Optimal Database Access for TV White Space

Marcello Caleffi, *Member, IEEE*, and Angela Sara Cacciapuoti, *Member, IEEE*

***Abstract*—In TV White Space, the unlicensed users are required to periodically access a database to acquire information on the spectrum usage of the licensed users. In addition, the unlicensed users can access the database on-demand, whenever they believe convenient, to update the spectrum availability information. In this paper, we design the *optimal database access strategy*, i.e., the strategy allowing the unlicensed users to jointly: 1) maximize the expected overall communication opportunities through on-demand accesses; and 2) respect the regulatory specifications. To this aim, we develop a stochastic analytical framework that allows us to account for: 1) the PU activity dynamics; 2) the quality dynamics among the different channels; and 3) the overhead induced by the database access. Specifically, at first, we prove that the database access problem can be modeled as a Markov decision process, and we show that it cannot be solved through brute-force search. Then, we prove that the optimal strategy exhibits a threshold structure, and we exploit this threshold property to design an algorithm able to efficiently compute the optimal strategy. The analytical results are finally validated through simulations.**

***Index Terms*—Spectrum sharing, TV White Space, cognitive radio, database, strategy.**

I. INTRODUCTION

VERY recently, several regulations and standards have approved or are underway to approve the dynamic access of unlicensed users to the TV White Space (TVWS) spectrum [1]–[3]. All the existing rulings obviated the spectrum sensing [4]–[6] as the mechanism for the unlicensed users to recognize and exploit portions of the TVWS spectrum whenever they are vacated by the licensed users, referred to as Primary Users (PUs). Instead, they require the unlicensed users to periodically access to a geolocated database service for acquiring the spectrum availability with a fixed timeframe, referred to as *database access period*. In addition, the regulators allow the unlicensed users to access to the database service on-demand within the database access period, whenever they believe convenient to update the spectrum availability information. However, the specifications of the on-demand access are not yet detailed by the regulations.

Manuscript received September 26, 2014; revised February 13, 2015, June 11, 2015, and September 14, 2015; accepted November 2, 2015. Date of publication November 6, 2015; date of current version January 14, 2016. This work was supported in part by the PON projects "FERSAT: studio di un sistema di segnalamento FERroviario basato sull'innovativo utilizzo delle tecnologie SATellitari e della loro integrazione con le tecnologie terrestri" and DATABANC "CHIS: Cultural Heritage Information System," and in part by the Campania POR project myOpenGov. The preliminary version of this work was presented at IEEE International Conference on Sensing, Communication, and Networking (SECON 2014). The associate editor coordinating the review of this paper and approving it for publication was A. Nallanathan.

The authors are with the Department of Electrical Engineering and Information Technologies, University of Naples Federico II, 80125 Naples, Italy (e-mail: marcello.caleffi@unina.it; angelasara.cacciapuoti@unina.it).

Digital Object Identifier 10.1109/TCOMM.2015.2498607

Clearly, the choice of the database access strategy, i.e., to decide whenever or not to access on-demand to the database, is crucial for the performance of any TVWS Cognitive Radio (CR) network, since it allows the unlicensed users, referred to as CR users, to determine the communication opportunities. To this aim, several factors must be taken into account for a proper access strategy design:

– *PU Traffic Pattern*
The longer is a spectrum band available to the CR user, the less frequent should be the on-demand database accesses.

– *Spectrum Characteristics*
The higher is the overall communication quality provided by the spectrum bands available to the CR user, the less frequent should be the on-demand database accesses. The CR user should access to the database service if and only if the communication opportunities enabled by updating the spectrum availability information exceed with high probability the communication opportunities provided by the already available spectrum bands.

– *Database Access Overhead*
The higher is the communication overhead associated with a database access, the less frequent should be the database accesses. The CR user should access to the database service if and only if the communication opportunities enabled by updating the spectrum availability information compensates with high probability the induced overhead.

In this paper, we propose an *optimal* database access strategy for CR networks operating in TVWS with the objective to account for all the aforementioned key factors. Specifically, the proposed database access strategy is optimal since it allows the CR users to jointly:

i) respect the requirements imposed by the existing rulings in term of periodic mandatory access to the database service;
ii) maximize the expected overall communication opportunities through the on-demand access by accounting for the aforementioned key factors.

More in detail, at first, we develop an analytical framework to model the choice of the database access strategy for an arbitrary CR user as a probabilistic decision process [7], where: i) the reward models the overall communication opportunities, i.e., the *overall spectrum characteristics*; ii) the cost models the *database access overhead*; iii) the transition probabilities of the decision process account for the *PU traffic pattern*. Then, we derive the closed-form expressions of the transition probabilities, and we prove that the decision can be modeled as a Markov Decision Process. Furthermore, we derive the computational complexity of the problem, showing that it can not be solved through brute-force search. Stemming from these results, we prove that the optimal strategy exhibits a threshold structure with respect to the reward. Specifically, whenever the reward exceeds a threshold value, the optimal strategy for the CR user is not to access to the database. Differently, whenever the reward does not exceed the threshold value, the optimal strategy is to access to the database. Finally, we exploit this threshold structure to design a computational-efficient algorithm for finding the optimal strategy.

The rest of the paper is organized as follows. In Sec. II, we describe the network model along with some preliminaries. In Sec. III, we design the optimal database access strategy, whereas in Sec. IV we validate the analytical framework through numerical simulations. In Sec. V, we conclude the paper, and, finally, some proofs are gathered in the Appendix.

II. Network Model and Preliminaries

In this section, we first describe the system model in Sec. II-A. Then, in Sec. II-B, we collect several definitions that will be used through the paper.

A. System Model

According to the current active regulations [8], [9] and standards [10], [11], a CR user[1] must obey to the following rules:

Rule 1: Each CR user must periodically[2]access to a database service to obtain the spectrum availability information.

Remark: The communication between the CR user and the database must not occur within the TV bands unless the CR user has already been authorized by the database.

Rule 2: Within a database access period, each CR user can access on-demand to a database service to update the spectrum availability information.

Rule 3: The spectrum availability information consists of a list of channels within which the CR user is authorized to operate and, for each channel, the duration of such an authorization.

Remark: By organizing the spectrum in M distinct channels, denoted by the set $\Omega = \{1, 2, \ldots, M\}$,[3] the spectrum availability information consists of the channel statuses $\{S_i(n)\}_{i\in\Omega}$ and the channel availabilities $\{N_i(n)\}_{i\in\Omega}$, in agreement with the existing regulations and standards[4]. Specifically, $S_i(n) = 1$ denotes the availability[5] of the i-th channel to the CR network in n-th time slot, whereas $N_i(n) = n_i$ denotes the availability of the i-th channel to the CR networks for n_i consecutive time slots starting at the n-th time slot. Clearly, it results $N_i(n) > 0$ if and only if $S_i(n) = 1$.

[1]In the following, the term *CR user* denotes a fixed unlicensed TV Bands Device (TVBD) in [8], a master White Space Device (WSD) in [9], or a Base Station (BS) in [10].

[2]See 10.6.2 in [10], 3.7 in [9] and 15.711.*b*.3 in [8].

[3]In the following, the channel $i = 1$ denotes a fictitious channel characterized by: i) being always available; ii) providing no communication opportunities, i,e., providing a null reward (see Assumption 2). In such a way, we can easily model the absence of available TV white space channels for the CR user transmissions.

[4]See 15.711.*b*.3 in [8], 3.23 in [9], and 10.7.1.6 in [10].

[5]We assume the channel statuses $S_i(n)$ and $S_j(n)$ independent $\forall i \neq j$. Such an assumption is not restrictive, since when the channel statuses are correlate, e.g., due to adjacent channel use restriction [9], [10], $S_i(n)$ represents either the status of the best channel belonging to a set of correlated channels (if the CR user is equipped with a single radio interface) or the status of a set of correlated channels (if the CR user is equipped with multiple radio interfaces).

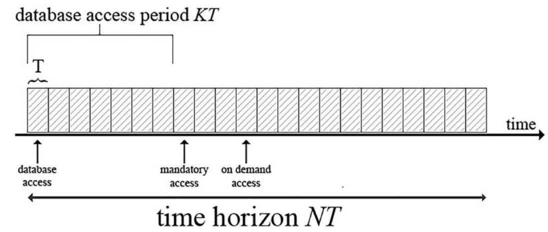

Fig. 1. CR time framing: within the time horizon NT, a database access is mandatory required every K time slots and on-demand access can occur freely to obtain updated spectrum availability information.

B. Assumptions and Definitions

In the following, we give some assumptions and definitions adopted through the paper.

Assumption 1 (Cognitive User Time): The CR user time is organized into N slots of duration T, with KT denoting the database access period, i.e., the maximum time interval between two mandatory database accesses (Rule 1). Hence, K represents the number of slots between two mandatory database accesses, as shown in Fig. 1.

Remark: Physically, the time horizon NT represents the time interval during which the CR user plans to opportunistically use the TVWS spectrum.

Definition 1 (Action Set): The *action set* A is the set of actions for the CR user:[6]

$$A = \{\tilde{a}_1, \ldots, \tilde{a}_M, a_1, \ldots, a_M\}, \quad (1)$$

where:

- Action $\tilde{a}_i$ denotes the event "the CR user can use the available i-th channel for packet transmission during the current time slot and a database access occurs within the current time slot".
- Action a_i denotes the event "the CR user can use the available i-th channel for packet transmission during the current time slot and a database access does not occur within the current time slot".

Remark: Clearly, multiple CR users can choose to concurrently use the same channel, since it has been reported as available from the database. We have two cases:

a) The CR users belong to the same secondary network. In such a case, we have a canonical shared-channel access problem, already handled within the TVWS standards [10].
b) The CR users belong to different and heterogenous secondary networks. In such a case, we can not rely on traditional access schemes since over-the-air communications among heterogeneous TVWS standards is still missing [12]. Nevertheless, very recently, some efforts trying to overcome this issue have been made in [13], [14].

[6]By accounting for Footnote 3, actions $\tilde{a}_1$ and a_1 are characterized by no packet transmissions.

Definition 2 (Cognitive User State): At the n-th time slot, the *CR user state* is defined as the pair:

$$(\tilde{n}_n, \mathbf{d}_n), \tag{2}$$

where $\tilde{n}_n$ denotes, at time slot n, the time slot of the previous database access, and $\mathbf{d}_n = (d_n^1, \ldots, d_n^M)$ denotes the M-tuple defined as follows:[7]

$$d_n^i = \begin{cases} n_i & \text{if } S_i(\tilde{n}_n) = 1 \wedge N_i(\tilde{n}_n) = n_i \\ 0 & \text{otherwise} \end{cases}, \tag{3}$$

with $S_i(\tilde{n}_n)$ denoting the realization of the i-th channel status process acquired through a database access at time slot $\tilde{n}_n$, $N_i(\tilde{n}_n) = n_i$ denoting the realization of the i-th channel availability process acquired through a database access at time slot $\tilde{n}_n$, and $\wedge$ denoting the logical operator *and*. In the following, Σ denotes the set of CR user states.

Remark: According to Definition 2, d_n^i denotes the number of consecutive time slots, starting from the last database access at $\tilde{n}_n$, in which the channel i will be available. Hence, at the n-th time slot, $d_n^i \neq 0$ if and only if during last database access at time slot $\tilde{n}_n$ the database revealed channel i as available, and it results $d_n^i = N_i(\tilde{n}_n)$.

Definition 3 (Allowed Action Set): The *allowed action set* $A_{\tilde{n}_n,\mathbf{d}_n}$ is the set of actions available at the CR user when the CR user state is $(\tilde{n}_n, \mathbf{d}_n)$ at time slot n, and it results:

$$\begin{aligned} &\tilde{a}_i \in A_{\tilde{n}_n,\mathbf{d}_n} \Longleftrightarrow \tilde{n}_n + d_n^i >= n, \\ &a_i \in A_{\tilde{n}_n,\mathbf{d}_n} \Longleftrightarrow \tilde{n}_n + d_n^i >= n \wedge \tilde{n}_n + K >= n, \end{aligned} \tag{4}$$

with K denoting the number of slots between two mandatory database accesses.

Remark: The *allowed action set* $A_{\tilde{n}_n,\mathbf{d}_n}$ denotes the set of actions available at the CR user at the arbitrary time slot n. Through $A_{\tilde{n}_n,\mathbf{d}_n}$, we are able to model the regulatory requirement of a mandatory database access every K slots. Specifically, $a_i, \tilde{a}_i \in A_{\tilde{n}_n,\mathbf{d}_n}$, i.e., the database access is discretionary, when:

a) the i-th channel is available at time slot n, i.e., $\tilde{n}_n + d_n^i \geq n$;
b) the mandatory database access requirement is satisfied, i.e., $\tilde{n}_n + K \geq n$.

On the other hand, $\tilde{a}_i \in A_{\tilde{n}_n,\mathbf{d}_n}$ and $a_i \notin A_{\tilde{n}_n,\mathbf{d}_n}$, i.e., the database access is mandatory, when:

a) the i-th channel is available at time slot n, i.e., $\tilde{n}_n + d_n^i \geq n$;
b) the mandatory database access requirement is not satisfied, i.e., $\tilde{n}_n + K < n$.

Remark: Since, by accounting for Footnote 7 it results $\tilde{a}_1 \in A_{\tilde{n}_n,\mathbf{d}_n}$ for any $(\tilde{n}_n, \mathbf{d}_n) \in \Sigma$, the CR user operates in agreement with Rule 2, i.e., the CR user can access to the database service on-demand. Furthermore, since we assume $N_i(\tilde{n}_n) = n_i \leq K$ for any $i \in \Omega$ and for any $\tilde{n}_n$, the CR user operates in agreement with Rule 2, i.e., the CR user accesses to the database service at most every K time slots.

[7]Clearly, by accounting for Footnote 3, it results $d_n^1 = K$ for any $\tilde{n}_n$.

Assumption 2 (Reward and Cost): We model the average communication quality (i.e., RSSI, SINR, throughput, etc.) of the i-th channel[8] with the dimensionless non-negative quantity r_i, referred to as *channel reward*. Furthermore, we model the average communication overhead induced by a database access with the dimensionless non-negative quantity c, referred to as *database access cost*.

Remark: As an example, if r_i models the average number of bits successfully transmitted through channel i during an arbitrary time slot, c_i models the induced communication overhead, i.e., the average number of bits exchanged during a database access. Nevertheless, by abstracting the communication opportunities and overheads from the particulars through the general notions of reward and cost, the adopted model achieves the following two key features: a) it restricts our attention on the effects of the database access strategy; b) it allows us to measure the performance of a database access strategy, and thus it allows us to quantitatively compare different strategies.

Remark: As mentioned within Section I, the reward models the *overall spectrum characteristics*, whereas the cost models the *database access overhead*. Furthermore, we account for the time variable PU traffic activity with the stochastic decision model developed in Section III.

Definition 4 (Reward): By choosing action $a \in A_{\tilde{n}_n,\mathbf{d}_n}$ when the CR user state is $(\tilde{n}_n, \mathbf{d}_n)$ at time slot n, the CR user reward $r_a(\tilde{n}_n, \mathbf{d}_n)$ is:

$$r_a(\tilde{n}_n, \mathbf{d}_n) = \begin{cases} r_i - c & \text{if } a = \tilde{a}_i,\ i \in \Omega \\ r_i & \text{if } a = a_i,\ i \in \Omega \end{cases}. \tag{5}$$

Assumption 3 (Ordered Channel Set): We assume the channel set Ω ordered according to the channel rewards[9]:

$$r_i \geq r_{i-1}\ \forall\, i = 2, \ldots, M. \tag{6}$$

III. Optimal Database Access Strategy

At first, in Sec. III-A, we formulate the *optimal database access problem* and we discuss its computational complexity (Theorem 2). Then, in Sec. III-B, we prove that the optimal strategy is monotone over the set of actions $\{a_1, \ldots, a_M\}$ (Theorem 3). Stemming from this, in Sec. III-C, we prove that the optimal strategy exhibits a threshold structure (Theorem 4) and we exploit this threshold structure to design a computational-efficient algorithm for solving the *optimal database access problem* (Theorem 5).

A. Optimal Database Access Problem

Here, we formulate the *optimal database access problem* in Definition 6 and we discuss its computational complexity in Theorem 2. To this aim, we prove in Theorem 1 that the problem of choosing the database access strategy can be modeled as a Markov Decision Process, where the reward models

[8]If the CR user is equipped with multiple network interfaces, then the reward r_i denotes the cumulative reward provided by a set of channels.

[9]Clearly, according to the definition of fictitious channel given in Footnote 3, it results $r_1 = 0$.

the discovered communication opportunities and the cost models the overhead associated with a database access. Theorem 1 requires the preliminary result stated in Lemma 1.

Lemma 1 (Transition Probability): Given that the CR user chooses action $a \in A_{\tilde{n}_n,\mathbf{d}_n}$, when the CR user state is $(\tilde{n}_n, \mathbf{d}_n)$ at time slot n, the probability of the CR user state being $(\tilde{n}_{n+1}, \mathbf{d}_{n+1})$ at time slot $n+1$ is conditionally dependent of only the current state and action, and it results (7), shown at the bottom of the page, where $p_a(\tilde{n}_{n+1}, \mathbf{d}_{n+1}|\tilde{n}_n, \mathbf{d}_n)$ denotes the transition probability when action a is chosen, and $p_{d^m_{n+1}|d^m_n} = P\left\{N_m(n+1) = d^m_{n+1}|N_m(\tilde{n}_n) = d^m_n\right\}$ denotes the conditional probability of the availability of the m-th channel:

$$p_{d^m_{n+1}|d^m_n,\tilde{n}_n} = \begin{cases} 0 & \text{if } \tilde{n}_n + d^m_n > n+1+d^m_{n+1}, \\ (p_m)^{k_m} & \text{otherwise} \end{cases} \tag{8}$$

with $p_m = P\{S_m(n) = 1\}$ denoting the probability of the m-th channel being available to the CR network in n-th time slot and $k_m = \min\{d^m_{n+1}, n+1+d^m_{n+1} - \tilde{n}_n - d^m_n\}$.

Proof: See Appendix A. ■

Theorem 1: The database access problem is a Markov Decision Process.

Proof: The proof follows from Lemma 1 by accounting for the Markov property of the transition probabilities [15]. ■

Definition 5 (Strategy): A strategy π is a function that maps the set of CR user states over the set of allowed actions:

$$\forall (\tilde{n}_n, \mathbf{d}_n) \in \Sigma : \pi(\tilde{n}_n, \mathbf{d}_n) \in A_{\tilde{n}_n,\mathbf{d}_n}. \tag{9}$$

In the following, Π denotes the set of strategies.

Corollary 1 (Expected Total Reward): Given the time horizon N, the expected total reward $v_\pi(0, \mathbf{0})$ obtained by the CR user starting from the state $(0, \mathbf{0})$ at time slot 1 and following the strategy π is recursively defined as:

$$v_\pi(0, \mathbf{0}) = r_{\tilde{a}_1}(0, \mathbf{0}) + \sum_{(\tilde{n}_2,\mathbf{d}_2)\in\Sigma} p_{\tilde{a}_1}(\tilde{n}_2, \mathbf{d}_2|0, \mathbf{0}) v_\pi(\tilde{n}_2, \mathbf{d}_2), \tag{10}$$

where $\mathbf{0} = (0, \ldots, 0)$ denotes the null M-tuple, and $v_\pi(\tilde{n}_n, \mathbf{d}_n)$ denotes the expected remaining reward at time slot n defined as in (11), shown at the bottom of the page. Specifically, $v_\pi(\tilde{n}_n, \mathbf{d}_n)$ denotes the reward obtained by the CR user starting from the state $(\tilde{n}_n, \mathbf{d}_n)$ at time slot n and following the strategy π, and $p_{\pi(\tilde{n}_n,\mathbf{d}_n)}(\tilde{n}_{n+1}, \mathbf{d}_{n+1}|\tilde{n}_n, \mathbf{d}_n)$ denotes the transition probability when action $\pi(\tilde{n}_n, \mathbf{d}_n)$ is chosen according to strategy π.

Proof: According to Theorem 1, the optimal database access problem is a Markov Decision Process. Hence, the proof follows from Theorem 4.2 in [15]. ■

Remark: The expected total reward $v_\pi(0, \mathbf{0})$ has been defined as a recursive function. More in detail, $v_\pi(\tilde{n}_n, \mathbf{d}_n)$ at the recursive step n is function of: i) the CR user state $(\tilde{n}_n, \mathbf{d}_n)$; ii) the strategy π through the action $\pi(\tilde{n}_n, \mathbf{d}_n)$; iii) the reward $v_\pi(\tilde{n}_{n+1}, \mathbf{d}_{n+1})$ obtained at the next time slot through the transition probability $p_{\pi(\tilde{n}_n,\mathbf{d}_n)}(\tilde{n}_{n+1}, \mathbf{d}_{n+1}|\tilde{n}_n, \mathbf{d}_n)$. At time instant $n = 1$ the only allowed action is $\tilde{a}_1$ since no spectrum opportunities have been previously discovered, i.e., the CR user state is $(0, \mathbf{0})$.

Definition 6 (Optimal Database Access Problem): Given the channel set $\Omega = \{1, 2, \ldots, M\}$, the statistics on the channel statuses $\{S_i(n)\}_{i\in\Omega}$ and the channel availabilities $\{N_i(n)\}_{i\in\Omega}$, the channel rewards $\{r_i\}_{i\in\Omega}$ and the database access cost c, the time horizon N and the mandatory database access parameter K, the goal is to choose the strategy $\pi^* \in \Pi$ that maximizes the expected total reward:

$$v_{\pi^*}(0, \mathbf{0}) = \sup_{\pi\in\Pi} \{v_\pi(0, \mathbf{0})\}, \tag{12}$$

and we refer to π^* as the optimal strategy.

Corollary 2: Given the channel set $\Omega = \{1, 2, \ldots, M\}$, the statistics on the channel statuses $\{S_i(n)\}_{i\in\Omega}$ and the channel availabilities $\{N_i(n)\}_{i\in\Omega}$, the channel rewards $\{r_i\}_{i\in\Omega}$ and the database access cost c, the time horizon N and the mandatory database access parameter K, the optimal database access strategy π^* is given by:

$$v_{\pi^*}(0, \mathbf{0}) = \max_{\pi\in\Pi} \{v_\pi(0, \mathbf{0})\}. \tag{13}$$

Proof: According to Theorem 1, the optimal database access problem is a Markov Decision Process. Hence, since the sets Σ and A are finite, it always exists a deterministic strategy achieving the supremum in (12) (Lemma 4.3.1 in [15]). ■

Theorem 2 (Problem Complexity): Given the channel set $\Omega = \{1, 2, \ldots, M\}$, the time horizon N and the mandatory database access parameter K, the number of strategies $|\Pi|$ that need to be evaluated to find the optimal strategy through brute-force search is equal to $|\Pi| = (2M)^{N\cdot K\cdot(K+1)^M}$.

$$p_a(\tilde{n}_{n+1}, \mathbf{d}_{n+1}|\tilde{n}_n, \mathbf{d}_n) = \begin{cases} = \begin{cases} \prod_{m=1}^{M} p_{d^m_{n+1}|d^m_n,\tilde{n}_n} & \text{if } \tilde{n}_{n+1} = n+1 \\ 0 & \text{otherwise} \end{cases} & \text{if } a = \tilde{a}_i, i \in \Omega \\ = \begin{cases} 1 & \text{if } \begin{cases} \tilde{n}_{n+1} = \tilde{n}_n \\ d^m_{n+1} = d^m_n \end{cases} \\ 0 & \text{otherwise} \end{cases} & \text{if } a = a_i, i \in \Omega \end{cases}. \tag{7}$$

$$v_\pi(\tilde{n}_n, \mathbf{d}_n) = \begin{cases} r_{\pi(\tilde{n}_n,\mathbf{d}_n)}(\tilde{n}_n, \mathbf{d}_n) + \sum_{(\tilde{n}_{n+1},\mathbf{d}_{n+1})\in\Sigma} p_{\pi(\tilde{n}_n,\mathbf{d}_n)}(\tilde{n}_{n+1}, \mathbf{d}_{n+1}|\tilde{n}_n, \mathbf{d}_n) v_\pi(\tilde{n}_{n+1}, \mathbf{d}_{n+1}) & \text{if } 1 < n < N \\ r_{\pi(\tilde{n}_N,\mathbf{d}_N)}(\tilde{n}_N, \mathbf{d}_N) & \text{if } n = N \end{cases}. \tag{11}$$

Proof: See Appendix B. ∎

Remark: The exponential time complexity of the *optimal database access problem* makes the problem computationally intractable even for small instances of the problem. As an example, when $M = K = 2$ and $N = 3$, it results $|\Pi| = 4^{54}$. Nevertheless, in the next section, we prove that the considered problem admits an optimal deterministic strategy with the following appealing structure:

$$\pi^*(\mathbf{s}) = \begin{cases} \tilde{a}_0 & \text{if } \mathbf{s} < \mathbf{s}* \\ a_0 & \text{if } \mathbf{s} \geq \mathbf{s}* \end{cases} \tag{14}$$

where s is the CR user state defined in Corollary 5, $\tilde{a}_0$ and a_0 are the actions defined in Corollary 6, and s^* is the control limit defined in Theorem 4. Hence, by exploiting such a optimal strategy structure, we are able to design a computational-effective *optimal database access algorithm*.

B. Properties of the Optimal Strategy

Here, we first prove in Theorem 3 that the expected remaining reward $v_\pi(\tilde{n}_n, \mathbf{d}_n)$ is monotone for $\pi(\tilde{n}_n, \mathbf{d}_n) \in \{a_1, \ldots, a_M\}$. Then, we reformulate the decision problem in Lemma 3. The proof of Theorem 3 requires a preliminary result (Lemma 2).

Lemma 2: Given that the CR user chooses action $a \in \{a_1, \ldots, a_M\}$ when the CR user state is $(\tilde{n}_n, \mathbf{d}_n)$ at time slot n, the transition probability $p_a(\tilde{n}_{n+1}, \mathbf{d}_{n+1}|\tilde{n}_n, \mathbf{d}_n)$ is conditionally independent of action a:

$$\begin{aligned} &p_{a_i}(\tilde{n}_{n+1}, \mathbf{d}_{n+1}|\tilde{n}_n, \mathbf{d}_n) \\ &\quad = p_{a_j}(\tilde{n}_{n+1}, \mathbf{d}_{n+1}|\tilde{n}_n, \mathbf{d}_n)\ \forall a_i, a_j \in \{a_1, \ldots, a_M\}\ . \end{aligned} \tag{15}$$

Proof: The proof follows from (7) and (8) in Lemma 1. ∎

Corollary 3: Given that the CR user chooses action $a \in \{\tilde{a}_1, \ldots, \tilde{a}_M\}$ when the CR user state is $(\tilde{n}_n, \mathbf{d}_n)$ at time slot n, the transition probability $p_a(\tilde{n}_{n+1}, \mathbf{d}_{n+1}|\tilde{n}_n, \mathbf{d}_n)$ is conditionally independent of action a.

Proof: The proof follows by adopting the same reasoning of Lemma 2. ∎

Theorem 3: The expected remaining reward $v_\pi(\tilde{n}_n, \mathbf{d}_n)$ is not decreasing for $\pi(\tilde{n}_n, \mathbf{d}_n) \in \{a_1, \ldots, a_M\}$ for any CR user state $(\tilde{n}_n, \mathbf{d}_n) \in \Sigma$, i.e.:

$$v_\pi(\tilde{n}_n, \mathbf{d}_n) \geq v_{\pi'}(\tilde{n}_n, \mathbf{d}_n), \tag{16}$$

where:

$$\pi, \pi' \in \Pi : \begin{cases} \pi(\tilde{n}_n, \mathbf{d}_n) = a_i \in \{a_1, \ldots, a_M\}, i > 1 \\ \pi'(\tilde{n}_n, \mathbf{d}_n) = a_{i-1} \in \{a_1, \ldots, a_M\} \\ \pi(\tilde{n}_k, \mathbf{d}_k) = \pi'(\tilde{n}_k, \mathbf{d}_k)\ \forall k > n \end{cases} . \tag{17}$$

Proof: See Appendix C. ∎

Insight 1: The rationale of the proof of Theorem 3 is the following. We first focus our attention over a specific subset of actions, i.e., the actions $\{a_1, \ldots, a_M\}$ not involving a database access. Clearly, the reward $r_{a_i}(\tilde{n}_n, \mathbf{d}_n)$ does depend on the quality of the i-th channel via action a_i. However, by accounting for Lemma 2 and by reasoning with backward induction, we prove that the expected remaining reward $v_\pi(\tilde{n}_n, \mathbf{d}_n)$ does not depend on the chosen action. Hence, by choosing the action a_i maximizing the reward $r_{a_i}(\tilde{n}_n, \mathbf{d}_n)$ we maximize the expected total reward $v_\pi(0, \mathbf{0})$ as well.

Remark: With Theorem 3, we proved that the expected remaining reward is monotone not-decreasing over the subset of actions not involving a database access. With Corollary 4, we will prove that the expected remaining reward is monotone not-decreasing also over the subset of actions involving a database access.

Corollary 4: The expected remaining reward $v_\pi(\tilde{n}_n, \mathbf{d}_n)$ is not decreasing for $\pi(\tilde{n}_n, \mathbf{d}_n) \in \{\tilde{a}_1, \ldots, \tilde{a}_M\}$ for any CR user state $(\tilde{n}_n, \mathbf{d}_n) \in \Sigma$.

Proof: The proof follows by adopting the same reasoning of Theorem 3. ∎

Insight 2: With reference to a CR user aiming not to access to the database, Theorem 3 states that, whenever multiple channels are available to the CR user at time slot n, the expected remaining reward achievable by using the available channel with the highest reward is never lower than the the expected remaining reward achievable by using any other available channel. A similar result is derived in Corollary 4 with reference to a CR user aiming to access to the database. Hence, we can exploit this result to simplify the considered problem:

i) by storing within the CR user state only the channel with the highest reward known to be available in the current time slot and in the future $K - 1$ time slots (see Corollary 5);
ii) by reducing the available actions to the CR user to either "to access to the database" or "not to access to the database" (see Corollary 6).

Corollary 5: At time slot n, the CR user state $(\tilde{n}_n, \mathbf{d}_n)$ can be represented by the K-tuple $\mathbf{s}_n = (s_n^1, \ldots, s_n^K) \in \Sigma' = \{-1, 1 \ldots, M\}^K$ with s_n^k defined as:

$$s_n^k = \begin{cases} i & \text{if } \tilde{n}_n + K \geq n + k \\ -1 & \text{if } \tilde{n}_n + K < n + k \end{cases} , \tag{18}$$

where $i = \max_{m \in \Omega} \left\{ m : \tilde{n}_n + d_n^m \geq n + k \right\}$.

Proof: The proof follows from Theorem 3. ∎

Remark: If a mandatory database access is required at time slot $n + k - 1$, then $s_n^k = -1$. Otherwise, s_n^k denotes the channel with the highest reward known to be available at time slot $n + k - 1$.

Corollary 6: At any time slot n, the action set $A_{\mathbf{s}_n}$ when the CR user state is $\mathbf{s}_n$ is:

$$A_{\mathbf{s}_n} = \begin{cases} \{\tilde{a}_0\} & \text{if } s_n^1 = -1 \\ \{\tilde{a}_0, a_0\} & \text{otherwise} \end{cases} , \tag{19}$$

where:

- Action $\tilde{a}_0$ denotes the event "the CR user updates the spectrum availability information in the current time slot by accessing the database".
- Action a_0 denotes the event "the CR user does not update the spectrum availability information in the current time slot by accessing the database".

Proof: The proof follows from Theorem 3. ■

By means of Theorem 3 and the related corollaries, we can finally reformulate the *optimal database access problem* in Lemma 3.

Lemma 3: For any time slot n and for any CR user state $(\tilde{n}_n, \mathbf{d}_n)$, we have (20), shown at the bottom of the page, where $\mathbf{s}_n$ is defined in Corollary 5, $r_{\pi(\mathbf{s}_n)}(\mathbf{s}_n)$ is equal to:

$$r_{\pi(\mathbf{s}_n)}(\mathbf{s}_n) = \begin{cases} r_{s_n^1} - c & \text{if } \pi(\mathbf{s}_n) = \tilde{a}_0 \\ r_{s_n^1} & \text{if } \pi(\mathbf{s}_n) = a_0 \end{cases}, \tag{21}$$

and $p_{\pi(\mathbf{s}_n)}(\mathbf{s}_{n+1}|\mathbf{s}_n)$ is defined as in (22), shown at the bottom of the page, with:

$$q_1(\mathbf{s}_{n+1}|\mathbf{s}_n) = \begin{cases} p_{s_{n+1}^1} \prod\limits_{j > s_{n+1}^1} \bar{p}_j & \text{if } s_{n+1}^1 > s_n^2 \\ \prod\limits_{j > s_{n+1}^1} \bar{p}_j & \text{if } s_{n+1}^1 = s_n^2 \\ 0 & \text{otherwise} \end{cases}, \tag{23}$$

$$q_{k>1}(\mathbf{s}_{n+1}|\mathbf{s}_n) = \begin{cases} p_{s_{n+1}^k} & \text{if } s_{n+1}^k = s_{n+1}^{k-1} \\ & \text{and } s_{n+1}^k > s_n^{k+1} \\ p_{s_{n+1}^k} \prod\limits_{j = s_{n+1}^k + 1}^{s_{n+1}^{k-1}} \bar{p}_j & \text{if } s_{n+1}^k < s_{n+1}^{k-1} \\ & \text{and } s_{n+1}^k > s_n^{k+1} \\ \prod\limits_{j = s_{n+1}^k + 1}^{s_{n+1}^{k-1}} \bar{p}_j & \text{if } s_{n+1}^k = s_n^{k+1} \\ 0 & \text{otherwise} \end{cases}. \tag{24}$$

Proof: See Appendix D. ■

Remark: By adopting the same reasoning of Theorem 1, it can be proved that time complexity of the reformulated problem is still exponential, since the number of strategies $|\Pi|$ that need to be evaluated to find the optimal strategy is equal to $|\Pi| = 2^{N(M+1)^K}$.

Algorithm 1. Optimal Database Access Algorithm

//$A^*(s)$ optimal action for CU state s
// $v^*(s)$ maximum expected remaining reward at CU state s
// base step: $n = N$
$A = \{\tilde{a}_0, a_0\}$
$\Sigma_N = \{0, 1, \ldots, M\}$
for $s = 0 : M$ **do**
    $A^*(s) = \{a_0\}$
    $v^*(s) = r(s, 1)$
end for
// inductive step: $n < N$
for $n = N - 1 : 1$ **do**
    for $x = \{x^1, \ldots, x^{K-1}\} \in \Sigma_{n+1}$ **do**
        // s^*: threshold according to Theorem 4.2
        for $s^1 = 0 : M$ **do**
            // $s = \{s^1, x^1, \ldots, x^{K-1}\}$
            if $s_1 > s^*$ **then**
                $A^*(s) = \{a_0\}$
            else
                $A^*(s) = \{\tilde{a}_0\}$
            end if
        end for
    end for
end for

C. Threshold Property of the Optimal Strategy

Here, we first prove in Theorem 4 that the optimal strategy exhibits a threshold structure with respect to the channel rewards $\{r_i\}_{i \in \Omega}$. Then, we exploit this structure to design a computational-efficient database access algorithm (Algorithm 1), and we prove in Theorem 5 that such an algorithm effectively finds the optimal strategy π^*. The proof of Theorem 4 requires some preliminaries (Definition 7 and Lemma 4).

Definition 7 (Cognitive User State Order): Given two CR user states $\mathbf{s}_n, \mathbf{x}_n \in \Sigma'$ at time slot n, we define the following partial order over Σ':

$$\mathbf{x}_n = (x_n^1, \ldots, x_n^K) \geq \mathbf{s}_n = (s_n^1, \ldots, s_n^K) \iff \\ \iff x_n^1 \geq s_n^1 \wedge s_n^k = x_n^k \; \forall k > 1. \tag{25}$$

$$v_{\pi^*(\tilde{n}_n, \mathbf{d}_n)}(\tilde{n}_n, \mathbf{d}_n) = v_{\pi^*(\mathbf{s}_n)}(\mathbf{s}_n) = \max_{\pi(\mathbf{s}_n) \in A_{\mathbf{s}_n}} \left\{ r_{\pi(\mathbf{s}_n)}(\mathbf{s}_n) + \sum_{\mathbf{s}_{n+1} \in \Sigma'} p_{\pi(\mathbf{s}_n)}(\mathbf{s}_{n+1}|\mathbf{s}_n) v_{\pi^*(\mathbf{s}_{n+1})}(\mathbf{s}_{n+1}) \right\}. \tag{20}$$

$$p_{\pi(\mathbf{s}_n)}(\mathbf{s}_{n+1}|\mathbf{s}_n) = \begin{cases} = \begin{cases} 1 & \text{if } \begin{cases} s_{n+1}^k = s_n^{k+1} \; \forall k \in [1, K-1] \\ s_{n+1}^K = -1 \end{cases} \\ 0 & \text{otherwise} \end{cases} & \text{if } \pi(\mathbf{s}_n) = a_0 \\ \prod\limits_{k=1}^{K} q_k(\mathbf{s}_{n+1}|\mathbf{s}_n) & \text{if } \pi(\mathbf{s}_n) = \tilde{a}_0 \end{cases}. \tag{22}$$

Lemma 4: Given the CR user state $\mathbf{s}_n \in \Sigma'$ at time slot n, it results:

$$\pi^*(\mathbf{s}_n) = a_0 \Longrightarrow \pi^*(\mathbf{x}_n) = a_0 \ \forall \mathbf{x}_n > \mathbf{s}_n. \tag{26}$$

Proof: See Appendix E. ■

Insight 3: (26) provides a practical rule for designing the optimal database access strategy. Specifically, if action a_0 is optimal when the CR user state is $\mathbf{s}_n = \{s_n^1, s_n^2, \ldots, s_n^K\}$, then action a_0 is optimal for any CR user state $\mathbf{x}_n = \{x_n^1, s_n^2, \ldots, s_n^K\} \in \Sigma'$ with x_n^1 greater than s_n^1, where the relation order "greater of" is defined by (25).

Corollary 7: Given the CR user state $\mathbf{s}_n \in \Sigma'$ at time slot n, it results:

$$\pi^*(\mathbf{s}_n) = \tilde{a}_0 \Longrightarrow \pi^*(\mathbf{x}_n) = \tilde{a}_0 \ \forall \mathbf{x}_n < \mathbf{s}_n. \tag{27}$$

Proof: It is straightforward to prove the corollary by following the same reasoning adopted in Appendix E. ■

Insight 4: (27) provides a practical rule for designing the optimal database access strategy. Specifically, if action $\tilde{a}_0$ is optimal when the CR user state is is $\mathbf{s}_n = \{s_n^1, s_n^2, \ldots, s_n^K\}$, then action $\tilde{a}_0$ is optimal for any CR user state $\mathbf{x}_n = \{x_n^1, s_n^2, \ldots, s_n^K\} \in \Sigma'$ with $x_n^1 \leq s_n^1$.

Theorem 4: Given the CR user state $\mathbf{s}_n = \left\{s_n^1, \ldots, s_n^K\right\}$ at time slot n and two strategies $\pi, \pi' \in \Pi : \pi(\mathbf{s}_n) = a_0, \pi'(\mathbf{s}_n) = \tilde{a}_0 \wedge \pi(\mathbf{s}_k) = \pi'(\mathbf{s}_k) \ \forall k > n$, it results:

$$\begin{aligned} & v_\pi(\mathbf{s}_n) > v_{\pi'}(\mathbf{s}_n) \Longleftrightarrow \\ & \quad r_{s_n^1} > \sum_{\mathbf{s}_{n+1} \in \Sigma'} p_{\pi'(\mathbf{s}_n)}(\mathbf{s}_{n+1}|\mathbf{s}_n) v_{\pi'}(\mathbf{s}_{n+1}) - \\ & \qquad - v_\pi\left(\mathbf{s}_{n+1} = \left\{s_n^2, \ldots, s_n^K, -1\right\}\right) - c. \end{aligned} \tag{28}$$

Proof: The proof follows from Lemma 3 and Corollary 7. ■

Remark: Theorem 4 proves that the *maximum expected remaining reward* $v_{\pi^*}(\left\{s_n^1, \ldots, s_n^K\right\})$ exhibits a threshold structure with respect to the channel reward $r_{s_n^1}$. More in detail, whenever $r_{s_n^1}$ is greater than a certain threshold, then action a_0 is optimal. Differently, whenever $r_{s_n^1}$ is lower than the same threshold, then action $\tilde{a}_0$ is optimal. Clearly, from (28) it results:

$$\begin{aligned} & v_\pi(\mathbf{s}_n) < v_{\pi'}(\mathbf{s}_n) \Longleftrightarrow \\ & \quad r_{s_n^1} < \sum_{\mathbf{s}_{n+1} \in \Sigma'} p_{\pi'(\mathbf{s}_n)}(\mathbf{s}_{n+1}|\mathbf{s}_n) v_{\pi'}(\mathbf{s}_{n+1}) - \\ & \qquad - v_\pi\left(\mathbf{s}_{n+1} = \left\{s_n^2, \ldots, s_n^K, -1\right\}\right) - c. \end{aligned} \tag{29}$$

Insight 5: (28) provides a practical rule for designing the optimal database access strategy. Specifically, the optimal action at time slot n depends on: i) the maximum expected remaining reward $v_\pi(\{s_n^2, \ldots, s_n^K, -1\})$ when the CR user state is $\{s_n^2, \ldots, s_n^K, -1\}$ at time slot $n+1$; ii) the transition probabilities $p_{\tilde{a}_0}(\mathbf{s}_{n+1}|\mathbf{s}_n)$ when action $\tilde{a}_0$ is chosen at time slot n; iii) the maximum expected remaining reward $v_\pi(\mathbf{s}_{n+1})$ when the CR user state is $\mathbf{s}_{n+1}$ at time slot $n+1$.

Theorem 5 (Optimal Database Access Algorithm): Algorithm 1 solves the *optimal database access problem*.

Proof: The proof follows from Theorem 4 by reasoning with backward induction (see 4.5 and 4.7.6 in [15]). ■

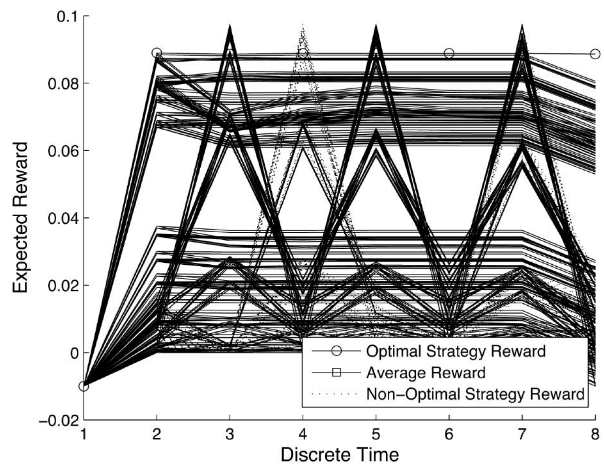

Fig. 2. Expected remaining reward vs time for $M = 2$, $K = 2$, $N = 8$, $c = 0.01$, and $p_i = 0.1$.

IV. Performance Evaluation

In this section, we validate the theoretical results derived in Sec. III by simulating a CR network operating in the TV White Space according to Rules 1-3.

In the first experiment, we compare the performance of Algorithm 1 with those obtained through brute-force search. More specifically, Fig. 2 presents the expected reward as a function of the discrete time. The adopted simulation set, summarized in Table I, is as follows: $M = 2$, $K = 2$, $N = 8$, $c = 0.01$, $r_i = (2i+1)/(2M)$ and $p_i = 0.1$ for any $i \in \Omega$, and we consider the following rewards: i) the *Optimal Strategy Reward*, i.e., the reward achieved by the strategy derived in Algorithm 1; ii) the *Average Reward*, i.e., the reward achievable by averaging over the rewards provided by any admissible strategy; ii) the *Non-Optimal Strategy Reward*, i.e., the reward achievable by an admissible strategy different from the optimal strategy. First, we note that there exist time slots such as $N = 3$ in which the reward achieved by the optimal strategy is not the maximum of the achievable rewards. This behavior is reasonable: Algorithm 1 aims at discovering the strategy assuring the highest expected total reward $v_\pi(0, \mathbf{0})$, i.e., the highest expected reward over the whole time horizon N. In fact, by averaging over the whole time horizon N, the optimal strategy achieves the highest expected reward, as confirmed with the next experiment.

Fig. 3 presents the expected total reward as a function of the PU inactivity probability for the same simulation set adopted for Fig. 2. We first observe that, for any considered value of the PU inactivity probability p_i, Algorithm 1 assures the highest expected total reward. This result clearly confirms Theorem 5, i.e., the optimality of the strategy obtained through Algorithm 1. Furthermore, we observe that the higher is the PU inactivity probability p_i, the higher is the expected total reward achieved by the optimal strategy. This result is reasonable: the optimal strategy aims probabilistically at exploiting any available opportunity, and the higher is p_i, the most likely the i-th channel will provide a transmission opportunity during an arbitrary time slot and the longer this opportunity will last. Finally, we observe that, the higher is the probability p_i of channel i being

TABLE I
PERFORMANCE EVALUATION PARAMETER SETTING

Symbol	Definition	Fig. 2	Fig. 3	Fig. 4	Fig. 5	Fig. 6	Fig. 7
M	number of TVWS channels	2	2	2-10	4	4	2-6
K	number of slots between two database accesses	2	2	2-4	4	4	4
N	total number of slots	8	8	8	30	30	30
c	database access cost	0.01	0.01	-	0.01	0.001-1	0.01
p_i	i-th TVWS channel availability probability	0.1	0.01-0.5	-	0.1	0.1	0.1
r_i	i-th TVWS channel reward	$\frac{2i+1}{2M}$	$\frac{2i+1}{2M}$	-	$\frac{2i+1}{2M}$	$\frac{2i+1}{2M}$	$\frac{2i+1}{2M}$

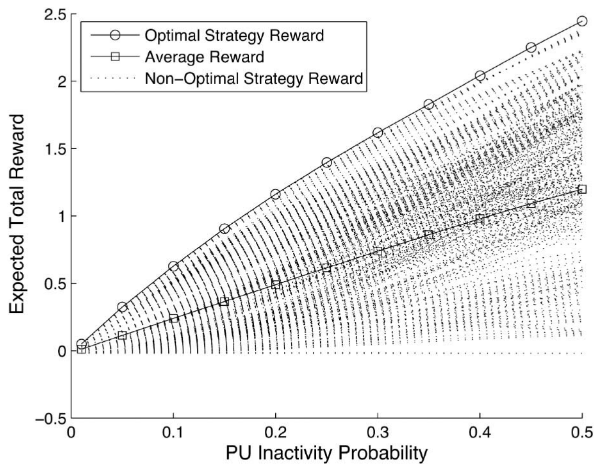

Fig. 3. Expected total reward vs PU inactivity probability for $M = 2$, $K = 2$, $N = 8$, and $c = 0.01$.

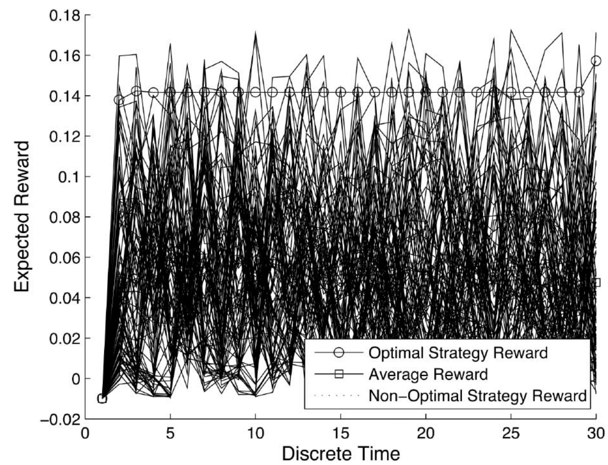

Fig. 5. Expected reward vs time for $M = 4$, $K = 4$, $N = 8$, $c = 0.01$, and $p_i = 0.1$.

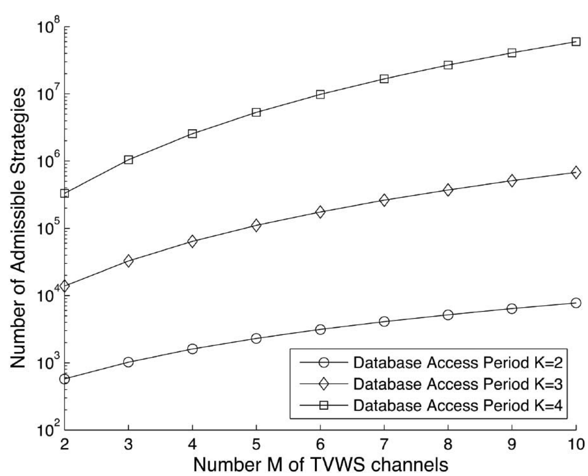

Fig. 4. Exhaustive search complexity vs number M of TVWS channels for $K = 2$, and $N = 8$. Logarithmic scale for axis y.

available, the larger is the range of the total expected rewards achievable by the admissible strategies. This result is reasonable: the higher is the the probability p_i, the higher is the impact of a good strategy in terms of achievable reward.

As mentioned in Sec. III-B, the complexity of the exhaustive search of the optimal solution is unfeasible even for small values of M. This is confirmed with Fig. 4, which presents the number $|\Pi|$ of admissible strategies as a function of the number M of TVWS channels for different values of K. As detailed in the remark following Lemma 3, the total number of strategies $|\Pi|$ grows exponentially with $\mathcal{O}(2^{NM^K})$. Hence, by fixing N and K, $|\Pi|$ grows exponentially with a polynomial factor in M as confirmed by Fig. 4, where $|\Pi|$ as function of M is reported in logarithmic scale. As instance, for $M = 4$, $K = 4$ and $N = 8$, we have $|\Pi| = 2^{18750}$ admissible strategies. Hence, in the following experiments, we guarantee a feasible computation by focus our attention on 100 randomly-generated admissible strategies.

In Fig. 5 we present the expected reward as a function of the discrete time for $M = 4$, $K = 4$, and $N = 30$. We first observe that the the highest instantaneous reward assured by the optimal strategy occurs at $N = 30$. This is reasonable: during the last time slot a CR user can not benefit from additional opportunities discovered through a on-demand database access, hence the optimal action is always to use any available channel unless a mandatory database access is required. Furthermore, we note that the optimal strategy assures a steady reward with respect to the discrete time. This is reasonable since the developed model aims at maximizing expected quantities, i.e., the expected reward. On the other hand, the non-optimal strategies present a high variability in the achieved expected rewards. This is reasonable. In fact, since these strategies are randomly generated and they are not optimal in terms of expected total reward, their expected reward values are deeply affected by channel statuses realizations.

Fig. 6 presents the expected total reward as a function of the database access cost for $M = 4$, $K = 4$, and $N = 30$. We first observe that the expected reward linearly decreases as the database access cost increases. This is reasonable and it agrees with the derived analysis (21). Moreover, we observe that the optimal strategy significantly outperforms the randomly generated strategies for any considered value of the access

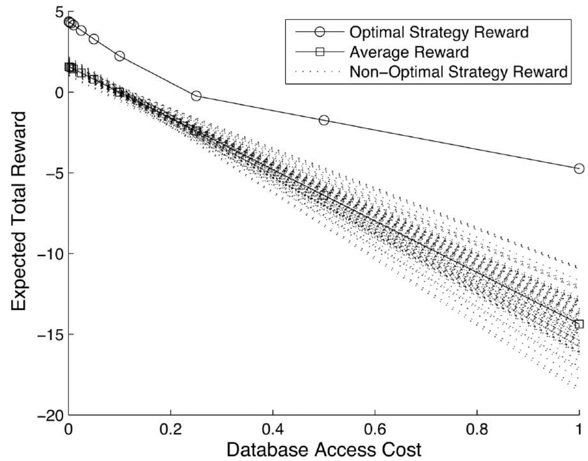

Fig. 6. Expected total reward vs database access cost c for $M = 4$, $K = 4$, $N = 8$, and $p_i = 0.1$.

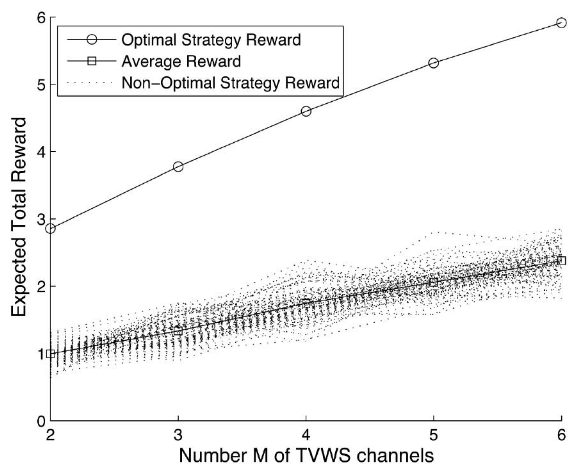

Fig. 7. Expected Total Reward vs number M of TVWS channels for $K = 4$, $N = 8$, $c = 0.01$ and $p_i = 0.1$.

cost. This confirms that: i) the choice of the database access strategy is crucial for the performance of any TVWS Cognitive Radio (CR) network; ii) a significant performance gain can be obtained with the optimal strategy design. Finally, we observe that, the higher is the access cost c, the larger is the range of the total expected rewards achievable by the non-optimal strategies. This agrees with the intuition, since the higher is the the access cost, the higher is the impact of a bad strategy in terms of achievable reward.

Finally, Fig. 7 presents the expected total reward as a function of the number of TVWS channels. We observe that, the higher is the number M of channels, the higher is the expected reward. This result is reasonable: the higher is M, the likely there exists a channel free from PU activity, and hence the more are the degrees of freedom that a strategy can exploit to maximize the expected total reward. Furthermore, we observe that the higher is the number M of channels, the more the optimal strategy outperforms the randomly-generated ones. This result is reasonable: the higher is M, the higher is the impact of a good strategy in terms of achievable reward. Clearly, this result substantiates the importance of an optimal strategy design to maximize the performance of the TVWS CR network.

V. Conclusions

Very recently, several regulations and standards have approved or are underway to approve the dynamic access of unlicensed users to the TV White Space spectrum. All the existing rulings rely on a periodic access to a database service as the primary mechanism for the unlicensed users to determine the White Space availability. Hence, in this paper, we address the problem of accessing to the database service by designing an *optimal database access strategy*, i.e., by designing a strategy allowing the unlicensed users to jointly: i) respect the requirements imposed by the existing rulings in term of periodic mandatory access to the database service; ii) maximize the expected overall communication opportunities through on-demand accesses to the database service. By modeling the database access problem as a Markov Decision Process, and by showing that the optimal strategy exhibits a threshold structure, we were able to design a computational-efficient algorithm for the optimal strategy. The numerical validation of the proposed strategy through a case study confirmed the optimality property.

Appendix

A. Proof of Lemma 1

Proof:

Case 1: $a = a_i, i \in \Omega$. By choosing action a_i at time slot n, the CR user does not acquire any updated spectrum availability information. Hence, the next CR user state $(\tilde{n}_{n+1}, \mathbf{d}_{n+1})$ is univocally determined by $(\tilde{n}_n, \mathbf{d}_n)$. Specifically, by accounting for Definition 2, it results:

$$\begin{aligned} \tilde{n}_{n+1} &= \tilde{n}_n \ , \\ d^i_{n+1} &= d^i_n \ , \end{aligned} \tag{30}$$

and the thesis follows.

Case 2: $a = \tilde{a}_i, i \in \Omega$. By choosing action $\tilde{a}_i$ at time slot n, the CR user does update the spectrum availability information. Hence, it results $\tilde{n}_{n+1} = n + 1$. Furthermore, with reference to the i-th channel, we have $d^i_{n+1} \neq 0$ if and only if $S_i(n + 1) = 1$ and $N_i(n + 1) = d^i_{n+1}$. We have two cases: i) if $\tilde{n}_n + d^i_n \geq n + 1$, i.e., if during the previous database access at time slot $\tilde{n}_n$ channel i was reported as available up to time slot $n + 1$, then $P\{N_i(n + 1) = d^i_{n+1} | N_i(\tilde{n}_n) = d^i_n\} = p_i^{k_i}, k_i = n + 1 + d^i_{n+1} - \tilde{n}_n - d^i_n$; ii) otherwise, $P\{N_i(n + 1) = d^i_{n+1} | N_i(\tilde{n}_n) = d^i_n\} = P\{N_i(n + 1) = d^i_{n+1}\} = p_i^{d^i_{n+1}}$. Finally, we have $d^i_{n+1} = 0$ if and only if $S_i(n + 1) = 0$ and $\tilde{n}_n + d^i_n < n + 1$, i.e., if during the previous database access at time slot $\tilde{n}_n$ channel i was not reported as available up to time slot $n + 1$. ■

B. Proof of Theorem 2

Proof: By observing that, at n-th time slot at least one database access must occurred within the previous K slots (Rule 1), i.e., $\tilde{n}_n \in [n - K, n)$, it results that the cardinality $|\Sigma|$ of the CR user state is equal to $K(K + 1)^M$. Since $M + 1$ actions are available for each CR user state and since the time horizon is N, i.e., a strategy defines a sequence of N different actions for each CR user state, we have $|\Pi| = (M+1)^{N\cdot K\cdot(K+1)^M}$. ■

C. Proof of Theorem 3

Proof: We prove the theorem through backward induction. ■

Case N: The *base step* follows by Assumption 3 and Definition 4, since we have:

$$v_\pi(\tilde{n}_N, \mathbf{d}_N) = r_{a_i}(\tilde{N}, \mathbf{d}_N) = r_i \geq \\ \geq r_{i-1} = r_{a_{i-1}}(\tilde{n}_N, \mathbf{d}_N) = v_{\pi'}(\tilde{n}_N, \mathbf{d}_N)\ . \quad (31)$$

Case $n < N$: Let us suppose that the theorem holds for $n+1$, and let us consider n. We prove the inductive step with a *reductio ad absurdum* by supposing that there exists $\pi, \pi' \in \Sigma$ satisfying (17) so that:

$$v_\pi(\tilde{n}_n, \mathbf{d}_n) < v_{\pi'}(\tilde{n}_n, \mathbf{d}_n). \quad (32)$$

As in (31), we have $r_{a_i}(\tilde{n}_n, \mathbf{d}_n) \geq r_{a_{i-1}}(\tilde{n}_n, \mathbf{d}_n)$. Thus, by accounting for (11) it results:

$$v_\pi(\tilde{n}_n, \mathbf{d}_n) < v_{\pi'}(\tilde{n}_n, \mathbf{d}_n) \Longrightarrow \\ \Longrightarrow \sum_{(\tilde{n}_{n+1}, \mathbf{d}_{n+1})} p_{a_i}(\tilde{n}_{n+1}, \mathbf{d}_{n+1}|\tilde{n}_n, \mathbf{d}_n) v_\pi(\tilde{n}_{n+1}, \mathbf{d}_{n+1}) < \\ < \sum_{(\tilde{n}_{n+1}, \mathbf{d}_{n+1})} p_{a_{i-1}}(\tilde{n}_{n+1}, \mathbf{d}_{n+1}|\tilde{n}_n, \mathbf{d}_n) v_{\pi'}(\tilde{n}_{n+1}, \mathbf{d}_{n+1}) \\ \Longrightarrow v_\pi(\tilde{n}_{n+1}, \mathbf{d}_{n+1}) < v_{\pi'}(\tilde{n}_{n+1}, \mathbf{d}_{n+1})\ . \quad (33)$$

where the last implication follows from (15) in Lemma 2. Hence, since by accounting for the second condition in (17) it results $v_\pi(\tilde{n}_{n+1}, \mathbf{d}_{n+1}) = v_{\pi'}(\tilde{n}_{n+1}, \mathbf{d}_{n+1})$, (33) constitutes a *reductio ab absurdum*.

D. Proof of Lemma 3

Proof: The expression (21) of the reward $r_{\pi(\mathbf{s}_n)}(\mathbf{s}_n)$ follows from (18) by accounting for Definition 4. The expression (22) of the transition probability $p_{a_0}(\mathbf{s}_{n+1}|\mathbf{s}_n)$ follows from (18) by accounting for (7). We consider now the expression (22) of the transition probability $p_{\tilde{a}_0}(\mathbf{s}_{n+1}|\mathbf{s}_n)$. With reference to s^1_{n+1}, at time slot $n+1$ we have $s^1_{n+1} = i$ if the following three conditions hold simultaneously: i) $S_j(n+1) = 0; \forall\, j > i$, i.e., all the channels with a reward greater then r_i are not available at time slot $n+1$; ii) $S_i(n+1) = 1$, i.e., channel i is available at time slot $n+1$; iii) $i \geq s_2$, i.e., the channel reported at time slot n as available at time slot $n+1$ is still available. Hence, (23) follows by exploiting the channel independence assumption [16]–[18]. Similarly, with reference to s^k_{n+1}, $k > 1$, we have $s^k_{n+1} = i$ if the following four conditions hold simultaneously: i) $S_j(n+1) = 0; \forall\, j > i$; ii) $S_i(n+1) = 1$; iii) $i \geq s^{k+1}_n$; iv) $i \leq s^{k-1}_{n+1}$, i.e., a channel reported at time slot $n+1$ as unavailable for time slot $n+k-1$ is not available at time slot $n+k$. As a consequence, (24) holds. By accounting for the previous results and for Corollary 5 and 6, we have finally the thesis (20). ■

E. Proof of Lemma 4

Proof: We prove the theorem with a *reductio ad absurdum* by supposing that:

$$\exists\, x^1_n \in \Omega, x^1_n > s^1_n : v_\pi(\mathbf{x}_n) < v_{\pi'}(\mathbf{x}_n), \quad (34)$$

where $\mathbf{s}_n = \left\{s^1_n, s^2_n, \ldots, s^K_n\right\}$ and $\mathbf{x}_n = \left\{x^1_n, s^2_n, \ldots, s^K_n\right\}$, and where $\pi, \pi' \in \Sigma$ satisfying (17) with $\pi^*(\mathbf{s}_n) = \pi(\mathbf{s}_n)$ and $\pi^*(\mathbf{x}_n) = \pi'(\mathbf{s}_n)$. ■

By accounting for (22), we note that:

$$p_a(\mathbf{s}_{n+1}|\mathbf{s}_n) = p_a(\mathbf{s}_{n+1}|\mathbf{x}_n)\ \forall\, \mathbf{s}_{n+1} \in \Sigma', \forall\, a \in A'. \quad (35)$$

Hence, it results:

$$v_\pi(\mathbf{x}_n) < v_{\pi'}(\mathbf{x}_n) \Longrightarrow \\ \Longrightarrow r_{x^1_n} + v_\pi\left(\left\{s^2_n, \ldots, s^K_n, -1\right\}\right) < \\ < r_{x^1_n} - c + \sum_{\mathbf{s}_{n+1}\in\Sigma'} p_{\tilde{a}_0}(\mathbf{s}_{n+1}|\mathbf{x}_n)\, v_{\pi'}(\mathbf{x}_{n+1}) \\ \Longrightarrow v_\pi\left(\left\{s^2_n, \ldots, s^K_n, -1\right\}\right) < \\ < -c + \sum_{\mathbf{s}_{n+1}\in\Sigma'} p_{\tilde{a}_0}(\mathbf{s}_{n+1}|\mathbf{x}_n)\, v_{\pi'}(\mathbf{x}_{n+1})\ . \quad (36)$$

Similarly, it results:

$$v_{\pi'}(\mathbf{s}_n) < v_\pi(\mathbf{s}_n) \Longrightarrow \\ \Longrightarrow r_{s^1_n} - c + \sum_{\mathbf{s}_{n+1}\in\Sigma'} p_{\tilde{a}_0}(\mathbf{s}_{n+1}|\mathbf{s}_n) v_{\pi'}(\mathbf{s}_{n+1}) < \\ < r_{s^1_n} + v_\pi\left(\left\{s^2_n, \ldots, s^K_n, -1\right\}\right) \\ \Longrightarrow -c + \sum_{\mathbf{s}_{n+1}\in\Sigma'} p_{\tilde{a}_0}(\mathbf{s}_{n+1}|\mathbf{s}_n) v_{\pi'}(\mathbf{s}_{n+1}) < \\ < v_\pi\left(\left\{s^2_n, \ldots, s^K_n, -1\right\}\right), \quad (37)$$

and, by accounting for (35), (36) and (37) constitute a *reductio ab absurdum*.

REFERENCES

[1] M. Nekovee, T. Irnich, and J. Karlsson, "Worldwide trends in regulation of secondary access to white spaces using cognitive radio," *IEEE Wireless Commun.*, vol. 19, no. 4, pp. 32–40, Aug. 2012.

[2] A. Flores, R. Guerra, E. Knightly, P. Ecclesine, and S. Pandey, "IEEE 802.11af: A standard for TV white space spectrum sharing," *IEEE Commun. Mag.*, vol. 51, no. 10, pp. 92–100, Oct. 2013.

[3] C.-S. Sum *et al.*, "Cognitive communication in TV white spaces: An overview of regulations, standards, and technology," *IEEE Commun. Mag.*, vol. 51, no. 7, pp. 138–145, Jul. 2013.

[4] A. Cacciapuoti, M. Caleffi, and L. Paura, "Widely linear cooperative spectrum sensing for cognitive radio networks," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, 2010, pp. 1–5.

[5] I. F. Akyildiz, B. F. Lo, and R. Balakrishnan, "Cooperative spectrum sensing in cognitive radio networks: A survey," *Phys. Commun.*, vol. 4, no. 1, pp. 40–62, 2011.

[6] A. S. Cacciapuoti, M. Caleffi, D. Izzo, and L. Paura, "Cooperative spectrum sensing techniques with temporal dispersive reporting channels," *IEEE Trans. Wireless Commun.*, vol. 10, no. 10, pp. 3392–3402, Oct. 2011.

[7] M. Caleffi and A. S. Cacciapuoti, "Database access strategy for TV white space cognitive radio networks," in *Proc. IEEE Int. Conf. Sens. Commun. Netw. (SECON)*, 2014, pp. 1–5.

[8] FCC, "ET docket 10-174: Second memorandum opinion and order in the matter of unlicensed operation in the TV broadcast bands," FCC, Active Regulation, Sep. 2012.
[9] OfCom, "Regulatory requirements for white space devices in the UHF TV band," OfCom, Active Regulation, Jul. 2012.
[10] 802.22 Working Group on Wireless Regional Area Networks, *Part 22: Cognitive Wireless RAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Policies and Procedures for Operation in the TV Bands*, IEEE, Official Standard, 2011.
[11] 802.15 Wireless Personal Area Network Working Group, *802.15.4 m-2014—Part 15.4: Low-Rate Wireless Personal Area Networks (LR-WPANs) Amendment 6: TV White Space Between 54 MHz and 862 MHz Physical Layer*, IEEE Active Standard, Dec. 2014.
[12] *IEEE Standard for Information Technology—Telecommunications and Information Exchange Between Systems–Local and Metropolitan Area Networks—Specific Requirements—Part 19: Tv White Space Coexistence Methods*, IEEE Active Standard 802.19.1-2014, 2014.
[13] M. Caleffi and A. S. Cacciapuoti, "Interference analysis for secondary coexistence in tv white space," *IEEE Commun. Lett.*, vol. 19, no. 3, pp. 383–386, Mar. 2015.
[14] M. Caleffi, A. S. Cacciapuoti, and L. Paura, "On the coexistence of cognitive radio ad hoc networks in tv white space," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2015, pp. 1–5.
[15] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Hoboken, NJ, USA: Wiley, 1994.
[16] A. Cacciapuoti, M. Caleffi, F. Marino, and L. Paura, "Routing update period in cognitive radio ad hoc networks," in *Proc. IEEE Int. Workshop Meas. Netw.*, 2013, pp. 100–104.
[17] Y. Han, E. Ekici, H. Kremo, and O. Altintas, "A survey of MAC issues for TV white space access," *Ad Hoc Netw.*, vol. 27, pp. 195–218, 2015.
[18] A. Cacciapuoti, M. Caleffi, F. Marino, and L. Paura, "Maximizing the route capacity in cognitive radio networks," in *Proc. IEEE Int. Conf. Sens. Commun. Netw. (SECON)*, 2014, pp. 19–23.

Marcello Caleffi (M'12) received the Dr. Eng. degree (*summa cum laude*) in computer science engineering from the University of Lecce, Lecce, Italy, and the Ph.D. degree in electronic and telecommunications engineering from the University of Naples Federico II, Naples, Italy, in 2005 and 2009, respectively. Currently, he is an Assistant Professor with DIETI Department, University of Naples Federico II, where he teaches telematics and mobile communication systems classes and has supervised and graduated more than 20 B.S. and M.S. students. From 2010 to 2011, he was a Visiting Researcher at the Broadband Wireless Networking Laboratory, Georgia Institute of Technology, Atlanta, GA, USA, and, in 2011, at the NaNoNetworking Center in Catalunya (N3Cat), Universitat Politcnica de Catalunya (UPC), Barcelona, Spain. His research interests include cognitive radio networks, biological networks, and nanonetworks. He has coauthored over 40 journal articles (the IEEE ACM TRANSACTIONS ON NETWORKING, the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, the IEEE TRANSACTIONS ON COMMUNICATIONS, the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, and the IEEE INTERNET OF THINGS) and conference publications (the IEEE GLOBECOM, the IEEE ICC, the IEEE SECON, and ACM NANOCOM). Since 2014, he has been serving as an Area Editor for *Ad Hoc Networks* (Elsevier).

Angela Sara Cacciapuoti (M'10) received the Dr. Eng. degree (*summa cum laude*) in telecommunications engineering and the Ph.D degree (with score "excellent") in electronic and telecommunications engineering from the University of Naples Federico II, Naples, Italy, in 2005 and 2009, respectively. Currently, she is an Assistant Professor with DIETI Department, University of Naples Federico II. From 2010 to 2011, she was a Visiting Researcher at the Broadband Wireless Networking Laboratory, Georgia Institute of Technology, Atlanta, GA, USA, and, in 2011, at the NaNoNetworking Center in Catalunya (N3Cat), Universitat Politcnica de Catalunya (UPC), Barcelona, Spain. Her current research interests include cognitive radio networks, nanonetworks, neuronal models, and Internet of Nano-Things. Since 2014, she has been an Area Editor of *Computer Networks* (Elsevier).