On Optimality of Myopic Policy in Multi-channel Opportunistic Access

Kehao Wang
Key Lab of Fiber Optic Sensing Tech. & Information Processing
Wuhan University of Technology, Hubei, P.R.C.
Email: kehao.wang@whut.edu.cn

Lin Chen, Jihong Yu
Université de Paris-Sud
Orsay, France
Email: {lin.chen,jihong.yu}@lri.fr

***Abstract*—We consider the channel access problem arising in opportunistic scheduling over fading channels, cognitive radio networks, and server scheduling. The multi-channel communication system consists of N channels. Each channel evolves as a time-nonhomogeneous multi-state Markov process. At each time instant, a user chooses M channels to transmit information. Some reward depending on the states of the chosen channels is obtained for each transmission. The objective is to design an access policy that maximizes the expected accumulated discounted reward over a finite or infinite horizon. The considered problem can be cast into a restless multi-armed bandit (RMAB) problem with PSPACE-hardness. A natural alternative is to consider the easily implementable myopic policy. In this paper, we perform an theoretical analysis on the considered RMAB problem, and establish a set of closed-form conditions to guarantee the optimality of the myopic policy.**

I. Introduction

Consider a communication system composed of N independent channels each of which is modeled as a *time-nonhomogeneous* X-state Markov chain with known matrix of transition probabilities. At each time period a user selects M channels to transmit information. Some reward depending on the states of those selected channels is obtained for each transmission. The objective is to design a channel access policy that maximizes the expected accumulated discounted reward (respectively, the expected accumulated reward) collected over a finite (respectively, infinite) time horizon. Mathematically, the considered channel access problem can be cast into the restless multi-armed bandit (RMAB) problem of fundamental importance in decision theory [1]. As well as we know, RMAB problems arise in many areas, such as wired and wireless communication systems, manufacturing systems, economic systems, statistics, biomedical engineering, and information systems etc. [1, 2].

There exist two major thrusts in the research of the RMAB problem. Since the optimality of the myopic policy is not generally guaranteed, the first research thrust is to analyze the performance difference between the optimal policy and approximation policy [7–9]. Specifically, a simple myopic policy, also called greedy policy, is developed in [7] which yields a factor 2 approximation of the optimal policy for a subclass of scenarios referred to as *Monotone MAB*. The second thrust is to establish sufficient conditions to guarantee the optimality of the myopic policy in some specific instances of restless bandit scenarios, particularly in the context of opportunistic communications [10–14].

For the case of *two-state*, Zhao *et al.* [10] established the structure of the myopic policy, and partly obtained the optimality for the case of i.i.d. channels. After that, Ahmad and Liu *et al.* [15] derived the optimality of the myopic sensing policy for the positively correlated i.i.d. channels for accessing one channel (i.e., $k = 1$) each time, and further extended the optimality to access multiple i.i.d. channels ($k > 1$) [12]. From another point, in [14], the authors extended i.i.d. channels [15] to non i.i.d. ones, and focused on a class of so-called *regular* functions, and derived closed-form sufficient conditions to guarantee the optimality of myopic sensing policy. For the complicated case of *multi-state*, the authors in [19] established the sufficient conditions for the optimality of myopic sensing policy in multi-state homogeneous channels with a set of assumptions.

In this paper, we consider heterogeneous multi-state channels, different from the scenario of homogeneous channels in [19], and construct a set of conditions to guarantee the optimality of myopic policy. In particular, the contributions of this paper include: 1) The structure of the myopic policy is shown to be a simple queue determined by the availability probability vector of channels provided that certain condition is satisfied for the transition matrix of multi-state channels. Further, a set of conditions is obtained under which the myopic policy is proved to be optimal; 2) Our derivation demonstrates the advantage of branch-and-bound and the directed comparison based optimization approach. The results of this paper are a generic contribution to the state of the art of the theory of restless bandit problems, although the structure of the optimal policy of generic restless bandit is not known.

II. Model and the Optimization Problem

A. System Model

We consider a time-slotted multi-channel communication system consisting of N channels, denoted as $\mathcal{N}$, and the slot index is t. Each of channel, i.e., n-th channel, is modeled as a *time-nonhomogeneous* X-state Markov chain with matrix of

978-1-4799-6664-6/16/$31.00 ©2016 IEEE

transition probabilities $\mathbf{P}^{(n)}(t)$,

$$\mathbf{P}^{(n)}(t) = \begin{pmatrix} p_{11}^{(n)}(t) & p_{12}^{(n)}(t) & \cdots & p_{1X}^{(n)}(t) \\ p_{21}^{(n)}(t) & p_{22}^{(n)}(t) & \cdots & p_{2X}^{(n)}(t) \\ \vdots & \vdots & \ddots & \vdots \\ p_{X1}^{(n)}(t) & p_{X2}^{(n)}(t) & \cdots & p_{XX}^{(n)}(t) \end{pmatrix} = \begin{pmatrix} P_1^{(n)}(t) \\ P_2^{(n)}(t) \\ \vdots \\ P_X^{(n)}(t) \end{pmatrix}, \tag{1}$$

where, $P_1^{(n)}(t), \cdots, P_X^{(n)}(t)$ are row vectors.

We want to use this communication system to transmit information. For that matter, at each time $t = 0, 1, 2, \cdots, T$, we can select M channels, observe their states, and use them to transmit information.

Let $S_n(t)$ denote the state of channel n at time t, and let $\mathcal{A}(t)$ denote the decision made at time t where $\mathcal{A}(t) \subseteq \mathcal{N}$ and $|\mathcal{A}(t)| = M$. Then we have the state vector $\mathbf{S}(t) = [S_1(t), \cdots, S_N(t)]$.

Initially, before any channel selection is made, we assume that we have probabilistic information about the state of each of the N channels. Specifically, we assume that at $t = 0$, the decision-make knows the probability mass function on the state space of each of the N channels; that is, the decision-maker knows

$$\mathbf{\Omega}(0) = [\mathbf{w}_1(0), \mathbf{w}_2(0), \cdots, \mathbf{w}_N(0)],$$

where,

$$\mathbf{w}_n(0) \triangleq [\mathbf{w}_{n1}(0), \mathbf{w}_{n2}(0), \cdots, \mathbf{w}_{nX}(0)], \quad n \in \mathcal{N},$$
$$\mathbf{w}_{nx}(0) \triangleq \mathbb{P}(S_n(0) = x), \quad x \in \mathcal{X}.$$

In general,

$$\mathcal{A}(0) = \rho\big(\mathbf{\Omega}(0)\big),$$
$$\mathcal{A}(t) = \rho\big(\mathcal{O}_{t-1}, \mathcal{A}_{t-1}, \mathbf{\Omega}(0)\big),$$

where,

$$\mathcal{O}_{t-1} \triangleq (\mathcal{O}(0), \mathcal{O}(1), \cdots, \mathcal{O}(t-1)),$$
$$\mathcal{A}_{t-1} \triangleq (\mathcal{A}(0), \mathcal{A}(1), \cdots, \mathcal{A}(t-1)),$$
$$\mathcal{O}(t) \triangleq (O_{\sigma_1}(t), \cdots, O_{\sigma_M}(t)),$$

and $O_{\sigma_m}(t) = S_{\sigma_m}(t)$ ($\sigma_m \in \mathcal{A}(t)$) denotes the observation state of channel σ_m at t.

We assume that the reward obtained from accessing a channel at slot t depends on the state of the channel chosen at t, formally defined as follows:

$$R(S_n(t)) = r_x \text{ if } S_n(t) = x, \tag{2}$$

where, $r_X \geq \cdots \geq r_1$ indicates that the reward obtained in the high SINR channel state is larger than that in the low SINR, and $\mathbf{r} \triangleq [r_1, \cdots, r_X]$ is an X-dimensional row vector.

B. Optimization Problem

The objective is to seek the optimal policy ρ^* that maximizes the expected accumulated discounted reward over a finite horizon:

$$\rho^* = \operatorname*{argmax}_{\rho} \mathbb{E}^{\rho} \left[\sum_{t=0}^{T} \beta^{t-1} R_{\rho_t}(\mathbf{\Omega}(t)) \middle| \mathbf{\Omega}(0) \right], \tag{3}$$

where, $R_{\rho_t}(\Omega(t))$ is the reward collected in slot t under the policy ρ_t with the initial belief vector $\Omega(0)$, β is the discount factor ($0 \leq \beta \leq 1$), and $\rho = (\rho_0, \rho_1, \cdots, \rho_T)$ are such that

$$\mathcal{A}(t) = \rho_t\big(\mathbf{\Omega}(t)\big), \forall t,$$
$$\mathbf{\Omega}(t) = [\mathbf{w}_1(t), \mathbf{w}_2(t), \cdots, \mathbf{w}_N(t)],$$
$$\mathbf{w}_n(t) \triangleq [\mathbf{w}_{n1}(t), \mathbf{w}_{n2}(t), \cdots, \mathbf{w}_{nX}(t)], \quad n \in \mathcal{N},$$
$$\mathbf{w}_{nx}(t) = \mathbb{P}(S_n(t) = x | \mathcal{O}_{t-1}, \mathcal{A}_{t-1}), \quad x \in \mathcal{X},$$

and $\mathbf{w}_n(t+1)$ is updated recursively using the following rule:

$$\mathbf{w}_n(t+1) = \begin{cases} P_x^{(n)}(t), & n \in \mathcal{A}(t), O_n(t) = x \\ \mathbf{w}_n(t)\mathbf{P}^{(n)}(t), & n \notin \mathcal{A}(t). \end{cases} \tag{4}$$

To get more insight on the structure of (3), we rewrite it in the language of dynamic programming as follows:

$$\begin{cases} V_T(\mathbf{\Omega}(T)) = \max_{\mathcal{A}(T)} \mathbb{E}\Big[\sum_{n\in\mathcal{A}(T)} \mathbf{w}_n(T)\mathbf{r}'\Big], \\ V_t(\mathbf{\Omega}(t)) = \max_{\mathcal{A}(t)} \mathbb{E}\Big[\sum_{n\in\mathcal{A}(t)} \mathbf{w}_n(t)\mathbf{r}' \\ \qquad + \beta \underbrace{\Sigma(\mathcal{A}(t), \mathbf{\Omega}(t)) V_{t+1}(\Omega(t+1))}_{F(\mathcal{A}(t), \mathbf{\Omega}(t))}\Big], \end{cases} \tag{5}$$

where,

$$\Sigma(\mathcal{A}(t), \mathbf{\Omega}(t)) \triangleq \sum_{\bigcup_{x=1}^{X} A_x = \mathcal{A}(t)}^{\bigcap_{x=1}^{X} A_x = \varnothing} \prod_{i\in A_1} \mathbf{w}_{i1}(t) \cdots \prod_{j \in A_X} \mathbf{w}_{jX}(t),$$

$V_t(\mathbf{\Omega}(t))$ is the value function corresponding to the maximal expected reward from time slot t to T ($1 \leq t \leq T$), $\mathbf{\Omega}(t+1)$ follows the evolution in (4) given that the channels in the subset A_x ($x \in \mathcal{X}$) are observed in state x. In particular, the term $F(\mathcal{A}(t), \mathbf{\Omega}(t))$ corresponds to the expected accumulated discounted reward starting from slot $t+1$ to T, calculated over all possible realizations of the selected channels (i.e., the channels in $\mathcal{A}(t)$).

C. Myopic Policy

Theoretically, the optimal policy can be obtained by solving the dynamic programming (5). It is infeasible, however, due to the tight coupling between the current action and the future reward, and in fact obtaining the optimal solution directly from the recursive equations (5) is computationally prohibitive. Henceforce, a natural alternative is to seek a simple myopic policy maximizing the immediate reward while ignoring the impact of the current action on the future reward, which is easy to compute and implement, formally defined as follows:

$$\hat{\mathcal{A}}(t) = \operatorname*{argmax}_{\mathcal{A}(t)} \mathbb{E}\Big[\sum_{n\in\mathcal{A}(t)} \mathbf{w}_n \mathbf{r}'\Big]. \tag{6}$$

For the purpose of tractable analysis, we introduce some partial orders used in the following sections.

Definition 1 (first order stochastic dominance, [20])**.** Let $\Pi(X) \triangleq \{(w_1, \cdots, w_X) : \sum_{i=1}^{X} w_i = 1, w_1, \cdots, w_X \geq 0\}$. For $\mathbf{w}_1$, $\mathbf{w}_2 \in \Pi(X)$, then $\mathbf{w}_1$ first order stochastically dominates $\mathbf{w}_2$—denoted as $\mathbf{w}_1 \geq_s \mathbf{w}_2$, if $\sum_{i=j}^{X} \mathbf{w}_{1i} \geq \sum_{i=j}^{X} \mathbf{w}_{2i}$ for $j = 1, 2, \cdots, X$.

Definition 2 (first order stochastic dominance matrix)**.** Let $\mathbf{w}_1, \cdots, \mathbf{w}_X \in \Pi(X)$ be any X belief vectors. Then the matrix $Q = [\mathbf{w}_1 \ \cdots \ \mathbf{w}_X]'$ is a first order stochastic dominance matrix if $\mathbf{w}_1 \leq_s \mathbf{w}_2 \leq_s \cdots \leq_s \mathbf{w}_X$.

Based on the first order stochastic dominance, we have the special structure of the myopic policy of (6), stated in the following.

Definition 3 (Myopic Policy)**.** The myopic policy $\hat{\rho} := (\hat{\rho}_0, \hat{\rho}_1, \cdots, \hat{\rho}_T)$ is the policy that selects the best M channels (in the sense of first order stochastic dominance order) at each time. That is, if $\mathbf{w}_{\sigma_1}(t) \geq_s \cdots \geq_s \mathbf{w}_{\sigma_N}(t)$, then the myopic policy at t is

$$\hat{\mathcal{A}}(t) = \hat{\rho}_t(\mathbf{\Omega}(t)) = \{\sigma_1, \cdots, \sigma_M\}.$$

III. Analysis on Optimality of Myopic Policy

To analyze the performance of the myopic policy conveniently, we first introduce an auxiliary value function [18] and then prove a critical feature of the auxiliary value function. Next, we give a simple assumption about transition matrix, and show its special stochastic order based on the assumption. Finally, by deriving the bounds of different policies, we get some important bounds, which serves as a basis to prove the optimality of the myopic policy.

A. Value Function and its Properties

First, we define the auxiliary value function (AVF) as follows:

$$\begin{cases} W_T^{\hat{\mathcal{A}}}(\mathbf{\Omega}(T)) = \sum_{n \in \hat{\mathcal{A}}(T)} \mathbf{w}_n(T)\mathbf{r}', \\ W_\tau^{\hat{\mathcal{A}}}(\mathbf{\Omega}(\tau)) = \sum_{n \in \hat{\mathcal{A}}(\tau)} \mathbf{w}_n(\tau)\mathbf{r}' \\ \qquad + \beta \underbrace{\Sigma(\mathcal{A}(\tau), \mathbf{\Omega}(\tau)) W_{\tau+1}^{\hat{\mathcal{A}}}(\mathbf{\Omega}(\tau))}_{F(\hat{\mathcal{A}}(\tau), \mathbf{\Omega}(\tau))}, \quad t+1 \leq \tau \leq T \\ W_t^{\mathcal{A}}(\mathbf{\Omega}(t)) = \sum_{n \in \mathcal{A}(t)} \mathbf{w}_n(t)\mathbf{r}' \\ \qquad + \beta \underbrace{\Sigma(\mathcal{A}(t), \mathbf{\Omega}(t)) W_{t+1}^{\hat{\mathcal{A}}}(\mathbf{\Omega}(t+1))}_{F(\mathcal{A}(t), \mathbf{\Omega}(t))}, \end{cases} \tag{7}$$

Remark. (1) AVF characterizes the expected discounted accumulated reward of the following special policy: at slot t, the first M channels in $\mathcal{A}(t)$ are accessed, and then the channels in $\hat{\mathcal{A}}(r)$ $(t+1 \leq r \leq T)$ are accessed; that is, the special policy $(\rho_t, \hat{\rho}_{t+1}, \cdots, \hat{\rho}_T)$ is adopted from slot t to T.

(2) If $\mathcal{A}(t) = \hat{\mathcal{A}}(t)$ (i.e., $\rho_t = \hat{\rho}_t$), then $W_t^{\mathcal{A}}(\mathbf{\Omega}(t)) = W_t^{\hat{\mathcal{A}}}(\mathbf{\Omega}(t))$ is the total reward from slot t to T under the myopic policy $\hat{\rho}$.

Lemma 1 (Decomposability)**.** $W_t^{\mathcal{A}}(\mathbf{\Omega}(t))$ is decomposable for all $t = 1, 2, \cdots, T$, i.e.,

$$\begin{aligned} & W_t^{\mathcal{A}}(\mathbf{w}_1, \cdots, \mathbf{w}_i, \cdots, \mathbf{w}_N) \\ & = \sum_{j=1}^{X} \omega_{ij} W_t^{\mathcal{A}}(\mathbf{w}_1, \cdots, \mathbf{e}_j, \cdots, \mathbf{w}_N), \end{aligned}$$

where, $\mathbf{e}_i = [\underbrace{0, \cdots, 0}_{i-1}, 1, \underbrace{0, \cdots, 0}_{X-i}]$.

B. Structural Properties of Matrix of Transition Probabilities

In this section, we give an assumption on the matrix of transition probabilities, and then points out some important properties of the matrix which serve as a basis of deriving the optimality of the myopic policy.

Proposition 1. Suppose that transition matrix $\mathbf{P}$ has X eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_X$ and the corresponding eigenvectors are $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_X$, then we have

1) $\lambda_1 = 1$ and $\mathbf{v}_1 = \mathbf{I}_X := [\underbrace{1 \cdots 1}_{X}]$;
2) If $\mathbf{w}_m, \mathbf{w}_n \in \Pi(X)$ and $\mathbf{w}_m \geq_s \mathbf{w}_n$, then the following holds for any λ

$$\begin{aligned} 0 = & (\mathbf{w}_m - \mathbf{w}_n)(\mathbf{v}_1 \ \cdots \ \mathbf{v}_X)' \\ & \left[\begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_4 \end{pmatrix} - \begin{pmatrix} \lambda & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_4 \end{pmatrix} \right] \end{aligned} \tag{8}$$

Assumption 1. Assume that

1) $\lambda_2^{(n)}(t) = \cdots = \lambda_X^{(n)}(t) \triangleq \lambda^{(n)}(t) > 0$ for $\mathbf{P}^{(n)}(t)$ $(n \in \mathcal{N}, t \geq 1)$.
2) At any t, $P_X^{(\varsigma_i^t)}(t) \leq P_1^{(\varsigma_{i+1}^t)}(t)$ $(i = 1, \cdots, N-1)$, where $\varsigma_1^t, \cdots, \varsigma_n^t$ is one permutation of $\{1, 2, \cdots, N\}$ at slot t.

Remark. The first part of Assumption 1 states the special structure of transmission matrix, i.e., having the eigenvalue $\lambda^{(n)}(t)$ with $X-1$ times, while the second part guarantees monotonic structure in the sense of stochastic order in terms of $\mathbf{w}_1(t), \cdots, \mathbf{w}_N(t)$ at any slot t; that is, the information states of all channels can be ordered stochastically at all slots.

Proposition 2. Under Assumption 1, $\mathbf{P}^{(n)}(t)$ is a first order stochastic dominance matrix.

Proposition 3. Under Assumption 1, at any slot t, $\{\mathbf{w}_1(t), \cdots, \mathbf{w}_N(t)\}$ can be ordered in the sense of first order stochastic order; that is, $\mathbf{w}_{\varsigma_1^t}(t) \leq_s \mathbf{w}_{\varsigma_2^t}(t) \leq_s \cdots \leq_s \mathbf{w}_{\varsigma_N^t}(t)$, where, $\{\varsigma_1^t, \varsigma_2^t, \cdots, \varsigma_N^t\}$ is a permutation of $\{1, 2, \cdots, N\}$ at slot t.

Proof: By (4), we have $P_1^{(\varsigma_i^t)}(t) \leq_s \mathbf{w}_{\varsigma_i^t}(t+1) \leq_s P_X^{(\varsigma_i^t)}(t)$ $(i = 1, \cdots, N)$. Combining with Assumption 1, then $\mathbf{w}_{\varsigma_1^t}(t+1) \leq_s \mathbf{w}_{\varsigma_2^t}(t+1) \leq_s \cdots \leq_s \mathbf{w}_{\varsigma_N^t}(t+1)$. ■

C. Optimality of Myopic Policy

Here, we derive some important bounds in the following Lemma 2 and then establish the sufficient condition, based on these bounds, to guarantee the optimality of the myopic policy. Specifically, in Lemma 2, we consider two belief vectors $\mathbf{\Omega}_l = (\mathbf{\Omega}_{-l}, \mathbf{w}_l)$ and $\mathbf{\Omega}_l' = (\mathbf{\Omega}_{-l}, \tilde{\mathbf{w}}_l)$ that differ only in one element, i.e., $\mathbf{w}_l \leq_s \tilde{\mathbf{w}}_l$ and gives the lower bound and the upper bound on $W_t^{\mathcal{A}'}(\mathbf{\Omega}_l') - W_t^{\mathcal{A}}(\mathbf{\Omega}_l)$.

Lemma 2. Under Assumption 1, $\overline{\lambda} \triangleq \max\{\lambda^{(i)}(t) : i \in \mathcal{N}, 1 \leq t \leq T\}$, $\mathbf{\Omega}_l \triangleq (\mathbf{\Omega}_{-l}, \mathbf{w}_l)$, $\mathbf{\Omega}'_l \triangleq (\mathbf{\Omega}_{-l}, \tilde{\mathbf{w}}_l)$, $\mathbf{w}_l \leq_s \tilde{\mathbf{w}}_l$, we have for $1 \leq t \leq T$

(A1): if $\mathcal{A}' = \mathcal{A}$, $l \in \mathcal{A}'$ and $l \in \mathcal{A}$,

$$(\tilde{\mathbf{w}}_l - \mathbf{w}_l)\mathbf{r}' \leq W_t^{\mathcal{A}'}(\mathbf{\Omega}'_l) - W_t^{\mathcal{A}}(\mathbf{\Omega}_l) \leq \sum_{i=0}^{T-t}(\beta\overline{\lambda})^i(\tilde{\mathbf{w}}_l - \mathbf{w}_l)\mathbf{r}';$$

(A2): if $\mathcal{A}' = \mathcal{A}$, $l \notin \mathcal{A}'$ and $l \notin \mathcal{A}$,

$$0 \leq W_t^{\mathcal{A}'}(\mathbf{\Omega}'_l) - W_t^{\mathcal{A}}(\mathbf{\Omega}_l) \leq \sum_{i=1}^{T-t}(\beta\overline{\lambda})^i(\tilde{\mathbf{w}}_l - \mathbf{w}_l)\mathbf{r}';$$

(A3): if $\mathcal{A}' \setminus \{l\} \subset \mathcal{A}$, $l \in \mathcal{A}'$ and $l \notin \mathcal{A}$,

$$0 \leq W_t^{\mathcal{A}'}(\mathbf{\Omega}'_l) - W_t^{\mathcal{A}}(\mathbf{\Omega}_l) \leq \sum_{i=0}^{T-t}(\beta\overline{\lambda})^i(\tilde{\mathbf{w}}_l - \mathbf{w}_l)\mathbf{r}'.$$

Proof: The proof is given in Appendix A. ■

Given $\mathbf{\Omega}$, in the following lemma, we consider two policies $\mathcal{A}_l$ and $\mathcal{A}_m$ which differ in one element; that is, $l \in \mathcal{A}_l$, $m \in \mathcal{A}_m$, $\mathcal{A}_l \setminus \{l\} = \mathcal{A}_m \setminus \{m\}$, and $\mathbf{w}_l >_s \mathbf{w}_m$, and establish sufficient condition such that $W_t^{\mathcal{A}_l}(\mathbf{\Omega}) > W_t^{\mathcal{A}_m}(\mathbf{\Omega})$.

Lemma 3. Under Assumption 1, given $m \in \mathcal{A}_m$, $l \in \mathcal{A}_l$, $\mathbf{w}_l >_s \mathbf{w}_m$, and $\mathcal{A}_l \setminus \{l\} = \mathcal{A}_m \setminus \{m\}$, if $\sum_{i=1}^{T-t}(\beta\overline{\lambda})^i \leq 1$, then $W_t^{\mathcal{A}_l}(\mathbf{\Omega}) > W_t^{\mathcal{A}_m}(\mathbf{\Omega})$.

Proof: Let $\mathbf{\Omega}'$ denote the set of channel belief vectors with $\tilde{\mathbf{w}}_l = \mathbf{w}_m$ and $\mathbf{w}'_i = \mathbf{w}_i$ for $\forall i \neq l$ and $i \in \mathcal{N}$, then $W_t^{\mathcal{A}_l}(\mathbf{\Omega}') = W_t^{\mathcal{A}_m}(\mathbf{\Omega}')$. By Lemma 2, we have

$$\begin{aligned}
&W_t^{\mathcal{A}_l}(\mathbf{\Omega}) - W_t^{\mathcal{A}_m}(\mathbf{\Omega}) \\
=&[W_t^{\mathcal{A}_l}(\mathbf{\Omega}) - W_t^{\mathcal{A}_l}(\mathbf{\Omega}')] - [W_t^{\mathcal{A}_m}(\mathbf{\Omega}) - W_t^{\mathcal{A}_m}(\mathbf{\Omega}')] \\
\geq&(\mathbf{w}_l - \mathbf{w}_m)\mathbf{r}' - \sum_{i=1}^{T-t}(\beta\overline{\lambda})^i(\mathbf{w}_l - \mathbf{w}_m)\mathbf{r}' \\
=&(\mathbf{w}_l - \mathbf{w}_m)\mathbf{r}'\Big(1 - \sum_{i=1}^{T-t}(\beta\overline{\lambda})^i\Big) \geq 0.
\end{aligned}$$

■

Now, the main optimal theory about the myopic policy is stated in the following.

Theorem 1. Under Assumption 1, the myopic policy is optimal if $\sum_{i=1}^{T-1}(\beta\overline{\lambda})^i \leq 1$ specifically, if $T \to \infty$, $\beta\overline{\lambda} \leq \frac{1}{2}$.

Proof: When $T \nrightarrow \infty$, we prove the theorem by backward induction. The theorem holds trivially for T. Assume that it holds for $T-1, \cdots, t+1$, i.e., the optimal accessing policy is to access the best channels (in the sense of stochastic dominance in terms of available probability vector) from time slot $t+1$ to T. We now show that it holds for t.

Suppose, by contradiction, that given $\mathbf{\Omega} \triangleq \{\mathbf{w}_{i_1}, \cdots, \mathbf{w}_{i_N}\}$ and $\mathbf{w}_1 >_s \mathbf{w}_2 >_s \cdots >_s \mathbf{w}_N$, the optimal policy is to access the best channels from time slot $t+1$ to T, and thus, at slot t, to access channels $\mathcal{A}(t) = \{i_1, \cdots, i_M\} \neq \hat{\mathcal{A}}(t) = \{1, \cdots, M\}$, given that the latter, $\hat{\mathcal{A}}(t)$, includes the best M channels at slot t. There must exist i_m and i_l at slot t such that $m \leq M < l$ and $\mathbf{w}_{i_m} < \mathbf{w}_{i_M} \leq \mathbf{w}_{i_l}$. It then follows from Lemma 3 that $W_t^{\{i_1,\cdots,i_M\}}(\mathbf{\Omega}) < W_t^{\{i_1,\cdots,i_{m-1},i_l,i_{m+1},\cdots,i_M\}}(\mathbf{\Omega})$, which contradicts with the assumption that the latter is the optimal policy. This contradiction completes our proof.

When $T \to \infty$, the proof follows straightforwardly by noticing that $\sum_{i=1}^{\infty} q^i = q/(1-q)$ for any $q \in (0,1)$. ■

Corollary 1. When $X = 2$ and $\mathbf{P}^{(n)}(t) = \mathbf{P}^{(n)}$ for any t, if $0 \leq p_{22}^{(n)} - p_{12}^{(n)} \leq \frac{1}{2}$, then the myopic policy is optimal.

Proof: Given $X = 2$, Assumption 1.1 is satisfied automatically. Meanwhile, Assumption 1.2 is not necessary since in this case the stochastic order (one kind of partial order) structure of belief vector is degenerated into the total order structure. In this case, $0 \leq \lambda^{(n)} = p_{22}^{(n)} - p_{12}^{(n)} \leq \frac{1}{2}$, and then the myopic policy is optimal by Theorem 1. ■

IV. Conclusion

In this paper, we have investigated the scheduling problem of multi-state channels arising in opportunistic communications. Generally, the problem can be formulated as a partially observable Markov decision process or restless multi-armed bandit, which is proved to be PSPACE-hard. In this paper, for heterogeneous i.i.d. multi-state channels, we have derived a set of closed form conditions to guarantee the optimality of the myopic policy (choosing the best channels) in the sense of stochastic dominance order. Specifically, the obtained conditions only depend on discount factor and the eigenvalues of all probability transmission matrices.

V. Acknowledgment

This work is supported in part by National NSF of China (61303027), NSF of Hubei Province (2015CFB585), China Postdoctoral Science Foundation (2013M531753, 2014T70748),

Appendix A
Proof of Lemma 2

We prove the lemma by backward induction. For slot T, we have

1) For $l \in \mathcal{A}'$, $l \in \mathcal{A}$, it holds that $W_T^{\mathcal{A}'}(\mathbf{\Omega}'_l) - W_T^{\mathcal{A}}(\mathbf{\Omega}_l) = \mathbf{r}(\tilde{\mathbf{w}}_l - \mathbf{w}_l)$;
2) For $l \notin \mathcal{A}'$, $l \notin \mathcal{A}$, it holds that $W_T^{\mathcal{A}'}(\mathbf{\Omega}'_l) - W_T^{\mathcal{A}}(\mathbf{\Omega}_l) = 0$;
3) For $l \in \mathcal{A}'$, $l \notin \mathcal{A}$, it exists at least one channel m such that $\mathbf{w}'_l \geq \mathbf{w}_m \geq \mathbf{w}_l$. It then holds that $0 \leq W_T^{\mathcal{A}'}(\mathbf{\Omega}'_l) - W_T^{\mathcal{A}}(\mathbf{\Omega}_l) = (\tilde{\mathbf{w}}_l - \mathbf{w}_m)\mathbf{r}' \leq (\tilde{\mathbf{w}}_l - \mathbf{w}_l)\mathbf{r}'$.

Therefore, Lemma 2 holds for slot T.

Assume that Lemma 2 holds for $T-1, \cdots, t+1$, then we prove the lemma for slot t.

We first prove the first case: $l \in \mathcal{A}'$, $l \in \mathcal{A}$. By developing $\mathbf{w}_l(t+1)$ in $\Omega(t+1)$ according to Lemma 1, we have:

$$F(\mathcal{A}', \mathbf{\Omega}'_l) = \Sigma(\mathcal{A}' \setminus \{l\}, \mathbf{\Omega}'_l)\Big[\sum_{j \in \mathcal{X}} \tilde{\omega}_{lj}(t) W_{t+1}(\mathbf{\Omega}_{-l}, \mathbf{e}_j \mathbf{P}^{(l)}(t))\Big],$$

$$F(\mathcal{A}, \mathbf{\Omega}_l) = \Sigma(\mathcal{A} \setminus \{l\}, \mathbf{\Omega}_l)\Big[\sum_{j \in \mathcal{X}} \omega_{lj}(t) W_{t+1}(\mathbf{\Omega}_{-l}, \mathbf{e}_j \mathbf{P}^{(l)}(t))\Big].$$

Furthermore, we have considering $\Sigma(\mathcal{A}' \setminus \{l\}, \mathbf{\Omega}'_l) = \Sigma(\mathcal{A} \setminus \{l\}, \mathbf{\Omega}_l)$

$$\begin{aligned}
&\digamma(\mathcal{A}', \mathbf{\Omega}'_l) - \digamma(\mathcal{A}, \mathbf{\Omega}_l) \\
\overset{(a)}{=}&\Sigma(\mathcal{A} \setminus \{l\}, \mathbf{\Omega}_l) \sum_{j\in\mathcal{X}-\{1\}} \Big[(\tilde{\omega}_{lj}(t) - \omega_{lj}(t)) \\
&\Big(W_{t+1}(\mathbf{\Omega}_{-l}, \mathbf{e}_j\mathbf{P}^{(l)}(t)) - W_{t+1}(\mathbf{\Omega}_{-l}, \mathbf{e}_1\mathbf{P}^{(l)}(t))\Big)\Big] \\
=&\Sigma(\mathcal{A} \setminus \{l\}, \mathbf{\Omega}_l) \sum_{j\in\mathcal{X}-\{1\}} \Big[\sum_{i=j}^{X}(\tilde{\omega}_{li}(t) - \omega_{li}(t)) \\
&\Big(W_{t+1}(\mathbf{\Omega}_{-l}, \mathbf{e}_j\mathbf{P}^{(l)}(t)) - W_{t+1}(\mathbf{\Omega}_{-l}, \mathbf{e}_{j-1}\mathbf{P}^{(l)}(t))\Big) \\
&+ \omega_{li}(t)\Big(W_{t+1}(\mathbf{\Omega}_{-l}, \mathbf{e}_j\mathbf{P}^{(l)}(t)) - W_{t+1}(\mathbf{\Omega}_{-l}, \mathbf{e}_1\mathbf{P}^{(l)}(t))\Big)\Big]
\end{aligned} \tag{9}$$

where, the equality (a) is due to $\omega_{l_1}(t) = 1 - \sum_{j\in\mathcal{X}-\{1\}} \omega_{l_j}(t) = 1 - \sum_{j=2}^{X} \omega_{l_j}(t)$.

Next, we analyze the term in the bracket, $W_{t+1}(\mathbf{\Omega}_{-l}, \mathbf{e}_j\mathbf{P}^{(l)}(t)) - W_{t+1}(\mathbf{\Omega}_{-l}, \mathbf{e}_1\mathbf{P}^{(l)}(t))$, of RHS of (9) through three cases:

Case 1: if $l \in \mathcal{A}'$, $l \in \mathcal{A}$, according to the induction hypothesis, we have

$$\begin{aligned}
0 &\le W_{t+1}(\mathbf{\Omega}_{-l}, \mathbf{e}_j\mathbf{P}^{(l)}(t)) - W_{t+1}(\mathbf{\Omega}_{-l}, \mathbf{e}_1\mathbf{P}^{(l)}(t)) \\
&\le \sum_{i=0}^{T-t-1} (\beta\overline{\lambda})^i(\mathbf{e}_j - \mathbf{e}_1)\mathbf{P}^{(l)}(t)\mathbf{r}'.
\end{aligned} \tag{10}$$

Case 2: if $l \notin \mathcal{A}'$, $l \notin \mathcal{A}$, according to the induction hypothesis, we have

$$\begin{aligned}
0 &\le W_{t+1}(\mathbf{\Omega}_{-l}, \mathbf{e}_j\mathbf{P}^{(l)}(t)) - W_{t+1}(\mathbf{\Omega}_{-l}, \mathbf{e}_1\mathbf{P}^{(l)}(t)) \\
&\le \sum_{i=1}^{T-t-1} (\beta\overline{\lambda})^i(\mathbf{e}_j - \mathbf{e}_1)\mathbf{P}^{(l)}(t)\mathbf{r}'.
\end{aligned} \tag{11}$$

Case 3: if $l \in \mathcal{A}'$, $l \notin \mathcal{A}$, according to the induction hypothesis, we have

$$\begin{aligned}
0 &\le W_{t+1}(\mathbf{\Omega}_{-l}, \mathbf{e}_j\mathbf{P}^{(l)}(t)) - W_{t+1}(\mathbf{\Omega}_{-l}, \mathbf{e}_1\mathbf{P}^{(l)}(t)) \\
&\le \sum_{i=0}^{T-t-1} (\beta\overline{\lambda})^i(\mathbf{e}_j - \mathbf{e}_1)\mathbf{P}^{(l)}(t)\mathbf{r}'.
\end{aligned} \tag{12}$$

Combining Case 1–3, we obtain the following:

$$\begin{aligned}
0 &\le W_{t+1}(\mathbf{\Omega}_{-l}, \mathbf{e}_j\mathbf{P}^{(l)}(t)) - W_{t+1}(\mathbf{\Omega}_{-l}, \mathbf{e}_1\mathbf{P}^{(l)}(t)) \\
&\le \sum_{i=0}^{T-t-1} (\beta\overline{\lambda})^i(\mathbf{e}_j - \mathbf{e}_1)\mathbf{P}^{(l)}(t)\mathbf{r}'.
\end{aligned} \tag{13}$$

Therefore, combining (9) and (13), we have

$$\begin{aligned}
0 \le& W_t^{\mathcal{A}'}(\mathbf{\Omega}'_l) - W_t^{\mathcal{A}}(\mathbf{\Omega}_l) \\
=&(\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t))\mathbf{r}' + \beta(\digamma(\mathcal{A}', \mathbf{\Omega}'_l) - \digamma(\mathcal{A}, \mathbf{\Omega}_l)) \\
\overset{(a)}{\le}&(\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t))\mathbf{r}' + \beta\Big[\sum_{i=0}^{T-t-1} (\beta\overline{\lambda})^i(\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t))(\overline{\lambda}\mathbf{E})\mathbf{r}'\Big] \\
=&(\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t))\mathbf{r}' + \sum_{i=1}^{T-t}(\beta\overline{\lambda})^i(\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t))\mathbf{r}' \\
=&\sum_{i=0}^{T-t}(\beta\overline{\lambda})^i(\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t))\mathbf{r}',
\end{aligned}$$

where, $\mathbf{E} := [e_1, \cdots, e_X]'$, and the inequality (a) is due to Assumption 1 and Proposition 1.

To the end, we complete the proof of the first part, $l \in \mathcal{A}'$, $l \in \mathcal{A}$, of Lemma 2.

Secondly, **we prove the second case** $l \notin \mathcal{A}'$, $l \notin \mathcal{A}$, which implies that in this case, $\mathcal{A}'(t) = \mathcal{A}(t)$. Assuming $k \in \mathcal{A}(t)$, we have:

$$\digamma(\mathcal{A}', \mathbf{\Omega}'_l) =\Sigma(\mathcal{A}(t) \setminus \{k\}, \mathbf{\Omega}'_l)\Big[\sum_{j\in\mathcal{X}} \omega_{kj}(t)W_{t+1}(\mathbf{\Omega}'_{-k}, \mathbf{e}_j\mathbf{P}^{(k)}(t))\Big],$$

$$\digamma(\mathcal{A}, \mathbf{\Omega}_l) =\Sigma(\mathcal{A}(t) \setminus \{k\}, \mathbf{\Omega}_l)\Big[\sum_{j\in\mathcal{X}} \omega_{kj}(t)W_{t+1}(\mathbf{\Omega}_{-k}, \mathbf{e}_j\mathbf{P}^{(k)}(t))\Big].$$

Thus, considering $\Sigma(\mathcal{A}(t)\setminus\{k\}, \mathbf{\Omega}_l) = \Sigma(\mathcal{A}(t)\setminus\{k\}, \mathbf{\Omega}'_l)$, we have

$$\begin{aligned}
&\digamma(\mathcal{A}', \mathbf{\Omega}'_l) - \digamma(\mathcal{A}, \mathbf{\Omega}_l) \\
&= \Sigma(\mathcal{A}(t) \setminus \{k\}, \mathbf{\Omega}_l)\Big[\sum_{j\in\mathcal{X}} \omega_{kj}(t) \\
&\Big(W_{t+1}(\mathbf{\Omega}'_{-k}, \mathbf{e}_j\mathbf{P}^{(k)}(t)) - W_{t+1}(\mathbf{\Omega}_{-k}, \mathbf{e}_j\mathbf{P}^{(k)}(t))\Big)\Big].
\end{aligned} \tag{14}$$

For the term in the bracket of RHS of (14), if channel l is never chosen for $W_{t+1}(\mathbf{\Omega}'_{-k}, \mathbf{e}_j\mathbf{P}^{(k)}(t))$ and $W_{t+1}(\mathbf{\Omega}_{-k}, \mathbf{e}_j\mathbf{P}^{(k)}(t))$ from the slot $t+1$ to the end of time horizon of interest T. That is to say, $l \notin \mathcal{A}'(r)$ and $l \notin \mathcal{A}(r)$ for $t+1 \le r \le T$, and further, we have $W_{t+1}(\mathbf{\Omega}'_{-k}, \mathbf{e}_j\mathbf{P}^{(k)}(t)) - W_{t+1}(\mathbf{\Omega}_{-k}, \mathbf{e}_j\mathbf{P}^{(k)}(t)) = 0$; otherwise, it exists t^o ($t+1 \le t^o \le T$) such that one of the following three cases holds.

Case 1: $l \notin \mathcal{A}'(r)$ and $l \notin \mathcal{A}(r)$ for $t \le r \le t^0 - 1$ while $l \in \mathcal{A}'(t^0)$ and $l \in \mathcal{A}(t^0)$;

Case 2: $l \notin \mathcal{A}'(r)$ and $l \notin \mathcal{A}(r)$ for $t \le r \le t^0 - 1$ while $l \notin \mathcal{A}'(t^0)$ and $l \in \mathcal{A}(t^0)$ (Note that this case does not exist according to the first order stochastic dominance of transition matrix $\mathbf{P}^{(i)}(t)$);

Case 3: $l \notin \mathcal{A}'(r)$ and $l \notin \mathcal{A}(r)$ for $t \le r \le t^0 - 1$ while $l \in \mathcal{A}'(t^0)$ and $l \notin \mathcal{A}(t^0)$.

For Case 1, according to the hypothesis ($l \in \mathcal{A}'$ and $l \in \mathcal{A}$), we have

$$W_{t^0}(\mathbf{\Omega}'_l(t^o)) - W_{t^0}(\mathbf{\Omega}_l(t^o)) \le \overline{\lambda} \sum_{i=0}^{T-t-1} (\beta\overline{\lambda})^i(\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t)))\mathbf{r}',$$

where, the inequality is due to Assumption 1 and Proposition 1.

For Case 2–3, by the induction hypothesis ($l \in \mathcal{A}'$, $l \notin \mathcal{A}$ or $l \in \mathcal{A}$, $l \notin \mathcal{A}'$), we have the similar results with Case 1.

Combing the results of the three cases, we obtain

$$\begin{aligned}
&W_{t+1}(\mathbf{\Omega}'_{-k}, \mathbf{e}_j\mathbf{P}^{(k)}(t)) - W_{t+1}(\mathbf{\Omega}_{-k}, \mathbf{e}_j\mathbf{P}^{(k)}(t)) \\
&\le \overline{\lambda} \sum_{i=0}^{T-t-1} (\beta\overline{\lambda})^i(\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t)))\mathbf{r}'.
\end{aligned} \tag{15}$$

Combing (15) and (14), we have

$$W_t^{\mathcal{A}'}(\mathbf{\Omega}'_l) - W_t^{\mathcal{A}}(\mathbf{\Omega}_l) = \beta(\digamma(\mathcal{A}', \mathbf{\Omega}'_l) - \digamma(\mathcal{A}, \mathbf{\Omega}_l))$$

$$\leq \beta\overline{\lambda} \sum_{i=0}^{T-t-1} (\beta\overline{\lambda})^i (\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t)))\mathbf{r}'$$
$$\leq \sum_{i=1}^{T-t} (\beta\overline{\lambda})^i (\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t)))\mathbf{r}',$$

which completes the proof of Lemma 2 when $l \notin \mathcal{A}'$ and $l \notin \mathcal{A}$.

Last, **we prove the third case** $l \in \mathcal{A}'(t)$ **and** $l \notin \mathcal{A}(t)$, then it exists at least one channel, and its belief vector denoted as $\mathbf{w}_m$, such that $\mathbf{w}'_l \geq_s \mathbf{w}_m \geq_s \mathbf{w}_l$. We have

$$\begin{aligned} & W_t^{\mathcal{A}'}(\mathbf{\Omega}'_l(t)) - W_t^{\mathcal{A}}(\mathbf{\Omega}_l(t)) \qquad (16) \\ =& W_t^{\mathcal{A}'}(\mathbf{w}_1, \cdots, \tilde{\mathbf{w}}_l, \cdots, \mathbf{w}_N) - W_t^{\mathcal{A}}(\mathbf{w}_1, \cdots, \mathbf{w}_l, \cdots, \mathbf{w}_N) \\ =& W_t^{\mathcal{A}'}(\mathbf{w}_1, \cdots, \tilde{\mathbf{w}}_l, \cdots, \mathbf{w}_N) - W_t^{\mathcal{A}'}(\mathbf{w}_1, \cdots, \mathbf{w}_l = \mathbf{w}_m, \cdots, \mathbf{w}_N) \\ +& W_t^{\mathcal{A}}(\mathbf{w}_1, \cdots, \mathbf{w}_l = \mathbf{w}_m, \cdots, \mathbf{w}_N) - W_t^{\mathcal{A}}(\mathbf{w}_1, \cdots, \mathbf{w}_l, \cdots, \mathbf{w}_N) \end{aligned}$$

According to the induction hypothesis ($l \in \mathcal{A}'$ and $l \in \mathcal{A}$), the first term of the RHS of (16) can be bounded as follows:

$$\begin{aligned} 0 \leq & W_t^{\mathcal{A}'}(\mathbf{w}_1, \cdots, \tilde{\mathbf{w}}_l, \cdots, \mathbf{w}_N) - W_t^{\mathcal{A}'}(\mathbf{w}_1, \cdots, \mathbf{w}_l = \mathbf{w}_m, \cdots, \mathbf{w}_N) \\ \leq & \sum_{i=0}^{T-t} (\beta\overline{\lambda}) i (\tilde{\mathbf{w}}_l(t) - \mathbf{w}_m(t))\mathbf{r}' \qquad (17) \end{aligned}$$

Meanwhile, the second term of the RHS of (16) is inducted by hypothesis ($l \notin \mathcal{A}'$ and $l \notin \mathcal{A}$):

$$\begin{aligned} 0 \leq & W_t^{\mathcal{A}}(\mathbf{w}_1, \cdots, \mathbf{w}_l = \mathbf{w}_m, \cdots, \mathbf{w}_N) - W_t^{\mathcal{A}}(\mathbf{w}_1, \cdots, \mathbf{w}_l, \cdots, \mathbf{w}_N) \\ \leq & \sum_{i=1}^{T-t} (\beta\overline{\lambda})^i (\mathbf{w}_m(t) - \mathbf{w}_l(t))\mathbf{r}' \qquad (18) \end{aligned}$$

Therefore, we have, combining (16), (17) and (18),

$$0 \leq W_t^{\mathcal{A}'}(\mathbf{\Omega}'_l(t)) - W_t^{\mathcal{A}}(\mathbf{\Omega}_l(t)) \leq \sum_{i=0}^{T-t} (\beta\overline{\lambda})^i (\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t))\mathbf{r}'.$$

Thus, we complete the proof of the third part, $l \in \mathcal{A}'(t)$ and $l \notin \mathcal{A}(t)$, of Lemma 2.

To the end, Lemma 2 is concluded.

References

[1] P. Whittle. Restless bandits: activity allocation in a changing world. *Journal of Applied Probability*, (Special Vol. 25A):287–298, 1988.

[2] J. Gittins, K. Glazebrook, and R. Webber. *Multi-Armed Bandit Allocation Indices*. Blackwell, Oxford, U.K., 2011.

[3] Q. Zhao, L. Tong, A. Swami, and Y. Chen. Decentralized cognitive mac for opportunistic spectrum access in ad hoc networks: A pomdp framework. *IEEE JSAC*, 25(3), Apr. 2010.

[4] J. C. Gittins and D.M. Jones. A Dynamic Allocation Index For the Sequential Design of Experiments. *Progress in Statistics*, pages 241–266, 1974.

[5] J. C. Gittins. Bandit Processes and Dynamic Allocation Indices. *Journal of the Royal Statistical Society*, 41(2):148–177, 1979.

[6] C. H. Papadimitriou and J. N. Tsitsiklis. The complexity of optimal queueing network control. *Mathematics of Operations Research*, 24(2):293–305, 1999.

[7] S. Guha and K. Munagala. Approximation algorithms for partial-information based stochastic control with markovian rewards. In *Proc. FOCS*, Providence, RI, Oct. 2007.

[8] S. Guha and K. Munagala. Approximation algorithms for restless bandit problems. In *Proc. ACM-SIAM SODA*, New York, Jan. 2009.

[9] D. Bertsimas and J. E. Nino-Mora. Restless bandits, linear programming relaxations, and a primal-dual heuristic. *Operations Research*, 48(1):80–90, 2000.

[10] Q. Zhao, and B. Krishnamachari, and K. Liu. On myopic sensing for multi-channel opportunistic access: Structure, optimality, and performance. *IEEE Transactions Wireless Communication*, 7(3):5413–5440, Dec. 2008.

[11] S. Ahmand, and M. Liu, and T. Javidi, and Q. zhao and B. Krishnamachari. Optimality of myopic sensing in multichannel opportunistic access. *IEEE Transactions on Information Theory*, 55(9):4040–4050, Sep. 2009.

[12] S. Ahmad and M. Liu. Multi-channel opportunistic access: a case of restless bandits with multiple plays. In *Allerton Conference*, Monticello, Il, Spet.-Oct. 2009.

[13] K. Liu, and Q. Zhao, and B. Krishnamachari. Dynamic multichannel access with imperfect channel state detection. *IEEE Transactions on Signal Processing*, 58(5):2795–2807, May 2010.

[14] K. Wang and L. Chen. On optimality of myopic policy for restless multi-armed bandit problem: An axiomatic approach. *IEEE Transactions on Signal Processing*, 60(1):300–309, 2012.

[15] T. Javidi S. H. Ahmad, M. Liu, Q. Zhao, and B. Krishnamachari. Optimality of myopic sensing in multi-channel opportunistic access. *IEEE Transactions on Information Theory*, 55(9):4040–4050, 2009.

[16] Fabio E. Lapiccirella, Keqin Liu and Zhi Ding. Multi-channel opportunistic access based on primary arq messages overhearing. In *Proceedings of IEEE ICC 2011*, Kyoto, Jun. 2011.

[17] K. Wang, Q. Liu, and Francis C.M. Lau. Multichannel opportunistic access by overhearing primary arq messages. *IEEE Transactions on Vehicular Technology*, 62(7):3486–3492, 2013.

[18] K. Wang, L. Chen, and Q. Liu. Opportunistic Spectrum Access by Exploiting Primary User Feedbacks in Underlay Cognitive Radio Systems: An Optimality Analysis. *IEEE Journal of Selected Topics in Signal Processing*, 7(5):869–882, 2013.

[19] Y. Ouyang and D. Teneketzis. On the optimality of myopic sensing in multi-state channels. *IEEE Transactions on Information Theory*, 60:681–696, 2014.

[20] A. Muller and D. Stoyan. *Comparison Methods for Stochastic Models and Risk*. Wiley, New York, 2002.