On Optimality of Myopic Sensing Policy with Imperfect Sensing in Multi-channel Opportunistic Access

Kehao Wang Lin Chen Quan Liu Khaldoun Al Agha

Abstract

We consider the channel access problem under imperfect sensing of channel state in a multi-channel opportunistic communication system, where the state of each channel evolves as an independent and identically distributed Markov process. The considered problem can be cast into a restless multi-armed bandit (RMAB) problem that is of fundamental importance in decision theory. It is well-known that solving the RMAB problem is PSPACE-hard, with the optimal policy usually intractable due to the exponential computation complexity. A natural alternative is to consider the easily implementable myopic policy that maximizes 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 optimality of the myopic policy under imperfect sensing for the considered RMAB problem. Specifically, for a family of generic and practically important utility functions, we establish the closed-form conditions under which the myopic policy is guaranteed to be optimal even under imperfect sensing. Despite our focus on the opportunistic channel access, the obtained results are generic in nature and are widely applicable in a wide range of engineering domains.

Index Terms

Restless multi-armed bandit (RMAB) problem, myopic policy, imperfect sensing, opportunistic spectrum access (OSA)

K. Wang, L. Chen and K. Al Agha are with the Laboratoire de Recherche en Informatique (LRI), Department of Computer Science, the University of Paris-Sud XI, 91405 Orsay, France (e-mail: {Kehao.Wang, Lin.Chen, Khaldoun.Alagha}@lri.fr). K. Wang and Q. Liu is with the school of Information Engineering, Wuhan University of Technology, 430070 Hubei, China (e-mail: {Kehao.wang, Quan.Liu}@whut.edu.cn).

I. Introduction

We consider an opportunistic multi-channel communication system in which a user has access to multiple channels, but is limited to sense and transmit only on a subset of them at a time. The fundamental problem we study is how the sender can exploit past observations and the knowledge of the stochastic properties of the channels to maximize its utility (e.g., expected throughput) by switching opportunistically across channels.

Formally, the considered channel access problem can be cast into the restless multi-armed bandit (RMAB) problem, one of the most well-known generalizations of the classic multi-armed bandit (MAB) problem, which is of fundamental importance in stochastic decision theory. The standard formulation of the RMAB problem can be briefly summarized as follows: There is a bandit of N independent arms, each evolving as a two-state Markov process. At each time slot, a player chooses k ($1 \leq k \leq N$) of the N arms to play and receives a certain amount of reward depending on the state of the played arms. Given the initial state of the system, the goal of the player is to find the optimal policy of playing the k arms at each slot so as to maximize the aggregated discounted long-term reward.

Despite the significant research efforts in the field, the RMAB problem in its generic form still remains open. Until today, very little result is reported on the structure of the optimal policy. Obtaining the optimal policy for a general RMAB problem is often intractable due to the exponential computation complexity. Hence, a natural alternative is to seek a simple myopic policy maximizing the short-term reward. Due to its simple and robust structure, the myopic sensing policy has begun to attract significant research attention, especially on the optimality of the myopic sensing policy.

The vast majority of studies in the area assume perfect observation of channel states. However, sensing or observation errors are inevitable in practical scenario (e.g., due to noise and system limitations), especially in wireless communication systems which is the focus of our work. More specifically, a good (bad, respectively) channel may be sensed as bad (good) and accessing a bad channel leads to zero reward. In such context, it is crucial to study the structure and the optimality of the myopic sensing policy with imperfect observation. We would like to emphasize that the presence of sensing error brings two difficulties when studying the myopic sensing policy in this new context.

- The channel state evolves as a non-linear mapping (w.r.t. the current channel state) instead of a linear one in the perfect sensing case.
- In the non-perfect sensing case, the state transition of a channel depends not only on the channel evolution itself, but also on the observation outcome, meaning that the transition is not deterministic.

Due to the above particularities[1], our problem requires an original study on the optimality of the myopic sensing policy that cannot draw on existing results in the perfect sensing case. We would like to report that despite its practical importance and particularities, very few work has been done on the impact of sensing error on the performance of the myopic sensing policy, or more generically, on the RMAB problem under imperfect observation. To the best of our knowledge, [1] is the only work in this area, where the optimality of the myopic policy is proved for the case of two channels with a particular utility function. In this paper, we derive closed-form conditions under which the myopic sensing policy is optimal under imperfect sensing for arbitrary N and generic utility functions. As shown in Section III-C, the result obtained in this paper can cover the result of [1]. Moreover, this paper also significantly extends our previous work [2], focusing on perfect sensing scenario in which the analysis cannot be applied in the imperfect sensing scenario due to the non-trivial particularities introduced by sensing error as mentioned previously. In this regard, our work in this paper contributes the existing literature by developing an adapted analysis on the RMAB problem under imperfect sensing under the generic framework proposed in [2].

The rest of the paper is organized as follows: Our model is formulated in Section II. Section III studies the optimality of the myopic sensing policy and illustrates the application of the derived results via two typical examples. A detailed discussion on the related work is given in Section IV. Finally, the paper is concluded by Section V.

II. Problem Formulation

A. Multi-channel Opportunistic Access with Imperfect Sensing

As outlined in the Introduction, we consider a multi-channel opportunistic communication system, in which a user is able to access a set $\mathcal{N}$ of N independent and statistically identical channels, each characterized by a Markov chain of two states, *good/idle* (1) and *bad/busy* (0). The state transmission probabilities are given by $\{p_{i,j}\}, i, j = 0, 1$. We assume that the system operates in a synchronously time slotted fashion with the time slot indexed by t ($t = 1, 2, \cdots, T$), where T is the time horizon of interest. Each channel goes through state transition at the beginning of each slot t. This generic multi-channel opportunistic communication model can be naturally cast into the opportunistic spectrum access (OSA) problem in cognitive radio systems where an unlicensed secondary user can opportunistically access the

[1]Please refer to the remark of (1) for a detailed analysis

temporarily unused channels of the licensed primary users, with the availability of each channel evolving as an independent Markov chain.

Limited by hardware constraints and energy cost, the user is allowed to sense only k ($1 \leq k \leq N$) of the N channels at each slot t. We denote the set of channels chosen by the user at slot t by $\mathcal{A}(t)$ where $\mathcal{A}(t) \in \mathcal{N}$ and $|\mathcal{A}(t)| = k$. We assume that the user makes the channel selection decision at the beginning of each slot after the channel state transition. Moreover, we are interested in the imperfect sensing scenario where channel sensing is subject to errors, i.e., a good channel may be sensed as bad one and vice versa. Let $\mathbf{S}(t) \triangleq [S_1(t), \cdots, S_N(t)]$ denote the channel state vector where $S_i(t) \in \{0, 1\}$ is the state of channel i in slot t and let $\mathbf{S}'(t) \triangleq \{S'_i(t), i \in A(t)\}$ denote the sensing outcome vector where $S'_i(t) = 0$ (1) means that the channel i is sensed bad (good) in slot t. Using such notation, the performance of channel state detection is characterized by two system parameters: the probability of false alarm $\epsilon_i(t)$ and the probability of miss detection $\delta_i(t)$, formally defined as follows:

$$\epsilon_i(t) \triangleq \Pr\{S'_i(t) = 1 | S_i(t) = 0\},$$

$$\delta_i(t) \triangleq \Pr\{S'_i(t) = 0 | S_i(t) = 1\}.$$

In our analysis, we consider the case where $\epsilon_i(t)$ and $\delta_i(t)$ are independent w.r.t. t and i. More specifically, we defined ϵ and δ as the system-wide false alarm rate and miss detection rate. We also assume that when the receiver successfully receives a packet from a channel, it sends an acknowledgement to the transmitter over the same channel at the end of the slot. The absence of an ACK signifies that the transmitter does not transmit over this channel or transmitted but the channel is busy in this slot.

Obviously, by sensing only k out of N channels, the user cannot observe the state information of the whole system. Hence, the user has to infer the channel states from its past decision and observation history so as to make its future decision. To this end, we define the *channel state belief vector* (hereinafter referred to as *belief vector* for briefness) $\Omega(t) \triangleq \{\omega_i(t), i \in \mathcal{N}\}$, where $0 \leq \omega_i(t) \leq 1$ is the conditional probability that channel i is in state good (i.e., $S_i(t) = 1$) at slot t given all past states, actions and observations[2]. Due to the Markovian nature of the channel model, the belief vector can be updated recursively using Bayes Rule as shown in (1).

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

[2]The initial belief $\omega_i(1)$ can be set to $\frac{p_{01}}{p_{01}+1-p_{11}}$ if no information about the initial system state is available.

where $ACK = 1$ denotes the case where an ACK is received (successful transmission, i.e., $S_i'(i) = 1$ and $S_i(t) = 1$) and $ACK = 0$ denotes the case where no ACK is received (failed transmission or no transmission, i.e., $S_i'(i) = 1$ $S_i(t) = 0$ or $S'(t) = 0$), $\varphi(\omega_i) = \frac{\epsilon\omega_i(t)}{\epsilon\omega_i(t)+1-\omega_i(t)}$ and

$$\tau(\omega_i(t)) \triangleq \omega_i(t)p_{11} + [1 - \omega_i(t)]p_{01} \tag{2}$$

denotes the operator for the one-step belief update.

Remark. We would like to emphasize that in contrast to the perfect sensing case [2] where $\omega_i(t+1)$ is a linear function of $\omega_i(t)$ whether i in sensed or not, in the imperfect sensing case, the mapping from $\omega_i(t)$ to $\omega_i(t+1)$ is no longer linear due to the sensing error (cf. the second line of equation (1)). Moreover, the state transition of a channel depends not only on the channel evolution itself, but also on the observation outcome, i.e., $\omega_i(t+1) = p_{11}$ for $i \in \mathcal{A}(t), ACK = 1$ and $\omega_i(t+1) = \tau(\varphi(\omega_i(t)))$ for $i \in \mathcal{A}(t), ACK = 0$. As will be shown later, these differences make the analysis for the imperfect sensing more complicated.

To conclude this subsection, we state some structural properties of $\tau(\omega_i(t))$ and $\varphi(\omega_i(t))$ that are useful in the subsequent proofs.

Lemma 1. *If* $\forall i$, $p_{01} < p_{11}$, *then*

- $\tau(\omega_i(t))$ *is monotonically increasing in* $\omega_i(t)$*;*
- $p_{01} \leq \tau(\omega_i(t)) \leq p_{11}$, $\forall\ 0 \leq \omega_i(t) \leq 1$.

Proof: Lemma 1 follows from $\tau(\omega_i(t)) = (p_{11} - p_{01})\omega_i(t) + p_{01}$ straightforwardly. ■

Lemma 2. *If* $0 \leq \epsilon \leq \frac{(1-p_{11})p_{01}}{p_{11}(1-p_{01})}$, *then*

- $\varphi(\omega_i(t))$ *increases monotonically in* $\omega_i(t)$ *with* $\varphi(0) = 0$ *and* $\varphi(1) = 1$*;*
- $\varphi(\omega_i(t)) \leq p_{01}$, $\forall p_{01} \leq \omega_i(t) \leq p_{11}$.

Proof: Noticing that $\varphi(\omega_i) = \frac{\epsilon\omega_i(t)}{\epsilon\omega_i(t)+1-\omega_i(t)}$, Lemma 2 follows straightforwardly. ■

B. Optimal Sensing Problem Formulation and Myopic Sensing Policy

Given the imperfect sensing context, we are interested in the user's optimization problem to find the optimal sensing policy π^* that maximizes the expected total discounted reward over a finite horizon. Mathematically, a sensing policy π is defined as a mapping from the belief vector $\Omega(t)$ to the action (i.e., the set of channels to sense) $\mathcal{A}(t)$ in each slot t: $\pi:\ \Omega(t) \rightarrow \mathcal{A}(t), |\mathcal{A}(t)| = k,\ t = 1, 2, \cdots, T$.

The following gives the formal definition of the optimal sensing problem:

$$\pi^* = \underset{\pi}{\operatorname{argmax}}\, \mathbb{E}\left[\sum_{t=1}^{T} \beta^t R_\pi(\Omega(t)) \middle| \Omega(1)\right] \tag{3}$$

where $R_\pi(\Omega(t))$ is the reward collected in slot t under the sensing policy π with the initial belief vector $\Omega(1)$, $0 \leq \beta \leq 1$ is the discounting factor characterizing the feature that the future rewards are less valuable than the immediate reward. By treating the belief value of each channel as the state of each arm of a bandit, the user's optimization problem can be cast into a restless multi-armed bandit problem.

In order to get more insight on the structure of the optimization problem formulated in (3) and the complexity to solve it, we derive the dynamic programming formulation of (3) as follows:

$$V_T(\Omega(t)) = \max_{\pi} \mathbb{E}[R_\pi(\Omega(T))] = \max_{\substack{\mathcal{A}(T)\subseteq\mathcal{N} \\ |\mathcal{A}(T)|=k}} \mathbb{E}[R_\pi(\Omega(T))],$$

$$V_t(\Omega(t)) = \max_{\substack{\mathcal{A}(t)\subseteq\mathcal{N} \\ |\mathcal{A}(t)|=k}} \mathbb{E}\left[R_\pi(\Omega(t)) + \beta \sum_{\mathcal{E}\subseteq\mathcal{A}(t)} \prod_{i\in\mathcal{E}} (1-\epsilon)\omega_i(t) \right.$$

$$\left. \prod_{j\in\mathcal{A}(t)\setminus\mathcal{E}} [1-(1-\epsilon)\omega_j(t)] V_{t+1}(\Omega(t+1))\right].$$

In the above equations, $V_t(\Omega(t))$ is the value function corresponding to the maximal expected reward from time slot t to T ($1 \leq t \leq T$) with the believe vector $\Omega(t+1)$ following the evolution described in (1) given that the channels in the subset $\mathcal{E}$ are sensed in state good and the channels in $\mathcal{A}(t)\setminus\mathcal{E}$ are sensed in state bad.

Theoretically, the optimal policy can be obtained by solving the above dynamic programming. Unfortunately, due to the impact of the current action on the future reward and the unaccountable space of the belief vector, obtaining the optimal solution directly from the above recursive equations is computationally prohibitive. Hence, a natural alternative is to seek simple myopic sensing policy which is easy to compute and implement that maximizes the expected immediate reward $F(\Omega(t))$, formally defined as follows:

$$\mathcal{A}(t) = \underset{\mathcal{A}(t)\subseteq\mathcal{N}}{\operatorname{argmax}}\, \Sigma_{i\in\mathcal{A}(t)} F(\Omega(t)). \tag{4}$$

In this paper, we focus on a class of generic and practically important functions defined in [2] as *regular* functions. More specifically, the expected immediate reward function $F(\Omega(t))$ studied in this paper are assumed to be symmetrical, monotonically non-decreasing and decomposable, defined by the three axioms in [2]. Under this condition, the myopic policy consists of choosing the k channels with the largest value of ω. In the following sections we focus on the structure and the optimality of the myopic

sensing policy under imperfect sensing. As pointed out in the remark following equations (1) and (2), the main technical difficulties compared with the perfect sensing case are the non-linearity of the mapping from $\omega_i(t)$ to $\omega_i(t+1)$ and the dependency of the channel state transition on the observation outcome.

III. Analysis on Optimality of Myopic Sensing Policy under Imperfect Sensing

The goal of this section is to establish closed-form conditions under which the myopic sensing policy, despite of its simple structure, achieves the system optimum under imperfect sensing. To this end, we set up by defining an auxiliary function and studying the structural properties of the auxiliary function, which serve as a basis in the study of the optimality of the myopic sensing policy. We then establish the main result on the optimality followed by the illustration on how the obtained result can be applied via two concrete application examples.

For the convenience of discussion, we firstly state some notations before presenting the analysis:

- The believe vector $\Omega(t)$ is sorted to $[\omega_1(t), \cdots, \omega_N(t)]$ at each slot t such that $\mathcal{A} = \{1, 2, \cdots, k\}$ [3];
- $\mathcal{N}(m) \triangleq \{1, \cdots, m\}$ $(m \leq N)$ denotes the first m channels in $\mathcal{N}$;
- Given $\mathcal{E} \subseteq \mathcal{M} \subseteq \mathcal{N}$, $Pr(\mathcal{M}, \mathcal{E}) \triangleq \prod_{i \in \mathcal{E}} (1-\epsilon)\omega_i(t) \prod_{j \in \mathcal{M} \setminus \mathcal{E}} [1-(1-\epsilon)\omega_j(t)]$, herein, $Pr(\mathcal{M}, \mathcal{E})$ denotes the expected probability that the channels in $\mathcal{E}$ are sensed in the good state, while the channels in $\mathcal{M} \setminus \mathcal{E}$ are sensed in the bad state, given that the channels in $\mathcal{M}$ are sensed;
- $\mathbf{P}_{\mathbf{11}}^{\mathcal{E}}$ denotes the vector of length $|\mathcal{E}|$ with each element being p_{11};
- $\mathbf{\Phi}(l, m) \triangleq [\tau(\omega_i(t)), l \leq i \leq m]$ where the components are sorted by channel index. $\mathbf{\Phi}(l, m)$ characterizes the updated belief values of the channels between l and m if they are not sensed;
- Given $\mathcal{E} \subseteq \mathcal{M} \subseteq \mathcal{N}$, $\mathbf{Q}^{\mathcal{M},\mathcal{E}} \triangleq [\tau(\varphi(\omega_i(t))), i \in \mathcal{M} \setminus \mathcal{E}]$ where the components are sorted by channel index. $\mathbf{Q}^{\mathcal{M},\mathcal{E}}$ characterizes the updated belief values of the channels in $\mathcal{M} \setminus \mathcal{E}$ if they are sensed in the bad state; $\overline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}} \triangleq [\tau(\varphi(\omega_i(t))), i \in \mathcal{M} \setminus \mathcal{E} \text{ and } i < l]$ characterizes the updated belief values of the channels in $\mathcal{M} \setminus \mathcal{E}$ if they are sensed in the bad state with the channel index smaller than l; $\underline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}} \triangleq [\tau(\varphi(\omega_i(t))), i \in \mathcal{M} \setminus \mathcal{E} \text{ and } i > l]$ characterizes the updated belief values of the channels in $\mathcal{M} \setminus \mathcal{E}$ if they are sensed in the bad state with the channel index larger than l;
- Let $\omega_{-i} \triangleq \{\omega_j, j \in \mathcal{A}, j \neq i\}$ and

$$\begin{cases} \Delta_{max} \triangleq \max\limits_{\omega_{-i} \in [0,1]^{k-1}} \{F(1, \omega_{-i}) - F(0, \omega_{-i})\}, \\ \Delta_{min} \triangleq \min\limits_{\omega_{-i} \in [0,1]^{k-1}} \{F(1, \omega_{-i}) - F(0, \omega_{-i})\}. \end{cases}$$

[3]For presentation simplicity, by slightly abusing the notations without introducing ambiguity, we drop the time slot index t.

A. Definition and Properties of Auxiliary Value Function

In this subsection, inspired by the form of the value function $V_t(\Omega(t))$ and the analysis in [3], we first define the auxiliary value function with imperfect sensing and then derive several fundamental properties of the auxiliary value function, which are crucial in the study on the optimality of the myopic sensing policy.

Definition 1 (Auxiliary Value Function under Imperfect Sensing)**.** *The auxiliary value function, denoted as $W_t(\Omega)$ ($t = 1, 2, \cdots, T$) is recursively defined as follows:*

$$W_T(\Omega(T)) = F(\omega_1(T), \cdots, \omega_k(T)); \tag{5}$$

$$W_t(\Omega(t)) = F(\omega_1(t), \cdots, \omega_k(t)) + \beta \sum_{\mathcal{E} \subseteq \mathcal{N}(k)} Pr(\mathcal{N}(k), \mathcal{E}) W_{t+1}(\Omega_{\mathcal{E}}(t+1)), \tag{6}$$

where $\Omega_{\mathcal{E}}(t+1) \triangleq (\mathbf{P}_{\mathbf{11}}^{\mathcal{E}}, \mathbf{\Phi}(k+1, N), \mathbf{Q}^{\mathcal{N}(\mathbf{k}), \mathcal{E}})$ denotes the belief vector generated by $\Omega(t)$ based on (1)*.*

The above recursively defined auxiliary value function gives the expected cumulated reward of the following sensing policy: in slot t, sense the first k channels; if a channel i is correctly sensed idle ($S_i' = 1$ and $S_i = 1$), then put it on the top of the list to be sensed in next slot, otherwise drop it to the bottom of the list. Recall Lemma 1 and Lemma 2, under the condition $0 \leq \epsilon \leq \frac{(1-p_{11})p_{01}}{p_{11}(1-p_{01})}$, if the belief vector $\Omega(t)$ is ordered decreasingly in slot t, the above sensing policy is the myopic sensing policy with $W_t(\Omega(t))$ being the total reward from slot t to T.

In the subsequent analysis of this subsection, we prove some structural properties of the auxiliary value function.

Lemma 3 (Symmetry)**.** *If the expected reward function F is regular, the correspondent auxiliary value function $W_t(\Omega)$ is symmetrical in any two channel $i, j \leq k$ for all $t = 1, 2, \cdots, T$, i.e.,*

$$W_t(\omega_1, \cdots, \omega_i, \cdots, \omega_j, \cdots, \omega_N) = W_t(\omega_1, \cdots, \omega_j, \cdots, \omega_i, \cdots, \omega_N), \quad \forall i, j \leq k. \tag{7}$$

Proof: The lemma can be easily shown by backward induction noticing that $(\omega_1, \cdots, \omega_i, \cdots, \omega_j, \cdots, \omega_N)$ and $(\omega_1, \cdots, \omega_j, \cdots, \omega_i, \cdots, \omega_N)$ generate the same belief vector $\Omega_{\mathcal{E}}(t+1)$ for any $\mathcal{E}$. ■

Lemma 4 (Decomposability)**.** *If the expected reward function F is regular, then the correspondent auxiliary value function $W_t(\Omega(t))$ is decomposable for all $t = 1, 2, \cdots, T$, i.e.,*

$$W_t(\omega_1,\cdots,\omega_i,\cdots,\omega_N)=\omega_i W_t(\omega_1,\cdots,1,\cdots,\omega_N)+$$
$$(1-\omega_i)W_t(\omega_1,\cdots,0,\cdots,\omega_N),\quad \forall i\in\mathcal{N}.$$

Proof: The proof is given in the appendix. ■

Lemma 4 can be applied one step further to prove the following corollary.

Corollary 1. *If the expected reward function F is regular, then for any $l,m\in\mathcal{N}$ it holds that*

$$W_t(\omega_1,\cdots,\omega_l,\cdots,\omega_m,\cdots,\omega_N)-$$
$$W_t(\omega_1,\cdots,\omega_m,\cdots,\omega_l,\cdots,\omega_N)$$
$$=(\omega_l-\omega_m)\Big[W_t(\omega_1,\cdots,1,\cdots,0,\cdots,\omega_N)-$$
$$W_t(\omega_1,\cdots,0,\cdots,1,\cdots,\omega_N)\Big],\quad t=1,2,\cdots,T.$$

Lemma 5 (Monotonicity)**.** *If the expected reward function F is regular, the correspondent auxiliary value function $W_t(\Omega)$ is monotonously non-decreasing in ω_l, $\forall l\in\mathcal{N}$, i.e.,*

$$\omega_l'\geq\omega_l\Longrightarrow W_t(\omega_1,\cdots,\omega_l',\cdots,\omega_N)\geq W_t(\omega_1,\cdots,\omega_l,\cdots,\omega_N).$$

Proof: The proof is given in the appendix. ■

B. Optimality of Myopic Sensing under Imperfect Sensing

In this section, we study the optimality of the myopic sensing policy under imperfect sensing. We start by showing the following important auxiliary lemmas (Lemma 6 and Lemma 7) and then establish the sufficient condition under which the optimality of the myopic sensing policy is guaranteed.

Lemma 6. *Given that (1) F is regular, (2) $\epsilon<\frac{p_{01}(1-p_{11})}{P_{11}(1-p_{01})}$, and (3) $\beta\leq\frac{\Delta_{min}}{\Delta_{max}\left[(1-\epsilon)(1-p_{01})+\frac{\epsilon(p_{11}-p_{01})}{1-(1-\epsilon)(p_{11}-p_{01})}\right]}$, if $p_{11}\geq\omega_l\geq\omega_m\geq p_{01}$ where $l<m$, then it holds that*

$$W_t(\omega_1,\cdots,\omega_l,\cdots,\omega_m,\cdots,\omega_N)\geq$$
$$W_t(\omega_1,\cdots,\omega_m,\cdots,\omega_l,\cdots,\omega_N),\quad t=1,\cdots,T.$$

Lemma 7. *Given that (1) F is regular, (2) $\epsilon<\frac{p_{01}(1-p_{11})}{P_{11}(1-p_{01})}$, and (3) $\beta\leq\frac{\Delta_{min}}{\Delta_{max}\left[(1-\epsilon)(1-p_{01})+\frac{\epsilon(p_{11}-p_{01})}{1-(1-\epsilon)(p_{11}-p_{01})}\right]}$, if $p_{11}\geq\omega_1\geq\cdots\geq\omega_N\geq p_{01}$, for any $1\leq t\leq T$, it holds that*

$$W_t(\omega_1,\cdots,\omega_{N-1},\omega_N) - W_t(\omega_N,\omega_1,\cdots,\omega_{N-1}) \leq (1-\omega_N)\Delta_{max},$$

$$W_t(\omega_1,\omega_2,\cdots,\omega_{N-1},\omega_N) - W_t(\omega_N,\omega_2,\cdots,\omega_{N-1},\omega_1) \leq (p_{11}-p_{01})\Delta_{max}\frac{1-[\beta(1-\epsilon)(p_{11}-p_{01})]^{T-t+1}}{1-\beta(1-\epsilon)(p_{11}-p_{01})}.$$

Lemma 6 states that by swapping two elements in Ω with the former larger than the latter, the user does not increase the total expected reward. Lemma 7, on the other hand, gives the upper bound on the difference of the total reward of the two swapping operations, swapping ω_N and ω_k ($k = N-1,\cdots,1$) and swapping ω_1 and ω_N, respectively. For clarity of presentation, the detailed proofs of the two lemmas are deferred to the Appendix. From a technical point of view, it is insightful to compare the methodology in the proof with that in the analysis presented in [4] for the perfect sensing case with $k = 1$. The key point of the analysis in [4] lies in the coupling argument leading to Lemma 3 in [4]. This analysis, however, cannot be directly applied in the generic case with imperfect sensing due to the non-linearity of the belief vector update as stated in the remark after equation (1). Hence, we base our analysis on the intrinsic structure of the auxiliary value function W and investigate the different "branches" of channel realizations to derive the relevant bounds, which are further applied to study the optimality of the myopic sensing policy, as stated in the following theorem.

Theorem 1. *If $p_{01} \leq \omega_i(1) \leq p_{11}, 1 \leq i \leq N$, the myopic sensing policy is optimal if the following conditions hold: (1) $F(\Omega)$ is regular; (2) $\epsilon < \frac{p_{01}(1-p_{11})}{P_{11}(1-p_{01})}$; (3) $\beta \leq \frac{\Delta_{min}}{\Delta_{max}\left[(1-\epsilon)(1-p_{01})+\frac{\epsilon(p_{11}-p_{01})}{1-(1-\epsilon)(p_{11}-p_{01})}\right]}$.*

Proof: It suffices to show that for $t = 1,\cdots,T$, by sorting $\Omega(t)$ in decreasing order such that $\omega_1 \geq \cdots \geq \omega_N$, it holds that $W_t(\omega_1,\cdots,\omega_N) \geq W_t(\omega_{i_1},\cdots,\omega_{i_N})$, where $(\omega_{i_1},\cdots,\omega_{i_N})$ is any permutation of $(1,\cdots,N)$.

We prove the above inequality by contradiction. Assume, by contradiction, the maximum of W_t is achieved at $(\omega_{i_1^*},\cdots,\omega_{i_N^*}) \neq (\omega_1,\cdots,\omega_N)$, i.e.,

$$W_t(\omega_{i_1^*},\cdots,\omega_{i_N^*}) > W_t(\omega_1,\cdots,\omega_N). \tag{8}$$

However, run a bubble sort algorithm on $(\omega_{i_1^*},\cdots,\omega_{i_N^*})$ by repeatedly stepping through it, comparing each pair of adjacent element $\omega_{i_l^*}$ and $\omega_{i_{l+1}^*}$ and swapping them if $\omega_{i_l^*} < \omega_{i_l^*+1}$. Note that when the algorithm terminates, the channel belief vector are sorted decreasingly, that is to say, it becomes $(\omega_1,\cdots,\omega_N)$. By applying Lemma 6 at each swapping, we have $W_t(\omega_{i_1^*},\cdots,\omega_{i_N^*}) \leq W_t(\omega_1,\cdots,\omega_N)$, which contradicts to (8). Theorem 1 is thus proven. ■

As noted in [1], when the initial belief ω_i is set to $\frac{p_{01}}{p_{01}+1-p_{11}}$ as is often the case in practical systems, it can be checked that $p_{01} \leq \omega_i(1) \leq p_{11}$ holds. Moreover, even the initial belief does not fall in $[p_{01}, p_{11}]$, all the the belief values are bounded in the interval from the second slot following Lemma 1. Hence our results can be extended by treating the first slot separately from the future slots.

C. Discussion

In this subsection, we illustrate the application of the result obtained above in two concrete scenarios and compare our work with the existing results.

Consider the channel access problem in which the user is limited to sense k channels and gets one unit of reward if a sensed channel is in the good state, i.e., the utility function can be formulated as $F(\Omega_A) = (1-\epsilon)\sum_{i\in A}\omega_i$. Note that the optimality of the myopic sensing policy under this model is studied in [1] for a subset of scenarios where $k = 1$, $N = 2$. We now study the generic case with $k, N \geq 2$. To that end, we apply Theorem 1. Notice in this example, we have $\Delta_{min} = \Delta_{max} = 1-\epsilon$. We can then verify that when $\epsilon < \frac{p_{01}(1-p_{11})}{P_{11}(1-p_{10})}$, it holds that $\frac{\Delta_{min}}{\Delta_{max}[(1-\epsilon)(1-p_{01})+\frac{\epsilon(p_{11}-p_{01})}{1-(1-\epsilon)(p_{11}-p_{01})}]} > 1$. Therefore, when the condition 1 and 2 holds, the myopic sensing policy is optimal for any β. This result in generic cases significantly extends the results obtained in [1] where the optimality of the myopic policy is proved for the case of two channels and only conjectured for general cases.

Next consider another scenario where the user can sense k channels but can only choose one of them to transmit its packets. Under this model, the user wants to maximize its expected throughput. More specifically, the slot utility function $F = F(\Omega_A) = 1 - \Pi_{i\in\mathcal{A}}[1-(1-\epsilon)\omega_i]$, which is regular. In this context, we have $\Delta_{max} = (1-\epsilon)^{k-1}p_{11}^{k-1}$ and $\Delta_{min} = (1-\epsilon)^{k-1}p_{01}^{k-1}$. The third condition on for the myopic policy to be optimal becomes $\beta \leq \frac{p_{01}^{k-1}}{p_{11}^{k-1}[(1-\epsilon)(1-p_{01})+\frac{\epsilon(p_{11}-p_{01})}{1-(1-\epsilon)(p_{11}-p_{01})}]}$. Particularly, when $\epsilon = 0$, $\beta \leq \frac{p_{01}^{k-1}}{p_{11}^{k-1}(1-p_{01})}$. It can be noted that even when there is no sensing error, the myopic policy is not ensured to be optimal, which confirms our findings in previous work [5] on perfect sensing scenarios.

IV. Related Work

Due to its application in numerous engineering problems, the restless multi-armed bandit (RMAB) problem is of fundamental importance in stochastic decision theory. However, finding the optimal policy in the generic RMAB problem is shown to be PSPACE-hard by Papadimitriou *et al.* in [6]. Whittle proposed a heuristic index policy, called Whittle index policy [7] which are shown to be asymptotically optimal in certain limited regime under some specific constraints [8]. Unfortunately, not every RMAB problem has a well-defined Whittle index. Moreover, computing the Whittle index can be prohibitively

complex. In this regard, Liu *et al.* studied in [9] the indexability of a class of RMAB problems relevant to dynamic multi-channel access applications. However, the optimality of the myopic policy based on Whittle index is not ensured in the general cases, especially when the arms follow non-identical Markov chains.

A natural alternative, given that the RMAB problem is not tractable, is to seek simple myopic policies maximizing the short-term reward. In this line of research, significant research efforts have been devoted to studying the performance gap between the myopic policy and the optimal one and designing approximation algorithms and heuristic policies (cf. [10], [11], [12]). Specifically, a simple myopic policy, termed as greedy policy, is developed in [10] that yields a factor 2 approximation of the optimal policy for a subclass of scenarios referred to as *Monotone bandits*. Recently, the RMAB problem finds its application in the opportunistic channel access and has motivated the study of the myopic sensing policy in this context. More specifically, the structure of the myopic sensing policy is studied in [13]. The optimality of the myopic sensing policy is derived in [4] for the positively correlated channels when the sender is limited to choose one channel each time (i.e., $k = 1$). The result is further extended in to the case of sensing multiple channels ($k \geq 1$) channels in [3] for a particular form of utility function modeling the fact that the user gets one unit of reward for each channel sensed good. A separation principle has been established in [11] which reveals the optimality of the myopic approach in the design of the channel state detector and the access policy. Our previous work [2] [14] adopts another line of research by focusing a family of generic and practically important utility functions and deriving closed-form conditions under which the myopic sensing policy is ensured to be optimal. In the context of imperfect sensing, the optimality of the myopic sensing policy is proved for the case of $N = 2$ and $k = 1$ in [1]. Our work presented in this paper contributes the literature by deriving the closed-form conditions on the optimality of the myopic sensing policy with imperfect sensing in the general case.

V. Conclusion

In this paper, we have investigated the problem of opportunistic channel access under imperfect channel state sensing. We have derived closed-form conditions under which the myopic sensing policy is ensured to be optimal. Due to the generic RMAB formulation of the problem, the obtained results and the analysis methodology presented in this paper are widely applicable in a wide range of domains.

APPENDIX A
PROOF OF LEMMA 4

We proceed the proof by backward induction. Firstly, it is easy to verify that the lemma holds for slot T.

Assume that the lemma holds from slots $t+1, \cdots, T$, we now prove it also holds for slot t by the following two different cases.

- Case 1: channel l is not sensed in slot t, i.e. $l \geq k+1$. Let $\mathcal{M} \triangleq \mathcal{N}(k) = \{1, \cdots, k\}$, $\omega_l = 0$ and 1, respectively, we have

$$\begin{aligned} W_t(\omega_1, \cdots, \omega_l, \cdots, \omega_n) &= F(\omega_1, \cdots, \omega_k) + \beta \sum_{\mathcal{E} \subseteq \mathcal{M}} Pr(\mathcal{M}, \mathcal{E}) W_{t+1}(\Omega_l^{\mathcal{E}}(t+1)), \\ W_t(\omega_1, \cdots, 0, \cdots, \omega_n) &= F(\omega_1, \cdots, \omega_k) + \beta \sum_{\mathcal{E} \subseteq \mathcal{M}} Pr(\mathcal{M}, \mathcal{E}) W_{t+1}(\Omega_{l,0}^{\mathcal{E}}(t+1)), \\ W_t(\omega_1, \cdots, 1, \cdots, \omega_n) &= F(\omega_1, \cdots, \omega_k) + \beta \sum_{\mathcal{E} \subseteq \mathcal{M}} Pr(\mathcal{M}, \mathcal{E}) W_{t+1}(\Omega_{l,1}^{\mathcal{E}}(t+1)), \end{aligned}$$

  where

$$\begin{aligned} \Omega_l^{\mathcal{E}}(t+1) &= (\mathbf{P}_{\mathbf{11}}^{\mathcal{E}}, \mathbf{\Phi}(k+1, l-1), \tau(\omega_l), \mathbf{\Phi}(l+1, N), \mathbf{Q}^{\mathcal{M},\mathcal{E}}), \\ \Omega_{l,0}^{\mathcal{E}}(t+1) &= (\mathbf{P}_{\mathbf{11}}^{\mathcal{E}}, \mathbf{\Phi}(k+1, l-1), p_{01}, \mathbf{\Phi}(l+1, N), \mathbf{Q}^{\mathcal{M},\mathcal{E}}), \\ \Omega_{l,1}^{\mathcal{E}}(t+1) &= (\mathbf{P}_{\mathbf{11}}^{\mathcal{E}}, \mathbf{\Phi}(k+1, l-1), p_{11}, \mathbf{\Phi}(l+1, N), \mathbf{Q}^{\mathcal{M},\mathcal{E}}). \end{aligned}$$

  To prove the lemma in this case, it is sufficient to prove

$$W_{t+1}(\Omega_l^{\mathcal{E}}(t+1)) = (1-\omega_l) W_{t+1}(\Omega_{l,0}^{\mathcal{E}}(t+1)) + \omega_l W_{t+1}(\Omega_{l,1}^{\mathcal{E}}(t+1)) \tag{9}$$

  According to induction result, we have

$$\begin{aligned} W_{t+1}(\Omega_l^{\mathcal{E}}(t+1)) =& \tau(\omega_l) \cdot W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}}, \mathbf{\Phi}(k+1, l-1), 1, \mathbf{\Phi}(l+1, N), \mathbf{Q}^{\mathcal{M},\mathcal{E}}) \\ &+ (1-\tau(\omega_l)) \cdot W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}}, \mathbf{\Phi}(k+1, l-1), 0, \mathbf{\Phi}(l+1, N), \mathbf{Q}^{\mathcal{M},\mathcal{E}}) \end{aligned} \tag{10}$$

$$\begin{aligned} W_{t+1}(\Omega_{l,0}^{\mathcal{E}}(t+1)) =& p_{01} \cdot W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}}, \mathbf{\Phi}(k+1, l-1), 1, \mathbf{\Phi}(l+1, N), \mathbf{Q}^{\mathcal{M},\mathcal{E}}) \\ &+ (1-p_{01}) \cdot W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}}, \mathbf{\Phi}(k+1, l-1), 0, \mathbf{\Phi}(l+1, N), \mathbf{Q}^{\mathcal{M},\mathcal{E}}) \end{aligned} \tag{11}$$

$$\begin{aligned} W_{t+1}(\Omega_{l,0}^{\mathcal{E}}(t+1)) =& p_{11} \cdot W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}}, \mathbf{\Phi}(k+1, l-1), 1, \mathbf{\Phi}(l+1, N), \mathbf{Q}^{\mathcal{M},\mathcal{E}}) \\ &+ (1-p_{11}) \cdot W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}}, \mathbf{\Phi}(k+1, l-1), 0, \mathbf{\Phi}(l+1, N), \mathbf{Q}^{\mathcal{M},\mathcal{E}}) \end{aligned} \tag{12}$$

  Combing (10), (11), (12), we have (9).

- Case 2: channel l is sensed in slot t, i.e. $l \leq k$. Let $\mathcal{M} \triangleq \mathcal{N}(k) \setminus \{l\} = \{1, \cdots, l-1, l+1, \cdots, k\}$, we have according to (6)

$$\begin{aligned} W_t(\Omega(t)) =& F(\omega_1, \cdots, \omega_l, \cdots, \omega_k) \\ &+ \beta(1-\epsilon)\omega_l \sum_{\mathcal{E}\subseteq\mathcal{M}} Pr(\mathcal{M}, \mathcal{E}) W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}}, p_{11}, \mathbf{\Phi}(k+1, N), \overline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}}, \underline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}}) \\ &+ \beta[1-(1-\epsilon)\omega_l] \sum_{\mathcal{E}\subseteq\mathcal{M}} Pr(\mathcal{M}, \mathcal{E}) W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}}, \mathbf{\Phi}(k+1, N), \overline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}}, \tau(\varphi(\omega_l)), \underline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}}) \end{aligned}$$

Let $\omega_l = 0$ and 1, respectively, we have

$$\begin{aligned} W_t(\omega_1, \cdots, 0, \cdots, \omega_n) =& F(\omega_1, \cdots, 0, \cdots, \omega_k) \\ &+ \beta \sum_{\mathcal{E}\subseteq\mathcal{M}} Pr(\mathcal{M}, \mathcal{E}) W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}}, \mathbf{\Phi}(k+1, N), \overline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}}, p_{01}, \underline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}}), \end{aligned}$$

$$\begin{aligned} W_t(\omega_1, \cdots, 1, \cdots, \omega_n) =& F(\omega_1, \cdots, 1, \cdots, \omega_k) \\ &+ \beta(1-\epsilon) \sum_{\mathcal{E}\subseteq\mathcal{M}} Pr(\mathcal{M}, \mathcal{E}) W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}}, p_{11}, \mathbf{\Phi}(k+1, N), \overline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}}, \underline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}}) \\ &+ \beta\epsilon \sum_{\mathcal{E}\subseteq\mathcal{M}} Pr(\mathcal{M}, \mathcal{E}) W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}}, \mathbf{\Phi}(k+1, N), \overline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}}, p_{11}, \underline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}}) \end{aligned}$$

To prove the lemma in this case, it is sufficient to show

$$\begin{aligned} &[1-(1-\epsilon)\omega_l] W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}}, \mathbf{\Phi}(k+1, N), \overline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}}, \tau(\varphi(\omega_l)), \underline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}}) \\ &\quad = (1-\omega_l) W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}}, \mathbf{\Phi}(k+1, N), \overline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}}, p_{01}, \underline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}}) \\ &\qquad + \epsilon\omega_l W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}}, \mathbf{\Phi}(k+1, N), \overline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}}, p_{11}, \underline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}}) \end{aligned} \quad (13)$$

According to induction result, we have

$$\begin{aligned} &W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}}, \mathbf{\Phi}(k+1, N), \overline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}}, \tau(\varphi(\omega_l)), \underline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}}) \\ &\quad = \tau(\varphi(\omega_l)) W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}}, \mathbf{\Phi}(k+1, N), \overline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}}, 1, \underline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}}) \\ &\qquad + (1-\tau(\varphi(\omega_l))) W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}}, \mathbf{\Phi}(k+1, N), \overline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}}, 0, \underline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}}) \end{aligned} \quad (14)$$

$$\begin{aligned} &W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}}, \mathbf{\Phi}(k+1, N), \overline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}}, p_{01}, \underline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}}) \\ &\quad = p_{01} W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}}, \mathbf{\Phi}(k+1, N), \overline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}}, 1, \underline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}}) \\ &\qquad + (1-p_{01}) W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}}, \mathbf{\Phi}(k+1, N), \overline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}}, 0, \underline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}}) \end{aligned} \quad (15)$$

$$W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}}, \mathbf{\Phi}(k+1,N), \overline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}}, p_{11}, \underline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}})$$
$$= p_{11} W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}}, \mathbf{\Phi}(k+1,N), \overline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}}, 1, \underline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}})$$
$$+ (1-p_{11}) W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}}, \mathbf{\Phi}(k+1,N), \overline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}}, 0, \underline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}}) \quad (16)$$

Combing (14), (15), (16), we have (13).

Combing the above analysis in two cases, we thus prove Lemma 4.

APPENDIX B
PROOF OF LEMMA 5

We proceed the proof by backward induction. Firstly, it is easy to verify that the lemma holds for slot T.

Assume that the lemma holds from slots $t+1, \cdots, T$, we now prove that it also holds for slot t by distinguishing the following two cases.

- Case 1: channel l is not sensed in slot t, i.e., $l \geq k+1$. In this case, the immediate reward is unrelated to ω_l and ω_l'. Moreover, let $\Omega(t+1)$ and $\Omega'(t+1)$ denote the belief vector generated by $\Omega(t) = (\omega_1, \cdots, \omega_l, \cdots, \omega_N)$ and $\Omega'(t) = (\omega_1, \cdots, \omega_l', \cdots, \omega_N)$, respectively, it can be noticed that $\Omega(t+1)$ and $\Omega'(t+1)$ differ in only one element: $\omega_l'(t+1) \geq \omega_l(t+1)$. By induction, it holds that $W_{t+1}(\Omega'(t+1)) \geq W_{t+1}(\Omega(t+1))$. Noticing (6), it follows that $W_t(\Omega'(t)) \geq W_t(\Omega(t))$.
- Case 2: channel l is sensed in slot t, i.e., $l \leq k$. Following Lemma 4 and after some straightforward algebraic operations, we have

$$W_t(\omega_1, \cdots, \omega_l', \cdots, \omega_N) - W_t(\omega_1, \cdots, \omega_l, \cdots, \omega_N) =$$
$$(\omega_l' - \omega_l)[W_t(\omega_1, \cdots, 1, \cdots, \omega_N) - W_t(\omega_1, \cdots, 0, \cdots, \omega_N)].$$

Let $\mathcal{M} \triangleq \mathcal{N}(k) \setminus \{l\} = \{1, \cdots, l-1, l+1, \cdots, k\}$, by developing $W_t(\Omega(t))$ as a function of ω_l, we have

$$W_t(\Omega(t)) = F(\omega_1(t), \cdots, \omega_k(t)) + \beta(1-\epsilon)\omega_l \sum_{\mathcal{E} \subseteq \mathcal{M}} Pr(\mathcal{M}, \mathcal{E}) W_{t+1}(\Omega_{\mathcal{E}}(t+1))$$
$$+ \beta[1-(1-\epsilon)\omega_l] \sum_{\mathcal{E} \subseteq \mathcal{M}} Pr(\mathcal{M}, \mathcal{E}) W_{t+1}(\Omega_{\mathcal{E}}(t+1)).$$

Let $\omega_l = 0$ and 1, respectively, we have

$$W_t(\omega_1, \cdots, 0, \cdots, \omega_n) \quad = \quad F(\omega_1, \cdots, 0, \cdots, \omega_n) + \beta \sum_{\mathcal{E} \subseteq \mathcal{M}} Pr(\mathcal{M}, \mathcal{E}) W_{t+1}(\Omega_0^{\mathcal{E}}(t+1)),$$

$$\begin{aligned} W_t(\omega_1,\cdots,1,\cdots,\omega_n) &= F(\omega_1,\cdots,1,\cdots,\omega_n)+\beta(1-\epsilon)\sum_{\mathcal{E}\subseteq\mathcal{M}} Pr(\mathcal{M},\mathcal{E})W_{t+1}(\Omega^{\mathcal{E}}_{1-\epsilon}(t+1)) \\ &+ \beta\epsilon\sum_{\mathcal{E}\subseteq\mathcal{M}} Pr(\mathcal{M},\mathcal{E})W_{t+1}(\Omega^{\mathcal{E}}_{\epsilon}(t+1)), \end{aligned}$$

where

$$\begin{aligned} \Omega^{\mathcal{E}}_0(t+1) &= (\mathbf{P}^{\mathcal{E}}_{\mathbf{11}},\mathbf{\Phi}(k+1,N),\overline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}},p_{01},\underline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}}), \\ \Omega^{\mathcal{E}}_{1-\epsilon}(t+1) &= (\mathbf{P}^{\mathcal{E}}_{\mathbf{11}},p_{11},\mathbf{\Phi}(k+1,N),\overline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}},\underline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}}), \\ \Omega^{\mathcal{E}}_{\epsilon}(t+1) &= (\mathbf{P}^{\mathcal{E}}_{\mathbf{11}},\mathbf{\Phi}(k+1,N),\overline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}},p_{11},\underline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}}). \end{aligned}$$

It can be checked that $\Omega^{\mathcal{E}}_{1-\epsilon}(t+1)\geq\Omega^{\mathcal{E}}_0(t+1)$ and $\Omega^{\mathcal{E}}_{\epsilon}(t+1)\geq\Omega^{\mathcal{E}}_0(t+1)$. It then follows from induction that given $\mathcal{E}$, $W_{t+1}(\Omega^{\mathcal{E}}_{1-\epsilon}(t+1))\geq W_{t+1}(\Omega^{\mathcal{E}}_0(t+1))$ and $W_{t+1}(\Omega^{\mathcal{E}}_{1-\epsilon}(t+1))\geq W_{t+1}(\Omega^{\mathcal{E}}_0(t+1))$. Noticing that F is increasing, we then have

$$\begin{aligned} W_t(\omega_1,\cdots,1,\cdots,\omega_n)-W_t(\omega_1,\cdots,0,\cdots,\omega_n) &= F(\omega_1,\cdots,1,\cdots,\omega_n)-F(\omega_1,\cdots,0,\cdots,\omega_n) \\ &+\beta(1-\epsilon)\sum_{\mathcal{E}\subseteq\mathcal{M}} Pr(\mathcal{M},\mathcal{E})[W_{t+1}(\Omega^{\mathcal{E}}_{1-\epsilon}(t+1))-W_{t+1}(\Omega^{\mathcal{E}}_0(t+1))] \\ &+\beta\epsilon\sum_{\mathcal{E}\subseteq\mathcal{M}} Pr(\mathcal{M},\mathcal{E})[W_{t+1}(\Omega^{\mathcal{E}}_{\epsilon}(t+1))-W_{t+1}(\Omega^{\mathcal{E}}_0(t+1))]\geq 0. \end{aligned}$$

Combining the above analysis in two cases completes our proof.

APPENDIX C
PROOF OF LEMMA 6 AND LEMMA 7

Due to the dependency between the two lemmas, we prove them together by backward induction.

We first show that Lemma 6 and Lemma 7 hold for slot T. It is easy to verify that Lemma 6 holds.

We then prove Lemma 7. Noticing that $p_{01}\leq\omega_N\leq\omega_k\leq p_{11}\leq 1$, we have

$$\begin{aligned} W_T(\omega_1,\cdots,\omega_N)-W_T(\omega_N,\omega_1,\cdots,\omega_{N-1}) &= F(\omega_1,\cdots,\omega_k)-F(\omega_N,\omega_1,\cdots,\omega_{k-1}) \\ &= (\omega_k-\omega_N)[F(\omega_1,\cdots,\omega_{k-1},1)-F(\omega_1,\cdots,\omega_{k-1},0)]\leq(1-\omega_N)\Delta_{max}, \\ W_T(\omega_1,\cdots,\omega_N)-W_T(\omega_N,\omega_2,\cdots,\omega_{N-1},\omega_1) &= F(\omega_1,\cdots,\omega_k)-F(\omega_N,\omega_2,\cdots,\omega_{k-1}) \\ &= (\omega_1-\omega_N)[F(1,\omega_2,\cdots,\omega_k)-F(0,\omega_2,\cdots,\omega_k)]\leq(p_{11}-p_{01})\Delta_{max}. \end{aligned}$$

Lemma 7 thus holds for slot T.

Assume that Lemma 6 and Lemma 7 hold for slots $T,\cdots,t+1$, we now prove that it holds for slot t.

We first prove Lemma 6. We distinguish the following three cases considering $l < m$:

- Case 1: $l \geq k+1$. In this case, we have

$$\begin{aligned}
&W_t(\omega_1,\cdots,\omega_l,\cdots,\omega_m,\cdots,\omega_N) - W_t(\omega_1,\cdots,\omega_m,\cdots,\omega_l,\cdots,\omega_N)\\
&\quad= (\omega_l-\omega_m)[W_t(\omega_1,\cdots,1,\cdots,0,\cdots,\omega_N) - W_t(\omega_1,\cdots,0,\cdots,1,\cdots,\omega_N)]\\
&\quad= (\omega_l-\omega_m)\beta \sum_{\mathcal{E}\subseteq\mathcal{N}(k)} Pr(\mathcal{N}(k),\mathcal{E})[W_{t+1}(\Omega_{\mathcal{E}}(t+1)) - W_{t+1}(\Omega'_{\mathcal{E}}(t+1))],
\end{aligned}$$

  where

$$\begin{aligned}
\Omega_{\mathcal{E}}(t+1) &= (\mathbf{P}^{\mathcal{E}}_{\mathbf{11}}, \tau(\omega_{k+1}),\cdots,p_{11},\cdots,p_{01},\cdots,\tau(\omega_N),\mathbf{Q}^{\mathcal{N}(\mathbf{k}),\mathcal{E}}),\\
\Omega'_{\mathcal{E}}(t+1) &= (\mathbf{P}^{\mathcal{E}}_{\mathbf{11}}, \tau(\omega_{k+1}),\cdots,p_{01},\cdots,p_{11},\cdots,\tau(\omega_N),\mathbf{Q}^{\mathcal{N}(\mathbf{k}),\mathcal{E}}).
\end{aligned}$$

  It follows from the induction result that $W_{t+1}(\Omega_{\mathcal{E}}(t+1)) \geq W_{t+1}(\Omega'_{\mathcal{E}}(t+1))$. Hence

$$W_t(\omega_1,\cdots,\omega_l,\cdots,\omega_m,\cdots,\omega_N) \geq W_t(\omega_1,\cdots,\omega_m,\cdots,\omega_l,\cdots,\omega_N).$$

- Case 2: $l \leq k$ and $m \geq k+1$. In this case, denote $\mathcal{M} \triangleq \mathcal{N}(k)\setminus\{l\}$, it can be noted that $\mathbf{Q}^{\mathcal{M},\mathcal{E}} = \underline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}} + \overline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}}$. In this case, we have

$$\begin{aligned}
&W_t(\omega_1,\cdots,\omega_l,\cdots,\omega_m,\cdots,\omega_N) - W_t(\omega_1,\cdots,\omega_m,\cdots,\omega_l,\cdots,\omega_N)\\
&\quad= (\omega_l-\omega_m)[W_t(\omega_1,\cdots,1,\cdots,0,\cdots,\omega_N) - W_t(\omega_1,\cdots,0,\cdots,1,\cdots,\omega_N)]\\
&\quad= (\omega_l-\omega_m)[F(\omega_1,\cdots,1,\cdots,\omega_k) - F(\omega_1,\cdots,0,\cdots,\omega_k)+\\
&\qquad\beta\sum_{\mathcal{E}\subseteq\mathcal{M}} Pr(\mathcal{M},\mathcal{E})[(1-\epsilon)W_{t+1}(\mathbf{P}^{\mathcal{E}}_{\mathbf{11}},p_{11},\tau(\omega_{k+1}),\cdots,p_{01},\cdots,\tau(\omega_N),\mathbf{Q}^{\mathcal{M},\mathcal{E}})+\\
&\qquad\epsilon W_{t+1}(\mathbf{P}^{\mathcal{E}}_{\mathbf{11}},\tau(\omega_{k+1}),\cdots,p_{01},\cdots,\tau(\omega_N),\overline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}},p_{11},\underline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}})-\\
&\qquad W_{t+1}(\mathbf{P}^{\mathcal{E}}_{\mathbf{11}},\tau(\omega_{k+1}),\cdots,p_{11},\cdots,\tau(\omega_N),\overline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}},p_{01},\underline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}})]\\
&\quad\geq (\omega_l-\omega_m)[\Delta_{min}+\beta\sum_{\mathcal{E}\subseteq\mathcal{M}} Pr(\mathcal{M},\mathcal{E})\cdot[(1-\epsilon)W_{t+1}(p_{01},\mathbf{P}^{\mathcal{E}}_{\mathbf{11}},p_{11},\tau(\omega_{k+1}),\cdots,\tau(\omega_N),\mathbf{Q}^{\mathcal{M},\mathcal{E}})+\\
&\qquad\epsilon W_{t+1}(p_{01},\mathbf{P}^{\mathcal{E}}_{\mathbf{11}},\tau(\omega_{k+1}),\cdots,\tau(\omega_N),\mathbf{Q}^{\mathcal{M},\mathcal{E}},p_{11})-\\
&\qquad W_{t+1}(\mathbf{P}^{\mathcal{E}}_{\mathbf{11}},p_{11},\tau(\omega_{k+1}),\cdots,\tau(\omega_N),\mathbf{Q}^{\mathcal{M},\mathcal{E}},p_{01})]\\
&\quad\geq (\omega_l-\omega_m)\Bigg[\Delta_{min}-\beta\sum_{\mathcal{E}\subseteq\mathcal{M}} Pr(\mathcal{M},\mathcal{E})\cdot\\
&\qquad\left((1-\epsilon)(1-p_{01})\Delta_{max}+\epsilon(p_{11}-p_{01})\Delta_{max}\frac{1-[\beta(1-\epsilon)(p_{11}-p_{01})]^{T-t}}{1-\beta(1-\epsilon)(p_{11}-p_{01})}\right)\Bigg]
\end{aligned}$$

$$\geq (\omega_l - \omega_m) \sum_{\mathcal{E}\subseteq\mathcal{M}} Pr(\mathcal{M},\mathcal{E})\cdot$$
$$\left[\Delta_{min} - \beta\left((1-\epsilon)(1-p_{01})\Delta_{max} + \epsilon(p_{11}-p_{01})\Delta_{max}\frac{1}{1-(1-\epsilon)(p_{11}-p_{01})}\right)\right] \geq 0,$$

where the first inequality follows the induction result of Lemma 6, the second inequality follows the induction result of Lemma 7, the third inequality follows the condition in the lemma.

- Case 3: $l, m \geq k$. This case follows Lemma 3.

Lemma 6 is thus proven for slot t.

We then proceed to prove Lemma 7. We start with the first inequality. We develop W_t w.r.t. ω_k and ω_N according to Lemma 4 as follows:

$$\begin{aligned}
&W_t(\omega_1,\cdots,\omega_{k-1},\omega_k,\cdots,\omega_{n-1},\omega_n) - W_t(\omega_n,\omega_1,\cdots,\omega_{k-1},\omega_k,...,\omega_{n-1})\\
&= \omega_k\omega_n[W_t(\omega_1,\cdots,\omega_{k-1},1,\omega_{k+1},\cdots,\omega_{n-1},1) - W_t(1,\omega_1,\cdots,\omega_{k-1},1,\omega_{k+1},\cdots,\omega_{n-1})]\\
&+\omega_k(1-\omega_n)[W_t(\omega_1,\cdots,\omega_{k-1},1,\omega_{k+1},\cdots,\omega_{n-1},0) - W_t(0,\omega_1,\cdots,\omega_{k-1},1,\omega_{k+1},\cdots,\omega_{n-1})]\\
&+(1-\omega_k)\omega_n[W_t(\omega_1,\cdots,\omega_{k-1},0,\omega_{k+1},\cdots,\omega_{n-1},1) - W_t(1,\omega_1,\cdots,\omega_{k-1},0,\omega_{k+1},\cdots,\omega_{n-1})]\\
&+(1-\omega_k)(1-\omega_n)[W_t(\omega_1,\cdots,\omega_{k-1},0,\omega_{k+1},\cdots,\omega_{n-1},0) - W_t(0,\omega_1,\cdots,\omega_{k-1},0,\omega_{k+1},\cdots,\omega_{n-1})]
\end{aligned} \tag{17}$$

We proceed the proof by upbounding the four terms in (17).

For the first term, we have

$$\begin{aligned}
&W_t(\omega_1,\cdots,\omega_{k-1},1,\omega_{k+1},\cdots,\omega_{n-1},1) - W_t(1,\omega_1,\cdots,\omega_{k-1},1,\omega_{k+1},\cdots,\omega_{n-1})\\
&\quad= \beta \sum_{\mathcal{E}\subseteq\mathcal{N}(k-1)} Pr(\mathcal{N}(k-1),\mathcal{E})\cdot[(1-\epsilon)W_{t+1}(\mathbf{P}^{\mathcal{E}}_{\mathbf{11}},p_{11},\mathbf{\Phi}(k+1,N-1),p_{11},\mathbf{Q}^{\mathcal{N}(\mathbf{k-1}),\mathcal{E}})\\
&\qquad\qquad+\epsilon W_{t+1}(\mathbf{P}^{\mathcal{E}}_{\mathbf{11}},\mathbf{\Phi}(k+1,N-1),p_{11},\mathbf{Q}^{\mathcal{N}(\mathbf{k-1}),\mathcal{E}},p_{11})\\
&\qquad\qquad-(1-\epsilon)W_{t+1}(p_{11},\mathbf{P}^{\mathcal{E}}_{\mathbf{11}},p_{11},\mathbf{\Phi}(k+1,N-1),\mathbf{Q}^{\mathcal{N}(\mathbf{k-1}),\mathcal{E}})\\
&\qquad\qquad-\epsilon W_{t+1}(\mathbf{P}^{\mathcal{E}}_{\mathbf{11}},p_{11},\mathbf{\Phi}(k+1,N-1),p_{11},\mathbf{Q}^{\mathcal{N}(\mathbf{k-1}),\mathcal{E}})] \leq 0
\end{aligned}$$

where, the inequality follows the induction of Lemma 6.

For the second term, we have

$$\begin{aligned}
&W_t(\omega_1,\cdots,\omega_{k-1},1,\omega_{k+1},\cdots,\omega_{n-1},0) - W_t(0,\omega_1,\cdots,\omega_{k-1},1,\omega_{k+1},\cdots,\omega_{n-1})\\
&= F(\omega_1,\cdots,\omega_{k-1},1) - F(0,\omega_1,\cdots,\omega_{k-1})
\end{aligned}$$

$$
\begin{aligned}
&+\beta \sum_{\mathcal{E}\subseteq\mathcal{N}(k-1)} Pr(\mathcal{N}(k-1),\mathcal{E})\cdot[(1-\epsilon)W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}},p_{11},\mathbf{\Phi}(k+1,N-1),p_{01},\mathbf{Q}^{\mathcal{N}(\mathbf{k-1}),\mathcal{E}})\\
&+\epsilon W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}},\mathbf{\Phi}(k+1,N-1),p_{01},\mathbf{Q}^{\mathcal{N}(\mathbf{k-1}),\mathcal{E}},p_{11})-W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}},p_{11},\mathbf{\Phi}(k+1,N-1),p_{01},\mathbf{Q}^{\mathcal{N}(\mathbf{k-1}),\mathcal{E}})]\\
&=F(\omega_1,\cdots,\omega_{k-1},1)-F(0,\omega_1,\cdots,\omega_{k-1})+\beta\sum_{\mathcal{E}\subseteq\mathcal{N}(k-1)} Pr(\mathcal{N}(k-1),\mathcal{E})\cdot\\
&[\epsilon W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}},\mathbf{\Phi}(k+1,N-1),p_{01},\mathbf{Q}^{\mathcal{N}(\mathbf{k-1}),\mathcal{E}},p_{11})-\epsilon W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}},p_{11},\mathbf{\Phi}(k+1,N-1),p_{01},\mathbf{Q}^{\mathcal{N}(\mathbf{k-1}),\mathcal{E}})]\\
&\leq \Delta_{max}
\end{aligned}
$$

following the induction of Lemma 6.

For the third term, we have

$$
\begin{aligned}
&W_t(\omega_1,\cdots,\omega_{k-1},0,\omega_{k+1},\cdots,\omega_{n-1},1)-W_t(1,\omega_1,\cdots,\omega_{k-1},0,\omega_{k+1},\cdots,\omega_{n-1})\\
&\quad=F(\omega_1,\cdots,\omega_{k-1},0)-F(1,\omega_1,\cdots,\omega_{k-1})+\\
&\quad\beta\sum_{\mathcal{E}\subseteq\mathcal{N}(k-1)} Pr(\mathcal{N}(k-1),\mathcal{E})[W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}},\mathbf{\Phi}(k+1,N-1),p_{11},\mathbf{Q}^{\mathcal{N}(\mathbf{k-1}),\mathcal{E}},p_{01})-\\
&\quad(1-\epsilon)W_{t+1}(p_{11},\mathbf{P}_{\mathbf{11}}^{\mathcal{E}},p_{01},\mathbf{\Phi}(k+1,N-1),\mathbf{Q}^{\mathcal{N}(\mathbf{k-1}),\mathcal{E}})-\epsilon W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}},p_{01},\mathbf{\Phi}(k+1,N-1),p_{11},\mathbf{Q}^{\mathcal{N}(\mathbf{k-1}),\mathcal{E}})]\\
&\leq -\Delta_{min}+\beta\sum_{\mathcal{E}\subseteq\mathcal{N}(k-1)} Pr(\mathcal{N}(k-1),\mathcal{E})[W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}},p_{11},\mathbf{\Phi}(k+1,N-1),\mathbf{Q}^{\mathcal{N}(\mathbf{k-1}),\mathcal{E}},p_{01})-\\
&\quad(1-\epsilon)W_{t+1}(p_{01},p_{11},\mathbf{P}_{\mathbf{11}}^{\mathcal{E}},\mathbf{\Phi}(k+1,N-1),\mathbf{Q}^{\mathcal{N}(\mathbf{k-1}),\mathcal{E}})-\epsilon W_{t+1}(p_{01},\mathbf{P}_{\mathbf{11}}^{\mathcal{E}},\mathbf{\Phi}(k+1,N-1),\mathbf{Q}^{\mathcal{N}(\mathbf{k-1}),\mathcal{E}},p_{11})]\\
&\leq -\Delta_{min}+\beta\sum_{\mathcal{E}\subseteq\mathcal{N}(k-1)} Pr(\mathcal{N}(k-1),\mathcal{E})\left[(1-\epsilon)(1-p_{01})\Delta_{max}+\epsilon(p_{11}-p_{01})\Delta_{max}\frac{1-[\beta(1-\epsilon)(p_{11}-p_{01})]^{T-t}}{1-\beta(1-\epsilon)(p_{11}-p_{01})}\right]\\
&\leq \sum_{\mathcal{E}\subseteq\mathcal{N}(k-1)} Pr(\mathcal{N}(k-1),\mathcal{E})\left[-\Delta_{min}+\beta\left[(1-\epsilon)(1-p_{01})\Delta_{max}+\epsilon(p_{11}-p_{01})\Delta_{max}\frac{1}{1-(1-\epsilon)(p_{11}-p_{01})}\right]\right]\leq 0
\end{aligned}
$$

where the first inequality follows the induction result of Lemma 6, the second equality follows the induction result of Lemma 7, the forth inequality is due the condition in Lemma 7.

For the fourth term, we have

$$
\begin{aligned}
&W_t(\omega_1,\cdots,\omega_{k-1},0,\omega_{k+1},\cdots,\omega_{n-1},0)-W_t(0,\omega_1,\cdots,\omega_{k-1},0,\omega_{k+1},\cdots,\omega_{n-1})\\
&=\beta\sum_{\mathcal{E}\subseteq\mathcal{N}(k-1)} Pr(\mathcal{N}(k-1),\mathcal{E})[W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}},\mathbf{\Phi}(k+1,N-1),p_{01},\mathbf{Q}^{\mathcal{N}(\mathbf{k-1}),\mathcal{E}},p_{01})\\
&-W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}},p_{01},\mathbf{\Phi}(k+1,N-1),\mathbf{Q}^{\mathcal{N}(\mathbf{k-1}),\mathcal{E}},p_{01})]\\
&=\beta\sum_{\mathcal{E}\subseteq\mathcal{N}(k-1)} Pr(\mathcal{N}(k-1),\mathcal{E})[W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}},\mathbf{\Phi}(k+1,N-1),p_{01},\mathbf{Q}^{\mathcal{N}(\mathbf{k-1}),\mathcal{E}},p_{01})\\
&-W_{t+1}(p_{01},\mathbf{P}_{\mathbf{11}}^{\mathcal{E}},\mathbf{\Phi}(k+1,N-1),\mathbf{Q}^{\mathcal{N}(\mathbf{k-1}),\mathcal{E}},p_{01})]
\end{aligned}
$$

$$\leq \beta \sum_{\mathcal{E}\subseteq\mathcal{N}(k-1)} Pr(\mathcal{N}(k-1),\mathcal{E})[W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}},\mathbf{\Phi}(k+1,N-1),\mathbf{Q}^{\mathcal{N}(\mathbf{k-1}),\mathcal{E}},p_{01},p_{01})$$
$$-W_{t+1}(p_{01},\mathbf{P}_{\mathbf{11}}^{\mathcal{E}},\mathbf{\Phi}(k+1,N-1),\mathbf{Q}^{\mathcal{N}(\mathbf{k-1}),\mathcal{E}},p_{01})]$$
$$\leq (1-p_{01})\beta\Delta_{max}$$

where, the second equality follows Lemma 3, the first inequality follows the induction result of Lemma 6 and the second inequality follows the induction result of Lemma 7.

Combing the above results of the four terms, we have

$$W_t(\omega_1,\cdots,\omega_N) - W_t(\omega_n,\omega_1,\cdots,\omega_{N-1})$$
$$\leq \omega_k(1-\omega_N)\cdot\Delta_{max} + (1-\omega_k)(1-\omega_N)\cdot(1-p_{01})\beta\Delta_{max}$$
$$\leq \omega_k(1-\omega_N)\Delta_{max} + (1-\omega_k)(1-\omega_N)\Delta_{max} \leq (1-\omega_N)\Delta_{max},$$

which completes the proof of the first part of Lemma 7.

Finally, we prove the second part of Lemma 7. To this end, denote $\mathcal{M} \triangleq \{2,\cdots,k\}$, we have

$$W_t(\omega_1,\cdots,\omega_N) - W_t(\omega_N,\omega_2,\cdots,\omega_{N-1},\omega_1)$$
$$= (\omega_1-\omega_N)[W_t(1,\omega_2,\cdots,\omega_{N-1},0) - W_t(0,\omega_2,\cdots,\omega_{N-1},1)]$$
$$= (\omega_1-\omega_N)(F(1,\omega_2,\cdots,\omega_k) - F(0,\omega_2,\cdots,\omega_k) + \beta\sum_{\mathcal{E}\subseteq\mathcal{M}} Pr(\mathcal{M},\mathcal{E})\cdot$$
$$[(1-\epsilon)W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}},p_{11},\mathbf{\Phi}(k+1,N-1),p_{01},\mathbf{Q}^{\mathcal{M},\mathcal{E}}) + \epsilon W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}},\mathbf{\Phi}(k+1,N-1),p_{01},p_{11},\mathbf{Q}^{\mathcal{M},\mathcal{E}})$$
$$-W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}},\mathbf{\Phi}(k+1,N-1),p_{11},p_{01},\mathbf{Q}^{\mathcal{M},\mathcal{E}})]$$
$$\leq (\omega_1-\omega_N)(\Delta_{max} + \beta\sum_{\mathcal{E}\subseteq\mathcal{M}} Pr(\mathcal{M},\mathcal{E})[(1-\epsilon)W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}},p_{11},\mathbf{\Phi}(k+1,N-1),p_{01},\mathbf{Q}^{\mathcal{M},\mathcal{E}})$$
$$+\epsilon W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}},\mathbf{\Phi}(k+1,N-1),p_{01},p_{11},\mathbf{Q}^{\mathcal{M},\mathcal{E}}) - W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}},\mathbf{\Phi}(k+1,N-1),p_{01},p_{11},\mathbf{Q}^{\mathcal{M},\mathcal{E}})])$$
$$= (\omega_1-\omega_N)(\Delta_{max} + \beta\sum_{\mathcal{E}\subseteq\mathcal{M}} Pr(\mathcal{M},\mathcal{E})[(1-\epsilon)W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}},p_{11},\mathbf{\Phi}(k+1,N-1),p_{01},\mathbf{Q}^{\mathcal{M},\mathcal{E}})$$
$$-(1-\epsilon)W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}},\mathbf{\Phi}(k+1,N-1),p_{01},p_{11},\mathbf{Q}^{\mathcal{M},\mathcal{E}}])$$
$$\leq (\omega_1-\omega_N)(\Delta_{max} + \beta\sum_{\mathcal{E}\subseteq\mathcal{M}} Pr(\mathcal{M},\mathcal{E})[(1-\epsilon)W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}},p_{11},\mathbf{\Phi}(k+1,N-1),\mathbf{Q}^{\mathcal{M},\mathcal{E}},p_{01}) -$$
$$(1-\epsilon)W_{t+1}(p_{01},\mathbf{P}_{\mathbf{11}}^{\mathcal{E}},\mathbf{\Phi}(k+1,N-1),\mathbf{Q}^{\mathcal{M},\mathcal{E}},p_{11})])$$
$$\leq (p_{11}-p_{01})\left[\Delta_{max} + \beta\sum_{\mathcal{E}\subseteq\mathcal{M}} Pr(\mathcal{M},\mathcal{E})(1-\epsilon)\frac{1-[\beta(1-\epsilon)(p_{11}-p_{01})]^{T-t}}{1-\beta(1-\epsilon)(p_{11}-p_{01})}(p_{11}-p_{01})\Delta_{max}\right]$$
$$= \frac{1-[\beta(1-\epsilon)(p_{11}-p_{01})]^{T-t+1}}{1-\beta(1-\epsilon)(p_{11}-p_{01})}(p_{11}-p_{01})\Delta_{max}$$

where the first two inequalities follows the recursive application of the induction result of Lemma 6, the third inequality follows the induction result of Lemma 7.

We thus complete the whole process of proving Lemma 6 and Lemma 7.

REFERENCES

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

[2] K. Wang and L. Chen. On optimality of myopic policy for restless multi-armed bandit problem: An axiomatic approach. *IEEE Transactions on Signal Processing*, 99, 2011.

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

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

[5] K. Wang and L. Chen. On the optimality of myopic sensing in multi-channel opportunistic access: the case of sensing multiple channels. *In submission to IEEE Transactions on Communication, available on Computing Research Repository (CoRR) arXiv:1103.1784v1*, 2011.

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

[8] R. R. Weber and G. Weiss. On an index policy for restless bandits. *Journal of Applied Probability*, 27(1):637–648, 1990.

[9] K. Liu and Q. Zhao. Indexability of restless bandit problems and optimality of whittle index for dynamic multichannel access. *IEEE Transactions on Information Theory*, 56(11):5547–5567, 2010.

[10] S. Guha and K. Munagala. Approximation algorithms for partial-information based stochastic control with markovian rewards. In *Proc. IEEE Symposium on Foundations of Computer Science (FOCS)*, Providence, RI, 2007.

[11] S. Guha and K. Munagala. Approximation algorithms for restless bandit problems. In *Proc. ACM-SIAM Symposium on Discrete Algorithms (SODA)*, New York, 2009.

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

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

[14] K. Wang Q. Liu and L. Chen. On optimality of greedy policy for a class of standard reward function of restless multi-armed bandit problem. *available on Computing Research Repository (CoRR) arXiv:1104.53911*, 2011.