Enhanced Cognitive Radio Operation through Belief-based Decision Making

J. Pérez-Romero, A. Raschellà, O. Sallent, A. Umbert
Dept. of Signal Theory and Communications
Universitat Politècnica de Catalunya (UPC)
Barcelona, Spain
e-mail: [jorperez, alessandror, sallent, annau]@tsc.upc.edu

***Abstract*— This paper focuses on the analysis of the trade-off existing between achieved performance and observation requirements in a decision making framework based on the cognitive cycle. For that purpose, it presents a general belief based decision making framework that can be particularized for different observation strategies that decide when measurements are carried out to characterize the dynamics of the radio environment. While the concept can be applied to different cognitive radio problems, the paper focuses on the spectrum selection to establish a set of radio links. The considered observation strategies and associated decision making are assessed under different environment conditions in terms of the traffic generation patterns. Results reveal that both the session generation rate and the session duration play a key role to choose the adequate observation strategy that balances the achieved performance and the measurement requirements.**

Keywords—cognitive radio; decision making; belief vector; spectrum selection

I. INTRODUCTION

Cognitive Radio (CR) [1] has emerged in the last years as a paradigm that envisages an intelligent radio able to automatically adjust its behavior based on the active monitoring of its environment. From a general perspective CR makes use of the cognitive cycle that involves observations of the environment, analysis of these observations, decision making to smartly configure certain radio parameters and finally execution of the decisions by means of actions. Analysis and decision can be supported by means of learning mechanisms that exploit the knowledge obtained from the execution of prior decisions. This general approach can be applied for efficiently and adaptively modifying different radio operational parameters such as frequency, transmit power, modulation scheme, etc.

Focusing on the observation stage, it typically involves making measurements at several nodes of a CR network. Then, these measurements need to be reported to the node in charge of the decision making. This is usually done through signaling procedures supported by cognitive control channels [2][3]. As a result, the observation stage can be very costly in terms of practical requirements such as signaling overhead, battery consumption, etc. Consequently, decision making strategies able to efficiently operate with the minimum amount of measurements become of high interest for enhancing CR operation.

Different approaches exist in the literature addressing decision making schemes in CR networks able to operate with reduced measurement needs. For instance, several research works rely on Partially Observable Markov Decision Processes (POMDPs) [4] as a decision making tool that combines partial observations of the radio environment at specific periods of time with a statistical characterization of the system dynamics. They have been used in [5] for implementing a cognitive Medium Access Control (MAC) that enables opportunistic spectrum access in ad-hoc networks. In [6][7] the use of POMDP for spectrum selection in CR networks is proposed. Some other works rely on the restless multi-armed bandit problem to find the optimal policy of sensing channels so as to maximize the expected throughput [8]. In [9] a classification of decision making techniques is given depending on the a priori knowledge they have and their sensitivity to sensing errors.

Under the above framework, this paper focuses on the analysis of the trade-off existing between performance and observation requirements in a CR decision-making framework that exploits the use of the so-called *belief vector* to predict the environment dynamics. The belief vector assesses the probability that the radio environment is under specific conditions (e.g. interference levels) at a certain time based on past measurements. As long as the belief vector predicts with sufficient accuracy the existing conditions at the decision making time, proper decisions can be made with minimum requirements in terms of observations.

While the concept can be applied to different types of problems in CR networks (e.g. transmit power adjustment, route selection, etc.), the paper will focus on the spectrum selection problem in a scenario where multiple spectrum blocks exhibiting different interference conditions are available to establish a set of radio links. For that purpose, and starting from the prior work in [6][7], this paper proposes a novel formulation that incorporates, under a single framework, a general belief-based decision making approach that can be particularised to different observation strategies. Moreover, the paper will analyse different elements reflecting the dynamics of the environment, in particular the traffic generation patterns, in order to better assess the convenience of one or another observation strategy.

The rest of the paper is organized as follows. Section II presents the considered system model and formulates the general belief-based decision making approach. This general approach is particularized for different observation strategies in Section III. Section IV presents the considered simulation model to evaluate the proposed approach along with the performance results. Finally, Section V points out concluding remarks and future works.

II. System Model and Problem Formulation

The system model considered in this paper assumes a set of j=1,..., L radio links, each one intended to support data transmission between either a pair of terminals or between a terminal and an infrastructure node. The j–th radio link will be supporting a certain data service characterized by a required bit rate $R_{req,j}$ and will be generating data transmission sessions of a certain duration D_j.

The potential spectrum to be assigned to the different radio links is organized in a set of i=1,..., M spectrum blocks (SBs). Each one is characterized by a certain central frequency and bandwidth. In general, the SBs can belong to different spectrum bands subject to different interference conditions and to different regulatory regimes (e.g., unlicensed bands, shared bands, licensed bands, etc.). Interference in each SB is assumed to come from other transmitters that have the same rights of use than the considered radio links (e.g., in case of shared bands, such as TVWS, it is assumed that the interference comes from other secondary devices).

The available bit rate for the j-th link in the i-th SB $R_{j,i}$ will depend on both the propagation conditions between the j-th link transmitter and receiver as well as on the interference in the i-th SB experienced at the receiver. Then, the problem considered here consists in performing an efficient allocation of the SBs to the radio links by properly matching the bit rate requirements with the achievable bit rate in each SB. For that purpose, it is assumed that the different radio links are controlled by a centralized management entity residing at the infrastructure side in charge of deciding the spectrum to be used by each radio link. Some illustrative use cases where this system model can be applicable are: (i) a Digital Home scenario in which different devices need to communicate, (ii) a set of cognitive small cells deployed in a cellular network that make use of additional spectrum to increase the network capacity, and (iii) an opportunistic Device-To-Device (D2D) radio link created to extend the coverage of certain cellular terminals that are outside the coverage area of the cellular infrastructure.

The spectrum selection decision-making will take a so-called *action*, corresponding to the allocation of a SB to a radio link, anytime that a data transmission session is initiated on this radio link. The action made for the j–th link at time t is denoted as $a_j(t)\in\{1,..., M\}$ and corresponds to the selected SB among those currently available.

The considered interference model denotes as $I_{j,i}(t)=I_{max,j,i}\cdot\sigma_i(t)$ the interference spectral density measured by the receiver of the j-th link in the i-th SB at a given time due to other external transmitters (i.e. not belonging to the L radio links). In order to capture that interfering sources may exhibit time-varying characteristics, $\sigma_i(t)$ is a SB-specific term between 0 and 1 (i.e. $\sigma_i(t)$=0 when no interference exists and $\sigma_i(t)$=1 when the interference reaches its maximum value $I_{max,j,i}$).

For modelling purposes, it is considered that the set of possible values of $\sigma_i(t)$ is translated into a discrete set of interference states $S^{(i)}(t)\in\{0,1,..., K\}$ where state $S^{(i)}(t)=k$ corresponds to $\sigma_{k-1}<\sigma_i(t)\leq\sigma_k$ for k>0 and to $\sigma_i(t)=\sigma_0=0$ for k=0. Note also that $\sigma_K=1$.

The interference evolution for the i-th block is modelled as an ergodic discrete-time Markov process with the state transition probability from being in state k at time t and moving to state k' in the next time step t+1 given by:

$$p_{k,k'}^{(i)} = \Pr\left[S^{(i)}\left(t+1\right)=k' \middle| S^{(i)}\left(t\right)=k \right] \tag{1}$$

It is assumed that the state of the i-th SB, $S^{(i)}(t)$, evolves independently from the other SBs, and that the state evolution is independent from the assignments made by the spectrum selection algorithm. Note that, without loss of generality, the time axis is assumed to be discretized in time steps. Then, the state transition probability matrix for the i-th SB is defined as:

$$\mathbf{P}^{(\mathrm{i})} = \begin{bmatrix} p_{0,0}^{(i)} & p_{0,1}^{(i)} & \cdots & p_{0,K}^{(i)} \\ p_{1,0}^{(i)} & p_{1,1}^{(i)} & & p_{1,K}^{(i)} \\ \vdots & & \ddots & \vdots \\ p_{K,0}^{(i)} & p_{K,1}^{(i)} & \cdots & p_{K,K}^{(i)} \end{bmatrix} \tag{2}$$

Moreover, let define $\pi^{(\mathrm{i})}=[\pi_0^{(i)} \quad \pi_1^{(i)} \quad \cdots \quad \pi_K^{(i)}]^{\mathrm{T}}$, where superscript T denotes transpose operation, as the steady state probability vector whose k-th component $\pi_k^{(i)}$ is the probability that the i-th SB is in the k-th interference state.

Each radio link with a data session in course (referred to as an active link) will obtain a reward that measures the obtained performance depending on the interference state of the allocated SB at each time. Then, let denote $r_{j,k}^{(i)}$ the reward that the j-th link gets when using its allocated SB i and the interference state is $S^{(i)}(t)=k$. The reward is a metric between 0 and 1 capturing how suitable the i-th SB is for the j-th radio link, depending on the bit rate that can be achieved in this SB with respect to the bit rate required by the application $R_{req,j}$. Using vector notation, the reward vector of the j-th link in the different interference states of the i-th SB is defined as $\mathrm{r}_{\mathrm{j}}^{(\mathrm{i})}=[r_{j,0}^{(i)} \quad r_{j,1}^{(i)} \quad \cdots \quad r_{j,K}^{(i)}]^{\mathrm{T}}$. It is worth mentioning that many possible definitions of the reward metric as a function of the bit rate may exist (e.g. sigmoid functions, linear functions, etc.).

The average reward experienced on the j-th link and i-th SB along a session starting to transmit data at time t+1 and ending after a certain duration D_j time steps, is given by:

$$r_{SESSION,j}^{(i)} = \frac{1}{D_j}\sum_{n=1}^{D_j} r_{j,S^{(i)}(t+n)}^{(i)} \tag{3}$$

With all the above foundations, the spectrum selection policy executed at time t for the j-th radio link will target the maximization of the expected reward that the session will experience along its duration:

$$a_j(t) = \arg \max_{\substack{i\in\{1,\ldots,M\} \\ i \text{ available}}} E\left[\frac{1}{D_j}\sum_{n=1}^{D_j} r_{j,S^{(i)}(t+n)}^{(i)}\right] \quad (4)$$

where the selection is made among the subset of available SBs, i.e. those that are not allocated to any other radio link at the *decision* making time t.

The *analysis* of the future evolution of the reward in each of the SBs until the session end will exploit measurements (observations) of the interference state of the different SBs carried out at specific time instants in the past, together with the statistical characterisation of the interference dynamics in each SB. In particular, denoting as $o^{(i)}(t-m^{(i)})$ the *observation* (measurement) on the i-th SB conducted at time step $t-m$ that provides the value of the interference state of the i-th SB that was measured at time step $t-m^{(i)}$, i.e. $o^{(i)}(t-m^{(i)}) = S^{(i)}(t-m^{(i)})$, the criterion of (4) can be reformulated in order to exploit knowledge from the observations in the past as:

$$a_j(t) = \arg \max_{\substack{i\in\{1,\ldots,M\} \\ i \text{ available}}} \Phi_j^{(i)}(t) \quad (5)$$

where $\Phi_j^{(i)}(t)$ is the SB-dependent decision function to be maximized, given by:

$$\Phi_j^{(i)}(t) = \frac{1}{\overline{D_j}}\sum_{n=1}^{\overline{D_j}} E\left[r_{j,S^{(i)}(t+n)}^{(i)} \middle| o^{(i)}\left(t-m^{(i)}\right) = S^{(i)}\left(t-m^{(i)}\right)\right] \quad (6)$$

Notice that, since the session duration D_j will usually be random and unknown at the decision making time t, it has been characterised in (6) statistically in terms of its average value $\overline{D_j}$. In (6), the estimation of the expected reward achieved in the i-th SB at future time instants $t+n$ based on the past observation at $t-m^{(i)}$ will rely on the statistical characterization of the interference dynamics given by the so-called *belief vector*. It is defined as $\mathrm{b}^{(i)}(t) = [b_0^{(i)}(t) \;\; b_1^{(i)}(t) \;\; \cdots \;\; b_K^{(i)}(t)]^{\mathrm{T}}$ where component $b_k^{(i)}(t)$ is the conditional probability that the i-th block will be in state $S^{(i)}(t)=k$ at time t given the last observation of the actual interference state that was taken at time step $t-m^{(i)}$, that is:

$$b_k^{(i)}(t) = \Pr\left[S^{(i)}(t) = k \middle| o^{(i)}\left(t-m^{(i)}\right) = S^{(i)}\left(t-m^{(i)}\right)\right] \quad (7)$$

Then, the expected reward obtained in the i-th SB at time $t+n$ can be expressed in terms of the belief vector as:

$$E\left[r_{j,S^{(i)}(t+n)}^{(i)} \middle| o^{(i)}\left(t-m^{(i)}\right) = S^{(i)}\left(t-m^{(i)}\right)\right] = \\ = \mathbf{b}^{(\mathbf{i})\mathbf{T}}(t+n)\,\mathbf{r}_\mathbf{j}^{(\mathbf{i})} \quad (8)$$

By making use of (8) the estimation of the average reward achieved in the i-th SB along the session duration of the j-th radio link given by decision function (6) can be expressed in terms of the belief vector as:

$$\Phi_j^{(i)}(t) = \frac{1}{\overline{D_j}}\sum_{n=1}^{\overline{D_j}} \mathbf{b}^{(\mathbf{i})\mathbf{T}}(t+n)\,\mathbf{r}_\mathbf{j}^{(\mathbf{i})} \quad (9)$$

The computation of the belief vector of the i-th SB at a certain time instant t is done recursively starting from the last observation of the actual interference state that was taken at time step $t-m^{(i)}$. Specifically, considering that $o^{(i)}(t-m^{(i)}) = S^{(i)}(t-m^{(i)})$ the components of the belief vector at time $t-m^{(i)}$ are given by:

$$b_k^{(i)}\left(t-m^{(i)}\right) = \begin{cases} 1 & \text{if } \; k = S^{(i)}\left(t-m^{(i)}\right) \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

This is expressed in vector notation as:

$$\mathbf{b}^{(\mathbf{i})}\left(t-m^{(i)}\right) = \mathbf{x}\left(S^{(i)}\left(t-m^{(i)}\right)\right) \quad (11)$$

where $\mathbf{x}(k)$ is defined as a column vector of K+1 components numbered from 0 to K that has all of them equal to 0 except the k-th component that is equal to 1.

Then, the belief vector at a time instant $t > t-m^{(i)}$ can be obtained from the belief vector at the previous time step t - 1 making use of the state transition probability matrix as:

$$\mathbf{b}^{(\mathbf{i})\mathbf{T}}(t) = \mathbf{b}^{(\mathbf{i})\mathbf{T}}(t-1)\cdot\mathbf{P}^{(\mathbf{i})} \quad (12)$$

By recursively applying (12) for the last $m^{(i)}$ time steps and by making use of (11) the belief vector at time t as a function of the last observation is given by:

$$\mathbf{b}^{(\mathbf{i})\mathbf{T}}(t) = \mathbf{b}^{(\mathbf{i})\mathbf{T}}\left(t-m^{(i)}\right)\left[\mathbf{P}^{(\mathbf{i})}\right]^{m^{(i)}} = \mathbf{x}^{\mathbf{T}}\left(S^{(i)}\left(t-m^{(i)}\right)\right)\left[\mathbf{P}^{(\mathbf{i})}\right]^{m^{(i)}} \quad (13)$$

III. Observation Strategies

The observation strategy specifies the time instants when the actual interference state in each SB is measured. The observation strategy should make sure that the time $m^{(i)}$ elapsed between the last observation and the spectrum selection decision making time t is adequate enough to compute the belief vector $\mathbf{b}^{(\mathbf{i})\mathbf{T}}(t+n)$ and make accurate decisions based on (9). Then, this paper considers the following three observation strategies:

a) *Instantaneous Measurements* (*IM*) *strategy*: This strategy consists in performing instantaneous measurements of the interference states in all the spectrum blocks at the time t when a new session has to be established, i.e. at the time when the spectrum selection decision-making is executed. In this specific case, the belief vector will always be computed with $m^{(i)} = 0$ and therefore it will capture the exact interference state at time t. Then, the belief vector at t will be given by:

$$\mathbf{b}^{(\mathrm{i})\mathrm{T}}(t)=\mathbf{x}^{\mathrm{T}}\left(S^{(i)}(t)\right) \tag{14}$$

Correspondingly, the decision function becomes:

$$\Phi_j^{(i)}(t)=\frac{1}{\overline{D_j}}\mathbf{x}^{\mathrm{T}}\left(S^{(i)}(t)\right)\left(\sum_{n=1}^{\overline{D_j}}\left[\mathbf{P}^{(\mathrm{i})}\right]^n\right)\mathbf{r}_{\mathbf{j}}^{(\mathbf{i})} \tag{15}$$

b) *Periodic Measurements (PM) strategy*: This strategy consists in performing periodic measurements of the *i*-th SB with observation period $T_{obs}^{(i)}$. In this way, the elapsed time $m^{(i)}$ between the last observation of the *i*-th SB and the decision making time *t* will always be upper bounded by $m^{(i)} \le T_{obs}^{(i)}$. As a further refinement of this periodic approach, it will be assumed that only the SBs that are not allocated to any link will be measured, since they are the only SBs that can be considered in the decision making process. In turn, when a SB is released, it will also be measured in case that the time since the last observation exceeds $T_{obs}^{(i)}$.

c) *Steady-state* (*StS*) *strategy*: This is the simple case in which no actual observations are performed. In this case, $m^{(i)} \to \infty$ and it can be easily proved making use of the properties of ergodic discrete time Markov processes [10] that the values of the belief vector will be equal to the steady-state probabilities $\boldsymbol{\pi}^{(\mathrm{i})\mathrm{T}}$. Then the general decision function becomes:

$$\Phi_j^{(i)}(t)=\lim_{m^{(i)}\to\infty}\frac{1}{\overline{D_j}}\sum_{n=1}^{\overline{D_j}}\mathbf{b}^{(\mathrm{i})\mathrm{T}}(t+n)\mathbf{r}_{\mathbf{j}}^{(\mathbf{i})}= \\ =\frac{1}{\overline{D_j}}\sum_{n=1}^{\overline{D_j}}\boldsymbol{\pi}^{(\mathrm{i})\mathrm{T}}\mathbf{r}_{\mathbf{j}}^{(\mathbf{i})}=\boldsymbol{\pi}^{(\mathrm{i})\mathrm{T}}\mathbf{r}_{\mathbf{j}}^{(\mathbf{i})} \tag{16}$$

IV. PERFORMANCE EVALUATION

A. Simulation parameters

A set of M = 5 SBs has been considered. Blocks B1 and B5 belong to the ISM band at 2.4 GHz with bandwidth 20 MHz. SBs B2, B3 and B4 belong to the white spaces in the TV band operated at frequencies 400, 800 and 600 MHz, respectively. Their bandwidths are 16, 24 and 16 MHz, respectively. Three different interference states are considered for the five SBs. The average durations of these states for each SB are presented in Table I.

TABLE I. DURATIONS OF THE INTERFERENCE STATES

State	B1	B2	B3	B4	B5
$S^{(i)}$=0	480 steps	120 steps	480 steps	240 steps	360 steps
$S^{(i)}$=1	120 steps	120 steps	120 steps	120 steps	160 steps
$S^{(i)}$=2	120 steps	480 steps	160 steps	60 steps	60 steps

A set of L = 3 links has been considered. For each link session durations and session generation rates of the data transmissions are exponentially distributed with averages, $\overline{D_j}$ (time steps) and ρ_j (sessions/time step), respectively, which are varied in the different simulations. The bit rate requirement for the link 1 is 200 Mb/s, while for links 2 and 3 it is 100 Mb/s. The reward considered in the simulations makes use of the formulation defined in [6]. In particular, the reward increases with the available bit rate up to the maximum at $R_{req,j}$ and then it starts to smoothly decrease reflecting that it becomes less efficient from a system perspective to have an available bit rate much higher than the required one. Table II presents the values of the achievable bit rates and associated rewards $r_{j,k}^{(i)}$ for each link in the different SBs and interference states. Moreover, different values of T_{obs} for PM strategy and simulation time T_{SIM}=10000 time steps have been taken into account.

TABLE II. BIT RATES AND REWARD VALUES OF THE LINKS IN THE DIFFERENT SPECTRUM BLOCKS

Link	SB	State $S^{(i)}$=0		State $S^{(i)}$=1		State $S^{(i)}$=2	
		$R_{j,i}$ (Mb/s)	$r_{j,0}^{(i)}$	$R_{j,i}$ (Mb/s)	$r_{j,1}^{(i)}$	$R_{j,i}$ (Mb/s)	$r_{j,2}^{(i)}$
1	B1	264	0.92	150	0.85	87	0.21
	B2	297	0.86	246	0.95	87	0.21
	B3	365	0.74	308	0.84	73	0.11
	B4	281	0.89	228	0.98	70	0.10
	B5	264	0.92	69	0.09	20	0.00
2, 3	B1	145	0.87	40	0.16	8	0.00
	B2	204	0.68	151	0.85	12	0.00
	B3	263	0.55	184	0.68	6	0.00
	B4	185	0.73	132	0.92	6	0.00
	B5	145	0.87	4	0.00	0.45	0.00

B. Key Performance Indicators (KPIs)

The assessment of the proposed framework has been carried out in terms of the following KPIs:

- *Average system reward:* It is the reward experienced by the active links depending on their allocated SBs and corresponding interference state averaged along the total simulation time T_{SIM}. Moreover, the result is averaged for all the L links.
- *Average satisfaction probability:* It is the fraction of time that the established sessions in the links achieve a bit rate higher or equal than the requirement $R_{req,j}$.
- *Observation rate:* It is the average number of observations per step that are performed to determine the interference state of the different SBs. This KPI is only applicable to IM and PM policies, while StS strategy does not require observations of the system.

C. Results

The performance of the different strategies as a function of the session rates ρ is presented in Fig. 1, Fig. 2 and Fig. 3 in terms of average system reward, satisfaction probability for the different links and observation rate, respectively. As an additional baseline reference, the random algorithm is also considered, in which the SB is selected randomly among the available ones without making any observation. $\overline{D_j}$ = 15 time steps and T_{obs}=10, 50, 100 and 200 time steps in case of PM strategy are considered in these results. $\overline{D_j}$ and ρ are the same for all the links.

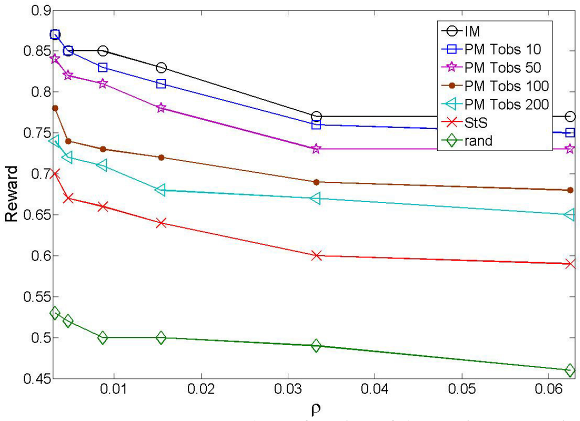

Fig. 1. Average system reward as a function of the session generation rate ρ

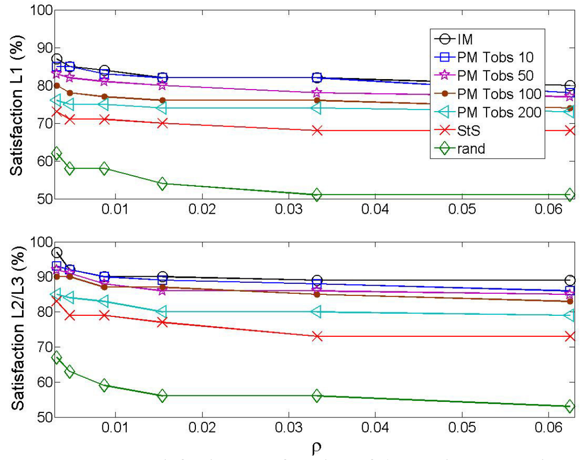

Fig. 2. Average satisfaction as a function of the session generation rate ρ

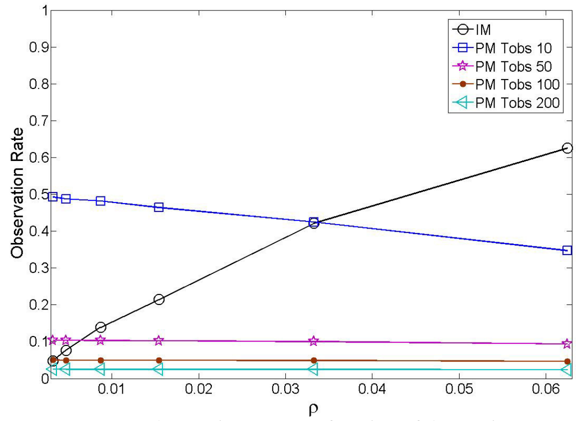

Fig. 3. Average observation rate as a function of the session generation rate ρ

Firstly it can be observed how all the proposed strategies allow achieving a clear improvement in terms of both reward and satisfaction probability with respect to the random selection of the SB. As for IM strategy, which makes decisions based on the most recent information at the decision making time, it achieves the best performance in terms of reward and satisfaction probability. However, it can be observed in Fig. 3 that the observation rate increases linearly with the session generation rate, because each time a new session arrives it requires performing observations on all the available SBs. Regarding StS strategy, from the perspective of observation rate it would be the best strategy because it does not require observations at all. However, its performance in terms of reward/satisfaction is highly degraded with respect to IM in the considered scenario (reward reductions of around 35% can be observed in Fig. 1) because decisions are made without considering the real interference state of the SBs.

Concerning PM strategy, the figures reflect that the proper setting of the observation period T_{obs} should result from the trade-off between reward and observation rate. Low values such as T_{obs}=10 lead to a large reward at the expense of an increase in the observation rate, as seen in Fig. 1 and Fig. 3, respectively. In turn, when increasing the value of T_{obs} the observation rate can be substantially reduced. However, this is at the expense of degrading the reward, because the time elapsed between the measurements and the decision making time increases with T_{obs} and thus the belief becomes less accurate. Taking as a reference the reward degradation with respect to IM, values of T_{obs} between 50 and 100 achieve a good trade-off between reward/observation rate in this scenario, because the reward reduction with respect to IM is around 10%, while a very significant reduction in terms of observation rate is achieved, particularly for large session generation rates ρ. It is also worth mentioning that, while the observation rate in IM increases linearly with the session generation rate ρ (see Fig. 3), in the case of PM it decreases slightly with ρ. This slight decrease, particularly noticeable in Fig. 3 for the case T_{obs}=10, is due to the fact that observations in PM are only carried out in the SBs that are not allocated to any link, so observation rate decreases when increasing the SB occupation (i.e. when increasing ρ).

Fig. 4 and Fig. 5 present the performance comparison between the different strategies as a function of the average session duration $\overline{D_j}$. In this case T_{obs}=50 and 100 time steps are considered. The time between the end of a session and the beginning of the next one is exponentially distributed with average 50 time steps, so the session generation rate will be $\rho=1/(50+\overline{D_j})$. From the figures it can be noticed how, for short session durations the comparison between the different techniques leads to similar conclusions as those obtained in Fig. 1 to Fig. 3. However, for long session durations it is observed that IM, PM and StS techniques tend to converge towards similar values of the reward. The reason is that, when a SB is allocated to a link for a long time, the link will tend to experience the steady-state conditions in this SB. Therefore, the reward estimation based on the steady-state probabilities made by the StS at the decision making time becomes a good estimate of the actual performance that will be achieved. Correspondingly, for long session durations, StS becomes the most adequate strategy because it is capable of properly estimating the performance without requiring any observations.

Then, based on the obtained results it can be concluded that the traffic generation pattern plays a key role when deciding the most adequate observation strategy in a belief-based decision making approach. On the one hand, for long session durations (in the considered scenario for $\overline{D_j}$ higher than approximately 300 time steps), the best approach is the decision making based on steady-state conditions because it allows properly estimating the performance without requiring dynamic observations of the environment. On the contrary, for shorter session durations the choice between IM and PM is related to the session arrival rate ρ that reflects the rate at which the spectrum selection functionality is triggered. In particular, a belief-based decision making with periodic observations becomes a good approach for large session generation rates ρ as long as the observation period is properly set, because it allows achieving good performance in terms of reward while significantly reducing the observation rate requirements. In the considered scenario, a proper setting of the observation period is between 50 and 100 time steps, and in this case, the periodic approach performs better than IM for session generation rates ρ approximately above 0.008, while for lower session generation rates it is more convenient the IM approach.

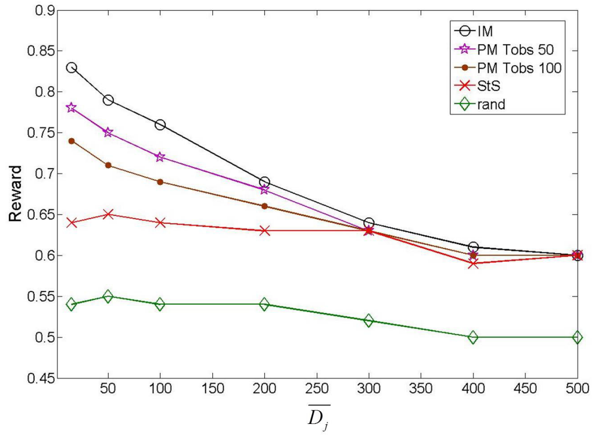

Fig. 4. Average system reward as a function of the average session duration $\overline{D_j}$

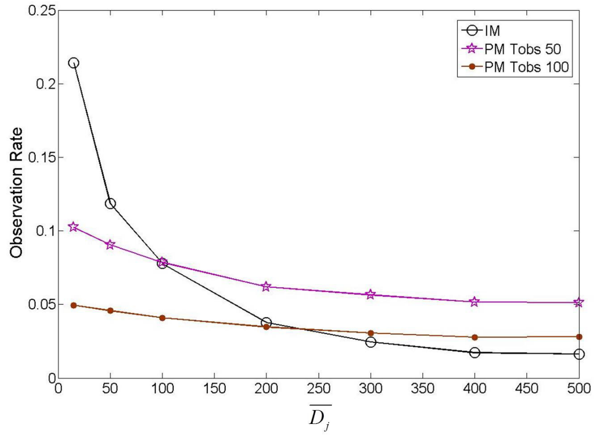

Fig. 5. Average observation rate as a function of the average session duration $\overline{D_j}$

V. CONCLUSIONS

This paper has presented a belief-based framework for decision making in CR networks, focusing on the spectrum selection problem where a number of radio links with different requirements have to be established. It exploits the belief vector concept to predict the environment dynamics at the decision making time and in later instants based on past measurements. In this context, the paper has analyzed the trade-off existing between performance and observation requirements of the cognitive cycle. For that purpose, a general formulation of the belief-based decision making has been presented and has been particularized for different observation strategies. They have been evaluated to assess the impact of the environment dynamics in terms of the traffic generation patterns. Results have demonstrated that, for long session durations a steady state-based strategy that does not require dynamic observations becomes the best approach. In turn, for short session durations the use of periodic measurements achieves a good trade-off between reward and observation rate for large session generation rates, while for low session generation rates the use of instantaneous measurements made at the decision making time becomes adequate.

Future work will deal with studying also the impact of the dynamicity of the radio environment on the observation strategies in terms of the interference states durations. Moreover, the proposed framework will be implemented in a USRP-based platform [11] that already provides some needed functionalities such as spectrum sensing and dynamic spectrum selection.

ACKNOWLEDGMENT

This work has been supported by the Spanish Research Council and FEDER funds under ARCO grant (ref. TEC2010-15198).

REFERENCES

[1] J. Mitola III, “Cognitive radio: an integrated agent architecture for software defined radio,” Ph.D. dissertation, KTH Royal Institute of Technology, 2000.

[2] ETSI TR 102 684 v1.1.1, “Reconfigurable Radio Systems (RRS); Feasibility Study on Control Channels for Cognitive Radio Systems”, April 2012.

[3] V. Stavroulaki, et al. “Cognitive Control Channels: From Concept to Identification of Implementation Options”, IEEE Communications Magazine, Vol. 50, No. 3, pp. 96-108, July 2012.

[4] K.P. Murphy, “A Survey of POMDP Solution Techniques”, available at http://http.cs.berkeley.edu/~murphyk/Papers/pomdp.ps.gz, 2000.

[5] Q. Zhao, L. Tong, A. Swami, 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.

[6] A. Raschellà, J. Pérez-Romero, O. Sallent, A. Umbert, “On the use of POMDP for Spectrum Selection in Cognitive Radio Networks”, CROWNCOM 2013, Washington DC, United States, July 2013.

[7] A. Raschellà, J. Pérez-Romero, O. Sallent, A. Umbert, “On the impact of the Observation Strategy in a POMDP-based framework for Spectrum Selection”, PIMRC 2013, London, UK, September 2013.

[8] K. Wang, L. Chen, Q.Liu, K. Al Agha, “On Optimality of Myopic Sensing Policy with Imperfect Sensing in Multi-Channel Opportunistic Access”, IEEE Transactions on Communications, Vol. 61, No. 9, pp. 3854-3862, September 2013.

[9] W. Jouini, C. Moy, J. Palicot, “Decision making for cognitive radio equipment: analysis of the first 10 years of exploration”, EURASIP Journal on Wireless Communications and Networking, 2012:26, pp. 1-16.

[10] L. Kleinrock, *Queueing Systems. Volume I: Theory*, John Wiley & Sons, 1975.

[11] A. Raschellà, A. Umbert, J. Pérez-Romero, O. Sallent, “A Testbed Platform to Demonstrate Spectrum Selection in Opportunistic Networks”, WMNC 2013.