Opportunistic Spectrum Access by Exploiting Primary User Feedbacks in Underlay Cognitive Radio Systems: An Optimality Analysis

Kehao Wang, Lin Chen, *Member, IEEE*, and Quan Liu

***Abstract*—We consider an underlay cognitive radio (CR) communication system in which a cognitive secondary user (SU) can access multiple primary spectrum channels only when its interference to the primary users (PU) is limited. To identify and exploit instantaneous transmission opportunities, the SU probes a subset of primary channels by overhearing the primary feedback signals so as to learn the primary receiver's channel condition and the interference tolerance level, then chooses appropriate power to transmit its data. In such context, the SU cannot probe all the channels for its limited number of receiving antennas, then a crucial optimization problem faced by the SU is to probe which channel(s) in order to maximize the long-term throughput given the past probing history. In this paper, we tackle this optimization problem by casting it into a restless multi-armed bandit (RMAB) problem that is of fundamental importance in decision theory. Given the specific and practical constraints posed by the problem, we analyze the myopic probing policy which consists of probing the best channels based on the past observation.We perform an analytical study on the optimality of the developed myopic probing policy. Specifically, for a family of generic and practically important utility functions, we establish the closed-form conditions to guarantee the optimality of the myopic probing policy, and also illustrate our analytical results via simulations on several typical network scenarios.**

***Index Terms*—Cognitive radio, opportunistic spectrum access, network feedback, restless multi-armed bandit problem**

I. Introduction

A. OSA in Underlay Cognitive Radio Systems

Due to the rapid growth of wireless communications in recent years, so far almost all the exploitable spectrum bands have been allocated for various wireless applications in different regions. Meanwhile, significant under-utilization of licensed spectrum bands at a given time or a given location has been widely observed. To increase spectrum utilization efficiency, the idea of cognitive radio (CR) has been proposed, whose core idea is opportunistic spectrum access (OSA) where unlicensed secondary users (SU) can utilize the spectrum when the licensed primary users (PU) are not using it (the overlay CR approach [1], [2], [3], [4]) or the interference generated by the SUs to the traffic of PUs is limited under certain threshold (the underlay CR approach [5], [6], [7]).

In our work, we focus on an underlay CR communication system composed of a set of primary channels where PU transmitting packets continuously[1]. An SU, equipped with one or multiple receiving and transmitting antennas, can transmit its data packets on one or a subset of the primary channels opportunistically when the interference it generates to PUs is limited. In order to exploit instantaneous transmission opportunities in such underlay CR network, the SU should learn the instantaneous channel information of the primary channel(s). To this end, it probes a subset of primary channels by overhearing the feedback signals on them so as to learn the primary receiver's channel condition and the interference tolerance level, and then chooses appropriate power to transmit its data. Note that today most practical wireless communication systems have built-in receiver feedback such as the transmission power level control signals in the IS-95 cellular systems, the ACK/NACK feedback packets in WiFi and the CQI-CDI messages in 4G wireless systems. Under such context, since the number of channels probed by the SU is usually limited (by the number of receiving and transmitting antennas), a crucial optimization problem is which channel(s) to probe in order to maximize the long-term utility (e.g., expected throughput), given the past observations.

B. Restless Multi-Armed Bandit Formulation

To formulate the problem posed above, we model each primary channel as an i.i.d. two-state discrete-time Markov chain. The state *good* corresponds to the situation with high signal to interference and noise ratio (SINR) while the state *bad* represents the low SINR situation due to fading or high background noise. The SU seeks a set of primary channels to probe by exploiting past observations and the knowledge of the stochastic properties of the channels with the objective of maximizing its long-term utility (e.g., expected throughput). Mathematically, the considered channel probing problem can be cast into the restless multi-armed bandit (RMAB) problem of fundamental importance in decision theory.

This work was supported in part by the ANR (Agence Nationale de la Recherche) under the grant Green-Dyspan (ANR-12-IS03), the National Science Foundation of China under Grant 51175389 and 60970019.

K. Wang and Q. Liu are with Key Laboratory of Fiber Optic Sensing Technology and Information Processing, and the school of Information, the Wuhan University of technology, 430070 Hubei, China (e-mail: Kehao.wang@whut.edu.cn;quanliu@whut.edu.cn).

L. Chen is with the Laboratoire de Recherche en Informatique (LRI), Department of Computer Science, the University of Paris-Sud, 91405 Orsay, France (e-mail: lin.chen@lri.fr).

[1]We adopt a worst-case assumption that PUs transmit all the time which is commonly used in analyzing underlay CR systems. In other words, our analysis does not rely on detection and exploitation of spectrum white space, which is the case of overlay CR systems widely investigated recently.

In the classic RMAB problem, a player chooses k out of N arms, each evolving as a Markov chain, to activate each time and receives a reward determined by the states of the activated arms. The objective is to maximize the long-run reward over an infinite horizon by choosing which k arms to activate each time. If only the activated arms change their states, the problem is degenerated to the multi-armed bandit (MAB) problem [8].

The MAB problem is solved by Gittins by showing that the optimal policy has an index structure [8], [9]. However, the RMAB problem is proved to be PSPACE-Hard [10]. Hence, a natural alternative is to seek a simple myopic policy maximizing the short-term reward. However, the optimality of a myopic policy is not always guaranteed. In our work, given the specific and practical constraints posed by the channel probing problem, we analyze the myopic probing policy which consists of probing the best channels based on the past observations. Especially, for a family of generic utility functions, we establish the closed-form conditions to guarantee the optimality of the myopic probing policy.

C. Related Work

Recently there are two major thrusts in the study of the myopic policy of the RMAB problem. Since the optimality of the myopic policy is not generally guaranteed, the first research thrust is to study how far it is to the optimal and design approximation algorithms as well as heuristic policies. The works [11], [12], [13] follow this line of research. Specifically, a simple myopic policy, also called greedy policy, is developed in [11] that yields a factor 2 approximation of the optimal policy for a subclass of scenarios referred to as *Monotone bandits*. The other thrust, more application-oriented, consists of establishing the optimality of the myopic policy in some specific application scenarios, particularly in the context of OSA. The works [14], [15], [16], [17], [18] belong to this category by focusing on specific forms of reward functions. Especially, Zhao *et al.* [14] established the structure of the myopic sensing policy, analyzed the performance, and partly obtained the optimality for the case of i.i.d. channels. Ahmad and Liu *et al.* [15] derived the optimality of the myopic sensing policy for the positively correlated i.i.d. channels when the SU is limited to access one channel (i.e., $k = 1$) each time, and further extended the optimality to the case of sensing multiple i.i.d. channels ($k > 1$) [16] for the scenario where the SU gets one unit of reward for each channel sensed good. Liu and Zhao *et al.* [17] proved the optimality of the myopic policy for the case of two channels with a particular utility function and conjectured it for arbitrary N. In our previous work [19], [20], [18], we first showed that the myopic policy is not optimal generally [19], then focused on a family of *regular* functions [20], and extended i.i.d. channels to non i.i.d. ones [18], and derived closed-form conditions under which the myopic sensing policy is ensured to be optimal.

Related works on exploiting primary control feedback are presented in [21], [22], [23], [24], [25]. Specifically, the authors [21] introduced an information-theoretic model of this basic observation and developed a generic strategy where the SU monitors the PU's effective packet rate by listening to ARQs and only transmits when that rate is above a threshold. That is, the work [21] is focused on the problem when to transmit while our work presented in this paper addresses the problem which channel(s) to probe and transmit. The authors [22] studied the myopic channel probing policy for the similar scenario proposed, but only established its optimality in the particular case of probing one channel ($k = 1$) each time. In our previous work [23], we established the optimality of myopic policy for the case of probing $N-1$ of N channels each time and analyzed the performance of the myopic probing policy by domination theory. However, this work studies the generic case of arbitrary k and also derives more strong conditions on the optimality by dropping one of the non-trivial conditions of [22]. From the angle of renewal theory, [24] studied the discovery of spectrum opportunities and the delay in finding an available channel, while [25] studied the interference to PU from the dynamic access of SU in the context of unknown primary behavior.

Compared with the previous work [18], our work treats a different scenario where the major technical difficulty lies in the non-linearity of the belief value update function, as detailed in Section II, and consequently a novel analysis is proposed to establish the closed-form condition on the optimality of the myopic channel probing policy and does not lose the optimality for the reward function with linear combination of the states of the selected channels. But in [18], part of the optimality is sacrificed, as stated in Section VII of [18], in order to cover the heterogeneous channels. The technical difference from [26] is that the belief vector is divided into 'value belief vector' and 'policy belief vector', which reflects the essence of decomposability. From the viewpoint of the RMAB problem, the optimality condition derived in this paper can be degenerated to those obtained in the literature [14], [15], [16], [17], [26] by relaxing some constraints.

The rest of the paper is organized as follows: Section II formulates the model and establishes the myopic channel probing policy. Section III studies the optimality of the myopic channel probing policy. Section IV illustrates the analytical results via a set of extensive simulation study on several typical network scenarios. The paper is concluded by Section V.

II. Problem Formulation

In this section, we describe the system model of the spectrum access in underlay CR model, based on which we formulate the RMAB-based channel probing problem and derive the myopic channel probing policy.

A. System Model

Primary System

As outlined in the Introduction, we consider a slotted multi-channel underlay CR communication system composed of N primary channels, each evolving as an i.i.d. Markov chain of two states, *good* (1) and *bad* (0), corresponding to the situation with high (low, respectively) SINR as shown in Fig. 1. The channel state transition matrix $\mathbf{P}$ is given as follows

$$\mathbf{P} = \begin{bmatrix} p_{11} & 1-p_{11} \\ p_{01} & 1-p_{01} \end{bmatrix}.$$

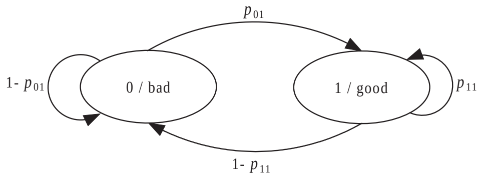

Fig. 1. Markov Channel Model

In our work, we focus on the positively correlated channel setting (i.e., $p_{11} > p_{01}$) which corresponds to the realistic scenarios where the channel states are observed to evolve gradually over time.

On the primary traffic, we adopt the commonly used worst-case assumption that primary users transmit all the time on each channels with the primary transmitter and receiver on channel i denoted as PTx i and PRx i, respectively.

Let $\mathbf{S}(t) \triangleq [S_1(t), \cdots, S_N(t)]$ denote the channel state vector where $S_i(t) \in \{0 \text{ (bad)}, 1 \text{ (good)}\}$ is the state of channel i in slot t. We define *outage* as data-packet decoding failure at PRx i. We denote the probability of the outage event as a function of the channel state $S_i(t)$:

$$O_s(i) \triangleq \Pr(\text{decoding failure} \mid S_i(t) = s), \ s \in \{1, 0\}, \ i \in \mathcal{N}.$$

It can be straightforwardly noted that $0 \leq O_0(i) < O_1(i) \leq 1$, $\forall i \in \mathcal{N}$. In our analysis, we focus on the case where $O_s(i)$ is independent w.r.t. i, and denote O_s as the system outage probability. Each PRx i sends an acknowledgement (ACK) to the corresponding PTx i on channel i at the end of each slot if the packet is successfully decoded. Thus the absence of an acknowledgement (denoted as negative ACK or NACK) signifies that the *outage* event happens on channel i at slot t.

Underlay Secondary System

We consider an SU, equipped with k ($1 \leq k < N$) receiving and transmitting antennas (denoted as STx and RTx), can transmit its data packets on k channels opportunistically when the interference that it generates to PUs is limited. In order to exploit instantaneous transmission opportunities, the SU probes k primary channels by overhearing the feedback signals on them so as to learn the primary receivers channel condition and the interference tolerance level before deciding whether it can transmit its data on the probed channels.

Specifically, when the SU decides to probe channel i, it exploits the primary feedbacks by overhearing the ACK/NACK packets to estimate the primary channel condition. We denote the set of channels probed by the SU at slot t as $\mathcal{A}(t)$ ($|\mathcal{A}(t)| = k$). The observation of a probed channel i in slot t is defined as $K_i(t) \in \{0 \text{ (NACK)}, 1 \text{ (ACK)}\}$. Throughout our analysis we assume that the SU can perfectly overhear the ACK/NACK packets on channel i once it decides to probe channel i. This is a reasonable assumption as the ACK/NACK packets are usually transmitted in a more robust way at lower data rate. We leave the generic case of imperfect overhearing for future investigation.

B. Optimal Channel Probing Problem Formulation and Myopic Probing Policy

Since the SU can only probe k channel each slot, the channel state vector $\mathbf{S}(t)$ is only partially observable to the SU for its decision. In this regard, we define the channel state belief vector $\Omega(t) \triangleq \{\omega_i(t), i \in \mathcal{N}\}$ (referred to as belief vector) to denote the estimation at the SU on the channel state for next slot, where $\omega_i(t)$ is the estimated conditional probability that the primary channel $i \in \mathcal{N}$ is in *good* state ($S_i(t) = 1$) given all the past observations and decisions of the SU. Given the belief vector $\Omega(t)$ and the probed channel set $\mathcal{A}(t)$, the belief vector can be updated recursively based on the primary feedback observation $\{K_i(t) : i \in \mathcal{A}(t)\}$ using the following Bayes rule (1).

$$\omega_i(t+1) = \begin{cases} \mathcal{T}(\phi(\omega_i(t))), & i \in \mathcal{A}(t), K_i(t) = 1 \\ \mathcal{T}(\varphi(\omega_i(t))), & i \in \mathcal{A}(t), K_i(t) = 0 \\ \mathcal{T}(\omega_i(t)), & i \notin \mathcal{A}(t), \end{cases} \tag{1}$$

where, the operators $\phi(\cdot)$, $\varphi(\cdot)$ and $\mathcal{T}(\cdot)$ are defined as follows:

$$\phi(x) \triangleq \frac{(1-O_1)x}{(1-O_1)x + (1-O_0)(1-x)},$$
$$\varphi(x) \triangleq \frac{O_1 x}{O_1 x + O_0(1-x)},$$
$$\mathcal{T}(x) \triangleq (p_{11} - p_{01})x + p_{01}.$$

Note that the numerator and denominator of $\phi(x)$ ($\varphi(x)$) represents the probability of successful decoding (decoding failure) with $S_i(t) = 1$ and that of successful decoding (decoding failure), respectively. $\mathcal{T}(x)$ is the Markovian evolving rule.

Remark. We emphasize that the mapping from $\omega_i(t)$ to $\omega_i(t+1)$ is not linear (cf. the first and second lines of equation (1)), and depends not only on the channel evolution but also on the observation outcome. As will be shown later, this non-linearity makes the analysis on the optimal channel probing policy much more involved and calls for an original study.

A *channel probing policy* π is defined as a sequence of mappings $\pi = [\pi_1, \pi_2, \cdots, \pi_T]$ where π_t maps the belief vector $\Omega(t)$ to the action $\mathcal{A}(t)$ (i.e., the set of channels to probe) in each slot t: i.e.,

$$\pi_t : \ \Omega(t) \mapsto \mathcal{A}(t), \ |\mathcal{A}(t)| = k.$$

In such underlay CR paradigm, we are interested in the SU's optimization problem to find the optimal policy π^* that maximizes the expected accumulated discounted reward over a finite time horizon:

$$\pi^* = \operatorname*{argmax}_{\pi} \mathbb{E}\left[\sum_{t=1}^{T} \beta^{t-1} R(\pi_t(\Omega(t))) \middle| \Omega(1) \right] \tag{2}$$

where $R(\pi_t(\Omega(t)))$ is the reward in slot t under the policy π_t with the initial belief vector $\Omega(1)^2$, $0 \leq \beta \leq 1$ is the discount factor characterizing the feature that future reward is less valuable than immediate reward. By treating the belief value of each channel as the state of each arm of a bandit, the SU's optimization problem can be cast into a RMAB problem.

To get more insight on the structure of the optimization problem (2) and the complexity to solve it, we derive the dynamic programming formulation of (2) as follows:

$$V_T(\Omega(T)) = \max_{\mathcal{A}(T)} \mathbb{E}\big[R(\pi_t(\Omega(T)))\big],$$

[2]If no information on the initial system state is available, each entry of $\Omega(1)$ can be set to the stationary distribution $\omega_0 = \frac{p_{01}}{1+p_{01}-p_{11}}$.

$$V_t(\Omega(t)) = \max_{\mathcal{A}(t)} \mathbb{E}\Big[R(\pi_t(\Omega(t))) + \beta \sum_{\mathcal{E}\subseteq\mathcal{A}(t)} \prod_{i\in\mathcal{E}} [1 - O_1\omega_j(t) - O_0(1-\omega_j(t))] \cdot \prod_{j\in\mathcal{A}(t)\setminus\mathcal{E}} [O_1\omega_i(t) + O_0(1-\omega_i(t))] V_{t+1}(\Omega(t+1))\Big].$$

where, $V_t(\Omega(t))$ is the value function corresponding to the maximal expected reward from time slot t to T ($1 \le t \le T$) and $\Omega(t+1)$ follows the evolution described in (1) given that the channels in the subset $\mathcal{E}$ are observed in *good* state and the channels in $\mathcal{A}(t)\setminus\mathcal{E}$ are observed in *bad* state.

Theoretically, the optimal policy of (2) can be obtained by solving the above dynamic programming. It is infeasible, however, due to the impact of the current action on the future reward and the unaccountable space of the belief vector, and in fact obtaining the optimal solution directly from the above recursive equations is computationally prohibitive. Hence, a natural alternative is to seek a simple myopic policy maximizing the immediate reward which is easy to compute and implement, formally defined as follows:

Definition 1 (Myopic Channel Probing Policy)**.** *Let $F(\Omega_A(t)) \triangleq \mathbb{E}[R(\pi_t(\Omega(t)))]$ denote the expected immediate reward obtained in slot t under the policy π_t with $\Omega_A(t) \triangleq \{\omega_i(t) : i \in \mathcal{A}(t)\}$, the myopic channel probing policy consists of probing the k channels that maximizes $F(\Omega_A(t))$, i.e., $\overline{\mathcal{A}}(t) = \mathrm{argmax}_{\mathcal{A}(t)\subseteq\mathcal{N}} F(\Omega_A(t))$.*

To make our analysis more generic, we focus on a class of practically important reward functions $F(\Omega_A(t))$, termed as regular functions defined in [18]. More specifically, the expected immediate reward function $F(\Omega_A(t))$ studied in this paper is assumed to be symmetrical, monotonically non-decreasing and decomposable, as characterized by the three axioms in [18]. Under this condition, the myopic channel probing policy consists of choosing the k channels with the largest belief value. In the following sections we study the optimality of the myopic channel probing policy with the particularities posed by the underlay CR systems presented previously.

III. Analysis on Optimality of Myopic Channel Probing Policy

The goal of this section is to establish the closed-form condition under which the myopic channel probing policy, despite of its simple structure, achieves the system optimum. To facilitate the presentation and analysis, we first state the notations and parameters. We then define the auxiliary function and adjugate auxiliary function and study their structural properties which pave the way of the optimality analysis. The core result on the optimality of the myopic channel probing policy is then presented, followed by a discussion to illustrate the results in a concrete network scenario.

A. Notations

For the convenience of presentation, we first state the notations and parameters employed in the following analysis.

1) $\mathcal{N}(k) \triangleq \{1, \cdots, k\}$ denotes the first k channels in $\mathcal{N}$;
2) Given $\mathcal{E} \subseteq \mathcal{M} \subseteq \mathcal{N}$,
$$\mathcal{C}^{\mathcal{E}}_{\mathcal{M}} \triangleq \prod_{i\in\mathcal{E}} [1 - O_1\omega_j(t) - O_0(1-\omega_j(t))] \times \prod_{j\in\mathcal{M}(t)\setminus\mathcal{E}} [O_1\omega_i(t) + O_0(1-\omega_i(t))],$$
$$\widehat{\mathcal{C}}^{\mathcal{E}}_{\mathcal{M}} \triangleq \prod_{i\in\mathcal{E}} [1 - O_1\widehat{\omega}_j(t) - O_0(1-\widehat{\omega}_j(t))] \times \prod_{j\in\mathcal{M}(t)\setminus\mathcal{E}} [O_1\widehat{\omega}_i(t) + O_0(1-\widehat{\omega}_i(t))].$$
$\mathcal{C}^{\mathcal{E}}_{\mathcal{M}}$ ($\widehat{\mathcal{C}}^{\mathcal{E}}_{\mathcal{M}}$) denotes the expected probability that the channels in $\mathcal{E}$ are observed good, whereas those in $\mathcal{M}\setminus\mathcal{E}$ bad, given that the channels in $\mathcal{M}$ are observed;
3) Given $\mathcal{E} \subseteq \mathcal{M} \subseteq \mathcal{N}$, $\mathbf{\Phi}(\mathcal{E}) \triangleq [\mathcal{T}(\phi(\widehat{\omega}_i(t))), i \in \mathcal{E}]$ characterizes the updated belief values of the channels in $\mathcal{E}$ if they are observed in the good state; $\mathbf{\Phi}^l(\mathcal{E}) \triangleq [\mathcal{T}(\phi(\widehat{\omega}_i(t))), i \in \mathcal{E}, i < l]$ characterizes the updated belief values of the channels in $\mathcal{E}$ if they are observed in the good state with the channel index smaller than l; $\mathbf{\Phi}_m(\mathcal{E}) \triangleq [\mathcal{T}(\phi(\widehat{\omega}_i(t))), i \in \mathcal{E}, m \le i]$ characterizes the updated belief values of the channels in $\mathcal{E}$ if they are observed in the good state with the channel index larger than m; $\mathbf{\Phi}^l_m(\mathcal{E}) \triangleq [\mathcal{T}(\phi(\widehat{\omega}_i(t))), i \in \mathcal{E}, m \le i < l]$;
4) $\mathbf{\Upsilon}^m_l \triangleq [\mathcal{T}(\widehat{\omega}_i(t)), l \le i \le m]$ is the updated belief values of the channels between l and m if they are not observed;
5) Given $\mathcal{E} \subseteq \mathcal{M} \subseteq \mathcal{N}$, $\mathbf{\Psi}(\mathcal{M},\mathcal{E}) \triangleq [\mathcal{T}(\varphi(\widehat{\omega}_i(t))), i \in \mathcal{M} \setminus \mathcal{E}]$ characterizes the updated belief values of the channels in $\mathcal{M}\setminus\mathcal{E}$ if they are observed in the bad state; $\mathbf{\Psi}^l(\mathcal{M},\mathcal{E}) \triangleq [\mathcal{T}(\varphi(\widehat{\omega}_i(t))), i \in \mathcal{M}\setminus\mathcal{E}, i < l]$ characterizes the updated belief values of the channels in $\mathcal{M}\setminus\mathcal{E}$ if they are observed in the bad state with the channel index smaller than l; $\mathbf{\Psi}_m(\mathcal{M},\mathcal{E}) \triangleq [\mathcal{T}(\varphi(\widehat{\omega}_i(t))), i \in \mathcal{M}\setminus\mathcal{E}, m \le i]$ characterizes the updated belief values of the channels in $\mathcal{M}\setminus\mathcal{E}$ if they are observed in the bad state with the channel index larger than m; $\mathbf{\Psi}^l_m(\mathcal{M},\mathcal{E}) \triangleq [\mathcal{T}(\varphi(\widehat{\omega}_i(t))), i \in \mathcal{M}\setminus\mathcal{E}, m \le i < l]$;
6) Let $\widehat{\omega}_{-i} \triangleq \{\widehat{\omega}_j : j \in \mathcal{A}, j \ne i\}$ and
$$\begin{cases} \Delta_{max} \triangleq \max\limits_{\widehat{\omega}_{-i}\in[0,1]^{k-1}} \{F(1,\widehat{\omega}_{-i}) - F(0,\widehat{\omega}_{-i})\}, \\ \Delta_{min} \triangleq \min\limits_{\widehat{\omega}_{-i}\in[0,1]^{k-1}} \{F(1,\widehat{\omega}_{-i}) - F(0,\widehat{\omega}_{-i})\}, \end{cases}$$
7) $\dot{\Omega} = (\dot{\omega}_1, \cdots, \dot{\omega}_N)$, where $p_{11} \ge \dot{\omega}_1 \ge \cdots \ge \dot{\omega}_N \ge p_{01}$.

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

Lemma 1. *If $p_{11} > p_{01}$, it holds that*

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

Proof: Lemma 1 holds straightforwardly from $\mathcal{T}(\omega_i(t)) = (p_{11} - p_{01})\omega_i(t) + p_{01}$. ∎

Lemma 2. *If $0 \le \frac{O_1}{O_0} \le \frac{(1-p_{11})p_{01}}{p_{11}(1-p_{01})}$, it holds that*

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

Proof: Noticing that $\varphi(\omega_i) = \frac{O_1\omega_i(t)}{O_1\omega_i(t)+O_0(1-\omega_i(t))}$, Lemma 2 follows straightforwardly. ■

Lemma 3. *If* $0 \le \frac{O_1}{O_0} \le \frac{(1-p_{11})p_{01}}{p_{11}(1-p_{01})}$, *let* $\zeta = \dfrac{1-O_1}{1-O_0}$, *it holds:*

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

Proof: Noticing that $\phi(\omega_i) = \frac{\zeta\omega_i(t)}{\zeta\omega_i(t)+1-\omega_i(t)}$ and $\zeta > 1$, Lemma 3 follows straightforwardly. ■

We would like to point out that when the initial belief $\omega_i(1)$ is set to $\frac{p_{01}}{p_{01}+1-p_{11}}$ as is often the case in practical systems, it can be checked that $p_{01} \le \omega_i(1) \le p_{11}$ 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 we always assume that the initial belief is located in $[p_{01}, p_{11}]$ for the ease of discussion.

B. Auxiliary and Adjugate Auxiliary Value Functions: Definition and Properties

In this subsection, we first define the auxiliary and the adjugate auxiliary value functions and then derive several fundamental properties of them, which are crucial in the study on the optimality of the myopic channel probing policy.

Definition 2 (Auxiliary Value Function (AVF) and Adjugate Auxiliary Value Function (AAVF))**.** *The AVF* $W_t(\Omega(t))$ *and AAVF* $\widehat{W}_t(\Omega(t);\widehat{\Omega}(t))$ *(*$1 \le t \le T$, $t+1 \le r \le T$*) are defined as follows:*

$$\text{AVF}\begin{cases} W_T(\Omega(T)) = F(\Omega_{\overline{\mathcal{A}}}(T)) \\ W_r(\Omega(r)) = F(\Omega_{\overline{\mathcal{A}}}(r)) \\ \qquad +\beta\sum_{\mathcal{E}\subseteq\overline{\mathcal{A}}(r)} \mathcal{C}^{\mathcal{E}}_{\overline{\mathcal{A}}(r)} W_{r+1}(\Omega_{\mathcal{E}}(r+1)) \\ W_t(\Omega(t)) = F(\Omega_{\mathcal{N}(k)}(t)) \\ \qquad +\beta\sum_{\mathcal{E}\subseteq\mathcal{N}(k)} \mathcal{C}^{\mathcal{E}}_{\mathcal{N}(k)} W_{t+1}(\Omega_{\mathcal{E}}(t+1)), \end{cases} \tag{3}$$

$$\text{AAVF}\begin{cases} \widehat{W}_T(\Omega(T);\widehat{\Omega}(T)) = F(\widehat{\Omega}_{\overline{\mathcal{A}}}(T)) \\ \widehat{W}_r(\Omega(r);\widehat{\Omega}(r)) = F(\widehat{\Omega}_{\overline{\mathcal{A}}}(r)) \\ \quad +\beta\sum_{\mathcal{E}\subseteq\overline{\mathcal{A}}(r)} \widehat{\mathcal{C}}^{\mathcal{E}}_{\overline{\mathcal{A}}(r)} \widehat{W}_{r+1}(\Omega_{\mathcal{E}}(r+1);\widehat{\Omega}_{\mathcal{E}}(r+1)) \\ \widehat{W}_t(\Omega(t);\widehat{\Omega}(t)) = F(\widehat{\Omega}_{\mathcal{N}(k)}(t)) \\ \quad +\beta\sum_{\mathcal{E}\subseteq\mathcal{N}(k)} \widehat{\mathcal{C}}^{\mathcal{E}}_{\mathcal{N}(k)} \widehat{W}_{t+1}(\Omega_{\mathcal{E}}(t+1);\widehat{\Omega}_{\mathcal{E}}(t+1)), \end{cases} \tag{4}$$

where

1) $\Omega_{\mathcal{E}}(t+1)$ *and* $\Omega_{\mathcal{E}}(r+1)$ *are generated by* $\langle\Omega(t), \mathcal{N}(k), \mathcal{E}\rangle$ *and* $\langle\Omega(r), \overline{\mathcal{A}}(r), \mathcal{E}\rangle$*, respectively, according to* (1)*, and then sorted by belief value;*
2) $\widehat{\Omega}_{\mathcal{E}}(t+1)$ *and* $\widehat{\Omega}_{\mathcal{E}}(r+1)$ *are generated by* $\langle\widehat{\Omega}(t), \mathcal{N}(k), \mathcal{E}\rangle$ *and* $\langle\widehat{\Omega}(r), \overline{\mathcal{A}}(r), \mathcal{E}\rangle$*, respectively, according to* (1)*, and the order of channel index keeps consistent with that of* $\Omega_{\mathcal{E}}(t+1)$ *and* $\Omega_{\mathcal{E}}(r+1)$*, respectively.*
3) $\overline{\mathcal{A}}(r)$ *and* $\mathcal{N}(k)$ *of AAVF are the same with that of AVF.*

Remark. It is insightful to note the following engineering implications that hinge behind the above definition:

1) AVF gives the expected discounted accumulated reward of the following policy: in slot t probe the first k channels in the belief vector and then probe the channels in $\overline{\mathcal{A}}(r)$ $(t+1 \le r \le T)$ (i.e., adopt the myopic channel probing policy from slot $t+1$ to T). If $\overline{\mathcal{A}}(t) = \mathcal{N}(k)$, then $W_t(\Omega(t))$ is the total reward from slot t to T under the myopic channel probing policy.
2) AAVF, as an adjugate of AVF, is introduced to show the different functions of 'policy' and 'value'. That is, its policy is determined by the policy belief vector $\Omega(t)$ while its value is determined by the value belief vector $\widehat{\Omega}(t)$. If $\widehat{\Omega}(t) = \Omega(t)$, then AAVF degenerates to AVF.

In the subsequent analysis, we derive some important structural properties of AVF and AAVF.

Lemma 4 (Belief Substitution of AAVF)**.** *Given two policy belief vectors* $\Omega = (\omega_1, \omega_2, \cdots, \omega_N)$, $\Omega' = (\omega'_1, \omega'_2, \cdots, \omega'_N)$, *if* $p_{11} \ge \omega_1 \ge \omega_2 \ge \cdots \ge \omega_N \ge p_{01}$ *and* $p_{11} \ge \omega'_1 \ge \omega'_2 \ge \cdots \ge \omega'_N \ge p_{01}$, *it holds that* $\widehat{W}_t(\Omega;\widehat{\Omega}) = \widehat{W}_t(\Omega';\widehat{\Omega})$.

Proof: It is easily to verify that the lemma holds for slot T. Assume that it holds for $t+1, \cdots, T-1$. We now prove it also holds for slot t.

Recall (4), it suffices to prove $\widehat{W}_{t+1}(\Omega_{\mathcal{E}}(t+1);\widehat{\Omega}_{\mathcal{E}}(t+1)) = \widehat{W}_{t+1}(\Omega'_{\mathcal{E}}(t+1);\widehat{\Omega}_{\mathcal{E}}(t+1))$ under given $\mathcal{E}$. According to the induction hypothesis, it is sufficient to show that when $\Omega_{\mathcal{E}}(t+1)$ and $\Omega'_{\mathcal{E}}(t+1)$ are sorted in the decreasing order of their elements, they have the same order of channel index. Let $\{\sigma_1, \cdots, \sigma_k\}$ denote any permutation of $\{1, \cdots, k\}$ and $\mathcal{E} = \{\sigma_1, \cdots, \sigma_m\}$ at slot t $(0 \le m \le k,\ \sigma_1 \le \cdots \le \sigma_m,\ \sigma_{m+1} \le \cdots \le \sigma_k)$, according to Lemma 1, 2 and 3, we have $\mathcal{T}(\phi(\omega_{\sigma_1})) \ge \cdots \ge \mathcal{T}(\phi(\omega_{\sigma_m})) \ge \mathcal{T}(\phi(\omega_k)) > \mathcal{T}(\omega_k) \ge \mathcal{T}(\omega_{k+1}) \ge \cdots \ge \mathcal{T}(\omega_N) \ge p_{01} \ge \mathcal{T}(\varphi(\omega_{\sigma_{m+1}})) \ge \cdots \ge \mathcal{T}(\varphi(\omega_{\sigma_k}))$, that is to say, the order of channel index of $\Omega_{\mathcal{E}}(t+1)$ is $(\sigma_1, \cdots, \sigma_m, k+1, \cdots, N, \sigma_{m+1}, \cdots, \sigma_k)$. Similarly, the order of channel index of $\Omega'_{\mathcal{E}}(t+1)$ is also $(\sigma_1, \cdots, \sigma_m, k+1, \cdots, N, \sigma_{m+1}, \cdots, \sigma_k)$. ■

Remark. Lemma 4 implies that by substituting a decreasingly sorted policy belief vector by another one, AAVF remains the same and hence generates the same reward.

Lemma 5 (Symmetry of AVF)**.** *If* $0 \le \frac{O_1}{O_0} \le \frac{(1-p_{11})p_{01}}{p_{11}(1-p_{01})}$ *and* F *is regular, it holds that* $W_t(\Omega(t))$ *is symmetrical in* ω_i, ω_j *for any* $i, j \in \mathcal{A}(t)$ *or* $i, j \notin \mathcal{A}(t)$ *for all* $t = 1, 2, \cdots, T$, *i.e.,*

$$\begin{aligned} &W_t(\omega_1, \cdots, \omega_i, \cdots, \omega_j, \cdots, \omega_N) \\ &= W_t(\omega_1, \cdots, \omega_j, \cdots, \omega_i, \cdots, \omega_N). \end{aligned}$$

Proof: The proof is given in Appendix A. ■

Remark. Lemma 5 implies that the reward generated by AVF remains the same against any channel permutation within the probed channels and within the non-probed channels.

Lemma 6 (Decomposability of AAVF)**.** *If* $0 \le \frac{O_1}{O_0} \le \frac{(1-p_{11})p_{01}}{p_{11}(1-p_{01})}$ *and* F *is regular, for any policy belief vector* Ω, *it holds that* $\widehat{W}_t(\Omega;\widehat{\Omega}(t))$ *is decomposable for all* $t = 1, 2, \cdots, T$, *i.e.,*

$$\widehat{W}_t(\Omega;\widehat{\Omega}) = \widehat{\omega}_l W_t(\Omega;\widehat{\Omega}_1) + (1-\widehat{\omega}_l) W_t(\Omega;\widehat{\Omega}_0), \forall l \in \mathcal{N},$$

where, $\widehat{\Omega} = (\widehat{\omega}_1, \cdots, \widehat{\omega}_l, \cdots, \widehat{\omega}_N)$,
$\widehat{\Omega}_0 = (\widehat{\omega}_1, \cdots, 0, \cdots, \widehat{\omega}_N)$, $\widehat{\Omega}_1 = (\widehat{\omega}_1, \cdots, 1, \cdots, \widehat{\omega}_N)$.

Proof: The proof is given in Appendix B. ■

Remark. Lemma 6 states that given the probing policy, the reward attained from AAVF can be decomposed into two terms with deterministic realizations 0 and 1 in any channel of the value belief vector.

Lemma 6 can be applied one step further to prove the following corollary.

Corollary 1. *If* $0 \leq \frac{O_1}{O_0} \leq \frac{(1-p_{11})p_{01}}{p_{11}(1-p_{01})}$ *and* F *is regular, for any belief vector* Ω*, it holds that* $\forall l, m \in \mathcal{N}, t = 1, \cdots, T$

$$\widehat{W}_t(\Omega;\widehat{\Omega}_0) - \widehat{W}_t(\Omega;\widehat{\Omega}_1) = (\widehat{\omega}_l - \widehat{\omega}_m)\Big[W_t(\Omega;\widehat{\Omega}_2) - W_t(\Omega;\widehat{\Omega}_3)\Big]$$

where

$$\widehat{\Omega}_0 = (\widehat{\omega}_1, \cdots, \widehat{\omega}_l, \cdots, \widehat{\omega}_m, \cdots, \widehat{\omega}_N),$$
$$\widehat{\Omega}_1 = (\widehat{\omega}_1, \cdots, \widehat{\omega}_m, \cdots, \widehat{\omega}_l, \cdots, \widehat{\omega}_N),$$
$$\widehat{\Omega}_2 = (\widehat{\omega}_1, \cdots, 1, \cdots, 0, \cdots, \widehat{\omega}_N),$$
$$\widehat{\Omega}_3 = (\widehat{\omega}_1, \cdots, 0, \cdots, 1, \cdots, \widehat{\omega}_N).$$

Lemma 7 (Monotonicity of AAVF)**.** *If* $0 \leq \frac{O_1}{O_0} \leq \frac{(1-p_{11})p_{01}}{p_{11}(1-p_{01})}$ *and* F *is regular, it holds that* $\widehat{W}_t(\dot{\Omega};\Omega(t))$ *is monotonously non-decreasing in* ω_l, $\forall l \in \mathcal{N}$*, i.e.,*

$$\widehat{\omega}_l' \geq \widehat{\omega}_l \Longrightarrow \widehat{W}_t(\dot{\Omega};\widehat{\Omega}_0) \geq W_t(\dot{\Omega};\widehat{\Omega}_1),$$

where,

$$\widehat{\Omega}_0 = (\widehat{\omega}_1, \cdots, \widehat{\omega}_l', \cdots, \widehat{\omega}_N), \ \widehat{\Omega}_1 = (\widehat{\omega}_1, \cdots, \widehat{\omega}_l, \cdots, \widehat{\omega}_N).$$

Proof: The proof is given in Appendix C. ■

Remark. Lemma 7 states that given the probing policy, the reward attained from AAVF increases with any element of the value belief vector.

C. Optimality of Myopic Channel Probing Policy

We are now ready to study the optimality of the myopic channel probing policy. We proceed by showing the following important auxiliary lemmas (Lemma 8–10 and Corollary 2) and then establish the sufficient condition under which the optimality of the myopic channel probing policy is guaranteed.

Lemma 8. *Given that (1)* F *is regular, (2)* $\frac{O_1}{O_0} < \frac{p_{01}(1-p_{11})}{P_{11}(1-p_{01})}$*, and (3)* $\beta \leq \frac{\Delta_{min}/\Delta_{max}}{(1-\frac{O_1}{O_0})(1-p_{01})+\frac{O_1(p_{11}-p_{01})}{1-(1-O_1)(p_{11}-p_{01})}}$*, if* $p_{01} \leq \widehat{\omega}_i \leq p_{11}$ $(1 \leq i \leq N)$ *and* $p_{11} \geq \widehat{\omega}_l \geq \widehat{\omega}_m \geq p_{01}$*, it holds that*

$$W_t(\widehat{\Omega}_0) \geq W_t(\widehat{\Omega}_1), \ 1 \leq t \leq T,$$

where,

$$\widehat{\Omega}_0 = (\widehat{\omega}_1, \cdots, \widehat{\omega}_l, \cdots, \widehat{\omega}_m, \cdots, \widehat{\omega}_N),$$
$$\widehat{\Omega}_1 = (\widehat{\omega}_1, \cdots, \widehat{\omega}_m, \cdots, \widehat{\omega}_l, \cdots, \widehat{\omega}_N).$$

Lemma 9. *Given that (1)* F *is regular, (2)* $\frac{O_1}{O_0} < \frac{p_{01}(1-p_{11})}{P_{11}(1-p_{01})}$*, and (3)* $\beta \leq \frac{\Delta_{min}/\Delta_{max}}{(1-\frac{O_1}{O_0})(1-p_{01})+\frac{O_1(p_{11}-p_{01})}{1-(1-O_1)(p_{11}-p_{01})}}$*, if* $p_{01} \leq \widehat{\omega}_i \leq p_{11}$ $(1 \leq i \leq N)$ *and* $p_{11} \geq \widehat{\omega}_1 \geq \widehat{\omega}_N \geq p_{01}$*, it holds that*

$$\widehat{W}_t(\dot{\Omega};\widehat{\Omega}_0) - \widehat{W}_t(\dot{\Omega};\widehat{\Omega}_1) \leq \frac{1-p_{01}}{O_0}\Delta_{max}, \ 1 \leq t \leq T$$

where $\widehat{\Omega}_0 = (\widehat{\omega}_1, \cdots, \widehat{\omega}_{N-1}, \widehat{\omega}_N)$, $\widehat{\Omega}_1 = (\widehat{\omega}_N, \widehat{\omega}_1, \cdots \widehat{\omega}_{N-1})$.

Lemma 10. *Given that (1)* F *is regular, (2)* $\frac{O_1}{O_0} < \frac{p_{01}(1-p_{11})}{P_{11}(1-p_{01})}$*, and (3)* $\beta \leq \frac{\Delta_{min}/\Delta_{max}}{(1-\frac{O_1}{O_0})(1-p_{01})+\frac{O_1(p_{11}-p_{01})}{1-(1-O_1)(p_{11}-p_{01})}}$*, if* $p_{01} \leq \widehat{\omega}_i \leq p_{11}$ $(1 \leq i \leq N)$ *and* $p_{11} \geq \widehat{\omega}_1 \geq \widehat{\omega}_N \geq p_{01}$*, it holds that for* $1 \leq t \leq T$

$$\begin{aligned}&\widehat{W}_t(\dot{\Omega};\widehat{\Omega}_0) - \widehat{W}_t(\dot{\Omega};\widehat{\Omega}_1)\\ &\leq (p_{11}-p_{01})\Delta_{max}\frac{1-[\beta(1-O_1)(p_{11}-p_{01})]^{T-t+1}}{1-\beta(1-O_1)(p_{11}-p_{01})},\end{aligned}$$

where

$$\widehat{\Omega}_0 = (\widehat{\omega}_1, \widehat{\omega}_2, \cdots, \widehat{\omega}_{N-1}, \widehat{\omega}_N), \widehat{\Omega}_1 = (\widehat{\omega}_N, \widehat{\omega}_2, \cdots, \widehat{\omega}_{N-1}, \widehat{\omega}_1).$$

For the clarity of presentation, the proofs of the above three lemmas are deferred to the Appendix. Technically speaking, the proof is based on the intrinsic structure of AVF and AAVF, and the development of different "branches" of the channel realizations to derive the relevant bounds, which are further applied to study the optimality of the myopic channel probing policy in Theorem 1.

Remark. We would like to provide a note on the engineering implications of the above three lemmas. Lemma 8 states that the SU cannot increase the expected reward by visiting a channel with a smaller belief value. Lemma 9 derives the upper bound on the difference of the total reward by swapping ω_N and ω_k $(k = N-1, \cdots, 1)$. Lemma 10, on the other hand, derives the upper bound on the difference of the total reward by swapping ω_N and ω_1.

Based on Lemma 8 and 10, we have the following corollary.

Corollary 2. *Given that (1)* F *is regular, (2)* $\frac{O_1}{O_0} < \frac{p_{01}(1-p_{11})}{P_{11}(1-p_{01})}$*, and (3)* $\beta \leq \frac{\Delta_{min}/\Delta_{max}}{(1-\frac{O_1}{O_0})(1-p_{01})+\frac{O_1(p_{11}-p_{01})}{1-(1-O_1)(p_{11}-p_{01})}}$*, if* $p_{01} \leq \widehat{\omega}_i \leq p_{11}$ $(1 \leq i \leq N)$ *and* $\widehat{\omega}_i = \max\{\widehat{\omega}_j : j \in \mathcal{N}\}$ *and* $\widehat{\omega}_N = \min\{\widehat{\omega}_j : j \in \mathcal{N}\}$*, it holds that for* $1 \leq t \leq T$

$$\begin{aligned}&\widehat{W}_t(\dot{\Omega};\widehat{\Omega}_0) - \widehat{W}_t(\dot{\Omega};\widehat{\Omega}_1)\\ &\leq (p_{11}-p_{01})\Delta_{max}\frac{1-[\beta(1-O_1)(p_{11}-p_{01})]^{T-t+1}}{1-\beta(1-O_1)(p_{11}-p_{01})},\end{aligned}$$

where

$$\widehat{\Omega}_0 = (\widehat{\omega}_1, \cdots, \widehat{\omega}_i, \cdots, \widehat{\omega}_{N-1}, \widehat{\omega}_N),$$
$$\widehat{\Omega}_1 = (\widehat{\omega}_1, \cdots, \widehat{\omega}_N, \cdots, \widehat{\omega}_{N-1}, \widehat{\omega}_i).$$

Proof: It can be developed that:

$$\begin{aligned}&\widehat{W}_t(\dot{\Omega};(\widehat{\omega}_1, \cdots, \widehat{\omega}_{i-1}, \widehat{\omega}_i, \widehat{\omega}_{i+1}, \cdots, \widehat{\omega}_{N-1}, \widehat{\omega}_N))\\ &\quad - \widehat{W}_t(\dot{\Omega};(\widehat{\omega}_1, \cdots, \widehat{\omega}_{i-1}, \widehat{\omega}_N, \widehat{\omega}_{i+1}, \cdots, \widehat{\omega}_{N-1}, \widehat{\omega}_i))\\ &\leq \widehat{W}_t(\dot{\Omega};(\widehat{\omega}_i, \widehat{\omega}_1, \cdots, \widehat{\omega}_{i-1}, \widehat{\omega}_{i+1}, \cdots, \widehat{\omega}_{N-1}, \widehat{\omega}_N))\\ &\quad - \widehat{W}_t(\dot{\Omega};(\widehat{\omega}_N, \widehat{\omega}_1, \cdots, \widehat{\omega}_{i-1}, \widehat{\omega}_{i+1}, \cdots, \widehat{\omega}_{N-1}, \widehat{\omega}_i))\\ &\leq (p_{11}-p_{01})\Delta_{max}\frac{1-[\beta(1-O_1)(p_{11}-p_{01})]^{T-t+1}}{1-\beta(1-O_1)(p_{11}-p_{01})},\end{aligned}$$

where, the first inequality follows from Lemma 8 and the second one Lemma 10. ■

The following theorem establishes the main result of our work by stating the condition under which the myopic channel probing policy is guaranteed to achieve the system optimum.

Theorem 1. *If* $p_{01} \leq \omega_i(1) \leq p_{11}, 1 \leq i \leq N$*, the myopic probing policy is optimal if (1)* $F(\Omega)$ *is regular; (2)* $\frac{O_1}{O_0} < \frac{p_{01}(1-p_{11})}{P_{11}(1-p_{01})}$*; (3)* $\beta \leq \frac{\Delta_{min}/\Delta_{max}}{(1-\frac{O_1}{O_0})(1-p_{01})+\frac{O_1(p_{11}-p_{01})}{1-(1-O_1)(p_{11}-p_{01})}}$.

Proof: It suffices to show that for any 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.,

$$W_t(\omega_{i_1^*}, \cdots, \omega_{i_N^*}) > W_t(\omega_1, \cdots, \omega_N). \tag{5}$$

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 8 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 (5). Theorem 1 is thus proven. ■

D. Discussion: a Case Study

To illustrate the application of the obtained result, we study a concrete underlay CR system where the SU can transmit at rate r_1 if the channel probed is observed in the *good* state and r_0 ($r_0 \leq r_1$) for the *bad* state. In this scenario, the utility function can be formulated as $F(\Omega_{\mathcal{A}}) = \sum_{i \in \mathcal{A}} [r_1 \cdot \omega_i + r_0 \cdot (1 - \omega_i)]$. Note that the optimality of the myopic policy for this model is studied in [22] for the case of $k = 1$ and very strict conditions are obtained on the optimality result. We now study the generic case with arbitrary k. To that end, recall Theorem 1, we can see that condition (1) holds, and we have $\Delta_{min} = \Delta_{max} = r_1 - r_0$. We can then verify that when $\frac{O_1}{O_0} < \frac{p_{01}(1-p_{11})}{P_{11}(1-p_{01})}$, it holds that $\frac{\Delta_{min}/\Delta_{max}}{(1-\frac{O_1}{O_0})(1-p_{01}) + \frac{O_1(p_{11}-p_{01})}{1-(1-O_1)(p_{11}-p_{01})}} > 1$. Therefore, when the condition (2) holds, the myopic channel probing policy is optimal for any β. This result in the generic case significantly generalizes the results of [22] by dropping one of the conditions of [22] which may be too stringent in practical scenarios and extends to arbitrary k.

IV. Numerical Analysis

A. Simulation Setting

In this section, we demonstrate some of the theoretical results derived in previous sections and gain further insights on the channel probing problem in underlay CR systems.

Specifically, we conduct a numerical analysis on the scenario studied in Section III-D by focusing on two typical parameter settings:

- The strong positively correlated case where p_{11} significantly outweighs p_{01}, indicating a system with strong positively correlated channels: concretely in our simulation we set $p_{11} = 0.8$, $p_{01} = 0.2$. Other parameters are set to $r_1 = 0.9$, $r_0 = 0.5$, $O_0 = 0.5$, $O_1 = 0.025$.
- The weakly positively correlated case where p_{11} only slightly outweighs p_{01}, indicating a system with weakly positively correlated channels. That is, $p_{11} = 0.6$, $p_{01} = 0.4$. Other parameters are set to $r_1 = 0.9$, $r_0 = 0.5$, $O_0 = 0.1$, $O_1 = 0.005$.

For comparison reference, we implement three policies:

- Optimal policy;
- Sub-optimal policy: given the complexity of finding the optimal policy, we investigate the performance of the sub-optimal policy which is regarded as the best policy chosen from 10000 random policies;
- Myopic policy, studied in previous analysis.

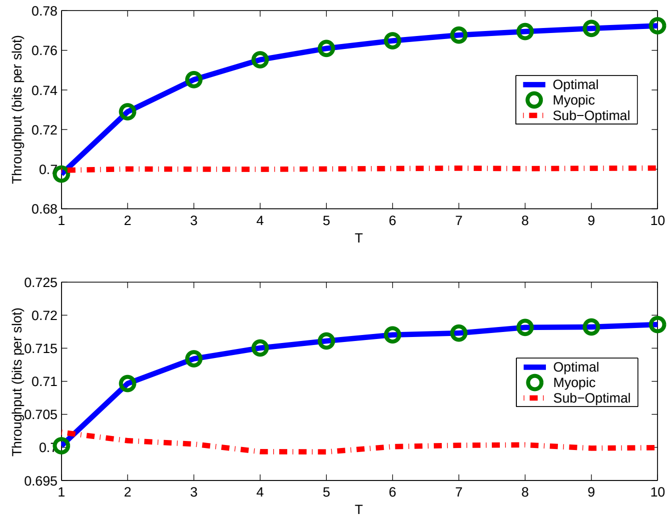

Fig. 2. Performance comparison ($k = 1$): upper plot: strong positively correlated case; lower plot: weakly positively correlated case

We conduct our numerical analysis for two network scenarios: (1) $N = 3$, $k = 1$ and (2) $N = 3$, $k = 2$, i.e., the scenarios where the SU can probe one and two channel(s) each time.

B. Simulation Results

In our numerical studies, we are interested in the performance of the average throughput (i.e., reward) of the three policies. The results obtained in this section provide a complementary quantitative study on the performance of the myopic channel probing policy although it is explicitly addressed in the analytical part. In Figure 2 and Figure 3, we plot the average throughput as a function of the time horizon $T = 10$ for $k = 1$ and $k = 2$, respectively.

It can be illustrated from the simulation results that the average throughput obtained by the myopic policy perfectly matches that of the optimal policy, which confirms our analytical findings in previous sections. We can also observe that the myopic policy outperforms the sub-optimal policy to various extents depending on the system parameter settings. Given the exponential complexity of obtaining the optimal policy and the large number of trials in the sub-optimal policy, the benefit of the myopic policy is well demonstrated.

V. Conclusion

We have investigated the problem of opportunistic spectrum access in underlay CR systems where an SU can exploit the primary feedback signals to obtain the channel information on the primary channels. We have formulated the channel probing problem in which the SU chooses the set of channels to probe so as to maximize its long-term reward. By casting the problem into the RMAB problem, we have established the closed-form condition to ensure the optimality of the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

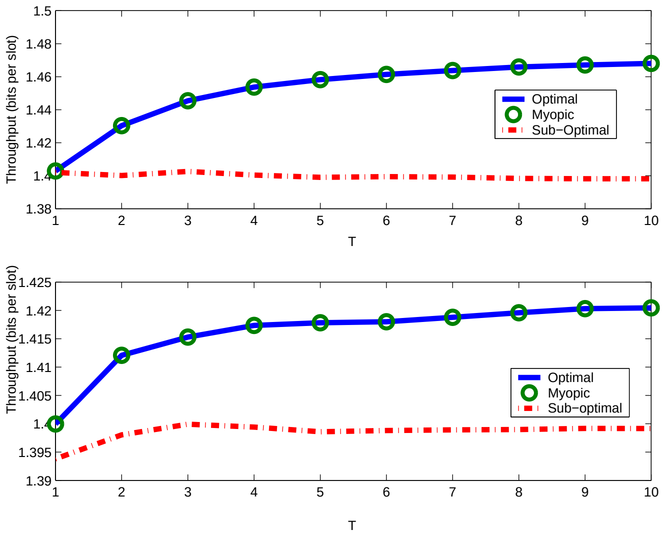

Fig. 3. Performance comparison ($k=2$): upper plot: strong positively correlated case; lower plot: weakly positively correlated case

myopic channel probing policy, a natural policy consisting of probing the best channels based on current belief values of the channel condition. Due to the generic nature of the addressed problem, we believe that the obtained results and the analysis methodology employed in our analysis are widely applicable in a wide range of engineering domains.

Appendix A
Proof of Lemma 5

Recall $W_t(\Omega(t)) = \beta\sum_{\mathcal{E}\subseteq\mathcal{N}(k)}\mathcal{C}^{\mathcal{E}}_{\mathcal{N}(k)}W_{t+1}(\Omega_{\mathcal{E}}(t+1)) + F(\Omega_{\mathcal{N}(k)}(t))$, we prove the lemma by distinguishing the following two cases:

Case 1: $i,j\in\mathcal{A}(t)$. Noticing that (1) both $F(\Omega_{\mathcal{N}(k)}(t))$ and $\sum_{\mathcal{E}\subseteq\mathcal{N}(k)}\mathcal{C}^{\mathcal{E}}_{\mathcal{N}(k)} = \sum_{\mathcal{E}\subseteq\mathcal{A}(t)}\mathcal{C}^{\mathcal{E}}_{\mathcal{A}(t)}$ are symmetrical w.r.t. ω_i and ω_j, (2) $(\omega_1,\cdots,\omega_i,\cdots,\omega_j,\cdots,\omega_N)$ and $(\omega_1,\cdots,\omega_j,\cdots,\omega_i,\cdots,\omega_N)$ generate the same belief vector $\Omega_{\mathcal{E}}(t+1)$ for any $\mathcal{E}$, and (3) myopic policy is adopted from slot $t+1$ to T, it holds that $W_t(\Omega_(t))$ is symmetrical w.r.t. ω_i and ω_j.

Case 2: $i,j\notin\mathcal{A}(t)$. Noticing that (1) $F(\Omega_{\mathcal{N}(k)}(t))$ and $\sum_{\mathcal{E}\subseteq\mathcal{N}(k)}\mathcal{C}^{\mathcal{E}}_{\mathcal{N}(k)} = \sum_{\mathcal{E}\subseteq\mathcal{A}(t)}\mathcal{C}^{\mathcal{E}}_{\mathcal{A}(t)}$ are unrelated with ω_i and ω_j, (2) $(\omega_1,\cdots,\omega_i,\cdots,\omega_j,\cdots,\omega_N)$ and $(\omega_1,\cdots,\omega_j,\cdots,\omega_i,\cdots,\omega_N)$ generate the same belief vector $\Omega_{\mathcal{E}}(t+1)$ for any $\mathcal{E}$, and (3) myopic policy is adopted from slot $t+1$ to T, it holds that $W_t(\Omega(t))$ is symmetrical w.r.t. ω_i and ω_j.

Combing the analysis completes the proof.

Appendix B
Proof of Lemma 6

We proceed the proof by backward induction. First, 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 observed in slot t, i.e. $l\geq k+1$. Let $\mathcal{M}\triangleq\mathcal{N}(k)=\{1,\cdots,k\}$, $\widehat{\omega}_l=0$ and 1, respectively, we have according to Lemma 4

$$
\begin{aligned}
&\widehat{W}_t(\Omega;(\widehat{\omega}_1,\cdots,\widehat{\omega}_l,\cdots,\widehat{\omega}_n))\\
&=F(\widehat{\omega}_1,\cdots,\widehat{\omega}_k)+\beta\sum_{\mathcal{E}\subseteq\mathcal{M}}\widehat{\mathcal{C}}^{\mathcal{E}}_{\mathcal{M}}\widehat{W}_{t+1}(\dot{\Omega};\widehat{\Omega}^{\mathcal{E}}_l(t+1)),\\
&\widehat{W}_t(\Omega;(\widehat{\omega}_1,\cdots,0,\cdots,\widehat{\omega}_n))\\
&=F(\widehat{\omega}_1,\cdots,\widehat{\omega}_k)+\beta\sum_{\mathcal{E}\subseteq\mathcal{M}}\widehat{\mathcal{C}}^{\mathcal{E}}_{\mathcal{M}}\widehat{W}_{t+1}(\dot{\Omega};\widehat{\Omega}^{\mathcal{E}}_{l,0}(t+1)),\\
&\widehat{W}_t(\Omega;(\widehat{\omega}_1,\cdots,1,\cdots,\widehat{\omega}_n))\\
&=F(\widehat{\omega}_1,\cdots,\widehat{\omega}_k)+\beta\sum_{\mathcal{E}\subseteq\mathcal{M}}\widehat{\mathcal{C}}^{\mathcal{E}}_{\mathcal{M}}\widehat{W}_{t+1}(\dot{\Omega};\widehat{\Omega}^{\mathcal{E}}_{l,1}(t+1)),
\end{aligned}
$$

where

$$
\begin{aligned}
\widehat{\Omega}^{\mathcal{E}}_l(t+1)&=(\mathbf{\Phi}(\mathcal{E}),\mathbf{\Upsilon}^{l-1}_{k+1},\mathcal{T}(\widehat{\omega}_l),\mathbf{\Upsilon}^{N}_{l+1},\mathbf{\Psi}(\mathcal{M},\mathcal{E})),\\
\widehat{\Omega}^{\mathcal{E}}_{l,0}(t+1)&=(\mathbf{\Phi}(\mathcal{E}),\mathbf{\Upsilon}^{l-1}_{k+1},p_{01},\mathbf{\Upsilon}^{N}_{l+1},\mathbf{\Psi}(\mathcal{M},\mathcal{E})),\\
\widehat{\Omega}^{\mathcal{E}}_{l,1}(t+1)&=(\mathbf{\Phi}(\mathcal{E}),\mathbf{\Upsilon}^{l-1}_{k+1},p_{11},\mathbf{\Upsilon}^{N}_{l+1},\mathbf{\Psi}(\mathcal{M},\mathcal{E})).
\end{aligned}
$$

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

$$
\begin{aligned}
\widehat{W}_{t+1}(\dot{\Omega};\widehat{\Omega}^{\mathcal{E}}_l(t+1)) =&(1-\widehat{\omega}_l)\widehat{W}_{t+1}(\dot{\Omega};\widehat{\Omega}^{\mathcal{E}}_{l,0}(t+1))\\
&+\widehat{\omega}_l\widehat{W}_{t+1}(\dot{\Omega};\widehat{\Omega}^{\mathcal{E}}_{l,1}(t+1)) \qquad (6)
\end{aligned}
$$

According to induction hypothesis, we have

$$
\begin{aligned}
&\widehat{W}_{t+1}(\dot{\Omega};\widehat{\Omega}^{\mathcal{E}}_l(t+1))\\
&=(1-\mathcal{T}(\widehat{\omega}_l))\widehat{W}_{t+1}(\dot{\Omega};(\mathbf{\Phi}(\mathcal{E}),\mathbf{\Upsilon}^{l-1}_{k+1},0,\mathbf{\Upsilon}^{N}_{l+1},\mathbf{\Psi}(\mathcal{M},\mathcal{E})))\\
&\quad+\mathcal{T}(\widehat{\omega}_l)\widehat{W}_{t+1}(\dot{\Omega};(\mathbf{\Phi}(\mathcal{E}),\mathbf{\Upsilon}^{l-1}_{k+1},1,\mathbf{\Upsilon}^{N}_{l+1},\mathbf{\Psi}(\mathcal{M},\mathcal{E}))) \quad (7)\\
&\widehat{W}_{t+1}(\dot{\Omega};\widehat{\Omega}^{\mathcal{E}}_{l,0}(t+1))\\
&=(1-p_{01})\widehat{W}_{t+1}(\dot{\Omega};(\mathbf{\Phi}(\mathcal{E}),\mathbf{\Upsilon}^{l-1}_{k+1},0,\mathbf{\Upsilon}^{N}_{l+1},\mathbf{\Psi}(\mathcal{M},\mathcal{E})))\\
&\quad+p_{01}\widehat{W}_{t+1}(\dot{\Omega};(\mathbf{\Phi}(\mathcal{E}),\mathbf{\Upsilon}^{l-1}_{k+1},1,\mathbf{\Upsilon}^{N}_{l+1},\mathbf{\Psi}(\mathcal{M},\mathcal{E}))) \quad (8)\\
&\widehat{W}_{t+1}(\dot{\Omega};\widehat{\Omega}^{\mathcal{E}}_{l,1}(t+1))\\
&=(1-p_{11})\widehat{W}_{t+1}(\dot{\Omega};\mathbf{\Phi}(\mathcal{E}),\mathbf{\Upsilon}^{l-1}_{k+1},0,\mathbf{\Upsilon}^{N}_{l+1},\mathbf{\Psi}(\mathcal{M},\mathcal{E}))\\
&\quad+p_{11}\widehat{W}_{t+1}(\dot{\Omega};\mathbf{\Phi}(\mathcal{E}),\mathbf{\Upsilon}^{l-1}_{k+1},1,\mathbf{\Upsilon}^{N}_{l+1},\mathbf{\Psi}(\mathcal{M},\mathcal{E})) \quad (9)
\end{aligned}
$$

Combing (7), (8), (9), we have (6).

Case 2: channel l is observed 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 eq. (4) and Lemma 4

$$
\begin{aligned}
&\widehat{W}_t(\Omega;\widehat{\Omega}(t))=F(\widehat{\omega}_1,\cdots,\widehat{\omega}_l,\cdots,\widehat{\omega}_k)\\
&+\beta[1-O_0(1-(1-\epsilon)\widehat{\omega}_l)]\sum_{\mathcal{E}\subseteq\mathcal{M}}\widehat{\mathcal{C}}^{\mathcal{E}}_{\mathcal{M}}\\
&\widehat{W}_{t+1}(\dot{\Omega};(\mathbf{\Phi}^l(\mathcal{E}),\mathcal{T}(\phi(\widehat{\omega}_l)),\mathbf{\Phi}_l(\mathcal{E}),\mathbf{\Upsilon}^{N}_{k+1},\mathbf{\Psi}^l(\mathcal{M},\mathcal{E}),\mathbf{\Psi}_l(\mathcal{M},\mathcal{E})))\\
&+\beta[O_0(1-(1-\epsilon)\widehat{\omega}_l)]\sum_{\mathcal{E}\subseteq\mathcal{M}}\widehat{\mathcal{C}}^{\mathcal{E}}_{\mathcal{M}}\\
&\widehat{W}_{t+1}(\dot{\Omega};(\mathbf{\Phi}^l(\mathcal{E}),\mathbf{\Phi}_l(\mathcal{E}),\mathbf{\Upsilon}^{N}_{k+1},\mathbf{\Psi}^l(\mathcal{M},\mathcal{E}),\mathcal{T}(\varphi(\widehat{\omega}_l)),\mathbf{\Psi}_l(\mathcal{M},\mathcal{E})))
\end{aligned}
$$

Let $\widehat{\omega}_l=0$ and 1, respectively, we have

$$
\begin{aligned}
&\widehat{W}_t(\Omega;\widehat{\Omega}_{l=0}(t))=F(\widehat{\omega}_1,\cdots,0,\cdots,\widehat{\omega}_k)\\
&+\beta[1-O_0]\sum_{\mathcal{E}\subseteq\mathcal{M}}\widehat{\mathcal{C}}^{\mathcal{E}}_{\mathcal{M}}\\
&\widehat{W}_{t+1}(\dot{\Omega};(\mathbf{\Phi}^l(\mathcal{E}),p_{01},\mathbf{\Phi}_l(\mathcal{E}),\mathbf{\Upsilon}^{N}_{k+1},\mathbf{\Psi}^l(\mathcal{M},\mathcal{E}),\mathbf{\Psi}_l(\mathcal{M},\mathcal{E})))\\
&+\beta O_0\sum_{\mathcal{E}\subseteq\mathcal{M}}\widehat{\mathcal{C}}^{\mathcal{E}}_{\mathcal{M}}\\
&\widehat{W}_{t+1}(\dot{\Omega};(\mathbf{\Phi}^l(\mathcal{E}),\mathbf{\Phi}_l(\mathcal{E}),\mathbf{\Upsilon}^{N}_{k+1},\mathbf{\Psi}^l(\mathcal{M},\mathcal{E}),p_{01},\mathbf{\Psi}_l(\mathcal{M},\mathcal{E}))),\\
&\widehat{W}_t(\Omega;\widehat{\Omega}_{l=1}(t))=F(\widehat{\omega}_1,\cdots,1,\cdots,\widehat{\omega}_k)\\
&+\beta[1-O_1]\sum_{\mathcal{E}\subseteq\mathcal{M}}\widehat{\mathcal{C}}^{\mathcal{E}}_{\mathcal{M}}
\end{aligned}
$$

$$\widehat{W}_{t+1}(\dot{\Omega};(\mathbf{\Phi}^l(\mathcal{E}),p_{11},\mathbf{\Phi}_l(\mathcal{E}),\mathbf{\Upsilon}_{k+1}^N,\mathbf{\Psi}^l(\mathcal{M},\mathcal{E}),\mathbf{\Psi}_l(\mathcal{M},\mathcal{E})))$$
$$+\beta O_1\sum_{\mathcal{E}\subseteq\mathcal{M}}\widehat{\mathcal{C}}_{\mathcal{M}}^{\mathcal{E}}$$
$$\widehat{W}_{t+1}(\dot{\Omega};(\mathbf{\Phi}^l(\mathcal{E}),\mathbf{\Phi}_l(\mathcal{E}),\mathbf{\Upsilon}_{k+1}^N,\mathbf{\Psi}^l(\mathcal{M},\mathcal{E}),p_{11},\mathbf{\Psi}_l(\mathcal{M},\mathcal{E})))$$

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

$$[1-O_0(1-(1-\epsilon)\widehat{\omega}_l)]$$
$$\widehat{W}_{t+1}(\dot{\Omega};(\mathbf{\Phi}^l(\mathcal{E}),\mathcal{T}(\phi(\widehat{\omega}_l)),\mathbf{\Phi}_l(\mathcal{E}),\mathbf{\Upsilon}_{k+1}^N,\mathbf{\Psi}^l(\mathcal{M},\mathcal{E}),\mathbf{\Psi}_l(\mathcal{M},\mathcal{E})))$$
$$+[O_0(1-(1-\epsilon)\widehat{\omega}_l)]$$
$$\widehat{W}_{t+1}(\dot{\Omega};(\mathbf{\Phi}^l(\mathcal{E}),\mathbf{\Phi}_l(\mathcal{E}),\mathbf{\Upsilon}_{k+1}^N,\mathbf{\Psi}^l(\mathcal{M},\mathcal{E}),\mathcal{T}(\varphi(\widehat{\omega}_l)),\mathbf{\Psi}_l(\mathcal{M},\mathcal{E})))$$
$$=(1-\widehat{\omega}_l)(1-O_0)$$
$$\widehat{W}_{t+1}(\dot{\Omega};(\mathbf{\Phi}^l(\mathcal{E}),p_{01},\mathbf{\Phi}_l(\mathcal{E}),\mathbf{\Upsilon}_{k+1}^N,\mathbf{\Psi}^l(\mathcal{M},\mathcal{E}),\mathbf{\Psi}_l(\mathcal{M},\mathcal{E})))$$
$$+(1-\widehat{\omega}_l)O_0$$
$$\widehat{W}_{t+1}(\dot{\Omega};(\mathbf{\Phi}^l(\mathcal{E}),\mathbf{\Phi}_l(\mathcal{E}),\mathbf{\Upsilon}_{k+1}^N),\mathbf{\Psi}^l(\mathcal{M},\mathcal{E}),p_{01},\mathbf{\Psi}_l(\mathcal{M},\mathcal{E})))$$
$$+\widehat{\omega}_l(1-O_1)$$
$$\widehat{W}_{t+1}(\dot{\Omega};(\mathbf{\Phi}^l(\mathcal{E}),p_{11},\mathbf{\Phi}_l(\mathcal{E}),\mathbf{\Upsilon}_{k+1}^N,\mathbf{\Psi}^l(\mathcal{M},\mathcal{E}),\mathbf{\Psi}_l(\mathcal{M},\mathcal{E})))$$
$$+\widehat{\omega}_lO_1$$
$$\widehat{W}_{t+1}(\dot{\Omega};(\mathbf{\Phi}^l(\mathcal{E}),\mathbf{\Phi}_l(\mathcal{E}),\mathbf{\Upsilon}_{k+1}^N,\mathbf{\Psi}^l(\mathcal{M},\mathcal{E}),p_{11},\mathbf{\Psi}_l(\mathcal{M},\mathcal{E}))) \tag{10}$$

According to induction hypothesis, we have

$$\widehat{W}_{t+1}(\dot{\Omega};\mathbf{\Phi}^l(\mathcal{E}),x,\mathbf{\Phi}_l(\mathcal{E}),\mathbf{\Upsilon}_{k+1}^N,\mathbf{\Psi}^l(\mathcal{M},\mathcal{E}),\mathbf{\Psi}_l(\mathcal{M},\mathcal{E}))$$
$$=(1-x)$$
$$\widehat{W}_{t+1}(\dot{\Omega};(\mathbf{\Phi}^l(\mathcal{E}),0,\mathbf{\Phi}_l(\mathcal{E}),\mathbf{\Upsilon}_{k+1}^N,\mathbf{\Psi}^l(\mathcal{M},\mathcal{E}),\mathbf{\Psi}_l(\mathcal{M},\mathcal{E})))$$
$$+x\widehat{W}_{t+1}(\dot{\Omega};(\mathbf{\Phi}^l(\mathcal{E}),1,\mathbf{\Phi}_l(\mathcal{E}),\mathbf{\Upsilon}_{k+1}^N,\mathbf{\Psi}^l(\mathcal{M},\mathcal{E}),\mathbf{\Psi}_l(\mathcal{M},\mathcal{E}))) \tag{11}$$

$$\widehat{W}_{t+1}(\dot{\Omega};\mathbf{\Phi}^l(\mathcal{E}),\mathbf{\Phi}_l(\mathcal{E}),\mathbf{\Upsilon}_{k+1}^N,\mathbf{\Psi}^l(\mathcal{M},\mathcal{E}),y,\mathbf{\Psi}_l(\mathcal{M},\mathcal{E}))$$
$$=(1-y)$$
$$\widehat{W}_{t+1}(\dot{\Omega};(\mathbf{\Phi}^l(\mathcal{E}),\mathbf{\Phi}_l(\mathcal{E}),\mathbf{\Upsilon}_{k+1}^N,\mathbf{\Psi}^l(\mathcal{M},\mathcal{E}),0,\mathbf{\Psi}_l(\mathcal{M},\mathcal{E})))$$
$$+y\widehat{W}_{t+1}(\dot{\Omega};(\mathbf{\Phi}^l(\mathcal{E}),\mathbf{\Phi}_l(\mathcal{E}),\mathbf{\Upsilon}_{k+1}^N,\mathbf{\Psi}^l(\mathcal{M},\mathcal{E}),1,\mathbf{\Psi}_l(\mathcal{M},\mathcal{E}))) \tag{12}$$

Let $x,y=\mathcal{T}(\phi(\widehat{\omega}_l)),\mathcal{T}(\varphi(\widehat{\omega}_l)),p_{11}$ and p_{01}, respectively, we can easily obtain (10) by some simple arithmetic operations.

Combing the above analysis, we thus prove Lemma 6.

Appendix C
Proof of Lemma 7

We proceed the proof by backward induction. First, 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 observed in slot t, i.e., $l\geq k+1$. In this case, the immediate reward is unrelated to $\widehat{\omega}_l$ and $\widehat{\omega}_l'$. Moreover, let $\widehat{\Omega}(t+1)$ and $\widehat{\Omega}'(t+1)$ denote the belief vector generated by $\widehat{\Omega}(t)=(\widehat{\omega}_1,\cdots,\widehat{\omega}_l,\cdots,\widehat{\omega}_N)$ and $\widehat{\Omega}'(t)=(\widehat{\omega}_1,\cdots,\widehat{\omega}_l',\cdots,\widehat{\omega}_N)$, respectively, it can be noticed that $\widehat{\Omega}(t+1)$ and $\widehat{\Omega}'(t+1)$ differ in only one element: $\widehat{\omega}_l'(t+1)\geq\widehat{\omega}_l(t+1)$. By induction hypothesis, it holds that $\widehat{W}_{t+1}(\dot{\Omega};\widehat{\Omega}'(t+1))\geq\widehat{W}_{t+1}(\dot{\Omega};\widehat{\Omega}(t+1))$. Noticing (4), it follows that $\widehat{W}_t(\dot{\Omega};\widehat{\Omega}'(t))\geq\widehat{W}_t(\dot{\Omega};\widehat{\Omega}(t))$.

Case 2: channel l is observed in slot t, i.e., $l\leq k$. Following Lemma 6 and after some straightforward algebraic operations, we have

$$\widehat{W}_t(\dot{\Omega};(\widehat{\omega}_1,\cdots,\widehat{\omega}_l',\cdots,\widehat{\omega}_N))-\widehat{W}_t(\dot{\Omega};(\widehat{\omega}_1,\cdots,\widehat{\omega}_l,\cdots,\widehat{\omega}_N))$$
$$=(\widehat{\omega}_l'-\widehat{\omega}_l)$$
$$[\widehat{W}_t(\dot{\Omega};(\widehat{\omega}_1,\cdots,1,\cdots,\widehat{\omega}_N))-\widehat{W}_t(\dot{\Omega};(\widehat{\omega}_1,\cdots,0,\cdots,\widehat{\omega}_N))].$$

Let $\mathcal{M}\triangleq\mathcal{N}(k)\setminus\{l\}=\{1,\cdots,l-1,l+1,\cdots,k\}$, by developing $\widehat{W}_t(\dot{\Omega};\Omega(t))$ as a function of $\widehat{\omega}_l$, we have

$$\widehat{W}_t(\dot{\Omega};\widehat{\Omega}(t))=F(\widehat{\omega}_1(t),\cdots,\widehat{\omega}_k(t))$$
$$+\beta[1-O_0(1-(1-\frac{O_1}{O_0})\widehat{\omega}_l)]\sum_{\mathcal{E}\subseteq\mathcal{M}}\widehat{\mathcal{C}}_{\mathcal{M}}^{\mathcal{E}}\widehat{W}_{t+1}(\dot{\Omega};\widehat{\Omega}_{\mathcal{E}}(t+1))$$
$$+\beta[O_0(1-(1-\frac{O_1}{O_0})\widehat{\omega}_l)]\sum_{\mathcal{E}\subseteq\mathcal{M}}\widehat{\mathcal{C}}_{\mathcal{M}}^{\mathcal{E}}\widehat{W}_{t+1}(\dot{\Omega};\widehat{\Omega}_{\mathcal{E}}(t+1)).$$

Let $\widehat{\omega}_l=0$ and 1, respectively, we have

$$W_t(\dot{\Omega};\widehat{\Omega}_{l=0}(t))=F(\widehat{\omega}_1(t),\cdots,0,\cdots,\widehat{\omega}_k(t))$$
$$+\beta[1-O_0]\sum_{\mathcal{E}\subseteq\mathcal{M}}\widehat{\mathcal{C}}_{\mathcal{M}}^{\mathcal{E}}W_{t+1}(\dot{\Omega};\widehat{\Omega}_{0,1}^{\mathcal{E}}(t+1))$$
$$+\beta O_0\sum_{\mathcal{E}\subseteq\mathcal{M}}\widehat{\mathcal{C}}_{\mathcal{M}}^{\mathcal{E}}\widehat{W}_{t+1}(\dot{\Omega};\widehat{\Omega}_{0,0}^{\mathcal{E}}(t+1)),$$

$$W_t(\dot{\Omega};\widehat{\Omega}_{l=1}(t))=F(\widehat{\omega}_1(t),\cdots,1,\cdots,\widehat{\omega}_k(t))$$
$$+\beta[1-O_1]\sum_{\mathcal{E}\subseteq\mathcal{M}}\widehat{\mathcal{C}}_{\mathcal{M}}^{\mathcal{E}}W_{t+1}(\dot{\Omega};\widehat{\Omega}_{1,1}^{\mathcal{E}}(t+1))$$
$$+\beta O_1\sum_{\mathcal{E}\subseteq\mathcal{M}}\widehat{\mathcal{C}}_{\mathcal{M}}^{\mathcal{E}}\widehat{W}_{t+1}(\dot{\Omega};\widehat{\Omega}_{1,0}^{\mathcal{E}}(t+1)),$$

where

$$\widehat{\Omega}_{0,1}^{\mathcal{E}}(t+1)=(\mathbf{\Phi}^l(\mathcal{E}),p_{01},\mathbf{\Phi}_l(\mathcal{E}),\mathbf{\Upsilon}_{k+1}^N,\mathbf{\Psi}^l(\mathcal{M},\mathcal{E}),\mathbf{\Psi}_l(\mathcal{M},\mathcal{E})),$$
$$\widehat{\Omega}_{0,0}^{\mathcal{E}}(t+1)=(\mathbf{\Phi}^l(\mathcal{E}),\mathbf{\Phi}_l(\mathcal{E}),\mathbf{\Upsilon}_{k+1}^N,\mathbf{\Psi}^l(\mathcal{M},\mathcal{E}),p_{01},\mathbf{\Psi}_l(\mathcal{M},\mathcal{E})),$$
$$\widehat{\Omega}_{1,1}^{\mathcal{E}}(t+1)=(\mathbf{\Phi}^l(\mathcal{E}),p_{11},\mathbf{\Phi}_l(\mathcal{E}),\mathbf{\Upsilon}_{k+1}^N,\mathbf{\Psi}^l(\mathcal{M},\mathcal{E}),\mathbf{\Psi}_l(\mathcal{M},\mathcal{E})),$$
$$\widehat{\Omega}_{1,0}^{\mathcal{E}}(t+1)=(\mathbf{\Phi}^l(\mathcal{E}),\mathbf{\Phi}_l(\mathcal{E}),\mathbf{\Upsilon}_{k+1}^N,\mathbf{\Psi}^l(\mathcal{M},\mathcal{E}),p_{11},\mathbf{\Psi}_l(\mathcal{M},\mathcal{E})).$$

It can be checked that $\widehat{\Omega}_{1,1}^{\mathcal{E}}(t+1)\geq\widehat{\Omega}_{0,1}^{\mathcal{E}}(t+1)$ and $\widehat{\Omega}_{1,0}^{\mathcal{E}}(t+1)\geq\widehat{\Omega}_{0,0}^{\mathcal{E}}(t+1)$. It then follows from induction that given $\mathcal{E}$, $\widehat{W}_{t+1}(\dot{\Omega};\widehat{\Omega}_{1,1}^{\mathcal{E}}(t+1))\geq\widehat{W}_{t+1}(\dot{\Omega};\widehat{\Omega}_{0,1}^{\mathcal{E}}(t+1))$ and $\widehat{W}_{t+1}(\dot{\Omega};\widehat{\Omega}_{1,0}^{\mathcal{E}}(t+1))\geq\widehat{W}_{t+1}(\dot{\Omega};\widehat{\Omega}_{0,0}^{\mathcal{E}}(t+1))$. Noticing that F is increasing, we then have

$$\widehat{W}_t(\dot{\Omega};(\widehat{\omega}_1,\cdots,1,\cdots,\widehat{\omega}_n))-\widehat{W}_t(\dot{\Omega};(\widehat{\omega}_1,\cdots,0,\cdots,\widehat{\omega}_n))$$
$$=F(\widehat{\omega}_1,\cdots,1,\cdots,\widehat{\omega}_n)-F(\widehat{\omega}_1,\cdots,0,\cdots,\widehat{\omega}_n)$$
$$+\beta[1-O_1]\sum_{\mathcal{E}\subseteq\mathcal{M}}\widehat{\mathcal{C}}_{\mathcal{M}}^{\mathcal{E}}\widehat{W}_{t+1}(\dot{\Omega};\widehat{\Omega}_{1,1}^{\mathcal{E}}(t+1))$$
$$+\beta O_1\sum_{\mathcal{E}\subseteq\mathcal{M}}\widehat{\mathcal{C}}_{\mathcal{M}}^{\mathcal{E}}\widehat{W}_{t+1}(\dot{\Omega};\widehat{\Omega}_{1,0}^{\mathcal{E}}(t+1))$$
$$-\beta[1-O_0]\sum_{\mathcal{E}\subseteq\mathcal{M}}\widehat{\mathcal{C}}_{\mathcal{M}}^{\mathcal{E}}\widehat{W}_{t+1}(\dot{\Omega};\widehat{\Omega}_{0,1}^{\mathcal{E}}(t+1))$$
$$-\beta O_0\sum_{\mathcal{E}\subseteq\mathcal{M}}\widehat{\mathcal{C}}_{\mathcal{M}}^{\mathcal{E}}\widehat{W}_{t+1}(\dot{\Omega};\widehat{\Omega}_{0,0}^{\mathcal{E}}(t+1))$$
$$\geq\beta[1-O_1]\sum_{\mathcal{E}\subseteq\mathcal{M}}\widehat{\mathcal{C}}_{\mathcal{M}}^{\mathcal{E}}\widehat{W}_{t+1}(\dot{\Omega};\widehat{\Omega}_{1,1}^{\mathcal{E}}(t+1))$$
$$+\beta O_1\sum_{\mathcal{E}\subseteq\mathcal{M}}\widehat{\mathcal{C}}_{\mathcal{M}}^{\mathcal{E}}\widehat{W}_{t+1}(\dot{\Omega};\widehat{\Omega}_{1,0}^{\mathcal{E}}(t+1))$$

$$
\begin{aligned}
&-\beta[1-O_0]\sum_{\mathcal{E}\subseteq\mathcal{M}}\widehat{\mathcal{C}}_{\mathcal{M}}^{\mathcal{E}}\widehat{W}_{t+1}(\dot{\Omega};\widehat{\Omega}_{1,1}^{\mathcal{E}}(t+1))\\
&-\beta O_0\sum_{\mathcal{E}\subseteq\mathcal{M}}\widehat{\mathcal{C}}_{\mathcal{M}}^{\mathcal{E}}\widehat{W}_{t+1}(\dot{\Omega};\widehat{\Omega}_{1,0}^{\mathcal{E}}(t+1))\\
&=\beta[O_0-O_1]\sum_{\mathcal{E}\subseteq\mathcal{M}}\widehat{\mathcal{C}}_{\mathcal{M}}^{\mathcal{E}}\\
&\Big[\widehat{W}_{t+1}(\dot{\Omega};\widehat{\Omega}_{1,1}^{\mathcal{E}}(t+1))-\widehat{W}_{t+1}(\dot{\Omega};\widehat{\Omega}_{1,0}^{\mathcal{E}}(t+1))\Big]\geq 0.
\end{aligned}
$$

Combining the above analysis in two cases completes our proof.

Appendix D
Proof of Lemma 8–Lemma 10

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

We first show that Lemma 8–10 hold for slot T. It is easy to verify that Lemma 8 holds.

We then prove Lemma 9–10. Noticing that $p_{01}\leq\widehat{\omega}_N\leq\widehat{\omega}_k\leq p_{11}\leq 1$ and $p_{01}\leq\widehat{\omega}_N\leq\widehat{\omega}_1\leq p_{11}$, we have

$$
\begin{aligned}
&\widehat{W}_T(\dot{\Omega};(\widehat{\omega}_1,\cdots,\widehat{\omega}_N))-\widehat{W}_T(\dot{\Omega};(\widehat{\omega}_N,\widehat{\omega}_1,\cdots,\widehat{\omega}_{N-1}))\\
&=F(\widehat{\omega}_1,\cdots,\widehat{\omega}_k)-F(\widehat{\omega}_N,\widehat{\omega}_1,\cdots,\widehat{\omega}_{k-1})\\
&=(\widehat{\omega}_k-\widehat{\omega}_N)[F(\widehat{\omega}_1,\cdots,\widehat{\omega}_{k-1},1)-F(\widehat{\omega}_1,\cdots,\widehat{\omega}_{k-1},0)]\\
&\leq(1-\widehat{\omega}_N)\Delta_{max},\\
&\widehat{W}_T(\dot{\Omega};(\widehat{\omega}_1,\cdots,\widehat{\omega}_N))-\widehat{W}_T(\dot{\Omega};(\widehat{\omega}_N,\widehat{\omega}_2,\cdots,\widehat{\omega}_{N-1},\widehat{\omega}_1))\\
&=F(\widehat{\omega}_1,\cdots,\widehat{\omega}_k)-F(\widehat{\omega}_N,\widehat{\omega}_2,\cdots,\widehat{\omega}_{k-1})\\
&=(\widehat{\omega}_1-\widehat{\omega}_N)[F(1,\widehat{\omega}_2,\cdots,\widehat{\omega}_k)-F(0,\widehat{\omega}_2,\cdots,\widehat{\omega}_k)]\\
&\leq(p_{11}-p_{01})\Delta_{max}.
\end{aligned}
$$

Lemma 9–10 thus hold for slot T.

Assume that Lemma 8–10 hold for slots $T,\cdots,t+1$, we now prove that they hold for slot t.

We first prove Lemma 8. We distinguish the following three cases considering $l<m$:

Case 1: $l\geq k+1$. This case follows Lemma 5.

Case 2: $l\leq k$ and $m\geq k+1$. In this case, denote $\mathcal{M}\triangleq\mathcal{N}(k)\setminus\{l\}$, we have

$$
\begin{aligned}
&W_t(\widehat{\omega}_1,\cdots,\widehat{\omega}_l,\cdots,\widehat{\omega}_m,\cdots,\widehat{\omega}_N)\\
&\quad-W_t(\widehat{\omega}_1,\cdots,\widehat{\omega}_m,\cdots,\widehat{\omega}_l,\cdots,\widehat{\omega}_N)\\
=&\widehat{W}_t(\dot{\Omega};(\widehat{\omega}_1,\cdots,\widehat{\omega}_l,\cdots,\widehat{\omega}_m,\cdots,\widehat{\omega}_N))\\
&\quad-\widehat{W}_t(\dot{\Omega};(\widehat{\omega}_1,\cdots,\widehat{\omega}_m,\cdots,\widehat{\omega}_l,\cdots,\widehat{\omega}_N))\\
=&(\widehat{\omega}_l-\widehat{\omega}_m)\Big[\widehat{W}_t(\dot{\Omega};(\widehat{\omega}_1,\cdots,1,\cdots,0,\cdots,\widehat{\omega}_N))\\
&\qquad\qquad-\widehat{W}_t(\dot{\Omega};(\widehat{\omega}_1,\cdots,0,\cdots,1,\cdots,\widehat{\omega}_N))\Big]\\
=&(\widehat{\omega}_l-\widehat{\omega}_m)\Big\{F(\widehat{\omega}_1,\cdots,1,\cdots,\widehat{\omega}_k)-F(\widehat{\omega}_1,\cdots,0,\cdots,\widehat{\omega}_k)\\
&\quad+\beta\sum_{\mathcal{E}\subseteq\mathcal{M}}\widehat{\mathcal{C}}_{\mathcal{M}}^{\mathcal{E}}\cdot\\
&\quad\Big[(1-O_1)\widehat{W}_{t+1}(\dot{\Omega};(\mathbf{\Phi}^l(\mathcal{E}),p_{11},\mathbf{\Phi}_l(\mathcal{E}),\mathcal{T}(\widehat{\omega}_{k+1}),\cdots,\\
&\qquad p_{01},\cdots,\mathcal{T}(\widehat{\omega}_N),\mathbf{\Psi}^l(\mathcal{M},\mathcal{E}),\mathbf{\Psi}_l(\mathcal{M},\mathcal{E}))\\
&\quad+O_1\widehat{W}_{t+1}(\dot{\Omega};(\mathbf{\Phi}^l(\mathcal{E}),\mathbf{\Phi}_l(\mathcal{E}),\mathcal{T}(\widehat{\omega}_{k+1}),\cdots,\\
&\qquad p_{01},\cdots,\mathcal{T}(\widehat{\omega}_N),\mathbf{\Psi}^l(\mathcal{M},\mathcal{E}),p_{11},\mathbf{\Psi}_l(\mathcal{M},\mathcal{E}))\\
&\quad-(1-O_0)\widehat{W}_{t+1}(\dot{\Omega};(\mathbf{\Phi}^l(\mathcal{E}),p_{01},\mathbf{\Phi}_l(\mathcal{E}),\mathcal{T}(\widehat{\omega}_{k+1}),\cdots,\\
&\qquad p_{11},\cdots,\mathcal{T}(\widehat{\omega}_N),\mathbf{\Psi}^l(\mathcal{M},\mathcal{E}),\mathbf{\Psi}_l(\mathcal{M},\mathcal{E}))\\
&\quad-O_0\widehat{W}_{t+1}(\dot{\Omega};(\mathbf{\Phi}^l(\mathcal{E}),\mathbf{\Phi}_l(\mathcal{E}),\mathcal{T}(\widehat{\omega}_{k+1}),\cdots,\\
&\qquad p_{11},\cdots,\mathcal{T}(\widehat{\omega}_N),\mathbf{\Psi}^l(\mathcal{M},\mathcal{E}),p_{01},\mathbf{\Psi}_l(\mathcal{M},\mathcal{E}))\Big]\Big\}\\
\geq&(\widehat{\omega}_l-\widehat{\omega}_m)\Big\{\Delta_{min}+\beta\sum_{\mathcal{E}\subseteq\mathcal{M}}\widehat{\mathcal{C}}_{\mathcal{M}}^{\mathcal{E}}\cdot\\
&\quad\Big[(1-O_1)\widehat{W}_{t+1}(\dot{\Omega};(\mathbf{\Phi}^l(\mathcal{E}),p_{11},\mathbf{\Phi}_l(\mathcal{E}),\mathcal{T}(\widehat{\omega}_{k+1}),\cdots,\\
&\qquad p_{01},\cdots,\mathcal{T}(\widehat{\omega}_N),\mathbf{\Psi}^l(\mathcal{M},\mathcal{E}),\mathbf{\Psi}_l(\mathcal{M},\mathcal{E}))\\
&\quad+O_1\widehat{W}_{t+1}(\dot{\Omega};(\mathbf{\Phi}^l(\mathcal{E}),\mathbf{\Phi}_l(\mathcal{E}),\mathcal{T}(\widehat{\omega}_{k+1}),\cdots,\\
&\qquad p_{01},\cdots,\mathcal{T}(\widehat{\omega}_N),\mathbf{\Psi}^l(\mathcal{M},\mathcal{E}),p_{11},\mathbf{\Psi}_l(\mathcal{M},\mathcal{E}))\\
&\quad-(1-O_0)\widehat{W}_{t+1}(\dot{\Omega};(\mathbf{\Phi}^l(\mathcal{E}),p_{11},\mathbf{\Phi}_l(\mathcal{E}),\mathcal{T}(\widehat{\omega}_{k+1}),\cdots,\\
&\qquad p_{01},\cdots,\mathcal{T}(\widehat{\omega}_N),\mathbf{\Psi}^l(\mathcal{M},\mathcal{E}),\mathbf{\Psi}_l(\mathcal{M},\mathcal{E}))\\
&\quad-O_0\widehat{W}_{t+1}(\dot{\Omega};(\mathbf{\Phi}^l(\mathcal{E}),\mathbf{\Phi}_l(\mathcal{E}),\mathcal{T}(\widehat{\omega}_{k+1}),\cdots,\\
&\qquad p_{11},\cdots,\mathcal{T}(\widehat{\omega}_N),\mathbf{\Psi}^l(\mathcal{M},\mathcal{E}),p_{01},\mathbf{\Psi}_l(\mathcal{M},\mathcal{E}))\Big]\Big\}\\
=&(\widehat{\omega}_l-\widehat{\omega}_m)\Big\{\Delta_{min}+\beta\sum_{\mathcal{E}\subseteq\mathcal{M}}\widehat{\mathcal{C}}_{\mathcal{M}}^{\mathcal{E}}\cdot\\
&\quad\Big[(O_0-O_1)\widehat{W}_{t+1}(\dot{\Omega};(\mathbf{\Phi}^l(\mathcal{E}),p_{11},\mathbf{\Phi}_l(\mathcal{E}),\mathcal{T}(\widehat{\omega}_{k+1}),\cdots,\\
&\qquad p_{01},\cdots,\mathcal{T}(\widehat{\omega}_N),\mathbf{\Psi}^l(\mathcal{M},\mathcal{E}),\mathbf{\Psi}_l(\mathcal{M},\mathcal{E}))\\
&\quad+O_1\widehat{W}_{t+1}(\dot{\Omega};(\mathbf{\Phi}^l(\mathcal{E}),\mathbf{\Phi}_l(\mathcal{E}),\mathcal{T}(\widehat{\omega}_{k+1}),\cdots,\\
&\qquad p_{01},\cdots,\mathcal{T}(\widehat{\omega}_N),\mathbf{\Psi}^l(\mathcal{M},\mathcal{E}),p_{11},\mathbf{\Psi}_l(\mathcal{M},\mathcal{E}))\\
&\quad-O_0\widehat{W}_{t+1}(\dot{\Omega};(\mathbf{\Phi}^l(\mathcal{E}),\mathbf{\Phi}_l(\mathcal{E}),\mathcal{T}(\widehat{\omega}_{k+1}),\cdots,\\
&\qquad p_{11},\cdots,\mathcal{T}(\widehat{\omega}_N),\mathbf{\Psi}^l(\mathcal{M},\mathcal{E}),p_{01},\mathbf{\Psi}_l(\mathcal{M},\mathcal{E}))\Big]\Big\}\\
\geq&(\widehat{\omega}_l-\widehat{\omega}_m)\Big\{\Delta_{min}+\beta\sum_{\mathcal{E}\subseteq\mathcal{M}}\widehat{\mathcal{C}}_{\mathcal{M}}^{\mathcal{E}}\cdot\\
&\quad\Big[(O_0-O_1)\widehat{W}_{t+1}(\dot{\Omega};(p_{01},\mathbf{\Phi}^l(\mathcal{E}),p_{11},\mathbf{\Phi}_l(\mathcal{E}),\mathcal{T}(\widehat{\omega}_{k+1}),\cdots,\\
&\qquad \mathcal{T}(\widehat{\omega}_N),\mathbf{\Psi}^l(\mathcal{M},\mathcal{E}),\mathbf{\Psi}_l(\mathcal{M},\mathcal{E}))\\
&\quad+O_1\widehat{W}_{t+1}(\dot{\Omega};(\mathbf{\Phi}^l(\mathcal{E}),p_{01},\mathbf{\Phi}_l(\mathcal{E}),\mathcal{T}(\widehat{\omega}_{k+1}),\cdots,\\
&\qquad \mathcal{T}(\widehat{\omega}_N),\mathbf{\Psi}^l(\mathcal{M},\mathcal{E}),\mathbf{\Psi}_l(\mathcal{M},\mathcal{E}),p_{11}))\\
&\quad-O_0\widehat{W}_{t+1}(\dot{\Omega};(\mathbf{\Phi}^l(\mathcal{E}),p_{11},\mathbf{\Phi}_l(\mathcal{E}),\mathcal{T}(\widehat{\omega}_{k+1}),\cdots,\\
&\qquad \mathcal{T}(\widehat{\omega}_N),\mathbf{\Psi}^l(\mathcal{M},\mathcal{E}),\mathbf{\Psi}_l(\mathcal{M},\mathcal{E}),p_{01})\Big]\Big\}\\
\geq&(\widehat{\omega}_l-\widehat{\omega}_m)\Big[\Delta_{min}-\beta\sum_{\mathcal{E}\subseteq\mathcal{M}}\widehat{\mathcal{C}}_{\mathcal{M}}^{\mathcal{E}}\cdot\Big((O_0-O_1)\frac{1-p_{01}}{O_0}\Delta_{max}\\
&\quad+O_1(p_{11}-p_{01})\Delta_{max}\frac{1-[\beta(1-O_1)(p_{11}-p_{01})]^{T-t}}{1-\beta(1-O_1)(p_{11}-p_{01})}\Big)\Big]\\
\geq&(\widehat{\omega}_l-\widehat{\omega}_m)\sum_{\mathcal{E}\subseteq\mathcal{M}}\widehat{\mathcal{C}}_{\mathcal{M}}^{\mathcal{E}}\cdot\Big[\Delta_{min}-\\
&\quad\beta\cdot\Big((1-\frac{O_1}{O_0})(1-p_{01})\Delta_{max}+\frac{O_1(p_{11}-p_{01})\Delta_{max}}{1-(1-O_1)(p_{11}-p_{01})}\Big)\Big]\\
\geq&0,
\end{aligned}
$$

where the first two inequalities follow the induction result of Lemma 8, the third inequality follows the induction result of Lemma 9–10 and Corollary 2, the fourth inequality follows the condition in the lemma.

Case 3: $l,m\geq k$. This case follows Lemma 5.

Lemma 6 is thus proven for slot t.

We then proceed to prove Lemma 9. We start with the first inequality. We develop $\widehat{W}_t$ w.r.t. $\widehat{\omega}_k$ and $\widehat{\omega}_N$ according to Lemma 6 as follows:

$$
\begin{aligned}
&\widehat{W}_t(\dot{\Omega};(\widehat{\omega}_1,\cdots,\widehat{\omega}_{k-1},\widehat{\omega}_k,\cdots,\widehat{\omega}_{n-1},\widehat{\omega}_n))\\
&-\widehat{W}_t(\dot{\Omega};(\widehat{\omega}_n,\widehat{\omega}_1,\cdots,\widehat{\omega}_{k-1},\widehat{\omega}_k,...,\widehat{\omega}_{n-1}))\\
&=\widehat{\omega}_k\widehat{\omega}_n\Big[\widehat{W}_t(\dot{\Omega};(\widehat{\omega}_1,\cdots,\widehat{\omega}_{k-1},1,\widehat{\omega}_{k+1},\cdots,\widehat{\omega}_{n-1},1))\\
&-\widehat{W}_t(\dot{\Omega};(1,\widehat{\omega}_1,\cdots,\widehat{\omega}_{k-1},1,\widehat{\omega}_{k+1},\cdots,\widehat{\omega}_{n-1}))\Big]\\
&+\widehat{\omega}_k(1-\widehat{\omega}_n)\Big[\widehat{W}_t(\dot{\Omega};(\widehat{\omega}_1,\cdots,\widehat{\omega}_{k-1},1,\widehat{\omega}_{k+1},\cdots,\widehat{\omega}_{n-1},0))\\
&-\widehat{W}_t(\dot{\Omega};(0,\widehat{\omega}_1,\cdots,\widehat{\omega}_{k-1},1,\widehat{\omega}_{k+1},\cdots,\widehat{\omega}_{n-1}))\Big]\\
&+(1-\widehat{\omega}_k)\widehat{\omega}_n\Big[\widehat{W}_t(\dot{\Omega};(\widehat{\omega}_1,\cdots,\widehat{\omega}_{k-1},0,\widehat{\omega}_{k+1},\cdots,\widehat{\omega}_{n-1},1))\\
&-\widehat{W}_t(\dot{\Omega};(1,\widehat{\omega}_1,\cdots,\widehat{\omega}_{k-1},0,\widehat{\omega}_{k+1},\cdots,\widehat{\omega}_{n-1}))\Big]\\
&+(1-\widehat{\omega}_k)(1-\widehat{\omega}_n)\Big[\widehat{W}_t(\dot{\Omega};(\widehat{\omega}_1,\cdots,\widehat{\omega}_{k-1},0,\widehat{\omega}_{k+1},\cdots,\widehat{\omega}_{n-1},0))\\
&-\widehat{W}_t(\dot{\Omega};(0,\widehat{\omega}_1,\cdots,\widehat{\omega}_{k-1},0,\widehat{\omega}_{k+1},\cdots,\widehat{\omega}_{n-1}))\Big] \qquad (13)
\end{aligned}
$$

Let $\mathcal{M}=\{1,\cdots,k-1\}$. We proceed the proof by upbounding the four terms in (13).

For the first term, we have

$$
\begin{aligned}
&\widehat{W}_t(\dot{\Omega};(\widehat{\omega}_1,\cdots,\widehat{\omega}_{k-1},1,\widehat{\omega}_{k+1},\cdots,\widehat{\omega}_{n-1},1))\\
&\quad-\widehat{W}_t(\dot{\Omega};(1,\widehat{\omega}_1,\cdots,\widehat{\omega}_{k-1},1,\widehat{\omega}_{k+1},\cdots,\widehat{\omega}_{n-1}))\\
&=\beta\sum_{\mathcal{E}\subseteq\mathcal{M}}\widehat{\mathcal{C}}_{\mathcal{M}}^{\mathcal{E}}\\
&\quad\Big[(1-O_1)\widehat{W}_{t+1}(\dot{\Omega};(\mathbf{\Phi}(\mathcal{E}),p_{11},\mathbf{\Upsilon}_{k+1}^{N-1},p_{11},\mathbf{\Psi}(\mathcal{M},\mathcal{E})))\\
&\quad+O_1\widehat{W}_{t+1}(\dot{\Omega};(\mathbf{\Phi}(\mathcal{E}),\mathbf{\Upsilon}_{k+1}^{N-1},p_{11},\mathbf{\Psi}(\mathcal{M},\mathcal{E}),p_{11}))\\
&\quad-(1-O_1)\widehat{W}_{t+1}(\dot{\Omega};(p_{11},\mathbf{\Phi}(\mathcal{E}),p_{11},\mathbf{\Upsilon}_{k+1}^{N-1},\mathbf{\Psi}(\mathcal{M},\mathcal{E})))\\
&\quad-O_1\widehat{W}_{t+1}(\dot{\Omega};(\mathbf{\Phi}(\mathcal{E}),p_{11},\mathbf{\Upsilon}_{k+1}^{N-1},p_{11},\mathbf{\Psi}(\mathcal{M},\mathcal{E})))\Big]\\
&\le 0
\end{aligned}
$$

where, the inequality follows the induction of Lemma 8.

For the second term, we have

$$
\begin{aligned}
&\widehat{W}_t(\dot{\Omega};(\widehat{\omega}_1,\cdots,\widehat{\omega}_{k-1},1,\widehat{\omega}_{k+1},\cdots,\widehat{\omega}_{n-1},0))\\
&\quad-\widehat{W}_t(\dot{\Omega};(0,\widehat{\omega}_1,\cdots,\widehat{\omega}_{k-1},1,\widehat{\omega}_{k+1},\cdots,\widehat{\omega}_{n-1}))\\
&=F(\widehat{\omega}_1,\cdots,\widehat{\omega}_{k-1},1)-F(0,\widehat{\omega}_1,\cdots,\widehat{\omega}_{k-1})+\beta\sum_{\mathcal{E}\subseteq\mathcal{M}}\widehat{\mathcal{C}}_{\mathcal{M}}^{\mathcal{E}}\cdot\\
&\quad\Big[(1-O_1)\widehat{W}_{t+1}(\dot{\Omega};(\mathbf{\Phi}(\mathcal{E}),p_{11},\mathbf{\Upsilon}_{k+1}^{N-1},p_{01},\mathbf{\Psi}(\mathcal{M},\mathcal{E})))\\
&\quad+O_1\widehat{W}_{t+1}(\dot{\Omega};(\mathbf{\Phi}(\mathcal{E}),\mathbf{\Upsilon}_{k+1}^{N-1},p_{01},\mathbf{\Psi}(\mathcal{M},\mathcal{E}),p_{11}))\\
&\quad-(1-O_0)\widehat{W}_{t+1}(\dot{\Omega};(p_{01},\mathbf{\Phi}(\mathcal{E}),p_{11},\mathbf{\Upsilon}_{k+1}^{N-1},\mathbf{\Psi}(\mathcal{M},\mathcal{E})))\\
&\quad-O_0\widehat{W}_{t+1}(\dot{\Omega};(\mathbf{\Phi}(\mathcal{E}),p_{11},\mathbf{\Upsilon}_{k+1}^{N-1},p_{01},\mathbf{\Psi}(\mathcal{M},\mathcal{E})))\Big]\\
&=F(\widehat{\omega}_1,\cdots,\widehat{\omega}_{k-1},1)-F(0,\widehat{\omega}_1,\cdots,\widehat{\omega}_{k-1})+\beta\sum_{\mathcal{E}\subseteq\mathcal{M}}\widehat{\mathcal{C}}_{\mathcal{M}}^{\mathcal{E}}\cdot\\
&\quad\Big[(1-O_0)\big[\widehat{W}_{t+1}(\dot{\Omega};(\mathbf{\Phi}(\mathcal{E}),p_{11},\mathbf{\Upsilon}_{k+1}^{N-1},p_{01},\mathbf{\Psi}(\mathcal{M},\mathcal{E})))\\
&\quad-\widehat{W}_{t+1}(\dot{\Omega};(p_{01},\mathbf{\Phi}(\mathcal{E}),p_{11},\mathbf{\Upsilon}_{k+1}^{N-1},\mathbf{\Psi}(\mathcal{M},\mathcal{E})))\big]\\
&\quad+O_1\big[\widehat{W}_{t+1}(\dot{\Omega};(\mathbf{\Phi}(\mathcal{E}),\mathbf{\Upsilon}_{k+1}^{N-1},p_{01},\mathbf{\Psi}(\mathcal{M},\mathcal{E}),p_{11}))\\
&\quad-\widehat{W}_{t+1}(\dot{\Omega};(\mathbf{\Phi}(\mathcal{E}),p_{11},\mathbf{\Upsilon}_{k+1}^{N-1},p_{01},\mathbf{\Psi}(\mathcal{M},\mathcal{E})))\big]\Big]\\
&\le\Delta_{max}+\beta\sum_{\mathcal{E}\subseteq\mathcal{M}}\widehat{\mathcal{C}}_{\mathcal{M}}^{\mathcal{E}}\cdot\Big[(1-O_0)\\
&\quad\big[\widehat{W}_{t+1}(\dot{\Omega};(\mathbf{\Phi}(\mathcal{E}),p_{11},\mathbf{\Upsilon}_{k+1}^{N-1},\mathbf{\Psi}(\mathcal{M},\mathcal{E}),p_{01}))\\
&\quad-\widehat{W}_{t+1}(\dot{\Omega};(p_{01},\mathbf{\Phi}(\mathcal{E}),p_{11},\mathbf{\Upsilon}_{k+1}^{N-1},\mathbf{\Psi}(\mathcal{M},\mathcal{E})))\big]\Big]\\
&\le\Delta_{max}+\beta\cdot(1-O_0)\cdot\frac{1-p_{01}}{O_0}\Delta_{max}
\end{aligned}
$$

following the induction of Lemma 8–9.

For the third term, we have

$$
\begin{aligned}
&\widehat{W}_t(\dot{\Omega};(\widehat{\omega}_1,\cdots,\widehat{\omega}_{k-1},0,\widehat{\omega}_{k+1},\cdots,\widehat{\omega}_{n-1},1))\\
&\quad-\widehat{W}_t(\dot{\Omega};(1,\widehat{\omega}_1,\cdots,\widehat{\omega}_{k-1},0,\widehat{\omega}_{k+1},\cdots,\widehat{\omega}_{n-1}))\\
&=F(\widehat{\omega}_1,\cdots,\widehat{\omega}_{k-1},0)-F(1,\widehat{\omega}_1,\cdots,\widehat{\omega}_{k-1})+\beta\sum_{\mathcal{E}\subseteq\mathcal{M}}\widehat{\mathcal{C}}_{\mathcal{M}}^{\mathcal{E}}\cdot\\
&\quad\Big[(1-O_0)\widehat{W}_{t+1}(\dot{\Omega};(\mathbf{\Phi}(\mathcal{E}),p_{01},\mathbf{\Upsilon}_{k+1}^{N-1},p_{11},\mathbf{\Psi}(\mathcal{M},\mathcal{E})))\\
&\quad+O_0\widehat{W}_{t+1}(\dot{\Omega};(\mathbf{\Phi}(\mathcal{E}),\mathbf{\Upsilon}_{k+1}^{N-1},p_{11},\mathbf{\Psi}(\mathcal{M},\mathcal{E}),p_{01}))-\\
&\quad-(1-O_1)\widehat{W}_{t+1}(\dot{\Omega};(p_{11},\mathbf{\Phi}(\mathcal{E}),p_{01},\mathbf{\Upsilon}_{k+1}^{N-1},\mathbf{\Psi}(\mathcal{M},\mathcal{E})))\\
&\quad-O_1\widehat{W}_{t+1}(\dot{\Omega};(\mathbf{\Phi}(\mathcal{E}),p_{01},\mathbf{\Upsilon}_{k+1}^{N-1},p_{11},\mathbf{\Psi}(\mathcal{M},\mathcal{E})))\Big]\\
&\le F(\widehat{\omega}_1,\cdots,\widehat{\omega}_{k-1},0)-F(1,\widehat{\omega}_1,\cdots,\widehat{\omega}_{k-1})+\beta\sum_{\mathcal{E}\subseteq\mathcal{M}}\widehat{\mathcal{C}}_{\mathcal{M}}^{\mathcal{E}}\cdot\\
&\quad\Big[(1-O_0)\widehat{W}_{t+1}(\dot{\Omega};(p_{01},p_{11},\mathbf{\Phi}(\mathcal{E}),\mathbf{\Upsilon}_{k+1}^{N-1},\mathbf{\Psi}(\mathcal{M},\mathcal{E})))\\
&\quad+O_0\widehat{W}_{t+1}(\dot{\Omega};(p_{11},\mathbf{\Phi}(\mathcal{E}),\mathbf{\Upsilon}_{k+1}^{N-1},\mathbf{\Psi}(\mathcal{M},\mathcal{E}),p_{01}))-\\
&\quad-(1-O_1)\widehat{W}_{t+1}(\dot{\Omega};(p_{01},p_{11},\mathbf{\Phi}(\mathcal{E}),\mathbf{\Upsilon}_{k+1}^{N-1},\mathbf{\Psi}(\mathcal{M},\mathcal{E})))\\
&\quad-O_1\widehat{W}_{t+1}(\dot{\Omega};(p_{01},\mathbf{\Phi}(\mathcal{E}),\mathbf{\Upsilon}_{k+1}^{N-1},\mathbf{\Psi}(\mathcal{M},\mathcal{E}),p_{11}))\Big]\\
&\le-\Delta_{min}+\beta\sum_{\mathcal{E}\subseteq\mathcal{M}}\widehat{\mathcal{C}}_{\mathcal{M}}^{\mathcal{E}}\cdot O_0\cdot\\
&\quad\Big[\widehat{W}_{t+1}(\dot{\Omega};(p_{11},\mathbf{\Phi}(\mathcal{E}),\mathbf{\Upsilon}_{k+1}^{N-1},\mathbf{\Psi}(\mathcal{M},\mathcal{E}),p_{01}))\\
&\quad-(1-\frac{O_1}{O_0})\widehat{W}_{t+1}(\dot{\Omega};(p_{01},p_{11},\mathbf{\Phi}(\mathcal{E}),\mathbf{\Upsilon}_{k+1}^{N-1},\mathbf{\Psi}(\mathcal{M},\mathcal{E})))\\
&\quad-\frac{O_1}{O_0}\widehat{W}_{t+1}(\dot{\Omega};(p_{01},\mathbf{\Phi}(\mathcal{E}),\mathbf{\Upsilon}_{k+1}^{N-1},\mathbf{\Psi}(\mathcal{M},\mathcal{E}),p_{11}))\Big]\\
&\le-\Delta_{min}+\beta\sum_{\mathcal{E}\subseteq\mathcal{M}}\widehat{\mathcal{C}}_{\mathcal{M}}^{\mathcal{E}}\cdot O_0\cdot\Big[(1-\frac{O_1}{O_0})\frac{1-p_{01}}{O_0}\Delta_{max}\\
&\quad+\frac{O_1}{O_0}(p_{11}-p_{01})\Delta_{max}\frac{1-[\beta(1-O_1)(p_{11}-p_{01})]^{T-t}}{1-\beta(1-O_1)(p_{11}-p_{01})}\Big]\\
&\le\sum_{\mathcal{E}\subseteq\mathcal{M}}\widehat{\mathcal{C}}_{\mathcal{M}}^{\mathcal{E}}\cdot\Big[-\Delta_{min}+\beta\cdot\big[(1-\frac{O_1}{O_0})(1-p_{01})\Delta_{max}\\
&\quad+\Delta_{max}\frac{O_1(p_{11}-p_{01})}{1-(1-O_1)(p_{11}-p_{01})}\big]\Big]\le 0
\end{aligned}
$$

where the first two inequalities follow the induction result of Lemma 8, the third equality follows the induction result of Lemma 9, the forth inequality is due the condition in Lemma 10.

For the fourth term, we have

$$
\begin{aligned}
&\widehat{W}_t(\dot{\Omega};(\widehat{\omega}_1,\cdots,\widehat{\omega}_{k-1},0,\widehat{\omega}_{k+1},\cdots,\widehat{\omega}_{n-1},0))\\
&\quad-\widehat{W}_t(\dot{\Omega};(0,\widehat{\omega}_1,\cdots,\widehat{\omega}_{k-1},0,\widehat{\omega}_{k+1},\cdots,\widehat{\omega}_{n-1}))\\
&=\beta\sum_{\mathcal{E}\subseteq\mathcal{M}}\widehat{\mathcal{C}}_{\mathcal{M}}^{\mathcal{E}}\cdot\\
&\quad\Big[(1-O_0)\widehat{W}_{t+1}(\dot{\Omega};(\mathbf{\Phi}(\mathcal{E}),p_{01},\mathbf{\Upsilon}_{k+1}^{N-1},p_{01},\mathbf{\Psi}(\mathcal{M},\mathcal{E})))
\end{aligned}
$$

$$
\begin{aligned}
&+ O_0\widehat{W}_{t+1}(\dot{\Omega}; (\mathbf{\Phi}(\mathcal{E}), \mathbf{\Upsilon}_{k+1}^{N-1}, p_{01}, \mathbf{\Psi}(\mathcal{M},\mathcal{E}), p_{01}))\\
&- (1-O_0)\widehat{W}_{t+1}(\dot{\Omega}; (p_{01}, \mathbf{\Phi}(\mathcal{E}), p_{01}, \mathbf{\Upsilon}_{k+1}^{N-1}, \mathbf{\Psi}(\mathcal{M},\mathcal{E})))\\
&- O_0\widehat{W}_{t+1}(\dot{\Omega}; (\mathbf{\Phi}(\mathcal{E}), p_{01}, \mathbf{\Upsilon}_{k+1}^{N-1}, p_{01}, \mathbf{\Psi}(\mathcal{M},\mathcal{E})))\Big]\\
\le&\beta \sum_{\mathcal{E}\subseteq\mathcal{M}} \widehat{\mathcal{C}}_{\mathcal{M}}^{\mathcal{E}}\cdot\\
&\Big[(1-O_0)\widehat{W}_{t+1}(\dot{\Omega}; (\mathbf{\Phi}(\mathcal{E}), p_{01}, \mathbf{\Upsilon}_{k+1}^{N-1}, \mathbf{\Psi}(\mathcal{M},\mathcal{E}), p_{01}))\\
&+ O_0\widehat{W}_{t+1}(\dot{\Omega}; (\mathbf{\Phi}(\mathcal{E}), \mathbf{\Upsilon}_{k+1}^{N-1}, p_{01}, \mathbf{\Psi}(\mathcal{M},\mathcal{E}), p_{01}))\\
&- (1-O_0)\widehat{W}_{t+1}(\dot{\Omega}; (p_{01}, \mathbf{\Phi}(\mathcal{E}), p_{01}, \mathbf{\Upsilon}_{k+1}^{N-1}, \mathbf{\Psi}(\mathcal{M},\mathcal{E})))\\
&- O_0\widehat{W}_{t+1}(\dot{\Omega}; (p_{01}, \mathbf{\Phi}(\mathcal{E}), \mathbf{\Upsilon}_{k+1}^{N-1}, p_{01}, \mathbf{\Psi}(\mathcal{M},\mathcal{E})))\Big]\\
\le&\beta\cdot(1-O_0)\cdot\frac{1-p_{01}}{O_0}\Delta_{max} + \beta\cdot O_0\cdot\frac{1-p_{01}}{O_0}\Delta_{max}\\
=&\beta\frac{1-p_{01}}{O_0}\cdot\Delta_{max}
\end{aligned}
$$

where, the first inequality follows the induction result of Lemma 8 and the second inequality follows the induction result of Lemma 9.

Combing the above results of the four terms, we have

$$
\begin{aligned}
&\widehat{W}_t(\dot{\Omega}; (\widehat{\omega}_1, \cdots, \widehat{\omega}_N)) - \widehat{W}_t(\dot{\Omega}; (\widehat{\omega}_N, \widehat{\omega}_1, \cdots, \widehat{\omega}_{N-1}))\\
\le&\widehat{\omega}_k(1-\widehat{\omega}_N)\cdot[\Delta_{max} + \beta\cdot(1-O_0)\cdot\frac{1-p_{01}}{O_0}\Delta_{max}]\\
&+ (1-\widehat{\omega}_k)(1-\widehat{\omega}_N)\cdot\beta\frac{1-p_{01}}{O_0}\Delta_{max}\\
\le&\widehat{\omega}_k(1-\widehat{\omega}_N)\cdot[\Delta_{max} + (1-O_0)\cdot\frac{1-p_{01}}{O_0}\Delta_{max}]\\
&+ (1-\widehat{\omega}_k)(1-\widehat{\omega}_N)\cdot\frac{1-p_{01}}{O_0}\Delta_{max}\\
=&\Delta_{max}\frac{1-\widehat{\omega}_N}{O_0}[\widehat{\omega}_k O_0 + (1-p_{01})(1-\widehat{\omega}_k O_0)]\\
\le&\Delta_{max}\frac{1-\widehat{\omega}_N}{O_0}[\widehat{\omega}_k O_0 + (1-\widehat{\omega}_k O_0)] \le \frac{1-\widehat{\omega}_N}{O_0}\Delta_{max}\\
\le&\frac{1-p_{01}}{O_0}\Delta_{max},
\end{aligned}
$$

which completes the proof of the Lemma 9.

Finally, **We then proceed to prove Lemma 10**. To this end, denote $\mathcal{M} \triangleq \{2, \cdots, k\}$, we have

$$
\begin{aligned}
&\widehat{W}_t(\dot{\Omega}; (\widehat{\omega}_1, \widehat{\omega}_2 \cdots, \widehat{\omega}_{N-1}, \widehat{\omega}_N))\\
&\qquad\qquad - \widehat{W}_t(\dot{\Omega}; (\widehat{\omega}_N, \widehat{\omega}_2, \cdots, \widehat{\omega}_{N-1}, \widehat{\omega}_1))\\
&= (\widehat{\omega}_1 - \widehat{\omega}_N)\\
&\Big[\widehat{W}_t(\dot{\Omega}; (1, \widehat{\omega}_2, \cdots, \widehat{\omega}_{N-1}, 0)) - \widehat{W}_t(\dot{\Omega}; (0, \widehat{\omega}_2, \cdots, \widehat{\omega}_{N-1}, 1))\Big]\\
&= (\widehat{\omega}_1 - \widehat{\omega}_N)\Big\{F(1, \widehat{\omega}_2, \cdots, \widehat{\omega}_k) - F(0, \widehat{\omega}_2, \cdots, \widehat{\omega}_k)\\
&+ \beta\sum_{\mathcal{E}\subseteq\mathcal{M}} \widehat{\mathcal{C}}_{\mathcal{M}}^{\mathcal{E}}\cdot\\
&\Big[(1-O_1)\widehat{W}_{t+1}(\dot{\Omega}; (p_{11}, \mathbf{\Phi}(\mathcal{E}), \mathbf{\Upsilon}_{k+1}^{N-1}, p_{01}, \mathbf{\Psi}(\mathcal{M},\mathcal{E})))\\
&+ O_1\widehat{W}_{t+1}(\dot{\Omega}; (\mathbf{\Phi}(\mathcal{E}), \mathbf{\Upsilon}_{k+1}^{N-1}, p_{01}, p_{11}, \mathbf{\Psi}(\mathcal{M},\mathcal{E})))\\
&- (1-O_0)\widehat{W}_{t+1}(\dot{\Omega}; (p_{01}, \mathbf{\Phi}(\mathcal{E}), \mathbf{\Upsilon}_{k+1}^{N-1}, p_{11}, \mathbf{\Psi}(\mathcal{M},\mathcal{E})))\\
&- O_0\widehat{W}_{t+1}(\dot{\Omega}; (\mathbf{\Phi}(\mathcal{E}), \mathbf{\Upsilon}_{k+1}^{N-1}, p_{11}, p_{01}, \mathbf{\Psi}(\mathcal{M},\mathcal{E})))\Big]\Big\}\\
&\le (\widehat{\omega}_1 - \widehat{\omega}_N)\Big\{\Delta_{max} + \beta\sum_{\mathcal{E}\subseteq\mathcal{M}} \widehat{\mathcal{C}}_{\mathcal{M}}^{\mathcal{E}}\cdot\\
&\Big[(1-O_1)\widehat{W}_{t+1}(\dot{\Omega}; (p_{11}, \mathbf{\Phi}(\mathcal{E}), \mathbf{\Upsilon}_{k+1}^{N-1}, p_{01}, \mathbf{\Psi}(\mathcal{M},\mathcal{E})))\\
&+ O_1\widehat{W}_{t+1}(\dot{\Omega}; (\mathbf{\Phi}(\mathcal{E}), \mathbf{\Upsilon}_{k+1}^{N-1}, p_{01}, p_{11}, \mathbf{\Psi}(\mathcal{M},\mathcal{E})))\\
&- (1-O_0)\widehat{W}_{t+1}(\dot{\Omega}; (p_{01}, \mathbf{\Phi}(\mathcal{E}), \mathbf{\Upsilon}_{k+1}^{N-1}, p_{11}, \mathbf{\Psi}(\mathcal{M},\mathcal{E})))\\
&- O_0\widehat{W}_{t+1}(\dot{\Omega}; (\mathbf{\Phi}(\mathcal{E}), \mathbf{\Upsilon}_{k+1}^{N-1}, p_{01}, p_{11}, \mathbf{\Psi}(\mathcal{M},\mathcal{E})))\Big]\Big\}\\
&= (\widehat{\omega}_1 - \widehat{\omega}_N)\Big\{\Delta_{max} + \beta\sum_{\mathcal{E}\subseteq\mathcal{M}} \widehat{\mathcal{C}}_{\mathcal{M}}^{\mathcal{E}}\cdot\\
&\Big[(1-O_1)[\widehat{W}_{t+1}(\dot{\Omega}; (p_{11}, \mathbf{\Phi}(\mathcal{E}), \mathbf{\Upsilon}_{k+1}^{N-1}, p_{01}, \mathbf{\Psi}(\mathcal{M},\mathcal{E})))\\
&- \widehat{W}_{t+1}(\dot{\Omega}; (p_{01}, \mathbf{\Phi}(\mathcal{E}), \mathbf{\Upsilon}_{k+1}^{N-1}, p_{11}, \mathbf{\Psi}(\mathcal{M},\mathcal{E})))]\\
&+ (O_0 - O_1)[\widehat{W}_{t+1}(\dot{\Omega}; (p_{01}, \mathbf{\Phi}(\mathcal{E}), \mathbf{\Upsilon}_{k+1}^{N-1}, p_{11}, \mathbf{\Psi}(\mathcal{M},\mathcal{E})))\\
&- \widehat{W}_{t+1}(\dot{\Omega}; (\mathbf{\Phi}(\mathcal{E}), \mathbf{\Upsilon}_{k+1}^{N-1}, p_{01}, p_{11}, \mathbf{\Psi}(\mathcal{M},\mathcal{E})))]\Big]\Big\}\\
&\le (\widehat{\omega}_1 - \widehat{\omega}_N)\Big\{\Delta_{max} + \beta\sum_{\mathcal{E}\subseteq\mathcal{M}} \widehat{\mathcal{C}}_{\mathcal{M}}^{\mathcal{E}}\cdot\\
&\Big[(1-O_1)[\widehat{W}_{t+1}(\dot{\Omega}; (p_{11}, \mathbf{\Phi}(\mathcal{E}), \mathbf{\Upsilon}_{k+1}^{N-1}, p_{01}, \mathbf{\Psi}(\mathcal{M},\mathcal{E})))\\
&- \widehat{W}_{t+1}(\dot{\Omega}; (\mathbf{\Phi}(\mathcal{E}), p_{01}, \mathbf{\Upsilon}_{k+1}^{N-1}, p_{11}, \mathbf{\Psi}(\mathcal{M},\mathcal{E})))]\Big]\Big\}\\
&\le (\widehat{\omega}_1 - \widehat{\omega}_N)\Big\{\Delta_{max} + \beta\sum_{\mathcal{E}\subseteq\mathcal{M}} \widehat{\mathcal{C}}_{\mathcal{M}}^{\mathcal{E}}\cdot\\
&\Big[(1-O_1)(\widehat{W}_{t+1}(\dot{\Omega}; (p_{11}, \mathbf{\Phi}(\mathcal{E}), \mathbf{\Upsilon}_{k+1}^{N-1}, \mathbf{\Psi}(\mathcal{M},\mathcal{E}), p_{01}))\\
&- \widehat{W}_{t+1}(\dot{\Omega}; (p_{01}, \mathbf{\Phi}(\mathcal{E}), \mathbf{\Upsilon}_{k+1}^{N-1}, \mathbf{\Psi}(\mathcal{M},\mathcal{E}), p_{11})))\Big]\Big\}\\
&\le (p_{11} - p_{01})\Big[\Delta_{max} + \sum_{\mathcal{E}\subseteq\mathcal{M}} \widehat{\mathcal{C}}_{\mathcal{M}}^{\mathcal{E}}\cdot\beta(1-O_1)(p_{11}-p_{01})\cdot\\
&\frac{1-[\beta(1-O_1)(p_{11}-p_{01})]^{T-t}}{1-\beta(1-O_1)(p_{11}-p_{01})}\Delta_{max}\Big]\\
&= \sum_{\mathcal{E}\subseteq\mathcal{M}} \widehat{\mathcal{C}}_{\mathcal{M}}^{\mathcal{E}}\cdot\Big[1 + \beta(1-O_1)(p_{11}-p_{01})\cdot\\
&\frac{1-[\beta(1-O_1)(p_{11}-p_{01})]^{T-t}}{1-\beta(1-O_1)(p_{11}-p_{01})}\Big](p_{11}-p_{01})\Delta_{max}\\
&= \frac{1-[\beta(1-O_1)(p_{11}-p_{01})]^{T-t+1}}{1-\beta(1-O_1)(p_{11}-p_{01})}(p_{11}-p_{01})\Delta_{max}
\end{aligned}
$$

where the first three inequalities follows the recursive application of the induction result of Lemma 8, the fourth inequality follows the induction result of Lemma 10.

We thus complete the whole process of Lemma 8–10.

REFERENCES

[1] J. Mitola and G. Q. Maguire. Cognitive radios: making software radios more personal. *IEEE Personal Comm.*, 6:13–18, Aug. 1999.

[2] S. Haykin. Cognitive Radio: Brain-Empowered Wireless Communications. *IEEE JSAC*, 23(2):201–220, 2005.

[3] R. Etkin, A. Parekh, and D. Tse. Spectrum sharing for unlicensed bands. *IEEE J. Select. Areas Commun.*, 25:517–528, Apr. 2007.

[4] Y. Xing, R. Chandramouli, S. Mangold, and S. Shankar N. Dynamic spectrum access in open spectrum wireless networks. *IEEE J. Select. Areas Commun.*, 24:626–637, Mar. 2006.

[5] Federal Communications Commission. Spectrum policy task force report (etdocket no.02-135). Nov. 2002.

[6] A. Ghasemi and E. S. Sousa. Fundamental limits of spectrum sharing in fading environments. *IEEE Trans. Wireless Commun.*, 6:649–658, Feb. 2007.

[7] R. Zhang and Y.-C. Liang. Exploiting multi-antennas for opportunistic spectrum sharing in cognitive radio networks. *IEEE J. Select. Topics Signal Processing*, 2:88–102, Feb. 2008.

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

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

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

[11] 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, Oct. 2007.

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

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

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

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

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

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

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

[19] K. Wang, L. Chen, K. Al Agha and Q. Liu. On optimality of myopic policy in opportunistic spectrum access: the case of sensing multiple channels and accessing one channel. *IEEE Wireless Communications Letters*, 1(5):452–455, 2012.

[20] 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. *IET Signal Processing*, 6(6):584–593, 2011.

[21] Michael Gasptar Krishnan Eswaran and Kannan Ramchandran. Cognitive radio through primary control feedback. *IEEE Journal on Seletec Areas in Communications*, 29(2):384–393, Feb. 2011.

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

[23] K. Wang, Q. Liu and Francis C.M. Lau. Multi-channel opportunistic access by overhearing primary arq messages. *IEEE Transactions on Vehicular Technology*, 99:1, 2013.

[24] H. Kim and K. G. Shin. Efficient discovery of spectrum opportunities with mac-layer sensing in cognitive radio networks. *IEEE Transactions on Mobile Computing*, 7:533–545, 2008.

[25] Chunxiao Jiang, Yan Chen, K.J.R. Liu and Yong Ren. Renewal-theoretical dynamic spectrum access in cognitive radio network with unknown primary behavior. *IEEE J. Select. Areas Commun.*, 31:406–416, 2013.

[26] K. Wang, L. Chen, Q. Liu, and Khaldoun Al Agha. On optimality of myopic sensing policy with imperfect sensing in multi-channel opportunistic access. *In submission to IEEE Transactions on Communications*, 2011.

Kehao Wang received the B.S degree in Electrical Engineering, M.s degree in Communication and Information System from Wuhan University of Technology, Wuhan, China, in 2003 and 2006, respectively, and Ph.D in the Department of Computer Science, the University of Paris-Sud XI, Orsay, France, in 2012. His research interests are cognitive radio networks, wireless network resource management, and data hiding.

Lin Chen received his B.E. degree in Radio Engineering from Southeast University, China in 2002 and the Engineer Diploma, Ph.D from Telecom ParisTech, Paris in 2005 and 2008, respectively. He also holds a M.S. degree of Networking from the University of Paris 6. He currently works as associate professor in the department of computer science of the 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.

Quan Liu is currently a Professor and the Dean of School of Information Engineering, Wuhan University of Technology, Wuhan, China. Her research interests include nonlinear system analysis and signal processing, etc. She has published more than 90 papers.