On the use of POMDP for Spectrum Selection in Cognitive Radio Networks

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

***Abstract*—Dynamic Spectrum Access is a key capability of Cognitive Radio (CR) networks to increase the efficiency in the use of the available spectrum resources. In this respect, this paper focuses on the spectrum selection problem when a number of radio links has to be established in a CR network to support applications with different bit rate requirements. A novel strategy based on a Partially Observable Markov Decision Process (POMDP) is proposed, whose target is to maximize a reward function that reflects the suitability of the available spectrum blocks to the application requirements. The proposed strategy combines partial observations of the interference state in the different spectrum blocks together with a statistical characterization of the interference dynamics. Thanks to this feature, the performance comparison of the algorithm against different reference strategies reveals that it achieves a very similar performance than a strategy operating under full knowledge of the real interference state of all the spectrum blocks, while at the same time it has much less requirements in terms of measurement needs and associated signaling.**

Keywords-spectrum selection; Partially Observable Markov Decision Process (POMDP)

I. INTRODUCTION

Spectrum management is defined as the process of developing and executing policies, regulations, procedures, and techniques used to allocate, assign, and authorize frequencies in the radio spectrum to specific services and users. Regulatory bodies at international, European and national levels are actively working towards efficient and flexible spectrum regulation by fostering technology and service neutral spectrum management, spectrum trading and promotion of collective use of spectrum as well as shared use of spectrum [1]. In such regulatory framework, spectrum usage efficiency can be enhanced through the combination of Dynamic Spectrum Access (DSA) and CR (Cognitive Radio) technology [2][3]. CR has emerged as an intelligent radio that automatically adjusts its behavior based on the active monitoring of its environment. In that respect, spectrum selection refers to choosing the most appropriate portion of radio electrical spectrum to be used in DSA/CR communication systems. Several research works have addressed the spectrum selection problem highlighting the importance of having efficient decision-making criteria. Some of these works rely on databases that record historical information about the occupation in the different channels [4][5]. This type of information can be used to build predictive models on spectrum availability [6]. In [7] an adaptive spectrum decision framework is presented taking into account different type of applications while in [8] a radio resource management method using both long and short term history information is analysed. Finally, in [9] the use of reinforcement learning for the detection of spectral resources in a multi-band CR scenario was investigated.

In order to perform an efficient spectrum selection, the cognitive cycle paradigm that includes observation, analysis, decision and action is exploited in this paper. The observation of the radio environment and the analysis of such observations will lead to acquire knowledge about the state of the potential spectrum blocks that can be selected (e.g. the amount of measured interference, their occupation, etc.) as well as their dynamic behavior (e.g. how the interference changes with time). Observations of the radio environment typically involve making measurements at the terminal side and reporting back to the infrastructure side, then resulting very costly in terms of signaling overhead, battery consumption, etc. Consequently, decision-making strategies able to efficiently operate with the minimum amount of measurements would be of high interest. In this respect, Partially Observable Markov Decision Processes (POMDPs) [10] become a powerful decision making tool since they allow achieving an optimized performance by combining observations at specific periods of time with a statistical characterization of the system dynamics. Some works in the literature have used POMDPs in similar contexts. In [11] an opportunistic spectrum access approach to channels that can be either busy or idle is proposed, assuming a single unlicensed user. In [12] the problem was extended to a multi-user scenario through a collaborative approach in which users need to exchange information about their belief vectors at each time slot to generate consistent actions.

In this framework, this paper proposes an algorithm that enables an efficient spectrum selection in the presence of external interference variations in different candidate spectrum blocks. The proposed solution considers a centralized entity, which is in charge of deciding the appropriate spectrum block to be assigned to a number of radio links intended to support different applications with specific bit rate requirements. The problem is formulated as a POMDP in which the agent responsible for the spectrum selection decisions does not have a full knowledge about the state of all the available spectrum portions, but it relies only on observations at some instants.

The contributions of this paper and novelties with respect to previous works in this area can be summarized as follows:

CROWNCOM 2013, July 08-10, Washington DC, United States

DOI 10.4108/icst.crowncom.2013.251983

(i) a POMDP framework for multi-band spectrum selection is presented that, as a difference from previous works, such as [11][12], does not rely only on binary (i.e., idle/occupied) measurements but it considers a generalization in which the temporal variation of each spectrum block is able to capture different degrees of interference; *(ii)* the proposed framework inherently considers heterogeneity of the requirements for the different users accessing the spectrum, so that not all the channels are equally appropriate depending on application needs; *(iii)* the proposed framework captures the multi-user perspective in a centralized way hence having a single decision making point and avoiding the inter-terminal information exchange required in other collaborative decentralized approaches such as [12].

The rest of the paper is organized as follows: in Section II the system model is described and the considered spectrum selection problem is formulated as a POMDP, presenting the corresponding spectrum selection policy. Section III presents the considered simulation model to evaluate the proposed approach. Results are presented in Section IV. Finally, Section V points out concluding remarks and future works.

II. System Model and Problem Formulation

Fig. 1 illustrates the system model together with the functional entities related with the spectrum selection problem considered in this paper. The system is characterized by a set of links j=1,..., L each one intended to support data transmission between a pair of terminals and/or infrastructure nodes. The radio link j will be characterized by a required bit rate $R_{req,j}$. The different links are controlled by a centralized management entity residing at the infrastructure side.

The potential spectrum to be assigned to the different radio links is organized in a set of i=1,..., M spectrum blocks. Each one is characterized by a central frequency and a certain bandwidth. From a general perspective, the spectrum blocks can belong to different spectrum bands subject to different interference conditions.

The available bit rate for the j-th link in the i-th spectrum block $R_{j,i}$ will depend on both the propagation conditions between the j-th link transmitter and receiver as well as on the interference experienced by the receiver in the i-th block. Then, the spectrum selection problem considered here consists in performing an efficient allocation of the spectrum blocks to the radio links by properly matching the bit rate requirements with the achievable bit rate in each spectrum block.

As illustrated in Fig. 1, the spectrum selection decision making is executed in a centralized entity in the infrastructure node that controls the existing links in the network. The overall process follows the steps of the classical cognitive cycle, in which the spectrum selection decisions are supported by the information stored in a Knowledge Database (KD) that includes the knowledge resulting from the analysis of the measurements (observations) made on the different spectrum blocks. Decisions made are translated into actions to configure the existing links with the corresponding spectrum allocation.

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 spectrum block at a given time due to other external transmitters (i.e. outside the control of the decision making entity). In order to capture that interfering sources may exhibit time-varying characteristics, $\sigma_i(t)$ is a spectrum block-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 $I_{max,j,i}$).

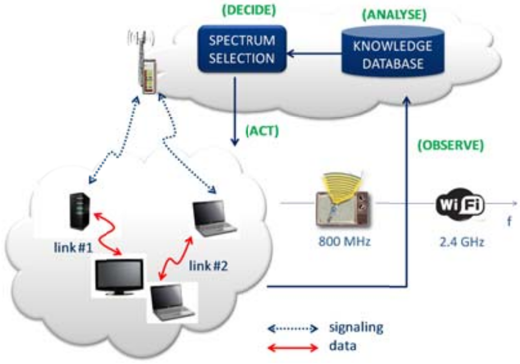

Figure 1. System Model.

For modeling 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 σ_K=1. The system state at time t is then given by the M-column vector $\mathbf{S}(\mathrm{t})=[S_i(t)]$.

Moreover, assuming that the state of each spectrum block remains the same for a period Δt, the interference evolution for the i-th block is modeled as a discrete-time Markov process with the state transition probability from state k to k' given by:

$$p^i_{k,k'}=\Pr\left[S_i\left(t+\Delta t\right)=k'\middle|S_i\left(t\right)=k\right] \quad (1)$$

It is assumed that the state of the i-th spectrum block $S_i(t)$ evolves independently from the other blocks, and that the state evolution is independent from the assignments made by the spectrum selection algorithm.

The execution of the spectrum selection decision-making algorithm results into *actions* corresponding to the allocation of spectrum blocks to the different radio links. The action made for link j at time t is denoted as $a_j(t)\in\{1,..., M\}$ and corresponds to the selected spectrum block among those currently available (i.e. not allocated to other links). It is assumed that an action is taken for a given link at any time that a data transmission session is initiated on this radio link.

As a consequence of the different actions and resulting spectrum block assignments, each radio link with a data session in course (i.e., an active link) will obtain a reward that measures the obtained performance depending on the interference state of the spectrum block at each time. Then, let denote $r_{j,i,S_i(t)}$ the reward that the j-th link gets at time t when using its allocated spectrum block i and the interference state is $S_i(t)$. The total system reward $T_R(t)$ is then given by the sum of rewards of all the active links at time t.

As a general target, the spectrum selection decision making should follow the optimal policy that maximizes the performance in terms of the expected long-term total system reward $T_R(t)$ accumulated over a certain time horizon tending to infinity. For this purpose, the decision-making entity would ideally need to know the actual interference state of all the spectrum blocks at time t. However, this would impact in terms of increasing signaling overheads and battery consumption to perform all the required observations (i.e., measurements) and report them to the decision-making entity. To overcome this issue, this paper proposes to make the decisions based on a statistical characterization of the interference state of the different spectrum blocks rather than on actual exhaustive observations. In the proposed solution, observations about the interference state of the spectrum blocks are carried out only at specific time instants defined according to a certain observation strategy. In this case, due to the partial knowledge that the decision making process has about the actual interference state of the spectrum blocks, the spectrum selection process can be modeled as a POMDP and the statistical characterization of the spectrum blocks at time t is given in terms of the so-called belief vector $\Upsilon(t)=[b_{i,k}(t)]$ where component $b_{i,k}(t)$ is the probability that the i-th block will be in state $S_i(t)=k$ at time t.

In a POMDP the complexity associated to finding the optimal policy that maximizes the expected long-term system reward is usually prohibitive, mainly because the number of states $(K+1)^M$ grows exponentially with the number of spectrum blocks. Consequently, this paper proposes to use instead the so-called *Myopic Policy* that maximizes the immediate system reward $T_R(t+\Delta t)$. It is worth mentioning that myopic policies have been found in some works to be optimal under certain conditions [13]. More specifically, considering that the spectrum block selection is made in time t for just one link j and among the set of available blocks so the selection will not impact on the immediate reward of any other link, the myopic spectrum selection policy becomes:

$$a_j\left(t\right)=\arg\max_{\substack{i\in\{1,\ldots,M\}\\ i\text{ available}}} E\left[T_R\left(t+\Delta t\right)\right]=\arg\max_{\substack{i\in\{1,\ldots,M\}\\ i\text{ available}}} E\left[r_{j,i,S_i(t+\Delta t)}\right] \quad (2)$$

The expected reward $E\left[r_{j,i,S_i(t+\Delta t)}\right]$ is computed using the belief vector values at time t and the state transition probabilities that the spectrum block i is in state k at time t and jumps to state k' in the next period $t+\Delta t$. Then, the decision policy is formulated as:

$$a_j\left(t\right)=\arg\max_{\substack{i\in\{1,\ldots,M\}\\ i\text{ available}}} \sum_{k=0}^{K} b_{i,k}\left(t\right)\sum_{k'=0}^{K} p_{k,k'}^{i}\cdot r_{j,i,k'} \quad (3)$$

The reward is a metric between 0 and 1 capturing how suitable the i-th spectrum block is for the j-th radio link/application, depending on the bit rate that can be achieved in this block with respect to the bit rate required by the application $R_{req,j}$. Based on the formulation defined in [14], the reward function considered in this paper is given by:

$$r_{j,i,k}=\frac{1-e^{-\frac{\Gamma\cdot U_{j,i,k}}{(\xi-1)^{1/\xi}\left(R_{j,i,k}/R_{req,j}\right)}}}{\lambda} \quad (4)$$

where $R_{j,i,k}$ denotes the achievable bit rate by the j-th link in the i-th spectrum block given that it is in state k. The relationship between achievable bit rate and interference state is a decreasing function assumed to be known for each link. $U_{j,i,k}$ is the following utility function that relates the achievable and the required bit rates:

$$U_{j,i,k}=\frac{\left(\xi-1\right)\left(R_{j,i,k}/R_{req,j}\right)^{\xi}}{1+\left(\xi-1\right)\left(R_{j,i,k}/R_{req,j}\right)^{\xi}} \quad (5)$$

Γ and ξ are shaping parameters to capture different degrees of elasticity with respect to the bit rate requirements and λ is a normalization factor given by:

$$\lambda=1-e^{-\frac{\Gamma}{(\xi-1)^{1/\xi}+(\xi-1)^{(1-\xi)/\xi}}} \quad (6)$$

The proposed formulation of the reward function $r_{j,i,k}$ increases with the available bit rate $R_{j,i,k}$ 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.

Based on all the above, the implementation of the spectrum selection decision making following (3) requires that the KD in Fig. 1 stores the state transition probabilities for the different spectrum blocks $p_{k,k'}^{i}$, the values of the reward $r_{j,i,k}$ that the different radio links can obtain in each spectrum block for each interference state, and the belief vector values $b_{i,k}(t)$.

Concerning $p_{k,k'}^{i}$ and $r_{j,i,k}$, they could in practice be obtained based on some initial acquisition mechanisms including measurements of the different links and spectrum blocks. The details on how to perform this acquisition as well as the capability to update the stored values whenever relevant changes are detected are out of the scope of this paper. Just as a reference, some previous works that have addressed the dynamic acquisition of unknown transition probabilities in POMDP systems are [15][16].

Concerning the belief vector values $b_{i,k}(t)$, they should be dynamically updated with time resolution Δt in accordance with the discrete-time Markov process that models the interference state in each spectrum block. To perform this update, the knowledge about the real interference of the spectrum blocks obtained through observations (i.e. measurements) performed at certain time instants can be exploited to obtain a more accurate estimation of the probability that the i-th block will be in state k at a later time. More precisely, let define as $o_i(t)$ the observation made at time t in the spectrum block i. This observation provides the actual interference state of the spectrum block, that is $o_i(t)=k$. Using the available observations $o_i(t)$, the values of $b_{i,k}(t)$ are updated for all the spectrum blocks every Δt as follows:

$$b_{i,k'}\left(t+\Delta t\right)=\begin{cases} p_{k,k'}^{i} & \text{if } (o_i\left(t\right)=k) \\ \sum_{n=0}^{K} p_{n,k'}^{i}\cdot b_{i,n}\left(t\right) & \text{otherwise}\end{cases} \quad (7)$$

The first condition in (7) corresponds to the spectrum blocks for which an observation is performed at time t providing the actual interference state of the spectrum block (i.e. o_i $(t)=k$). Then, the probability $b_{i,k'}(t+\Delta t)$ that spectrum block *i* will be in state *k'* in the next time period $t+\Delta t$ is simply given by the state transition probability $p^{i}_{k,k'}$. In turn, the second condition in (7) corresponds to those spectrum blocks for which no observation has been performed at time *t*. In this case, the actual interference state is not known and thus the value $b_{i,k'}(t+\Delta t)$ is computed probabilistically from the belief values $b_{i,n}(t)$ and the state transition probabilities to state *k*'.

According to the above, an observation strategy is required to determine the time instants in which the observations of the different spectrum blocks are carried out. In the context of this paper, it is assumed that observations are executed periodically every T_{obs} for all spectrum blocks.

It is worth mentioning that this paper assumes that the network operates in a stationary environment, so that the values of the state transition probabilities and the rewards for the different links/spectrum blocks do not change. In case of non-stationary environments, some additional mechanisms would be needed to detect that the operational conditions of the network have changed and to trigger the necessary acquisition mechanisms to obtain the new values of these parameters. However, such mechanisms are out of the scope of this paper and are left for future work.

III. Evaluation Scenario

This section describes the specific scenario and simulation assumptions that have been considered to evaluate the performance achieved by the proposed algorithm.

A. Simulation parameters

A set of M = 5 spectrum blocks have been considered. Blocks B1 and B5 belong to the ISM band at 2.4 GHz with bandwidth 20 MHz. Spectrum blocks B2, B3 and B4 belong to the white spaces in the TV band operated opportunistically 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 spectrum blocks. The average durations of these states for each spectrum block are presented in Table I.

TABLE I. Durations of the Interference States for the different Spectrum Blocks

State	B1	B2	B3	B4	B5
S_i=0	40 min	4 min	4 min	40 min	32 min
S_i=1	12 min	4 min	12 min	4 min	4 min
S_i=2	12 min	40 min	24 min	12 min	4 min

A set of L = 3 links is considered in the evaluation. Each link generates sessions whose duration is exponentially distributed with average T=30 s. The time between the end of a session and the beginning of the next one is also exponentially distributed with average T_{inter} varied in different simulations. The bit rate requirement for the link 1 is $R_{req,1}$=200 Mb/s, while for links 2 and 3 it is $R_{req,2}$= $R_{req,3}$=100 Mb/s. Table II presents the values of the achievable bit rates $R_{j,i,k}$ and associated rewards $r_{j,i,k}$ for each link in the different spectrum blocks and interference states. Parameters Γ=1 and ξ=5 have been considered to compute the reward.

The periodicity T_{obs} between observations in the POMDP-based approach is varied in the different simulations. Performance has been obtained with the simulator operating in steps of Δt=1 s during T_{SIM}=604800 time steps.

TABLE II. Bit rates (Mb/s) and Reward values of the Links in the different spectrum blocks

Link	Spectrum Block	State S_i=0		State S_i=1		State S_i=2	
		$R_{j,i,0}$	$r_{j,i,0}$	$R_{j,i,1}$	$r_{j,i,1}$	$R_{j,i,2}$	$r_{j,i,2}$
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. Benchmarking

The performance of the proposed POMDP-based approach following selection policy of (3) has been compared against the following reference strategies:

- *Full Observation spectrum selection algorithm (FO)*. This algorithm performs an observation of the actual interference state $S_i(t)$ for all the available spectrum blocks (i.e. those that are not allocated to any link) whenever a new link establishment is required. Then, the spectrum block that provides the highest reward is allocated, that is:

$$a_j(t) = \arg \max_{\substack{i\in\{1,\dots,M\} \\ i \text{ available}}} r_{j,i,S_i(t)} \tag{8}$$

- *Steady state probabilities-based spectrum selection algorithm (PR).* This algorithm makes the decisions based on static information stored in the database about the statistic behavior of the different spectrum blocks. This is captured by means of the steady state probabilities π^{i}_{k} that measure the probability that the spectrum block *i* will be in state *k*. The decision policy is then formulated in a similar way as for the POMDP-based algorithm in (3) but with the static values of π^{i}_{k} instead of the dynamic values of the belief vector, that is:

$$a_j(t) = \arg \max_{\substack{i\in\{1,\dots,M\} \\ i \text{ available}}} \sum_{k=0}^{K} \pi^{i}_{k} \sum_{k'=0}^{K} p^{i}_{k,k'} \cdot r_{j,i,k'} \tag{9}$$

- *Random spectrum selection algorithm*. In this case, whenever a new link has to be established, the spectrum block is selected randomly among those that are not currently allocated to any other link.

C. Key Performance Indicators (KPIs)

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

- 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 $R_{req,j}$. The result is the average for all the links along the total simulation time.

- Average system reward: It is the reward obtained by the active links depending on their allocated spectrum blocks and corresponding interference state averaged along the total simulation time T_{SIM}. The result is averaged for all the L links, that is:

$$T_{R\_Avg} = \frac{1}{L}\sum_{j=1}^{L}\frac{1}{T_{ACT}(j)}\sum_{\substack{t=1 \\ \text{link j active at t}}}^{T_{SIM}} r_{j,i,S_i(t)} \qquad (10)$$

where $T_{ACT}(j)$ represents the total number of simulation time steps in which the j-th link has been active.

- Observation rate: It is the average number of observations per second that are performed to determine the interference state of the different spectrum blocks. This KPI is only applicable to FO and POMDP algorithms, while PR and Random strategies do not require observations of the system during their operation.

IV. Performance Evaluation Results

The performance of the different algorithms in terms of average reward and satisfaction probability is presented in Fig. 2 and Fig. 3, respectively, as a function of the observation period T_{obs} that is a key parameter of the POMDP-based spectrum selection algorithm (hence, the selected T_{obs} values do not affect the performance of the other strategies). T_{inter}=10 s is considered in these results. As it can be observed, the performance obtained by the POMDP-based algorithm is a decreasing function of the observation period T_{obs}. In particular, for low values of T_{obs} the reward and the satisfaction probabilities are very similar to the ones obtained by the FO strategy that has a perfect knowledge of the different spectrum blocks. However, this similar performance is achieved by the POMDP with much less requirements in terms of observations, as it can be noticed in Fig. 4 that depicts the observation rate as a function of T_{obs} for both FO and POMDP algorithms. For instance, when T_{obs} is 60 s, POMDP achieves a reduction of around 68% with respect to FO in terms of observation rate, while the performance of POMDP in terms of reward and satisfaction probability is only about 3% smaller than the performance achieved by FO. Further reductions in terms of observation rate can be achieved when increasing T_{obs}, as seen in Fig. 4.

Concerning the comparison against the PR algorithm, it can be observed in both Fig. 2 and Fig. 3 that for low values of the observation period, POMDP achieves a significant improvement in both the reward and the satisfaction (e.g. for T_{obs}=60 s there is an improvement of 32%). Then, for large values of T_{obs} the performance of both PR and POMDP tends to converge to similar values. The reason is that the dynamic update of the belief vector values for the POMDP-based algorithm according to (7) tends to converge towards the steady state probabilities when there are very large periods without any observations, that is $b_{i,k}(t) \rightarrow \pi_k^i$ for $T_{obs} \rightarrow \infty$. Consequently, in such a case the decision making criteria of POMDP and PR given by (3) and (9) become the same.

Finally, focusing now on the comparison against the Random algorithm, it can be observed in Fig. 2 and Fig. 3 that POMDP achieves a very significant improvement in terms of both reward and satisfaction probability, in the order of 43% and 46%, respectively, for the case of T_{obs}=60 s.

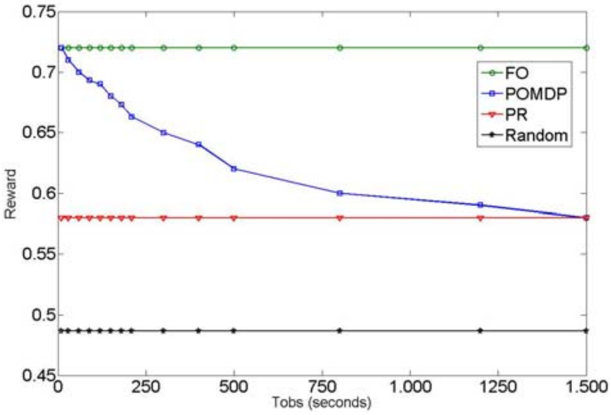

Figure 2. Average reward as a function of the observation period T_{obs}

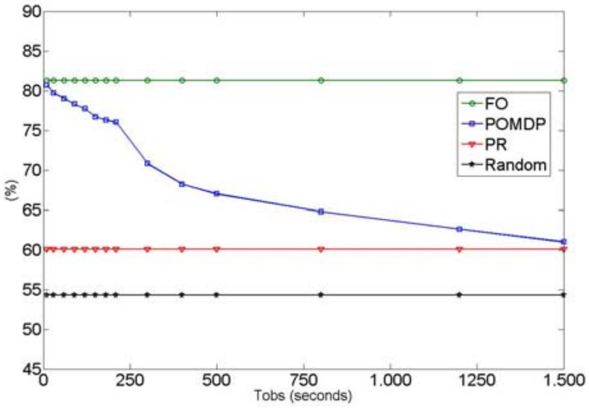

Figure 3. Average satisfaction probability as a function of the observation period T_{obs}

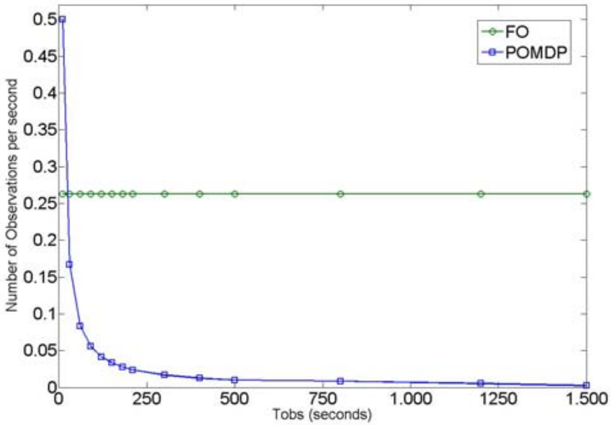

Figure 4. Observation rate as a function of the observation period T_{obs}

Fig. 5 and Fig. 6 present the impact of varying the time T_{inter} between the end of a session and the beginning of the next one, in terms of average reward and observation rate, respectively, for all the considered algorithms. For the POMDP-based algorithm, T_{obs}=60 s and T_{obs}=180 s are considered. It can be observed in Fig. 5 that the reward tends to increase with T_{inter} for all the strategies and that the reward obtained by the

POMDP strategy is still very close to the one achieved by FO, while requiring a much lower observation rate as seen in Fig. 6. Moreover, the POMDP allows still obtaining a significant improvement with respect to PR and Random algorithms for both values of T_{obs}.

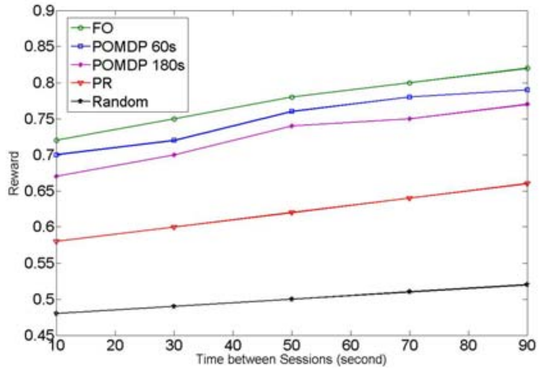

Figure 5. Average reward as a function of the time between sessions

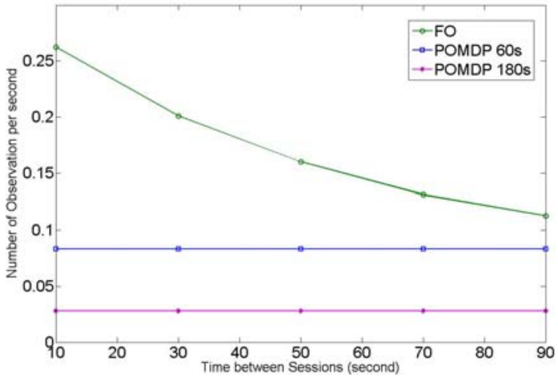

Figure 6. Observation rate as a function of the time between sessions

V. Conclusions

In this paper a novel POMDP-based framework for spectrum selection in wireless cognitive radio networks has been presented. The proposed framework operates in a centralized way making use of the knowledge stored in a database that contains the statistical characterization of the interference variations existing in the available spectrum blocks. The proposed approach inherently considers heterogeneity in the bit rate requirements of the applications to be established by maximizing a reward function that considers the different suitability of each spectrum block to each radio link/application. To highlight the efficiency of the proposed approach, a comparison has been performed against different references. It has been obtained that the proposed POMDP-based algorithm allows obtaining similar performance in terms of reward and satisfaction as the full observation scheme that makes decisions based on knowing the real interference state in all the available spectrum blocks, in spite of requiring a much lower measurement rate since only partial observations of the system at specific time instants are carried out. In addition it achieves a significant performance gain in terms of reward with respect to a random spectrum selection and to a strategy that makes decisions based on static knowledge of the spectrum block statistics.

Based on the promising results obtained, future work will deal with performing a further optimization of the observation strategy, with the inclusion of spectrum handover mechanisms in the proposed approach and with the development of strategies for dynamically acquiring and maintaining the state transition probability values stored in the knowledge database.

Acknowledgment

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

References

[1] IEEE Std 1900.1TM-2008, “IEEE Standard Definitions and Concepts for Dynamic Spectrum Access: Terminology Relating to Emerging Wireless Networks, System Functionality, and Spectrum Management”

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

[3] I.F. Akyildiz, W.-Y. Lee, M.C. Vuran, S. Mohanty, “Next generation/dynamic spectrum access/cognitive radio wireless networks: a survey”, Comput. Networks (Elsevier) 50 (13) (2006) 2127–2159

[4] J. Vartiainen, M. Hoyhtya, J. Lehtomaki, and T. Braysy, “Priority channel selection based on detection history database,” CROWNCOM 2010, June 2010.

[5] Y. Li, Y. Dong, H. Zhang, H. Zhao, H. Shi, and X. Zhao, “QoS provisioning spectrum decision algorithm based on predictions in cognitive radio networks,” WiCOM 2010, Sept. 2010.

[6] P. A. K. Acharya, S. Singh, and H. Zheng, “Reliable open spectrum communications through proactive spectrum access,” in IN PROC. OF TAPAS, 2006.

[7] W.-Y. Lee and I. Akyldiz, “A spectrum decision framework for cognitive radio networks,” Mobile Computing, IEEE Transactions, vol. 10, 2011, pp. 161–174.

[8] M. Höyhtyä, J. Vartiainen, H. Sarvanko, and A. Mämmelä, “Combination of short term and long term database for cognitive radio resource management,” 3rd International Symposium on Applied Sciences in Biomedical and Communication Technologies (ISABEL) , Nov. 2010, pp. 1–5 .

[9] U. Berthold, F. Fu, M. van der Schaar, F. K. Jondral, “Detection of Spectral Resources in Cognitive Radios Using Reinforcement Learning”, DySPAN 2008, Oct. 2008.

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

[11] 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, Apr. 2007.

[12] H. Liu, B. Krishnamachari, Q. Zhao, “Cooperation and Learning in Multiuser Opportunistic Spectrum Access”, ICC 2008, May 2008.

[13] Q. Zhao, B. Krishnamachari, K. Liu, “On Myopic Sensing for Multi-Channel Opportunistic Access: Structure, Optimality, and Performance”, IEEE Transactions on Wireless Communications, Vol. 7, No. 12, December, 2008, pp. 5431-5440.

[14] F. Bouali, O. Sallent, J. Pérez-Romero, R. Agusti, “Exploiting Knowledge Management for Supporting Spectrum Selection in Cognitive Radio Networks”, 7th International Conference on Cognitive Radio Oriented Wireless Networks (CrownCom) 2012, Stockholm, Sweden, June. 2012

[15] E. Fernández-Gaucherand, A. Arapostathis, S. I. Marcus, “On the Adaptive Control of a Partially Observable Markov Decision Process”, Proc. of the 27th Conference on Decision and Control, Austin, Texas, December, 1988.

[16] S. Ross “Bayes-Adaptive POMDPs: Toward an Optimal Policy for Learning POMDPs with Parameter Uncertainty”, Course Project report, 2006, available at http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.132.7043