Throughput Optimal Opportunistic Channel Switching in Cognitive Radio Networks

M. Santhoshkumar, Dony J. Muttath, and K. Premkumar, *Member, IEEE*

***Abstract*—In this letter, we consider a problem of throughput optimum channel switching for channel access by a secondary user in multichannel Cognitive Radio Networks in which there is a delay involved in switching from one channel to another. We consider an imperfect sensing model that takes into account false alarm and detection probabilities. We show that the problem can be posed in the framework of Partially Observable Markov Decision Process (POMDP), and obtain an average throughput optimal policy for which we provide structural results. Also, we propose a low complex algorithm ASTUTE for channel access based on one stage reward. We compare the performance of the optimum policy with ASTUTE. Our results show that the average throughput performance of ASTUTE is very close to that of the optimal solution.**

***Index Terms*—Average cost MDP, cognitive radio networks, POMDP, reconfiguration delay, switching cost, wideband sensing.**

I. Introduction

The demand for spectrum has led to the study of exploration and exploitation of underutilized spectrum by cognitive radios. Thus, the secondary users (SUs) opportunistically sense and access primary users' (PUs) channels, when available. Only recently, there is an interest in dynamic spectrum access with switching cost [1], [2]. When the SU switches from one channel to another, it has to switch off the current RF chain and switch on a different RF chain [1] resulting in a reconfiguration delay, and thus, reducing the available throughput to the SU. In this letter, we study a channel switching problem that maximizes throughput of SU considering switching across channels.

A. Previous Work

A number of works have been reported in the literature on channel sensing and access. The problem of sequential channel sensing in a multichannel network has been studied [3], [4], and wideband spectrum sensing has been studied in [5]. The problem of maximizing the spectrum utilization with handoff has been studied [6]. However, most of the works (e.g., [3], [4], [6]) do not address the switching cost across channels.

To the best of our knowledge, [7] is the first work on spectrum access with switching cost. A block channel access (BCA) policy is designed for a block of time-slots in which each SU chooses a channel for sensing/access at the beginning of each block, and if found idle, uses the same channel for the whole block; if there is any collision among SUs, all the colliding SUs switch randomly to another channel. The block-length is designed such that the BCA achieves a logarithmic regret. But, this policy depends on perfect sensing.

In [2], the authors study the switching energy consumption in industrial cognitive radio networks (CRN), and found that wider the gap in frequencies, larger is the switching energy consumed. Thus, a joint design of frequency assignment and scheduling is needed, which is shown to be NP-Hard.

In [8], Agarwal and De study the energy efficiency of single channel and multi channel dynamic spectrum access schemes. It is shown that in the single channel scheme, the SUs wait on the current channels for spectrum access, whereas in multi channel scheme, SUs immediately switch to other channels. Hence, for delay tolerant data, a single channel scheme can be more energy-efficient than multi channel scheme. This work led to the study of channel switching having N channels and an SU with a finite volume of traffic in [1], [5], [9].

In [5], Wang *et al.* consider homogeneous PUs, and a wideband spectrum sensing by the SU with which it knows the list of all idle channels. In case of the current channel being busy, the SU can wait on the current channel with probability P_s, or switches on to an idle channel with probability $1 - P_s$, irrespective of the current channel.

In [9], Wu *et al.* consider a CRN where the SUs predict the future state of the channel, and hence, can proactively switch to minimize the energy consumption. Though, this scheme is inherently energy-efficient, it employs the same switching strategy irrespective of the channel that it currently stays on.

In [1], Ding *et al.* consider a finite schedule length over which all packets of SU need to be delivered. The authors design an energy-efficient switching scheme called current channel aware (CCA) policy based on Q-learning, which does not require traffic statistics of the PUs. In CCA, the optimum action achieves the minimum expected total energy consumption, and exploits the heterogeneous statistics of channels. They show the tradeoff between delay tolerance of SU's data and the channel switching cost.

All the previous work so far consider energy efficient channel switching, and not on maximizing the throughput of the SU. To the best of our knowledge, the problem of throughput optimal switching has not been studied in the literature. This problem is of interest, as switching causes a reconfiguration delay, which causes a dip in throughput, which we study in this letter.

B. Contributions of the Paper

In this letter, we report the following contributions.

Manuscript received May 17, 2021; accepted June 16, 2021. Date of publication June 22, 2021; date of current version September 9, 2021. The associate editor coordinating the review of this article and approving it for publication was R. Wang. *(Corresponding author: K. Premkumar.)*

The authors are with the Communications Research Lab, Department of Electronics and Communication Engineering, Indian Institute of Information Technology, Design and Manufacturing, Kancheepuram, Chennai 600 127, India (e-mail: kpk@iiitdm.ac.in).

Digital Object Identifier 10.1109/LWC.2021.3091635

1) We propose a simple sensing model that abstracts any sensing procedure (e.g., energy detector, likelihood based detector, etc). Only the probabilities of false alarm and detection are used for modeling imperfect sensing.
2) We formulate the channel switching problem with reconfiguration delay in the framework of POMDP, and obtain an optimal policy that maximizes the average throughput, and provide a structure of the optimum policy.
3) We propose a low-complexity algorithm ASTUTE, and show that its average throughput performance is very close to that of the optimal policy.

C. Organization of the Paper

In Section II, we describe the PUs' traffic and the sensing model in CRN. In Section III, we formulate the optimal channel switching problem, and show that the problem is modeled as an average reward optimal problem in the framework of Partially Observable Markov Decision Process (POMDP). In Section IV, we obtain the solution to the problem, and also provide a computationally simple solution. We provide numerical results in Section V. Finally, we conclude in Section VI.

II. SYSTEM MODEL

In this Section, we provide the details of the CRN and the sensing model.

A. Network Model

We consider a CRN comprising N channels indexed by $i \in \mathcal{C} = \{1, 2, 3, \ldots, N\}$ with each channel i being used by primary user (PU). The PU who uses channel i is also denoted by i. We consider one secondary user (SU) who can use any channel $i \in \mathcal{C}$, when PU i is not using the channel. The SU is perpetual and always has elastic data to send. Hence, whenever there is an opportunity for the SU to transmit, it can explore and exploit that.

We consider a discrete-time system in which time is slotted with the length of each time-slot being T in which the initial τ_s is used by the SU for wideband sensing of all N channels, and can potentially use the remaining time $T - \tau_s$ for its data transmission [5]. A time-slot is denoted by $k \in \mathbb{Z}_+ = \{0, 1, 2, \ldots, \}$. During time-slot k, a PU i either has a packet to transmit which is called *busy* state (denoted by 0), or remains in *idle* state (denoted by 1). Let $\Theta_i[k] \in \{0, 1\}$ denote the *busy/idle* state of PU i, during time-slot k. The PU state $\{\Theta_i[k], k \in \mathbb{Z}_+\}$ that we consider in this work follows a discrete-time Markov chain with the transition probability matrix of ith PU being

$$\boldsymbol{P}_i = \begin{bmatrix} p_{i,00}, & p_{i,01} \\ p_{i,10}, & p_{i,11} \end{bmatrix}. \tag{1}$$

We assume that during any time-slot k, the *busy/idle* states of PUs, $\Theta_1[k], \Theta_2[k], \ldots, \Theta_N[k]$ are independent of each other. Let $\boldsymbol{\Theta}[k] \triangleq [\Theta_1[k], \Theta_2[k], \ldots, \Theta_N[k]] \in \{0, 1\}^N$ denote the vector of traffic states of all PUs during time-slot k.

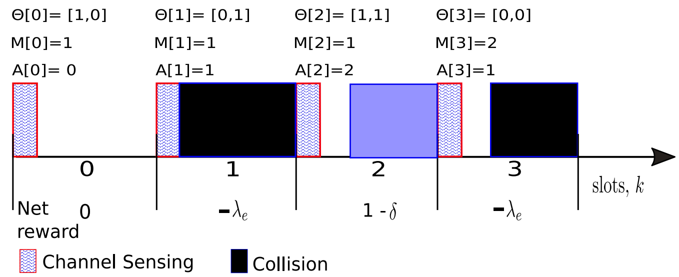


Fig. 1. An illustration of channel switching. At $k = 0$, SU chooses to continue in channel 1 with out transmission ($A[0] = 0$) incurring no throughput. At $k = 1$, SU chooses to transmit in channel 1 ($A[1] = 1$) and collides with PU 1's transmission (as $\Theta_1[1] = 0$) losing a throughput of $-\lambda_e$ as penalty. At $k = 2$, SU switches to channel 2, getting a throughput of $1 - \delta$ due to reconfiguration delay. At $k = 3$, the SU's transmission collides with that of PU, incurring a penalty of λ_e (whether the SU uses a full slot or $1 - \delta$, the penalty is the same). If at slot 0, the SU had chosen $A[0] = 1$, it would have got a throughput of 1. Note that an initial τ_s portion of a time-slot is used for sensing.

B. Sensing Model

During a time-slot k, the SU senses all channels using a wideband spectrum sensing method, and obtains noisy observations of channel state from each channel i. Let $O_i[k] \in \{0, 1\}$ be the state of channel i as sensed/observed by the SU. The sensing model is given by

$$\mathbb{P}\{O_i[k] = 1 \mid \Theta_i[k] = \theta\} = \begin{cases} \alpha, & \text{if } \theta = 0, \\ \beta, & \text{if } \theta = 1. \end{cases} \tag{2}$$
$$= 1 - \mathbb{P}\{O_i[k] = 0 \mid \Theta_i[k] = \theta\}.$$

Thus, the actual state $\Theta_m[k]$ can not be observed directly, and the state can only be observed through $O_m[k]$. At the beginning of time-slot k, the vector of observations that the SU observes is given by $\boldsymbol{O}[k] \triangleq [O_1[k], O_2[k], \ldots, O_N[k]]$. From the current and past observations, the SU makes a decision of accessing a channel a during time-slot k (see Fig. 1).

III. PROBLEM FORMULATION

In this Section, we formulate an average throughput optimal problem with sensing and switching costs for optimum channel sensing and access.

At the beginning of time-slot k, let the RF chain of the SU is tuned to the channel $M[k] \in \mathcal{C}$. We define the state of the system by $\boldsymbol{S}[k] \triangleq [\boldsymbol{\Theta}[k], M[k]]$. At time-slot k, the SU makes a wide-band sensing of all channels (see [5]), and based on the system state $\boldsymbol{S}[k]$, it makes the following decision: whether to switch and access channel $A[k]$ during time-slot k ($A[k]$ may be the same as $M[k]$), or remain silent (and stay in channel $M[k]$ during time-slot k) [1], i.e., the action (or decision) that the SU takes is $A[k] \in \mathcal{A} = \mathcal{C} \cup \{0\}$. If the SU accesses a channel, then $A[k] \in \mathcal{C}$. For the case, the SU refrains from transmission during time-slot k, $A[k] = 0$; in this case, the SU stays in the same channel $M[k]$ during time-slot k.

If $A[k] \in \mathcal{C}$ and is different from $M[k]$, then the SU would have to wait for a delay of $\Delta < T - \tau_s$ for re-configuring its RF chain [5]. Let $0 \leq \delta := \Delta/(T - \tau_s) < 1$ be the fraction of time that is spent for reconfiguration. The throughput available to SU is normalized as 1 if there is no switching and is $1 - \delta$ if there is switching. If $A[k]$ is the same as $M[k]$ or $A[k] = 0$, then there is no switching, and hence, no loss.

In this work, we assume that the reconfiguration delay is the same for switching from any channel to any other channel. However, this assumption can be relaxed.

We define the reward of making a decision on choosing an action a for access in the current time-slot (when the state of the system is $[\boldsymbol{\theta}, m]$) by

$$G_a([\boldsymbol{\theta}, m], a) = \begin{cases} 0, & \text{if } a = 0, \\ 1, & \text{if } a \in \mathcal{C}, \theta_a = 1, \\ -\lambda_e, & \text{if } a \in \mathcal{C}, \theta_a = 0, \end{cases} \quad (3)$$

where $\lambda_e > 0$ is the penalty imposed on the SU when it causes interference to PUs' transmission. It is inferred that when the SU interferes with the PU, it is made to lose a throughput of λ_e. We define the cost of making a decision on switching to channel a for access in the current time-slot by

$$L_s([\boldsymbol{\theta}, m], a) = \begin{cases} \delta, & \text{if } a \in \mathcal{C}, a \neq m, \text{ and } \theta_a = 1, \\ 0, & \text{otherwise}, \end{cases} \quad (4)$$

as the SU loses a fraction of throughput δ due to switching from channel m to channel $a \neq m$. From (3) and (4), the net gain of action a in the current time-slot is given by

$$G([\boldsymbol{\theta}, m], a) = G_a([\boldsymbol{\theta}, m], a) - L_s([\boldsymbol{\theta}, m], a). \quad (5)$$

Note that the action chosen at time-slot k is $A[k] = a$ (see Fig. 1). Since $\boldsymbol{\theta}[k]$ is observable only partially through $\boldsymbol{O}[k]$, the *information vector* available to the SU at time-slot k is

$$\boldsymbol{I}[k] = [\boldsymbol{O}[0:k], A[0:k-1]], \quad (6)$$

where the notation $x[m:n]$ means $[x[m], x[m+1], \ldots, x[n]]$, and $\boldsymbol{I}[0] = [\boldsymbol{O}[0]]$. A policy $\boldsymbol{\mu}$ is a sequence of functions $\boldsymbol{\mu} = (\mu_0, \mu_1, \ldots,)$, where μ_k maps the information vector $\boldsymbol{I}[k]$ into the action space $\mathcal{A}$ (see Chap. 5, pp. 218–219 [10]). The expected one-stage reward during time-slot k is

$$\begin{aligned} \mathbb{E}[G(\boldsymbol{S}[k], A[k])] &= \mathbb{E}\mathbb{E}[G(\boldsymbol{S}[k], A[k]) \mid \boldsymbol{I}[k]] \\ &= \mathbb{E}[g([\boldsymbol{\Pi}[k], M[k]], A[k])], \end{aligned} \quad (7)$$

where $\boldsymbol{\Pi}[k] = [\Pi_1[k], \Pi_2[k], \ldots, \Pi_N[k]]$ is the a posteriori probability vector, $\Pi_i[k] \triangleq \mathbb{P}\{\Theta_i[k] = 1 \mid \boldsymbol{I}[k]\}$, and $g([\boldsymbol{\Pi}[k], M[k]], A[k]) \triangleq \mathbb{E}[G(\boldsymbol{S}[k], A[k]) \mid \boldsymbol{I}[k]]$. Also, it is clear that $\boldsymbol{\Pi}[k]$ is a sufficient statistic, which can be obtained recursively from $\boldsymbol{\Pi}[k-1]$ and the observation $\boldsymbol{O}[k]$ (see Chap. 5, pp. 251–252 [10]). Hence, given $\boldsymbol{\Pi}[k-1]$, $\boldsymbol{\Pi}[k]$ is conditionally independent of $\boldsymbol{\Pi}[k-2], \boldsymbol{\Pi}[k-3], \ldots, \boldsymbol{\Pi}[0]$.

Define $\boldsymbol{X}[k] \triangleq [\boldsymbol{\Pi}[k], M[k]]$. Note that the process $\{\boldsymbol{X}[k], k \in \mathbb{Z}_+\}$ is a controlled Markov process, with the control at time-slot k being $A[k]$. Thus, for the MDP that we propose, $\boldsymbol{X}[k]$ is the state of the system, and $A[k]$ is the action. The state space of this MDP is $\mathcal{X} \triangleq [0,1]^N \times \mathcal{C}$, and the action space is $\mathcal{A} \triangleq \{0\} \cup \mathcal{C}$. Define the expected gain function $g : \mathcal{X} \times \mathcal{A} \rightarrow \mathbb{R}$ as

$$g([\boldsymbol{\pi}, m], a) = \begin{cases} 0, & \text{if } a = 0, \\ \pi_a - \lambda_e(1-\pi_a), & \text{if } a \in \mathcal{C}, a = m, \\ (1-\delta)\pi_a - \lambda_e(1-\pi_a), & \text{if } a \in \mathcal{C}, a \neq m. \end{cases} \quad (8)$$

Note that the reward function in (8) captures the throughput that the SU obtains. A channel sensing and channel access policy is a sequence of functions $\boldsymbol{\mu} = (\mu_0, \mu_1, \mu_2, \ldots,)$, where each function $\mu_k : \mathcal{X} \rightarrow \mathcal{A}$. The average throughput of policy $\boldsymbol{\mu}$ with initial state $\boldsymbol{x}_0$ is

$$J_{\boldsymbol{\mu}}(\boldsymbol{x}_0) = \lim_{K \rightarrow \infty} \frac{1}{K} \mathbb{E}\left[\sum_{k=0}^{K-1} g(\boldsymbol{X}[k], \mu_k(\boldsymbol{X}[k]) \big| \boldsymbol{X}[0] = \boldsymbol{x}_0\right]. \quad (9)$$

A stationary policy is one for which $\mu_k = \mu$ for all k. Note that for a policy $\boldsymbol{\mu}$, $\mu(\boldsymbol{X}[k])$ is the action taken during time-slot k. We are interested in obtaining a stationary policy $\boldsymbol{\mu}^* = \{\mu^*, \mu^*, \ldots,\}$ that maximizes the average throughput.

IV. Optimum Average Throughput Policy

The problem defined in Section III is to obtain an optimum sequence of controls $A[k] = \mu^*(\boldsymbol{X}[k])$ (where $A[k]$ is the *switch-and-access* or *remain-silent* decision taken during each time-slot k) such that the average throughput is maximum, i.e.,

$$\boldsymbol{\mu}^*(\boldsymbol{x}_0) = \arg\max_{\boldsymbol{\mu}} J_{\boldsymbol{\mu}}(\boldsymbol{x}_0)$$

The solution to this is given by the following Bellman's equation (see [11, Th. 6.5.2])

$$g^* + V(\boldsymbol{x}) = \max_{a \in \mathcal{A}} \left[g(\boldsymbol{x}, a) + \mathbb{E}\left[V(\boldsymbol{x}') \big| \boldsymbol{x}, a\right]\right], \quad (10)$$

$$\mu^*(\boldsymbol{x}) = \arg\max_{a \in \mathcal{A}} \left[g(\boldsymbol{x}, a) + \mathbb{E}\left[V(\boldsymbol{x}') \big| \boldsymbol{x}, a\right]\right], \quad (11)$$

where $\boldsymbol{x}'$ is the next state when the current (state, action) pair is $(\boldsymbol{x}, a)$ (note: $\boldsymbol{x} = (\boldsymbol{\pi}, m)$) and $g^* = J_{\boldsymbol{\mu}^*}(\boldsymbol{x}_0)$ is the optimum average throughput which is independent of $\boldsymbol{x}_0$.

A. Solution by Value Iteration

Define the sequence of value functions $v_0(\boldsymbol{x}), v_1(\boldsymbol{x}), \ldots,$

$$\begin{aligned} v_0(\boldsymbol{x}) &:= 0 \\ v_{k+1}(\boldsymbol{x}) &:= \max_{a \in \mathcal{A}} \left[g(\boldsymbol{x}, a) + \mathbb{E}\left[v_k(\boldsymbol{x}') \big| \boldsymbol{x}, a\right]\right], k = 0, 1, 2, \ldots \end{aligned} \quad (12)$$

It has been shown (see [10, Sec. 4.3]) that the maximizer that maximizes (12) in the value iteration as $k \rightarrow \infty$ yields the optimum policy $\boldsymbol{\mu}^*(\boldsymbol{x})$. We provide the structure of the optimum policy in the Theorem below.

Theorem 1: If $A^*(\boldsymbol{\pi}, m)$, the optimum action for state $[\boldsymbol{\pi}, m]$, is $i \in \mathcal{C}$, then for all $0 \leq r \leq 1 - \pi_i$, $A^*(\boldsymbol{\pi} + r\boldsymbol{e}_i, m) = i$.

Proof: Please see the Technical Report available at the url http://web.iiitdm.ac.in/prem/2021-tr01.pdf.

B. ASTUTE

The value iteration shown in Section IV-A provides a numerical way of computing the maximum average throughput policy. But, it is computationally expensive. Hence, we propose an **A**lgorithm for **S**witching **T**o achieve **U**seful **T**hroughput **E**fficiency (ASTUTE) based on Theorem 1,

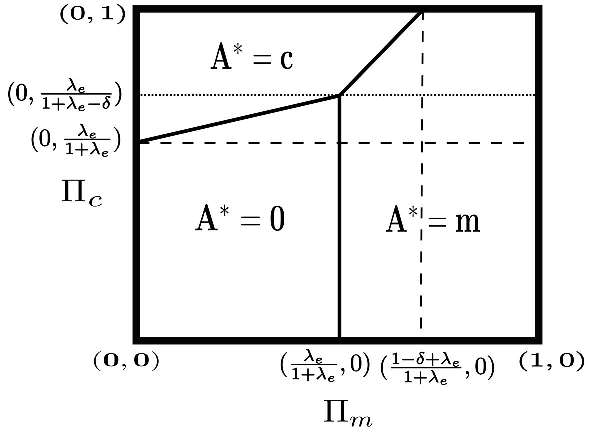


Fig. 2. The optimum decision regions for the actions $A^* \in \{0, m, c\}$ are illustrated here. The regions are obtained from Theorem 1, and from (8).

Algorithm 1: ASTUTE for *N* Channels

1 function sensing and access $([\boldsymbol{\pi}, m], a, \boldsymbol{o}')$;

Input :
1) $[\boldsymbol{\pi}, m]$: state of the system during time-slot $k-1$
2) a: action chosen during time-slot $k-1$
3) $\boldsymbol{o}'$: sensing observation made by SU at time-slot k

Output:
1) $[\boldsymbol{\pi}', m']$: state of the system during time-slot k
2) a': action chosen during time-slot k

2 For all channels $i \in \mathcal{C}$, compute the state π_i' as follows:

$$\hat{\pi}_i = \pi_i \cdot p_{i,11} + (1 - \pi_i) \cdot p_{i,01}, \text{and}$$
$$\pi_i' = \begin{cases} \frac{\beta\hat{\pi}_i}{\beta\hat{\pi}_i + \alpha(1-\hat{\pi}_i)}, & \text{if } o_i' = 1, \\ \frac{(1-\beta)\hat{\pi}_i}{(1-\beta)\hat{\pi}_i + (1-\alpha)(1-\hat{\pi}_i)}, & \text{if } o_i' = 0. \end{cases} \quad (13)$$

3 Find channel *c*, other than *m*, for which π_c' is the largest, i.e.,

$$c = \arg\max\{\pi_i' : i = 1, 2, \cdots, N, i \neq m\}.$$

It is to be noted that for any $i \neq m$, $\pi_c' \geqslant \pi_i'$, and hence, among channels $i \neq m$, switching to channel *c* is better than switching to any other channel $i \neq c$. Thus, the optimum action problem reduces to one of choosing between no-access (denoted by 0), continue to use channel *m*, and switch to channel *c*, i.e., the optimum action $A^* \in \{0, m, c\}$.

4 The decision regions for each of the actions $A^* = 0$, *m*, *c* are shown in Fig. 2. For a given state $\boldsymbol{\pi}'$, choose the action which is labeled for the point $\boldsymbol{\pi}'$.

which is computationally less intensive, and is presented in Algorithm 1.

In Algorithm 1, we note that Eqn. (13) gives the state of the system at time-slot *k*, i.e., $[\boldsymbol{\Pi}[k], M[k]] = [\boldsymbol{\pi}', m']$ (note: $m' = m$, if $a = 0$, and $m' = a$, if $a \neq 0$), based on which the SU is required to make a channel access decision A^* ($A^* \in \{0, m, c\}$). The optimum decision is given by the value iteration algorithm which is prohibitively complex, and a sub-optimal decision is given by the ASTUTE (see Fig. 2) which partitions the square region $[0, 1] \times [0, 1]$ into decision regions A_0^*, A_m^*, and A_c^*. If $\boldsymbol{\pi}' \in A_a^*$, then optimum access is $A^* = a$.

We now compare the computational complexity of value iteration and ASTUTE. For a finite state and finite action space MDP, let *S* be the size of the state space and *A* be the size of the action space. In each iteration step of value iteration, for *N* channels, the number of computations required, in general, is $4NAS^2$ multiplications and $4NAS^2$ additions. Thus, the number of computations in each iteration is $O(NAS^2)$. If we run the value iteration for a finite number of *m* iterations, then the computational complexity is $O(mNAS^2)$, which is the computational complexity of value iteration.

We now compute the complexity of ASTUTE. At time-slot *k*, computing the current state from the previous state and the current observation requires 4*N* additions and 4*N* multiplications, where *N* is the number of channels (see Eqn. (13)). Thus, updating the state requires *O*(*N*) computations. Thus, for a state space of size *S*, ASTUTE requires only *O*(*NS*) computations, which is much less complex than the value iteration algorithm.

V. Results and Discussion

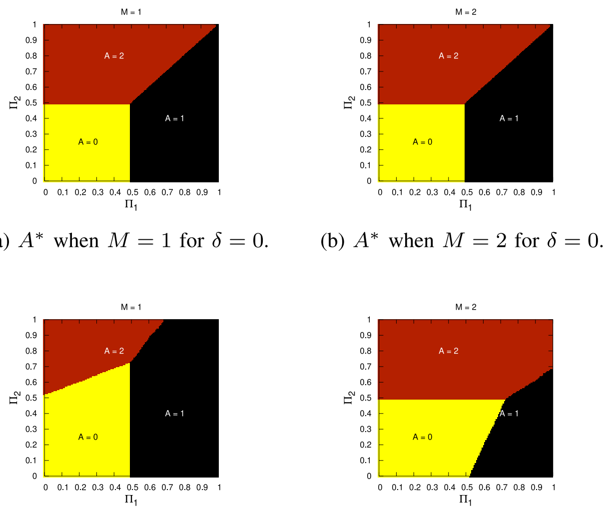
In this Section, we numerically evaluate the performance of optimum channel access with switching cost. Also, we compare the optimum policy with ASTUTE that we have described in Algorithm 1.

We consider a CRN with two channels (i.e., $N = 2$). The busy/idle state transition probabilities of the PUs are given by $p_{1,01} = p_{2,01} = 0.1$, and $p_{1,10} = p_{2,10} = 0.1$. The α and β of the observation model are taken as 0.1 and 0.8, respectively. The throughput penalty due to collision, λ_e is varied from 1 to 3. The normalized reconfiguration delay due to switching, δ is varied from 0.0 to 1.0. We obtain the throughput optimal policy by value iteration (VI) described in Section IV-A.

As we have discussed in Section III, switching of channels for access involves reconfiguration of RF chain from one channel to another, which happens with a normalized switching delay of $\delta = \Delta/(T - \tau_s)$, where $0 \leq \delta < 1$. This normalized delay appears in the dynamic optimization through single-slot gain function (see Eqns. (4), (5), (7) and (8)). Thus, the latency is an important parameter that makes the throughput drop if the number of switching is large, and it is important to study the throughput as a function of the normalized latency parameter δ. This is the motivation to study the throughput as a function of switching latency in simulation experiments.

The optimal policy is plotted in Fig. 3 for various values of δ. For $\delta = 0$ (no reconfiguration delay), it is clear that the access policy is the same whether the current RF chain is in channel $M = 1$, or $M = 2$. The optimum policy is to choose channel 1 when $\Pi_1 > \max\{\frac{\lambda_e}{1+\lambda_e}, \Pi_2\}$; the optimum policy is to choose channel 2 when $\Pi_2 > \max\{\frac{\lambda_e}{1+\lambda_e}, \Pi_1\}$.

For $\delta > 0$, it is intuitive to see that the regions of $A = 0$ and $A = 1$ extends into the region of the channel 2, for the case of $M = 1$ (and same is true for the case of $M = 2$). The polygons that represent the optimal decision regions can be described completely by means of hyperplanes, and in this example, it can be inferred that the straight lines $\Pi_2 = m_1\Pi_1 + c_1$ in the region $\Pi_1 \in [0, 0.5]$, and $\Pi_2 = m_2\Pi_1 + c_2$ in the region



(a) A^* when $M = 1$ for $\delta = 0$. (b) A^* when $M = 2$ for $\delta = 0$.

(c) A^* when $M = 1$ for $\delta = 0.9$. (d) A^* when $M = 2$ for $\delta = 0.9$.

Fig. 3. (a) and (c) show the optimum action A^* when the SU is tuned to channel $M = 1$, and (b) and (d) show the optimum action A^* when it is tuned to $M = 2$. The optimum policy is computed by value iteration using the following parameters: $N = 2$, $p_{1,01} = p_{2,01} = 0.1$, $p_{1,10} = p_{2,10} = 0.1$, $\alpha = 0.1$, $\beta = 0.8$, and $\lambda_e = 1$. (a) A^* when $M = 1$ for $\delta = 0$. (b) A^* when $M = 2$ for $\delta = 0$. (c) A^* when $M = 1$ for $\delta = 0.9$. (d) A^* when $M = 2$ for $\delta = 0.9$.

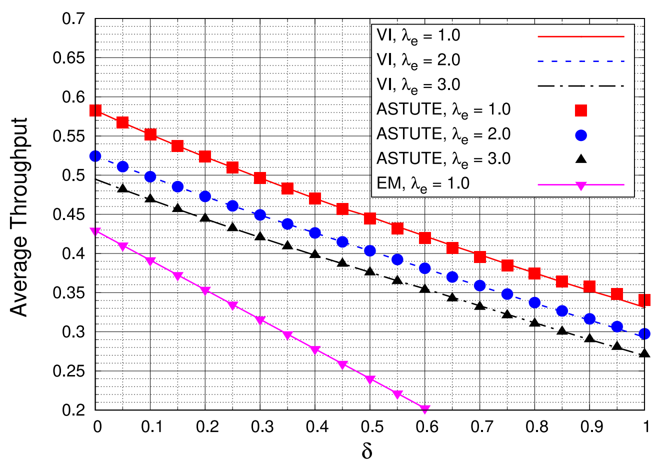


Fig. 4. Average throughput vs normalized reconfiguration delay δ for the optimal policy and ASTUTE. The parameters used are $N = 2$, $p_{1,01} = p_{2,01} = 0.1$, $p_{1,10} = p_{2,10} = 0.1$, $\alpha = 0.1$, and $\beta = 0.8$.

$\Pi_1 \in [0.5, \frac{1+\lambda_e-\delta}{1+\lambda_e}]$, where the parameters of the straight lines can be estimated.

Average throughput performance (see (9)) for the optimal policy obtained through value iteration is plotted in Fig. 4. For the parameters described, we also compute the average throughput of ASTUTE, and is plotted in Fig. 4. For λ_e taking values 1, 2, and 3, we see that there is a close match between ASTUTE and the optimal policy. Also, we note that as λ_e increases from 1.0 to 3.0, the average throughput decreases as the channel access becomes more conservative.

Also, one can have channel access based on $\arg\max_i \pi_i$, which is plotted as Expectation Maximization (EM) in Fig. 4. Note that for a given a posterior state $\pi_i[k]$, we can show that the expected number of time-slots PU i stays in state 1 is

$$\frac{(1-\pi_i[k])p_{i,01} + \pi_i[k]p_{i,11}}{p_{i,10}}.$$

Thus, EM corresponds to choosing channel that remains in idle state for the longest time. It is observed that it switches too often (as it is oblivious to delay δ), and thus loses throughput.

VI. Conclusion

In this letter, we describe the problem of channel access in a multichannel CRNs when there is a reconfiguration delay in switching across channels. We pose the problem as a POMDP, and obtain an average throughput optimum policy through value iteration. The optimum policy has been shown to have a threshold structure. Based on the structure, we propose an algorithm ASTUTE, and show numerically that our proposed algorithm performs very close to that of the optimum policy.

This work can be extended to M SUs. Here, each SU can make a local channel access decision by solving the optimum channel access problem that is reported in the paper, independent of other SUs. The resulting local decisions of the SUs can be reported to a fusion center which would make a global decision for channel access for each of the SUs. A distributed channel access scheme (without a fusion center) can also be designed in which the SUs can employ a MAC protocol to resolve the channel access contention.

Also, in this work, we consider the same cost for switching from any channel to any other channel. As a future direction, this can be extended to include different switching costs. Also, if the busy/idle traffic statistics of the PUs are not available to the decision maker, then, it becomes imperative to look for learning algorithms to solve this problem. This would be a good scope for future work.

References

[1] H. Ding, X. Li, Y. Ma, and Y. Fang, "Energy-efficient channel switching in cognitive radio networks: A reinforcement learning approach," *IEEE Trans. Veh. Technol.*, vol. 69, no. 10, pp. 12359–12362, Oct. 2020.

[2] S. Demirci and D. Gözüpek, "Switching cost-aware joint frequency assignment and scheduling for industrial cognitive radio networks," *IEEE Trans. Ind. Informat.*, vol. 16, no. 7, pp. 4365–4377, Jul. 2020.

[3] H. Shokri-Ghadikolaei and C. Fischione, "Analysis and optimization of random sensing order in cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 5, pp. 803–819, May 2015.

[4] H. Jiang, L. Lai, R. Fan, and H. V. Poor, "Optimal selection of channel sensing order in cognitive radio," *IEEE Trans. Wireless Commun.*, vol. 8, no. 1, pp. 297–307, Jan. 2009.

[5] S. Wang, Y. Wang, J. P. Coon, and A. Doufexi, "Energy-efficient spectrum sensing and access for cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 61, no. 2, pp. 906–912, Feb. 2012.

[6] F. Sheikholeslami, M. Nasiri-Kenari, and F. Ashtiani, "Optimal probabilistic initial and target channel selection for spectrum handoff in cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 1, pp. 570–584, Jan. 2015.

[7] L. Chen, S. Iellamo, and M. Coupechoux, "Opportunistic spectrum access with channel switching cost for cognitive radio networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kyoto, Japan, Jun. 2011, pp. 1–5.

[8] S. Agarwal and S. De, "Impact of channel switching in energy constrained cognitive radio networks," *IEEE Commun. Lett.*, vol. 19, no. 6, pp. 977–980, Jun. 2015.

[9] Y. Wu, Q. Yang, X. Liu, and K. S. Kwak, "Delay-constrained optimal transmission with proactive spectrum handoff in cognitive radio networks," *IEEE Trans. Commun.*, vol. 64, no. 7, pp. 2767–2779, Jul. 2016.

[10] D. P. Bertsekas, *Dynamic Programming and Optimal Control, Vol. I*, 3rd ed. Belmont, MA, USA: Athena Sci., 2007.

[11] V. Krishnamurthy, *Partially Observed Markov Decision Processes: From Filtering to Controlled Sensing*, 1st ed. Cambridge, U.K.: Cambridge Univ. Press, 2016.