On the impact of the Observation Strategy in a POMDP-based framework for Spectrum Selection

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, based on the Cognitive Radio (CR) paradigm, is a promising solution to improve the currently inefficient spectrum utilization. In this respect, this paper deals with the spectrum selection problem when a number of radio links has to be established in a CR network. A novel strategy based on a Partially Observable Markov Decision Process (POMDP) is proposed combining partial observations of the interference state together with a statistical characterization of the interference dynamics. In this context, the use of an efficient observation strategy is a key element to account for the trade-off between achieved performance and measurement requirements. For that purpose, the aim of the paper is to propose and evaluate different observation policies for the POMDP-based spectrum selection framework. Results show that the proposed framework is able to achieve very similar performance than a strategy operating under full knowledge of the interference state requiring much less associated signaling.**

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

I. INTRODUCTION

Spectrum management is 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 this context, spectrum usage efficiency can be enhanced through the combination of Dynamic Spectrum Access (DSA) and Cognitive Radio (CR) 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], which can be used to build predictive models on spectrum availability [6].

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 (SBs) 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) [7] 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 [8] an opportunistic spectrum access approach to channels that can be either busy or idle is proposed, assuming a single unlicensed user. In [9] 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. Based on the above, this paper formulates the spectrum selection problem in a scenario with heterogeneous application requirements and variable interference levels in the available SBs as a POMDP process. On the one side, as a difference from previous works in the literature, the framework presented in this paper is able to capture different levels of interference, while [8][9] are based on binary (i.e., idle/occupied) measurements. Moreover, the proposed approach considers the heterogeneity of requirements in the different applications to be supported, while in previous works the different suitability levels between spectral resources and application requirements have not been considered. This paper extends previous works from the authors in [10], where just a short contribution stating the main concepts was presented in the work in progress track, and [11], where the framework was comprehensively formulated and first results focusing on comparison against different references was performed. In that context, this paper represents a substantial step forward with respect to our previous works by focusing on the impact of different observation strategies, which constitute a key element in any POMDP framework that should balance the trade-off between measurement cost and achieved performance.

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. The considered observation strategies will be detailed in Section III. Section IV presents the considered simulation model to evaluate the proposed approaches. Results are presented in Section V. Finally, Section VI 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 potential spectrum to be assigned to the different radio links is organized in a set of $i=1,...,M$ SBs. Each one is characterized by a central frequency and a bandwidth. SBs can belong to different spectrum bands subject to different interference conditions.

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 experienced by the receiver in the i-th block. Then, the spectrum selection problem considered here consists in performing an efficient allocation of the SBs to the radio links by properly matching the required and achievable bit rates.

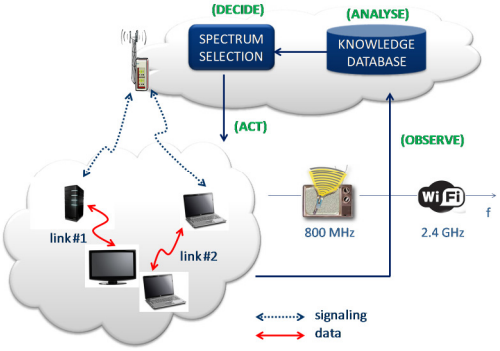

Figure 1. System Model.

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 observations (measurements) made on the different SBs. 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 SB at a given time due to other external transmitters. 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 $I_{max,j,i}$).

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 $\sigma_K=1$. Moreover, the interference evolution for the i-th block is modeled as a discrete-time Markov process that evolves in discrete time intervals of duration Δt with state transition probability from state k to k' given by:

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

It is assumed that the state of the i-th SB $S_i(t)$ 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 SBs 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 SB among those currently available. 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 SB assignments, each radio link with a data session in course will obtain a reward that measures the obtained performance depending on the interference state of the SB 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 SB 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 SBs at time t. However, this would impact in terms of increasing signaling overheads and battery consumption to perform all the required observations and report them to the decision-making entity. To overcome this issue, it is proposed to model the spectrum selection process as a POMDP that relies on (i) partial observations of certain SBs carried out at specific time instants defined according to an observation strategy, and (ii) a statistical characterization of the interference dynamics in the SBs 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 this context, the definition of a smart observation strategy becomes a key aspect to ensure that the knowledge of the current interference state is accurate enough to make the proper decisions, while at the same time reducing the cost associated to performing measurements.

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 SBs. Consequently, this paper proposes to use instead the so-called *Myopic Policy* that maximizes the immediate system reward $T_R(t+\Delta t)$. Myopic policies have been found in some works to be optimal under certain conditions [12]. More specifically, considering that the SB 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:

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

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 SB 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:

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

The reward is a metric between 0 and 1 capturing how suitable the i-th SB 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 [13], 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 SB 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:

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

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 a maximum of R_{req} 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 SBs $p_{k,k'}^{i}$, the values of the reward $r_{j,i,k}$ that the different radio links can obtain in each SB for each interference state, and the belief vector values $b_{i,k}(t)$.

Concerning $p_{k,k'}^{i}$ and $r_{j,i,k}$, they can be obtained based on some initial acquisition mechanisms including measurements of the different links and SBs. 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 [14][15].

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 SB. To perform this update, the POMDP exploits the knowledge about the real interference of the SBs obtained through observations performed at certain time instants according to the observation strategy. More precisely, let define as $o_i(t)$ the observation made at time t in the SB i that provides the actual interference state of the SB, that is $o_i(t)=k$. Then the values of $b_{i,k}(t)$ are updated for all the SBs every Δt as follows:

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

The first condition in (7) corresponds to the SBs for which an observation is performed at time t providing the actual interference state of the SB (i.e. $o_i(t)=k$). Then, the probability $b_{i,k'}(t+\Delta t)$ that SB i will be in state k' in the next time period $t+\Delta t$ is simply given by the state transition probability $p_{k,k'}^{i}$. In turn, the second condition in (7) corresponds to those SBs 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'.

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/SBs 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. OBSERVATION STRATEGIES

An observation strategy should specify the time instants when measurements have to be performed to obtain the actual interference state in the available SBs. If measurements are made very often, this will turn into a more accurate knowledge of the actual system state that will impact on making better decisions thus resulting in better performance. On the contrary, this will increase the cost in terms of sensing requirements and signaling overheads to report the measurement results. Hence, a trade-off arises between performance and measurement cost.

Another relevant aspect when deciding the observation strategy is to properly capture the dynamics of the different SBs to be measured. When the interference dynamics varies slowly a measurement taken at a certain time instant can be valid for a longer time horizon than when the interference exhibits a fast variation, which would require more frequent measurements to track the actual interference. Then, information about the dynamics in each SB is also a relevant issue of the observation strategy. To account for the above trade-offs, we consider the following observation strategies:

- *Full Observation strategy (FO).* In this strategy, it is assumed that an observation of the actual interference state $S_i(t)$ for all the available SBs is executed whenever a new link establishment is required. Hence, a perfect knowledge of the system state in all the SBs is available prior to the spectrum selection decision making. Note that in this case with full knowledge the decision making criterion does not rely on the POMDP approach but it simply allocates the SB that provides the highest reward, that is:

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

This strategy will be considered as the baseline for comparison with the POMDP-based approach.

- *Periodic Observation strategy (PO).* This strategy supports the POMDP-based decision making of (3) by means of observations performed periodically every T_{obs} in all the SBs that are not allocated to any link.
- *Adaptive Observation strategy (AO).* This strategy supports the POMDP-based decision making of (3) by means of observations whose periodicity is adaptively varied depending on the dynamics of each SB. In particular, assuming that the last observation made for the i-th SB at time t indicated that the real interference state was k, the next observation will be performed at time $t+T_{obs}(i,k)$. The period $T_{obs}(i,k)$ is computed based on the expected duration of the k-th interference state obtained from the transition probabilities for the i-th SB:

$$T_{obs}(i,k) = \rho \frac{\Delta t}{1 - p_{k,k}^{i}} \tag{9}$$

where ρ is a coefficient to be selected (0< ρ ≤1).

Moreover, in *PO* and *AO* approaches, measurements are made for non-used SBs, while allocated blocks will be measured at the time when they are released if the time elapsed since the last observation is higher than $T_{obs}(i,k)$.

IV. Evaluation Scenario

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

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	10 min	10 min	4 min	30 min	10 min
S_i=1	50 min	10 min	90 min	50 min	50 min
S_i=2	10 min	80 min	4 min	10 min	50 min

A set of L = 3 links has been considered. 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 10 s. The bit rate requirement for the link 1 is 200 Mb/s, while for links 2 and 3 it is 100 Mb/s. Other parameters are Γ=1 and ξ=5. Performance has been obtained in steps of Δt=1 s during T_{SIM}=604800 time steps.

B. 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 SBs and interference state averaged along the total simulation time T_{SIM}. The result is averaged for all the L links.
- Observation rate: It is the average number of observations per second that are performed to determine the interference state of the different SBs.

V. Performance Evaluation Results

The performance of the different strategies in terms of average reward, satisfaction probability and observation rate is presented respectively in Figs. 2, 3 and 4 as a function of the parameter ρ of the AO strategy. 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. In case of PO, T_{obs}=60 s and T_{obs}=150 s have been considered.

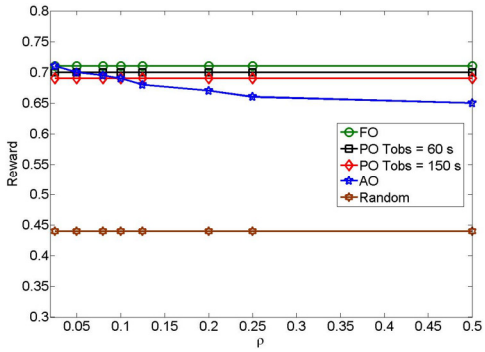

Figure 2. Performance in terms of Reward as a function of ρ

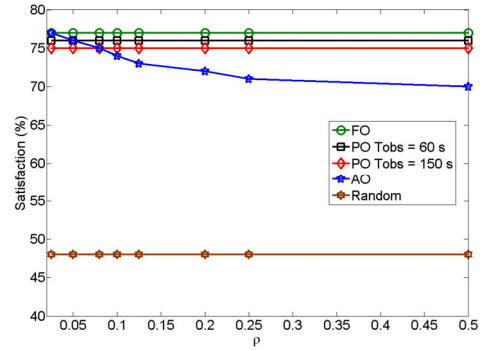

Figure 3. Performance in terms of Satisfaction as a function of ρ

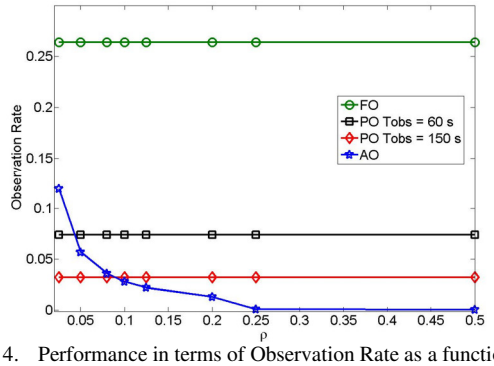

Figure 4. Performance in terms of Observation Rate as a function of ρ

From Fig. 2 and Fig. 3 it can be observed that all the strategies achieve a very significant improvement of around 60% in terms of reward and satisfaction probability with

respect to the random spectrum selection algorithm. Besides, the reward and satisfaction achieved by both POMDP-based approaches PO and AO is very similar to the FO-based algorithm, particularly for low values of ρ roughly up to 0.1.

However, as seen in Fig. 4, this is achieved with a very significant reduction in the observation rate with respect to FO. For instance, when ρ is 0.1 with AO the observation rate is reduced in 89% with respect to FO, while for PO the reduction is 72% and 88% for the cases T_{obs}=60 s and T_{obs}=150 s, respectively. In turn, the reward and satisfaction achieved with AO and ρ=0.1 is only 3% less than with FO, while the reduction with PO is between 1.5% and 3% depending on T_{obs}. Then, AO strategy with setting ρ=0.1 is a good trade-off between the considered fixed T_{obs} period values of PO strategy. Moreover, increasing factor ρ tends to degrade the performance of AO because of the longer time between observations.

To further gain insight in the capability of the AO strategy to adapt to interference dynamics, the impact of varying the average durations of the interference states for each SB has been analyzed. Hence, the durations of Table I have been multiplied by a factor varied between 0.5 and 3. Fig. 5 presents the corresponding performance in terms of observation rate for the different strategies, considering ρ = 0.1 for the AO one.

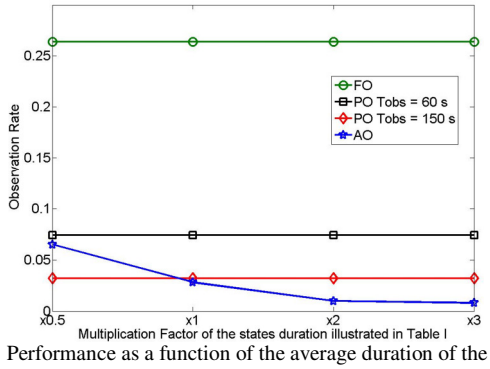

Figure 5. Performance as a function of the average duration of the SB states.

From the figure it can be observed that the observation rate requirements for AO are significantly reduced when the state durations are longer, which leads to further improvements in comparison with FO and PO. For instance, when the durations are multiplied by a factor 3, AO strategy allows a reduction in the observation rate of 75% with respect to PO with T_{obs}=150 s and of 89% with T_{obs}=60 s. This reduction is achieved without having a significant impact in terms of neither reward nor satisfaction. Specifically, the obtained values for all the considered strategies range from 0.68 to 0.70 in the case of the reward, and from 72% to 76% in the case of the satisfaction.

VI. CONCLUSIONS

In this paper a novel POMDP-based framework for spectrum selection in cognitive radio networks has been presented. The framework makes use of the knowledge stored in a database that contains the statistical characterization of the interference variations in the SBs. The approach 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 SB to each radio link/application. The main focus of the paper has been on the impact of the observation strategies that determine the instants in which the SBs are measured. It has been obtained that the POMDP-based algorithm operating with partial observations executed following either a periodical or an adaptive strategy is able to achieve a very close performance with respect to a strategy with full observation capabilities, but with an important reduction of more than 70% in terms of observation rate. Moreover, it has been obtained that the adaptive observation strategy is able to modify the observation rate requirements in accordance with the observed interference dynamics, thus allowing a further reduction in the observation rate with respect to the periodical approach.

Future work will deal with performing a further optimization of the POMDP-based algorithm with the inclusion of spectrum handover mechanisms and with the development of strategies for dynamically acquiring and maintaining the state transition probability values stored in the KD.

ACKNOWLEDGMENT

This work is 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 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] K.P. Murphy, "A Survey of POMDP Solution Techniques", available at http://http.cs.berkeley.edu/~murphyk/Papers/pomdp.ps.gz, 2000.

[8] 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.

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

[10] J. Pérez-Romero, A. Raschellà, O. Sallent, A. Umbert, "Multi-band Spectrum Selection Framework based on Partial Observation", WoWMoM 2103, Madrid, Spain, June 2013.

[11] 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.

[12] Q. Zhao, B. Krishnamachari, K. Liu, "On Myopic Sensing for Multi-Channel Opportunistic Access: Structure, Optimality, and Performance", IEEE Trans. on Wireless Communications, Vol. 7, No. 12, Dec., 2008.

[13] F. Bouali, O. Sallent, J. Pérez-Romero, R. Agusti, "Exploiting Knowledge Management for Supporting Spectrum Selection in Cognitive Radio Networks", CROWNCOM 2012, Stockholm, Sweden, June, 2012.

[14] E. Fernández-Gaucherand, A. Arapostathis, S. I. Marcus, "On the Adaptive Control of a Partially Observable Markov Decision Process", CDC 1988, Austin, Texas, December, 1988.

[15] 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.