On Optimality of Myopic Policy in Multi-Channel Opportunistic Access

Kehao Wang, Lin Chen, and Jihong Yu

***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, and obtains some reward, i.e., throughput, based on the states of the chosen channels. The objective is to design an access policy, i.e., which channels should be accessed at each time instant, such that the expected accumulated discounted reward is maximised over a finite or infinite horizon. The considered problem can be cast into a restless multi-armed bandit (RMAB) problem, which is PSPACE-hard, with the optimal policy usually intractable due to the exponential computation complexity. Hence, a natural alternative is to consider the easily implementable myopic policy that only maximises the immediate reward but ignores the impact of the current strategy on the future reward. In this paper, we perform an analytical study on the performance of the myopic policy for the considered RMAB problem, and establish a set of closed-form conditions to guarantee the optimality of the myopic policy.**

***Index Terms*— Restless bandit, myopic policy, optimality, stochastic order, multivariate analysis.**

I. Introduction

A. Motivation

CONSIDER a communication system composed of N independent channels each of which is modelled as a *time-nonhomogeneous* X-state Markov chain with known probability transition matrices. At each time period a user opportunistically selects M channels to transmit information. A 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 collected over a finite or 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 we know, RMAB problems arise in many areas, such as wireless communication systems, manufacturing systems, economic systems, statistics, biomedical engineering, and information systems etc. [1], [2].

The considered problem can also be formulated as a Partially Observed Markov Decision Process (POMDP) [3], which can be solved by numerical methods for any channel transmission matrix and reward process. However, the numerical approach does not provide any meaningful insight into optimal policy. Moreover, this numerical approach has huge computational complexity. For the two reasons, we study some instances of the generic RMAB in which the optimal policy has a simple structure. Specially, we develop some sufficient conditions to guarantee the optimality of the myopic policy; that is, the optimal policy is to access the best channels each time in the sense of stochastic dominance order.

B. Related Work

In the classic RMAB problem, a player chooses M out of N arms, each evolving as a Markov chain, to activate each time, and receives certain reward determined by the states of the activated arms. The objective is to maximize the long-run reward over an infinite (or finite) horizon by choosing which M arms to activate each time. If only the activated arms change their states, the problem is degenerated to the multi-armed bandit (MAB) problem [4]. The MAB problem is solved by Gittins by showing that the optimal policy has an index structure [4], [5]. However, the RMAB problem is proved to be PSPACE-Hard [6].

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 of which some focus on the case of two-state channel while others on multi-state channels.

For the case of *two-state*, Zhao *et al.* [10] partly obtained the optimality for the case of independently and identically distributed (i.i.d.) channels by analyzing the structure of the

Manuscript received March 30, 2016; revised July 18, 2016 and September 5, 2016; accepted November 6, 2016. Date of publication November 15, 2016; date of current version February 14, 2017. This work is supported in part by National NSF of China (61672395, 61303027, 61603283), the International Science & Technology Cooperation Program of China (Grant No. 2015DFA70340), NSF of Hubei Province (2015CFB585), China Postdoctoral Science Foundation (2013M531753, 2014T70748). This work is presented in ICC 2016. The associate editor coordinating the review of this paper and approving it for publication was W. Saad. *(Corresponding author: Jihong Yu.)*

K. Wang is with the Key Laboratory of Fiber Optic Sensing Technology and Information Processing, School of Information Engineering, Wuhan University of Technology, Hubei 430070, China (e-mail: kehao.wang@whut.edu.cn).

L. Chen and J. Yu are with the Laboratoire de Recherche en Informatique, Department of Computer Science, University of Paris-Sud XI, 91405 Orsay, France (e-mail: chen@lri.fr; jihong.yu@lri.fr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCOMM.2016.2628899

myopic policy. Then Ahmad *et al.* [15] derived the optimality of the myopic sensing policy for the positively correlated i.i.d. channels for accessing one channel each time, and further extended the optimality to access multiple i.i.d. channels [12]. From another point, in [14], we extended [15] which used i.i.d. channels to non i.i.d. ones, focused on a class of so-called *regular* functions, and derived closed-form sufficient conditions to guarantee the optimality of myopic policy. The authors in [16] studied the myopic channel probing policy for the similar scenario proposed, but only established its optimality in the particular case of probing one channel ($M = 1$) each time. In our previous work [17], we established the optimality of myopic policy for the case of probing $M = N - 1$ of N channels at each time instant and analyzed the performance of the myopic probing policy by domination theory. Furthermore, in [18], we studied the generic case of probing arbitrary M channels at each time instant, and derived closed form conditions on the optimality by dropping one of the non-trivial conditions of [16].

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 non-trivial assumptions. In [20], we studied the same model, and showed the optimality of myopic policy in accessing $N-1$ of N channels without the fourth assumption. In [21], we studied a special instance of multi-state case, and proved the optimality of myopic policy under a set of conditions.

C. Contribution of the Paper

Although the multi-state case is also the focus of our work, there exists huge difference between [19], [20] and our work from the viewpoint of channel model. Specifically, on one hand, the channels are modelled to be heterogeneous in our work, which means that probability transition matrix of each channel is different from those of other channels, while in [19] and [20] the probability transition matrices of all channels are identical. On the other hand, the probability transition matrix of each channel is time-nonhomogeneous, i.e., the transition matrix at each time slot is different from those matrices at other slots, while in [19], [20] the probability transition matrix of each channel keeps the same in the whole time horizon. Compared to our previous work [21], the special instance considered in [21] is only one of three instances studied in this work.

The difference in channel model brings about major difficulties in optimising the expected accumulated discount reward in heterogeneous channels from two aspects: i) how to obtain a non-trivial upper performance bound for a pair of special policies depending on multiple different stochastic matrices under multivariate reward (corresponding to multi-state) case; ii) how to determine the stochastic order of belief vectors which characterise the available probabilities of all channels when the transition matrices are non-homogeneous in time horizon.

The two issues are resolved, respectively, by i) assuming that each transmission matrix has a non-trivial eigenvalue with $X-1$ times, under which the first-order stochastic dominance is preserved and meanwhile, the upper performance bound of each pair of special policies involved in transition matrix is characterized by its non-trivial eigenvalue; ii) assuming that there exists a deterministic stochastic dominance order of transmission matrices at any time instance.

In particular, the contributions of this paper include:

- The structure of the myopic policy is shown to be a decreasing-order list determined by the availability probability vectors of all channels provided that certain conditions are satisfied for the probability transition matrices of these channels.
- Multiple set of conditions concerning the structures of probability transition matrices are obtained for different scenarios to guarantee the optimality of the myopic policy.
- The optimisation approach adopted in this work demonstrates the advantage of branch-and-bound and directed comparison.

D. Organization

The rest of the paper is organized as follows: Our model is formulated in Section II. Section III studies the optimality of the myopic policy. Section IV extends the optimality to other cases. Then a numerical evaluation is verified in Section V. Finally, the paper is concluded by Section VI.

Notation: $(\cdot)^{\mathsf{T}}$ and $(\cdot)^{-\mathbf{1}}$ are the transpose and inverse of matrix (or vector). $\mathbf{I}_X$ is the row vector with 1 in all elements. $\mathbf{e}_i$ is the row vector with 1 in the i-th element and 0 in other elements. $\mathbf{E} = [\mathbf{e}_1, \cdots, \mathbf{e}_X]^{\mathsf{T}}$ is an unit matrix.

II. Model and the Optimization Problem

A. System Model

We consider a time-slotted multi-channel communication system consisting of N channels, denoted as $\mathcal{N} = \{1, 2, \cdots, N\}$, and the slot index is t. Each of channel, i.e., n-th channel, is modelled as a *time-nonhomogeneous* X-state Markov chain with probability transition matrix $\mathbf{P}^{(n)}(t)$,

$$\mathbf{P}^{(n)}(t) = \begin{pmatrix} p_{11}^{(n)}(t) & p_{12}^{(n)}(t) & \cdots & p_{1X}^{(n)}(t) \\ p_{21}^{(n)}(t) & p_{22}^{(n)}(t) & \cdots & p_{2X}^{(n)}(t) \\ \vdots & \vdots & \ddots & \vdots \\ p_{X1}^{(n)}(t) & p_{X2}^{(n)}(t) & \cdots & p_{XX}^{(n)}(t) \end{pmatrix} = \begin{pmatrix} \mathbf{P}_1^{(n)}(t) \\ \mathbf{P}_2^{(n)}(t) \\ \vdots \\ \mathbf{P}_X^{(n)}(t) \end{pmatrix},$$

where, $\mathbf{P}_1^{(n)}(t), \cdots, \mathbf{P}_X^{(n)}(t)$ 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, then we have the state vector $\mathbf{S}(t) = [S_1(t), \cdots, S_N(t)]$. Let $\mathcal{A}(t)$ denote the decision made at time t where $\mathcal{A}(t) \subseteq \mathcal{N}$ and $|\mathcal{A}(t)| = M$.

Initially, before any channel selection is made, we assume that we have probabilistic information about the state of each of the N channels, i.e., obtaining the information by observing their states of all channels. Specifically, we assume that at

$t = 0$, the decision-maker knows the probability mass function on the state space of each of the N channels, i.e., the decision-maker knows

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

where,

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

In general,

$$\begin{aligned} \mathcal{A}(0) &= \rho\big(\Omega(0)\big), \\ \mathcal{A}(t) &= \rho\big(\mathcal{O}_{t-1}, \mathcal{A}_{t-1}, \Omega(0)\big), \end{aligned}$$

where, ρ is the mapping to current policy from belief vector, observation history, and decision history,

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

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:

$$R(S_n(t)) = r_x \text{ if } S_n(t) = x, \tag{1}$$

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:

$$\rho^* = \underset{\rho}{\operatorname{argmax}}\, \mathbb{E}^{\rho}\left[\sum_{t=0}^{T} \beta^{t-1} R_{\rho_t}(\Omega(t)) \,\middle|\, \Omega(0)\right], \tag{2}$$

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

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

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

$$\mathbf{w}_n(t+1) = \begin{cases} \mathbf{P}_x^{(n)}(t), & n \in \mathcal{A}(t), O_n(t) = x \\ \mathbf{w}_n(t)\mathbf{P}^{(n)}(t), & n \notin \mathcal{A}(t). \end{cases} \tag{3}$$

To get more insight on the structure of (2), we rewrite it in the language of dynamic programming as follows:

$$\begin{cases} V_T(\Omega(T)) = \max_{\mathcal{A}(T)} \mathbb{E}\Big[\sum_{n\in\mathcal{A}(T)} \mathbf{w}_n(T)\mathbf{r}^{\mathsf{T}}\Big], \\ V_t(\Omega(t)) \quad = \max_{\mathcal{A}(t)} \ \mathbb{E}\Big[\sum_{n\in\mathcal{A}(t)} \mathbf{w}_n(t)\mathbf{r}^{\mathsf{T}} \\ \qquad\qquad + \beta \underbrace{\Sigma(\mathcal{A}(t), \Omega(t)) V_{t+1}(\Omega(t+1))}_{F(\mathcal{A}(t),\Omega(t))}\Big], \end{cases} \tag{4}$$

where,

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

$V_t(\Omega(t))$ is the value function corresponding to the maximal expected reward from time slot t to T ($0 \leq t \leq T$), $\Omega(t+1)$ follows the evolution in (3) given that the channels in the subset A_x ($x \in \mathcal{X}$) are observed in state x. In particular, the term $F(\mathcal{A}(t), \Omega(t))$ 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 (4). It is difficult, 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 (4) 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:

$$\mathcal{A}(t) = \underset{\mathcal{A}(t)}{\operatorname{argmax}}\, \mathbb{E}\Big[\sum_{n\in\mathcal{A}(t)} \mathbf{w}_n\mathbf{r}^{\mathsf{T}}\Big]. \tag{5}$$

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

Definition 1 (first order stochastic dominance, [22]): 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 the following exists for $j = 1, 2, \cdots, X$,

$$\sum_{i=j}^{X} \mathbf{w}_{1i} \geq \sum_{i=j}^{X} \mathbf{w}_{2i}.$$

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 $\mathbf{Q} = [\mathbf{w}_1 \ \cdots \ \mathbf{w}_X]^{\mathsf{T}}$ 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 by (5), 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 $\mathcal{A}(t) = \hat{\rho}_t(\Omega(t)) = \{\sigma_1, \cdots, \sigma_M\}$.

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 comparing different policies, we get some important bounds, which serve 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:

$$\begin{cases} W_T^{\hat{\mathcal{A}}}(\Omega(T)) = \sum_{n\in\hat{\mathcal{A}}(T)} \mathbf{w}_n(T)\mathbf{r}^{\mathsf{T}}, \\ W_\tau^{\hat{\mathcal{A}}}(\Omega(\tau)) = \sum_{n\in\hat{\mathcal{A}}(\tau)} \mathbf{w}_n(\tau)\mathbf{r}^{\mathsf{T}} \\ \qquad + \beta \underbrace{\Sigma(\hat{\mathcal{A}}(\tau), \Omega(\tau)) W_{\tau+1}^{\hat{\mathcal{A}}}(\Omega(\tau))}_{F(\hat{\mathcal{A}}(\tau),\Omega(\tau))}, \\ \qquad\qquad t+1 \leq \tau \leq T \\ W_t^{\mathcal{A}}(\Omega(t)) = \sum_{n\in\mathcal{A}(t)} \mathbf{w}_n(t)\mathbf{r}^{\mathsf{T}} \\ \qquad + \beta \underbrace{\Sigma(\mathcal{A}(t), \Omega(t)) W_{t+1}^{\hat{\mathcal{A}}}(\Omega(t+1))}_{F(\mathcal{A}(t),\Omega(t))}, \end{cases} \tag{6}$$

Remark 1: (1) AVF characterizes the expected accumulated discounted reward of the following special policy: at slot t, the first M channels in $\mathcal{A}(t)$ are accessed, and then the channels in $\hat{\mathcal{A}}(r)$ $(t+1 \leq \tau \leq T)$ 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 $W_t^{\mathcal{A}}(\Omega(t)) = W_t^{\mathcal{Q}}(\Omega(t))$ is the total reward from slot t to T under the myopic policy $\hat{\rho}$.

Lemma 1 (Decomposability): The auxiliary value function $W_t^{\mathcal{A}}(\Omega(t))$ is decomposable for all $t = 1, 2, \cdots, T$, i.e.,

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

Proof: See Appendix A. □

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 $\mathbf{v}_1 = \frac{\mathbf{I}_X}{\sqrt{X}}$;

2) If $\mathbf{w}_m, \mathbf{w}_n \in \Pi(X)$, then the following holds for any λ

$$\begin{aligned} &(\mathbf{w}_m - \mathbf{w}_n)(\mathbf{v}_1 \ \cdots \ \mathbf{v}_X)^{\mathsf{T}} \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_X \end{pmatrix} \\ &= (\mathbf{w}_m - \mathbf{w}_n)(\mathbf{v}_1 \ \cdots \ \mathbf{v}_X)^{\mathsf{T}} \begin{pmatrix} \lambda & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_X \end{pmatrix} \end{aligned} \tag{7}$$

Proof: See Appendix B. □

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 slot 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 2: The first part of Assumption 1 states the special structure of transition matrix, i.e., having the positive 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.

Proof: See Appendix C. □

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: Considering $\mathbf{e}_1 \leq_s \mathbf{w}_{\varsigma_i^t}(t) \leq_s \mathbf{e}_X$, we have $P_1^{(\varsigma_i^t)}(t) \leq_s \mathbf{w}_{\varsigma_i^t}(t+1) \leq_s P_X^{(\varsigma_i^t)}(t)$ $(i = 1, \cdots, N)$ by (3). Combining with Assumption 1, we have $\mathbf{w}_{\varsigma_1^t}(t+1) \leq_s \mathbf{w}_{\varsigma_2^t}(t+1) \leq_s \cdots \leq_s \mathbf{w}_{\varsigma_N^t}(t+1)$. □

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 $\Omega_l = (\Omega_{-l}, \mathbf{w}_l)$ and $\Omega_l' = (\Omega_{-l}, \tilde{\mathbf{w}}_l)$ that differ only in one element, i.e., $\mathbf{w}_l \leq_s \tilde{\mathbf{w}}_l$, and gives the lower bound as well as the upper bound on $W_t^{\mathcal{A}'}(\Omega_l') - W_t^{\mathcal{A}}(\Omega_l)$.

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

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

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

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

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

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

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

Proof: The proof is given in Appendix D. □

Remark 3: (A1) achieves its lower bound when l is chosen at slot t while never chosen after t, and achieves the upper bound when l is chosen from t to T. (A2) achieves its lower bound when l is never chosen from t, and achieves the upper bound when l is not chosen at t but always chosen from $t+1$ to T.

Given Ω, 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 $W_t^{\mathcal{A}_l}(\Omega) > W_t^{\mathcal{A}_m}(\Omega)$.

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 $\sum_{i=1}^{T-t} (\beta\overline{\lambda})^i \le 1$, then $W_t^{\mathcal{A}_l}(\Omega) > W_t^{\mathcal{A}_m}(\Omega)$.

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

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

□

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 $\sum_{i=1}^{T-1} (\beta\overline{\lambda})^i \le 1$ specifically, if $T \to \infty$, $\beta\overline{\lambda} \le \frac{1}{2}$.

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 $\Omega \triangleq \{\mathbf{w}_{i_1}, \cdots, \mathbf{w}_{i_N}\}$ 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 $\mathcal{A}(t) = \{i_1, \cdots, i_M\} \ne \hat{\mathcal{A}}(t) = \{1, \cdots, M\}$, 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 $m \le M < l$ and $\mathbf{w}_{i_m} < \mathbf{w}_{i_M} \le \mathbf{w}_{i_l}$. It then follows from Lemma 3 that $W_t^{\{i_1,\cdots,i_M\}}(\Omega) < W_t^{\{i_1,\cdots,i_{m-1},i_l,i_{m+1},\cdots,i_M\}}(\Omega)$, 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 $0 \le p_{22}^{(n)} - p_{12}^{(n)} \le \frac{1}{2}$, 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, we have

$$\begin{pmatrix} p_{22}^{(n)} & 1-p_{22}^{(n)} \\ p_{12}^{(n)} & 1-p_{12}^{(n)} \end{pmatrix} = \begin{pmatrix} 1 & 1-p_{22}^{(n)} \\ 1 & -p_{12}^{(n)} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & p_{22}^{(n)} - p_{12}^{(n)} \end{pmatrix} \begin{pmatrix} 1 & 1-p_{22}^{(n)} \\ 1 & -p_{12}^{(n)} \end{pmatrix}^{-1}.$$

Therefore, when $0 \le \lambda^{(n)} = p_{22}^{(n)} - p_{12}^{(n)} \le \frac{1}{2}$, the myopic policy is optimal by Theorem 1. □

D. Discussion

In [19], the authors considered the scheduling problem with homogeneous channels, and proved the optimality of myopic policy under four assumptions. In this paper, we study the access problem with heterogeneous channels and their probability transition matrices are non-homogeneous in time slots. Therefore, the problem considered in this paper is more generic from the viewpoint of restless bandit theory, and accordingly, the approach adopted in this paper is different from [19] to a large extent. In fact, due to different assumptions concerning the structure of probability transition matrices, our channel model cannot degenerate to that in [19] even though heterogeneity is neglected in both channels and time slots. One special case is $X = 2$ in which the results in [19] show that the myopic policy is optimal for *homogeneous* channels when $\lambda \ge 0$, while our results show that the myopic policy is optimal for *heterogeneous* channels when $0 \le \lambda^{(n)} \le 0.5$ $(n \in \mathcal{N})$.

IV. Optimality Extension

In this section, we extend the optimality of myopic policy to two cases: 1) each transition matrix has negative eigenvalues, and 2) each matrix has negative or positive eigenvalues, except the trivial eigenvalue '1' stated in Proposition 1.

A. Optimality of Myopic Policy for Transition Matrix With Negative Eigenvalues

Assumption 2: Assume that

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

Proposition 4: Under Assumption 2, $\mathbf{w}_m, \mathbf{w}_l \in \Pi(X)$, and $\mathbf{w}_m \ge_s \mathbf{w}_l$, we have

- $\mathbf{w}_m \prod_{\tau=t}^{t+2i-1} \mathbf{P}^{(n)}(\tau) \ge_s \mathbf{w}_l \prod_{\tau=t}^{t+2i-1} \mathbf{P}^{(n)}(\tau), i = 1, \cdots;$

$$\mathbf{P}[1]=\begin{pmatrix}0.5&0.2&0.3\\0.1&0.6&0.3\\0.1&0.2&0.7\end{pmatrix},\mathbf{P}[2]=\begin{pmatrix}0.6&0.2&0.2\\0.3&0.5&0.2\\0.3&0.2&0.5\end{pmatrix},\mathbf{P}[3]=\begin{pmatrix}0.6&0.2&0.2\\0.5&0.3&0.2\\0.5&0.2&0.3\end{pmatrix},\mathbf{P}[4]=\begin{pmatrix}0.3&0.2&0.5\\0.1&0.4&0.5\\0.1&0.2&0.7\end{pmatrix}$$
$$\mathbf{P}[5]=\begin{pmatrix}0.4&0.2&0.4\\0.5&0.1&0.4\\0.5&0.2&0.3\end{pmatrix},\mathbf{P}[6]=\begin{pmatrix}0.3&0.4&0.3\\0.4&0.3&0.3\\0.4&0.4&0.2\end{pmatrix},\mathbf{P}[7]=\begin{pmatrix}0.5&0.2&0.3\\0.6&0.1&0.3\\0.6&0.2&0.2\end{pmatrix},\mathbf{P}[8]=\begin{pmatrix}0.1&0.4&0.5\\0.3&0.2&0.5\\0.3&0.4&0.3\end{pmatrix}$$

- $\mathbf{w}_m\prod_{\tau=t}^{t+2i}\mathbf{P}^{(n)}(\tau)\leq_s\mathbf{w}_l\prod_{\tau=t}^{t+2i}\mathbf{P}^{(n)}(\tau)$, $i=0,1,\cdots$.

Proof: See Appendix E. □

Based on Assumption 2 and Proposition 4, we have the following lemma.

Lemma 4: Under Assumption 2, $\overline{\lambda}\triangleq\max\{-\lambda^{(i)}(t):\ i\in\mathcal{N},1\leq t\leq T\}$, $\Omega_l=(\Omega_{-l},\mathbf{w}_l)$, $\Omega_l'=(\Omega_{-l},\tilde{\mathbf{w}}_l)$, $\mathbf{w}_l\leq_s\tilde{\mathbf{w}}_l$, and $\sum_{i=1}^{T-t}(\beta\overline{\lambda})^i\leq 1$, we have for $1\leq t\leq T$

- (B1): if $\mathcal{A}'=\mathcal{A}$, $l\in\mathcal{A}'$ and $l\in\mathcal{A}$,

$$(1-\sum_{i=1}^{\lceil\frac{T-t}{2}\rceil}(\beta\overline{\lambda})^{2i-1})(\tilde{\mathbf{w}}_l-\mathbf{w}_l)\mathbf{r}^{\mathsf{T}}\leq W_t^{\mathcal{A}'}(\Omega_l')-W_t^{\mathcal{A}}(\Omega_l)\leq(1+\sum_{i=1}^{\lfloor\frac{T-t}{2}\rfloor}(\beta\overline{\lambda})^{2i})(\tilde{\mathbf{w}}_l-\mathbf{w}_l)\mathbf{r}^{\mathsf{T}};$$

- (B2): if $\mathcal{A}'=\mathcal{A}$, $l\notin\mathcal{A}'$ and $l\notin\mathcal{A}$,

$$-\sum_{i=1}^{\lceil\frac{T-t}{2}\rceil}(\beta\overline{\lambda})^{2i-1}(\tilde{\mathbf{w}}_l-\mathbf{w}_l)\mathbf{r}^{\mathsf{T}}\leq W_t^{\mathcal{A}'}(\Omega_l')-W_t^{\mathcal{A}}(\Omega_l)\leq\sum_{i=1}^{\lfloor\frac{T-t}{2}\rfloor}(\beta\overline{\lambda})^{2i}(\tilde{\mathbf{w}}_l-\mathbf{w}_l)\mathbf{r}^{\mathsf{T}};$$

- (B3): if $\mathcal{A}'\setminus\{l\}\subset\mathcal{A}$, $l\in\mathcal{A}'$ and $l\notin\mathcal{A}$,

$$-\sum_{i=1}^{\lceil\frac{T-t}{2}\rceil}(\beta\overline{\lambda})^{2i-1}(\tilde{\mathbf{w}}_l-\mathbf{w}_l)\mathbf{r}^{\mathsf{T}}\leq W_t^{\mathcal{A}'}(\Omega_l')-W_t^{\mathcal{A}}(\Omega_l)\leq(1+\sum_{i=1}^{\lfloor\frac{T-t}{2}\rfloor}(\beta\overline{\lambda})^{2i})(\tilde{\mathbf{w}}_l-\mathbf{w}_l)\mathbf{r}^{\mathsf{T}}.$$

Proof: Please refer to Appendix F. □

Remark 4: (B1) achieves its lower bound when l is chosen at slot $t,t+1,t+3,\cdots$, and the upper bounds when l is chosen from $t,t+2,t+4,\cdots$. (B2) achieves its lower bound when l is chosen at slot $t+1,t+3,\cdots$, and upper bounds when l is chosen at $t+2,t+4,\cdots$.

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

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

B. Optimality of Myopic Policy for Transition Matrix With Negative or Positive Eigenvalues

Assumption 3: Assume that

1) $\lambda_2^{(n)}(t)=\cdots=\lambda_X^{(n)}(t)\triangleq\lambda^{(n)}(t)$ for $\mathbf{P}^{(n)}(t)$ $(n\in\mathcal{N})$.
2) For any slot t,

$$\max\{P_1^{(\varsigma_i^t)}(t),P_X^{(\varsigma_i^t)}(t)\}\leq_s\min\{P_1^{(\varsigma_{i+1}^t)}(t),P_X^{(\varsigma_{i+1}^t)}(t)\},$$

for $i=1,\cdots,N-1$, where $\varsigma_1^t,\cdots,\varsigma_N^t$ is one permutation of $\mathcal{N}$ at slot t.

Combing the lower and upper bounds of both Lemma 2 and Lemma 4, we have

Lemma 5: Under Assumption 3, $\overline{\lambda}\triangleq\max\{|\lambda^{(i)}(t)|:\ i\in\mathcal{N},1\leq t\leq T\}$. Given $\Omega_l=(\Omega_{-l},\mathbf{w}_l)$, $\Omega_l'=(\Omega_{-l},\tilde{\mathbf{w}}_l)$, $\mathbf{w}_l\leq_s\tilde{\mathbf{w}}_l$, we have for $1\leq t\leq T$

- (C1): if $\mathcal{A}'=\mathcal{A}$, $l\in\mathcal{A}'$ and $l\in\mathcal{A}$,

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

- (C2): if $\mathcal{A}'=\mathcal{A}$, $l\notin\mathcal{A}'$ and $l\notin\mathcal{A}$,

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

- (C3): if $\mathcal{A}'\setminus\{l\}\subset\mathcal{A}$, $l\in\mathcal{A}'$ and $l\notin\mathcal{A}$,

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

Remark 5: (C1) achieves its lower bound when $\lambda^{(l)}(t)\leq 0$, $\lambda^{(l)}(\tau)\geq 0(t+1\leq\tau\leq T)$ and l is chosen at slot $\tau=t,t+1,t+2,\cdots,T$, and achieves the upper bound when $\lambda^{(l)}(\tau)\geq 0$ and l is chosen at slot $\tau=t,t+1,t+2,\cdots,T$. (C2) achieves its lower bound when $\lambda^{(l)}(t)\leq 0$, $\lambda^{(l)}(\tau)\geq 0(t+1\leq\tau\leq T)$ and l is chosen at slot $\tau=t+1,t+2,\cdots,T$, and upper bounds when $\lambda^{(l)}(\tau)\geq 0$ and l is chosen at slot $\tau=t+1,t+2,\cdots,T$.

Theorem 3: Under Assumption 3, the myopic policy is optimal if $\sum_{i=1}^{T-t}(\beta\overline{\lambda})^i\leq\frac{1}{2}$, specifically, if $T\to\infty$, $\beta\overline{\lambda}\leq\frac{1}{3}$.

Corollary 3: When $X=2$ and $\mathbf{P}^{(n)}(t)=\mathbf{P}^{(n)}$ for any t, if $-\frac{1}{3}\leq p_{22}^{(n)}-p_{12}^{(n)}\leq\frac{1}{3}$, then the myopic policy is optimal.

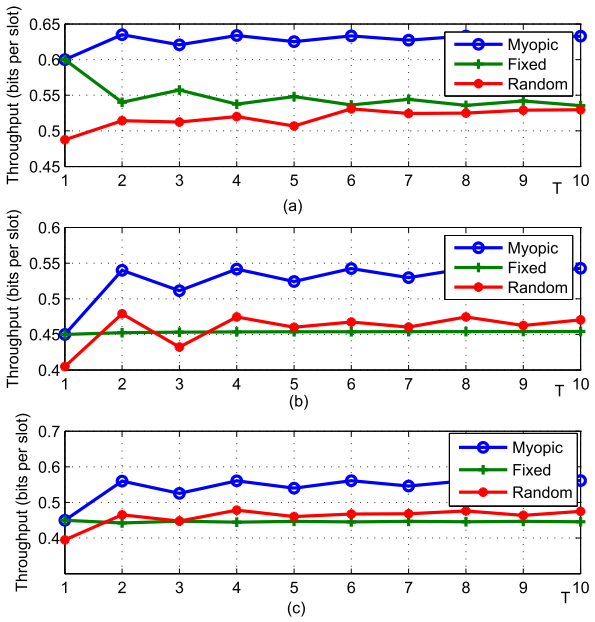

Fig. 1. Performance comparison of Myopic Policy, Fixed Policy, and Random Policy for Case 1 (Fig. 1(a)), Case 2 (Fig. 1(b)), and Case 3 (Fig. 1(c)), where $X = 3$, $N = 2$, $M = 1$, $\beta = 1$, $\mathbf{r} = (0.0 \quad 0.5 \quad 1.0)$, and probability transition matrices follow Tables. I, II, and III, respectively.

TABLE I

PROBABILITY TRANSITION MATRICES FOR CASE 1

n \ t	1	2	3	4	⋯	9	10	⋯
1	$\mathbf{P}[1]$	$\mathbf{P}[2]$	$\mathbf{P}[1]$	$\mathbf{P}[2]$	⋯	$\mathbf{P}[1]$	$\mathbf{P}[2]$	⋯
2	$\mathbf{P}[3]$	$\mathbf{P}[4]$	$\mathbf{P}[3]$	$\mathbf{P}[4]$	⋯	$\mathbf{P}[3]$	$\mathbf{P}[4]$	⋯

TABLE II

PROBABILITY TRANSITION MATRICES FOR CASE 2

n \ t	1	2	3	4	⋯	9	10	⋯
1	$\mathbf{P}[5]$	$\mathbf{P}[6]$	$\mathbf{P}[5]$	$\mathbf{P}[6]$	⋯	$\mathbf{P}[5]$	$\mathbf{P}[6]$	⋯
2	$\mathbf{P}[7]$	$\mathbf{P}[8]$	$\mathbf{P}[7]$	$\mathbf{P}[8]$	⋯	$\mathbf{P}[7]$	$\mathbf{P}[8]$	⋯

V. NUMERICAL SIMULATION

In this section, we study the average reward (corresponding to the expected accumulated discount reward divided by T when $\beta = 1$) performance of *Myopic policy*, *Random policy* (randomly choosing a channel at each time slot), and *Fixed policy* (choosing a fixed channel at all time slots) by three simplest scenarios in which $X = 3$, $N = 2$, $M = 1$, $\beta = 1$, $\mathbf{r} = (0.0 \quad 0.5 \quad 1.0)$, and

1) Case 1. The probability transition matrices are set according to Table I, herein, $\lambda^{(1)}(t) = 0.4$ or 0.3, $\lambda^{(2)}(t) = 0.1$ or 0.2, which corresponds to the conditions in Theorem 1.
2) Case 2. The probability transition matrices are set according to Table II, herein, $\lambda^{(1)}(t) = -0.1$, $\lambda^{(2)}(t) = -0.1$ or -0.2, which corresponds to the conditions in Theorem 2.
3) Case 3. The probability transition matrices are set according to Table III, herein, $\lambda^{(1)}(t) = 0.4$ or 0.3, $\lambda^{(2)}(t) = -0.1$.

Fig. 1 shows that the average reward obtained by the myopic policy outperforms both the fixed policy and the random policy to various extents, since the myopic policy is optimal in the above three cases. Considering the exponential complexity of obtaining the optimal policy, the benefit of the myopic policy is obvious.

TABLE III

PROBABILITY TRANSITION MATRICES FOR CASE 3

n \ t	1	2	3	4	⋯	9	10	⋯
1	$\mathbf{P}[1]$	$\mathbf{P}[2]$	$\mathbf{P}[1]$	$\mathbf{P}[2]$	⋯	$\mathbf{P}[1]$	$\mathbf{P}[2]$	⋯
2	$\mathbf{P}[5]$	$\mathbf{P}[6]$	$\mathbf{P}[5]$	$\mathbf{P}[6]$	⋯	$\mathbf{P}[5]$	$\mathbf{P}[6]$	⋯

VI. 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. Due to the generic RMAB formulation of the problem, the derived results and the analysis methodology proposed in this paper can be applied in a wide range of domains. Some future research directions include seeking the optimality of myopic policy under indirect (or imperfect) observation with error, seeking simple index policy for this problem under certain conditions, and finding policy for restless bandit problem when channels are correlated to certain extent.

APPENDIX A
PROOF OF LEMMA 1

We prove the lemma by backward induction. For T, we have $W_T^{\hat{\mathcal{A}}}(\Omega(T)) = \sum_{n\in\hat{\mathcal{A}}(T)} \mathbf{w}_n(T)\mathbf{r}^{\mathsf{T}}$ which is obviously decomposable for $\mathbf{w}_i$ $(i \in \mathcal{N})$. Suppose that the lemma holds for $T-1, \cdots, t+1$, we prove that it still holds for t by two different cases.

Case 1: $i \in \mathcal{A}(t)$.

$$
\begin{aligned}
&W_t^{\mathcal{A}}(\mathbf{w}_1, \cdots, \mathbf{w}_i, \cdots, \mathbf{w}_N) \\
&\quad = \sum_{n\in\mathcal{A}(t)} \mathbf{w}_n(t)\mathbf{r}^{\mathsf{T}} + \beta\,\Sigma(\mathcal{A}(t), \Omega(t)) W_{t+1}^{\hat{\mathcal{A}}}(\Omega(t+1)) \\
&\quad = \sum_{n\in\mathcal{A}(t)} \mathbf{w}_n(t)\mathbf{r}^{\mathsf{T}} + \beta\,\Sigma(\mathcal{A}(t)\setminus\{i\}, \Omega(t)) \\
&\qquad \times \sum_{j\in\mathcal{X}} \omega_{ij}(t) W_{t+1}^{\hat{\mathcal{A}}}(\Omega^{-i}(t+1), \mathbf{e}_j\mathbf{P}^{(i)}(t)). \qquad (8)
\end{aligned}
$$

$$
\begin{aligned}
&\sum_{j=1}^{X} \omega_{ij} W_t^{\mathcal{A}}(\mathbf{w}_1, \cdots, \mathbf{e}_j, \cdots, \mathbf{w}_N) \\
&\quad = \sum_{j=1}^{X} \omega_{ij} \Big[\sum_{n\in\mathcal{A}(t)} \mathbf{e}_j\mathbf{r}^{\mathsf{T}} \\
&\qquad + \beta\,\Sigma(\mathcal{A}(t)\setminus\{i\}, \Omega(t)) W_{t+1}^{\hat{\mathcal{A}}}(\Omega^{-i}(t+1), \mathbf{e}_j\mathbf{P}^{(i)}(t)) \Big]
\end{aligned}
$$

$$= \sum_{n\in\mathcal{A}(t)} \mathbf{w}_n(t)\mathbf{r}^\mathsf{T} + \beta\Sigma(\mathcal{A}(t)\setminus\{i\},\Omega(t)) \times \sum_{j\in\mathcal{X}} \omega_{ij}(t) W_{t+1}^{\hat{\mathcal{A}}}(\Omega^{-i}(t+1), \mathbf{e}_j\mathbf{P}^{(i)}(t)). \quad (9)$$

By (8) and (9), we have $W_t^{\mathcal{A}}(\mathbf{w}_1,\cdots,\mathbf{w}_i,\cdots,\mathbf{w}_N) = \sum_{j=1}^{X}\omega_{ij}W_t^{\mathcal{A}}(\mathbf{w}_1,\cdots,\mathbf{e}_j,\cdots,\mathbf{w}_N)$.

Case 2: $i \notin \mathcal{A}(t)$.

$$\begin{aligned}
&W_t^{\mathcal{A}}(\mathbf{w}_1,\cdots,\mathbf{w}_i,\cdots,\mathbf{w}_N)\\
&= \sum_{n\in\mathcal{A}(t)} \mathbf{w}_n(t)\mathbf{r}^\mathsf{T} + \beta\Sigma(\mathcal{A}(t),\Omega(t)) W_{t+1}^{\hat{\mathcal{A}}}(\Omega(t+1))\\
&= \sum_{n\in\mathcal{A}(t)} \mathbf{w}_n(t)\mathbf{r}^\mathsf{T} + \beta\Sigma(\mathcal{A}(t),\Omega(t)) W_{t+1}^{\hat{\mathcal{A}}}(\Omega^{-i}(t+1), \times \mathbf{w}_i(t)\mathbf{P}^{(i)}(t))\\
&= \sum_{n\in\mathcal{A}(t)} \mathbf{w}_n(t)\mathbf{r}^\mathsf{T} + \beta\Sigma(\mathcal{A}(t),\Omega(t)) \sum_{j\in\mathcal{X}} \mathbf{w}_i(t)\mathbf{P}^{(i)}(t)\mathbf{e}_j^\mathsf{T} W_{t+1}^{\hat{\mathcal{A}}}(\Omega^{-i}(t+1),\mathbf{e}_j).
\end{aligned} \quad (10)$$

$$\begin{aligned}
&\sum_{j=1}^{X}\omega_{ij}W_t^{\mathcal{A}}(\mathbf{w}_1,\cdots,\mathbf{e}_j,\cdots,\mathbf{w}_N)\\
&= \sum_{j=1}^{X}\omega_{ij}\Big[\sum_{n\in\mathcal{A}(t)} \mathbf{e}_j\mathbf{r}^\mathsf{T} + \beta\Sigma(\mathcal{A}(t),\Omega(t)) W_{t+1}^{\hat{\mathcal{A}}}(\Omega^{-i}(t+1), \times\mathbf{e}_j\mathbf{P}^{(i)}(t))\Big]\\
&= \sum_{n\in\mathcal{A}(t)} \mathbf{w}_n(t)\mathbf{r}^\mathsf{T} + \beta\Sigma(\mathcal{A}(t),\Omega(t)) \sum_{j\in\mathcal{X}} \omega_{ij}(t) W_{t+1}^{\hat{\mathcal{A}}}(\Omega^{-i}(t+1), \mathbf{e}_j\mathbf{P}^{(i)}(t))\\
&= \sum_{n\in\mathcal{A}(t)} \mathbf{w}_n(t)\mathbf{r}^\mathsf{T} + \beta\Sigma(\mathcal{A}(t),\Omega(t)) \times \sum_{j\in\mathcal{X}} \omega_{ij}(t) \sum_{k\in\mathcal{X}} \mathbf{e}_j\mathbf{P}^{(i)}(t)\mathbf{e}_k^\mathsf{T} W_{t+1}^{\hat{\mathcal{A}}}(\Omega^{-i}(t+1),\mathbf{e}_d)\\
&= \sum_{n\in\mathcal{A}(t)} \mathbf{w}_n(t)\mathbf{r}^\mathsf{T} + \beta\Sigma(\mathcal{A}(t),\Omega(t)) \sum_{k\in\mathcal{X}} \mathbf{w}_i(t)\mathbf{P}^{(i)}(t)\mathbf{e}_k^\mathsf{T} W_{t+1}^{\hat{\mathcal{A}}}(\Omega^{-i}(t+1),\mathbf{e}_k)
\end{aligned} \quad (11)$$

By (10) and (11), we have $W_t^{\mathcal{A}}(\mathbf{w}_1,\cdots,\mathbf{w}_i,\cdots,\mathbf{w}_N) = \sum_{j=1}^{X}\omega_{ij}W_t^{\mathcal{A}}(\mathbf{w}_1,\cdots,\mathbf{e}_j,\cdots,\mathbf{w}_N)$.

Combing Case 1-2, we prove the lemma.

Appendix B
Proof of Proposition 1

(1) For the property of $\lambda_1 = 1$ and $\mathbf{v}_1 = \frac{\mathbf{I}_X}{\sqrt{X}}$, it is easily verified, i.e.,

$$\mathbf{P}\mathbf{I}_X^\mathsf{T} = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1X}\\ p_{21} & p_{22} & \cdots & p_{2X}\\ \vdots & \vdots & \ddots & \vdots\\ p_{X1} & p_{X2} & \cdots & p_{XX}\end{pmatrix}\begin{pmatrix}1\\1\\\vdots\\1\end{pmatrix} = \begin{pmatrix}\sum_{j=1}^{X} p_{1,j}\\ \sum_{j=1}^{X} p_{2,j}\\ \vdots\\ \sum_{j=1}^{X} p_{X,j}\end{pmatrix} = \begin{pmatrix}1\\1\\\vdots\\1\end{pmatrix} = \mathbf{I}_X^\mathsf{T}.$$

(2) For the property of replacing λ_1 with any value λ, we have the LHS of (7)

$$\begin{aligned}
&(\mathbf{w}_m-\mathbf{w}_n)(\mathbf{v}_1\ \cdots\ \mathbf{v}_X)^\mathsf{T}\begin{pmatrix}\lambda_1 & 0 & \ldots & 0\\ 0 & \lambda_2 & \ldots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \ldots & \lambda_X\end{pmatrix}\\
&= [\lambda_1(\mathbf{w}_m-\mathbf{w}_n)\mathbf{v}_1^\mathsf{T}\ \ \lambda_2(\mathbf{w}_m-\mathbf{w}_n)\mathbf{v}_2^\mathsf{T}\ \ \cdots\ \ \lambda_X(\mathbf{w}_m-\mathbf{w}_n)\mathbf{v}_X^\mathsf{T}]\\
&= [\lambda_1(\mathbf{w}_m-\mathbf{w}_n)\mathbf{I}_X^\mathsf{T}\ \ \lambda_2(\mathbf{w}_m-\mathbf{w}_n)\mathbf{v}_2^\mathsf{T}\ \ \cdots\ \ \lambda_X(\mathbf{w}_m-\mathbf{w}_n)\mathbf{v}_X^\mathsf{T}]\\
&= [0\ \ \lambda_2(\mathbf{w}_m-\mathbf{w}_n)\mathbf{v}_2^\mathsf{T}\ \ \cdots\ \ \lambda_X(\mathbf{w}_m-\mathbf{w}_n)\mathbf{v}_X^\mathsf{T}].
\end{aligned} \quad (12)$$

For the RHS of (7), we have

$$\begin{aligned}
&(\mathbf{w}_m-\mathbf{w}_n)(\mathbf{v}_1\ \cdots\ \mathbf{v}_X)^\mathsf{T}\begin{pmatrix}\lambda & 0 & \ldots & 0\\ 0 & \lambda_2 & \ldots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \ldots & \lambda_X\end{pmatrix}\\
&= [\lambda(\mathbf{w}_m-\mathbf{w}_n)\mathbf{v}_1^\mathsf{T}\ \ \lambda_2(\mathbf{w}_m-\mathbf{w}_n)0\mathbf{v}_2^\mathsf{T}\ \ \cdots\ \ \lambda_X(\mathbf{w}_m-\mathbf{w}_n)\mathbf{v}_X^\mathsf{T}]\\
&= [\lambda(\mathbf{w}_m-\mathbf{w}_n)\mathbf{I}_X^\mathsf{T}\ \ \lambda_2(\mathbf{w}_m-\mathbf{w}_n)\mathbf{v}_2^\mathsf{T}\ \ \cdots\ \ \lambda_X(\mathbf{w}_m-\mathbf{w}_n)\mathbf{v}_X^\mathsf{T}]\\
&= [0\ \ \lambda_2(\mathbf{w}_m-\mathbf{w}_n)\mathbf{v}_2^\mathsf{T}\ \ \cdots\ \ \lambda_X(\mathbf{w}_m-\mathbf{w}_n)\mathbf{v}_X^\mathsf{T}].
\end{aligned} \quad (13)$$

By (12) and (13), we prove the equation (7).

Appendix C
Proof of Proposition 2

According to Assumption 1, we have the determinant $|\mathbf{P}^{(n)}(t)|$ of $\mathbf{P}^{(n)}(t)$ is not less than 0, i.e., $|\mathbf{P}^{(n)}(t)| = (\lambda^{(n)}(t))^{X-1} > 0$. Thus, $\mathbf{P}^{(n)}$ can be decomposed as follows

$$\begin{aligned}
&\mathbf{P}^{(n)}(t)\\
&= (\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_X)\begin{pmatrix}1 & 0 & \ldots & 0\\ 0 & \lambda & \ldots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \ldots & \lambda\end{pmatrix}(\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_X)^{-1}\\
&= \begin{pmatrix}\frac{1}{\sqrt{X}} & v_{21} & \ldots & v_{X1}\\ \frac{1}{\sqrt{X}} & v_{22} & \ldots & v_{X2}\\ \vdots & \vdots & \ddots & \vdots\\ \frac{1}{\sqrt{X}} & v_{2X} & \ldots & v_{XX}\end{pmatrix}\begin{pmatrix}1 & 0 & \ldots & 0\\ 0 & \lambda & \ldots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \ldots & \lambda\end{pmatrix}(\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_X)^{-1}\\
&= \begin{pmatrix}\frac{1}{\sqrt{X}} & v_{21} & \ldots & v_{X1}\\ \frac{1}{\sqrt{X}} & v_{22} & \ldots & v_{X2}\\ \vdots & \vdots & \ddots & \vdots\\ \frac{1}{\sqrt{X}} & v_{2X} & \ldots & v_{XX}\end{pmatrix}\begin{pmatrix}\lambda & 0 & \ldots & 0\\ 0 & \lambda & \ldots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \ldots & \lambda\end{pmatrix}(\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_X)^{-1}\\
&\quad + \begin{pmatrix}\frac{1}{\sqrt{X}} & v_{21} & \ldots & v_{X1}\\ \frac{1}{\sqrt{X}} & v_{22} & \ldots & v_{X2}\\ \vdots & \vdots & \ddots & \vdots\\ \frac{1}{\sqrt{X}} & v_{2X} & \ldots & v_{XX}\end{pmatrix}\begin{pmatrix}1-\lambda & 0 & \ldots & 0\\ 0 & 0 & \ldots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \ldots & 0\end{pmatrix}(\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_X)^{-1}
\end{aligned}$$

$$= \lambda \mathbf{E} + \frac{1}{\sqrt{X}} \begin{pmatrix} (1-\lambda)q_{11} & (1-\lambda)q_{12} & \dots & (1-\lambda)q_{1X} \\ (1-\lambda)q_{11} & (1-\lambda)q_{12} & \dots & (1-\lambda)q_{1X} \\ \vdots & \vdots & \ddots & \vdots \\ (1-\lambda)q_{11} & (1-\lambda)q_{12} & \dots & (1-\lambda)q_{1X} \end{pmatrix}$$

$$= \begin{pmatrix} P_1^{(n)}(t) \\ P_2^{(n)}(t) \\ \vdots \\ P_X^{(n)}(t) \end{pmatrix},$$

where,

$$(\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_X)^{-1} := \begin{pmatrix} q_{11} & q_{12} & \dots & q_{1X} \\ q_{21} & q_{22} & \dots & q_{2X} \\ \vdots & \vdots & \ddots & \vdots \\ q_{X1} & q_{X2} & \dots & q_{XX} \end{pmatrix}.$$

It is easily to verify that $P_1^{(n)}(t) \leq_s P_2^{(n)}(t) \leq_s \cdots \leq_s P_X^{(n)}(t)$ if $\lambda > 0$, and further $\mathbf{P}^{(n)}(t)$ is a first order stochastic dominance matrix.

Appendix D
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 $W_T^{\mathcal{A}'}(\Omega_l') - W_T^{\mathcal{A}}(\Omega_l) = \mathbf{r}(\tilde{\mathbf{w}}_l - \mathbf{w}_l)$;
2) For $l \notin \mathcal{A}'$, $l \notin \mathcal{A}$, it holds that $W_T^{\mathcal{A}'}(\Omega_l') - W_T^{\mathcal{A}}(\Omega_l) = 0$;
3) For $l \in \mathcal{A}'$, $l \notin \mathcal{A}$, it exists at least one channel m such that $\tilde{\mathbf{w}}_l \geq \mathbf{w}_m \geq \mathbf{w}_l$. It then holds that $0 \leq W_T^{\mathcal{A}'}(\Omega_l') - W_T^{\mathcal{A}}(\Omega_l) = (\tilde{\mathbf{w}}_l - \mathbf{w}_m)\mathbf{r}^{\mathsf{T}} \leq (\tilde{\mathbf{w}}_l - \mathbf{w}_l)\mathbf{r}^{\mathsf{T}}$.

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:

$$F(\mathcal{A}', \Omega_l') = \Sigma(\mathcal{A}' \setminus \{l\}, \Omega_l') \Big[\sum_{j \in \mathcal{X}} \tilde{\omega}_{lj}(t) W_{t+1}(\Omega_{-l}, \mathbf{e}_j \mathbf{P}^{(l)}(t)) \Big], \tag{14}$$

$$F(\mathcal{A}, \Omega_l) = \Sigma(\mathcal{A} \setminus \{l\}, \Omega_l) \Big[\sum_{j \in \mathcal{X}} \omega_{lj}(t) W_{t+1}(\Omega_{-l}, \mathbf{e}_j \mathbf{P}^{(l)}(t)) \Big]. \tag{15}$$

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

$$\begin{aligned} &F(\mathcal{A}', \Omega_l') - F(\mathcal{A}, \Omega_l) \\ &= \Sigma(\mathcal{A} \setminus \{l\}, \Omega_l) \Big[\sum_{j \in \mathcal{X}} \tilde{\omega}_{lj}(t) W_{t+1}(\Omega_{-l}, \mathbf{e}_j \mathbf{P}^{(l)}(t)) \\ &\quad - \sum_{j \in \mathcal{X}} \omega_{lj}(t) W_{t+1}(\Omega_{-l}, \mathbf{e}_j \mathbf{P}^{(l)}(t)) \Big] \\ &\overset{(a)}{=} \Sigma(\mathcal{A} \setminus \{l\}, \Omega_l) \sum_{j=2}^{X} \Big[(\tilde{\omega}_{lj}(t) - \omega_{lj}(t)) \\ &\quad \times \Big(W_{t+1}(\Omega_{-l}, \mathbf{e}_j \mathbf{P}^{(l)}(t)) - W_{t+1}(\Omega_{-l}, \mathbf{e}_1 \mathbf{P}^{(l)}(t)) \Big) \Big] \\ &= \Sigma(\mathcal{A} \setminus \{l\}, \Omega_l) \sum_{j=2}^{X} \Big[\sum_{i=j}^{X} (\tilde{\omega}_{li}(t) - \omega_{li}(t)) \\ &\quad \times \Big(W_{t+1}(\Omega_{-l}, \mathbf{e}_j \mathbf{P}^{(l)}(t)) - W_{t+1}(\Omega_{-l}, \mathbf{e}_{j-1} \mathbf{P}^{(l)}(t)) \Big) \Big] \end{aligned} \tag{16}$$

where, the equality (a) is due to $\omega_{l1}(t) = 1 - \sum_{j=2}^{X} \omega_{lj}(t)$.

Next, we analyze the term in the bracket, $W_{t+1}(\Omega_{-l}, \mathbf{e}_j \mathbf{P}^{(l)}(t)) - W_{t+1}(\Omega_{-l}, \mathbf{e}_{j-1} \mathbf{P}^{(l)}(t))$, of RHS of (16) through three cases:

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

$$\begin{aligned} 0 &\leq W_{t+1}(\Omega_{-l}, \mathbf{e}_j \mathbf{P}^{(l)}(t)) - W_{t+1}(\Omega_{-l}, \mathbf{e}_{j-1} \mathbf{P}^{(l)}(t)) \\ &\leq \sum_{i=0}^{T-t-1} (\beta \overline{\lambda})^i (\mathbf{e}_j - \mathbf{e}_{j-1}) \mathbf{P}^{(l)}(t) \mathbf{r}^{\mathsf{T}}. \end{aligned} \tag{17}$$

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

$$\begin{aligned} 0 &\leq W_{t+1}(\Omega_{-l}, \mathbf{e}_j \mathbf{P}^{(l)}(t)) - W_{t+1}(\Omega_{-l}, \mathbf{e}_{j-1} \mathbf{P}^{(l)}(t)) \\ &\leq \sum_{i=1}^{T-t-1} (\beta \overline{\lambda})^i (\mathbf{e}_j - \mathbf{e}_{j-1}) \mathbf{P}^{(l)}(t) \mathbf{r}^{\mathsf{T}}. \end{aligned} \tag{18}$$

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

$$\begin{aligned} 0 &\leq W_{t+1}(\Omega_{-l}, \mathbf{e}_j \mathbf{P}^{(l)}(t)) - W_{t+1}(\Omega_{-l}, \mathbf{e}_{j-1} \mathbf{P}^{(l)}(t)) \\ &\leq \sum_{i=0}^{T-t-1} (\beta \overline{\lambda})^i (\mathbf{e}_j - \mathbf{e}_{j-1}) \mathbf{P}^{(l)}(t) \mathbf{r}^{\mathsf{T}}. \end{aligned} \tag{19}$$

Combining Case 1–3, we obtain the following:

$$\begin{aligned} 0 &\leq W_{t+1}(\Omega_{-l}, \mathbf{e}_j \mathbf{P}^{(l)}(t)) - W_{t+1}(\Omega_{-l}, \mathbf{e}_{j-1} \mathbf{P}^{(l)}(t)) \\ &\leq \sum_{i=0}^{T-t-1} (\beta \overline{\lambda})^i (\mathbf{e}_j - \mathbf{e}_{j-1}) \mathbf{P}^{(l)}(t) \mathbf{r}^{\mathsf{T}}. \end{aligned} \tag{20}$$

Therefore, combining (16) and (20), we have

$$\begin{aligned} 0 &\leq W_t^{\mathcal{A}'}(\Omega_l') - W_t^{\mathcal{A}}(\Omega_l) \\ &= (\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t))\mathbf{r}^{\mathsf{T}} + \beta(F(\mathcal{A}', \Omega_l') - F(\mathcal{A}, \Omega_l)) \\ &\leq (\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t))\mathbf{r}^{\mathsf{T}} \\ &\quad + \beta \sum_{j=2}^{X} \Big[\sum_{i=j}^{X} (\tilde{\omega}_{li}(t) - \omega_{li}(t)) \\ &\quad \times \sum_{i=0}^{T-t-1} (\beta \overline{\lambda})^i (\mathbf{e}_j - \mathbf{e}_{j-1}) \mathbf{P}^{(l)}(t) \mathbf{r}^{\mathsf{T}} \Big] \\ &= (\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t))\mathbf{r}^{\mathsf{T}} + \beta \Big[\sum_{i=0}^{T-t-1} (\beta \overline{\lambda})^i (\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t)) \mathbf{P}^{(l)}(t) \mathbf{r} \Big] \\ &\overset{(a)}{\leq} (\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t))\mathbf{r}^{\mathsf{T}} + \beta \Big[\sum_{i=0}^{T-t-1} (\beta \overline{\lambda})^i (\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t)) (\overline{\lambda} \mathbf{E}) \mathbf{r}^{\mathsf{T}} \Big] \end{aligned}$$

$$= (\tilde{\mathbf{w}}_l(t)-\mathbf{w}_l(t))\mathbf{r}^{\mathsf{T}}+\sum_{i=1}^{T-t}(\beta\overline{\lambda})^i(\tilde{\mathbf{w}}_l(t)-\mathbf{w}_l(t))\mathbf{r}^{\mathsf{T}}$$

$$= \sum_{i=0}^{T-t}(\beta\overline{\lambda})^i(\tilde{\mathbf{w}}_l(t)-\mathbf{w}_l(t))\mathbf{r}^{\mathsf{T}},$$

where, 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:

$$F(\mathcal{A}', \Omega_l') = \Sigma(\mathcal{A}(t)\setminus\{k\}, \Omega_l')\Big[\sum_{j\in\mathcal{X}}\omega_{kj}(t)W_{t+1}(\Omega_{-k}', \mathbf{e}_j\mathbf{P}^{(k)}(t))\Big], \tag{21}$$

$$F(\mathcal{A}, \Omega_l) = \Sigma(\mathcal{A}(t)\setminus\{k\}, \Omega_l)\Big[\sum_{j\in\mathcal{X}}\omega_{kj}(t)W_{t+1}(\Omega_{-k}, \mathbf{e}_j\mathbf{P}^{(k)}(t))\Big]. \tag{22}$$

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

$$F(\mathcal{A}', \Omega_l') - F(\mathcal{A}, \Omega_l) = \Sigma(\mathcal{A}(t)\setminus\{k\}, \Omega_l)\Big[\sum_{j\in\mathcal{X}}\omega_{kj}(t)\Big(W_{t+1}(\Omega_{-k}', \mathbf{e}_j\mathbf{P}^{(k)}(t)) - W_{t+1}(\Omega_{-k}, \mathbf{e}_j\mathbf{P}^{(k)}(t))\Big)\Big]. \tag{23}$$

For the term in the bracket of RHS of (23), if channel l is never chosen for $W_{t+1}(\Omega_{-k}', \mathbf{e}_j\mathbf{P}^{(k)}(t))$ and $W_{t+1}(\Omega_{-k}, \mathbf{e}_j\mathbf{P}^{(k)}(t))$ 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 $W_{t+1}(\Omega_{-k}', \mathbf{e}_j\mathbf{P}^{(k)}(t)) - W_{t+1}(\Omega_{-k}, \mathbf{e}_j\mathbf{P}^{(k)}(t)) = 0$; 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

$$\begin{aligned} W_{t^0}(\Omega_l'(t^o)) - W_{t^0}(\Omega_l(t^o)) &\le \sum_{i=0}^{T-t^o}(\beta\overline{\lambda})^i(\tilde{\mathbf{w}}_l(t^o) - \mathbf{w}_l(t^o))\mathbf{r}^{\mathsf{T}} \\ &= \sum_{i=0}^{T-t^o}(\beta\overline{\lambda})^i(\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t))\prod_{\tau=t+1}^{t^0}\mathbf{P}^{(l)}(\tau)\mathbf{r}^{\mathsf{T}} \\ &\overset{(a)}{\le} \sum_{i=0}^{T-t^o}(\beta\overline{\lambda})^i(\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t))(\overline{\lambda}\mathbf{E})^{t^0-t}\mathbf{r}^{\mathsf{T}} \\ &= \sum_{i=0}^{T-t^o}(\beta\overline{\lambda})^i(\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t))(\overline{\lambda}\mathbf{E})^{t^0-t}\mathbf{r}^{\mathsf{T}} \\ &\overset{t^0=t+1}{\le} \sum_{i=0}^{T-t-1}(\beta\overline{\lambda})^i(\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t))(\overline{\lambda}\mathbf{E})\mathbf{r}' \\ &= \overline{\lambda}\sum_{i=0}^{T-t-1}(\beta\overline{\lambda})^i(\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t)))\mathbf{r}^{\mathsf{T}}, \end{aligned}$$

where, the equality (a) 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

$$\begin{aligned} &W_{t+1}(\Omega_{-k}', \mathbf{e}_j\mathbf{P}^{(k)}(t)) - W_{t+1}(\Omega_{-k}, \mathbf{e}_j\mathbf{P}^{(k)}(t)) \\ &\le \overline{\lambda}\sum_{i=0}^{T-t-1}(\beta\overline{\lambda})^i(\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t)))\mathbf{r}^{\mathsf{T}}. \end{aligned} \tag{24}$$

Combing (24) and (23), we have

$$\begin{aligned} W_t^{\mathcal{A}'}(\Omega_l') - W_t^{\mathcal{A}}(\Omega_l) &= \beta(F(\mathcal{A}', \Omega_l') - F(\mathcal{A}, \Omega_l)) \\ &\le \beta\overline{\lambda}\sum_{i=0}^{T-t-1}(\beta\overline{\lambda})^i(\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t)))\mathbf{r}^{\mathsf{T}} \\ &\le \sum_{i=1}^{T-t}(\beta\overline{\lambda})^i(\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t)))\mathbf{r}^{\mathsf{T}}, \end{aligned}$$

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 $\mathbf{w}_l' \ge_s \mathbf{w}_m \ge_s \mathbf{w}_l$. We have

$$\begin{aligned} &W_t^{\mathcal{A}'}(\Omega_l'(t)) - W_t^{\mathcal{A}}(\Omega_l(t)) \\ &= W_t^{\mathcal{A}'}(\mathbf{w}_1, \cdots, \tilde{\mathbf{w}}_l, \cdots, \mathbf{w}_N) - W_t^{\mathcal{A}}(\mathbf{w}_1, \cdots, \mathbf{w}_l, \cdots, \mathbf{w}_N) \\ &= W_t^{\mathcal{A}'}(\mathbf{w}_1, \cdots, \tilde{\mathbf{w}}_l, \cdots, \mathbf{w}_N) \\ &\quad - W_t^{\mathcal{A}'}(\mathbf{w}_1, \cdots, \mathbf{w}_l = \mathbf{w}_m, \cdots, \mathbf{w}_N) \\ &\quad + W_t^{\mathcal{A}}(\mathbf{w}_1, \cdots, \mathbf{w}_l = \mathbf{w}_m, \cdots, \mathbf{w}_N) \\ &\quad - W_t^{\mathcal{A}}(\mathbf{w}_1, \cdots, \mathbf{w}_l, \cdots, \mathbf{w}_N) \end{aligned} \tag{25}$$

According to the induction hypothesis ($l \in \mathcal{A}'$ and $l \in \mathcal{A}$), the first term of the RHS of (25) can be bounded as follows:

$$\begin{aligned} 0 \le\ & W_t^{\mathcal{A}'}(\mathbf{w}_1, \cdots, \tilde{\mathbf{w}}_l, \cdots, \mathbf{w}_N) \\ &- W_t^{\mathcal{A}'}(\mathbf{w}_1, \cdots, \mathbf{w}_l = \mathbf{w}_m, \cdots, \mathbf{w}_N) \\ \le\ & \sum_{i=0}^{T-t}(\beta\overline{\lambda})i(\tilde{\mathbf{w}}_l(t) - \mathbf{w}_m(t))\mathbf{r}^{\mathsf{T}} \end{aligned} \tag{26}$$

Meanwhile, the second term of the RHS of (25) is inducted by hypothesis ($l \notin \mathcal{A}'$ and $l \notin \mathcal{A}$):

$$\begin{aligned} 0 &\le W_t^{\mathcal{A}}(\mathbf{w}_1, \cdots, \mathbf{w}_l = \mathbf{w}_m, \cdots, \mathbf{w}_N) - W_t^{\mathcal{A}}(\mathbf{w}_1, \cdots, \mathbf{w}_l, \cdots, \mathbf{w}_N) \\ &\le \sum_{i=1}^{T-t}(\beta\overline{\lambda})^i(\mathbf{w}_m(t) - \mathbf{w}_l(t))\mathbf{r}^{\mathsf{T}} \end{aligned} \tag{27}$$

Therefore, we have, combining (25), (26) and (27),

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

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.

Appendix E
Proof of Proposition 4

According to the deriving process of Proposition 2, we have

$$\mathbf{P}^{(n)}(t) = \begin{pmatrix} c_1^n(t)+\lambda^{(n)}(t) & c_2^n(t) & \cdots & c_X^n(t) \\ c_1^n(t) & c_2^n(t)+\lambda^{(n)}(t) & \cdots & c_X^n(t) \\ \vdots & \vdots & \ddots & \vdots \\ c_1^n(t) & c_2^n(t) & \cdots & c_X^n(t)+\lambda^{(n)}(t) \end{pmatrix}. \tag{28}$$

Thus, under Assumption 2, it is easily to verify that $P_1^{(n)}(t) \ge_s P_2^{(n)}(t) \ge_s \cdots \ge_s P_X^{(n)}(t)$ if $\lambda^{(n)}(t) < 0$.

By some simple matrix operations, we have $\mathbf{w}_m \prod_{\tau=t}^{t+k-1}\mathbf{P}^{(n)}(\tau)$ and $\mathbf{w}_l \prod_{\tau=t}^{t+k-1}\mathbf{P}^{(n)}(\tau)$ in the top of the next page. (See (29) (30), at the top of the next page.)

Thus, if $k = 2i$, then $\prod_{\tau=t}^{t+2i-1}\lambda^{(n)}(\tau) > 0$, and further, $\mathbf{w}_m \prod_{\tau=t}^{t+2i-1}\mathbf{P}^{(n)}(\tau) \ge_s \mathbf{w}_l \prod_{\tau=t}^{t+2i-1}\mathbf{P}^{(n)}(\tau)$; otherwise, $\prod_{\tau=t}^{t+2i}\lambda^{(n)}(\tau) < 0$ and $\mathbf{w}_m \prod_{\tau=t}^{t+2i}\mathbf{P}^{(n)}(\tau) \le_s \mathbf{w}_l \prod_{\tau=t}^{t+2i}\mathbf{P}^{(n)}(\tau)$.

Appendix F
Proof of Lemma 4

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 $W_T^{\mathcal{A}'}(\Omega_l') - W_T^{\mathcal{A}}(\Omega_l) = (\tilde{\mathbf{w}}_l - \mathbf{w}_l)\mathbf{r}^{\mathsf{T}}$;
2) For $l \notin \mathcal{A}'$, $l \notin \mathcal{A}$, it holds that $W_T^{\mathcal{A}'}(\Omega_l') - W_T^{\mathcal{A}}(\Omega_l) = 0$;
3) For $l \in \mathcal{A}'$, $l \notin \mathcal{A}$, it exists at least one channel m such that $\mathbf{w}_l' \ge \mathbf{w}_m \ge \mathbf{w}_l$. It then holds that $0 \le W_T^{\mathcal{A}'}(\Omega_l') - W_T^{\mathcal{A}}(\Omega_l) \le (\tilde{\mathbf{w}}_l - \mathbf{w}_l)\mathbf{r}^{\mathsf{T}}$.

Therefore, Lemma 4 holds for slot T.

Assume that Lemma 4 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:

$$F(\mathcal{A}', \Omega_l') = \Sigma(\mathcal{A}' \setminus \{l\}, \Omega_l')\Big[\sum_{j\in\mathcal{X}}\omega_{lj}'(t)W_{t+1}(\Omega_{-l}, \mathbf{e}_j\mathbf{P}^{(l)}(t))\Big], \tag{31}$$

$$F(\mathcal{A}, \Omega_l) = \Sigma(\mathcal{A} \setminus \{l\}, \Omega_l)\Big[\sum_{j\in\mathcal{X}}\omega_{lj}(t)W_{t+1}(\Omega_{-l}, \mathbf{e}_j\mathbf{P}^{(l)}(t))\Big]. \tag{32}$$

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

$$\begin{aligned} &F(\mathcal{A}', \Omega_l') - F(\mathcal{A}, \Omega_l) \\ &= \Sigma(\mathcal{A} \setminus \{l\}, \Omega_l)\Big[\sum_{j\in\mathcal{X}}\tilde{\omega}_{lj}(t)W_{t+1}(\Omega_{-l}, \mathbf{e}_j\mathbf{P}^{(l)}(t)) \\ &\quad - \sum_{j\in\mathcal{X}}\omega_{lj}(t)W_{t+1}(\Omega_{-l}, \mathbf{e}_j\mathbf{P}^{(l)}(t))\Big] \\ &\overset{(a)}{=} \Sigma(\mathcal{A} \setminus \{l\}, \Omega_l)\sum_{j\in\mathcal{X}\setminus\{1\}}\Big[(\tilde{\omega}_{lj}(t) - \omega_{lj}(t)) \\ &\quad \times\Big(W_{t+1}(\Omega_{!-l}, \mathbf{e}_j\mathbf{P}^{(l)}(t)) - W_{t+1}(\Omega_{-l}, \mathbf{e}_1\mathbf{P}^{(l)}(t))\Big)\Big], \end{aligned} \tag{33}$$

where, the equality (a) is due to $\omega_{l1}(t) = 1 - \sum_{j=2}^{X}\omega_{lj}(t)$.

By Proposition 4, we have $\mathbf{e}_j\mathbf{P}^{(l)}(t) \le_s \mathbf{e}_1\mathbf{P}^{(l)}(t)$. Then we analyze the term in the bracket, $W_{t+1}(\Omega_{-l}, \mathbf{e}_j\mathbf{P}^{(l)}(t)) - W_{t+1}(\Omega_{-l}, \mathbf{e}_1\mathbf{P}^{(l)}(t))$, of RHS of (16) through three cases:

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

$$\begin{aligned} &-(1 + \sum_{i=1}^{\lfloor\frac{T-t-1}{2}\rfloor}(\beta\overline{\lambda})^{2i})(\mathbf{e}_1 - \mathbf{e}_j)\mathbf{P}^{(l)}(t)\mathbf{r}^{\mathsf{T}} \\ &\le W_{t+1}(\Omega_{-l}, \mathbf{e}_j\mathbf{P}^{(l)}(t)) - W_{t+1}(\Omega_{-l}, \mathbf{e}_1\mathbf{P}^{(l)}(t)) \\ &\le -(1 - \sum_{i=1}^{\lceil\frac{T-t-1}{2}\rceil}(\beta\overline{\lambda})^{2i-1})(\mathbf{e}_1 - \mathbf{e}_j)\mathbf{P}^{(l)}(t)\mathbf{r}^{\mathsf{T}}. \end{aligned} \tag{34}$$

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

$$\begin{aligned} &-\sum_{i=1}^{\lfloor\frac{T-t-1}{2}\rfloor}(\beta\overline{\lambda})^{2i}(\mathbf{e}_1 - \mathbf{e}_j)\mathbf{P}^{(l)}(t)\mathbf{r}^{\mathsf{T}} \\ &\le W_{t+1}(\Omega_{-l}, \mathbf{e}_j\mathbf{P}^{(l)}(t)) - W_{t+1}(\Omega_{-l}, \mathbf{e}_1\mathbf{P}^{(l)}(t)) \\ &\le \sum_{i=1}^{\lceil\frac{T-t-1}{2}\rceil}(\beta\overline{\lambda})^{2i-1}(\mathbf{e}_1 - \mathbf{e}_j)\mathbf{P}^{(l)}(t)\mathbf{r}^{\mathsf{T}}. \end{aligned} \tag{35}$$

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

$$\begin{aligned} &-\sum_{i=1}^{\lfloor\frac{T-t-1}{2}\rfloor}(\beta\overline{\lambda})^{2i}(\mathbf{e}_1 - \mathbf{e}_j)\mathbf{P}^{(l)}(t)\mathbf{r}^{\mathsf{T}} \\ &\le W_{t+1}(\Omega_{-l}, \mathbf{e}_j\mathbf{P}^{(l)}(t)) - W_{t+1}(\Omega_{-l}, \mathbf{e}_1\mathbf{P}^{(l)}(t)) \\ &\le -(1 - \sum_{i=1}^{\lceil\frac{T-t-1}{2}\rceil}(\beta\overline{\lambda})^{2i-1})(\mathbf{e}_1 - \mathbf{e}_j)\mathbf{P}^{(l)}(t)\mathbf{r}^{\mathsf{T}}. \end{aligned} \tag{36}$$

Combining Case 1–3, we obtain the following:

$$\begin{aligned} &-(1 + \sum_{i=1}^{\lfloor\frac{T-t-1}{2}\rfloor}(\beta\overline{\lambda})^{2i})(\mathbf{e}_1 - \mathbf{e}_j)\mathbf{P}^{(l)}(t)\mathbf{r}^{\mathsf{T}} \\ &\le W_{t+1}(\Omega_{-l}, \mathbf{e}_j\mathbf{P}^{(l)}(t)) - W_{t+1}(\Omega_{-l}, \mathbf{e}_1\mathbf{P}^{(l)}(t)) \\ &\le \sum_{i=1}^{\lceil\frac{T-t-1}{2}\rceil}(\beta\overline{\lambda})^{2i-1}(\mathbf{e}_1 - \mathbf{e}_j)\mathbf{P}^{(l)}(t)\mathbf{r}^{\mathsf{T}}. \end{aligned} \tag{37}$$

$$\mathbf{w}_m \prod_{\tau=t}^{t+k-1} \mathbf{P}^{(n)}(\tau) = \left[\sum_{\tau=t}^{t+k-1} c_1^n(\tau) \prod_{\tau'=\tau+1}^{t+k-1} \lambda^{(n)}(\tau') + \prod_{\tau=t}^{t+k-1} \lambda^{(n)}(\tau)\omega_{m1}, \cdots, \sum_{\tau=t}^{t+k-1} c_X^n(\tau) \prod_{\tau'=\tau+1}^{t+k-1} \lambda^{(n)}(\tau') + \prod_{\tau=t}^{t+k-1} \lambda^{(n)}(\tau)\omega_{mX} \right] \tag{29}$$

$$\mathbf{w}_l \prod_{\tau=t}^{t+k-1} \mathbf{P}^{(n)}(\tau) = \left[\sum_{\tau=t}^{t+k-1} c_1^n(\tau) \prod_{\tau'=\tau+1}^{t+k-1} \lambda^{(n)}(\tau') + \prod_{\tau'=\tau}^{t+k-1} \lambda^{(n)}(\tau')\omega_{l1}, \cdots, \sum_{\tau=t}^{t+k-1} c_X^n(\tau) \prod_{\tau'=\tau+1}^{t+k-1} \lambda^{(n)}(\tau') + \prod_{\tau=t}^{t+k-1} \lambda^{(n)}(\tau)\omega_{lX} \right] \tag{30}$$

Therefore, combining (33) and (37), we have the upper bound of (B1)

$$\begin{aligned}
&W_t^{\mathcal{A}'}(\Omega_l') - W_t^{\mathcal{A}}(\Omega_l) \\
&= (\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t))\mathbf{r}^{\mathsf{T}} + \beta(F(\mathcal{A}', \Omega_l') - F(\mathcal{A}, \Omega_l)) \\
&\le (\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t))\mathbf{r}^{\mathsf{T}} \\
&\quad + \beta \sum_{j=2}^{X} \Big[(\omega_{lj}'(t) - \omega_{lj}(t)) \\
&\quad \times (- \sum_{i=1}^{\lceil \frac{T-t-1}{2} \rceil} (\beta\overline{\lambda})^{2i-1} (\mathbf{e}_j - \mathbf{e}_1)\mathbf{P}^{(l)}(t)\mathbf{r}^{\mathsf{T}} \Big] \\
&= (\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t))\mathbf{r}^{\mathsf{T}} \\
&\quad - \beta \Big[\sum_{i=1}^{\lceil \frac{T-t-1}{2} \rceil} (\beta\overline{\lambda})^{2i-1} (\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t))\mathbf{P}^{(l)}(t)\mathbf{r}^{\mathsf{T}} \Big] \\
&\overset{(a)}{\le} (\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t))\mathbf{r}^{\mathsf{T}} \\
&\quad + \beta \Big[\sum_{i=1}^{\lceil \frac{T-t-1}{2} \rceil} (\beta\overline{\lambda})^{2i-1} (\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t))(\overline{\lambda}\mathbf{E})\mathbf{r}^{\mathsf{T}} \Big] \\
&= (\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t))\mathbf{r}^{\mathsf{T}} + \sum_{i=1}^{\lceil \frac{T-t-1}{2} \rceil} (\beta\overline{\lambda})^{2i} (\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t))\mathbf{r}^{\mathsf{T}} \\
&= (\mathbf{w}_l'(t) - \mathbf{w}_l(t))\mathbf{r}^{\mathsf{T}} + \sum_{i=1}^{\lfloor \frac{T-t}{2} \rfloor} (\beta\overline{\lambda})^{2i} (\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t))\mathbf{r}^{\mathsf{T}} \\
&= (1 + \sum_{i=1}^{\lfloor \frac{T-t}{2} \rfloor} (\beta\overline{\lambda})^{2i})(\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t))\mathbf{r}^{\mathsf{T}},
\end{aligned} \tag{38}$$

where, the inequality (a) is due to Assumption 2 and Proposition 4.

Then, the lower bound of (B1)

$$\begin{aligned}
&W_t^{\mathcal{A}'}(\Omega_l') - W_t^{\mathcal{A}}(\Omega_l) \\
&= (\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t))\mathbf{r}^{\mathsf{T}} + \beta(F(\mathcal{A}', \Omega_l') - F(\mathcal{A}, \Omega_l)) \\
&\ge (\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t))\mathbf{r}^{\mathsf{T}} + \beta \sum_{j \in \mathcal{X} - \{1\}} \Big[(\tilde{\omega}_{lj}(t) - \omega_{lj}(t)) \\
&\quad \times \Big[- (1 + \sum_{i=1}^{\lfloor \frac{T-t-1}{2} \rfloor} (\beta\overline{\lambda})^{2i})(\mathbf{e}_1 - \mathbf{e}_j)\mathbf{P}^{(l)}(t)\mathbf{r}^{\mathsf{T}} \Big] \Big] \\
&= (\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t))\mathbf{r}^{\mathsf{T}} \\
&\quad - \beta \Big[(1 + \sum_{i=1}^{\lfloor \frac{T-t-1}{2} \rfloor} (\beta\overline{\lambda})^{2i})(\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t))\mathbf{P}^{(l)}(t)\mathbf{r}^{\mathsf{T}} \Big] \\
&= (\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t))\mathbf{r}^{\mathsf{T}} \\
&\quad - \beta \Big[(1 + \sum_{i=1}^{\lfloor \frac{T-t-1}{2} \rfloor} (\beta\overline{\lambda})^{2i})(\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t))(\lambda^{(l)}\mathbf{E})\mathbf{r}^{\mathsf{T}} \Big] \\
&\ge (\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t))\mathbf{r}^{\mathsf{T}} \\
&\quad - \beta \Big[(1 + \sum_{i=1}^{\lfloor \frac{T-t-1}{2} \rfloor} (\beta\overline{\lambda})^{2i})(\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t))(\overline{\lambda}\mathbf{E})\mathbf{r}^{\mathsf{T}} \Big] \\
&= (1 - \beta\overline{\lambda} - \sum_{i=1}^{\lfloor \frac{T-t-1}{2} \rfloor} (\beta\overline{\lambda})^{2i+1})(\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t))\mathbf{r}^{\mathsf{T}} \\
&= (1 - \sum_{i=0}^{\lfloor \frac{T-t-1}{2} \rfloor} (\beta\overline{\lambda})^{2i+1})(\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t))\mathbf{r}^{\mathsf{T}} \\
&= (1 - \sum_{i=1}^{\lceil \frac{T-t}{2} \rceil} (\beta\overline{\lambda})^{2i-1})(\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t))\mathbf{r}^{\mathsf{T}}.
\end{aligned} \tag{39}$$

Combining (38) and (39), we complete the proof of the first part, $l \in \mathcal{A}'$, $l \in \mathcal{A}$, of Lemma 4.

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:

$$\begin{aligned}
F(\mathcal{A}', \Omega_l') = \Sigma(\mathcal{A}(t) \setminus \{k\}, \Omega_l') \Big[\sum_{j \in \mathcal{X}} \omega_{kj}(t) W_{t+1}(\Omega_{-k}', \\
\times e_j \mathbf{P}^{(k)}(t)) \Big],
\end{aligned} \tag{40}$$

$$\begin{aligned}
F(\mathcal{A}, \Omega_l) = \Sigma(\mathcal{A}(t) \setminus \{k\}, \Omega_l) \Big[\sum_{j \in \mathcal{X}} \omega_{kj}(t) W_{t+1}(\Omega_{-k}, \\
\times e_j \mathbf{P}^{(k)}(t)) \Big].
\end{aligned} \tag{41}$$

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

$$\begin{aligned}
&F(\mathcal{A}', \Omega_l') - F(\mathcal{A}, \Omega_l) \\
&= \Sigma(\mathcal{A}(t) \setminus \{k\}, \Omega_l) \Big[\sum_{j \in \mathcal{X}} \omega_{kj}(t) \\
&\quad \times \Big(W_{t+1}(\Omega_{-k}', \mathbf{e}_j \mathbf{P}^{(k)}(t)) - W_{t+1}(\Omega_{-k}, \mathbf{e}_j \mathbf{P}^{(k)}(t)) \Big) \Big].
\end{aligned} \tag{42}$$

For the term in the bracket of RHS of (23), if channel l is never chosen for $W_{t+1}(\Omega'_{-k}, \mathbf{e}_j\mathbf{P}^{(k)}(t))$ and $W_{t+1}(\Omega_{-k}, \mathbf{e}_j\mathbf{P}^{(k)}(t))$ 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 $W_{t+1}(\Omega'_{-k}, \mathbf{e}_j\mathbf{P}^{(k)}(t)) - W_{t+1}(\Omega_{-k}, \mathbf{e}_j\mathbf{P}^{(k)}(t)) = 0$; 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 $l \in \mathcal{A}(t^0)$;

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 the upper bound only when $\tilde{\mathbf{w}}_l(t^o) \le_s \mathbf{w}_l(t^o)$; that is, $t^0 \ge t+2$.

$$\begin{aligned}
&W_{t^0}(\Omega'_l(t^o)) - W_{t^0}(\Omega_l(t^o)) \\
&\le (1 + \sum_{i=1}^{\lfloor \frac{T-t^0}{2} \rfloor} (\beta\overline{\lambda})^{2i})(\tilde{\mathbf{w}}_l(t^o) - \mathbf{w}_l(t^o))\mathbf{r}^{\mathsf{T}} \\
&= (1 + \sum_{i=1}^{\lfloor \frac{T-t^0}{2} \rfloor} (\beta\overline{\lambda})^{2i})(\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t))\left(\mathbf{P}^{(l)}(t)\right)^{t^0-t}\mathbf{r}^{\mathsf{T}} \\
&\overset{(a)}{=} (1 + \sum_{i=1}^{\lfloor \frac{T-t^0}{2} \rfloor} (\beta\overline{\lambda})^{2i})(\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t))(\lambda^{(l)}\mathbf{E})^{t^0-t}\mathbf{r}^{\mathsf{T}} \\
&\overset{(b)}{\le} (1 + \sum_{i=1}^{\lfloor \frac{T-t-2}{2} \rfloor} (\beta\overline{\lambda})^{2i})(\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t))(\overline{\lambda})^2\mathbf{r}^{\mathsf{T}} \\
&= \overline{\lambda}^2(1 + \sum_{i=1}^{\lfloor \frac{T-t-2}{2} \rfloor} (\beta\overline{\lambda})^{2i})(\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t))\mathbf{r}^{\mathsf{T}},
\end{aligned}$$

where, the equality (a) is due to Assumption 3 and Proposition 1, and (b) is due to $t^0 = t+2$.

The lower bound is achieved when $\tilde{\mathbf{w}}_l(t^o) \le_s \mathbf{w}_l(t^o)$; that is, $t^0 \ge t+1$. Thus,

$$\begin{aligned}
&W_{t^0}(\Omega'_l(t^o)) - W_{t^0}(\Omega_l(t^o)) \\
&\ge (1 + \sum_{i=1}^{\lfloor \frac{T-t^0}{2} \rfloor} (\beta\overline{\lambda})^{2i})(\tilde{\mathbf{w}}_l(t^o) - \mathbf{w}_l(t^o))\mathbf{r}^{\mathsf{T}} \\
&= (1 + \sum_{i=1}^{\lfloor \frac{T-t^0}{2} \rfloor} (\beta\overline{\lambda})^{2i})(\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t))\left(\mathbf{P}^{(l)}(t)\right)^{t^0-t}\mathbf{r}^{\mathsf{T}} \\
&\overset{(a)}{=} (1 + \sum_{i=1}^{\lfloor \frac{T-t^0}{2} \rfloor} (\beta\overline{\lambda})^{2i})(\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t))(\lambda^{(l)}\mathbf{E})^{t^0-t}\mathbf{r}^{\mathsf{T}} \\
&\overset{(b)}{\ge} (1 + \sum_{i=1}^{\lfloor \frac{T-t-1}{2} \rfloor} (\beta\overline{\lambda})^{2i})(\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t))(\lambda^{(l)}\mathbf{E})\mathbf{r}^{\mathsf{T}} \\
&\ge -\overline{\lambda}(1 + \sum_{i=1}^{\lfloor \frac{T-t-1}{2} \rfloor} (\beta\overline{\lambda})^{2i})(\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t))\mathbf{r}^{\mathsf{T}},
\end{aligned}$$

where, the equality (a) is due to Assumption 2 and Proposition 1, and (b) is due to $t^0 = t+1$.

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

$$\begin{aligned}
&W_{t+1}(\Omega'_{-k}, \mathbf{e}_j\mathbf{P}^{(k)}(t)) - W_{t+1}(\Omega_{-k}, \mathbf{e}_j\mathbf{P}^{(k)}(t)) \\
&= \beta^{t^0-t-1}[W_{t^0}(\Omega'_l(t^o)) - W_{t^0}(\Omega_l(t^o))] \\
&\le \beta\overline{\lambda}^2(1 + \sum_{i=1}^{\lfloor \frac{T-t-2}{2} \rfloor} (\beta\overline{\lambda})^{2i})(\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t))\mathbf{r}^{\mathsf{T}}, \qquad (43)
\end{aligned}$$

where, the inequality is due to $t^0 = t+2$.

$$\begin{aligned}
&W_{t+1}(\Omega'_{-k}, \mathbf{e}_j\mathbf{P}^{(k)}(t)) - W_{t+1}(\Omega_{-k}, \mathbf{e}_j\mathbf{P}^{(k)}(t)) \\
&= \beta^{t^0-t-1}[W_{t^0}(\Omega'_l(t^o)) - W_{t^0}(\Omega_l(t^o))] \\
&\ge -\overline{\lambda}(1 + \sum_{i=1}^{\lfloor \frac{T-t-1}{2} \rfloor} (\beta\overline{\lambda})^{2i})(\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t))\mathbf{r}^{\mathsf{T}}, \qquad (44)
\end{aligned}$$

where, the inequality is due to $t^0 = t+1$.

Combing (43) and (44), we have

$$\begin{aligned}
&W_t^{\mathcal{A}'}(\Omega'_l) - W_t^{\mathcal{A}}(\Omega_l) \\
&= \beta(F(\mathcal{A}', \Omega'_l) - F(\mathcal{A}, \Omega_l)) \\
&\le (\beta\overline{\lambda})^2(1 + \sum_{i=1}^{\lfloor \frac{T-t-2}{2} \rfloor} (\beta\overline{\lambda})^{2i})(\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t))\mathbf{r}^{\mathsf{T}} \\
&= \sum_{i=1}^{\lfloor \frac{T-t}{2} \rfloor} (\beta\overline{\lambda})^{2i}(\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t))\mathbf{r}^{\mathsf{T}},
\end{aligned}$$

and

$$\begin{aligned}
W_t^{\mathcal{A}'}(\Omega'_l) - W_t^{\mathcal{A}}(\Omega_l) &= \beta(F(\mathcal{A}', \Omega'_l) - F(\mathcal{A}, \Omega_l)) \\
&\ge -\beta\overline{\lambda}(1 + \sum_{i=1}^{\lfloor \frac{T-t-1}{2} \rfloor} (\beta\overline{\lambda})^{2i})(\tilde{\mathbf{w}}_l(t) - \mathbf{w}_l(t))\mathbf{r}^{\mathsf{T}} \\
&= -\sum_{i=1}^{\lceil \frac{T-t}{2} \rceil} (\beta\overline{\lambda})^{2i-1}(\tilde{\mathbf{w}}_l - \mathbf{w}_l)\mathbf{r}^{\mathsf{T}}
\end{aligned}$$

which completes the proof of Lemma 4 when $l \notin \mathcal{A}'$ and $l \notin \mathcal{A}$.

The proof of the third case $l \in \mathcal{A}'(t)$ and $l \notin \mathcal{A}(t)$ is similar with the corresponding part of Lemma 2.

To the end, Lemma 4 is concluded.

REFERENCES

[1] P. Whittle, "Restless bandits: Activity allocation in a changing world," *J. Appl. Probab.*, vol. 25, pp. 287–298, Jan. 1988.

[2] J. Gittins, K. Glazebrook, and R. Webber, *Multi-Armed Bandit Allocation Indices*. Oxford, U.K.: Blackwell, 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 J. Sel. Areas Commun.*, vol. 25, no. 3, pp. 589–600, Apr. 2007.

[4] J. C. Gittins and D. M. Jones, "A dynamic allocation index for the sequential design of experiments," *Progress Statist.*, pp. 241–266, Jan. 1972.

[5] J. C. Gittins, "Bandit processes and dynamic allocation indices," *J. Roy. Statist. Soc.*, vol. 41, no. 2, pp. 148–177, 1979.

[6] C. H. Papadimitriou and J. N. Tsitsiklis, "The complexity of optimal queuing network control," *Math. Oper. Res.*, vol. 24, no. 2, pp. 293–305, May 1999.
[7] S. Guha and K. Munagala, "Approximation algorithms for partial-information based stochastic control with Markovian rewards," in *Proc. IEEE FOCS*, Providence, RI, USA, Oct. 2007, pp. 483–493.
[8] S. Guha, P. Shi, and K. Munagala, "Approximation algorithms for restless bandit problems," in *Proc. ACM-SIAM Symp. Discrete Algorithms (SODA)*, New York, NY, USA, P. 28–37, Jan. 2009.
[9] D. Bertsimas and J. E. Niño-Mora, "Restless bandits, linear programming relaxations, and a primal-dual index heuristic," *Oper. Res.*, vol. 48, no. 1, pp. 80–90, 2000.
[10] Q. Zhao, B. Krishnamachari, and K. Liu, "On myopic sensing for multi-channel opportunistic access: Structure, optimality, and performance," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 5413–5440, Dec. 2008.
[11] S. H. A. Ahmad, 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.
[12] S. H. A. Ahmad and M. Liu, "Multi-channel opportunistic access: A case of restless bandits with multiple plays," in *Proc. Allerton Conf.*, Monticello, Il, USA, Sep. 2009, pp. 1361–1368.
[13] K. Liu, Q. Zhao, and B. Krishnamachari, "Dynamic multichannel access with imperfect channel state detection," *IEEE Trans. Signal Process.*, vol. 58, no. 5, pp. 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 Trans. Signal Process.*, vol. 60, no. 1, pp. 300–309, Jan. 2012.
[15] S. H. A. Ahmad, 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.
[16] F. E. Lapiccirella, K. Liu, and Z. Ding, "Multi-channel opportunistic access based on primary ARQ messages overhearing," in *Proc. IEEE ICC*, Kyoto, Japan, Jun. 2011, pp. 1–5.
[17] K. Wang, Q. Liu, and F. C. M. Lau, "Multichannel opportunistic access by overhearing primary ARQ messages," *IEEE Trans. Veh. Technol.*, vol. 62, no. 7, pp. 3486–3492, Sep. 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 J. Sel. Topics Signal Process.*, vol. 7, no. 5, pp. 869–882, Oct. 2013.
[19] Y. Ouyang and D. Teneketzis, "On the optimality of myopic sensing in multi-state channels," *IEEE Trans. Inf. Theory*, vol. 60, no. 1, pp. 681–696, Jan. 2014.
[20] K. Wang, L. Chen, J. Yu, and D. Zhang, "Optimality of myopic policy for multistate channel access," *IEEE Commun. Lett.*, vol. 20, no. 2, pp. 300–303, Feb. 2016.
[21] K. Wang, L. Chen, and J. Yu, "On optimality of myopic policy in multi-channel opportunistic access," in *Proc. IEEE ICC*, Kuala Lumpur, Malaysia, May 2016, pp. 1–6.
[22] A. Müller and D. Stoyan, *Comparison Methods for Stochastic Models Risk*. New York, NY, USA: Wiley, 2002.

Kehao Wang received the B.S. degree in electrical engineering and the M.S. degree in communication and information system from the Wuhan University of Technology, Wuhan, China, in 2003 and 2006, respectively, and the Ph.D. degree from the Department of Computer Science, University of Paris-Sud XI, Orsay, France, in 2012. In 2013, he held a post-doctoral position with The Hong Kong Polytechnic University. Since 2013, he has been with the School of Information Engineering, Wuhan University of Technology, where he is currently an Associate Professor. Since 2015, he has been a Visiting Scholar with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA, USA. His research interests are stochastic optimization, operation research, scheduling, wireless network communications, and embedded operating system.

Lin Chen received the B.E. degree in radio engineering from Southeast University, China, in 2002, the M.S. degree in networking from the University of Paris 6, and the Engineer Diploma and Ph.D. degrees from Telecom ParisTech, Paris, in 2005 and 2008, respectively. He is currently an Associate Professor with the Department of Computer Science, University of Paris-Sud XI. His main research interests include modeling and control for wireless networks, security and cooperation enforcement in wireless networks, and game theory.

Jihong Yu received the B.E. degree in communication engineering and the M.E. degree in communication and information systems from the Chongqing University of Posts and Telecommunications, Chongqing, China, in 2010 and 2013, respectively. He is currently pursuing the Ph.D. degree in computer science with the University of Paris-Sud XI, Orsay, France. His research interests include wireless communications and networking and RFID technologies.