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. Citation information: DOI 10.1109/TVT.2018.2870882, IEEE Transactions on Vehicular Technology

On Optimality of Second-Highest Policy for Opportunistic Multichannel Access

Kehao Wang, Lin Chen, Jihong Yu, Qingju Fan, Yang Zhang, Wei Chen, Pan Zhou, Yi Zhong

***Abstract*—We consider an opportunistic communication system in which one transmitter communicates with one receiver by one of N two-state Markov channels. The transmitter probes a channel before access. In particular, the transmitter does not access the probed channel if it is found to be in a bad state. Taking into account the probing cost, the transmitter will transmit over the chosen channel for a fixed time interval after probing. To maximize the throughput of the transmitter, we propose the second-highest probing policy, i.e., probing the second-best channel in terms of available probabilities of those channels. Further, we present three sets of conditions to guarantee the optimality of the policy for three scenarios, respectively. The conditions show that the optimality of the policy is tightly coupled with initial belief vector and non-trivial eigenvalue of two-state transition matrix. In addition, we extend the optimality of the policy to two related scenarios from the standpoint of exploitation *vs* exploration.**

***Index Terms*—Opportunistic access; myopic policy; optimality; initial value**

I. Introduction

We consider a generic opportunistic communication system in which one transmitter communicates with one receiver by one of N channels each time. In particular, each channel is modeled as an independent and identically distributed (i.i.d.) two-state discrete-time Markov process. Considering the probing cost, the transmitter is assumed to probe one channel each time and get the state of the probed channel. Depending on the channel state, the transmitter decides to choose one channel to deliver information for a fixed time interval, and then gets a certain reward for the interval. The objective of the transmitter is to seek a joint probing and accessing policy which maximizes the expected discounted reward accrued over a finite time horizon.

The procedure of seeking an optimal joint probing and accessing policy can be transformed into a partially observable Markov decision process (POMDP) [1] or a restless multi-armed bandit (RMAB) [2] problem which is proved to be PSPACE-Hard [3]. Thus, there exists a huge computational complexity in obtaining the optimal policy for a generic POMDP or RMAB.

For this reason, a natural alternative for the transmitter is to seek a simple myopic policy. In this regard, the authors of [4, 5] proposed a myopic policy, i.e., sensing the best channel, for the case of sensing and accessing one channel in each time slot, and then extended this policy to the case of sensing and accessing multiple channels in [6]. Moreover, they proved that the myopic sensing policy is optimal if the state transition of the Markov channels is positively correlated. In [7], the authors considered an opportunistic access scenario in which one transmitter senses k of N channels and accesses one of the sensed channels, and showed that sensing the best k channels is the optimal policy if $k+1 = N$, while not optimal generally for $k+1 < N$, by constructing a counterexample. In [8], the authors considered an opportunistic communication system similar to [7] except sensing k channels and accessing m ($1 \leqslant m \leqslant k$) of those k sensed channels, and gave some conditions for the optimality of the myopic policy by comparing some lower and upper bounds. In [9], we considered the heterogeneous Markov channels and proposed a set of sufficient conditions to guarantee the optimality of the myopic policy. In [10–12], we further investigated the opportunistic access with imperfect state observation, and presented some conditions on the optimality of the myopic policy. In [13], the authors considered the homogeneous multi-state Markov channels and proposed a set of sufficient conditions to guarantee the optimality of the myopic policy. In [14, 15], we explored the multi-state Markov channels with imperfect state observation, and showed that the myopic policy is optimal under some closed-form conditions on the channel transition matrices.

From [5–15], we know that one of the common points of those policies for the transmitter is to always access channel(s) from those sensed or probed in the latest slot, regardless of the sensing or probing results. However, in order to obtain more rewards, a natural strategy for the transmitter is to avoid accessing those channels which have been found to be bad in the latest slot. That is to say, if the states of those channels have been found to be bad, the transmitter should choose channels from those which are found to be good or not sensed (probed)

This work is supported in part by National NSF of China (no. 61672395, 61603283, 61601336, 61401169, 61701356). (Corresponding author: jihong-yu@hotmail.com)

K. Wang and Y. Zhong are with School of Information Engineering, Wuhan University of Technology, Wuhan 430070, China (email: {kehao.wang, zhongyi}@whut.edu.cn). Lin Chen is with the Laboratoire de Recherche en Informatique (LRI), Universit Paris-Sud and the Institut Universitaire de France (IUF), France (e-mail: chen@lri.fr). J. Yu is with School of Information and Electronics, Beijing Institute of Technology, Beijing, China. (email: jihong.yu@bit.edu.cn). Q. Fan is with School of Sciences, Wuhan University of Technology, Wuhan 430070, China (email: fanqingju@hotmail.com). Y. Zhang is with School of Computer Science, Wuhan University of Technology, Wuhan 430070, China (email: yangzhang@whut.edu.cn). W. Chen is with School of Automation, Wuhan University of Technology, Wuhan 430070, China (email: greatchen@whut.edu.cn). P. Zhou is with School of Electrical and Information Communications, Huazhong University of Science and Technology, Wuhan 430074, China (email: panzhou@hust.edu.cn).

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. Citation information: DOI 10.1109/TVT.2018.2870882, IEEE Transactions on Vehicular Technology

in the latest slot.

This guidance is integrated into the design of the joint probing and accessing policy for opportunistic communication [16, 17]. In particular, the authors of [16] considered the decision-making problem in which one transmitter is probing one channel and accessing one channel of N channels for a fixed time interval. In particular, if the probed channel is found to be in a good state, the transmitter transmits over it; otherwise, the transmitter chooses another one. Moreover, a myopic policy, namely the second-best probing policy, is proposed and shown to be optimal for the positively correlated case of three channels ($N = 3$) and conjectured to be optimal for the generic case with multiple channels ($N > 3$). In our previous work [17], we studied the joint probing and accessing policy for opportunistic communication in which the transmitter is allowed to probe k channels and access only one channel each time. Then we proposed an extended second-best policy, i.e., probing k channels from the second-best one, and proved that the policy is optimal for the case of $k+2 = N$, while not optimal as shown by a counterexample for $k+2 < N$.

There are many literatures in spectrum sensing and antenna selection [18–20]. In [18], the authors, through the POMDP approach, studied antenna selection at a receiver equipped with multiple antenna elements but only a single radio frequency chain for packet reception. In [19], the authors, based on sub-Nyquist sampling and the POMDP framework, developed an adaptive constrained wideband spectrum sensing method with no requirement of any prior knowledge. In [20], the authors considered spectrum sensing under secondary user hardware limitation and then proposed a random spectrum sensing strategy to select the subchannels to sense in a totally random fashion.

In this paper, following the conjectures of [16, 17], we carry out a deep investigation into the joint probing and accessing policy for opportunistic communication in which the transmitter is allowed to probe one channel and access one channel for a fixed time interval each time. Further, we propose a joint probing and accessing policy in which the transmitter probes the second-best channel and chooses the channel for transmission if it is found to be good; otherwise, the transmitter chooses the best channel.

The difference of our work from [16, 17] is that, first, we derive some sufficient conditions for the optimality of the second-best policy to avoid justifying whether the myopic policy is optimal in a generic case, and, second, we discuss the optimality of some similar policies in other scenarios from the standpoint of exploitation *vs* exploration.

Our contributions in this paper include

- We obtain several sets of closed-form sufficient conditions to guarantee the optimality of the proposed policy for both positively correlated case and negatively correlated case.
- We find that the optimality of the proposed policy is coupled with the initial belief information of the opportunistic system and the non-trivial eigenvalue of the state transition matrix.
- We extend the optimality to two related scenarios, probing multiple channels and accessing one channel, and probing two channels and accessing the better one of the two.

The paper is organized as follows. Section II formulates the problem and gives the motivation. Section III defines pseudo value function and investigates its structural properties. Section IV establishes the conditions under which the myopic policy is optimal. Section V extends the optimality to other related scenarios. Section VI verifies the theoretical results by numerical simulation. Section VII concludes the paper.

II. System Model and Problem Formulation

A. System Model

We consider an opportunistic communication system which is composed of N channels, one transmitter and one receiver. The transmitter is allowed to deliver information to the receiver over one of N independent channels $\mathcal{N} = \{1, 2, \cdots, N\}$. Each channel has two states, i.e., *good* (1) and *bad* (0), and evolves according to a Markovian process. All the channels are homogeneous and have the same transition matrix

$$\boldsymbol{P} = \begin{bmatrix} 1-p_{01} & p_{01} \\ 1-p_{11} & p_{11} \end{bmatrix} = \begin{bmatrix} 1-p_{11}+\lambda & p_{01} \\ 1-p_{11} & p_{01}+\lambda \end{bmatrix}, \quad (1)$$

where $\lambda := p_{11} - p_{01}$ is the non-trivial eigenvalue and '1' is the trivial eigenvalue of $\boldsymbol{P}$.

The opportunistic system is assumed to work in a synchronous fashion and the time horizon is divided into mini-slots, indexed by ς ($\varsigma = 0, 1, \cdots, T'-1$), herein T' is the number of mini-slots. Considering the energy cost of probing, the transmitter is limited to probing one of the N channels each time and accessing one channel based on the probed result at each mini-slot. Especially, once probed at a mini-slot, the transmitter would suspend probing for a fixed number, $K-1$, of mini-slots to further reduce the probing cost; that is, the transmitter would access the same channel for K continuous mini-slots, as shown in Fig. 1. We assume that the transmitter does not receive feedback information for previous transmissions.

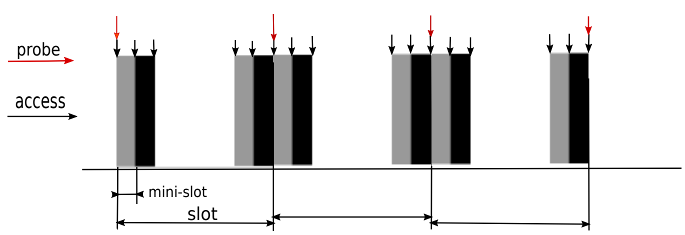


Fig. 1. Probing and Accessing model

Suppose that K continuous mini-slots constitute a slot. Thus, T' mini-slots are divided into $T := \lceil \frac{T'}{K} \rceil$ slots with slot index t ($t = 0, 1, \cdots, T-1$).

Let $s_i(t)$ be the state of channel i at the beginning of slot t, $b(t)$ be the channel probed at slot t, $o(t)$ be the observed state of channel $b(t)$, and $a(t)$ be the channel accessed at slot t. Denote by $A_t := \big(a(0), a(1), \cdots, a(t-1)\big)$ the access history,

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. Citation information: DOI 10.1109/TVT.2018.2870882, IEEE Transactions on Vehicular Technology

by $O_t := \big(o(0), o(1), \cdots, o(t-1)\big)$ the observation history, and by $B_t := \big(b(0), b(1), \cdots, b(t-1)\big)$ the probing history.

At each probing time, the transmitter only obtains the state of the probing channel and speculates the states of other channels from the past probing history and decision history information. Let $\omega_i(t)$ $(0 \le \omega_i(t) \le 1)$ denote the probability of channel i in 'good' state at the beginning of slot t given the history information[1]. Then we introduce a *belief vector* to characterize the opportunistic system state, i.e.,

$$\boldsymbol{w}(t) := \big(\omega_1(t), \omega_2(t), \cdots, \omega_N(t)\big). \tag{2}$$

Due to Markov feature, a belief value only depends on its latest value and is updated as follows

$$\omega_i(t+1) = \begin{cases} \phi(1), & i = b(t), o(t) = 1, \\ \phi(0), & i = b(t), o(t) = 0, \\ \phi(\omega_i(t)), & i \neq b(t), \end{cases} \tag{3}$$

where

$$\phi(\omega) := \tau^K(\omega) := \tau\big(\tau^{K-1}(\omega)\big), \tag{4}$$

$$\tau(\omega) := p_{11}\omega + p_{01}(1-\omega). \tag{5}$$

B. Decision Problem and Policy

Let $F(\omega_{a(t)}(t), \omega_{b(t)}(t))$ be the reward accrued in slot t (corresponding to K mini-slots), i.e.,

$$F(\omega_{a(t)}(t), \omega_{b(t)}(t)) := \sum_{k=0}^{K-1} \Big[\omega_{b(t)}(t)\tau^k(1) + (1-\omega_{b(t)}(t))\tau^k(\omega_{a(t)}(t))\Big]. \tag{6}$$

Proposition 1. $F(\omega_{a(t)}(t), \omega_{b(t)}(t))$ is an increasing function in both $\omega_{a(t)}(t)$ and $\omega_{b(t)}(t)$. Moreover, $F(\omega_{a(t)}(t), \omega_{b(t)}(t))$ is symmetric in $\omega_{a(t)}(t)$ and $\omega_{b(t)}(t)$, i.e.,

$$F(\omega_{a(t)}(t), \omega_{b(t)}(t)) = F(\omega_{b(t)}(t), \omega_{a(t)}(t)). \tag{7}$$

Proof. It is easy to check this result according to the definition of $F(\omega_{a(t)}(t), \omega_{b(t)}(t))$. □

Proposition 2. Given $\boldsymbol{w}(t) = (\omega_1(t), \omega_2(t), \cdots, \omega_N(t))$, if probing channel $b(t)$ and obtaining $o_t = 0$, then the optimal accessing policy at slot t is to access the channel

$$\bar{a}(t) = \operatorname*{argmax}_i \big\{\omega_i(t) : \ i \in \mathcal{N} - \{b(t)\}\big\}. \tag{8}$$

Proof. According to (11), (12), and Proposition 1, we know that $V_t(\boldsymbol{w}(t))$ is an increasing function in $\omega_{a(t)}(t)$, which leads to the proposition. □

According to Proposition 2, we know that seeking an optimal probing and accessing policy can be simplified into seeking an optimal probing policy. Thus, the objective of the transmitter is to find an optimal probing policy π^* which can maximize the expected reward collected over a finite number of slots (or mimi-slots).

Denote a probing policy by $\pi := (\pi_0, \pi_1, \cdots, \pi_{T-1})$ where π_t maps the belief vector $\boldsymbol{w}(t)$ to the action $b(t)$ in slot t, i.e.,

$$\pi_t : \ \boldsymbol{w}(t) \longmapsto b(t), \quad t = 0, 1, \cdots, T-1. \tag{9}$$

[1]Especially, $\omega_i(0) = \frac{p_{01}}{p_{01}+1-p_{11}}$ if no information is given.

Then we have the following optimal problem:

$$\pi^* = \operatorname*{argmax}_{\pi} \mathbb{E}\left\{\sum_{t=0}^{T-1} \beta^t R_{\pi_t}(\boldsymbol{w}(t)) \middle| \boldsymbol{w}(0)\right\}, \tag{10}$$

where β $(0 \le \beta \le 1)$ is a discount factor and $R_{\pi_t}(\boldsymbol{w}(t))$ denotes the reward accred under the mapping π_t in slot t and $\boldsymbol{w}(0)$.

For ease of analysis, we rewrite (10) in the language of dynamic programming,

$$V_T(\boldsymbol{w}(t)) = \max_{b(T)} \Big\{F(\omega_{\bar{a}(T)}(T), \omega_{b(T)}(T))\Big\}, \tag{11}$$

$$\begin{aligned} V_t(\boldsymbol{w}(t)) = \max_{b(t)} \Big\{ & F(\omega_{\bar{a}(t)}(t), \omega_{b(t)}(t)) \\ & + \beta\omega_{b(t)}(t)V_{t+1}(\boldsymbol{w}_{-b(t)}(t+1), \phi(1)) \\ & + \beta(1-\omega_{b(t)}(t))V_{t+1}(\boldsymbol{w}_{-b(t)}(t+1), \phi(0))\Big\}, \end{aligned} \tag{12}$$

where

$$\begin{aligned} \boldsymbol{w}_{-b(t)}(t+1) := \big(& \phi(\omega_1(t)), \cdots, \phi(\omega_{b(t)-1}(t)), \\ & \phi(\omega_{b(t)+1}(t)), \cdots, \phi(\omega_N(t))\big). \end{aligned}$$

Considering the huge difficulty from the recursive iteration in (12), we avoid computing the optimal policy π^* and turn to seek a simple myopic policy, which only maximizes the current reward, as follows:

$$\bar{b}(t) := \operatorname*{argmax}_{b(t) \in \mathcal{N}} \{F(\omega_{\bar{a}(t)}(t), \omega_{b(t)}(t))\}. \tag{13}$$

However, according to Proposition 1, if $\omega_1(t) \geqslant \omega_2(t) \geqslant \cdots \geqslant \omega_N(t)$, then $(b(t), a(t)) = (1, 2)$ or $(2, 1)$ generates the identically maximal slot reward; that is, the myopic policy is not unique. From the perspective of exploitation *vs.* exploration, probing channel $b(t) = 2$ would obtain more information from the system than $b(t) = 1$. Thus, the transmitter always probes channel $b(t) = 2$, and accesses channel $a(t) = 1$, which is defined as follows

Definition 1 (The Second-Highest Probing Policy)**.** The second-highest probing policy is to probe the second-best channel in terms of the belief values.

C. Motivation

Although the second-highest probing policy is easy to implement, we demonstrate by the following counterexample that the second-highest policy cannot be guaranteed to be optimal. This counterexample also shows that the conjecture in [17] is not true.

Counterexample. $N = 6$, $T = 3$, $K = 1$, $p_{11} = 0.5$, $p_{01} = 0.3$, $\boldsymbol{w}(0) = (0.999, 0.50, 0.49, 0.39, 0.25, 0.25)$. Let V_1 be the reward generated by the second-highest probing policy for $t = 0, 1, 2$, and V_2 be the reward generated by probing the third-highest channel for $t = 0$ and then probing the second-highest channel for $t = 1, 2$. Then we can get $V_2 - V_1 \approx 2.38338 - 2.38334 = 0.00004 > 0$, which shows the second-highest probing policy is not optimal in this setting.

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. Citation information: DOI 10.1109/TVT.2018.2870882, IEEE Transactions on Vehicular Technology

Thus, a natural question is under what condition the second-highest policy is optimal. The following sections will answer this question by seeking sufficient conditions to guarantee that the second-highest policy is optimal.

III. Pseudo Value Function

In this section, we first introduce the *pseudo value function* [9] and then derive the decomposability feature of the pseudo value function for (12). For convenience of presentation, in each slot t, $\boldsymbol{w}(t)$ is sorted in the descending order, i.e., $\omega_1(t) \geq \omega_2(t) \geq \cdots \geq \omega_N(t)$.

Definition 2. The pseudo value function for (12) can be written as

$$\begin{aligned} W_T(\boldsymbol{w}(T)) &= \omega_1(T) + \omega_2(T) - \omega_1(T)\omega_2(T), \\ W_r(\boldsymbol{w}(r)) &= \omega_1(r) + \omega_2(r) - \omega_1(r)\omega_2(r) \\ &\quad + \beta\omega_2(r)W_{r+1}(\boldsymbol{w}_{-2}(r+1), \phi(1)) \\ &\quad + \beta(1-\omega_2(r))W_{r+1}(\boldsymbol{w}_{-2}(r+1), \phi(0)), \\ W_t^{b(t)}(\boldsymbol{w}(t)) &= \omega_{\bar{a}(t)}(t) + \omega_{b(t)}(t) - \omega_{\bar{a}(t)}(t)\omega_{b(t)}(t) \\ &\quad + \beta\omega_{b(t)}(t)W_{t+1}(\boldsymbol{w}_{-b(t)}(t+1), \phi(1)) \\ &\quad + \beta(1-\omega_{b(t)}(t))W_{t+1}(\boldsymbol{w}_{-b(t)}(t+1), \phi(0)), \end{aligned}$$

where $t < r \leqslant T$.

$W_t^{b(t)}(\boldsymbol{w}(t))$ denotes the expected accumulated reward from slot t to T under the policy of probing the channel $b(t)$ for slot t and then probing the second-highest channel from slot $t+1$ to T. If $b(t) = 2$, then $W_t^{b(t)}(\boldsymbol{w}(t))$ is the total reward generated by the second-highest probing policy.

It is easy to show by backward induction that the second-highest policy is optimal if $W_t^{b(t)}(\boldsymbol{w}(t))$ achieves its maximum with $b(t) = 2$. Before establishing the optimality of the second-highest probing policy, we show the decomposability property of the pseudo value function for (12) in the following lemma.

Lemma 1. For $\forall\ i \in \mathcal{N}$ and $t = 0, 1, \cdots, T$, we have

$$\begin{aligned} W_t^{b(t)}(\omega_1, \cdots, \omega_i, \cdots, \omega_N) &= \omega_i W_t^{b(t)}(\omega_1, \cdots, 1, \cdots, \omega_N) \\ &\quad + (1-\omega_i)W_t^{b(t)}(\omega_1, \cdots, 0, \cdots, \omega_N). \end{aligned}$$

Proof. The lemma can be proven by backward induction following the same proof of [9]. □

Proposition 3. $F(\omega_{a(t)}, 1) - F(\omega_{a(t)}, 0)$ is decreasing in $\omega_{a(t)}$.

Proof.

$$\begin{aligned} F(\omega_{a(t)}, 1) - F(\omega_{a(t)}, 0) &= \sum_{k=0}^{K-1} \left[\tau^k(1) - \tau^k(\omega_{a(t)})\right] \\ &= (1-\omega_{a(t)}) \sum_{k=0}^{K-1} \lambda^k, \qquad (14) \end{aligned}$$

which is decreasing with $\omega_{a(t)}$. □

IV. Optimality Analysis

In this section, we investigate three scenarios, i.e., positively correlated channels, negatively correlated channels with odd K, and negatively correlated channels with even K, and propose three sets of conditions to guarantee the optimality of the second-highest policy for the three scenarios, respectively.

A. Positively Correlated Channels $\lambda \geqslant 0$

The following lemma gives some bounds on some exchange operations in different positions, under certain conditions on the initial belief values and the non-trivial eigenvalue of $\boldsymbol{P}$.

Lemma 2. Given $\lambda^K \leqslant 4\big[\frac{1-\phi(0)}{(1-\phi(1))(1-\omega_0)} - 1\big]$ and $\phi(0) \leqslant \omega_i \leqslant \phi(1)$ $(1 \leqslant i \leqslant n)$, we have for $0 \leqslant t \leqslant T-1$

1) if $\omega_i \geqslant \omega_{i+1}$ $(3 \leqslant i \leqslant n-1)$, then
$$W_t(\cdots, \omega_i, \omega_{i+1}, \cdots) \geqslant W_t(\cdots, \omega_{i+1}, \omega_i, \cdots). \qquad (15)$$
2) if $\omega_2 \geqslant \omega_3$, then
$$W_t(\omega_1, \omega_2, \omega_3, \omega_4, \cdots) \geqslant W_t(\omega_1, \omega_3, \omega_2, \omega_4, \cdots). \qquad (16)$$
3) if $\omega_1 \geqslant \omega_2$, then
$$\begin{aligned} 0 &\leqslant W_t(\omega_1, \omega_2, \omega_3, \cdots) - W_t(\omega_2, \omega_1, \omega_3, \cdots) \\ &\leqslant (\omega_1 - \omega_2)(F(\phi(1), 1) - F(\phi(1), 0)). \qquad (17) \end{aligned}$$
4) if $\omega_1 \geqslant \omega_2 \geqslant \cdots \geqslant \omega_n$, then
$$\begin{aligned} &W_t(\omega_1, \omega_2, \cdots, \omega_{n-1}, \omega_n) - W_t(\omega_1, \omega_n, \omega_2, \cdots, \omega_{n-1}) \\ &\leqslant F(\omega_1, 1) - F(\omega_1, 0). \qquad (18) \end{aligned}$$

Proof. See Appendix I. □

The following Lemma 3 is parallel to Lemma 2 under a different conditions on the non-trivial eigenvalue of $\boldsymbol{P}$.

Lemma 3. Given $\lambda^K \leqslant \frac{1-\phi(1)}{2-\phi(1)-\phi(0)}$ and $\phi(0) \leqslant \omega_i \leqslant \phi(1)$ $(1 \leqslant i \leqslant n)$, we have for $0 \leqslant t \leqslant T-1$

1) if $\omega_i \geqslant \omega_{i+1}$ $(3 \leqslant i \leqslant n-1)$, then
$$W_t(\cdots, \omega_i, \omega_{i+1}, \cdots) \geqslant W_t(\cdots, \omega_{i+1}, \omega_i, \cdots). \qquad (19)$$
2) if $\omega_2 \geqslant \omega_3$, then
$$W_t(\omega_1, \omega_2, \omega_3, \omega_4, \cdots) \geqslant W_t(\omega_1, \omega_3, \omega_2, \omega_4, \cdots). \qquad (20)$$
3) if $\omega_1 \geqslant \omega_2$, then
$$W_t(\omega_1, \omega_2, \omega_3, \cdots) - W_t(\omega_2, \omega_1, \omega_3, \cdots) \geqslant 0 \qquad (21)$$
4) if $\omega_1 \geqslant \omega_2 \geqslant \cdots \geqslant \omega_n$, then
$$\begin{aligned} &W_t(\omega_1, \omega_2, \cdots, \omega_{n-1}, \omega_n) - W_t(\omega_n, \omega_2, \cdots, \omega_{n-1}, \omega_1) \\ &\leqslant (\omega_1 - \omega_n)(F(\phi(0), 1) - F(\phi(0), 0))\frac{1}{1-\lambda^K}. \qquad (22) \end{aligned}$$

Proof. See Appendix II. □

Based on Lemma 2 and Lemma 3, we have the following.

Theorem 1. Given $\phi(0) \leqslant \omega_i(0) \leqslant \phi(1)$ $(1 \leqslant i \leqslant N)$, the second-highest probing policy is optimal if one of the following holds

1) $\lambda^K \leqslant 4\big[\frac{1-\phi(0)}{(1-\phi(1))(1-\omega_0)} - 1\big]$, or

2) $\lambda^K \leqslant \frac{1-\phi(1)}{2-\phi(1)-\phi(0)}$.

Proof. Case 1. $\lambda^K \leqslant 4\big[\frac{1-\phi(0)}{(1-\phi(1))(1-\omega_0)} - 1\big]$. In this case, we know that Lemma 2 holds. According to the rule of Bubble sort algorithm, we can easily check that $V_t(\omega_1, \omega_2, \cdots, \omega_N)$ is the maximal, which shows that the second-highest policy is optimal.

Case 2. $\lambda^K \leqslant \frac{1-\phi(1)}{2-\phi(1)-\phi(0)}$. In this case, Lemma 3 holds. Thus, following the similar proof, we know that the second-highest policy is optimal. □

Corollary 1. Given $\phi(0) \leqslant \omega_i(0) \leqslant \phi(1)$ $(1 \leqslant i \leqslant N)$, the second-highest probing policy is optimal if one of the following holds

1) $5 - 2\sqrt{5} \leqslant p_{01} \leqslant \omega_i(0) \leqslant p_{11}$, or
2) $p_{01} \leqslant \omega_i(0) \leqslant p_{11} \leqslant \frac{3-\sqrt{5}}{2}$.

Proof. Please refer to Appendix III. □

B. Negatively Correlated Channels $\lambda < 0$: odd K

For the case of negatively correlated channels with odd K, we have the following similar lemma.

Lemma 4. Given $\lambda^K \leqslant \frac{1-\phi(0)}{2-\phi(0)-\phi^2(0)}$ and $\phi(1) \leqslant \omega_i \leqslant \phi(0)$ $(1 \leqslant i \leqslant n)$, we have for $0 \leqslant t \leqslant T-1$

1) if $\omega_i \geqslant \omega_{i+1}$ $(3 \leqslant i \leqslant n-1)$, then
$$W_t(\cdots, \omega_i, \omega_{i+1}, \cdots) \geqslant W_t(\cdots, \omega_{i+1}, \omega_i, \cdots). \quad (23)$$
2) if $\omega_2 \geqslant \omega_3$, then
$$W_t(\omega_1, \omega_2, \omega_3, \omega_4, \cdots) \geqslant W_t(\omega_1, \omega_3, \omega_2, \omega_4, \cdots). \quad (24)$$
3) if $\omega_1 \geqslant \omega_2$, then
$$W_t(\omega_1, \omega_2, \omega_3, \cdots) \geqslant W_t(\omega_2, \omega_1, \omega_3, \cdots). \quad (25)$$
4) if $\phi(\omega_2) \geqslant \omega_3 \geqslant \cdots \geqslant \omega_n$, then
$$\begin{aligned} &W_t(\omega_1, \phi(\omega_2), \cdots, \omega_{n-1}, \omega_n) \\ &- W_t(\omega_n, \phi(\omega_2), \cdots, \omega_{n-1}, \omega_1) \\ &\leqslant (\omega_1 - \omega_n)[F(\phi^2(0), 1) - F(\phi^2(0), 0)]\frac{1-|\lambda|^{KT}}{1-|\lambda|^K}. \end{aligned} \quad (26)$$

Proof. See Appendix IV. □

Following the similar proof of Theorem 1, we obtain

Theorem 2. Given $\phi(1) \leqslant \omega_i(0) \leqslant \phi(0)$ $(1 \leqslant i \leqslant N)$, the second-highest policy is optimal if the following holds
$$|\lambda|^K \leqslant \frac{1-\phi(0)}{2-\phi(0)-\phi^2(0)}. \quad (27)$$

Corollary 2. Given $\phi(1) \leqslant \omega_i(0) \leqslant \phi(0)$ $(1 \leqslant i \leqslant N)$, the second-highest probing policy is optimal if the following holds
$$|\lambda| = p_{01} - p_{11} \leqslant \frac{1-p_{01}}{2-p_{01}-p_{11}}. \quad (28)$$

Proof. It is easy to check since $|\lambda|^K \leqslant |\lambda|$, $\phi(0) \leqslant p_{01}$, and $\phi^2(0) \geqslant p_{11}$. □

C. Negatively Correlated Channels $\lambda < 0$: even K

Following the similar proofs of Lemma 3, Lemma 4, and Theorem 2, we have the following theorem concerning the negatively correlated channels with even K.

Theorem 3. Given $\phi(0) \leqslant \omega_i(0) \leqslant \phi(1)$ $(1 \leqslant i \leqslant N)$, the second-highest probing policy is optimal if
$$\lambda^K \leqslant \frac{1-\phi(1)}{2-\phi(1)-\phi^2(0)}. \quad (29)$$

V. Extension

A. The Case of Probing Multiple Channels

For ease of presentation, we denote by S1 the scenario considered in the previous sections where the transmitter is allowed to probe and access one channel in each slot.

Suppose a scenario, denoted by S2, in which the transmitter is allowed to probe a number, M, of channels, and access one channel each time. In this case, the second-highest probing is extended to probe M channels from the second-highest one. This extended policy is also optimal under the same conditions in Theorems 1–3.

This can be explained simply. Compared to S1, the transmitter in S2 probes the other $M-1$ channels to obtain more information from the communication system such that the future decision can be made more precisely in S2 than S1. Considering the optimality of the second-highest probing policy for S1, the extended policy is certainly optimal for S2.

B. The Case of Probing Two Channels and Accessing One of the Two Probed Channels

Suppose a scenario, denoted by S3, in which the transmitter is allowed to probe two channels, denoted by $\mathcal{B}_t$, and access the better one $a(t)$ of $\mathcal{B}_t$. Letting $b(t) := \mathcal{B}_t - \{a(t)\}$, then the current reward is
$$\begin{aligned} &F(\omega_{a(t)}(t), \omega_{b(t)}(t)) \\ &= \sum_{k=0}^{K-1} \left[\omega_{b(t)}(t)\tau^k(1) + (1-\omega_{b(t)}(t))\tau^k(\omega_{a(t)}(t))\right], \end{aligned} \quad (30)$$
which is the same as (6) in S1.

In this case, the myopic policy is to probe the two best channels each time, which leads to the same $(\bar{b}_t, \bar{a}_t)$ as the second-highest probing policy in S1.

On the other hand, compared with S1, the transmitter in S3 has the same slot reward and policy, and moreover, probes more channels to obtain more information from the system. As a result, the myopic policy, probing the two best channels, is optimal for S3 under those conditions in Theorems 1–3.

VI. Numerical Simulation

In this section, we evaluate the performance of the second-highest probing policy under different scenarios by comparing with myopic probing policy (probing the first-best channel), random probing policy (randomly probing a channel in each slot), and the optimal probing policy which is obtained only for small T by brute force search in the whole policy space. Specifically, we compare the average accumulated reward of these policies with the function of the number of time slots.

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. Citation information: DOI 10.1109/TVT.2018.2870882, IEEE Transactions on Vehicular Technology

A. Positively correlated case ($\lambda \geqslant 0$)

First, we consider the scenarios with 4 channels and 10 slots, in which each slot only includes one mini-slot, i.e., $K = 1$.

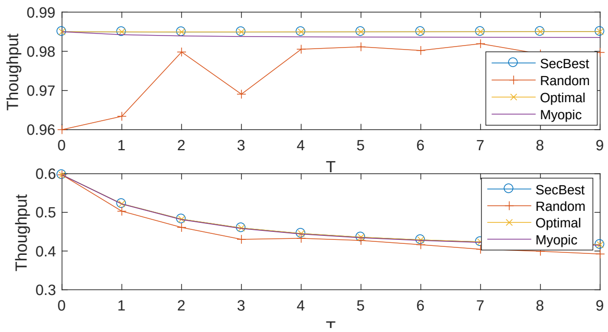


Fig. 2. Comparison of four polices. [Upper] $T = 10, K = 1, N = 4, p_{11} = 0.90, p_{01} = 0.54, \boldsymbol{w}(0) = [0.90,\ 0.85,\ 0.80,\ 0.60]$; [Lower] $T = 10, K = 1, N = 4, p_{11} = 0.38, p_{01} = 0.15, \boldsymbol{w}(0) = [0.38,\ 0.35,\ 0.30,\ 0.20]$.

The upper plot of Fig. 2 shows that the performance curve of the second-highest policy matches that of the optimal policy which verifies the theoretical result of the first part of Theorem 1, i.e., the second-highest policy is optimal under $\lambda^K \leqslant 4[\frac{1-\phi(0)}{(1-\phi(1))(1-\omega_0)} - 1]$ and $p_{01} \leqslant \omega_i(0) \leqslant p_{11}$. Meanwhile, we can observe that the second-highest policy is better than the myopic policy, although the performance difference is not obvious.

The lower plot of Fig. 2 also shows the second-highest policy is optimal which verifies the theoretical result in the second part of Theorem 1. Meanwhile, the second-highest outperforms both the myopic policy and the random policy.

Second, we evaluate the performance of these polices in the scenario with $K = 4$, $N = 4$ and $T = 10$. Similarly, we observe from Fig. 3 that the second-highest policy is optimal and outperforms the myopic policy and the random policy.

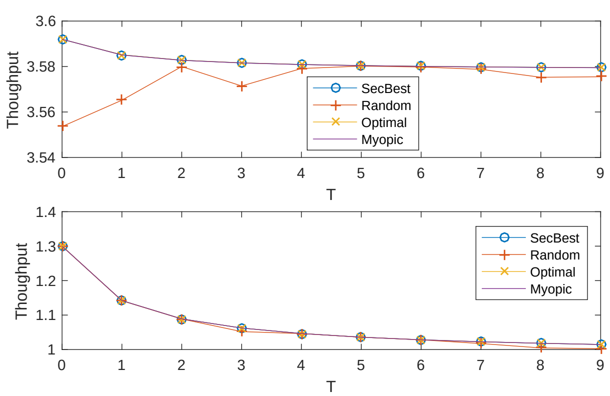


Fig. 3. Comparison of four polices. [Upper] $T = 10, K = 4, N = 4, p_{11} = 0.90, p_{01} = 0.54, \boldsymbol{w}(0) = [0.90,\ 0.85,\ 0.80,\ 0.60]$; [Lower] $T = 10, K = 4, N = 4, p_{11} = 0.38, p_{01} = 0.15, \boldsymbol{w}(0) = [0.38,\ 0.35,\ 0.30,\ 0.20]$.

B. Negatively correlated case ($\lambda < 0$)

For this case, we consider the cases with $K = 1$ and $K = 2$, which correspond to the odd K in Theorem 2 and the even K in Theorem 3, respectively. From Fig. 4, we observe that the second-highest policy is optimal under the corresponding sufficient conditions of Theorem 2 and Theorem 3.

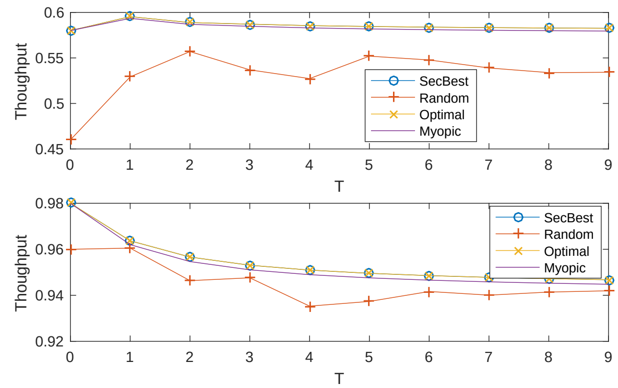


Fig. 4. Comparison of four polices. [Upper] $T = 10, K = 1, N = 4, p_{11} = 0.10, p_{01} = 0.40, \boldsymbol{w}(0) = [0.40,\ 0.30,\ 0.20,\ 0.10]$; [Lower] $T = 10, K = 2, N = 4, p_{11} = 0.10, p_{01} = 0.60, \boldsymbol{w}(0) = [0.60,\ 0.40,\ 0.30,\ 0.20]$.

VII. Conclusion

We consider the joint probing and accessing problem in an opportunistic communication system in which the transmitter probes one channel and accesses one channel for a fixed time interval based on the probed result. We formalize this problem in the language of RMAB and then propose the second-highest probing policy. First, by constructing a counterexample, we show that the second-highest probing policy is not alway optimal. Further, we present three sets of closed-form conditions under which the optimality of the second-highest policy is ensured. The sufficient conditions show that the optimality is tightly coupled with the initial belief vector and the non-trivial eigenvalue of the state transition matrix. From the viewpoint of exploitation *vs.* exploration, we extend the optimality to two related scenarios. Future work will include seeking the optimality of the policy for those channels with heterogeneous state transition matrices.

Appendix I
Proof of Lemma 2

The proving process is based on backward induction in three steps as follows:

1° In slot T, this lemma holds trivially, noticing $W_T(\boldsymbol{w}(T)) = F(\boldsymbol{w}(T))$.

For the first part of Lemma 2,

$$\begin{aligned} &W_T(\cdots,\omega_i,\omega_{i+1},\cdots) - W_T(\cdots,\omega_{i+1},\omega_i,\cdots) \\ &= F(\omega_1,\omega_2) - F(\omega_1,\omega_2) = 0 \end{aligned}$$

For the second part of Lemma 2,

$$\begin{aligned} &W_T(\omega_1,\omega_2,\omega_3,\cdots,\omega_n) - W_T(\omega_1,\omega_3,\omega_2,\cdots,\omega_n) \\ =&F(\omega_1,\omega_2) - F(\omega_1,\omega_3) \\ =&(\omega_2-\omega_3)(F(\omega_1,1) - F(\omega_1,0)) \\ \geqslant&(\omega_2-\omega_3)(F(\omega_1,1) - F(\omega_1,0)) \geqslant 0 \end{aligned}$$

For the third part of Lemma 2,

$$\begin{aligned}&W_t(\omega_1,\omega_2,\omega_3,\cdots,\omega_n)-W_t(\omega_2,\omega_1,\omega_3,\cdots,\omega_n)\\=&F(\omega_1,\omega_2)-F(\omega_2,\omega_1)=0\end{aligned}$$

For the fourth part of Lemma 2,

$$\begin{aligned}&W_T(\omega_1,\omega_2,\cdots,\omega_{n-1},\omega_n)-W_T(\omega_1,\omega_n,\omega_2,\cdots,\omega_{n-1})\\=&F(\omega_1,\omega_2)-F(\omega_1,\omega_n)\\=&(\omega_2-\omega_n)(F(\omega_1,1)-F(\omega_1,0))\\\leqslant&(F(\omega_1,1)-F(\omega_1,0))\end{aligned}$$

where the second equality is due to Lemma 1.

2° Assume at $T-1,\cdots,t+1$, Lemma 2 (the first to fourth parts are denoted by IH1–IH4, respectively) holds, we thus prove this Lemma also holds at slot t.

3° At slot t,

3°1 For the first part of Lemma 2:

$$\begin{aligned}&W_t(\cdots,\omega_i,\omega_{i+1},\cdots)-W_t(\cdots,\omega_{i+1},\omega_i,\cdots)\\&=(\omega_i-\omega_{i+1})W_t(\omega_1,\cdots,\omega_{i-1},1,0,\omega_{i+2},\cdots,\omega_n)\\&-(\omega_i-\omega_{i+1})W_t(\omega_1,\cdots,\omega_{i-1},0,1,\omega_{i+2},\cdots,\omega_n)\\&=(\omega_i-\omega_{i+1})\omega_2\Big\{\\&\quad W_{t+1}(\phi(1),\phi(\omega_1),\phi(\omega_3),\cdots,\phi(\omega_{i-1}),\\&\qquad\qquad\phi(1),\phi(0),\phi(\omega_{i+2}),\cdots,\phi(\omega_n))\\&-W_{t+1}(\phi(1),\phi(\omega_1),\phi(\omega_3),\cdots,\phi(\omega_{i-1}),\\&\qquad\qquad\phi(0),\phi(1),\phi(\omega_{i+2}),\cdots,\phi(\omega_n))\Big\}\\&+(\omega_i-\omega_{i+1})(1-\omega_2)\Big\{\\&\quad W_{t+1}(\phi(\omega_1),\phi(\omega_3),\cdots,\phi(\omega_{i-1}),\\&\qquad\qquad\phi(1),\phi(0),\phi(\omega_{i+2}),\cdots,\phi(\omega_n),\phi(0))\\&-W_{t+1}(\phi(\omega_1),\phi(\omega_3),\cdots,\phi(\omega_{i-1}),\\&\qquad\qquad\phi(0),\phi(1),\phi(\omega_{i+2}),\cdots,\phi(\omega_n),\phi(0))\Big\}\\&\overset{\text{IH1,2}}{\geqslant}0.\end{aligned}$$

3°2 For the second part of Lemma 2,

$$\begin{aligned}&W_t(\omega_1,\omega_2,\omega_3,\omega_4,\cdots,\omega_n)-W_t(\omega_1,\omega_3,\omega_2,\omega_4,\cdots,\omega_n)\\&=(\omega_2-\omega_3)\\&\times[W_t(\omega_1,1,0,\omega_4,\cdots,\omega_n)-W_t(\omega_1,0,1,\omega_4,\cdots,\omega_n)]\\&=(\omega_2-\omega_3)[F(\omega_1,1)-F(\omega_1,0)\\&+W_{t+1}(\phi(1),\phi(\omega_1),\phi(0),\phi(\omega_4),\cdots,\phi(\omega_n))\\&-W_{t+1}(\phi(\omega_1),\phi(1),\phi(\omega_4),\cdots,\phi(\omega_n),\phi(0))]\\&\overset{\text{IH2,3}}{\geqslant}(\omega_2-\omega_3)[F(\omega_1,1)-F(\omega_1,0)\\&+W_{t+1}(\phi(1),\phi(0),\phi(\omega_1),\phi(\omega_4),\cdots,\phi(\omega_n))\\&-W_{t+1}(\phi(1),\phi(\omega_1),\phi(\omega_4),\cdots,\phi(\omega_n),\phi(0))]\\&\overset{\text{IH4}}{\geqslant}(\omega_2-\omega_3)\\&\times[F(\omega_1,1)-F(\omega_1,0)-(F(\phi(1),1)-F(\phi(1),0))]\\&\overset{(a)}{\geqslant}0,\end{aligned}$$

where (a) is due to Proposition 3 with $\omega_1\leqslant\phi(1)$.

3°3 For the third part of Lemma 2,

$$\begin{aligned}0&\leqslant W_t(\omega_1,\omega_2,\omega_3,\cdots,\omega_n)-W_t(\omega_2,\omega_1,\omega_3,\cdots,\omega_n)\\&=(\omega_1-\omega_2)[W_t(1,0,\omega_3,\cdots,\omega_n)-W_t(0,1,\omega_3,\cdots,\omega_n)]\\&=(\omega_1-\omega_2)[F(1,0)-F(0,1)\\&+W_{t+1}(\phi(1),\phi(\omega_3),\cdots,\phi(\omega_n),\phi(0))\\&-W_{t+1}(\phi(1),\phi(0),\phi(\omega_3),\cdots,\phi(\omega_n))]\\&\overset{\text{IH4}}{\leqslant}(\omega_1-\omega_2)(F(\phi(1),1)-F(\phi(1),0)),\end{aligned}$$

3°4 For the fourth part of Lemma 2, according to Lemma 1, we have

$$\begin{aligned}&W_t(\omega_1,\omega_2,\cdots,\omega_{n-1},\omega_n)-W_t(\omega_1,\omega_n,\omega_2,\cdots,\omega_{n-1})\\&=\omega_2\omega_n\\&\times[W_t(\omega_1,1,\omega_3,\cdots,\omega_{n-1},1)-W_t(\omega_1,1,1,\omega_3,\cdots,\omega_{n-1})]\\&+\omega_2(1-\omega_n)\\&\times[W_t(\omega_1,1,\omega_3,\cdots,\omega_{n-1},0)-W_t(\omega_1,0,1,\omega_3,\cdots,\omega_{n-1})]\\&+(1-\omega_2)\omega_n\\&\times[W_t(\omega_1,0,\omega_3,\cdots,\omega_{n-1},1)-W_t(\omega_1,1,0,\omega_3,\cdots,\omega_{n-1})]\\&+(1-\omega_2)(1-\omega_n)\\&\times[W_t(\omega_1,0,\omega_3,\cdots,\omega_{n-1},0)-W_t(\omega_1,0,0,\omega_3,\cdots,\omega_{n-1})].\end{aligned}\tag{31}$$

In the following, we analyze (31) by four cases.

3°4°1 For the first term of (31), we have

$$\begin{aligned}&W_t(\omega_1,1,\omega_3,\cdots,\omega_{n-1},1)-W_t(\omega_1,1,1,\omega_3,\cdots,\omega_{n-1})\\&=F(\omega_1,1)-F(\omega_1,1)\\&+W_{t+1}(\phi(1),\phi(\omega_1),\phi(\omega_3),\cdots,\phi(\omega_{n-1}),\phi(1))\\&-W_{t+1}(\phi(1),\phi(\omega_1),\phi(1),\phi(\omega_3),\cdots,\phi(\omega_{n-1}))\\&\overset{\text{IH1,2}}{\leqslant}W_{t+1}(\phi(1),\phi(\omega_1),\phi(1),\phi(\omega_3),\cdots,\phi(\omega_{n-1}))\\&-W_{t+1}(\phi(1),\phi(\omega_1),\phi(1),\phi(\omega_3),\cdots,\phi(\omega_{n-1}))\\&=0.\end{aligned}$$

3°4°2 For the second term of (31), we have

$$\begin{aligned}&W_t((\omega_1,1,\omega_3,\cdots,\omega_{n-1},0)-W_t(\omega_1,0,1,\omega_3,\cdots,\omega_{n-1})\\&=F(\omega_1,1)-F(\omega_1,0)\\&+W_{t+1}(\phi(1),\phi(\omega_1),\phi(\omega_3),\cdots,\phi(\omega_{n-1}),\phi(0))\\&-W_{t+1}(\phi(\omega_1),\phi(1),\phi(\omega_3),\cdots,\phi(\omega_{n-1}),\phi(0))\\&\overset{\text{IH3}}{\leqslant}F(\omega_1,1)-F(\omega_1,0)\\&+(\phi(1)-\phi(\omega_1))(F(\phi(1),1)-F(\phi(1),0))\\&=F(\omega_1,1)-F(\omega_1,0)+\lambda^K(1-\phi(1))(F(\omega_1,1)-F(\omega_1,0))\end{aligned}$$

3°4°3 For the third term of (31), we have

$$\begin{aligned}&W_t((\omega_1,0,\omega_3,\cdots,\omega_{n-1},1)-W_t(\omega_1,1,0,\omega_3,\cdots,\omega_{n-1})\\&=F(\omega_1,0)-F(\omega_1,1)\\&+W_{t+1}(\phi(\omega_1),\phi(\omega_3),\cdots,\phi(\omega_{n-1}),\phi(1),\phi(0))\\&-W_{t+1}(\phi(1),\phi(\omega_1),\phi(0),\phi(\omega_3),\cdots,\phi(\omega_{n-1}))\\&\overset{\text{IH1-3}}{\leqslant}F(\omega_1,0)-F(\omega_1,1)\\&+W_{t+1}(\phi(1),\phi(\omega_1),\phi(\omega_3),\cdots,\phi(\omega_{n-1}),\phi(0))\end{aligned}$$

$$
\begin{aligned}
&- W_{t+1}(\phi(1), \phi(\omega_1), \phi(0), \phi(\omega_3), \cdots, \phi(\omega_{n-1}))\\
&\overset{\text{IH2}}{\leqslant} F(\omega_1, 0) - F(\omega_1, 1)\\
&+ W_{t+1}(\phi(1), \phi(\omega_1), \phi(\omega_3), \cdots, \phi(\omega_{n-1}), \phi(0))\\
&- W_{t+1}(\phi(1), \phi(0), \phi(\omega_1), \phi(\omega_3), \cdots, \phi(\omega_{n-1}))\\
&\overset{\text{IH4}}{\leqslant} F(\omega_1, 0) - F(\omega_1, 1) + F(\phi(1), 1) - F(\phi(1), 0)\\
&\leqslant 0.
\end{aligned}
$$

3°4°4 For the fourth term of (31), we have

$$
\begin{aligned}
&W_t(\omega_1, 0, \omega_3, \cdots, \omega_{n-1}, 0) - W_t(\omega_1, 0, 0, \omega_3, \cdots, \omega_{n-1})\\
&= W_{t+1}(\phi(\omega_1), \phi(\omega_3), \cdots, \phi(\omega_{n-1}), \phi(0), \phi(0))\\
&- W_{t+1}(\phi(\omega_1), \phi(0), \phi(\omega_3), \cdots, \phi(\omega_{n-1}), \phi(0))\\
&\overset{\text{IH4}}{\leqslant} F(\phi(\omega_1), 1) - F(\phi(\omega_1), 0)\\
&\overset{(g)}{\leqslant} F(\phi(\min\{\omega_0, \omega_1\}), 1) - F(\phi(\min\{\omega_0, \omega_1\}), 0)
\end{aligned}
$$

where (g) is due to Proposition 3.

Combining the results of $3°4°1 - 3°4°4$, we have

$$
\begin{aligned}
&W_t(\omega_1, \omega_2, \cdots, \omega_{n-1}, \omega_n) - W_t(\omega_1, \omega_n, \omega_2, \cdots, \omega_{n-1})\\
&\leqslant (F(\omega_1, \omega_2) - F(\omega_1, \omega_N))\\
&+ \omega_2(1 - \phi(1))(1 - \omega_n)\lambda^K[F(\omega_1, 1) - F(\omega_1, 0)]\\
&+ (1 - \omega_2)\omega_n[F(\phi(1), 1) - F(\phi(1), 0)]\\
&+ (1 - \omega_2)(1 - \omega_n)[F(\phi(\omega_1), 1) - F(\phi(\omega_1), 0)]\\
&\leqslant (1 - \omega_n)(1 + \omega_2(1 - \omega_1)\lambda^K)\\
&\times [F(\min\{\omega_0, \omega_1\}, 1) - F(\min\{\omega_0, \omega_1\}, 0)]\\
&\overset{(h)}{\leqslant} (1 - \omega_n)(1 + \frac{1}{4}\lambda^K)\\
&\times [F(\min\{\omega_0, \omega_1\}, 1) - F(\min\{\omega_0, \omega_1\}, 0)]\\
&\overset{(i)}{\leqslant} F(\omega_1, 1) - F(\omega_1, 0),
\end{aligned}
$$

where (h) is due to $\omega_2 + 1 - \omega_1 \leqslant 1$, and (i) is due to $\lambda^K \leqslant 4\big[\frac{1-\phi(0)}{(1-\phi(1))(1-\omega_0)} - 1\big]$.

To this end, we complete the proof of Lemma 2.

Appendix II
Proof of Lemma 3

The proof of Lemma 3 basically follows that of Lemma 2.

1° In slot T, this lemma holds trivially, noticing $W_T(\boldsymbol{w}(T) = F(\boldsymbol{w}(T)))$.

2° Assume at $T-1, \cdots, t+1$, Lemma 3 (the first to fourth parts are denoted by IH1–IH4, respectively) holds, we thus prove this Lemma also holds at slot t.

3° At slot t,

3°1 For the first part of Lemma 3:

$$
\begin{aligned}
&W_t(\cdots, \omega_i, \omega_{i+1}, \cdots) - W_t(\cdots, \omega_{i+1}, \omega_i, \cdots)\\
&= (\omega_i - \omega_{i+1})W_t(\omega_1, \cdots, \omega_{i-1}, 1, 0, \omega_{i+2}, \cdots, \omega_n)\\
&- (\omega_i - \omega_{i+1})W_t(\omega_1, \cdots, \omega_{i-1}, 0, 1, \omega_{i+2}, \cdots, \omega_n)\\
&= (\omega_i - \omega_{i+1})\omega_2\Big\{\\
&\quad W_{t+1}(\phi(1), \phi(\omega_1), \phi(\omega_3), \cdots, \phi(\omega_{i-1}),\\
&\qquad\qquad \phi(1), \phi(0), \phi(\omega_{i+2}), \cdots, \phi(\omega_n))\\
&- W_{t+1}(\phi(1), \phi(\omega_1), \phi(\omega_3), \cdots, \phi(\omega_{i-1}),\\
&\qquad\qquad \phi(0), \phi(1), \phi(\omega_{i+2}), \cdots, \phi(\omega_n))\Big\}\\
&+ (\omega_i - \omega_{i+1})(1 - \omega_2)\Big\{\\
&\quad W_{t+1}(\phi(\omega_1), \phi(\omega_3), \cdots, \phi(\omega_{i-1}),\\
&\qquad\qquad \phi(1), \phi(0), \phi(\omega_{i+2}), \cdots, \phi(\omega_n), \phi(0))\\
&- W_{t+1}(\phi(\omega_1), \phi(\omega_3), \cdots, \phi(\omega_{i-1}),\\
&\qquad\qquad \phi(0), \phi(1), \phi(\omega_{i+2}), \cdots, \phi(\omega_n), \phi(0))\Big\}\\
&\overset{\text{IH1,2}}{\geqslant} 0.
\end{aligned}
$$

3°2 For the second part of Lemma 3,

$$
\begin{aligned}
&W_t(\omega_1, \omega_2, \omega_3, \omega_4, \cdots, \omega_n) - W_t(\omega_1, \omega_3, \omega_2, \omega_4, \cdots, \omega_n)\\
&= (\omega_2 - \omega_3)\\
&\times [W_t(\omega_1, 1, 0, \omega_4, \cdots, \omega_n) - W_t(\omega_1, 0, 1, \omega_4, \cdots, \omega_n)]\\
&= (\omega_2 - \omega_3)[F(\omega_1, 1) - F(\omega_1, 0)\\
&+ W_{t+1}(\phi(1), \phi(\omega_1), \phi(0), \phi(\omega_4), \cdots, \phi(\omega_n))\\
&- W_{t+1}(\phi(\omega_1), \phi(1), \phi(\omega_4), \cdots, \phi(\omega_n), \phi(0))]\\
&\overset{\text{IH2,3}}{\geqslant} (\omega_2 - \omega_3)[F(\omega_1, 1) - F(\omega_1, 0)\\
&+ W_{t+1}(\phi(0), \phi(\omega_1), \phi(\omega_4), \cdots, \phi(\omega_n), \phi(1))\\
&- W_{t+1}(\phi(1), \phi(\omega_1), \phi(\omega_4), \cdots, \phi(\omega_n), \phi(0))]\\
&\overset{\text{IH4}}{\geqslant} (\omega_2 - \omega_3)\Big[F(\omega_1, 1) - F(\omega_1, 0)\\
&- (\phi(1) - \phi(0))\frac{F(\phi(0), 1) - F(\phi(0), 0)}{1 - \lambda^K}\Big]\\
&= (\omega_2 - \omega_3)\\
&\times \Big[F(\omega_1, 1) - F(\omega_1, 0) - \lambda^K\frac{F(\phi(0), 1) - F(\phi(0), 0)}{1 - \lambda^K}\Big]\\
&\overset{(a)}{\geqslant} 0,
\end{aligned}
$$

where (a) is due to $\lambda^K \leqslant \frac{1-\phi(1)}{2-\phi(1)-\phi(0)}$.

3°3 For the third part of Lemma 3,

$$
\begin{aligned}
&W_t(\omega_1, \omega_2, \omega_3, \cdots, \omega_n) - W_t(\omega_2, \omega_1, \omega_3, \cdots, \omega_n)\\
&= (\omega_1 - \omega_2)[W_t(1, 0, \omega_3, \cdots, \omega_n) - W_t(0, 1, \omega_3, \cdots, \omega_n)]\\
&= (\omega_1 - \omega_2)[F(1, 0) - F(0, 1)\\
&+ W_{t+1}(\phi(1), \phi(\omega_3), \cdots, \phi(\omega_n), \phi(0))\\
&- W_{t+1}(\phi(1), \phi(0), \phi(\omega_3), \cdots, \phi(\omega_n))]\\
&\overset{\text{IH1,2}}{\geqslant} 0,
\end{aligned}
$$

3°4 For the fourth part of Lemma 3,

$$
\begin{aligned}
&W_t(\omega_1, \omega_2, \cdots, \omega_{n-1}, \omega_n) - W_t(\omega_n, \omega_2, \cdots, \omega_{n-1}, \omega_1)\\
&= (\omega_1 - \omega_n)\\
&\times [W_t(1, \omega_2, \cdots, \omega_{n-1}, 0) - W_t(0, \omega_2, \cdots, \omega_{n-1}, 1)]\\
&= (\omega_1 - \omega_n)\Big[F(\omega_2, 1) - F(\omega_2, 0)\\
&+ \omega_2 W_{t+1}(\phi(1), \phi(1), \phi(\omega_3), \cdots, \phi(\omega_{n-1}), \phi(0))\\
&- \omega_2 W_{t+1}(\phi(1), \phi(0), \phi(\omega_3), \cdots, \phi(\omega_{n-1}), \phi(1))\\
&+ (1 - \omega_2)W_{t+1}(\phi(1), \phi(\omega_3), \cdots, \phi(\omega_{n-1}), \phi(0), \phi(0))\\
&- (1 - \omega_2)W_{t+1}(\phi(0), \phi(\omega_3), \cdots, \phi(\omega_{n-1}), \phi(1), \phi(0))\Big]
\end{aligned}
$$

$$\overset{\text{IH1,3}}{\leqslant} (\omega_1-\omega_n)\Big[F(\omega_2,1)-F(\omega_2,0)$$
$$+\omega_2 W_{t+1}(\phi(1),\phi(1),\phi(\omega_3),\cdots,\phi(\omega_{n-1}),\phi(0))$$
$$-\omega_2 W_{t+1}(\phi(0),\phi(1),\phi(\omega_3),\cdots,\phi(\omega_{n-1}),\phi(1))$$
$$+(1-\omega_2)W_{t+1}(\phi(1),\phi(\omega_3),\cdots,\phi(\omega_{n-1}),\phi(0),\phi(0))$$
$$-(1-\omega_2)W_{t+1}(\phi(0),\phi(\omega_3),\cdots,\phi(\omega_{n-1}),\phi(0),\phi(1))\Big]$$
$$\overset{\text{IH4}}{\leqslant} (\omega_1-\omega_n)\Big[F(\omega_2,1)-F(\omega_2,0)$$
$$+\omega_2(\phi(1)-\phi(0))(F(\phi(0),1)-F(\phi(0),0))\frac{1}{1-\lambda^K}$$
$$+(1-\omega_2)(\phi(1)-\phi(0))\frac{F(\phi(0),1)-F(\phi(0),0)}{1-\lambda^K}\Big]$$
$$\overset{(a)}{\leqslant} (\omega_1-\omega_n)\Big[(F(\phi(0),1)-F(\phi(0),0))$$
$$-\omega_2(F(\phi(0),1)-F(\phi(0),0))\frac{\lambda^K}{1-\lambda^K}$$
$$-(1-\omega_2)(F(\phi(0),1)-F(\phi(0),0))\frac{\lambda^K}{1-\lambda^K}\Big]$$
$$=(\omega_1-\omega_n)(F(\phi(0),1)-F(\phi(0),0))(1+\frac{\lambda^K}{1-\lambda^K})$$
$$=(\omega_1-\omega_n)(F(\phi(0),1)-F(\phi(0),0))\frac{1}{1-\lambda^K},$$

where (a) is due to Proposition 3 with $\omega_2 \geqslant \phi(0)$.

Appendix III
Proof of Corollary 1

Based on Theorem 1, we have

Case 1.
$$\lambda^K \leqslant 4\left[\frac{1-\phi(0)}{(1-\phi(1))(1-\omega_0)}-1\right]$$
$$\overset{(a)}{\Leftarrow} \lambda^K \leqslant \frac{8p_{01}-4p_{11}}{1-p_{01}}$$
$$\Leftarrow \lambda \leqslant \frac{8p_{01}-4p_{11}}{1-p_{01}}$$
$$\Leftrightarrow p_{11}-p_{01} \leqslant \frac{8p_{01}-4p_{11}}{1-p_{01}}$$
$$\Leftrightarrow p_{11} \leqslant \frac{p_{01}(9-p_{01})}{5-p_{01}}$$
$$\Leftarrow 1 \leqslant \frac{p_{01}(9-p_{01})}{5-p_{01}}$$
$$\Leftrightarrow p_{01} \geqslant 5-2\sqrt{5},$$

where (a) is due to $p_{01} \leqslant \phi(0) \leqslant \phi(1) \leqslant p_{11}$ and $\omega_0 = \frac{p_{01}}{1-p_{11}+p_{01}}$.

Case 2.
$$\lambda^K \leqslant \frac{1-\phi(1)}{2-\phi(1)-\phi(0)}$$
$$\overset{(b)}{\Leftarrow} \lambda^K \leqslant \frac{1-p_{11}}{2-p_{11}-p_{01}}$$
$$\Leftarrow p_{11}-p_{01} \leqslant \frac{1-p_{11}}{2-p_{11}-p_{01}}$$
$$\Leftrightarrow (1-p_{11})(2-p_{11}) \geqslant (1-p_{01})^2$$
$$\Leftarrow (1-p_{11})(2-p_{11}) \geqslant 1$$
$$\Leftrightarrow p_{11} \leqslant \frac{3-\sqrt{5}}{2},$$

where (b) is due to $p_{01} \leqslant \phi(0) \leqslant \phi(1) \leqslant p_{11}$.

Combining the two cases, we finish the proof.

Appendix IV
Proof of Lemma 4

The proving process is based on backward induction in three steps as follows:

1° In slot T, these Lemmas hold trivially, noticing $W_T(\boldsymbol{w}(T) = F(\boldsymbol{w}(T)))$.

For the first part of Lemma 4:

$$W_T(\omega_1,\cdots,\omega_i,\omega_{i+1},\cdots,\omega_n)-W_T(\omega_1,\cdots,\omega_{i+1},\omega_i,\cdots,\omega_n)$$
$$=F(\omega_1,\omega_2)-F(\omega_1,\omega_2)=0$$

For the second part of Lemma 4:

$$W_T(\omega_1,\omega_2,\omega_3,\cdots,\omega_n)-W_T(\omega_1,\omega_3,\omega_2,\cdots,\omega_n)$$
$$=F(\omega_1,\omega_2)-F(\omega_1,\omega_3)$$
$$=(\omega_2-\omega_3)(F(\omega_1,1)-F(\omega_1,0))$$
$$\geqslant(\omega_2-\omega_3)(F(\omega_1,1)-F(\omega_1,0))\geqslant 0$$

For the third part of Lemma 4,

$$W_t(\omega_1,\omega_2,\omega_3,\cdots,\omega_n)-W_t(\omega_2,\omega_1,\omega_3,\cdots,\omega_n)$$
$$=F(\omega_1,\omega_2)-F(\omega_2,\omega_1)=0$$

For the fourth part of Lemma 4:

$$W_T(\omega_1,\omega_2,\cdots,\omega_{n-1},\omega_n)-W_T(\omega_1,\omega_n,\omega_2,\cdots,\omega_{n-1})$$
$$=F(\omega_1,\omega_2)-F(\omega_1,\omega_n)$$
$$=(\omega_2-\omega_n)(F(\omega_1,1)-F(\omega_1,0))$$
$$\leqslant F(\omega_1,1)-F(\omega_1,0).$$

2° Assume at $T-1,\cdots,t+1$, Lemma 4 (the first to fourth parts are denoted by IH1–IH4, respectively) holds, we thus prove this Lemma also holds at slot t.

3° At slot t,

3°1 For the first part of Lemma 4:

$$W_t(\cdots,\omega_i,\omega_{i+1},\cdots)-W_t(\cdots,\omega_{i+1},\omega_i,\cdots)$$
$$=(\omega_i-\omega_{i+1})W_t(\omega_1,\cdots,\omega_{i-1},1,0,\omega_{i+2},\cdots,\omega_n)$$
$$-(\omega_i-\omega_{i+1})W_t(\omega_1,\cdots,\omega_{i-1},0,1,\omega_{i+2},\cdots,\omega_n)$$
$$=(\omega_i-\omega_{i+1})\omega_2\Big\{$$
$$W_{t+1}(\phi(\omega_n),\cdots,\phi(\omega_{i+2}),\phi(0),\phi(1),\phi(\omega_{i-1}),\cdots,\phi(\omega_3),\phi(\omega_1),\phi(1))$$
$$-W_{t+1}(\phi(\omega_n),\cdots,\phi(\omega_{i+2}),\phi(1),\phi(0),\phi(\omega_{i-1}),\cdots,\phi(\omega_3),\phi(\omega_1),\phi(1))\Big\}$$
$$+(\omega_i-\omega_{i+1})(1-\omega_2)\Big\{$$
$$W_{t+1}(\phi(0),\phi(\omega_n),\cdots,\phi(\omega_{i+2}),\phi(0),\phi(1),\phi(\omega_{i-1}),\cdots,\phi(\omega_3),\phi(\omega_1))$$
$$-W_{t+1}(\phi(0),\phi(\omega_n),\cdots,\phi(\omega_{i+2}),\phi(1),\phi(0),\phi(\omega_{i-1}),\cdots,\phi(\omega_3),\phi(\omega_1))\Big\}$$
$$\overset{\text{IH1-3}}{\geqslant} 0.$$

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. Citation information: DOI 10.1109/TVT.2018.2870882, IEEE Transactions on Vehicular Technology

3°2 For the second part of Lemma 4,

$$\begin{aligned}
&W_t(\omega_1,\omega_2,\omega_3,\omega_4,\cdots,\omega_n)-W_t(\omega_1,\omega_3,\omega_2,\omega_4,\cdots,\omega_n)\\
&=(\omega_2-\omega_3)\\
&\times[W_t(\omega_1,1,0,\omega_4,\cdots,\omega_n)-W_t(\omega_1,0,1,\omega_4,\cdots,\omega_n)]\\
&=(\omega_2-\omega_3)[F(\omega_1,1)-F(\omega_1,0)\\
&+W_{t+1}(\phi(\omega_n),\cdots,\phi(\omega_4),\phi(0),\phi(\omega_1),\phi(1))\\
&-W_{t+1}(\phi(0),\phi(\omega_n),\cdots,\phi(\omega_4),\phi(1),\phi(\omega_1))]\\
&\overset{\text{IH1-3}}{\geqslant}(\omega_2-\omega_3)[F(\omega_1,1)-F(\omega_1,0)\\
&+W_{t+1}(\phi(1),\phi(\omega_n),\cdots,\phi(\omega_4),\phi(\omega_1),\phi(0))\\
&-W_{t+1}(\phi(0),\phi(\omega_n),\cdots,\phi(\omega_4),\phi(\omega_1),\phi(1))]\\
&\overset{\text{IH4}}{\geqslant}(\omega_2-\omega_3)\Big[F(\omega_1,1)-F(\omega_1,0)\\
&-(\phi(0)-\phi(1))[F(\phi^2(0),1)-F(\phi^2(0),0)]\frac{1}{1-|\lambda|^K}\Big]\\
&\geqslant(\omega_2-\omega_3)\Big[F(\omega_1,1)-F(\omega_1,0)\\
&-\frac{(F(\phi^2(0),1)-F(\phi^2(0),0))|\lambda|^K}{1-|\lambda|^K}\Big]\\
&\overset{(a)}{\geqslant}0,
\end{aligned}$$

where (a) is due to $|\lambda|^K\leqslant\frac{1-\phi(0)}{2-\phi(0)-\phi^2(0)}$.

3°3 For the third part of Lemma 4,

$$\begin{aligned}
&W_t(\omega_1,\omega_2,\omega_3,\cdots,\omega_n)-W_t(\omega_2,\omega_1,\omega_3,\cdots,\omega_n)\\
&=(\omega_1-\omega_2)[W_t(1,0,\omega_3,\cdots,\omega_n)-W_t(0,1,\omega_3,\cdots,\omega_n)]\\
&=(\omega_1-\omega_2)[F(1,0)-F(0,1)\\
&+W_{t+1}(\phi(0),\phi(\omega_n),\cdots,\phi(\omega_3),\phi(1))\\
&-W_{t+1}(\phi(\omega_n),\cdots,\phi(\omega_3),\phi(0),\phi(1))]\\
&\overset{\text{IH1-3}}{\geqslant}0
\end{aligned}$$

3°4 For the fourth part of Lemma 4,

$$\begin{aligned}
&W_t(\omega_1,\phi(\omega_2),\cdots,\omega_{n-1},\omega_n)-W_t(\omega_n,\phi(\omega_2),\cdots,\omega_{n-1},\omega_1)\\
&=(\omega_1-\omega_n)\\
&\times[W_t(1,\phi(\omega_2),\cdots,\omega_{n-1},0)-W_t(0,\phi(\omega_2),\cdots,\omega_{n-1},1)]\\
&=(\omega_1-\omega_n)\Big[F(\phi(\omega_2),1)-F(\phi(\omega_2),0)\\
&+\phi(\omega_2)W_{t+1}(\phi(0),\phi(\omega_{n-1}),\cdots,\phi(\omega_3),\phi(1),\phi(1))\\
&-\phi(\omega_2)W_{t+1}(\phi(1),\phi(\omega_{n-1}),\cdots,\phi(\omega_3),\phi(0),\phi(1))\\
&+(1-\phi(\omega_2))W_{t+1}(\phi(0),\phi(0),\phi(\omega_{n-1}),\cdots,\phi(\omega_3),\phi(1))\\
&-(1-\phi(\omega_2))W_{t+1}(\phi(0),\phi(1),\phi(\omega_{n-1}),\cdots,\phi(\omega_3),\phi(0))\Big]\\
&\overset{\text{IH1-3}}{\leqslant}(\omega_1-\omega_n)\Big[F(\phi(\omega_2),1)-F(\phi(\omega_2),0)\\
&+\phi(\omega_2)W_{t+1}(\phi(0),\phi(\omega_{n-1}),\cdots,\phi(\omega_3),\phi(1),\phi(1))\\
&-\phi(\omega_2)W_{t+1}(\phi(1),\phi(\omega_{n-1}),\cdots,\phi(\omega_3),\phi(1),\phi(0))\\
&+(1-\phi(\omega_2))W_{t+1}(\phi(0),\phi(0),\phi(\omega_{n-1}),\cdots,\phi(\omega_3),\phi(1))\\
&-(1-\phi(\omega_2))W_{t+1}(\phi(1),\phi(0),\phi(\omega_{n-1}),\cdots,\phi(\omega_3),\phi(0))\Big]\\
&\overset{\text{IH4}}{\leqslant}(\omega_1-\omega_n)\Big[F(\phi(\omega_2),1)-F(\phi(\omega_2),0)\\
&+\phi(\omega_2)(\phi(0)-\phi(1))\frac{F(\phi^2(0),1)-F(\phi^2(0),0)}{1-|\lambda|^K}\\
&+(1-\phi(\omega_2))(\phi(0)-\phi(1))\frac{F(\phi^2(0),1)-F(\phi^2(0),0)}{1-|\lambda|^K}\Big]\\
&\overset{(a)}{\leqslant}(\omega_1-\omega_n)\Big[F(\phi^2(0),1)-F(\phi^2(0),0)\\
&-\phi(\omega_2)(F(\phi^2(0),1)-F(\phi^2(0),0))\frac{\lambda^K}{1-|\lambda|^K}\\
&-(1-\phi(\omega_2))(F(\phi^2(0),1)-F(\phi^2(0),0))\frac{\lambda^K}{1-|\lambda|^K}\Big]\\
&=(\omega_1-\omega_n)[F(\phi^2(0),1)-F(\phi^2(0),0)](1+\frac{|\lambda|^K}{1-|\lambda|^K})\\
&=(\omega_1-\omega_n)[F(\phi^2(0),1)-F(\phi^2(0),0)]\frac{1}{1-|\lambda|^K},
\end{aligned}$$

where (a) is due to Proposition 3 with $\phi(\omega_2)\geqslant\phi^2(0)$.

To this end, we complete the proof of Lemma 4.

REFERENCES

[1] Q. Zhao, L. Tong, A. Swami, and Y. Chen. Decentralized Cognitive MAC for Opportunistic Spectrum Access in Adhoc Networks: A POMDP Framework. *IEEE J. Sel. Areas Commun.*, vol. 25, no. 3, pp. 589–600, 2007.

[2] P. Whittle. Restless Bandits: Activity Allocation in A Changing World. *J. Appl. Probab.*, vol. 25A, pp. 287–298, 1988.

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

[4] 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, 2008.

[5] S. Ahmad, M. Liu, T. Javidi, Q. Zhao, and B. Krishnamachari. Optimality of Myopic Sensing in Multichannel Opportunistic Access. *IEEE Trans. Inf. Theory*, vol. 55, no. 9, pp. 4040–4050, 2009.

[6] S. Ahmad, M. Liu. Multi-channel Opportunistic Access: a Case of Restless Bandits with Multiple Plays. In *Allerton Conference*, Oct. 2009.

[7] 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 Commun. Lett.*, vol. 1, no. 5, pp. 452–455, 2012.

[8] Y. Liu, M. Liu, and S. Ahmad. Sufficient Conditions on the Optimality of Myopic Sensing in Opportunistic Channel Access: A Unifying Framework. *IEEE Trans. Inf. Theory*, vol. 60, no. 8, pp. 4922–4940, 2014.

[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, and Q. Liu, "Opportunistic Spectrum Access by Exploiting Primary User Feedbacks in Underlay Cognitive Radio Systems: An Optimality Analysis," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 5, pp. 869–882, Oct. 2013.

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. Citation information: DOI 10.1109/TVT.2018.2870882, IEEE Transactions on Vehicular Technology

[11] K. Wang, Q. Liu, and F. C. Lau, "Multichannel Opportunistic Access by Overhearing Primary ARQ Messages," *IEEE Trans. Veh. Technol.*, vol. 62, no. 7, pp. 3486–3492, Sep. 2013.

[12] K. Wang, L. Chen, and Q. Liu, "On Optimality of Myopic Policy for Opportunistic Access With Nonidentical Channels and Imperfect Sensing," *IEEE Trans. Veh. Technol.*, vol. 63, no. 5, pp. 2478–2483, May 2014.

[13] Y. Ouyang and D. Teneketzis, "On the optimality of myopic sensing in multi-state channels," *IEEE Trans. Inf. Theory*, vol. 60, no.1, pp. 681–696, Jan. 2014.

[14] K. Wang, L. Chen, J. Yu, and D. Zhang, "Optimality of Myopic Policy for Multistate Channel Access," *IEEE Commun. Lett.*, vol. 20, no. 2, pp. 300–303, Feb 2016.

[15] K. Wang, L. Chen, and J. Yu, "On Optimality of Myopic Policy in Multi-channel Opportunistic Access," *IEEE Trans. Commun.*, vol. 65, no. 2, pp. 677–690, 2017.

[16] M. Johnston, I. Keslassy, and E. Modiano. Channel Probing in Opportunistic Communication Systems. *IEEE Trans. Inf. Theory*, vol. 63, no. 11, pp. 7535–7552, 2017.

[17] K. Wang, Q. Liu, Q. Fan, and Q. Ai. Optimally Probing Channel in Opportunistic Communication Access. *IEEE Commun. Lett.*, vol. 22, no. 7, pp. 1246–1249, 2018.

[18] R. G. Stephen, C. R. Murthy, M. Coupechoux. A Markov Decision Theoretic Approach to Pilot Allocation and Receive Antenna Selection. *IEEE Trans. Wireless Commu.*, vol. 12, no. 8, pp. 3813–3823, 2013.

[19] X. Jiang, X. Wang, H. Xi. Finding Optimal Polices for Wideband Spectrum Sensing Based on Constrained POMDP Framework. *IEEE Trans. Wireless Commu.*, vol. 16, no. 8, pp. 5311–5324, 2017.

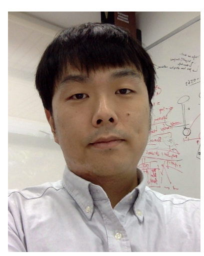
[20] T. Xiong, Y. Yao, Y. Ren, Z. Li. Multiband Spectrum Sensing in Cognitive Radio Networks With Secondary User Hardware Limitation: Random and Adaptive Spectrum Sensing Strategies. *IEEE Trans. Wireless Commu.*, vol. 17, no. 5, pp. 3018–3029, 2018.

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. From Feb. 2013 to Aug. 2013, he was a postdoc with the HongKong Polytechnic University. In 2013, he joined the School of Information Engineering at the Wuhan University of Technology, where he is currently an associate professor. From Dec. 2015, he has been a visiting scholar in the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA. USA. His research interests are stochastic optimization, operation research, scheduling, wireless network communications, and embedded operating system.

Lin Chen received 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.

Jihong Yu received the B.E. degree in communication engineering and M.E degree in communication and information systems from Chongqing University of Posts and Telecommunications, Chongqing, China, in 2010 and 2013, respectively, and the Ph.D. degree in computer science at the University of Paris-Sud, Orsay, France, in 2016. He was a research fellow from 2017 in the School of Computing Science, Simon Fraser University, British Columbia, Canada. Currently, he is an associate professor in Bejing Institute of Technology. His research interests include RFID technologies, wireless communications, and Internet of things.

Qingju Fan received Ph.D in the Department of Control Science and Engineering of Huazhong University of Technology in 2011. From Sep. 2014 to Sep. 2015, she has been a visiting scholar in the Department of Physics of Boston University. She is currently an associate professor in Wuhan University of Technology. Her research interests is fractal time series.



Yang Zhang (M'11) is currently an Associate Professor in the School of Computer Science and Technology, Wuhan University of Technology. He received the Ph.D. degree from the School of Computer Science and Engineering, Nanyang Technological University (NTU), Singapore, in 2015. He obtained his B.Eng. and M.Eng. degrees from Beihang University (BUAA), Beijing, China, in 2008 and 2011, respectively. His research interests are market-oriented modeling for wireless resource allocations, and optimal decision making problems.

Wei Chen received his Ph.D. degrees from school of electronic information and communications, Huazhong University of Science and Technology in 2005, China. He is now a professor and doctoral supervisor at School of Automation in Wuhan University of Technology, China. His major research interests include mobile communication signal processing and application, cooperative communication and wireless mesh network.

Pan Zhou (S'07–M'14) is currently an associate professor with School of Electronic Information and Communications, Huazhong University of Science and Technology (HUST), Wuhan, Hubei, China. He received his Ph.D. in the School of Electrical and Computer Engineering at the Georgia Institute of Technology (Georgia Tech), Atlanta, USA in 2011. He received his B.S. degree in the Advanced Class of HUST, and a M.S. degree in the Department of Electronics and Information Engineering from HUST, in 2006 and 2008, respectively. He held honorary degree in his bachelor and merit research award of HUST in his master study. He was a senior technical member at Oracle Inc, America during 2011 to 2013, and worked on Hadoop and distributed storage system for big data analytics at Oralce Cloud Platform. His current research interests include big data analytics and machine learning, security and privacy, and information networks.

Yi Zhong received his BEEE and MSEE both in the School of Information Engineering from Wuhan University of Technology, in 1999 and 2002 respectively. And he got his Ph.D. of Information and Communication Engineering in 2007. Yi has published numerous articles and papers in domestic and international journals, and has attended in many conferences on topics ranging from Embedded Control Systems to System Fault Diagnosis