policy μ providing the minimum at the right side of (3) is an optimal policy. $h(i)$ is known as relative or differential cost for each state i. It represents the minimum, over all policies, of the difference between the expected cost to reach n from i for the first time and the cost that would be incurred if the cost per stage were equal to the average λ at all states.

There are several computational methods for solving Bellman equation: the value iteration algorithm, the policy iteration algorithm and the linear programming method provide exact solutions to the problem (see Bertsekas (2007) and Puterman (2005)). However, when the dimension of the sets S and U is relatively large, the problem becomes so complex that solving it exactly may be computationally intractable. This is known as the *curse of dimensionality* in dynamic programming. In some situations, we are not able to compute all the transition probabilities $p_{ij}(u)$ of the model, therefore obtaining an exact solution is impossible. For these cases multiple approximate methods have been developed within the framework of approximate dynamic programming (see Powell (2005)) or reinforcement learning.

There are several variations for MDP problems. One of the most important ones refers to the time horizon over which the process is assumed to operate. It may be finite, when the optimization is done over a finite number of stages, or infinite, when the number of stages is assumed to be infinite. The latter type of problems present some theoretical difficulties, and some technical conditions must hold to be solvable. However, when these conditions are present, infinite-horizon problems require less computational effort than finite-horizon problems with similar dimension. Sometimes, more than one performance objective must be attained. In these cases, it is usual to set bounds in all the objectives except one, which should be optimized assuring that the other objectives remain within their bounds, *i.e.* the rest of objectives constitute constraints on the MDP problem. This strategy is known as constrained MDP (CMDP). To solve these problems, the most usual approaches are to re-formulate the problem as a linear-programming one or to use Lagrangian relaxation on the constraints. Finally, in some problems, the control decision at each state must be taken without complete knowledge of the state. Instead of directly observing the state, the controller observes an additional variable related with the state, so that the probability of each state can be inferred. These problems are known as Partially Observable MDP (POMDP) and are tractable, in general, only for small dimensional problems. The more complex versions of MDPs are, in fact, generalizations of the problem. As we will see, some problems must be formulated as Constrained POMDP, for which very few results are available so far and are generally addressed by heuristic methods.

3. MDP applications in cognitive radio

MDP has been frequently applied in the design of MAC protocols in cognitive radio. They can be classified into two classes: decentralized and centralized access protocols. In the decentralized case, each unlicensed user is responsible of performing spectrum sensing and spectrum access, in general with limited, and sometimes unreliable, information about the

spectrum usage. In consequence, it is usual to find partially observed MDP (POMDP) formulations of the problem, which easily become intractable when the dimension of the problem increases. The access of secondary users to the spectrum should have the less possible impact on licensed users. When including these restrictions on the formulation the resulting problem is a constrained POMDP. In the centralized case, a central device, generally referred to as spectrum broker, performs spectrum management, controlling the access of secondary users to idle spectrum channels. It is usually assumed that the spectrum broker has perfect information about the spectrum usage, therefore the problem is formulated as an MDP, or as a CMDP if constraints are included.

3.1 Decentralized access

In Zhao et. al. (2007), the activity of a licensed user is modeled as an on-off model represented by a two-state Markov chain. The problem of channel sensing and access in a spectrum overlay system was formulated as a POMDP. The actions consists on sensing and accessing a channel, and the channel sensing result is considered an observation. The reward is defined as the number of transmitted bits. The objective is to maximize the expected total number of transmitted bits in a certain number of time slots under the constraint that the collision probability with a licensed user should be maintained below a target level.

Geirhofer et. al. (2008) propose a cognitive radio that can coexist with multiple parallel WLAN channels, operating below a given interference constraint. The coexistence between conventional and cognitive radios is based on the prediction of WLAN's behavior by means of a continuous-time Markov chain model. The cognitive MAC is derived from this model by recasting the problem as a constrained Markov decision process (CMDP).

The goal in Chen et. al (2008) is to maximize the throughput of a secondary user while limiting the probability of colliding with primary users. The access mechanism comprises the following three basic components: a spectrum sensor that identifies spectrum opportunities, a sensing strategy that determines which channels to sense and an access strategy that decides whether to access based on potentially erroneous sensing outcomes. This joint design was formulated as a constrained partially observable Markov decision process (POMDP).

The approach in Li et. al. (2011) is to maximize the throughput of the secondary user subject to collision constraints imposed by the primary users. The formulation follows a constrained partially observable Markov decision process.

3.2 Centralized access

In Yu et. al. (2007) the spectrum broker controls the access of secondary users based on a threshold rule computed by means of an MDP formulation with the objective of minimizing the blocking probability of secondary users. In order to cope with the non-stationarity of traffic conditions, the authors propose a finite horizon MDP instead of an infinite horizon one. The drawback is that the policy cannot be computed off-line, imposing a high computational overhead on the system.

Tang et. al. (2009) study several admission control schemes at a centralized spectrum manager. The objective is to meet the traffic demands of secondary users, increasing spectrum utilization efficiency while assuring a grade of service in terms of blocking probability to primary users. Among the schemes analyzed, the best performing one is based on a constrained Markov decision process (CMDP).

Centralized access has received less attention than decentralized access in cognitive radio research in general and in the application of MDP in particular. On the one hand, decentralized access constitutes a harder research challenge because each agent only has partial and sometimes unreliable information about the wireless network and the spectrum bands. This leads to the harder POMDP problems. On the other hand, although centralized access relies on a spectrum broker which generally has full information about the system state, the dimension of the problem increases proportionally to the total number of managed channels. Therefore, although the MDP or CMDP problem may be solvable, its dimension imposes a serious computational overhead. This drawback may be overcome with an off-line computation of the policies. However, when traffic conditions are non-stationary this approach is not applicable and approximate solutions based on reinforcement learning strategies should be explored. In this work we focus on the application of MDP to centralized access and how it can be exploited to balance GoS of each class of user.

3.3 Other applications

Other applications of MDP have been found within the framework of cognitive radio. In Hoang et. al. (2010), authors propose an algorithm based on finite-horizon MDP to schedule the duration of spectrum sensing periods and data transmission periods at the cognitive users aiming to improve their throughput. Berthold et. al. (2008) formulate the spectral resource detection problem as an MDP allowing the cognitive users to select the frequency bands with the most available resources. Galindo-Serrano and Giupponi (2010) deals with the problem of aggregated interference generated by multiple cognitive radios at the receivers of primary (licensed) users. The problem is formulated as a POMDP and it is solved heuristically by means of an approximated dynamic programming method known as distributed Q-learning.

In this paper we highlight another application of MDP: dynamic trading of spectrum bands. While this issue has been typically addressed with a game-theoretic approach, we explore the use of MDP and CMDP formulations to balance benefit and grade of service for primary users in a centralized spectrum access framework.

4. System model

In this section we consider two models for coordinated spectrum access. In the first one, secondary users are accepted or rejected according to an admission policy that only considers the impact on the blocking probability for primary users. In this first model there is a trade-off between the blocking probability of licensed and unlicensed users. The second model includes a spectrum bidding procedure, in which secondary users offer a price, within a finite countable set of prices for mathematical tractability, for the use of a channel. In the second model the trade-off appears between the the blocking probability of licensed users and the expected benefit obtained from spectrum rental.

4.1 Priority-based access

This access is only based on priority, not in bidding price, *i.e.* licensed users are given higher priority than secondary users. Therefore the objective is to minimize the blocking probability of licensed users but also that of unlicensed users. The general rule is that primary users are always accepted if there are available channels but, depending on the available channels, the controller can deny access to secondary users. Once a secondary user occupies a channel, it is this user who decides when to release this channel and it can not be removed by the controller.

There are several approaches to address this type of problems. One of them is to formulate an MDP where the expected cost is obtained as a linear combination (more precisely a convex combination) of the blocking probability of each class of users. By adjusting the weighting factors we can compute a Pareto front for both blocking probabilities. A Pareto front is defined as the set of values corresponding to several coupled objective functions such that, for every point of the set, one objective cannot be improved without worsening the rest of objective values. In this type of access, the Pareto front allows to fix a blocking probability value for the licensed users and know the best possible performance for unlicensed users.

Incoming traffic is characterized by a classic Poisson model. Licensed users arrive with a rate of λ_L arrivals per unit of time. The arrival rate for unlicensed users is denoted by λ_U. The licensed spectrum managed by the central controller is assumed to be divided into channels (or bands) with equal bandwidth. Each user occupies a single channel. The average holding times for licensed and unlicensed users are given by $1/\mu_L$ and $1/\mu_U$ respectively, where μ_L and μ_U denote the departure rate for each class. Because a Poisson traffic model is considered, both the inter-arrival time and the channel holding times are exponentially distributed random variables for both user classes. The model can be easily extended including more user classes, the probability that a user occupies two or more channels, and so on. Essentially the procedure is the same, but the Markov chain would comprise more states as more features are considered in the model. In this model, the state of the Markov chain is determined by the number of channels k occupied by licensed users (LU), and the number of channels s occupied by secondary users (SU). Because spectrum is a limited resource, there is a finite number N of channels. Figure 1 depicts a diagram of the model and its parameters. Note that we can map all the possible combinations of (k, s) for $0 \leq k \leq N$, $0 \leq s \leq N$ and $k + s \leq N$ to a single integer i such that

$$0 \leq i \leq \frac{N(N+1)}{2} + N + 1. \tag{4}$$

The number in the right hand side of 4 is the total number of states. Let N_T denote this number.

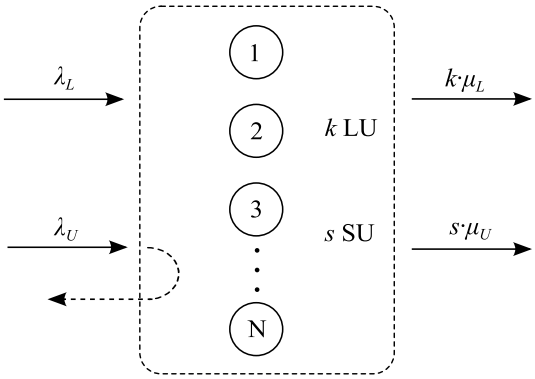


Fig. 1. Diagram of the priority based access model. The system has N channels that can be occupied by k licensed users (LU) and s secondary users (SU) such that $k + s \leq N$. The total departure rates for each type of users depend on k and s.

The model described above consists of a continuous-time Markov chain. In the framework of MDPs we have to define the actions and the costs of these actions. Let $g(i, u)$ denote the

instantaneous cost of taking action u at state i. In the system considered action u is simply defined as

$$u = \begin{cases} 1 & \text{, if incoming user is not accepted} \\ 0 & \text{, otherwise} \end{cases} \tag{5}$$

The above formula refers only to unlicensed users. It is assumed that licensed users are always accepted unless all channels are occupied. The function $g(i,u)$ is given by the convex combination of two per-stage cost functions, *i.e.* $g(i,u) = \alpha g_L(i,u) + (1-\alpha) g_U(i,u)$, where

$$g_L(i,u) = \begin{cases} 1 & \text{, if } i \equiv (k,s) \text{ and } k+s = N \\ 0 & \text{, otherwise} \end{cases} \tag{6}$$

where the symbol "$\equiv$" denotes equivalence, i.e. i maps a state (k,s) such that $k+s=N$. Similarly,

$$g_U(i,u) = \begin{cases} 1 & \text{, if } i \equiv (k,s) \text{ and } k+s = N \\ u & \text{, otherwise} \end{cases} \tag{7}$$

These functions determine the blocking probability per unit of time for each class of users. Note that the blocking probability is defined as the probability that the system does not provide a channel to an incoming user. The objective is to find a policy such that, for a relative importance given to each cost (determined by α), the expected average value of the combined cost is minimized. The function to minimize is then given by

$$\lim_{K\to\infty} \frac{1}{E\{t_K\}} E\left\{ \int_0^{t_K} g\left(x(t), u(t)\right) \right\} \tag{8}$$

where t_K is the completion time of the K-th transition. The problem can be solved by formulating its auxiliary discrete-time average cost problem. Let γ be a scalar greater than the transition rate at any state of the chain, *i.e.* $\gamma > v_i(u)$. We can compute the transitions probabilities $\tilde{p}_{i,j}(u)$ for the auxiliary discrete-time problem from the probabilities $p_{i,j}(u)$ of the original problem as

$$\tilde{p}_{i,j}(u) = \begin{cases} \frac{v_i(u)}{\gamma} p_{i,j}(u) & \text{, if } i \neq j \\ 1 - \frac{v_i(u)}{\gamma} & \text{, if } i = j \end{cases} \tag{9}$$

It is known (see Bertsekas (2007)) that if the scalar λ and the vector $\tilde{h}$ satisfy

$$\tilde{h}(i) = \min_{u\in\{0,1\}} \left[g(i,u) - \lambda + \sum_{j=1}^{N_T} \tilde{p}_{ij}(u) \tilde{h}(j) \right] \quad i = 1, \ldots, n \tag{10}$$

then λ and the vector h with components $h(i) = \gamma \tilde{h}(i)$ solve the original problem. It can be anticipated that the structure of this problem, essentially a connection admission control problem, requires a threshold type solution in which upcoming unlicensed users will only be admitted into the system if the number of occupied channels is below certain threshold.

4.2 Auction-based access

As explained in the introduction, public administrations assign the spectrum bands to wireless operators by a license scheme. Generally, operators gain spectrum licenses by bidding for them in public auction processes. We refer to this spectrum assignment framework as primary

market. The increasing demand of spectrum and the existence of spectrum holes have revealed the inefficiency of this mechanism. One practical and economically feasible way to solve this inefficiency is to allow spectrum owners to sell their spectrum opportunities in a secondary market. In contrast to the primary market, the secondary operates in real-time. Secondary users, that may be operators without a spectrum license, submit their bids for spectrum opportunities to the spectrum owner, who determines the winner or winners by giving them access to the band and charging them the bidding price.

The arrival processes are modeled, as in previous subsection, as independent Poisson processes. The arrival rates for licensed and unlicensed users are λ_L and λ_U respectively. The service rates are μ_L and μ_U. Again, it is assumed that each incoming user occupies a single channel. The system state is given by the number of primary users k and secondary users s holding a channel: (k,s) for $0 \leq k \leq N, 0 \leq s \leq N$ and $k+s \leq N$. Each state is mapped into an integer $i \equiv (k,s)$, so that $i = 0,1,\ldots N_T$, where N_T is given by 4. For mathematic tractability, the bidding prices are classified into a finite set of values: $\mathbb{B} = \{b_1, b2, \ldots b_m\}$ given in money charged per unit of time. Each price on this set has a probability p_i, $i = 1 \ldots m$ to be offered by an incoming user. Obviously $\sum_{i=1}^{m} p_i = 1$. Figure 2 depicts the model described.

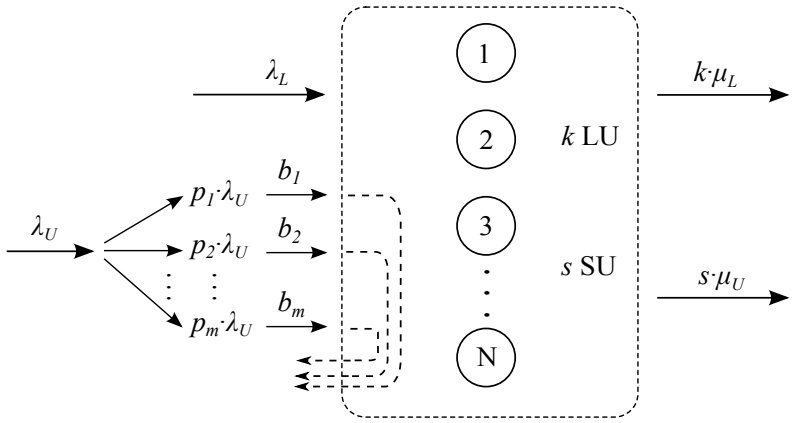


Fig. 2. Diagram of the auction based access model. Secondary users (SU) can offer up to m different bid prices. Each bid offer is assigned a probability. The access policy decides upon each offer according to the price offered and the system's state.

In this case, the objective of the MDP is to obtain the maximum economic profit with the minimum impact on the licensed users. The control u at each stage determines the admitted and rejected bidding prices. Logically, the control should be defined as a threshold, *i.e.* when $u = i$ only bids equal or above p_i are admitted. For notation convenience, the control $u = m+1$ indicates that no bid is accepted. The per-stage reward function $g(i,u)$ is given by the linear combination of $g_L(i,u)$ (defined in previous subsection) and $g_U(i,u)$ defined, in this model, as the expected benefit at stage i when decision u is made. Therefore $g(i,u) = \alpha g_L(i,u) + \beta g_U(i,u)$ where the scalars α and β are weighting factors. Note that $\beta < 0$ since the objective is to minimize the average expected cost given by $g(i,u)$. Let B_i denote the expected income when an unlicensed user whose bidding price is b_i is accepted. Since the average channel holding time for unlicensed users is $1/\mu_U$, then $B_i = b_i/\mu_U$. Given a control u, $P(r|u)$ denotes

the conditional probability that the bidding price of the next accepted secondary user is b_r.

$$P(r|u) = \begin{cases} \frac{p_r}{\sum_{j=u}^{m} p_j} & \text{, if } r \geq u \\ 0 & \text{, otherwise} \end{cases} \tag{11}$$

Let us define $\tilde{g}_U(i,u,j)$ as the average benefit associated to the transition from state i to state j. Its expression is

$$\tilde{g}_U(i,u,j) = \begin{cases} p_U \sum_{r=1}^{m} B_r P(r|u) & \text{, if } j = i+1 \\ 0 & \text{, otherwise} \end{cases} \tag{12}$$

where $p_U = \lambda_U/(\lambda_U + \lambda_L)$ denotes the probability that the next arrival corresponds to a secondary user. Therefore, the per-stage benefit $g_U(i,u)$ is given by

$$\begin{aligned} g_U(i,u) &= \sum_{j=1}^{N_T} \tilde{g}_U(i,u,j) p_{i,j}(u) \\ &= p_{i,i+1}(u) p_U \sum_{j=1}^{N_T} B_r P(r|u). \end{aligned} \tag{13}$$

We can formulate the auxiliary discrete-time average cost problem for the model described. The equation providing the optimum average cost λ is

$$\tilde{h}(i) = \min_{u \in \{0,1\}} \left[\alpha g_L(i,u) + \beta g_U(i,u) v_i(u) - \lambda + \sum_{j=1}^{N_T} \tilde{p}_{ij}(u) \tilde{h}(j) \right] \tag{14}$$

for $i = 1, \ldots, n$. The structure of this problem also anticipates a threshold-type solution. In this case, there will be a set of thresholds, one per bidding price. By properly adjusting the weighting factors α and β we can also compute a Pareto front allowing us to determine the maximum possible benefit for a given blocking objective for the licensed users.

4.3 Constrained MDP

So far, the approach to merge several objectives consisted on combining them into a single objective by means of a weighted sum and solving the problem as a conventional MDP. However, as explained in Section 2, when several objectives concur in an MDP problem, the formulation strategy may consist on optimizing one of them subject to constraints on the other objectives. This strategy results in a CMDP formulation of the problem. Solving MDPs by iterative methods such as policy or value iteration allows us to find deterministic policies, *i.e.* policies that associate each system's state $i \in S$ to a single control $u \in U(i)$, where $U(i)$ is a subset of U containing the controls allowed in state i. However, these policies do not, in general, solve CMDP problems. Instead, the solution of CMDPs is a randomized policy, defined as a function that associates each state to a probability distribution defined over the elements in $U(i)$.

There are mainly two approaches to solve CMDPs, linear programming (LP) and Lagrangian relaxation of the Bellman's equation. This paper follows the former one. Each feasible LP formulation relies on the use of the *dual* variables $\phi(i,u)$, defined as the stationary probability that the system is in state i and chooses action u under a given randomized stationary policy. The problems addressed in this paper result, under every stationary policy, in a truncated birth-death process, since primary users are always accepted. In consequence, every resulting Markov chain is *irreducible*, in other words, it is recurrent and there are not transient states.

Moreover, the state and action spaces are finite. Under these circumstances, as shown in Puterman (2005), every feasible solution of the LP problem corresponds to some randomized stationary policy. Therefore, if the constrained problem is feasible, then there exists an optimal randomized stationary policy.

The LP approach consists of expressing the objective and the constraints in terms of $\phi(i,u)$. Once the problem is discretized, the average cost is defined as

$$\lambda = \lim_{K\to\infty} \frac{1}{K} E\left\{ \sum_{k=0}^{K} g(x_k, u_k) \right\} \tag{15}$$

where k denotes the decision epoch of the process. The objective is to find the policy μ solving

$$\min_{\mu} \lambda \tag{16}$$

The constraints are defined similarly to the main objective: each constraint impose a bound on an average cost related to different per-stage cost. Each constraint has the following form:

$$c = \lim_{K\to\infty} \frac{1}{K} E\left\{ \sum_{k=0}^{K} c(x_k, u_k) \right\} \leq \beta \tag{17}$$

where $c(x(t), u(t))$ is the real-valued function providing the per-stage cost associated to the constraint β. Therefore the constrained average reward MDP with one constraint is defined as

$$\begin{array}{c} \min \lambda \\ \text{s.t.} \\ c \leq \beta \end{array} \tag{18}$$

Given the characteristics of the problem (finite state and action spaces and recurrent Markov chain under every policy), the limits in (15) and (17) exist and are equal to

$$\lambda = \sum_{i\in S} \sum_{u\in U(i)} g(i,u)\, \phi(i,u) \tag{19}$$

and

$$c = \sum_{i\in S} \sum_{u\in U(i)} c(i,u)\, \phi(i,u) \tag{20}$$

respectively. In addition, the following conditions must be hold by the *dual* variables:

$$\sum_{u\in U(j)} \phi(j,u) = \sum_{i\in S} \sum_{u\in U(i)} p_{i,j}(u)\, \phi(i,u) \tag{21}$$

for all $j \in S$, which is closely related to the balance equations of the Markov chain and

$$\sum_{i\in S} \sum_{u\in U(i)} \phi(i,u) = 1, \tag{22}$$

which, together with $\phi(j,u) \geq 1$ for $i \in S$ and $u \in U(i)$ correspond to the definition of $\phi(i,u)$ as a limiting average state action frequency. In consequence, the LP for the CMDP has the

following formulation

$$\begin{gathered}\min_{\phi} \sum_{i\in S}\sum_{u\in U(i)} g(i,u)\,\phi(i,u) \\ \text{s.t.} \\ \sum_{i\in S}\sum_{u\in U(i)} c(i,u)\,\phi(i,u) \le \beta \\ \sum_{u\in U(j)} \phi(j,u) - \sum_{i\in S}\sum_{u\in U(i)} p_{i,j}(u)\,\phi(i,u) = 0 \\ \sum_{i\in S}\sum_{u\in U(i)} \phi(i,u) = 1 \\ \phi(j,u) \ge 1\end{gathered} \tag{23}$$

Assuming that the problem is feasible and ϕ^* is the optimal solution of the LP problem above, the stationary randomized optimal policy μ^* is generated by

$$q_{\mu^*(i)}(u) = \frac{\phi^*(i,u)}{\sum_{u'\in U(i)} \phi^*(i,u')} \tag{24}$$

for cases where the sum in the denominator is nonzero. Otherwise, the state is transitory and the control is irrelevant. Note that $q_{\mu^*(i)}(u)$ denotes the probability of choosing action u at state i under policy μ^*.

Using the approach above in the problems described in previous section is straightforward:

- *Priority-based access*: in the LP problem (23) replace $g(i,u)$ by $g_U(i,u)$ defined in (7), and $c(i,u)$ by $g_L(i,u)$ defined in (6). For each value of β we obtain the point in the Pareto front corresponding to a blocking probability β for the licensed users.
- *Auction-based access*: in the LP problem (23) replace $g(i,u)$ by $g_U(i,u)$ defined in (13), and $c(i,u)$ by $g_L(i,u)$ defined in (6). As in previous case, for each value of β we obtain a point in the Pareto front.

5. Numerical results

In this section we provide examples of the Pareto front computation procedures described in previous section for each DSA type.

5.1 Priority based access

For this DSA scheme we will consider three scenarios characterized by the asymmetry between the traffic intensity of licensed and unlicensed users. In every scenario, the average holding time is equal for every user, independently of their type. Therefore the service rate $\mu_L = \mu_U = 5$. Assuming that the time unit is an hour, this results in an average holding time of 12 minutes per connection. The total traffic ($\lambda = \lambda_L + \lambda_U$) is 40 calls/h, which results in a total incoming traffic of 8 Erlangs. In a wireless cell covering 2.5 km^2 of urban area (cell radius equal to 400 m), with 2000 people per km^2 and a 10% aggregate market penetration (licensed and unlicensed users), the number of covered users is around 500, and the resulting traffic intensity is 0.016 Erlangs per user. The number of available channels is set to $N = 10$, in order to evaluate the system in a relatively congested situation. With the assumed traffic intensity we can estimate the blocking probability of the system for the aggregate traffic by means of

the well-known Erlang's B formula (see Kleinrock (1975)):

$$E(n,\rho) = \frac{\frac{\rho^n}{n!}}{\sum_{j=0}^{j=n} \frac{\rho^j}{j!}} \tag{25}$$

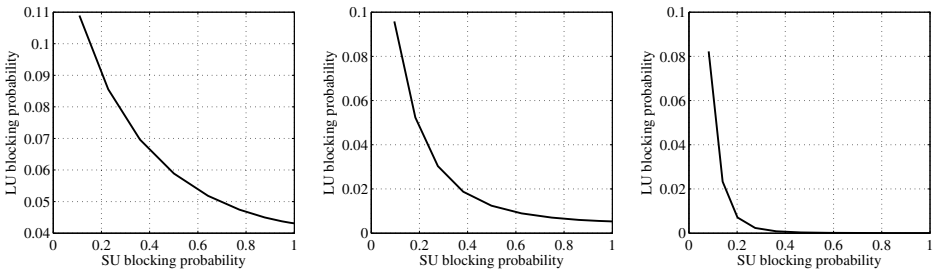
where n is the number of channels and ρ denotes the utilization factor. In our case $\rho=\lambda/\mu_L = \lambda/\mu_U$. According to this formula, if the system accepted every incoming user, the total blocking probability would be $E(10,8)$=0.12. As we will see, this probability is an upper bound for the blocking probability of the primary users, which are always accepted if the system has any available channel, and a lower bound for the secondary users.

The three scenarios are summarized in Table 1.

parameter	**scenario 1**	**scenario 2**	**scenario 3**
λ_L (calls/h)	30	20	10
λ_U (calls/h)	10	20	30
μ_L=μ_U (calls/h)	5	5	5
N	10	10	10

Table 1. Parameters values at the three scenarios of the priority based access problem.

First, we show in Fig. 3 the Pareto front obtained by means of an MDP where the blocking costs of licensed and unlicensed users were merged by means of a convex combination. The Pareto front was obtained by solving each MDP problem for 10000 values of the α parameter ranging from 0.01 to 1.



(a) scenario 1 (b) scenario 2 (c) scenario 3

Fig. 3. Pareto fronts obtained for the priority-based access in scenario 1 (a), scenario 2 (b) and scenario 3 (c)

All the three scenarios receive the same total traffic intensity. However, when the traffic intensity of the primary users is smaller, the Pareto front is closer to both axes, *i.e.* the performances of both the primary and secondary users improve. This is an expectable result since only the traffic of secondary users is controlled by the access policy. When the optimization affects to a higher portion of the total amount of traffic the improvement is also more noticeable, showing the benefits of the MDP formulation.

The Pareto fronts obtained by means of the CMDP formulation in previous scenarios are identical to those shown in Fig. 3, showing that both formulations are equivalent in terms of finding the Pareto front for the priority-based access problem. The only difference relies on practical considerations. The CMDP approach allows us to find a policy with a predefined

blocking probability for primary users while the MDP formulation implies the exploration of the Pareto front, since there is no a priori relationship between α and this blocking probability. On the other hand, implementing the policy solving the CMDP problem implies to randomize at least one control (it can be shown that the number of required randomized controls equals the number of constraints). While this is technically feasible, a stationary deterministic policy is simpler to implement.

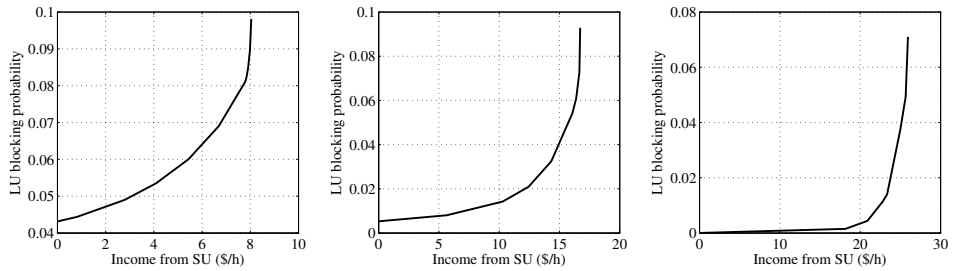
5.2 Auction based access

For the auction-based access we consider again the three scenarios defined in previous section. Additionally we define three classes of secondary users (SU), characterized by the price that they offer per minute of channel occupation. The bid offers per class are: class 1: 0.01 \$/m, class 2: 0.02 \$/m and class 3: 0.03 \$/m. Additionally, we define the probability of an SU incoming call being of each class. The SU class probability distribution is: class 1 probability: 0.5, class 2 probability: 0.3 and class 3 probability: 0.2. We summarize SU class definition in Table 2.

SU class	**class 1**	**class 2**	**class 3**
offered price (\$/m)	0.01	0.02	0.03
probability	0.5	0.3	0.2

Table 2. Classification of SU in terms their bid offers and their probabilities.

Note that both the offered prices and their probability distributions are static, *i.e.* they do not change over time and are independent of the system occupation. It is not completely unrealistic taking into account typical tariff policies of wireless operators. In this environment the class structure and the probability distribution may be seen as types of contracts for secondary users and market penetration of each type of contract respectively. However, for a more dynamical auction process, where bidders are able to change their bid offers adaptively, the model should be revised. One possibility would be to define one probability distribution for each state. More detailed modeling strategies would increase the complexity of the MDP solving algorithm or even make them intractable. This is a classic problem of MDP formulation, known as the *curse of dimensionality* and is typically addressed by means of the heuristic approach of approximate dynamic programming.



(a) scenario 1 (b) scenario 2 (c) scenario 3

Fig. 4. Pareto fronts obtained for the auction-based access in scenario 1 (a), scenario 2 (b) and scenario 3 (c)

Figure 4 shows the Pareto fronts for the auction-based system in the three scenarios. As in previous subsection, the MDP and the CMDP approaches provided similar results. It

can be observed that, for the same traffic intensity (the three scenarios receive 40 calls per unit of time) when the traffic share of the secondary users is higher (scenarios with higher number) the Pareto front moves away from the y-axis, *i.e.* the income obtained from secondary users increases and it also approaches the x-axis, *i.e.* the blocking probability of the licensed users diminishes. It is interesting to check that, especially in scenarios 2 and 3, a very small increment of the blocking probability of licensed users can multiply the benefit obtained from spectrum leasing by a factor of 2 or 3. On the other hand, these figures also indicate that once the income surpasses certain threshold, Pareto-optimal policies can only produce small increments of the income by dramatically rising the blocking probability.

6. Conclusions

This chapter has surveyed the use of MDP formulation within the framework of cognitive radio. We have reviewed the fundamentals of MDP and its generalizations, such as CMDP, POMDP and constrained POMDP. While most previous works focus on decentralized access, we focus on centralized access. The main difference between them is that when the access relies on a central controller or *spectrum broker*, it generally has full knowledge of the spectrum occupation, while in decentralized access decision have to be taken with partial and sometimes unreliable information about channel occupation. Therefore, centralized schemes are more suitable to MDP or CMDP modeling, while decentralized ones generally require POMDP or constrained POMDP which are intractable in many cases and require approximated or heuristic algorithms. We consider two types of access: one where only one type of secondary user tries to access the licensed spectrum and other where users are classified according to the price they are willing to pay for the use of the spectrum. The first one is referred to as priority-based access and the second one as auction-based access. The main issue of the problems addressed is that two contrary objectives coexist. In priority-based access, the controller tries to reduce the blocking probability of both types of users. In auction-based, the objectives are to reduce blocking probability for licensed users and to increase the income received from spectrum leasing. For these problems there does not exists an *optimal* policy, but a set of *Pareto optimal* policies. The performance of these policies lie on the Pareto front, defined as the set of points where one objective cannot be improved without worsening the other one. We have shown how to compute these Pareto fronts for each access scheme by weighting the objectives in an MDP problem and by formulating a CMDP. The first approach requires solving Bellman's equation and the second requires solving a linear program. We have obtained the Pareto fronts for several scenarios, showing the influence of traffic share on system's performance. The Pareto front is a very usual tool to determine the performance threshold for each objective upon which further increments on this objective require excessive degradation of the other one. MDP and CMDP are useful tools for developing centralized access policies for cognitive radio systems. One drawback is the so-called *curse of dimensionality*, that may render computationally intractable the problem as the sizes of the state and action spaces increase. In addition, although policies can be computed off-line, alleviating the computational overhead of the access controller, the system's parameters may be variable, requiring many pre-computed policies and thus imposing large memory requirements.

7. Acknowledgments

This research has been supported by the MICINN/FEDER project grant TEC2010-21405-C02-02/TCM (CALM) and it was also developed in the framework of

"Programa de Ayudas a Grupos de Excelencia de la Region de Murcia, Fundacion Seneca, Agencia de Ciencia y Tecnologia de la RM (Plan Regional de Ciencia y Tecnologia 2007/2010".

8. References

FCC Spectrum policy Task Force "Report on the spectrum efficiency group,", FCC Report, Nov. 2002.

D. P. Bertsekas, "Dynamic Programming and Optimal Control, vol. 2," Third Edition *Athenea Scientific*, 2007.

M. L. Puterman "Markov Decision Processes: Discrete Stochastic Dynamic Programming," First Edition *Wiley-Interscience*, 2005.

W. B. Powell "Approximate Dynamic Programming: Solving the Curses of Dimensionality," First Edition *Wiley-Interscience*, 2007.

Q. Zhao, L. Tong, A. Swami and Y. Chen, "Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: a POMDP framework," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 3, pp. 589-600, April 2007.

S. Geirhofer, L. Tong and B.M. Sadler, "Cognitive Medium Access: Constraining Interference Based on Experimental Models," *Selected Areas in Communications, IEEE Journal on* , vol.26, no.1, pp.95-105, Jan. 2008.

Y. Chen, Q. Zhao and A. Swami, "Joint Design and Separation Principle for Opportunistic Spectrum Access in the Presence of Sensing Errors," *Information Theory, IEEE Transactions on* , vol.54, no.5, pp.2053-2071, May 2008.

X. Li, Q. Zhao, X. Guan and L. Tong, "Optimal Cognitive Access of Markovian Channels under Tight Collision Constraints," *Selected Areas in Communications, IEEE Journal on*, vol.29, no.4, pp.746-756, April 2011.

O. Yu, E. Saric and A. Li, "Dynamic control of open spectrum management," in *Proceedings of IEEE Wireless Communications and Networking Conference* (WCNC), March 2007, pp. 127-132.

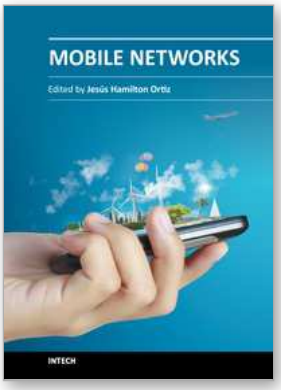
P. K. Tang, Y. H. Chew, W.-L. Yeow and L. C. Ong, "Performance Comparison of Three Spectrum Admission Control Policies in Coordinated Dynamic Spectrum Sharing Systems," *Vehicular Technology, IEEE Transactions on* , vol.58, no.7, pp.3674-3683, Sept. 2009.

A. Hoang, Y.-C. Liang and Y. Zeng., "Adaptive joint scheduling of spectrum sensing and data transmission in cognitive radio networks," *Communications, IEEE Transactions on* , vol.58, no.1, pp.235-246, Jan. 2010.

U. Berthold,F. Fangwen, M. van der Schaar and F.K. Jondral, "Detection of Spectral Resources in Cognitive Radios Using Reinforcement Learning," in Proc. *New Frontiers in Dynamic Spectrum Access Networks, 2008. DySPAN 2008. 3rd IEEE Symposium on* , vol., no., pp.1-5, 14-17 Oct. 2008.

A. Galindo-Serrano and L. Giupponi, "Distributed Q-Learning for Aggregated Interference Control in Cognitive Radio Networks," *Vehicular Technology, IEEE Transactions on* , vol.59, no.4, pp.1823-1834, May 2010.

L. Kleinrock, "Queuing Systems, Volume 1: Theory," *John Wiley & Sons*, New York, 1975.



Mobile Networks

Edited by Dr. Jesús Ortiz

ISBN 978-953-51-0593-0
Hard cover, 192 pages
Publisher InTech
Published online 09, May, 2012
Published in print edition May, 2012

The growth in the use of mobile networks has come mainly with the third generation systems and voice traffic. With the current third generation and the arrival of the 4G, the number of mobile users in the world will exceed the number of landlines users. Audio and video streaming have had a significant increase, parallel to the requirements of bandwidth and quality of service demanded by those applications. Mobile networks require that the applications and protocols that have worked successfully in fixed networks can be used with the same level of quality in mobile scenarios. Until the third generation of mobile networks, the need to ensure reliable handovers was still an important issue. On the eve of a new generation of access networks (4G) and increased connectivity between networks of different characteristics commonly called hybrid (satellite, ad-hoc, sensors, wired, WIMAX, LAN, etc.), it is necessary to transfer mechanisms of mobility to future generations of networks. In order to achieve this, it is essential to carry out a comprehensive evaluation of the performance of current protocols and the diverse topologies to suit the new mobility conditions.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Juan J. Alcaraz, Mario Torrecillas-Rodríguez, Luis Pastor-González and Javier Vales-Alonso (2012). Dynamic Spectrum Access in Cognitive Radio: An MDP Approach, Mobile Networks, Dr. Jesús Ortiz (Ed.), ISBN: 978-953-51-0593-0, InTech, Available from: http://www.intechopen.com/books/mobile-networks/dynamic-spectrum-access-in-cognitive-radio-an-mdp-approach

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821