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

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

***Abstract*—We consider the channel access problem in a multi-channel opportunistic communication system with imperfect channel sensing, where the state of each channel evolves as an independent and identically distributed Markov process. The considered problem can be cast into a restless multi-armed bandit (RMAB) problem that is of fundamental importance in decision theory. It is well-known that the optimal policy of RMAB problem is intractable for its exponential computation complexity. A natural alternative is to consider the easily implementable myopic policy that maximizes the immediate reward but ignores the impact of the current strategy on the future reward. In this paper, we perform an analytical study on the optimality of the myopic policy under imperfect sensing for the considered RMAB problem. Specifically, for a family of generic and practically important utility functions, we establish the closed-form conditions to guarantee the optimality of the myopic policy even under imperfect sensing. Despite our focus on the opportunistic channel access, the obtained results are generic in nature and are widely applicable in a wide range of engineering domains.**

***Index Terms*—Restless multi-armed bandit (RMAB), myopic policy, imperfect sensing, opportunistic spectrum access (OSA).**

I. Introduction

We consider an opportunistic multi-channel communication system where a user has access to multiple channels, but is limited to sense and transmit on a subset of channels each time. The fundamental problem to study is how the user can exploit past observations and the knowledge of the stochastic properties of the channels to maximize its expected accumulated throughput by switching channels opportunistically.

Formally, there are N independent and identically distributed (i.i.d.) channels, each evolving as a two-state Markov process where the state of a channel indicates the availability of this channel. At each time slot, a user chooses k ($1 \leq k \leq N$) of the N channels to sense, and then transmit a certain amount of data depending on the states of the chosen channels. Given the initial states of these channels, the goal of the user is to find the optimal policy of sensing channels at each slot so as to maximize the expected accumulated throughput. This problem can be cast into the restless multi-armed bandit (RMAB) problem [1] or partially observable Markov decision process (POMDP) [2].

Due to its application in numerous engineering problems, the RMAB problem is of fundamental importance in stochastic decision theory. However, finding the optimal policy of the generic RMAB problem is shown to be PSPACE-hard by Papadimitriou *et al.* in [3]. Whittle proposed a heuristic index policy, called Whittle index policy [1] which is shown to be asymptotically optimal in certain limited regime under some specific constraints [4]. In this regard, Liu *et al.* studied in [5] the indexability of a class of RMAB problems relevant to dynamic multi-channel access applications. However, the optimality of the index policy based on Whittle index is not ensured in the general cases, especially when the channels follow non-identical Markov chains. Unfortunately, not every RMAB problem has a well-defined Whittle index. Moreover, computing the Whittle index can be prohibitively complex.

A natural alternative, given the intractability of the RMAB problem, is to seek a simple myopic policy maximizing the short-term reward. In this line of research, significant research efforts have been devoted to studying the performance of the myopic policy, especially in the context of opportunistic spectrum access (OSA). Key contributions from recent works on this subject can be summarized as follows:

- Zhao *et al.* [6] 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.* [7] derived the optimality of the myopic policy for the positively correlated i.i.d. channels when the user is limited to access one channel (i.e., $k = 1$) each time.
- Ahmad and Liu [8] further extended the optimality result to the case of sensing multiple i.i.d. channels ($k > 1$) for a particular form of utility function modeling the fact that the user gets one unit of reward for each channel sensed to be idle.
- Our previous works studied the case of non i.i.d. channels and provided generic conditions on the reward function under which the myopic policy is optimal [9], and also

Manuscript received August 6, 2012; revised November 26, 2012 and April 16, 2013. The editor coordinating the review of this paper and approving it for publication was L. Badia.

Part of the work of L. Chen is supported by the ANR (Agence Nationale de la Recherche) under the grant Green-Dyspan (ANR-12-IS03). Part of the work of Q. Liu is supported by the National Science Foundation of China under Grant 51175389.

K. Wang is with Key Laboratory of Fiber Optic Sensing Technology and Information Processing, and the School of Information, Wuhan University of Technology, 430070 Hubei, China (e-mail: www.kehao@gmail.com).

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

Q. Liu is with the school of Information Engineering, Wuhan University of Technology, 430070 Hubei, China (e-mail: Quan.Liu@whut.edu.cn).

Part of the work of K. Wang is supported by the National Natural Science Foundation of China under Grant 61303027 and Wuhan International Scientific and Technological Cooperation Project 2013030409020112.

Digital Object Identifier 10.1109/TCOMM.2013.071213.120573

illustrated that when these conditions are not satisfied, the myopic policy may not be optimal [10].

The vast majority of previous studies (i.e., [2], [6], [7], [8], [9]) in the area of OSA assume perfect observation of channel states. However, sensing or observation errors are inevitable in practical scenario (e.g., due to noise and system limitations), especially in wireless communication systems. More specifically, an idle (busy, respectively) channel may be sensed to be busy (idle) and accessing a busy channel leads to zero reward. In such context, it is crucial to study the structure and the optimality of the myopic sensing policy with imperfect observation. We would like to emphasize that the presence of sensing error brings two difficulties when studying the myopic sensing policy in this new context.

- The belief value evolves as a non-linear mapping instead of a linear one in the perfect sensing case.
- In the imperfect sensing case, the belief value update of a channel depends not only on the channel evolution itself, but also on the observation outcome, meaning that the transition is not deterministic.

Therefore, our problem requires an original study on the optimality of the myopic sensing policy that cannot draw on existing results in the perfect sensing case. We would like to report that despite its practical importance, very few work has been done on the impact of sensing error on the performance of the myopic sensing policy, or more generically, on the RMAB problem under imperfect observation.

To the best of our knowledge, [11], [12] are the only work in this area. Chen, Zhao and Swami *et al.* [11] decoupled the design of the sensing strategy from that of the spectrum sensor and the access strategy, reduced the constrained POMDP to an unconstrained one, and showed that the myopic sensor operating and access policies are optimal for the joint design of OSA, but left the optimality of the myopic sensing policy unaddressed. After that Liu, Zhao and Krishnamachari [12] established the optimality of the myopic sensing policy for the case of two channels under certain conditions and conjectured the optimality for arbitrary N under the same conditions. In this paper, we derive closed-form conditions to guarantee the optimality of the myopic sensing policy for arbitrary N and for a class of utility functions. As shown in Section III-C, the result obtained in this paper can cover the result of [12]. Moreover, this paper also significantly extends our previous work [9], focusing on perfect sensing scenario. In this regard, our work in this paper contributes the existing literature by developing an adapted analysis on the RMAB problem under imperfect sensing.

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

II. Problem Formulation

A. Multi-channel Opportunistic Access with Imperfect Sensing

As outlined in the Introduction, we consider a multi-channel opportunistic communication system, in which a user is able to access a set $\mathcal{N}$ of N i.i.d. channels, each characterized by a Markov chain of two states, *idle* (1) and *busy* (0). The state transmission probabilities are given by $\{p_{i,j}\}_{,i,j=0,1}$.

We assume that the system operates in a synchronous time slot fashion with the time slot indexed by t $(t = 1, 2, \cdots, T)$, where T is the time horizon of interest. Each channel goes through state transition at the beginning of each slot t. This generic multi-channel opportunistic communication model can be naturally cast into OSA problem in cognitive radio systems where an unlicensed secondary user can opportunistically access the temporarily unused channels of the licensed primary users, with the availability of each channel evolving as an independent Markov chain.

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

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

In our analysis, we consider the case where $\epsilon_i(t)$ and $\delta_i(t)$ are independent w.r.t. t and i. More specifically, we defined ϵ and δ as the system-wide false alarm rate and miss detection rate.

We also assume that when the receiver successfully receives a packet from a channel, it sends an acknowledgement (ACK) to the transmitter over the same channel at the end of the slot. The absence of an ACK (NACK) signifies that the transmitter does not transmit over this channel or transmitted but the channel is busy in this slot. We assume that acknowledgement are received without error since acknowledgements are always transmitted over idle channels [12].

Obviously, by sensing only k channels, the user cannot observe the state of the whole system. Hence, the user has to infer the channel states from its past decision and observation history so as to make its future decision. To this end, we define the *information state* (hereinafter referred to as *belief vector* for briefness) $\Omega(t) \triangleq \{\omega_i(t), i \in \mathcal{N}\}$, where $0 \le \omega_i(t) \le 1$ is the conditional probability that channel i is in idle state (i.e., $S_i(t) = 1$) at slot t given all past states, actions and observations[1]. As stated in [12], in order to ensure that the user and its intended receiver tune to the same channel in each slot, channel selections should be based on common observations $\{0 \text{ (NACK)}, 1 \text{ (ACK)}\}^k$ rather than the detection

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

outcomes at the transmitter. Due to the Markovian nature of the channel model, given the action $\mathcal{A}(t)$ and the observations $\{ACK_i(t) \in \{0,1\} : i \in \mathcal{A}(t)\}$, the belief vector can be updated recursively using Bayes Rule as shown in (1).

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

Note that the belief update under $ACK_i(t) = 0$ results from the fact that the receiver cannot distinguish a failed transmission (i.e., collides with the primary user with probability $\delta(1-\omega_i(t))$) from no transmission (with probability $\epsilon\omega_i(t) + (1-\delta)(1-\omega_i(t))$) [12]. For convenience, we introduce two operators $\varphi(\omega_i) \triangleq \frac{\epsilon\omega_i(t)}{\epsilon\omega_i(t)+1-\omega_i(t)}$ and $\tau(\omega_i(t)) \triangleq \omega_i(t)p_{11} + [1-\omega_i(t)]p_{01}$.

Remark. We would like to emphasize that in contrast to the perfect sensing case [9] where $\omega_i(t+1)$ is a linear function of $\omega_i(t)$ whether i is sensed or not, in the imperfect sensing case, the mapping from $\omega_i(t)$ to $\omega_i(t+1)$ is no longer linear due to the sensing error (cf. the second line of equation (1)). Moreover, the belief value update of a channel depends not only on the channel evolution itself, but also on the observation outcome, i.e., $\omega_i(t+1) = p_{11}$ for $i \in \mathcal{A}(t), ACK_i(t) = 1$ and $\omega_i(t+1) = \tau(\varphi(\omega_i(t)))$ for $i \in \mathcal{A}(t), ACK_i(t) = 0$. As will be shown later, these differences make the analysis for the imperfect sensing more complicated.

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

Lemma 1. *Given $p_{01} < p_{11}$, then*

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

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

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

Proof: Lemma 1 follows from $\tau(\omega_i(t)) = (p_{11} - p_{01})\omega_i(t) + p_{01}$ straightforwardly. Lemma 2 follows from $\varphi(\omega_i) = \frac{\epsilon\omega_i(t)}{\epsilon\omega_i(t)+1-\omega_i(t)}$, Lemma 2. ■

B. Optimal Sensing Problem and Myopic Sensing Policy

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

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

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

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

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

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

$$V_t(\Omega(t)) = \max_{\mathcal{A}(t)\subseteq\mathcal{N}} \mathbb{E}\Big[R_\pi(\Omega(t)) + \beta \sum_{\mathcal{E}\subseteq\mathcal{A}(t)} \prod_{i\in\mathcal{E}} (1-\epsilon)\omega_i(t) \prod_{j\in\mathcal{A}(t)\backslash\mathcal{E}} [1-(1-\epsilon)\omega_j(t)] V_{t+1}(\Omega(t+1))\Big].$$

In the above equations, $V_t(\Omega(t))$ is the value function corresponding to the maximal expected reward from time slot t to T with the belief vector $\Omega(t+1)$ following the evolution described in (1) given that the channels in the subset $\mathcal{E}$ are observed in idle state (i.e., receiving ACK) and the channels in $\mathcal{A}(t)\backslash\mathcal{E}$ are observed in busy state.

Theoretically, the optimal policy can be obtained by solving the above dynamic programming. Unfortunately, due to the impact of the current action on the future reward and the unaccountable space of the belief vector, obtaining the optimal solution directly from the above recursive equations is computationally prohibitive. Hence, a natural alternative is to seek simple myopic sensing policy which is easy to compute and implement that maximizes the expected immediate reward, denoted by $F(\Omega_A(t)) \triangleq \mathbb{E}[R_\pi(\Omega(t))]$ with $\Omega_A(t) \triangleq \{\omega_i(t) : i \in \mathcal{A}(t)\}$. Let $\overline{\mathcal{A}}(t)$ denote the set of channels sensed at slot t under myopic policy, $\overline{\mathcal{A}}(t)$ can be formally defined as follows:

$$\overline{\mathcal{A}}(t) = \operatorname*{argmax}_{\mathcal{A}(t)\subseteq\mathcal{N}} F(\Omega_A(t)). \quad (3)$$

In this paper, we focus on a class of *regular* functions defined in [9]. Specifically, the expected immediate reward function $F(\Omega_A(t))$ is assumed to be symmetrical, monotonically non-decreasing and decomposable, defined by three axioms in [9]. Under this condition, the myopic policy consists of choosing the k channels with the largest belief value. In the following sections we focus on the optimality of the myopic sensing policy under imperfect sensing. As pointed out in the remark following equations (1), the main technical difficulties compared to the perfect sensing case are the non-linear mapping from $\omega_i(t)$ to $\omega_i(t+1)$ and the dependency of the belief value update on the observation outcome.

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

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

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

- $\mathcal{N}(m)$ denotes the first m channles in belief vector;
- Given $\mathcal{E} \subseteq \mathcal{M} \subseteq \mathcal{N}$, $Pr(\mathcal{M},\mathcal{E}) \triangleq \prod_{i\in\mathcal{E}}(1-\epsilon)\omega_i(t) \prod_{j\in\mathcal{M}\setminus\mathcal{E}}[1-(1-\epsilon)\omega_j(t)]$;
- $\mathbf{P_{11}^{\mathcal{E}}}$ denotes the vector of length $|\mathcal{E}|$ with each element being p_{11};
- $\mathbf{\Phi}(l,m) \triangleq [\tau(\omega_i(t)) : l \le i \le m]$ where the components are sorted by belief value; $\mathbf{\Phi}_i(l,m) \triangleq [\tau(\omega_j(t)) : l \le j \le m, j \ne i, \omega_j(t) \ge \omega_i(t)]$; $\mathbf{\Phi}^j(l,m) \triangleq [\tau(\omega_i(t)) : l \le i \le m, i \ne j, \omega_j(t) > \omega_i(t)]$; $\mathbf{\Phi}_i^j(l,m) \triangleq [\tau(\omega_h(t)) : l \le h \le m, h \ne i, h \ne j, \omega_j(t) > \omega_h(t) \ge \omega_i(t)]$;
- Given $\mathcal{E} \subseteq \mathcal{M} \subseteq \mathcal{N}$, $\mathbf{Q}^{\mathcal{M},\mathcal{E}} \triangleq [\tau(\varphi(\omega_i(t))) : i \in \mathcal{M}\setminus\mathcal{E}]$ where the components are sorted by belief value; $\overline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{1}} \triangleq [\tau(\varphi(\omega_i(t))) : i \in \mathcal{M}\setminus\mathcal{E}\setminus\{l\}$ and $\omega_i(t) \ge \omega_l(t)]$; $\underline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{1}} \triangleq [\tau(\varphi(\omega_i(t))) : i \in \mathcal{M}\setminus\mathcal{E}\setminus\{l\}$ and $\omega_i(t) < \omega_l(t)]$;
- Let $\omega_{-i} \triangleq \{\omega_j : j \in \mathcal{A}, j \ne i\}$ and

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

A. Definition and Properties of Auxiliary Value Function

Inspired by the form of the value function $V_t(\Omega(t))$ and the analysis in [8], we define the auxiliary value function with imperfect sensing.

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

$$\begin{cases}W_T(\Omega(T)) = F(\Omega_{\overline{\mathcal{A}}}(T));\\ W_r(\Omega(r)) = F(\Omega_{\overline{\mathcal{A}}}(r)) \\ \qquad +\beta\sum_{\mathcal{E}\subseteq\overline{\mathcal{A}}(r)} Pr(\overline{\mathcal{A}}(r),\mathcal{E})W_{r+1}(\Omega_{\mathcal{E}}(r+1));\\ W_t(\Omega(t)) = F(\Omega_{\mathcal{N}(k)}(t)) \\ \qquad +\beta\sum_{\mathcal{E}\subseteq\mathcal{N}(k)} Pr(\mathcal{N}(k),\mathcal{E})W_{t+1}(\Omega_{\mathcal{E}}(t+1)),\end{cases} \tag{4}$$

where $\Omega_{\mathcal{E}}(t+1)$ 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.*

The above recursively defined auxiliary value function gives the expected discounted accumulated reward of the following sensing policy: in slot t sense the first k channels in the belief vector and then sense the channels in $\overline{\mathcal{A}}(r)$ $(t+1 \le r \le T)$ (i.e., adopt the myopic policy from slot $t+1$ to T). If $\mathcal{N}(k) = \overline{\mathcal{A}}(t)$, the above sensing policy is the myopic sensing policy with $W_t(\Omega(t))$ being the total reward from slot t to T.

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

Lemma 3 (Symmetry)**.** *Given $0 \le \epsilon \le \frac{(1-p_{11})p_{01}}{p_{11}(1-p_{01})}$, if F is regular, the correspondent auxiliary value function $W_t(\Omega(t))$ is symmetrical in ω_i,ω_j where $i,j \in \mathcal{A}(t)$ or $i,j \notin \mathcal{A}(t)$ for all $t = 1,2,\cdots,T$:*

$$W_t(\omega_1,\cdots,\omega_i,\cdots,\omega_j,\cdots,\omega_N) \\ = W_t(\omega_1,\cdots,\omega_j,\cdots,\omega_i,\cdots,\omega_N).$$

Proof: The proof is given in the appendix. ■

Lemma 4 (Decomposability)**.** *Given $0 \le \epsilon \le \frac{(1-p_{11})p_{01}}{p_{11}(1-p_{01})}$, if F is regular, the correspondent auxiliary value function $W_t(\Omega(t))$ is decomposable for $t = 1,\cdots,T$ and $\forall l \in \mathcal{N}$:*

$$W_t(\omega_1,\cdots,\omega_l,\cdots,\omega_N) = \omega_l W_t(\omega_1,\cdots,1,\cdots,\omega_N) \\ + (1-\omega_l)W_t(\omega_1,\cdots,0,\cdots,\omega_N).$$

Proof: The proof is given in the appendix. ■

To demonstrate the property of decomposability of the auxiliary function which is crucial to the study of the optimality, we provide an illustrative example in the following.

Example 1. Consider a system with $k = 2, N = 3, T = 2$, $F(\Omega_A) = (1-\epsilon)\sum_{i\in\mathcal{A}}\omega_i$ which is regular. Given the belief value setting $\omega_1 < \omega_2 < \omega_3$, the auxiliary function $W_t(\omega_1,\omega_2,\omega_3)$ can be developed based on (4) as follows:

$$\begin{aligned}&W_t(\omega_1,\omega_2,\omega_3) = (1-\epsilon)(\omega_1+\omega_2)\\ &+[(1-\epsilon)\omega_1][(1-\epsilon)\omega_2]W_{t+1}(p_{11},p_{11},\tau(\omega_3))\\ &+[(1-\epsilon)\omega_1][1-(1-\epsilon)\omega_2]W_{t+1}(p_{11},\omega_3,\tau(\varphi(\omega_2)))\\ &+[1-(1-\epsilon)\omega_1][(1-\epsilon)\omega_2]W_{t+1}(p_{11},\omega_3,\tau(\varphi(\omega_1)))\\ &+[(1-1-\epsilon)\omega_1][1-(1-\epsilon)\omega_2]W_{t+1}(\omega_3,\tau(\varphi(\omega_2)),\tau(\varphi(\omega_1)))\\ &=(1-\epsilon)(\omega_1+\omega_2)+[(1-\epsilon)\omega_1][(1-\epsilon)\omega_2](1-\epsilon)(p_{11}+p_{11})\\ &+[(1-\epsilon)\omega_1][1-(1-\epsilon)\omega_2](1-\epsilon)(p_{11}+\omega_3)\\ &+[1-(1-\epsilon)\omega_1][(1-\epsilon)\omega_2](1-\epsilon)(p_{11}+\omega_3)\\ &+[(1-1-\epsilon)\omega_1][1-(1-\epsilon)\omega_2](1-\epsilon)(\omega_3+\tau(\varphi(\omega_2)).\end{aligned}$$

In the above function, let $\omega_2 = 0, 1$, we have

$$\begin{aligned}&W_t(\omega_1,0,\omega_3) = (1-\epsilon)(\omega_1+0)\\ &+[(1-\epsilon)\omega_1][(1-\epsilon)0]W_{t+1}(p_{11},p_{11},\tau(\omega_3))\\ &+[(1-\epsilon)\omega_1][1-(1-\epsilon)0]W_{t+1}(p_{11},\omega_3,\tau(\varphi(0)))\\ &+[1-(1-\epsilon)\omega_1][(1-\epsilon)0]W_{t+1}(p_{11},\omega_3,\tau(\varphi(\omega_1)))\\ &+[(1-1-\epsilon)\omega_1][1-(1-\epsilon)0]W_{t+1}(\omega_3,\tau(\varphi(0)),\tau(\varphi(\omega_1)))\\ &=(1-\epsilon)(\omega_1+0)+[(1-\epsilon)\omega_1][(1-\epsilon)0](1-\epsilon)(p_{11}+p_{11})\\ &+[(1-\epsilon)\omega_1][1-(1-\epsilon)0](1-\epsilon)(p_{11}+\omega_3)\\ &+[1-(1-\epsilon)\omega_1][(1-\epsilon)0](1-\epsilon)(p_{11}+\omega_3)\\ &+[(1-1-\epsilon)\omega_1][1-(1-\epsilon)0](1-\epsilon)(\omega_3+\tau(\varphi(0))\\ &W_t(\omega_1,1,\omega_3) = (1-\epsilon)(\omega_1+1)\\ &+[(1-\epsilon)\omega_1][(1-\epsilon)1]W_{t+1}(p_{11},p_{11},\tau(\omega_3))\\ &+[(1-\epsilon)\omega_1][1-(1-\epsilon)1]W_{t+1}(p_{11},\omega_3,\tau(\varphi(1)))\\ &+[1-(1-\epsilon)\omega_1][(1-\epsilon)1]W_{t+1}(p_{11},\omega_3,\tau(\varphi(\omega_1)))\\ &+[(1-1-\epsilon)\omega_1][1-(1-\epsilon)1]W_{t+1}(\omega_3,\tau(\varphi(1)),\tau(\varphi(\omega_1)))\\ &=(1-\epsilon)(\omega_1+1)+[(1-\epsilon)\omega_1][(1-\epsilon)1](1-\epsilon)(p_{11}+p_{11})\\ &+[(1-\epsilon)\omega_1][1-(1-\epsilon)1](1-\epsilon)(p_{11}+\omega_3)\\ &+[1-(1-\epsilon)\omega_1][(1-\epsilon)1](1-\epsilon)(p_{11}+\omega_3)\\ &+[(1-1-\epsilon)\omega_1][1-(1-\epsilon)1](1-\epsilon)(\omega_3+\tau(\varphi(1))).\end{aligned}$$

We can check the decomposability by verifying that

$$W_t(\omega_1,\omega_2,\omega_3) = \omega_2 W_t(\omega_1,1,\omega_3) + (1-\omega_2)W_t(\omega_1,0,\omega_3).$$

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

Corollary 1. *Given* $0 \le \epsilon \le \frac{(1-p_{11})p_{01}}{p_{11}(1-p_{01})}$, *if* F *is regular, then for any* $l, m \in \mathcal{N}$, $t = 1, 2, \cdots, T$, *it holds*

$$\begin{aligned} & W_t(\omega_1, \cdots, \omega_l, \cdots, \omega_m, \cdots, \omega_N) \\ & \qquad - W_t(\omega_1, \cdots, \omega_m, \cdots, \omega_l, \cdots, \omega_N) \\ & = (\omega_l - \omega_m)\Big[W_t(\omega_1, \cdots, 1, \cdots, 0, \cdots, \omega_N) \\ & \qquad - W_t(\omega_1, \cdots, 0, \cdots, 1, \cdots, \omega_N)\Big]. \end{aligned}$$

B. Optimality of Myopic Sensing under Imperfect Sensing

In this section, we study the optimality of the myopic sensing policy under imperfect sensing. We start by showing the following important auxiliary lemmas (Lemma 5, 7 and 8) and then establish the sufficient condition under which the optimality of the myopic sensing policy is guaranteed.

Lemma 5. *Given that (1)* F *is regular, (2)* $\epsilon < \frac{p_{01}(1-p_{11})}{P_{11}(1-p_{01})}$, *and (3)* $\beta \le \frac{\Delta_{min}}{\Delta_{max}\left[(1-\epsilon)(1-p_{01})+\frac{\epsilon(p_{11}-p_{01})}{1-(1-\epsilon)(p_{11}-p_{01})}\right]}$, *if* $p_{11} \ge \omega_i \ge p_{01}, i \in \mathcal{N}$, $l < m$ *and* $\omega_l > \omega_m$, $\forall\ 1 \le t \le T$, *it holds that*

$$\begin{aligned} & W_t(\omega_1, \cdots, \omega_l, \cdots, \omega_m, \cdots, \omega_N) \\ & \qquad \ge W_t(\omega_1, \cdots, \omega_m, \cdots, \omega_l, \cdots, \omega_N). \end{aligned}$$

Lemma 6. *Given that (1)* F *is regular, (2)* $\epsilon < \frac{p_{01}(1-p_{11})}{P_{11}(1-p_{01})}$, *and (3)* $\beta \le \frac{\Delta_{min}}{\Delta_{max}\left[(1-\epsilon)(1-p_{01})+\frac{\epsilon(p_{11}-p_{01})}{1-(1-\epsilon)(p_{11}-p_{01})}\right]}$, *if* $p_{11} \ge \omega_1 \ge \cdots \ge \omega_N \ge p_{01}$, *for any* $1 \le t \le T$, *it holds that*

$$\begin{aligned} & W_t(\omega_1, \cdots, \omega_{k-1}, \omega_k, \cdots, \omega_{N-1}, \omega_N) \\ & -W_t(\omega_1, \cdots, \omega_{k-1}, \omega_N, \omega_k, \cdots, \omega_{N-1}) \le (1-\omega_N)\Delta_{max}, \end{aligned}$$

Based on Lemma 3, $W_t(\omega_1, \cdots, \omega_{k-1}, \omega_N, \omega_k, \cdots, \omega_{N-1}) = W_t(\omega_N, \omega_1, \cdots, \omega_{k-1}, \omega_k, \cdots, \omega_{N-1})$, combined with Lemma 6, we have the following lemma (Lemma 7):

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

$$\begin{aligned} & W_t(\omega_1, \cdots, \omega_{k-1}, \omega_k, \cdots, \omega_{N-1}, \omega_N) \\ & -W_t(\omega_N, \omega_1, \cdots, \omega_{k-1}, \omega_k, \cdots, \omega_{N-1}) \le (1-\omega_N)\Delta_{max}, \end{aligned}$$

Lemma 8. *Given that (1)* F *is regular, (2)* $\epsilon < \frac{p_{01}(1-p_{11})}{P_{11}(1-p_{01})}$, *and (3)* $\beta \le \frac{\Delta_{min}}{\Delta_{max}\left[(1-\epsilon)(1-p_{01})+\frac{\epsilon(p_{11}-p_{01})}{1-(1-\epsilon)(p_{11}-p_{01})}\right]}$, *if* $p_{11} \ge \omega_1 \ge \cdots \ge \omega_N \ge p_{01}$, *for any* $1 \le t \le T$, *it holds that*

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

Lemma 5 states that by swapping two elements in Ω with the former larger than the latter, the user does not increase the total expected reward. Lemma 7 and 8, on the other hand, give the upper bounds on the difference of the total reward of the two swapping operations, swapping ω_N and ω_j ($j = N-1, \cdots, 1$) and swapping ω_1 and ω_N, respectively. For clarity of presentation, the detailed proofs of the three lemmas are deferred to the Appendix. From a technical point of view, it is insightful to compare the methodology in the proof with that in the analysis presented in [7] for the perfect sensing case with $k = 1$. The key point of the analysis in [7] lies in the coupling argument leading to Lemma 3 in [7]. This analysis, however, cannot be directly applied in the generic case with imperfect sensing due to the non-linear update of the belief vector as stated in the remark after equation (1). Hence, we base our analysis on the intrinsic structure of the auxiliary value function W and investigate the different 'branches' of channel realizations to derive the relevant bounds, which are further applied to study the optimality of the myopic sensing policy, as stated in the following theorem.

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

Proof: It suffices to show that for $t = 1, \cdots, T$, by sorting $\Omega(t)$ in decreasing order such that $\omega_1 \ge \cdots \ge \omega_N$, it holds that $W_t(\omega_1, \cdots, \omega_N) \ge W_t(\omega_{i_1}, \cdots, \omega_{i_N})$, 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 $(\omega_{i_1^*}, \cdots, \omega_{i_N^*}) \ne (\omega_1, \cdots, \omega_N)$, 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 5 at each swapping, we have $W_t(\omega_{i_1^*}, \cdots, \omega_{i_N^*}) \le W_t(\omega_1, \cdots, \omega_N)$, which contradicts to (5). Theorem 1 is thus proven. ■

As noted in [12], 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 value does not fall in $[p_{01}, p_{11}]$, all the the belief values are bounded in the interval from the second slot following Lemma 1. Hence our results can be extended by treating the first slot separately from the future slots.

C. Discussion

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

Consider the channel access problem in which the user is limited to sense k channels and gets one unit of reward if a sensed channel is in the idle state (i.e., receiving ACK), thus the utility function can be formulated as $F(\Omega_A) = (1-\epsilon)\sum_{i \in \mathcal{A}} \omega_i$. Note that the optimality of the myopic sensing policy under this model is studied in [12] for a subset of scenarios where $k = 1$, $N = 2$. We now study the generic case with $k, N \ge 2$. To that end, we apply Theorem 1. Notice in this example, we have $\Delta_{min} = \Delta_{max} = 1 - \epsilon$. We can then verify that when $\epsilon < \frac{p_{01}(1-p_{11})}{P_{11}(1-p_{01})}$, it holds that $\frac{\Delta_{min}}{\Delta_{max}[(1-\epsilon)(1-p_{01})+\frac{\epsilon(p_{11}-p_{01})}{1-(1-\epsilon)(p_{11}-p_{01})}]} > 1$. Therefore, when the condition 1 and 2 holds, the myopic sensing policy is optimal for any β. This result in generic cases significantly extends the results obtained in [12] where the optimality of the myopic policy is proved for the case of two channels and only conjectured for general cases.

Next consider another special scenario where the user can sense and access all channels that are sensed to be idle, and gets one unit of reward if any of the channels has a successful transmission. Under this model, the user wants to maximize its expected throughput. More specifically, the slot utility function $F = F(\Omega_A) = 1-\Pi_{i\in\mathcal{A}}[1-(1-\epsilon)\omega_i]$, which is regular. In this context, we have $\Delta_{max} = (1-\epsilon)^{k-1}p_{11}^{k-1}$ and $\Delta_{min} = (1-\epsilon)^{k-1}p_{01}^{k-1}$. The third condition on for the myopic policy to be optimal becomes $\beta \leq \frac{p_{01}^{k-1}}{p_{11}^{k-1}[(1-\epsilon)(1-p_{01})+\frac{\epsilon(p_{11}-p_{01})}{1-(1-\epsilon)(p_{11}-p_{01})}]}$. Particularly, when $\epsilon = 0$, $\beta \leq \frac{p_{01}^{k-1}}{p_{11}^{k-1}(1-p_{01})}$. It can be noted that even when there is no sensing error, the myopic policy is not ensured to be optimal for any β.

IV. Conclusion

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

Appendix A
Proof of Lemma 3

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

- *Case 1:* $i,j \in \mathcal{A}(t)$. Noticing that (1) both F and $\sum_{\mathcal{E}\subseteq\mathcal{N}(k)} Pr(\mathcal{N}(k),\mathcal{E}) = \sum_{\mathcal{E}\subseteq\mathcal{A}(t)} Pr(\mathcal{A}(t),\mathcal{E})$ 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 $W_{t+1}(\Omega_{\mathcal{E}}(t+1))$ is symmetrical w.r.t. ω_i and ω_j.
- *Case 2:* $i,j \notin \mathcal{A}(t)$. Noticing that (1) both F and $\sum_{\mathcal{E}\subseteq\mathcal{N}(k)} Pr(\mathcal{N}(k),\mathcal{E}) = \sum_{\mathcal{E}\subseteq\mathcal{A}(t)} Pr(\mathcal{A}(t),\mathcal{E})$ are unrelated to ω_i,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 $W_{t+1}(\Omega_{\mathcal{E}}(t+1))$ is symmetrical w.r.t. ω_i and ω_j.

Combing the analysis completes the proof.

Appendix B
Proof of Lemma 4

We prove the lemma by backward induction. Firstly, it can be checked that Lemma 4 holds for slot T.

Assume that Lemma 4 holds for slots $t+1,\cdots,T$, we now prove that it holds for slot t by distinguishing the following two cases.

Case 1: l is not sensed in slot t, i.e. $l \geq k+1$. In this case, let $\mathcal{M} \triangleq \mathcal{N}(k) = \{1,\cdots,k\}$, we have

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

where

$$\Omega_l^{\mathcal{E}}(t+1) = (\mathbf{P_{11}^{\mathcal{E}}}, \mathbf{\Phi}_l(k+1,N), \tau(\omega_l), \mathbf{\Phi}^l(k+1,N), \mathbf{Q}^{\mathcal{M},\mathcal{E}}).$$

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

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

where

$$\Omega_{l,0}^{\mathcal{E}}(t+1) = (\mathbf{P_{11}^{\mathcal{E}}}, \mathbf{\Phi}_l(k+1,N), p_{01}, \mathbf{\Phi}^l(k+1,N), \mathbf{Q}^{\mathcal{M},\mathcal{E}}),$$
$$\Omega_{l,1}^{\mathcal{E}}(t+1) = (\mathbf{P_{11}^{\mathcal{E}}}, \mathbf{\Phi}_l(k+1,N), p_{11}, \mathbf{\Phi}^l(k+1,N), \mathbf{Q}^{\mathcal{M},\mathcal{E}}).$$

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

$$W_{t+1}(\Omega_l^{\mathcal{E}}(t+1)) = (1-\omega_l)W_{t+1}(\Omega_{l,0}^{\mathcal{E}}(t+1)) + \omega_l W_{t+1}(\Omega_{l,1}^{\mathcal{E}}(t+1)). \quad (6)$$

From the induction result, we have

$$W_{t+1}(\Omega_l^{\mathcal{E}}(t+1)) \quad (7)$$
$$= \tau(\omega_l)\cdot W_{t+1}(\mathbf{P_{11}^{\mathcal{E}}}, \mathbf{\Phi}_l(k+1,N), 1, \mathbf{\Phi}^l(k+1,N), \mathbf{Q}^{\mathcal{M},\mathcal{E}}) + (1-\tau(\omega_l))W_{t+1}(\mathbf{P_{11}^{\mathcal{E}}}, \mathbf{\Phi}_l(k+1,N), 0, \mathbf{\Phi}^l(k+1,N), \mathbf{Q}^{\mathcal{M},\mathcal{E}}),$$
$$W_{t+1}(\Omega_{l,0}^{\mathcal{E}}(t+1)) \quad (8)$$
$$= p_{01}\cdot W_{t+1}(\mathbf{P_{11}^{\mathcal{E}}}, \mathbf{\Phi}_l(k+1,N), 1, \mathbf{\Phi}^l(k+1,N), \mathbf{Q}^{\mathcal{M},\mathcal{E}}) + (1-p_{01})W_{t+1}(\mathbf{P_{11}^{\mathcal{E}}}, \mathbf{\Phi}_l(k+1,N), 0, \mathbf{\Phi}^l(k+1,N), \mathbf{Q}^{\mathcal{M},\mathcal{E}}),$$
$$W_{t+1}(\Omega_{l,1}^{\mathcal{E}}(t+1)) \quad (9)$$
$$= p_{11}\cdot W_{t+1}(\mathbf{P_{11}^{\mathcal{E}}}, \mathbf{\Phi}_l(k+1,N), 1, \mathbf{\Phi}^l(k+1,N), \mathbf{Q}^{\mathcal{M},\mathcal{E}}) + (1-p_{11})W_{t+1}(\mathbf{P_{11}^{\mathcal{E}}}, \mathbf{\Phi}_l(k+1,N1), 0, \mathbf{\Phi}^l(k+1,N), \mathbf{Q}^{\mathcal{M},\mathcal{E}}).$$

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

Case 2: l is sensed in slot t, i.e. $l \leq k$. Let $\mathcal{M} \triangleq \mathcal{N}(k)\setminus\{l\} = \{1,\cdots,l-1,l+1,\cdots,k\}$, it follows (4) that

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

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

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

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

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

$$
\begin{aligned}
&[1-(1-\epsilon)\omega_l]\\
&W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}},\mathbf{\Phi}(k+1,N),\overline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}},\tau(\varphi(\omega_l)),\underline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}})\\
&=(1-\omega_l)W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}},\mathbf{\Phi}(k+1,N),\overline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}},p_{01},\underline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}})\\
&+\epsilon\omega_l W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}},\mathbf{\Phi}(k+1,N),\overline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}},p_{11},\underline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}}).
\end{aligned}\tag{10}
$$

From the induction result, we have

$$
\begin{aligned}
&W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}},\mathbf{\Phi}(k+1,N),\overline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}},\tau(\varphi(\omega_l)),\underline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}}) \quad (11)\\
&=\tau(\varphi(\omega_l))W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}},\mathbf{\Phi}(k+1,N),\overline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}},1,\underline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}})\\
&+(1-\tau(\varphi(\omega_l)))W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}},\mathbf{\Phi}(k+1,N),\overline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}},0,\underline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}}),\\
&W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}},\mathbf{\Phi}(k+1,N),\overline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}},p_{01},\underline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}}) \quad (12)\\
&=p_{01}W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}},\mathbf{\Phi}(k+1,N),\overline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}},1,\underline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}})\\
&+(1-p_{01})W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}},\mathbf{\Phi}(k+1,N),\overline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}},0,\underline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}}),\\
&W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}},\mathbf{\Phi}(k+1,N),\overline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}},p_{11},\underline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}}) \quad (13)\\
&=p_{11}W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}},\mathbf{\Phi}(k+1,N),\overline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}},1,\underline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}})\\
&+(1-p_{11})W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}},\mathbf{\Phi}(k+1,N),\overline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}},0,\underline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}}).
\end{aligned}
$$

Combing (11), (12), (13), we obtain (10).

Combing the above analysis completes our proof.

Appendix C
Proof of Lemma 5, Lemma 6, Lemma 7 and Lemma 8

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

We first show that Lemma 5 – 8 hold for slot T. It is easy to verify that Lemma 5 holds.

We then prove Lemma 6, 7 and 8. Noticing the conditions $p_{01}\le\omega_N\le\omega_k\le p_{11}\le 1$ in Lemma 7 and $p_{01}\le\omega_N\le\omega_1\le p_{11}$ in Lemma 8, we have

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

Lemma 6, 7 and 8 thus hold for slot T.

Assume that Lemma 5 – 8 hold for slots $T,\cdots,t+1$, we now prove that they hold for slot t.

We first prove Lemma 5. We distinguish the following three cases:

Case 1: $l,m\notin\mathcal{N}(k)$. This case follows Lemma 3.

Case 2: $l\in\mathcal{N}(k)$ and $m\notin\mathcal{N}(k)$. In this case, $\mathcal{M}\triangleq\mathcal{N}(k)\backslash\{l\}$, it can be noted that $\mathbf{Q}^{\mathcal{M},\mathcal{E}}=(\underline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}},\overline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}})=(\underline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{m}},\overline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{m}})$ and $(\mathbf{\Phi}_m(k+1,N),\mathbf{\Phi}^m(k+1,N))=(\mathbf{\Phi}_l(k+1,m-1),\mathbf{\Phi}_l(m+1,N),\mathbf{\Phi}^l(k+1,m-1),\mathbf{\Phi}^l(m+1,N))$. In this case, we have

$$
\begin{aligned}
&W_t(\omega_1,\cdots,\omega_l,\cdots,\omega_m,\cdots,\omega_N)-W_t(\omega_1,\cdots,\omega_m,\cdots,\omega_l,\cdots,\omega_N)\\
&=(\omega_l-\omega_m)\\
&[W_t(\omega_1,1,\cdots,0,\cdots,\omega_N)-W_t(\omega_1,\cdots,0,\cdots,1,\cdots,\omega_N)]\\
&=(\omega_l-\omega_m)\Big\{F(\omega_1,\cdots,1,\cdots,\omega_k)-F(\omega_1,\cdots,0,\cdots,\omega_k)\\
&+\beta\sum_{\mathcal{E}\subseteq\mathcal{M}}Pr(\mathcal{M},\mathcal{E})\Big[\\
&(1-\epsilon)W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}},p_{11},\mathbf{\Phi}_m(k+1,N),p_{01},\mathbf{\Phi}^m(k+1,N),\mathbf{Q}^{\mathcal{M},\mathcal{E}})\\
&+\epsilon W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}},\mathbf{\Phi}_m(k+1,N),p_{01},\mathbf{\Phi}^m(k+1,N),\overline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}},p_{11},\underline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{l}})\\
&-W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}},\mathbf{\Phi}_l(k+1,m-1),\mathbf{\Phi}_l(m+1,N),p_{11},\mathbf{\Phi}^l(k+1,m-1),\\
&\qquad\mathbf{\Phi}^l(m+1,N),\overline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{m}},p_{01},\underline{\mathbf{Q}}^{\mathcal{M},\mathcal{E},\mathbf{m}})\Big]\Big\}\\
&\ge(\omega_l-\omega_m)\Big\{\Delta_{min}+\beta\sum_{\mathcal{E}\subseteq\mathcal{M}}Pr(\mathcal{M},\mathcal{E})\\
&\Big[(1-\epsilon)W_{t+1}(p_{01},\mathbf{P}_{\mathbf{11}}^{\mathcal{E}},p_{11},\mathbf{\Phi}_m(k+1,N),\mathbf{\Phi}^m(k+1,N),\mathbf{Q}^{\mathcal{M},\mathcal{E}})\\
&+\epsilon W_{t+1}(p_{01},\mathbf{P}_{\mathbf{11}}^{\mathcal{E}},\mathbf{\Phi}_m(k+1,N),\mathbf{\Phi}^m(k+1,N),\mathbf{Q}^{\mathcal{M},\mathcal{E}},p_{11})\\
&-W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}},p_{11},\mathbf{\Phi}_l(k+1,m-1),\mathbf{\Phi}_l(m+1,N),\\
&\qquad\mathbf{\Phi}^l(k+1,m-1),\mathbf{\Phi}^l(m+1,N),\mathbf{Q}^{\mathcal{M},\mathcal{E}},p_{01})\Big]\Big\}\\
&=(\omega_l-\omega_m)\Big\{\Delta_{min}+\beta\sum_{\mathcal{E}\subseteq\mathcal{M}}Pr(\mathcal{M},\mathcal{E})\\
&\Big[(1-\epsilon)W_{t+1}(p_{01},\mathbf{P}_{\mathbf{11}}^{\mathcal{E}},p_{11},\mathbf{\Phi}_m(k+1,N),\mathbf{\Phi}^m(k+1,N),\mathbf{Q}^{\mathcal{M},\mathcal{E}})\\
&+\epsilon W_{t+1}(p_{01},\mathbf{P}_{\mathbf{11}}^{\mathcal{E}},\mathbf{\Phi}_m(k+1,N),\mathbf{\Phi}^m(k+1,N),\mathbf{Q}^{\mathcal{M},\mathcal{E}},p_{11})\\
&-W_{t+1}(\mathbf{P}_{\mathbf{11}}^{\mathcal{E}},p_{11},\mathbf{\Phi}_m(k+1,N),\mathbf{\Phi}^m(k+1,N),\mathbf{Q}^{\mathcal{M},\mathcal{E}},p_{01})\Big]\Big\}\\
&\ge(\omega_l-\omega_m)\Big[\Delta_{min}-\beta\sum_{\mathcal{E}\subseteq\mathcal{M}}Pr(\mathcal{M},\mathcal{E})\cdot\\
&\Big((1-\epsilon)(1-p_{01})\Delta_{max}\\
&+\epsilon(p_{11}-p_{01})\Delta_{max}\frac{1-[\beta(1-\epsilon)(p_{11}-p_{01})]^{T-t}}{1-\beta(1-\epsilon)(p_{11}-p_{01})}\Big)\Big]\\
&\ge(\omega_l-\omega_m)\sum_{\mathcal{E}\subseteq\mathcal{M}}Pr(\mathcal{M},\mathcal{E})\cdot\Big[\Delta_{min}\\
&-\beta\Big((1-\epsilon)(1-p_{01})\Delta_{max}+\frac{\epsilon(p_{11}-p_{01})\Delta_{max}}{1-(1-\epsilon)(p_{11}-p_{01})}\Big)\Big]\ge 0,
\end{aligned}
$$

where the first inequality follows the induction result of Lemma 5, the second inequality follows the induction result of Lemma 7 and 8, the forth inequality follows the condition in the lemma.

Case 3: $l,m\in\mathcal{N}(k)$. This case follows Lemma 3.

Lemma 5 is thus proven for slot t.

We then proceed to prove Lemma 6. We start with the first inequality. We develop W_t w.r.t. ω_k and ω_N according to Lemma 4 as follows:

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

$$
\begin{aligned}
&+\omega_k(1-\omega_n)[W_t(\omega_1,\cdots,\omega_{k-1},1,\omega_{k+1},\cdots,\omega_{n-1},0)\\
&-W_t(\omega_1,\cdots,\omega_{k-1},0,1,\omega_{k+1},\cdots,\omega_{n-1})]\\
&+(1-\omega_k)\omega_n[W_t(\omega_1,\cdots,\omega_{k-1},0,\omega_{k+1},\cdots,\omega_{n-1},1)\\
&-W_t(\omega_1,\cdots,\omega_{k-1},1,0,\omega_{k+1},\cdots,\omega_{n-1})]\\
&+(1-\omega_k)(1-\omega_n)[W_t(\omega_1,\cdots,\omega_{k-1},0,\omega_{k+1},\cdots,\omega_{n-1},0)\\
&-W_t(\omega_1,\cdots,\omega_{k-1},0,0,\omega_{k+1},\cdots,\omega_{n-1})]. \qquad (14)
\end{aligned}
$$

We proceed the proof by upbounding the four terms in (14).
For the first term, we have

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

where, the inequality follows the induction of Lemma 5.

For the second term, we have

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

following the induction of Lemma 5.

For the third term, we have

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

where the first inequality follows the induction result of Lemma 5, the second equality follows the induction result of Lemma 7 and 8, the forth inequality is due the condition in Lemma 7.

For the fourth term, we have

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

where the first inequality follows Lemma 5, the second follows the induction result of Lemma 7.

Combing the above results of the four terms, we have

$$
\begin{aligned}
&W_t(\omega_1,\cdots,\omega_N)-W_t(\omega_1,\cdots,\omega_{k-1},\omega_N,\omega_k,\cdots,\omega_{N-1})\\
\le&\omega_k(1-\omega_N)\cdot\Delta_{max}+(1-\omega_k)(1-\omega_N)\cdot(1-p_{01})\beta\Delta_{max}\\
\le&\omega_k(1-\omega_N)\Delta_{max}+(1-\omega_k)(1-\omega_N)\Delta_{max}\\
\le&(1-\omega_N)\Delta_{max},
\end{aligned}
$$

which completes the proof of Lemma 6.

Based on Lemma 3, $W_t(\omega_1,\cdots,\omega_{k-1},\omega_N,\omega_k,\cdots,\omega_{N-1}) = W_t(\omega_N,\omega_1,\cdots,\omega_{k-1},\omega_k,\cdots,\omega_{N-1})$, combined with Lemma 6, **we conclude the proof of Lemma 7**.

Finally, we prove Lemma 8. To this end, denote $\mathcal{M} \triangleq \{2,\cdots,k\}$, we have

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

where the first two inequalities follows the induction result of Lemma 5, the third inequality follows the induction result of Lemma 8.

We thus complete the whole process of proving Lemma 5–8.

REFERENCES

[1] P. Whittle, "Restless bandits: activity allocation in a changing world," *J. Applied Probability*, special vol. 25A, pp. 287–298, 1988.

[2] Q. Zhao, L. Tong, A. Swami, and Y. Chen, "Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: a POMDP framework," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 3, pp. 589–600, Apr. 2007.

[3] C. H. Papadimitriou and J. N. Tsitsiklis, "The complexity of optimal queueing network control," *Mathematics of Operations Research*, vol. 24, no. 2, pp. 293–305, 1999.

[4] R. R. Weber and G. Weiss, "On an index policy for restless bandits," *J. Applied Probability*, vol. 27, no. 1, pp. 637–648, 1990.

[5] K. Liu and Q. Zhao, "Indexability of restless bandit problems and optimality of whittle index for dynamic multichannel access," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5547–5567, 2010.

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

[7] S. H. Ahmad, M. Liu, T. Javidi, Q. Zhao, and B. Krishnamachari, "Optimality of myopic sensing in multi-channel opportunistic access," *IEEE Trans. Inf. Theory*, vol. 55, no. 9, pp. 4040–4050, 2009.

[8] S. H. Ahmad and M. Liu, "Multi-channel opportunistic access: a case of restless bandits with multiple plays," in *Proc. 2009 Allerton Conference on Communication, Control, and Computing*, pp. 1361–1368.

[9] K. Wang and L. Chen, "On optimality of myopic policy for restless multi-armed bandit problem: an axiomatic approach," *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp. 300–309, 2012.

[10] K. Wang, L. Chen, K. Al Agha, and Q. Liu, "On the optimality of myopic sensing in multi-channel opportunistic spectrum access: the case of sensing multiple channels and accessing one channel," *IEEE Wireless Commun. Lett.*, vol. 1, no. 5, pp. 452–455, Oct. 2011.

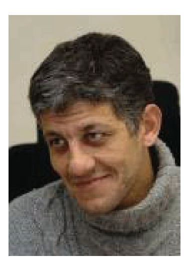
[11] Y. Chen, Q. Zhao, and A. Swami, "Joint design and separation principle for opportunistic spectrum access in the presence of sensing errors," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 2053–2071, May 2008.

[12] K. Liu, Q. Zhao, and B. Krishnamachari, "Dynamic multichannel access with imperfect channel state detection," *IEEE Trans. Signal Process.*, vol. 58, no. 5, pp. 2795–2807, May 2010.

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.



Khaldoun Al Agha (SM'03) received the Engineering degree from Supelec, Paris, France, in 1993, and the Master's and Ph.D. degrees from the University of Versailles, Versailles, France, in 1995 and 1993, respectively. He received the HDR from University of Paris-Sud11, Paris, in 2002. In 1998, he was an Assistant Professor with the University of Versailles. In 1999, he was with the French National Institute for Research in Computer Science and Control (INRIA) for one year. He is currently a Full Professor with the University of Paris-Sud11. He created and conducts the networking group at the LRI laboratory. He participates in different projects (BRAIN, SAMU, Arcade, SAFARI, OCARI, SARAH, RAF, NC2, etc.). His research interests are on resource allocation, security, and Quality of Service for cellular and ad hoc networks