On Optimality of Myopic Policy in Opportunistic Spectrum Access: The Case of Sensing Multiple Channels and Accessing One Channel

Kehao Wang, Lin Chen, *Member, IEEE,* Khaldoun Al Agha, *Senior Member, IEEE,* and Quan Liu

***Abstract*—Recent works have developed a simple and robust myopic sensing policy for multi-channel opportunistic communication systems where a secondary user (SU) can access one of N i.i.d. Markov channels. The optimality of the myopic sensing policy maximizing the SU's expected accumulative reward is established under certain conditions on channel parameters. This paper studies the scenario where the SU, equipped with one radio, can sense k channels but access only one channel each time slot. The objective of the SU is to maximize its throughput over the time horizon T. By characterizing the myopic sensing policy in this context, we establish analytically its optimality for a subset of specific system settings, notably the cases $k = 2$, $T = 2$ and $k = N - 1$ with arbitrary T. In the more generic case, we construct counterexamples to show that the myopic sensing policy is not optimal.**

***Index Terms*—Opportunistic spectrum access (OSA), myopic sensing policy, partially observed Markov decision process (POMDP), restless multi-armed bandit problem (RMAB).**

I. Introduction

THE concept of opportunistic spectrum access (OSA), first envisioned by J. Mitola in the seminal paper [1] on the software defined radio systems, has emerged in recent years as a promising paradigm to enable more efficient spectrum utilization. The basic idea of OSA is to exploit instantaneous spectrum availability by allowing the unlicensed secondary users (SU) to access the temporarily unused channels of the licensed primary users (PU) in an opportunistic fashion. In this context, a well-designed channel access policy is crucial to achieve efficient spectrum usage.

In this paper, we consider a generic OSA scenario where there are N slotted spectrum channels partially occupied by the PUs. Each channel evolves as an independent and identically distributed (i.i.d.), two-state discrete-time Markov chain. The two states for each channel, busy (state 0) and idle (state 1), indicate whether the channel is free for an SU to transmit its packet on that channel at a given slot. The state transition probabilities are given by $\{p_{ij}\}, i, j = 0, 1$. An SU equipped with one radio seeks a sensing policy that opportunistically exploits the temporarily unused channels to transmit its packets. To this end, in each slot, the SU is allowed to sense k channels based on its prior observations and obtain one unit as reward if one of the sensed channel is in the idle state, indicating that the SU can effectively send one packet. The objective of the SU is to find the optimal sensing policy maximizing the expected reward that it can obtain over a finite or infinite time horizon. We assume perfect sensing for the ease of presentation in this paper.

As stated in [2], the design of the optimal sensing policy can be formulated as a partially observable Markov decision process (POMDP), or a restless multi-armed bandit problem (RMAB), of which the application is far beyond the domain of cognitive radio systems[1]. Unfortunately, obtaining the optimal policy for a general POMDP or RMAB is often intractable due to the exponential computation complexity. Hence, a natural alternative is to seek simple myopic policies for the SU. In this line of research, a myopic sensing strategy is developed in [4], [5] where the SU is limited to sense one and multiple channels at each slot, and the myopic sensing policy is proven to be optimal when the state transitions of the Markov channels are positively correlated, i.e., $p_{11} \geq p_{01}$.

In this paper, we extend the proposed myopic policy to the scenario where the SU, equipped with one radio, can sense k channels but can access only one channel each time slot, and get one unit of reward if one of the sensed channels is in the idle state. Through mathematic analysis, we show that the generalized myopic sensing policy is optimal only for a small subset of cases, notably the cases $k = 2, T = 2$ and $k = N-1$ with arbitrary T. In other cases, we give counterexamples to show that the myopic sensing policy is not optimal.

It is insightful to note that the work most relevant to our study presented in this paper is [5] where the user can sense multiple channels and access all that are sensed idle. In our study, at most one available channel can be used due to system limitation. In terms of engineering implication, our work differs from [5] in that the sensing decision serves an immediate purpose, which is the reward in [5]; in our model as long as there is one available channel, the sensing outcome of the other channels serves a more long-term purpose, which is to keep the channel state information up to date. We believe that our results presented in this paper, together with [3], [4], [5], lead to more in-depth understanding of the intrinsic structure and the resulting optimality of the myopic sensing policy and will stimulate more profound research on this topic.

Manuscript received May 2, 2012. The associate editor coordinating the review of this letter and approving it for publication was H. Jiang.

The work presented in the paper is supported in part by the China Scholarship Council, the National Natural Science Foundation of China (No. grant 51175389), and the Centre National de la Recherche Scientifique (No. grant EDC25064).

K. Wang, L. Chen (corresponding author), and K. Al Agha are with the Laboratoire de Recherche en Informatique (LRI), Department of Computer Science, the University of Paris-Sud XI, 91405 Orsay, France (e-mail: {Kehao.Wang, Lin.Chen, Khaldoun.Alagha}@lri.fr).

K. Wang and Q. Liu are with the School of Information Engineering, Wuhan University of Technology, 430070 Hubei, China (e-mail: {Kehao.wang, Quan.Liu}@whut.edu.cn).

Digital Object Identifier 10.1109/WCL.2012.062512.120326

[1]Please refer to [3], [4] for more examples where this formulation is applicable. A summary on the related works on the analysis of this problem using the POMDP and RMAB approaches are presented in [4]. We thus do not provide a literature survey in this paper.

II. Problem Formulation

As explained in the Introduction, we are interested in a synchronously slotted cognitive radio network where an SU, equipped with one radio, can opportunistically sense a set $\mathcal{N}$ of N i.i.d. channels partially occupied by PUs and access only one channel. The state of each channel i in time slot t, denoted by $S_i(t)$, is modeled by a discrete time two-state Markov chain. At the beginning of each slot t, the SU selects a subset $\mathcal{A}(t)$ of k channels to sense sequentially. If at least one of the sensed channels is in the idle state (i.e., unoccupied by any PU), the SU transmits its packet using its radio and collects one unit of reward. Otherwise, the SU cannot transmit, thus obtaining no reward.

We use similar notations as those adopted in [3], [4]. More specifically, let $\omega_i(t)$ denote the conditional probability that $S_i(t)=1$ given the past actions and observations, based on the sensing policy $\mathcal{A}(t)$ in slot t and the sensing result. Define the channel belief vector $\Omega(t) \triangleq \{\omega_i(t), i\in\mathcal{N}\}$. $\Omega(t)$ can be updated using Bayes Rule as shown in (1).

$$\omega_i(t+1)=\begin{cases} p_{11}, & i\in\mathcal{A}(t), S_i(t)=1\\ p_{01}, & i\in\mathcal{A}(t), S_i(t)=0\\ \tau(\omega_i(t)), & i\notin\mathcal{A}(t)\end{cases}, \qquad (1)$$

where $\tau(\omega_i(t)) \triangleq \omega_i(t)p_{11}+[1-\omega_i(t)]p_{01}$ characterizes the evolution of the belief value of the non-sensed channels. It is easy to check the following property:

$$\begin{cases} p_{01}\le\tau(\omega_i)\le p_{11}, & p_{01}\le p_{11}\\ p_{11}\le\tau(\omega_i)\le p_{01}, & p_{01}> p_{11}\end{cases}. \qquad (2)$$

Formally, let $\mathbb{E}[R_{\pi_t}(t)]$ denote the expected reward obtained at slot t under a sensing policy π_t sensing channels in $\mathcal{A}(t)$ for slot t, we have

$$\mathbb{E}[R_{\pi_t}(\Omega(t))]=1-\prod_{i\in\mathcal{A}(t)}(1-\omega_i(t)).$$

The focus of our work is to study the optimal sensing policy of the SU in order to maximize the average reward over T slots. Formally, the optimization problem for the SU P_{SU}, when the SU is allowed to sense k channels given $\Omega(1)$[2], is formally defined as follows:

$$P_{SU}:\ \max_{\pi}\ \frac{1}{T}\mathbb{E}\left[\sum_{t=1}^{T}R_{\pi_t}(\Omega(t))|\Omega(1)\right]. \qquad (3)$$

As already shown in [3], [4], [5], the optimization problem P_{SU} is by nature a POMDP, or a restless multi-armed bandit problem, of which the optimal sensing policy is in general intractable. Hence, a natural alternative is to seek simple myopic sensing policy, i.e., the sensing policy maximizing the immediate reward based on current belief.

Definition 1 (Myopic Sensing)**.** *Sort $\{\omega_i(t)\}_{i=1}^N$ such that $\omega_1(t)\ge\cdots\ge\omega_N(t)$, the myopic policy consists of sensing channel 1 to channel k.*

The myopic sensing policy is easy to implement and maximizes the immediate payoff. In the next section, we show that the myopic sensing policy is optimal for the case $k=N-1$, $k=2$ and $T=2$ when $p_{11}\ge p_{01}$ and when $p_{11}<p_{01}$ and $N\le 4$. Beyond this small subset of parameter settings, we show that the myopic sensing policy is not optimal by constructing several representative counterexamples.

III. Optimality of Myopic Sensing Policy

In this section, we study the optimality of the myopic sensing policy. We structure our analysis into three cases from the particular case to the general case: (1) $T=2$, $k=2$; (2) $k=N-1$; (3) the general case.

A. Optimality of myopic sensing policy when $T=2$ and $k=2$

This subsection is focused on the case where the SU is allowed to sense two channels each slot and aims at maximizing the expected reward of the upcoming two slots. The following two theorems study the optimality of the myopic sensing policy for $p_{11}\ge p_{01}$ and $p_{11}<p_{01}$, respectively.

Theorem 1. *When $T=2$ and $k=2$, the myopic sensing policy is optimal for $p_{11}\ge p_{01}$.*

Proof: We sort $\{\omega_i(t)\}_{i=1}^N$ at the beginning of slot t such that $\omega_1\ge\cdots\ge\omega_N$[3]. Under this notation, the expected reward of the myopic sensing policy (i.e., sensing channels 1 and 2), denoted as R^*, is $R^*=R^*(t)+R^*(t+1)$, where $R^*(t)$, $R^*(t+1)$ denote the expected reward of the myopic sensing policy of slot t and $t+1$, respectively. $R^*(t)$ can be calculated as

$$R^*(t)=1-(1-\omega_1)(1-\omega_2).$$

We now derive $R^*(t+1)$ by distinguishing four cases:

- Case 1: in slot t, both channel 1 and channel 2 are sensed idle, which happens with probability $\omega_1\omega_2$. In this case, recall (1) and (2), the myopic sensing policy for slot $t+1$ is to still sense channels 1 and 2, resulting an expected reward $1-(1-p_{11})(1-p_{11})$ following (1).
- Case 2: in slot t, channel 1 is sensed idle and channel 2 busy, which happens with probability $\omega_1(1-\omega_2)$. In this case, recall (1) and (2), the myopic sensing policy for slot $t+1$ is to sense channel 1 and channel 3, resulting an expected reward $1-(1-p_{11})(1-\tau(\omega_3))$ following (1).
- Case 3: in slot t, channel 1 is sensed busy and channel 2 idle, this is the symmetrical scenario of Case 2. The expected reward in this case is also $1-(1-p_{11})(1-\tau(\omega_3))$.
- Case 4: in slot t, both channel 1 and channel 2 are sensed busy, which happens with probability $(1-\omega_1)(1-\omega_2)$. In this case, recall (1) and (2), we can further distinguish two subcases:
  - Case 4.1: $N\ge 4$. In this subcase, the myopic sensing policy for slot $t+1$ is to sense channels 3 and 4, resulting an expected reward $1-(1-\tau(\omega_3))(1-\tau(\omega_4))$ following (1).
  - Case 4.2: $N=3$. In this subcase, the myopic sensing policy for slot $t+1$ is to sense channels 3 and either

[2] $\omega_i(1)$ of $\Omega(1)$ can be set to $\frac{p_{01}}{p_{01}+p_{10}}$ if no information about the initial system state is available.

[3] For the simplicity of presentation, by slightly abusing the notations without introducing ambiguity, we drop the time slot index of $\omega_i(t)$.

channel 1 or channel 2 as their expected idle probabilities are both p_{01} following (1). In this subcase, the resulting expected reward is $1-(1-\tau(\omega_3))(1-p_{01})$.

Following the above analysis, we have

$$R^* = 1-(1-\omega_1)(1-\omega_2)+\omega_1\omega_2[1-(1-p_{11})(1-p_{11})] + \omega_1(1-\omega_2)[1-(1-p_{11})(1-\tau(\omega_3))] + (1-\omega_1)\omega_2[1-(1-p_{11})(1-\tau(\omega_3))] + (1-\omega_1)(1-\omega_2)[1-(1-\tau(\omega_3))(1-F)], \quad (4)$$

where $F = p_{01}$ when $N=3$ and $\tau(\omega_4)$ when $N \geq 4$.

We now show that sensing any two channels $\{i,j\} \neq \{1,2\}$ cannot bring the SU more reward. We proceed the proof by distinguishing the following two cases:

- Case 1: $\{i,j\}$ is partially overlapped with $\{1,2\}$, i.e., $\{i,j\}\bigcap\{1,2\} \neq \emptyset$;
- Case 2: $\{i,j\}$ is totally distinct to $\{1,2\}$, i.e., $\{i,j\}\bigcap\{1,2\} = \emptyset$;

Case 1. When $\{i,j\}$ is partially overlapped with $\{1,2\}$, without loss of generality, assume that $i=1$ and $j \geq 3$, we can derive the upper bound of the expected reward of sensing the channels $\{i,j\}=\{1,j\}$. Here by upper bound we mean that the SU senses channel i and j in slot t and the two channels with the largest idle probabilities for slot $t+1$, leading to the maximal reward that the SU can achieve.

- When $j=3$, following the similar analysis as that in (4), we can get the utility upper bound $\overline{R}_1$ when sensing the channels $\{i,j\}=\{1,3\}$. By some algebraic operations, we obtain

  $$R^* - \overline{R}_1 = (1-\omega_1)(\omega_2-\omega_3)(1-(1-p_{11})(F-p_{01})),$$

  where F is defined in (4). Noticing that $p_{01} \leq \tau(\omega_i) \leq p_{11}, \forall i \in \mathcal{N}$ following (1), it holds that $F \geq p_{01}$. Hence $R^* - \overline{R}_1 \geq 0$ holds for $j=3$.
- When $j \geq 4$,we obtain, by similar induction, that $R^* - \overline{R}_1 \geq 0$ holds for $j \geq 4$, too.

The above results show that any other sensing policy cannot outperform the myopic sensing policy in this case.

Case 2. When $\{i,j\}$ is totally distinct to $\{1,2\}$, implying $N \geq 4$, we can obtain the reward upper bound $\overline{R}_2$ of the sensing policy $\{i,j\}$. It then follows $R^* - \overline{R}_2 \geq 0$, meaning that sensing $\{i,j\}$ cannot outperform the myopic sensing policy in this case, either. Combining the results of both cases completes the proof of Theorem 1. ■

The following theorem studies the optimality of the myopic sensing policy when $p_{11} < p_{01}$. The proof follows the similar way as that of Theorem 1 and is thus omitted.

Theorem 2. *When $k=2$, $N \leq 4$ and $T=2$, the myopic policy is optimal for $p_{11} < p_{01}$.*

B. Optimality of myopic sensing policy when $k = N-1$

In this subsection, we show that the myopic sensing policy is optimal when the SU can sense $N-1$ out of N channels.

To that end, inspired by the analysis in [5], let $V_t(\Omega;\mathcal{A}(t))$ denote the expected accumulative reward obtained by sensing $\mathcal{A}(t)$ in slot t followed by the myopic policy in subsequent slots, and $p_{11}[n]$ ($p_{01}[n]$) denote the vector of length n with each element being p_{11} (p_{01}). To prove the optimality of the myopic sensing policy, we establish three lemmas regarding $V_t(\Omega;\mathcal{A}(t))$.

Lemma 1. $V_t(\omega_1,\cdots,\omega_i,\cdots,\omega_j,\cdots,\omega_N;\mathcal{A}(t)) = V_t(\omega_1,\cdots,\omega_j,\cdots,\omega_i,\cdots,\omega_N;\mathcal{A}(t))$ *for* $\forall i,j \in \mathcal{A}(t)$.

Proof: The proof is straightforward by noticing that both the immediate reward and the channel belief vector $\Omega(t+1)$ are unrelated with the sensing order of ω_i,ω_j. ■

Lemma 2. *It holds that V_t is an affine function, i.e.,*

$$V_t(\omega_1,\cdots,\omega_i,\cdots,\omega_N;\mathcal{A}(t)) = \omega_i V_t(\omega_1,\cdots,1,\cdots,\omega_N;\mathcal{A}(t)) + (1-\omega_i)V_t(\omega_1,\cdots,0,\cdots,\omega_N;\mathcal{A}(t)).$$

Proof: We prove the lemma by induction. It can be easily checked that the lemma holds for slot T. Assume that it holds for slot $T,\cdots,t+1$, we now prove it holds for slot t. We proceed by distinguishing the following two cases:

Case 1: $i \notin \mathcal{A}(t)$. In this case we have

$$V_t(\Omega;\mathcal{A}(t)) = 1-\prod_{j\in\mathcal{A}(t)}(1-\omega_j) + \sum_{e\in\mathcal{A}(t)}\prod_{p\in e}\omega_p\prod_{q\in\mathcal{A}(t)\setminus e}(1-\omega_q)\cdot V_{t+1}(p_{11}[|e|],\tau(\omega_i),p_{01}[N-1-|e|]).$$

By induction, $V_{t+1}(\Omega(t+1))$ is an affine function of $\tau(\omega_i)$, and meanwhile, $\tau(\omega_i)$ is an affine transform of ω_i, thus $V_{t+1}(\Omega(t+1))$ is an affine function of ω_i. It follows that $V_t(\Omega(t);\mathcal{A}(t))$ is also an affine function of ω_i.

Case 2: $i \in \mathcal{A}(t)$. Let j denote the channel not selected in slot t ($j \notin \mathcal{A}(t)$), and let $\mathcal{A}'(t) = \mathcal{A}(t)\setminus\{i\}$, we have

$$\begin{aligned} &V_t(\Omega;\mathcal{A}(t)) \\ =&1-\prod_{l\in\mathcal{A}(t)}(1-\omega_l) + \sum_{e\in\mathcal{A}(t)}\prod_{p\in e}\omega_p\prod_{q\in\mathcal{A}(t)\setminus e}(1-\omega_q)\cdot \\ &V_{t+1}(p_{11}[|e|],\tau(\omega_j),p_{01}[N-1-|e|]) \\ =&1-\prod_{l\in\mathcal{A}(t)}(1-\omega_l) + \sum_{e\in\mathcal{A}'(t)}\prod_{p\in e}\omega_p\prod_{q\in\mathcal{A}'(t)\setminus e}(1-\omega_q)\cdot \\ &\Big[\omega_i V_{t+1}(p_{11}[|e|+1],\tau(\omega_j),p_{01}[N-2-|e|]) \\ &+(1-\omega_i)V_{t+1}(p_{11}[|e|],\tau(\omega_j),p_{01}[N-1-|e|])\Big] \end{aligned}$$

Obviously, $1-\prod_{l\in\mathcal{A}(t)}(1-\omega_l) = 1-(1-\omega_i)\prod_{l\in\mathcal{A}'(t)}(1-\omega_l)$ is an affine function of ω_i, the second term is also an affine function of ω_i by induction. Therefore, $V_t(\Omega(t);\mathcal{A}(t))$ is an affine function of ω_i. ■

Lemma 3. *Let $\mathcal{A}(t) = \mathcal{N}\setminus\{j\}$ and $\mathcal{A}'(t) = \mathcal{N}\setminus\{i\}$ where $\omega_i(t) \geq \omega_j(t)$, it holds that $V_t(\Omega;\mathcal{A}(t)) \geq V_t(\Omega;\mathcal{A}'(t))$.*

Proof: We prove by backward induction. It can be easily checked that the lemma holds for slot T. Assume that it holds for slot $T,\cdots,t+1$, we now prove it holds for slot t.

In the case $p_{11} > p_{01}$, we have

$$\begin{aligned} &V_t(\Omega;\mathcal{A}(t)) - V_t(\Omega;\mathcal{A}'(t)) \\ =&V_t(\cdots,\omega_i,\omega_j;\mathcal{A}(t)) - V_t(\cdots,\omega_j,\omega_i;\mathcal{A}'(t)) \\ =&(\omega_i-\omega_j)\big[V_t(\cdots,1,0;\mathcal{A}(t)) - V_t(\cdots,0,1;\mathcal{A}'(t))\big] \\ =&(\omega_i-\omega_j)\Big[\omega_i\prod_{l=1,\neq i,j}^{N}(1-\omega_l) + \prod_{l=1,\neq i,j}^{N}\omega_l\cdot \end{aligned}$$

$$\Big[V_{t+1}(p_{11}[N-2],p_{11},p_{01})-V_{t+1}(p_{11}[N-2],p_{01},p_{11})\Big]\Big]$$

$$\geq(\omega_i-\omega_j)\Big[\omega_i\prod_{l=1,\neq i,j}^{N}(1-\omega_l)\Big]\geq 0$$

where, the first inequality follows the induction, the second equality follows Lemma 2, the third equality is due to the fact that if any channel l $(1 \leq l \leq N, l \neq i,j)$ is sensed busy (happened with $1-\prod_{l=1,\neq i,j}^{N}\omega_l$), then $V_{t+1}(p_{11}[N-m],p_{11},p_{01},p_{01}[m-2])-V_{t+1}(p_{11}[N-m],p_{01},p_{11},p_{01}[m-2])=0$ $(m\geq 3)$ according to Lemma 1.

The case of $p_{11} < p_{01}$ can be proved similarly. ■

Theorem 3. *The myopic policy is optimal if $k = N-1$.*

Proof: We prove the theorem by induction. In slot T, the optimality of the myopic policy is obvious. Assume that the myopic policy is also optimal for slot $T-1,...,t+1$. We prove it holds for slot t. To that end, we sort $\Omega(t)$ in the decreasing order such that $\omega_1 \geq \cdots \geq \omega_N$. To prove the optimality of myopic policy, we need to show that $V_t(\Omega;\mathcal{A}(t)) \geq V_t(\Omega;\mathcal{A}'(t))$ where $\mathcal{A}(t)=\{1,\cdots,N-1\}=\mathcal{N}\setminus\{N\}$ and $\mathcal{A}'(t)$ is any $N-1$ elements of $\mathcal{N}$. Without loss of generality, we assume $\mathcal{A}'(t)=\mathcal{N}\setminus\{l\}$. Noticing that $\omega_l \geq \omega_N$, it follows from Lemma 3 that $V_t(\Omega;\mathcal{A}(t)) \geq V_t(\Omega;\mathcal{A}'(t))$. ■

C. Non-optimality of myopic sensing policy in general cases

In this subsection, we show that the myopic sensing policy is not optimal for the general cases beyond those studied in the previous subsections by constructing several representative counterexamples.

Counterexample 1 ($k=3$, $T=2$, $N=6$, $p_{11}\geq p_{01}$)**.** Let R_{c1}^* denote the expected reward generated by the myopic sensing policy (sensing the 3 channels with highest elements in the believe vector at each slot, i.e., ω_1, ω_2, ω_3). Let R_{c1} denote the expected reward from the sensing policy that senses the 2 highest elements and the forth highest element in the believe vector (i.e., ω_1, ω_2 and ω_4 according to our notation) for the current slot t and senses the highest 3 elements in the believe vector for slot $t+1$. It can be calculated that under the setting $[\omega_1,\omega_2,\omega_3,\omega_4,\omega_5,\omega_6] = [0.99,0.5,0.4,0.39,0.25,0.25]$, $p_{11}=0.5$, $p_{01}=0.3$, it holds that $R_{c1}-R_{c1}^*=0.00005625>0$. The myopic sensing policy is not optimal for this counterexample.

Counterexample 2 ($k=2$, $T=3$, $N=4$, $p_{11}>p_{01}$)**.** Let R_{c2}^* denote the expected reward of the myopic sensing policy and R_{c2} that of the policy of sensing channels $1,3$ for the first slot and the 2 channels with the largest belief values for the following two slots. We have $R_{c2}-R_{c2}^*>0$ with the parameters $[\omega_1,\omega_2,\omega_3,\omega_4]=[1.0,0.8,0.7,0.6]$, $p_{11}=0.8$, $p_{01}=0.3$. The myopic sensing policy is not optimal for this counterexample, either.

Counterexample 3 ($k=2$, $T=2$, $N=5$, $p_{11}<p_{01}$)**.** Let R_{c3}^* denote the expected reward of the myopic sensing policy and R_{c3} that of the policy of sensing channels $1,3$ for the first slot and the 2 channels with the largest belief values for the second slot. We have $R_{c3}-R_{c3}^*=0.0005>0$ with the parameters $[\omega_1,\omega_2,\omega_3,\omega_4,\omega_5]=[1.0,0.8,0.7,0.5,0.0]$, $p_{11}=0.3$, $p_{01}=0.8$. The myopic sensing policy is not optimal for this counterexample, either.

To conclude this section, it is insightful to note that the major results of this paper on the optimality of the myopic sensing policy hinge on the fundamental trade-off between exploration, by sensing unexplored channels in order to learn and predict the future channel state, thus maximizing the long-term reward, and exploitation, by accessing the channel with the highest estimated idle probability based on currently available information (the belief vector) which greedily maximizes the immediate reward. For a short-sighted SU ($T=1$ and $T=2$), exploitation naturally dominates exploration (i.e., the immediate reward overweighs the potential gain in future reward) under certain system parameter settings, resulting the optimality of the myopic sensing policy in a subset of this scenario. When sensing $N-1$ of N, the SU can completely obtain enough information to justify the optimality of myopic policy. In contrast, to achieve maximal reward for $T\geq 3$ and $1<k<N-1$, the SU should strike a balance between exploration and exploitation, and thus the myopic sensing policy that greedily maximizes the immediate reward is no more optimal.

IV. Conclusion

In this paper, we study the optimality of the myopic sensing policy with perfect sensing in the generic scenario of opportunistic spectrum access in a multi-channel communication system where an SU senses a subset of channels partially occupied by licensed PUs. We show that the myopic sensing policy is optimal only for a small subset of cases. In the generic case, we give counterexamples to show that the myopic sensing policy is not optimal. Due to the generic nature of myopic policy, we believe that the results obtained in this paper lead to more in-depth understanding of the intrinsic structure and optimality of the myopic policy, and will stimulate more profound research on this topic.

References

[1] J. Mitola, "Cognitive radio for flexible mobile multimedia communications," in *1999 IEEE Int. Workshop on Mobile Multimedia Communications*.

[2] Q. Zhao, L. Tong, A. Swami, and Y. Chen, "Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: a POMDP framework," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 3, Apr. 2007.

[3] Q. Zhao, B. Krishnamachari, and K. Liu, "On myopic sensing for multichannel opportunistic access: structure, optimality, and performance," *IEEE Trans. Wireless Commun.*, vol. 7, no. 3, pp. 5413-5440, Dec. 2008.

[4] S. Ahmand, M. Liu, T. Javidi, Q. Zhao, and B. Krishnamachari, "Optimality of myopic sensing in multichannel opportunistic access," *IEEE Trans. Inf. Theory*, vol. 55, no. 9, pp. 4040-4050, Sep. 2009.

[5] S. Ahmad and M. Liu, "Multi-channel opportunistic access: a case of restless bandits with multiple plays," in *2009 Allerton Conference*.