Energy-efficient Opportunistic Spectrum Access in Cognitive Radio Networks

Contents

4.1 Introduction

The traditional spectrum management is based on agreements between the SP and PUs. CR is considered as the key technology that enables unlicensed users to access the licensed spectrum. Furthermore, the new spectrum-licensing paradigm, initiated by the FCC in 2008 [8], has promoted the idea of using the CR technology to face the spectrum scarcity problem. It allows unlicensed users to access the spectrum as long as they do not harm licensed users' transmissions.

Although the use of licensed bands by CR users is widely recognized, it is not well understood which applications are suitable for CR users, and what type of traffic a CR user may support. In fact, if CR users support multimedia applications, such as

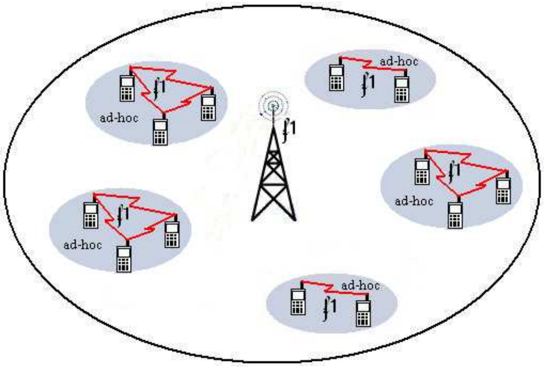

FIGURE 4.1: First use case: Using CR in ad-hoc communication. If the licensed frequency $f1$ is not used by PUs, SUs can communicate in ad-hoc mode using $f1$.

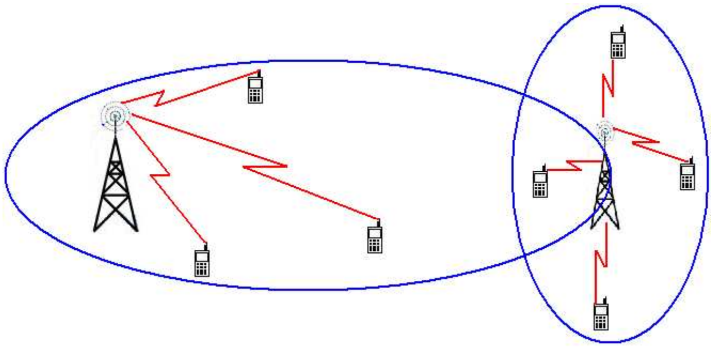

FIGURE 4.2: Second use case: using CR for BS's transmissions. If the licensed frequency $f1$ is not used by PUs, the BS serves its users using $f1$.

video streaming, VoIP or online gaming, they must be able to guarantee some QoS requirements. These motivations are behind the problem considered in this chapter.

The model that we are studying in this chapter is suited for several use-cases of the CR paradigm in wireless networks. Firstly, this model allows ad-hoc connections to use spectrum holes (frequencies that are not utilized by PUs), as illustrated in Figure 4.1. Second, we may consider that SUs are CR base stations, which are able to sense the activity of a primary base station, and take advantage of spectrum holes for transmitting on the downlink (see Figure 4.2).

Many works focused on the study of optimal OSA in CR networks (see [80], [81] and [82]). In [83], the authors focused on the OSA taking into account the energy consumption. They formulated their problem as a POMDP and derived some properties of the optimal OSA policy. Their control parameter is the duration of sensing used by a SU at each time slot in order to determine the PU's activity. They provided heuristic control policies by using grid-based approximations, which have low complexity but give suboptimal control policies. Authors of [53] incorporated the energy constraint in the design of the optimal OSA policy. They formulated the problem also using a POMDP with a finite horizon criterion. They established a threshold structure of the optimal policy for the single channel model without providing analytical expression of the threshold. The main difference between these works and ours is that we consider not only the energy consumption, but also the transmission delay. Moreover, we consider a POMDP with an average reward criterion. Authors of [84] analyzed the latency of DSA in CR networks by considering a dedicated control with embedded control channel. In [85], authors considered an adaptive modulation scheme in order to guarantee a delay for SUs. The difference between their work and our's is that they considered a dynamic spectrum sharing and we are considering an OSA context.

It is noteworthy that the impact of the energy consumption or the capacity of CR users to support additional QoS requirements such as the expected delay, to the best of our knowledge, has been somehow ignored in the literature, partially due to the difficulties in analyzing it. In fact, it is very important for today's wireless networks to garantee a certain level of QoS. As SUs are not licensed to use the spectrum, the transmission delay of their packets depends not only on the PUs' activity but also on the competition with each other.

Our main contribution is to consider, in this CR setting, an optimal OSA mechanism that takes into account energy consumption and transmission delay. Note that, taking into account the delay as well as the energy consumption significantly complicates the optimization problem. For instance, without considering the delay constraint, the SU achieves the best tradeoff between trying to access licensed channel and sleeping to conserve energy. However, the design of energy-QoS tradeoff lies among several conflicting objectives: gaining immediate access, gaining spectrum occupancy information, conserving energy, and minimizing packet delays. The novelty of this work is to study an energy-QoS tradeoff OSA mechanism for SUs in a CR network. The major contributions of this chapter are:

- The problem is formulated as an infinite horizon POMDP with average criterion. Usually, OSA mechanisms for CR networks were modeled using POMDPs with expected total discounted reward (see [80], [83] and [27] for some examples). However, as decisions are taken frequently by SUs (every time slot) the discount rate is very close to 1. Thus, the average expected reward is more suited to model OSA mechanisms [86].
- In order to gain insights into the energy-delay constrained OSA, we derive structural properties of the value function. We are able to show that the value function is increasing with the belief and decreasing with packet delays. These structural results not only give us insights about the optimal OSA policy, but also reduce

the computational complexity when seeking for the optimal policies. In fact, the value function can be approximated by simple functions (see [87]).

- We show that the SUs maximize their average rewards by adopting a simple threshold policy, and we derive closed-form expressions of these thresholds.
- Since SUs may use a dedicated channel for their packets, the optimal threshold policy guarantees a bounded delay.
- We propose some learning algorithm to estimate the RF environment on-the-fly.

The organization of this chapter is as follows. In the next section, we describe the primary and the secondary user models. Section 4.3 presents our Markov decision process framework. In Section 4.4, we study the existence of an optimal threshold policy for our opportunistic spectrum access with an energy-QoS tradeoff. We propose two learning based protocols for the estimation of state transition rates in Section 4.5. Before concluding the chapter, we present, in Section 4.6, some numeric illustrations.

4.2 Model

We consider a wireless network where N independent channels are licensed to PUs. The state of each channel $n \in \{1, \ldots, N\}$ is modeled by a time-homogeneous discrete Markov process $s_n(t)$. The state space is $\{0, 1\}$, where $s_n(t) = 0$ means that the channel n is free for SUs' access, and $s_n(t) = 1$ means that the channel n is occupied by PUs. The following matrix gives the transition probabilities of the channel n:

$$P_n = \begin{pmatrix} \alpha_n & 1-\alpha_n \\ \beta_n & 1-\beta_n \end{pmatrix}.$$

In fact, SUs observe a "good" channel (ON) if PUs are not using the licensed channel. On the other hand, the presence of PUs in the licensed channel results in a "bad" channel (OFF) for SUs. Therefore, the licensed channels can be modeled by the ON/OFF Gilbert-Elliot model [88], [89]. The transition rates evolve as illustrated in Figure 5.1. Note that this model was used in several works in the OSA area (see [80], [27], [83] and [53] for some examples).

The global state of the system, composed of the N channels, is denoted by the vector $\mathbf{s}(t) = [s_1(t), ..., s_N(t)]$, and the global state space is $\mathcal{S} = \{0, 1\}^N$. The transition probabilities can be determined by statistics of the PUs' traffic, and are assumed to be known by SUs. We present, in Section 4.5, some methods allowing the SU to estimate these transition probabilities on-the-fly.

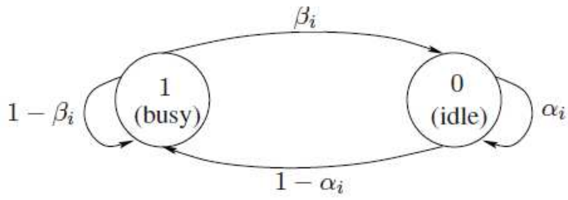

FIGURE 4.3: The channel transition probabilities for channel i.

We consider that all the N licensed channels are open for SUs' transmissions when PUs are not using them. The aim of SUs is to find licensed channels that are not used by PUs during a given time slot. Note that looking for opportunities in licensed channels may induce not only a large packet delay, but also higher energy consumption, spent for sensing and transmissions over licensed channels. This may be caused by high traffic of PUs or collisions between SUs. For this end, we consider an OSA that takes into account packet delay, throughput and energy consumption. In order to introduce some QoS guarantee for SUs, we assume that, at any time slot, SUs have access to the network through another technology referred to as dedicated channel. This assumption ensures a higher bound of packet delays, while benefiting from licensed spectrum holes. Indeed, the aim of SUs is to find a tradeoff between the following conflicting objectives: transmitting with a guaranteed delay, but with higher cost using the dedicated channel, or transmitting with a lower cost using the licensed channels, but without delay guarantee.

The objective of SUs is to minimize the transmission cost accounting for energy consumption and transmission delay, i.e. a QoS guarantee with the lowest possible cost. In order to achieve such goal, a SU has to choose at each time slot one of the following actions:

- to be inactive during the time slot in order to save energy,
- to sense a licensed channel and to transmit if the channel is available during the time slot, else to wait for next time slot,
- to sense a licensed channel and to transmit if the channel is available during the time slot, else to use the dedicated channel.

Figure 4.4 illustrates the action diagram for SUs. Our important contribution is to consider the average packet delay in the optimal decision. Moreover, we consider that sensing licensed channels has a cost for SUs, which models the energy spent when sensing licensed channels. Given these constraints, we seek for an optimal OSA policy for SUs in CR networks. In the remainder of this chapter, we focus on the model of one SU accessing opportunistically licensed channels. The multi-user context will be studied in the following chapter.

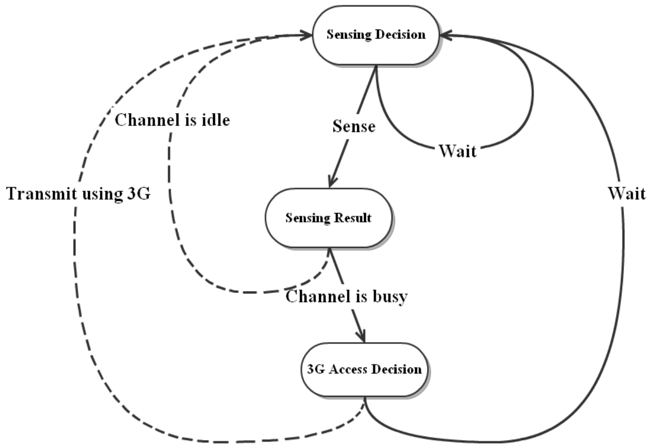

FIGURE 4.4: The action diagram for SUs. There are mainly two decision steps: the sensing decision and the access decision.

4.3 POMDP framework

The global system state $\mathbf{s}(t)$ cannot be directly observed by a SU. To overcome this difficulty, the SU infers the global state of the system based on observations that can be summarized in a belief vector $\omega(t) = \{\omega_1(t), ..., \omega_{2^N}(t)\}$, where $\omega_j(t)$ is the conditional probability (given observations and decisions history) that the system state $\mathbf{s}(t) = j$, at the time slot t. Since the N channels are independent, we may consider the following simpler belief vector:

$$\lambda(t) = [\lambda_1(t), .., \lambda_N(t)],$$

where $\lambda_i(t)$ is the conditional probability that the channel i is available at the time slot t. This approximation was used in several analysis such as [90] and [91]. Hence, we study the OSA for SUs in CR networks as a POMDP problem. Our OSA mechanism can be formulated using a POMDP framework described as follows:

State We define the state of the system as a composition of belief and delay $(\lambda(t), l(t))$. The delay of a packet held by a SU is denoted by $l(t)$. When the SU receives a new packet, its delay equals one, and increases by one every time slot, except when the SU transmits the packet. We assume that the SU does not accept a new packet until it transmits the held one. We take this assumption in order to evaluate the impact of the OSA mechanism on the delay of the packets. As part of my future work, I will analyze the effect of traffic characteristics such as throughput and traffic model on the performance of the OSA mechanism. Note that the SU does not have accurate information about the first part of the system state, i.e. the belief vector $\lambda(t)$, but has a perfect knowledge about the packet delay.

Action The SU makes a two-level decision. It chooses, first, whether to sense the licensed channels or not. Then, it decides about the transmission over the licensed channels or using the dedicated one. It is straightforward that the SU transmits over the licensed channels if idle. Therefore, the second step decision is taken when there are no opportunities in licensed bands. In fact, when the licensed channel was sensed as occupied, the SU has two options:

- to wait for the next slot;
- to transmit using the dedicated channel.

Without loss of generality, we assume that both decisions are made at the beginning of the time slot. For each time slot t and each state $(\lambda(t), l(t))$, we consider that the three possible actions for SUs are:

$$a(t) = \begin{cases} 0: & \text{to be inactive;} \\ 1: & \text{to sense and to transmit only if the channel is available during the slot;} \\ 2: & \text{to sense and to transmit if the channel is available during the time slot,} \\ & \text{else to transmit through the dedicated channel.} \end{cases}$$

Observation and belief When the SU decides to sense (i.e. to take action $a(t) \in \{1, 2\}$), one channel $n^*(t)$ is determined and the SU observes the channel occupancy state $s_{n^*(t)}(t) \in \{0, 1\}$. Let $\theta(t)$ be the observation outcome at the time slot t, where $\theta(t) = 0$ if the channel is sensed as idle, and $\theta(t) = 1$ otherwise. The SU takes into account the history of observations and actions by updating the belief vector based on observation outcomes. For each channel n, the conditional probability, $\lambda_n(t + 1) := Pr(s_n(t + 1) = 0|a(t), \theta(t))$, is updated as follows:

$$\lambda_n(t+1) = \begin{cases} \beta_n + (\alpha_n - \beta_n)\lambda_n(t) & \text{if } a(t) = 0 \text{ or } n \neq n^*(t), \\ \alpha_n & \text{if } \quad a(t) \neq 0, \text{or } \theta(t) = 0 \text{ and } n = n^*(t), \\ \beta_n & \text{if } \quad a(t) \neq 0, \text{or } \theta(t) = 1 \text{ and } n = n^*(t). \end{cases} \quad (4.1)$$

The belief update function depends mainly on licensed channels' transition rates α and β. Indeed, these statistics may not be available for SUs. Specifically, we propose, in Section 4.5, some learning methods that allow the SUs to estimate the RF environment on-the-fly. Note that we can extend easily our model to sense not only one licensed channel, but also a subset of the licensed channels.

Channel choice policy At a given time slot t, the SU chooses a licensed channel $n^*(t) \in N$ to sense based on its belief vector $\lambda(t)$. There exists several channel choice policies in the literature like total sensing (see [66] and [59]), opportunistic sensing (see

[80] and [53]), randomized sensing (see [69] and [70]), and periodic sensing (see [67] and [68]). An example of opportunistic and greedy channel choice policy is to sense the channel that has the highest probability to be idle, i.e. $n^*(t) := \arg\max_n(\lambda_n(t))$.

Policies We define a sensing and access policy μ as a vector $[\mu_1, \mu_2, \ldots]$, where μ_t is a mapping from a state $(\lambda(t), l(t))$ to an action $a(t)$ at the time slot t. We denote by Γ the set of all possible policies.

Reward and costs The SU tries to maximize its revenue by increasing the reward (obtained from successful transmissions) and decreasing the costs (spent for sensing and transmissions). Note that the cost of transmission over the dedicated channel is higher than the cost paid for transmission using the licensed channels. The different cost and rewards for SUs are denoted by:

- Reward: Let Φ be the reward representing the number of delivered bits when the SU transmits its packet.
- Costs: Let c_s be the energy cost function for sensing a licensed channel, measured as monetary units. This function depends on the action $a(t)$ taken by the SU as follows:

$$c_s(a(t)) = \begin{cases} c_s, & \text{if } a(t) > 0, \\ 0, & \text{if } a(t) = 0. \end{cases}$$

  The PU and the SP (for the dedicated access), charge a price for each successfully transmitted packet. Those prices are respectively P_p for a transmission over a licensed channel and P_{3G} for a transmission over the dedicated channel. Indeed, P_{3G} is higher than P_p. Therefore, when the SU transmits successfully a packet, it obtains the reward $z_t(a(t), \theta(t))$, which depends on the action $a(t)$ and the observation $\theta(t)$, and is expressed as follows:

$$z_t(a(t), \theta(t)) = \begin{cases} 0, & \text{if } a(t) = 0, \\ \Phi - P_p & \text{if } a(t) \geq 1 \text{ and } \theta(t) = 0, \\ \Phi - P_{3G}, & \text{if } a(t) = 2 \text{ and } \theta(t) = 1. \end{cases}$$

- Delay: In order to model the impact of the delay, we introduce an additional cost when a packet is not transmitted. This cost depends on the current delay $l(t)$ of the packet, and is defined by the function $f(l(t))$. This function is assumed to be increasing with $l(t)$, in order to growth the incentive of transmitting the packet when it becomes delayed.
- Instantaneous reward: At the time slot t, the instantaneous reward r_t of a SU depends on the system state $(\lambda(t), l(t))$ and the action $a(t)$, and is expressed as

follows:

$$r_t((\lambda(t), l(t)), a(t)) = z_t(a(t), \theta(t)) - f(l(t)) - c_s(a(t)).$$

The problem faced by the SU consists of maximizing its average expected reward:

$$\bar{R}(\mu) = \lim_{T\to\infty} \frac{1}{T} \mathbb{E}_\mu \left(\sum_{t=1}^{T} r_t((\lambda(t), l(t)), a(t)) | \lambda(0), l(0) \right),$$

while $\lambda(0)$ is the initial belief vector. It is very important to consider the average reward rather than the total reward or the discounted cost as the SU takes frequently decisions. Then, our objective is to find an optimal sensing policy μ^* that maximizes the average expected reward $\bar{R}(\mu)$:

$$\mu^* = arg \max_{\mu\in\Gamma} \lim_{T\to\infty} \frac{1}{T} \mathbb{E}_\mu \left(\sum_{t=1}^{T} r_t((\lambda(t), l(t)), a(t)) | \lambda(0), l(0) \right). \tag{4.2}$$

For simplicity reasons, and to get deep theoretical analysis, we may restrict our study to the set of stationary policies. A stationary policy is a mapping that specifies for each state, independently of time slots, an action to be chosen. Note that looking for stationary policies reduce significantly the computational complexity of the OSA problem. In some particular MDP and POMDP problems, we are able to determine an optimal policy in a smaller set reduced to stationary policies. We prove in the following proposition that there exists an optimal stationary policy for our POMDP problem.

Proposition 4.1. *There exists an optimal stationary policy for our POMDP formulation of the OSA problem described in Equation (4.2).*

Proof. The proof results from Theorems 8.10.9 and 8.10.7 of [92]. Note that we have a POMDP with a discrete state space.

First, the immediate reward $r_t((s, l), a)$ is finite, i.e. $-\infty < r_t((s, l), a) < +\infty$ (as all costs and rewards are finite). Second, we prove that there exists a stationary policy d^∞ for which the derived Markov chain is positive recurrent.

Let us focus on the following belief vector:

$$\Lambda_0 = (\lambda_1, \lambda_2, ..., \lambda_N) \quad \text{such that} \quad \lambda_j = \Omega^{j-1}(\beta_j|0), \quad \text{for} \quad j = 1, \dots, N,$$

where λ_j represents the belief of a channel that was not sensed for j successive slots.

Denote by d^∞ the stationary policy which senses licensed channels at every slot, with a greedy channel choice policy. Suppose that $\alpha \geq \beta$, the analysis of the other case is analogous. Let us prove that the derived Markov chain is positive recurrent. Note that if a SU senses the licensed channel i as busy, its belief equals β, the least belief probability over all the licensed channels, and therefore the SU will choose another licensed channel to sense in the next time slot as it is considering a greedy channel choice policy. The probability that the system returns to the initial belief from any state Λ is $p(\lambda) = \prod_{k=0}^{N}(1 - \Omega^n(\lambda_j)) > 0, n \in \{O, ..., N\}$, and then the return time to the initial belief τ_j follow a geometric distribution so that $E\{\tau_j\} = \frac{1}{p(\Lambda_j)}$. Therefore, all state are positive recurrent under d^∞.

Third, let us prove that $g^{d^\infty} > -\infty$ and the set $\{b \in \mathcal{S}_b : r_t((s,l),a) > g^{d^\infty}$ for some $a \in \mathcal{A}\}$ is finite and no empty. As the policy d^∞ senses licensed channels every slot, $g^{d^\infty} = -f(l(t)) - c_s - (f(l(t)) + P_p - \Phi)\lambda_{n^*}$. If we have the following inequality

$$-f(l(t)) - c_s - (f(l(t)) + P_p - \Phi)\lambda_{n^*} > \max\{-f(l(t)), \Phi - c_s - P_{3G} - (P_p - P_{3G})\lambda_{n^*}\}$$

for all belief b, the policy always sense licensed channels is optimal and we have achieved our goal. Otherwise, the set $\{b \in \mathcal{S}_b : r_t((s,l),a) > g^{d^\infty}$ for some $a \in \mathcal{A}\}$ is finite and no empty.

Finally, we obtain from the theorems 8.10.9 and 8.10.7 of [92] that there exists an average optimal stationary policy. □

Given this result, we can restrict our problem to the set Γ_S of stationary policies. Then, for the rest of this chapter, we omit the time index t, and we look for an optimal sensing policy that is a mapping from a system state (λ, l) to an action a, independently of the time slot t. Before seeking for the optimal OSA policy, we make an analysis of the value function of the POMDP problem.

We denote by $\Omega^{ns}(\lambda|\theta)$, the function that updates the belief vector λ when the user chooses to be inactive in the current slot, i.e. the SU takes action 0. The function $\Omega^s(\lambda|\theta)$ updates the belief vector λ when the SU senses a licensed channel in the current slot and observes θ, i.e. the SU takes the action 1 or 2.

We define, in the following, the value function $V(\lambda, l)$. Let us denote by $Q_a(\lambda, l)$ the action-value function of taking the action a in the current slot when the information state is (λ, l). Therefore, the value function is expressed as follows:

$$V(\lambda, l|\lambda_0, l_0) = \max_{a \in \mathcal{A}} \left(g_u(\lambda_0, l_0) + Q_a(\lambda, l|\lambda_0, l_0)\right), \tag{4.3}$$

where (λ_0, l_0) is the initial state of the system and $g_u(\lambda_0, l_0)$ is a constant that depends on the initial state. Note that for any stationary policy, the state of the SUs is an irreducible Markov chain with one ergodic class. Thus, a unique steady state probability exists, and we can omit the initial distribution. Thus, the value function for our POMDP problem can be expressed as follows:

$$g_u + V(\lambda, l) = \max_{a \in \mathcal{A}} Q_a(\lambda, l), \tag{4.4}$$

where g_u is a constant. The optimal policy for our POMDP problem is the one that chooses the following action in the state (λ, l) :

$$a^*(\lambda, l) = \arg\max_{a \in \mathcal{A}} Q_a(\lambda, l). \tag{4.5}$$

We determine the action-value function for each different action 0, 1 and 2. When the SU decides to wait, i.e. to take the action $a = 0$, we have:

$$Q_0(\lambda, l) = -f(l) + V(\Omega^{ns}(\lambda|\theta = 0), l + 1). \tag{4.6}$$

When the SU chooses to sense the channel n^* and decides to wait for the next time slot if the channel n^* is busy, i.e. to take action 1, we have:

$$\begin{aligned} Q_1(\lambda, l) \quad = \quad & -c_s + \lambda_{n^*}(\Phi - P_p + V(\Omega^s(\lambda|\theta = 0), 1)) \\ & +(1 - \lambda_{n^*})(-f(l) + V(\Omega^s(\lambda|\theta = 1), l + 1)). \end{aligned} \tag{4.7}$$

When the SU chooses to sense the channel n^* and to transmit using the dedicated channel if the channel n^* is busy, i.e. to take action 2, we have:

$$\begin{aligned} Q_2(\vec{\lambda}, l) \quad = \quad & \Phi - c_s + \lambda_{n^*}(-P_p + V(\Omega^s(\lambda|\theta = 0), 1)) \\ & +(1 - \lambda_{n^*})(-P_{3G} + V(\Omega^s(\lambda|\theta = 1), 1)). \end{aligned} \tag{4.8}$$

For the remainder of this chapter, we take some assumptions that simplify the analysis of the optimal policy. We focus on the case of one licensed channel. The multichannel case will be studied in Section 4.3.2. We take the assumption that there exists a packet delay l^* such that the SU transmits its packet using the dedicated channel if the observation is $\theta = 1$. In fact, this assumption is somehow realistic, as the SU has no interest to keep the file in its buffer indefinitely. We denote by α and β the transition rates of the licensed channel, and λ the belief of the SU.

4.3.1 The single channel model

To solve the POMDP problem, the belief vector is a key element as it gives us insights about the system state. Firstly, we analyze the belief update function. The following lemma gives us some properties of the belief update function Ω^{ns}. We consider that $\alpha \geq \beta$. When $\alpha \leq \beta$, the analysis is similar and results are analogous.

Lemma 4.2. *We have the following properties of the belief update function Ω^{ns}.*

1. *The update function $\Omega^{ns}(\lambda|\theta)$ is increasing with belief λ.*
2. *We have the following equivalence:*

$$\Omega^{ns}(\lambda|\theta) \geq \lambda \quad \Leftrightarrow \quad \lambda \leq \pi(0),$$

and

$$\Omega^{ns}(\lambda|\theta) \leq \lambda \quad \Leftrightarrow \quad \lambda \geq \pi(0),$$

where $\pi(0) = \frac{\beta}{1-\alpha+\beta}$ is the stationary probability that the licensed channel is idle.

Proof. First, the update function Ω^{ns} is linear with the belief because $\Omega^{ns}(\lambda) = \beta + (\alpha - \beta)\lambda$. As we have considered the case where $\alpha \geq \beta$, then the update function is increasing with the belief.

Second, let us prove that $\Omega^{ns}(\lambda) \geq \lambda$ for all beliefs $\lambda \leq \pi(0)$ by induction on the belief.

1. We have the initial condition: $\beta \leq \pi(0) = \frac{\beta}{1-\alpha+\beta}$ and $\Omega^{ns}(\beta) = \beta + (\alpha - \beta)\beta \geq \beta$.
2. Assume that we have: $\Omega^{ns}(\lambda) \geq \lambda$, for a given $\lambda \leq \pi(0)$.
3. The induction operator derives the following belief value: $\Omega^{ns}(\Omega^{ns}(\lambda)) = \beta + (\alpha - \beta)\Omega^{ns}(\lambda) \geq \beta + (\alpha - \beta)\lambda = \Omega^{ns}(\lambda)$.

Thus, $\Omega^{ns}(\lambda) \geq \lambda$ for all $\lambda \leq \pi(0)$. The analysis for $\lambda \geq \pi(0)$ is similar. □

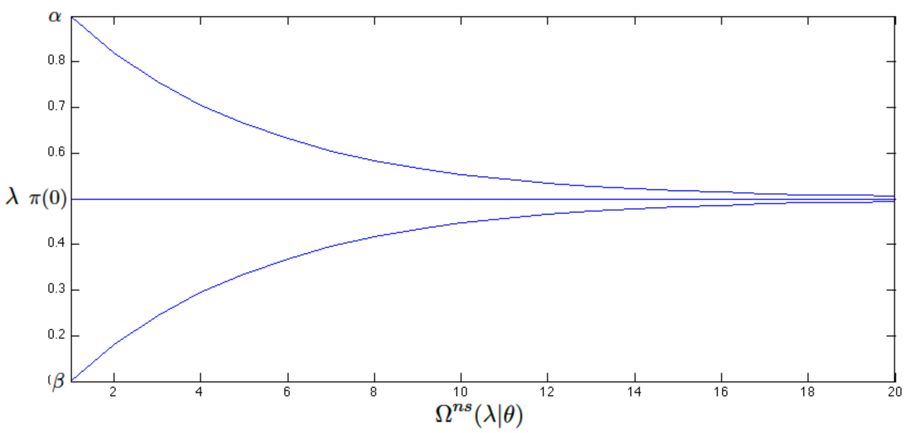

FIGURE 4.5: The belief update function Ω^{ns} with respect to number of time slots the channel was not sensed.

Figure 4.5 depicts the belief evolution. As shown in [34], the value function for a POMDP over a finite time horizon is PWLC with respect to the belief vector. In Proposition 4.3, we show that the value function for our POMDP problem over an infinite horizon with the average criterion has also this property.

Proposition 4.3. *The value function $V(\lambda, l)$, given by Equation (4.4), is PWLC with respect to the belief vector λ over an infinite horizon with average criterion.*

Proof. The proof of the Proposition 4.3 is similar to [34] where the authors considered the finite time horizon problem. Hence, we briefly describe the procedure for this proof. For all belief vectors λ, the value function $V(\lambda, l^*)$ is linear with the belief:

$$\begin{aligned} V(\lambda, l^*) &= Q_2(\lambda, l^*) - g_u, \\ &= -g_u + \Phi - c_s - P_{3G} + V(\Omega^s(\lambda|\theta=1), 1) + \\ &\quad \lambda_{n^*}(P_{3G} - P_p + V(\Omega^s(\lambda|\theta=0), 1) - V(\Omega^s(\lambda|\theta=1), 1)). \end{aligned}$$

Then the value function $V(\lambda, l^*)$ can be rewritten as an inner product of the belief vector and a Υ-vector. As $Q_2(\lambda, l) = Q_2(\lambda, l^*)$, for all l, the action-value function $Q_2(\lambda, l)$ can be also rewritten as an inner product of the belief vector and a Υ-vector. We suppose that Proposition 4.3 holds for all packet delays higher than $l+1$, and we prove that the proposition is true for packet delay l. After some algebra, we can rewrite the action-value functions given in Equations (4.6) and (4.8) in terms of Υ-vector as follows:

$$Q_0(\lambda, l) = -f(l) + \max_{\Upsilon \in \Gamma_{l+1}} < \Omega^{ns}(\lambda|\theta), \Upsilon >= -f(l) + \sum_{s \in \mathcal{S}} \omega_s \left[\sum_{s' \in \mathcal{S}} P(s'|s) \Upsilon_{l+1}^{\Omega^{ns}(\lambda|\theta)} \right], \quad (4.9)$$

and

$$\begin{aligned} Q_1(\lambda, l) &= -c_s + \lambda(\phi - P_p + V(\alpha, 1)) + (1-\lambda)(-f(l) + \max_{\Upsilon \in \Gamma_{l+1}} < \Omega^s(\lambda|\theta=1), \Upsilon >) \\ &= -c_s + \lambda(\phi - P_p + V(\alpha, 1)) + (1-\lambda) \left(-f(l) + \sum_{s \in \mathcal{S}} \omega_s \left[\sum_{s' \in \mathcal{S}} P(s'|s) \Upsilon_{l+1}^{\Omega^s(\lambda|\theta=1)} \right] \right), \quad (4.10) \end{aligned}$$

where $\Upsilon_{l+1}^{\Omega^{ns}(\lambda|\theta)}$ and $\Upsilon_{l+1}^{\Omega^s(\lambda|\theta=1)}$ are, respectively, the Υ-vectors for the regions containing belief vectors $\Omega^{ns}(\lambda|\theta)$ and $\Omega^s(\lambda|\theta=1)$, respectively. Each term in the square brackets of Equations (4.9) and (4.10) are elements $\Upsilon_{\lambda,l}$ of a Υ-vector Υ_l. Thus, the action-value functions can be rewritten as an inner product of the belief vector and a Υ-vector Υ_l. Moreover, there is only a finite number of such Υ-vector Υ_l, since we have a finite set of belief for all l. As the maximum of a finite set of piecewise linear and convex functions is also piecewise linear and convex, the Proposition 4.3 holds for all beliefs and packet delays. □

Note that monotonicity results help us for establishing the structure of the optimal policies (see [93] for an example) and provide insights into the underlying problem. The following propositions states monotonicity results of the value function with respect to each of its parameters.

Proposition 4.4. *For a given belief vector* λ*, the value function is monotonically decreasing with packet delays, i.e.* $V(\lambda, l) \leq V(\lambda, l')$ *for* $l \geq l'$*.*

Proof. Let us prove that the value function $V(\lambda, l)$ is monotonically decreasing with packet delays, for a given belief vector λ. Note that SUs take the action 2 for all beliefs λ when the packet delay is l^*. Therefore, we have:

$$V(\lambda, l^*) = \Phi - c_s + \lambda(-P_p + V(\alpha, 1)) + (1-\lambda)(-P_{3G} + V(\beta, 1)).$$

Note that the SU chooses the action that maximizes its average expected reward for the packet delay $l^* - 1$ and belief λ, as follows:

$$\begin{aligned} V(\lambda, l^*-1) &= \max_a Q_a(\lambda, l^*-1) - g_u \\ &\geq Q_2(\lambda, l^*-1) - g_u, \\ &\geq \Phi - c_s + \lambda(-P_p + V(\alpha, 1)) + (1-\lambda)(-P_{3G} + V(\beta, 1)) - g_u, \\ &\geq V(\lambda, l^*). \end{aligned}$$

Let us prove that this propriety holds for all packet delays using a backward induction on packet delays:

1. Initial condition: For all belief vector λ, we have that: $V(\lambda, l^*) \leq V(\lambda, l^*-1)$,
2. Suppose that $V(\lambda, l+2) \leq V(\lambda, l+1)$, $\forall \lambda$.
3. We have:

$$\begin{aligned} Q_0(\lambda, l) &= -f(l) + V(\Omega^{ns}(\lambda|\theta), l+1), \\ &\geq -f(l+1) + V(\Omega^{ns}(\lambda|\theta), l+2), \\ &= Q_0(\lambda, l+1). \end{aligned}$$

$$\begin{aligned}
Q_1(\lambda, l) &= -c_s + \lambda(\Phi - P_p + V(\alpha, 1)) + (1-\lambda)(-f(l) + V(\beta, l+1)), \\
&\geq -c_s + \lambda(\Phi - P_p + V(\alpha, 1)) + (1-\lambda)(-f(l+1) + V(\beta, l+2)), \\
&= Q_1(\lambda, l+1). \\
Q_2(\lambda, l) &= -c_s + \Phi - P_{3G} + V(\beta, 1) + \lambda(P_{3G} - P_p + V(\alpha, 1) - V(\beta, 1)), \\
&= Q_2(\lambda, l+1).
\end{aligned}$$

The inequalities come from the induction assumption and the monotonicity of the penalty function $f(l)$.

Finally, the value function is decreasing with packet delays. □

This result is intuitive as for the same belief λ and for a given packet delay, the maximum expected remaining reward that can be accrued is lower than the one the SU can get with a smaller packet delay. We present, in the following lemma, a result that will be useful for the proof of the monotonicity of the value function with respect to the belief.

Lemma 4.5. *We have the following inequality:*

$$-P_p + V(\alpha, 1) \geq -P_{3G} + V(\beta, 1).$$

Proof. We prove this lemma by contradiction. Suppose that $-P_p + V(\alpha, 1) < -P_{3G} + V(\beta, 1)$. Let us prove, in the following, that the constant g_u is higher than $\Phi - c_s - P_p$:

$$\begin{aligned}
g_u + V(\alpha, 1) &\geq Q_2(\alpha, 1), \\
g_u + V(\alpha, 1) &\geq -c_s + \alpha(\phi - P_p + V(\alpha, 1)) + (1-\alpha)(\phi - P_{3G} + V(\beta, 1)), \\
g_u + V(\alpha, 1) &\geq -c_s + \phi - P_p + V(\alpha, 1), \\
g_u &> \Phi - c_s - P_p.
\end{aligned}$$

We take the assumption that the immediate reward when the channel is idle is positive, i.e. $\Phi - c_s - P_p \geq 0$. We have already assumed that the SU takes the action 2 in the state (λ, l^*) for all belief vector λ, i.e $a^*(\lambda, l^*) = 2, \forall \lambda$. Therefore, we have:

$$g_u + V(\lambda, l^*) = -c_s + \lambda(\phi - P_p + V(\alpha, 1)) + (1-\lambda)(\phi - P_{3G} + V(\beta, 1)).$$

Let us focus on the packet delay $l^* - 1$. If $\lambda \leq \pi(0)$, the following inequality holds:

$$\begin{aligned}
Q_0(\lambda, l^* - 1) &= -f(l^* - 1) + V(\Omega^{ns}(\lambda), l^*), \\
&= -g_u - f(l^* - 1) - c_s + \Omega^{ns}(\lambda)(\phi - P_p + V(\alpha, 1)) \\
&+ (1 - \Omega^{ns}(\lambda))(\phi - P_{3G} + V(\beta, 1)), \\
&= V(\lambda, l^*) - f(l^* - 1) + (\Omega^{ns}(\lambda) - \lambda)(P_{3G} - P_p + V(\alpha, 1) - V(\beta, 1)), \\
&< V(\lambda, l^*).
\end{aligned}$$

The inequality is due to the assumption that $-P_p + V(\alpha, 1) < -P_{3G} + V(\beta, 1)$, the belief update function $\Omega^{ns}(\lambda) \geq \lambda$, and the delay penalty $f(l^* - 1)$ is positive. As the value function $V(\lambda, l)$ is decreasing with packet delays (see Proposition 4.4), then we have: $Q_0(\lambda, l^* - 1) < V(\lambda, l^*) < V(\lambda, l^* - 1)$. Note that we have already proved that g_u is positive. Thus, the SU does not take the action 0 when the packet delay is $l^* - 1$. Let us focus on the action 1, we have the following inequality:

$$\begin{aligned}
Q_1(\lambda, l^* - 1) &= -c_s + \lambda(\phi - P_p + V(\alpha, 1)) + (1 - \lambda)(-f(l^* - 1) + V(\beta, l^*)), \\
&= -c_s + \lambda(\phi - P_p + V(\alpha, 1)) + (1 - \lambda)\left(\phi - g_u - f(l^* - 1) - c_s\right. \\
&\quad \left. + \beta(-P_p + V(\alpha, 1)) + (1 - \beta)(-P_{3G} + V(\beta, 1))\right), \\
&< -c_s + \lambda(\phi - P_p + V(\alpha, 1)) \\
&+ (1 - \lambda)(\phi - g_u - f(l^* - 1) - c_s - P_{3G} + V(\beta, 1)), \\
&< -c_s + \lambda(\phi - P_p + V(\alpha, 1)) + (1 - \lambda)(\phi - P_{3G} + V(\beta, 1)), \\
&= Q_2(\lambda, l^* - 1).
\end{aligned}$$

The first inequality is due to the assumption that $-P_p + V(\alpha, 1) < -P_{3G} + V(\beta, 1)$, and the second one is because g_u, $f(l^* - 1)$ and c_s are positive. Thus, the optimal strategy is to take the action 2 when the packet delay is $l^* - 1$.

Let us prove by backward induction on l, that the optimal action is the action 2 for all belief vector $\lambda \leq \pi(0)$.

- If the SU takes the action 2 when the packet delay is l^*, then it takes also the action 2 when the packet delay is $l^* - 1$.
- We suppose that SU takes the action 2 when the packet delay is $l < l^* - 1$.

- We have the following inequality:

$$
\begin{aligned}
Q_0(\lambda, l-1) &= -f(l-1) + V(\Omega^{ns}(\lambda), l), \\
&= -g_u - f(l-1) - c_s + \Omega^{ns}(\lambda)(\phi - P_p + V(\alpha, 1)) \\
&+ (1 - \Omega^{ns}(\lambda))(\phi - P_{3G} + V(\beta, 1)), \\
&= V(\lambda, l) - f(l-1) + (\Omega^{ns}(\lambda) - \lambda)(P_{3G} - P_p + V(\alpha, 1) - V(\beta, 1)), \\
&< V(\lambda, l).
\end{aligned}
$$

The inequality is due to the assumption that $-P_p + V(\alpha, 1) < -P_{3G} + V(\beta, 1)$, $\Omega^{ns}(\lambda) \geq \lambda$, and $f(l-1)$ is positive. As the value function is decreasing with the packet delay (see Proposition 4.4), then $Q_0(\lambda, l-1) < V(\lambda, l-1) + g_u$, i.e. the SU does not take the action 0 with the packet delay $l-1$. Let us compare the action-value functions $Q_1(\lambda, l-1)$ and $Q_2(\lambda, l-1)$:

$$
\begin{aligned}
Q_1(\lambda, l-1) &= -c_s + \lambda(\phi - P_p + V(\alpha, 1)) + (1-\lambda)(-f(l-1) + V(\beta, l)), \\
&= -c_s + \lambda(\phi - P_p + V(\alpha, 1)) + (1-\lambda)\,(\,\phi - g_u - f(l-1) - c_s \\
&\quad +\beta(-P_p + V(\alpha, 1)) + (1-\beta)(-P_{3G} + V(\beta, 1))\,)\,, \\
&< -c_s + \lambda(\phi - P_p + V(\alpha, 1)) \\
&+ (1-\lambda)(\phi - g_u - f(l-1) - c_s - P_{3G} + V(\beta, 1)), \\
&< -c_s + \lambda(\phi - P_p + V(\alpha, 1)) + (1-\lambda)(\phi - P_{3G} + V(\beta, 1)), \\
&= Q_2(\lambda, l-1).
\end{aligned}
$$

The first inequality is due to the assumption that $-P_p + V(\alpha, 1) < -P_{3G} + V(\beta, 1)$ and the second one is because g_u, $f(l-1)$ and c_s are positive. Thus, The optimal strategy is to take action 2 when the delay of its packet equals $l-1$.

Finally, the SU takes action 2 for all packet delays and beliefs lower than $\pi(0)$.

Let us focus on the action-value function $Q_2(\alpha, 1)$, when the packet delay is $l = 1$, we have:

$$
\begin{aligned}
Q_2(\alpha, 1) &= -c_s + \alpha(\phi - P_p + V(\alpha, 1)) + (1-\alpha)(\phi - P_{3G} + V(\beta, 1)), \\
Q_2(\alpha, 1) &= \phi - c_s - P_{3G} + V(\beta, 1) + \alpha(P_{3G} - P_p + V(\alpha, 1) - V(\beta, 1)), \\
-g_u + Q_2(\alpha, 1) &= -g_u + V(\alpha, 1) - P_p + \phi - c_s + (\alpha - 1)(P_{3G} - P_p + V(\alpha, 1) - V(\beta, 1)).
\end{aligned}
$$

As the SU takes the action 2 also for the state $(\beta, 1)$, we have the following expression of the constant g_u:

$$\begin{aligned} g_u + V(\beta,1) &= -c_s + \beta(\phi - P_p + V(\alpha,1)) + (1-\beta)(\phi - P_{3G} + V(\beta,1)), \\ g_u + V(\beta,1) &= \phi - c_s - P_{3G} + V(\beta,1) + \beta(P_{3G} - P_p + V(\alpha,1) - V(\beta,1)), \\ g_u &= \phi - c_s - P_{3G} + \beta(P_{3G} - P_p + V(\alpha,1) - V(\beta,1)). \end{aligned}$$

Thus, we obtain:

$$-g_u + Q_2(\alpha,1) = V(\alpha,1) + P_{3G} - P_p + (\alpha - \beta - 1)(P_{3G} - P_p + V(\alpha,1) - V(\beta,1)).$$

As we have assumed that $P_{3G} - P_p + V(\alpha,1) - V(\beta,1) < 0$, and $P_{3G} > P_p$, we obtain: $V(\alpha,1) + g_u \leq Q_2(\alpha,1)$, and the SU takes also the action 2 in the state $(\alpha,1)$:

$$g_u + V(\alpha,1) = Q_2(\alpha,1) = -c_s + \alpha(\phi - P_p + V(\alpha,1)) + (1-\alpha)(\phi - P_{3G} + V(\beta,1)).$$

Finally, let us evaluate the difference between $V(\alpha,1)$ and $V(\beta,1)$. We have:

$$\begin{aligned} V(\alpha,1) - V(\beta,1) &= (\alpha-\beta)(P_{3G} - P_p + V(\alpha,1) - V(\beta,1)), \\ V(\alpha,1) - V(\beta,1) &< 0. \end{aligned}$$

and

$$\begin{aligned} V(\alpha,1) - V(\beta,1) &= (\alpha-\beta)(P_{3G} - P_p + V(\alpha,1) - V(\beta,1)), \\ (V(\alpha,1) - V(\beta,1))(1-\alpha+\beta) &= (\alpha-\beta)(P_{3G} - P_p), \\ V(\alpha,1) - V(\beta,1) &= \frac{(\alpha-\beta)(P_{3G} - P_p)}{1-\alpha+\beta}, \\ &> 0. \end{aligned}$$

which leads to a contradiction. Therefore, $-P_p + V(\alpha,1) \geq -P_{3G} + V(\beta,1)$. The analysis is similar when $\lambda > \pi(0)$. □

We study the monotonicity of the value function with respect to the belief. Intuitively, with a higher belief, for a given packet delay, the SU obtains better rewards. We prove, in the following proposition, that this intuition is true.

Proposition 4.6. *For a given packet delay l, the value function is monotonically increasing with beliefs λ, i.e. $V(\lambda, l) \geq V(\lambda', l)$ for $\lambda \geq \lambda'$.*

Proof. Let us prove that the value function $V(\lambda, l)$ is increasing with beliefs λ for a given packet delay l. For all $\lambda_1 \leq \lambda_2$, we have:

$$\begin{aligned} V(\lambda_1, l^*) &= -g_u - c_s + \Phi - P_{3G} + V(\beta, 1) + \lambda_1(P_{3G} - P_p + V(\alpha, 1) - V(\beta, 1)), \\ &\leq -g_u - c_s + \Phi - P_{3G} + V(\beta, 1) + \lambda_2(P_{3G} - P_p + V(\alpha, 1) - V(\beta, 1)), \\ &= V(\lambda_2, l^*). \end{aligned}$$

This inequality results from the Lemma 4.5. Let us prove that this propriety holds for all packet delays using backward induction:

- Initial condition: There exists a packet delay l^*, such that $V(\lambda_1, l^*) \leq V(\lambda_2, l^*)$, $\forall \lambda_1 \leq \lambda_2$,
- Suppose that $V(\lambda_1, l+1) \leq V(\lambda_2, l+1)$, $\forall \lambda_1 \leq \lambda_2$,
- we have the following expressions of the action-value functions:

$$\begin{aligned} Q_0(\lambda_1, l) &= -f(l) + V(\Omega^{ns}(\lambda_1|\theta), l+1), \\ &\leq -f(l) + V(\Omega^{ns}(\lambda_2|\theta), l+1), \\ &= Q_0(\lambda_2, l). \end{aligned}$$

The inequality is a direct result from the induction assumption and the Lemma 4.2. Moreover, we have:

$$\begin{aligned} Q_2(\lambda_1, l) &= -c_s + \Phi - P_{3G} + V(\beta, 1) + \lambda_1(P_{3G} - P_p + V(\alpha, 1) - V(\beta, 1)), \\ &\leq -c_s + \Phi - P_{3G} + V(\beta, 1) + \lambda_2(P_{3G} - P_p + V(\alpha, 1) - V(\beta, 1)), \\ &= Q_2(\lambda_2, l). \end{aligned}$$

The inequality comes also from the Lemma 4.5.

First case Assume that $\Phi + f(l) - P_p + V(\alpha, 1) - V(\beta, l+1) \geq 0$. Then, we have:

$$\begin{aligned} Q_1(\lambda_1, l) &= -c_s - f(l) + V(\beta, l+1) + \lambda_1(\Phi + f(l) - P_p + V(\alpha, 1) - V(\beta, l+1)), \\ &\leq -c_s - f(l) + V(\beta, l+1) + \lambda_2(\Phi + f(l) - P_p + V(\alpha, 1) - V(\beta, l+1)), \\ &= Q_1(\lambda_2, l). \end{aligned}$$

Finally, we have that $V(\lambda_1, l) \leq V(\lambda_2, l)$.

Second case Assume that $\Phi + f(l) - P_p + V(\alpha, 1) - V(\beta, l+1) < 0$. Then, for all beliefs λ, we have:

$$\begin{aligned} Q_1(\lambda, l) &= -c_s + \lambda(\phi - P_p + V(\alpha, 1)) + (1-\lambda)(-f(l) + V(\beta, l+1)), \\ &\leq -c_s - f(l) + V(\beta, l+1), \\ &\leq -c_s - f(l) + V(\Omega^{ns}(\lambda|\theta), l+1), \\ &\leq Q_0(\lambda, l). \end{aligned}$$

In fact, we have that $\beta \leq \Omega^{ns}(\lambda|\theta)$ for all beliefs, and the value function $V(\lambda, l)$ is increasing with beliefs for the packet delay $l+1$ (induction assumption). Thus, $g_u + V(\lambda, l) = \max\{Q_0(\lambda, l), Q_2(\lambda, l)\}$. Therefore, we have proved that $V(\lambda_1, l) \leq V(\lambda_2, l)$.

Finally, the value function is increasing with beliefs for all packet delays. □

In the following lemma, we prove that $g_u > -f(l)$.

Lemma 4.7. *The value function's constant g_u is higher than $-f(l)$.*

Proof. we have:

$$\begin{aligned} g_u + V(\alpha, 1) &\geq Q_0(\alpha, 1), \\ g_u + V(\alpha, 1) &\geq -f(l) + V(\Omega^{ns}(\alpha), l+1), \\ g_u + V(\alpha, 1) - V(\Omega^{ns}(\alpha), l+1) &\geq -f(l), \\ g_u &> -f(l). \end{aligned}$$

The inequality comes from the monotonicity of the value function and $\Omega^{ns}(\alpha) < \alpha$. □

Once we have studied the monotonicity of the value function with respect to both of its parameters, we are able to show that the optimal OSA policy has a threshold structure.

4.3.2 The multichannel model

Note that Lemma 4.2 holds for the multichannel model. In fact, if $\vec{\lambda}_1 \leq \vec{\lambda}_2$, then $\lambda_{n_1^*} \leq \lambda_{n_2^*}$ and $\Omega^{ns}(\lambda_{n_1^*}) \leq \Omega^{ns}(\lambda_{n_2^*})$, and therefore, $\Omega^{ns}(\vec{\lambda}_1) \leq \Omega^{ns}(\vec{\lambda}_2)$. Second, consider that $\lambda_{n^*} \leq \pi(0)$. Then, we have $\Omega^{ns}(\lambda_{n^*}) \geq \lambda_{n^*}$, and thus $\Omega^{ns}(\vec{\lambda}) \geq \vec{\lambda}$. Otherwise, we have $\Omega^{ns}(\vec{\lambda}) \leq \vec{\lambda}$.

The Proposition 4.3 can be straightforwardly extended to the multichannel model. Furthermore, we have studied, in Proposition 4.4, the monotonicity of the value function for a fixed belief value with respect to the packet delay. This proposition can be also straightforwardly extended to the multichannel model.

Let us focus on the Proposition 4.6. The monotonicity of the value function with respect to the belief vector depends on the order relationship over the belief set and also on the monotonicity of the belief update functions $\Omega^s(\lambda|\theta = 0)$ and $\Omega^s(\lambda|\theta = 1)$ depending on the belief vector. Note that the monotonicity of the value function with respect to the belief is the main difficulty for extending our study to the multichannel model, and will be considered as a part of our future works.

4.4 Optimal threshold policy

We determine, in this section, an optimal OSA policy for the SU, and we study the structure of such policy. An intuitive behavior of a SU that is accessing opportunistically the spectrum, and that is aware of both energy consumption and transmission delay, is:

- When the packet is recent, i.e. the delay of the packet is small, and the belief is small the SU chooses to wait for better opportunities at next time slots.
- For a delayed packet, the SU chooses to sense and access the dedicated channel if there are no free licensed channels.

We prove in this section, that the intuition is true, and there exists an optimal sensing policy, which has a threshold structure.

Note that SUs have a two-level decision. the first decision for a SU is whether to sense the licensed channels or to wait, depending on its belief, λ, and the current delay of the packet, l. Specifically, we have the following result, which gives us a threshold policy on the belief probability that answers this question.

Proposition 4.8. *For a given packet delay l, the optimal action for the SU is to wait for the next time slot, i.e. $a^*(\lambda, l) = 0$ if and only if $\lambda \leq \lambda^*$, where λ^* is the solution of the equation $\lambda^* = \max(0, \min\{Th1(\lambda^*, l), Th2(\lambda^*, l)\})$. The thresholds $Th1(\lambda^*, l)$ and $Th2(\lambda^*, l)$ are expressed as follows:*

$$Th1(\lambda^*, l) = \frac{V(\Omega^{ns}(\lambda^*|\theta), l+1) - V(\beta, l+1) + c_s}{f(l) + \Phi - P_p + V(\alpha, 1) - V(\beta, l+1)}, \quad \text{and}$$

$$Th2(\lambda^*, l) = \frac{V(\Omega^{ns}(\lambda^*|\theta), l+1) - V(\beta, 1) + c_s - f(l) - \Phi + P_{3G}}{-P_p + V(\alpha, 1) + P_{3G} - V(\beta, 1)}.$$

Proof. In this proposition, we determine explicitly the best action $a^*(\lambda, l)$ for the SU depending on the belief λ and the packet delay l. For a given information state (λ, l), the SU decides to take the action 0 if and only if $Q_0(\lambda, l) \geq \max\{Q_1(\lambda, l), Q_2(\lambda, l)\}$.

- First, we assume that $Q_1(\lambda, l) > Q_2(\lambda, l)$. Let us compare $Q_0(\lambda, l)$ and $Q_1(\lambda, l)$. The inequality $Q_0(\lambda, l) \geq Q_1(\lambda, l)$ is equivalent to:

$$\begin{aligned} -f(l) + V(\Omega^{ns}(\lambda|\theta), l+1) &\geq -c_s + \lambda(\Phi - P_p + V(\alpha, 1)) \\ &\quad +(1-\lambda)(-f(l) + V(\beta, l+1)), \\ V(\Omega^{ns}(\lambda|\theta), l+1) &\geq V(\beta, l+1) - c_s + \lambda(f(l) \\ &\quad +\Phi - P_p + V(\alpha, 1) - V(\beta, l+1)). \end{aligned}$$

As the value function $V(\lambda, l)$ is decreasing with packet delays and increasing with beliefs, we have that $V(\alpha, 1) \geq V(\beta, l+1)$. Moreover, we have already assumed that the immediate reward Φ is higher than the cost P_p. Thus, the expression $f(l) + \Phi - P_p + V(\alpha, 1) - V(\beta, l+1)$ is positive, and we obtain the following equivalence:

$$\begin{aligned} Q_0(\lambda, l) \geq Q_1(\lambda, l) \Leftrightarrow V(\Omega^{ns}(\lambda|\theta), l+1) &\geq V(\beta, l+1) - c_s + \lambda(f(l) + \Phi - P_p \\ &\quad +V(\alpha, 1) - V(\beta, l+1)). \end{aligned}$$

Define the functions F and G as follows:

$$\begin{aligned} F(\lambda, l) &= V(\Omega^{ns}(\lambda|\theta), l+1), \\ G(\lambda, l) &= V(\beta, l+1) - c_s + \lambda(f(l) + \Phi - P_p + V(\alpha, 1) - V(\beta, l+1)). \end{aligned}$$

Note that:

  - If $F(\lambda, l) \geq G(\lambda, l)$, then $Q_0(\lambda, l) \geq Q_1(\lambda, l)$, and the optimal action for the SU is $a(t) = 0$.
  - If $F(\lambda, l) < G(\lambda, l)$, then $Q_0(\lambda, l) < Q_1(\lambda, l)$, and the optimal action for the SU is $a(t) = 1$.

We have proved, in Proposition 4.3, that the value function is PWLC with beliefs. Therefore, for all packet delays, the function $F(\lambda, l)$ is PWLC and increasing with λ, and the function $G(\lambda, l)$ is linear and increasing with λ. Let us study the sign of $F(\lambda, l) - G(\lambda, l)$. Under these setting, six cases rise up:

  1. $F(\lambda, l)$ is always higher than $G(\lambda, l)$, see Figure (4.6, case 1).
  2. $F(\lambda, l)$ is always lower than $G(\lambda, l)$, see Figure (4.6, case 2).

3. $F(\lambda, l)$ and $G(\lambda, l)$ intersect once and $F(\beta, l) < G(\beta, l)$, see Figure (4.6, case3).
4. $F(\lambda, l)$ and $G(\lambda, l)$ intersect once and $F(\beta, l) \geq (\beta, l)$, see Figure (4.6, case 4).
5. $F(\lambda, l)$ and $G(\lambda, l)$ intersect twice and $F(\beta, l) \geq (\beta, l)$, see Figure (4.6, case 5).
6. $G(\lambda, l)$ is tangent to $F(\lambda, l)$, see Figure (4.6, case 6).

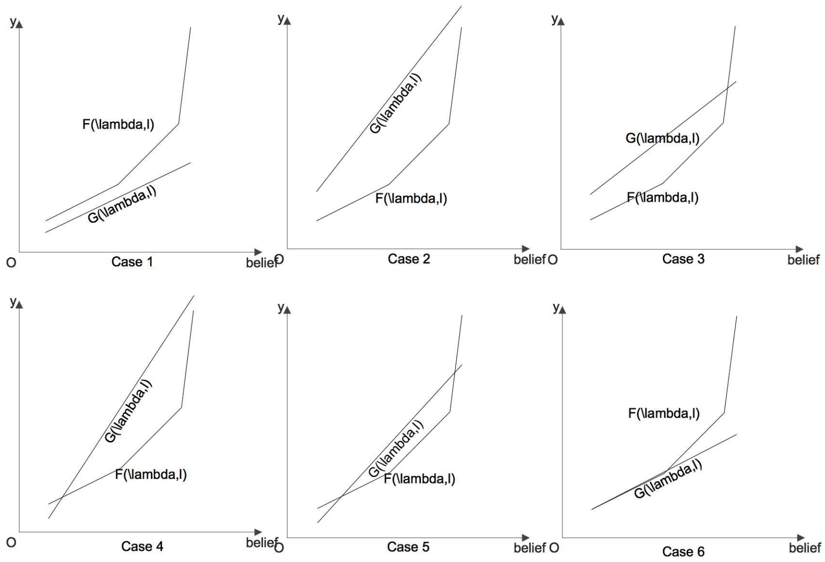

FIGURE 4.6: The analysis of the threshold: the functions $F(\lambda, l)$ and $G(\lambda, l)$.

Let us focus on the study of $F(\pi(0), l)$ and $G(\pi(0), l)$. Suppose that the SU chooses the action 0 for the state $(\pi(0), l)$. Then, we obtain the following inequality:

$$\begin{aligned} g_u + V(\pi(0), l) &= -f(l) + V(\Omega^{ns}(\pi(0)), l + 1), \\ g_u + V(\pi(0), l) &\leq -f(l) + V(\Omega^{ns}(\pi(0)), l), \\ g_u + V(\pi(0), l) &\leq -f(l) + V(\pi(0), l), \\ g_u &\leq -f(l). \end{aligned}$$

This leads to a contradiction as $g_u > -f(l)$ (see Lemma 4.7). It follows that $Q_0(\lambda, l) < Q_1(\lambda, l)$, and $F(\pi(0), l) < G(\pi(0), l)$. Thus, F and G intersect once for belief probability in $[\beta, \alpha]$. Finally, the optimal OSA policy is depicted in the following:

– The SU takes the action 0 for all beliefs lower than the following threshold:

$$Th1(\lambda, l) = \frac{V(\Omega^{ns}(\lambda|\theta), l + 1) - V(\beta, l + 1) + c_s}{f(l) + \Phi - P_p + V(\alpha, 1) - V(\beta, l + 1)},$$

and takes the action 1 otherwise.

- Second, we consider the case where $Q_2(\lambda, l) > Q_1(\lambda, l)$. Then, we have to compare the actions 0 and 2, which is equivalent to comparing the action-value functions $Q_0(\lambda, l)$ and $Q_2(\lambda, l)$. The SU takes the action 0 instead of the action 2 if and only if $Q_0(\lambda, l) \geq Q_2(\lambda, l)$, which is equivalent to:

$$\begin{aligned} -f(l) + V(\Omega^{ns}(\lambda|\theta), l+1) &\geq -c_s + \lambda(\Phi - P_p + V(\alpha, 1)) \\ &\quad +(1-\lambda)(\phi - P_{3G} + V(\beta, 1)), \\ V(\Omega^{ns}(\lambda|\theta), l+1) &\geq V(\beta, 1) + \Phi + f(l) - c_s - P_{3G} \\ &\quad +\lambda(P_{3G} - P_p + V(\alpha, 1) - V(\beta, 1)). \end{aligned}$$

  Note that we have, from Lemma 4.5, that $P_{3G} - P_p + V(\alpha, 1) - V(\beta, 1) \geq 0$. Then, we can provide the same analysis presented in the previous case with the function $F(\lambda, l) = V(\Omega^{ns}(\lambda|\theta), l+1)$ and the function $G(\lambda, l) = V(\beta, 1) + \Phi + f(l) - c_s - P_{3G} + \lambda(P_{3G} - P_p + V(\alpha, 1) - V(\beta, 1))$. The former is PWLC and increasing with λ, and latter is linear increasing with λ. Thus, we obtain the following threshold policy:

  - The SU takes the action 0 for all beliefs lower than the following threshold:

$$Th2(\lambda, l) = \frac{V(\Omega^{ns}(\lambda|\theta), l+1) - V(\beta, 1) - \Phi - f(l) + c_s + P_{3G}}{P_{3G} - P_p + V(\alpha, 1) - V(\beta, 1)},$$

    and takes the action 2 otherwise.

□

This proposition gives us a necessary and sufficient condition on the sensing decision of SUs depending on the belief probability λ. Consequently, if $\lambda > \lambda^*$, then the optimal action for the SU is to sense licensed channels, i.e. $a^*(\lambda, l) \neq 0$.

Furthermore, we have proved, in Lemma 4.2, that the belief vector may be decreasing with the belief update function $\Omega^{ns}(.)$. It follows that there are less opportunities, at the next time slot, to transmit the packet. Thus, the SU should never decide to wait, i.e. action 0, if the belief decreases with $\Omega^{ns}(.)$. The following proposition proves that this intuition is true.

Proposition 4.9. *For all $\lambda > \pi(0)$ and l, the SU never takes the action 0 and thus,* $Q_0(\lambda, l) < \max\{Q_1(\lambda, l), Q_2(\lambda, l)\}$.

Proof. We have from the Lemma 4.2 that if $\lambda > \pi(0)$ then $\Omega^{ns}(\lambda) \leq \lambda$. Suppose that the SU takes the action 0 for a belief λ and packet delay l. Thus we have:

$$\begin{aligned} g_u + V(\lambda, l) &= -f(l) + V(\Omega^{ns}(\lambda), l+1), \\ g_u + V(\lambda, l) &\leq -f(l) + V(\Omega^{ns}(\lambda), l), \\ g_u + V(\lambda, l) &\leq -f(l) + V(\lambda, l), \\ g_u &\leq -f(l). \end{aligned}$$

This leads to a contradiction as $g_u > -f(l)$. The first inequality is because the value function is decreasing with the packet delay, and the second one is because the value function is increasing with the belief and $\Omega^{ns}(\lambda) \leq \lambda$. Thus, if $\lambda > \pi(0)$, then the SU never takes the action 0 and then $Q_0(\lambda, l) < \max\{Q_1(\lambda, l), Q_2(\lambda, l)\}$. □

Remark 4.10. The SU never chooses the action 0 after it transmits a packet over the licensed channel because $\Omega^s(\lambda, \theta = 0) = \alpha > \pi(0)$.

Note that if licensed channels are often occupied, the SU would decide to transmit using the dedicated channel. We depict, in the following proposition, the threshold structure of the optimal decision about the use of the dedicated channel.

Proposition 4.11. *For all belief λ, the SU chooses to use the dedicated channel in spite of waiting for the next time slot if and only if the delay l of the current packet verifies:*

$$-f(l) - \Phi + P_{3G} + V(\beta, l+1) - V(\beta, 1) > 0.$$

Proof. Let us compare the value-action functions $Q_1(\lambda, l)$ and $Q_2(\lambda, l)$ for all belief vector λ and packet delay l. The SU waits for next time slot after sensing if $Q_1(\lambda, l) \geq Q_2(\lambda, l)$, which is equivalent to:

$$\begin{aligned} -c_s + \lambda(\Phi - P_p + V(\alpha, 1)) + (1-\lambda)(-f(l) + V(\beta, l+1)) &\geq -c_s + \lambda(\Phi - P_p + V(\alpha, 1)) \\ &\quad +(1-\lambda)(\phi - P_{3G} + V(\beta, 1)), \\ -f(l) + V(\beta, l+1)\phi - P_{3G} + V(\beta, 1) &\geq 0. \end{aligned}$$

Remark that this condition depends only on the packet delay l. □

Note that this expression depends neither on the cost of sensing C_s nor on the belief vector λ. That is obvious, as this expression determines the best action to do after sensing a channel. We have the last property about the optimal threshold policy.

Corollary 1 (Never Wait After Sensing). For all l, if the penalty cost $-f(l)$ is lower than $\Phi - P_{3G}$, then the SU transmits on the dedicated channel when the licensed channel is sensed as busy.

Proof. If $-f(l)$ is lower than $\Phi - P_{3G}$, then $-f(l) - \Phi + P_{3G} + V(\beta, l+1) - V(\beta, 1)$ is always negative. In fact, $V(\beta, 2) - V(\beta, 1)$ is negative, and $-f(l) - \Phi + P_p + V(\beta, l+1) - V(\beta, 1)$ is decreasing with l. Therefore, the previous expression is negative for all delays $l \geq 1$. □

Remark 4.12. We obtain a two-level threshold structure of the optimal OSA policy, one threshold for each decision step (see Figure 4.4). In fact, the SU has to choose between the two following options: to sleep (action 0), or to sense the licensed channels (action 1). This decision is made using the threshold expressed in Proposition 4.8, based on the belief vector. Thereafter, the SU makes decision about using the dedicated channel or not (action 2), if it decides to sense licensed channels, depending on its packet delay, regardless the belief, based on Proposition 4.11.

Obviously, the optimal OSA policy depends on the transition rates α and β of the PUs' activity. Most of researches in the OSA area assume that some information such that the statistics about the PUs' activity, or the licensed channel transition rates are priory known by the SUs, which may not be realistic in decentralized systems. In practice, an SDR that implement CR uses some learning methods to get insight about the RF environment. We present, in the following section, some learning methods that can be used in order to learn transition rates of the licensed channel on-the-fly.

4.5 Online learning of the RF environment

We have already proved that SUs have an optimal energy-delay constrained policy having a threshold structure, given perfect knowledge of channels transition rates. However, in practice, some information, such as transition rates α and β, are not available for the SU. In this section, we consider a model where the SU does not have external information about the state transition rates. In the following, we present two learning based protocols for SUs in order to estimate the licensed channels dynamics: rate estimator, and transition matrices estimator.

4.5.1 Rate estimator

In this approach, the SU starts with an initial arbitrary values of α and β. Then, it updates them every time slot depending on information about the system state. In fact, the SU computes its sensing policy based on the estimators $\hat{\alpha} = \{\hat{\alpha}_1, ..., \hat{\alpha}_N\}$ and

$\hat{\beta} = \{\hat{\beta}_1, ..., \hat{\beta}_N\}$, where $\hat{\alpha}_i$ (resp. $\hat{\beta}_i$) is the estimator of α_i (resp. β_i). In practice, the SU estimates the following parameters. First, the SU estimates $\hat{\alpha}_i$, which is the probability that the channel i is sensed as idle, given that it was idle in the previous slot. Second, the SU estimates $\hat{\pi}_i(0)$, the stationary probability that the licensed channel is sensed as idle. Finally, the SU obtains the estimated value of β_i based on the following relation:

$$\hat{\beta}_i = (1 - \hat{\alpha}_i)\frac{\hat{\pi}_i(0)}{1 - \hat{\pi}_i(0)}.$$

Formally, the licensed channels' transition rates are estimated based on the following counting processes:

- The vector $\hat{K} = \{\hat{K}_1, ..., \hat{K}_N\}$, where $\hat{K}_i$ represents the number of time slots a channel stays in the idle state, i.e. $\hat{K}_i$ is incremented if the channel i is sensed and is idle at time slots t and $t-1$.
- The vector $\hat{I} = \{\hat{I}_1, ..., \hat{I}_N\}$, where $\hat{I}_i$ represents the number of time slots that the channel is sensed and is idle.
- The vector $\hat{M} = \{\hat{M}_1, ..., \hat{M}_N\}$, where $\hat{M}_i$ represents the number of time slots that the channel is sensed.

Therefore, the SU estimates the state transition rates $\hat{\alpha}$ and $\hat{\pi}_i(0)$ based on the following expressions: $\hat{\alpha}_i = \frac{\hat{K}_i}{\hat{I}_i}$ and $\hat{\pi}_i(0) = \frac{\hat{I}_i}{\hat{M}_i}$.

The convergence of the previous estimators, $\hat{\alpha}$ and $\hat{\beta}$, depends on the occurrence of two successive sensing of the same channel. The SU may not sense frequently the same channel in two successive time slots. Therefore, this estimation method may be inaccurate, and may also harm the SU decision. We propose, in the next section, a more accurate, but also more complex, learning method named transition matrices.

4.5.2 Transition matrices estimator

We present, in this section, a learning protocol that estimates the transition matrices. We define the set of transition matrices $\{P_i(0), P_i(1), ...\}$, where $P_i(j)$ is the transition matrix of the channel i, when this channel was not sensed during j consecutive slots. For example, if the channel i was sensed, j slots before as idle, then the current belief on the state of this channel is $(1, 0) * P_i(j)$.

Similarly to the rate estimator, the transition matrices are estimated using counting processes. Note that the previous learning protocol is somehow a particular case of this

Table 4.1: Simulation parameters

Parameter	Value
P_{3G}	80
P_p	10
c_s	5
Φ	35

Table 4.2: Simulation scenarios

Scenario	Description	α	β
Scenario 1	Licensed channels are often occupied	0.15	0.1
Scenario 2	Licensed channels are often idle	0.85	0.7
Scenario 3	Licensed channels have low transition rates	0.95	0.05

approach. In fact, estimating α and β is equivalent to estimating the set of transition matrices such that the channel was sensed in the previous slot $\{P_1(0), ..., P_N(0)\}$. Indeed, this learning based protocol gives a more accurate estimation of PUs' activity. Specifically, transition matrices estimator method updates the set of transition matrices every time slot in contrast to the rate estimators method which updates the transition rates only if SU senses as idle of the same licensed channel channel for two successive time slots. However, it needs more memory and computational complexity compared to the rates estimators method. Depending on the computational capacity of the SU, it may choose to implement either the rates estimator or the transition matrices estimator method.

4.6 Numeric illustrations

We make extensive numerical experimentations over important number of packets, in order to evaluate the performance of the proposed OSA mechanism, and validate the threshold structure of such policy. We consider 4 i.i.d licensed channels, i.e. $N = 4$ (with 4 licensed channels, we have approximately 10^6 states). Furthermore, the system parameters are summarized in Table 4.1. We consider a model composed of four symmetric channels, and we simulate the system depicted in Figure 4.7. The three different scenarios studied in this chapter are illustrated in Table 4.2.

In this section, we describe the optimal threshold OSA policy, given perfect knowledge about the transition rates of licensed channels. We consider, first, the single channel case and then, we focus on the multichannel model. In the second part of this section, we present some results using estimated values of transition rates, and we compare the performance of the two proposed learning methods.

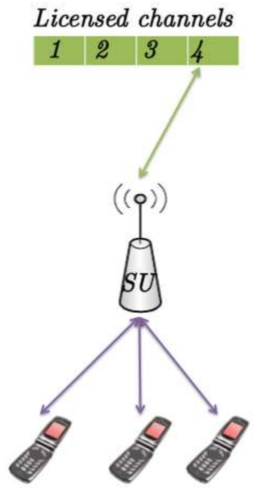

FIGURE 4.7: The simulation model.

4.6.1 Single channel model

In this section, we consider only one licensed channel having transition rates $\alpha = 0.15$ and $\beta = 0.1$. Figure 4.8 illustrates the optimal OSA policy of a SU depending on the belief and the packet delay. For each packet delay, the SU has a threshold policy depending on beliefs. We observe, in Figure 4.8, that the threshold belief probability λ^* is decreasing with packet delays. Furthermore, the maximum packets delay is 13 time slots, regardless the belief vector.

Consider the same scenario with transition rates $\alpha = 0.7$ and $\beta = 0.85$. We observe, in Figure 4.9, that the optimal OSA policy of the SU has also a threshold structure. Furthermore, a packet has at most a delay of 3 time slots regardless its belief. Indeed, the SU always choose the dedicated channel for packets having a delay of 3 slots.

We have proved, in this section, that our numerical results validate our analytic finding. Indeed, the optimal OSA policy has a two-level threshold structure. We study in the next section the OSA mechanism in a multiple licensed channels context.

4.6.2 The multichannel model

In this section, we consider a model composed of 4 licensed channels. Note that when there are multiple licensed channels, SUs have to decide which one they have to sense and access. We implement, in our simulations, a natural greedy channel choice policy. In fact, we consider that the SU chooses the channel that has the highest belief.

We simulate the first scenario depicted in Table 4.2 and we illustrate, in Figure 4.10, the optimal OSA policy for SUs, depending on packet delays. For each packet delay l,

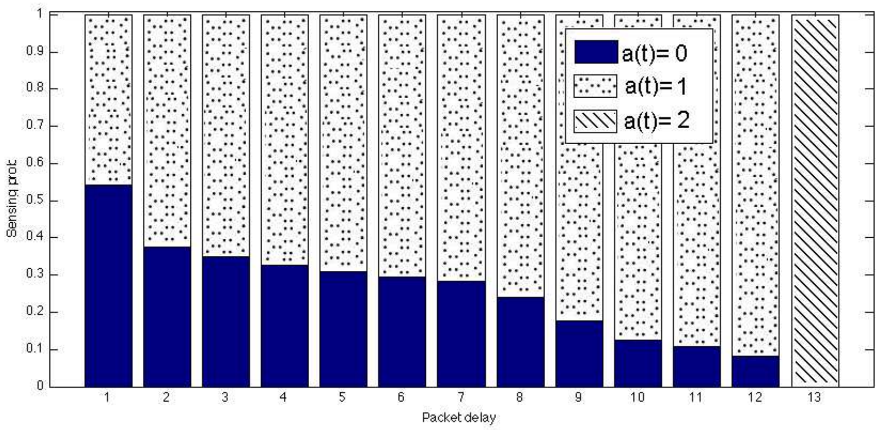

FIGURE 4.8: The optimal OSA policy with one licensed channel, with $\alpha = 0.15$ and $\beta = 0.1$.

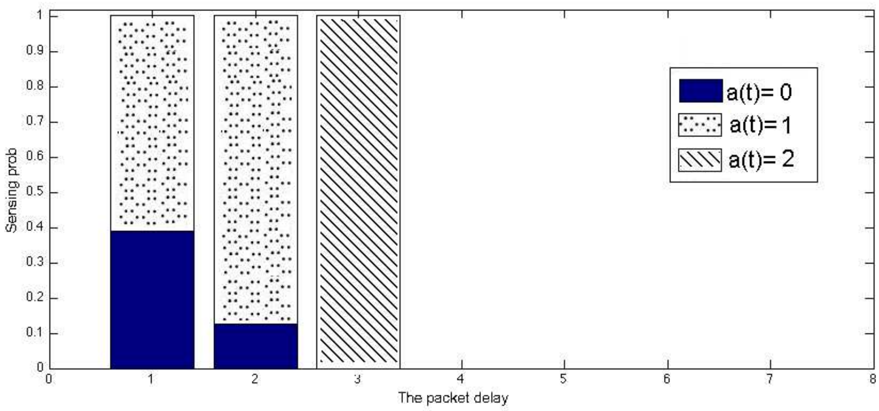

FIGURE 4.9: The optimal OSA policy with one licensed channel, with $\alpha = 0.7$ and $\beta = 0.85$.

the best action for the SU is to wait for the next slot if its belief probability is lower than λ^*. Otherwise, the SU decides to sense the licensed channels. We observe that the maximum packet delay l^* equals 9. Then, when the packet delay is $l = 9$, the SU decides to sense and to transmit using the dedicated channel if the sensed channel is occupied (action 2). This observation validates the result of Proposition 4.11, as the choice of the action 2 depends only on the packet delay, regardless the belief vector.

We illustrate the optimal OSA policy for SUs obtained through simulating the second scenario of Table 4.2, in Figure 4.11. We observe that the SU chooses to transmit over the dedicated channel if there are no opportunities in the licensed spectrum, when the delay of the packet equals 5 slots, regardless its belief. Otherwise, it senses the licensed channels if its belief is higher than the threshold λ^*, and wait if its belief is lower. This result is intuitive as in this scenario, licensed channels are more often idle, inducing a

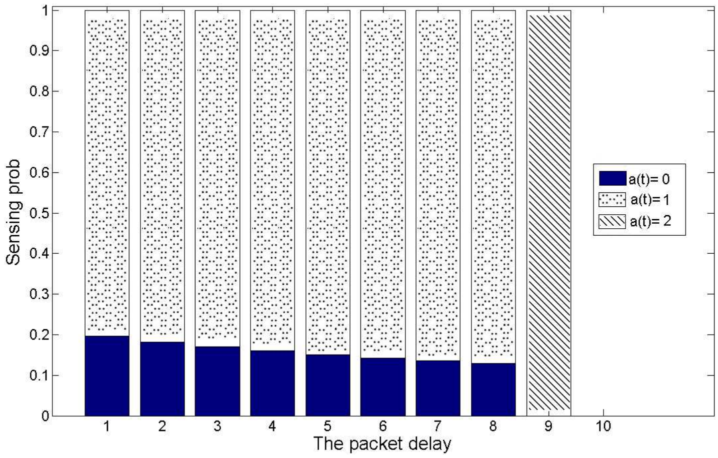

FIGURE 4.10: The optimal OSA policy in the scenario 1.

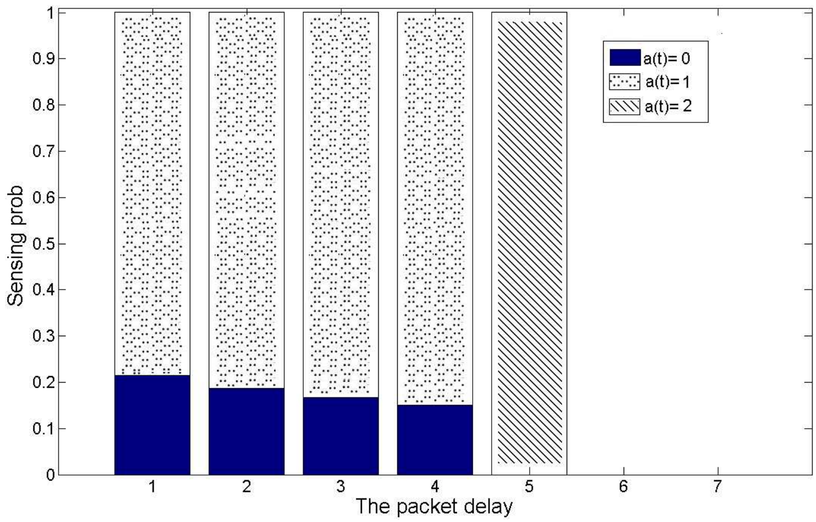

FIGURE 4.11: The optimal OSA policy in the scenario 2.

lower packet delay. Note that in both scenarios, the threshold belief λ^* is decreasing with the packet delays.

Finally, we consider the last scenario depicted in Table 4.2. We observe, in Figure 4.12, that the maximum packet delay equals at most 5 time slots. We further observe that the OSA policy for SUs has also a threshold structure. However, the threshold belief probability λ^* is not monotonous with packet delays. In fact, in this scenario, licensed channels are more static (the probability for each channel to stay occupied or idle is high enough). Thus, it appears one kind of periodic threshold strategy. One more observation, in this scenario, is that the SU changes the choice of the licensed channel to sense if it was sensed as occupied at the last time slot. Indeed, a channel sensed as occupied has a belief of β, the lowest possible belief, and will not be chosen at the next time slot, as we are using a greedy channel choice policy.

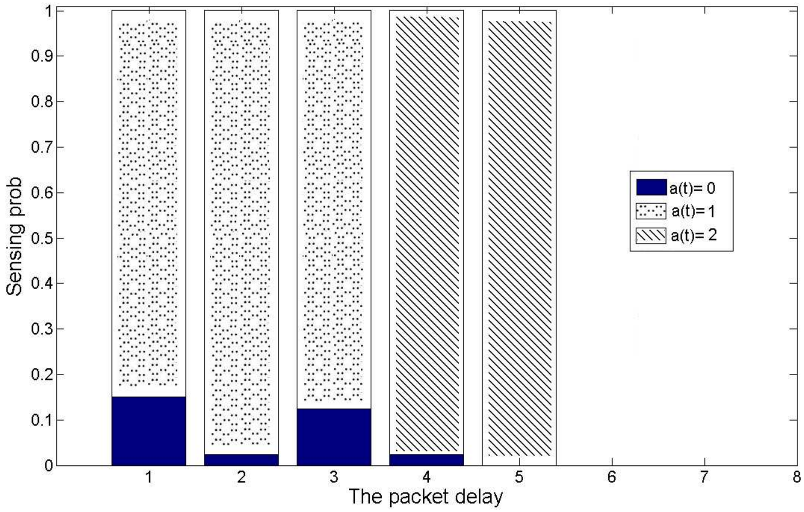

FIGURE 4.12: The optimal OSA policy in the scenario 3.

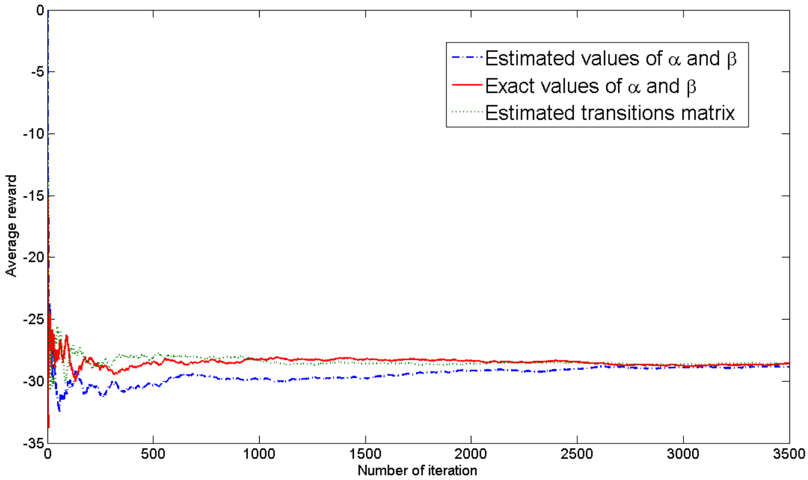

FIGURE 4.13: The average reward for scenario 1.

4.6.3 The multichannel model using estimated values of α and β

We simulate the three scenarios presented in Table 4.2, with estimated values of licensed channels' transition rates. Moreover, we consider both learning approaches presented in Section 4.5. In this section, we evaluate the performance of these learning methods using the two following metrics: The average reward and the average delay. We consider the model with known values of α and β (studied in Section 4.6.2) as a reference model.

Figures 4.15 and 4.16 show that both learning protocols converge. In fact, we observe that both protocols converge before 400 iterations. However, in Figures 4.13 and 4.14, we can observe that the transition matrices estimation method converge 3 times faster (about 1000 iterations) than the rate estimators method (about 3000 iterations).

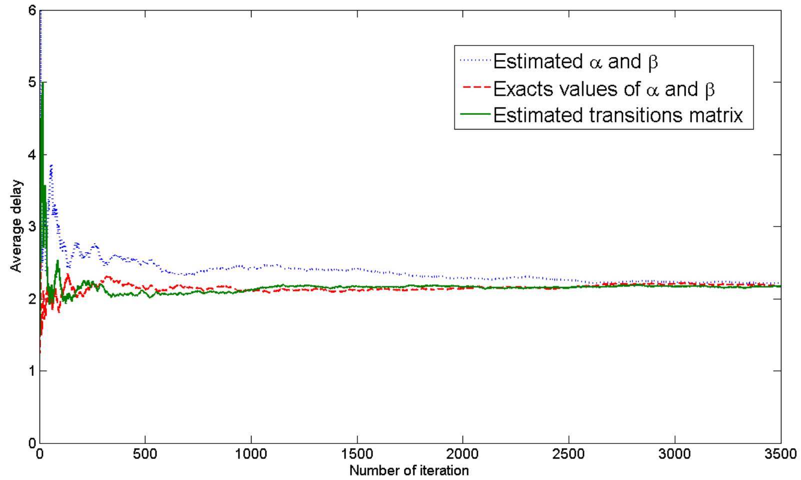

FIGURE 4.14: The average delay for scenario 1.

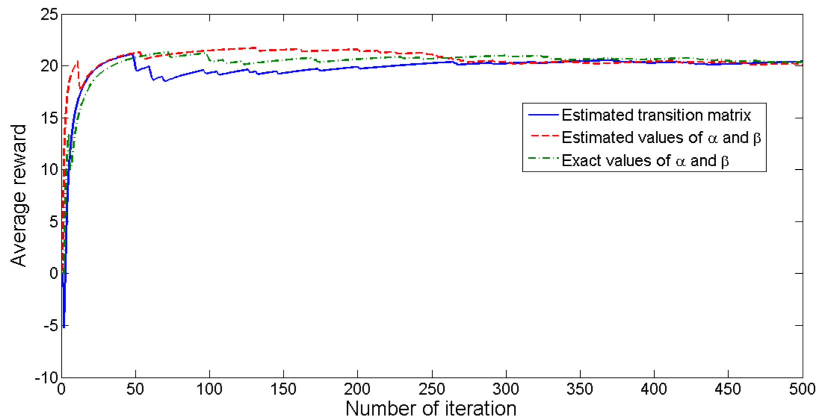

FIGURE 4.15: The average reward for scenario 2.

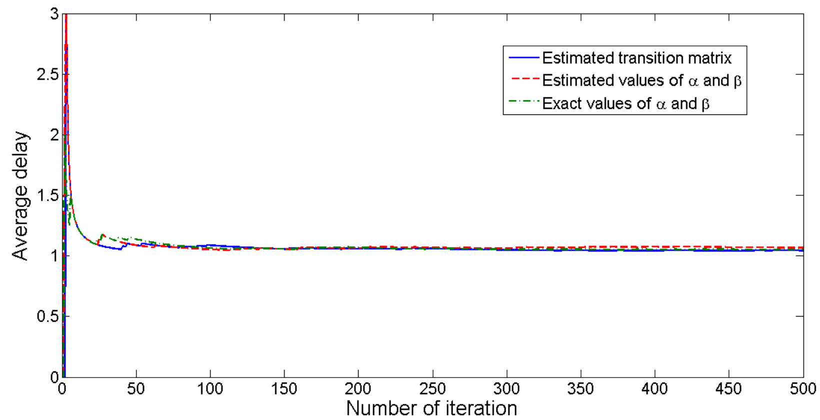

FIGURE 4.16: The average delay for scenario 2.

4.7 Conclusion

In this chapter, we have used a POMDP framework for designing an optimal OSA policy for CR networks taking into account an energy-delay tradeoff for SUs. Introducing a QoS metric in the OSA policy is very important, with the emergence of heterogeneous mobiles that are able to transmit their QoS-dependent traffic over different mediums of communication like 3G, WiFi and TV White Space. We have provided some structural properties of the value function and we have proved the existence of an optimal stationary OSA policy that has two-level threshold structure. We have been able to determine explicitly the threshold structure of the optimal policy.

Note that the interaction between several SUs has not been considered here, and in the literature very few, at the best of our knowledge. This perspective is also very important because if the channel choice policy is the same for all the SUs, there could have lots of collisions between several SUs that have sensed the same licensed channel. In the following chapter, we extend this study to the multichannel context. Indeed, we consider that SUs make decision individually and try to maximize their own benefits.