IET Communications

Research Article

IET Journals
The Institution of Engineering and Technology

Myopic policy for opportunistic access in cognitive radio networks by exploiting primary user feedbacks

ISSN 1751-8628
Received on 25th February 2014
Accepted on 23rd November 2014
doi: 10.1049/iet-com.2014.1026
www.ietdl.org

Kehao Wang[1] ✉, *Quan Liu*[1], *Fangmin Li*[1], *Lin Chen*[2], *Xiaolin Ma*[1]

[1]*Key Laboratory of Fiber Optic Sensing Technology and Information Processing, Ministry of Education, Wuhan University of Technology, Wuhan, 430070, People's Republic of China*
[2]*LRI, Department of Computer Science, University of Paris-Sud XI, Orsay, 91405, France*
✉ *E-mail: kehao.wang@whut.edu.cn*

Abstract: The authors consider a cognitive radio network overlaying on top of a legacy primary network in which a secondary user is allowed to access primary channel by overhearing feedback signals over the primary channels. Each channel is assumed to be a two state Makovian process. Aiming at maximising the expected accumulated discounted network throughput, the considered sequential decision-making problem can be cast into a restless multi-armed bandit (RMAB) problem which is well-known to be PSPACE-hard, and thus a natural alternative approach is to seek a simple myopic policy. This study presents a theoretical study on the optimality of the proposed myopic policy for the special RMAB problem by considering four different cases: negatively correlated homogeneous channels, heterogeneous channels, positively correlated heterogeneous channels and negatively correlated heterogeneous channels. More specifically, the authors establish the closed-form conditions to guarantee the optimality of the myopic policy for the four cases, respectively, which, combined with the case of positively correlated homogeneous channels, constitute a complete paradigm for the optimality of the myopic policy.

1 Introduction

Cognitive radio (CR) has been viewed as a promising approach to achieve a more spectrally efficient communication where secondary users (SUs) can opportunistically utilise the spectrum originally allocated to primary users (PUs) if limiting the interference to PUs under a tolerable level [1–8].

In this paper, we consider a CR communication system composed of a set of PUs and an SU. Each PU, with one receiving and transmitting antenna, occupies one of primary channels and is assumed to transmit packets continuously [We adopt a worst-case assumption that PUs transmit all the time which is commonly used in analysing underlay CR systems. In other words, our analysis does not rely on detection and exploitation of spectrum white space, which is the case of overlay CR systems widely investigated recently.]. The SU, equipped with one or multiple receiving and transmitting antennas, seeks opportunities to transmit data packets over one or a subset of the primary channels. To make full use of instantaneous transmission opportunities, the SU can learn the instantaneous channel state information (CSI) of the primary channel(s) by overhearing the feedback signals over a subset of primary channels, and then chooses appropriate power to transmit data over these primary channels without causing serious interference to PUs.

However, the number of channels overheard by the SU is usually limited by the number of receiving antennas, and then a natural optimisation problem for the SU is that given the past observations, which channel(s) should be overheard to attain information about the CR system to maximise the long-term utility (e.g. expected throughput).

Each primary channel is assumed to follow an identically and independently distributed (i.i.d.) two-state discrete-time Markov process in which one state 'good' corresponds to a channel with high signal to interference and noise ratio (SINR) while another state 'bad' represents a low SINR channel because of fading or high background noise.

Under this assumption, the considered channel overhearing problem can be cast into the restless multi-armed bandit (RMAB) problem in decision theory. However, the RMAB problem is proved to be PSPACE-hard [9]. Hence, a natural alternative is to seek a simple myopic policy maximising the short-term reward. However, the optimality of a myopic policy is not always guaranteed generally since the myopic policy cannot reflect the tradeoff between 'exploitation' and 'exploration' in a decision-making problem [10].

Therefore a few works on the performance of the myopic policy have been carried out along two directions. The first research direction is to design approximation algorithms and heuristic policies, and then study how far the performance of the proposed policy is to the optimum performance [11–13]. The other thrust focuses on the optimality of the myopic policy in some specific application scenarios, particularly in the context of OSA, that is, [10, 14–19].

For the similar scenario considered in this paper, the authors of [20] established the optimality of the myopic policy for the case of probing one channel ($k = 1$) each time. In our previous work [21], we established the optimality of myopic policy for the case of probing $N - 1$ of N channels and analysed its performance by domination theory, and then extended the optimality to probe multiple channels for positively correlated homogeneous channels (PCOC) [1]. Compared with the most relevant literature [1, 20, 21], in this paper we derive the sufficient conditions to guarantee the optimality of the myopic policy for four different cases: negatively correlated homogeneous channels (NCOC), heterogeneous channels (EC), positively correlated EC (PCEC) and negatively correlated EC (NCEC). Specifically, for NCOC, the reverse structure of belief vector preserves three critical exchange operations in branch and bound process concerning the derivation of the optimality of myopic policy. For heterogeneous cases (EC, PCEC, NCEC), the structure property (i.e. 'decomposability') of value function plays a critical role in deriving the bounds of some fixed policy. These structure properties are the key point for the optimality. The main contributions of this paper can be summarised as follows:

- We analyse the structure properties of belief vector and value function, and utilise them to establish sufficient conditions to

guarantee the optimality of myopic policy. The obtained optimality results combined with that of [1] (corresponding to PCOC) constitutes a complete diagram for the optimality of the myopic policy.

- For a specific scenario, we compare the relevant optimality results, which reflects the tradeoff between generic channel model and sufficient conditions, that is, the sufficient conditions sacrifice part of optimality to cover the generic channel model.
- From the viewpoint of the RMAB problem, the optimality conditions derived in this paper can be degenerated to those obtained in the literature [18, 22] by relaxing some constraints.

The rest of the paper is organised as follows: Section 3 introduces pseudo value function and its structural properties. Section 4 studies the optimality of the myopic channel probing policy. Finally, the paper is concluded in Section 5.

2 Problem formulation

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

2.1 System model

We consider a slotted multi-channel underlay CR communication system composed of N primary channels (denoted by $\mathcal{N}$), each evolving as an i.i.d. Markov chain of two states, 'good' (1) and 'bad' (0), corresponding to the situation with high (low, respectively) SINR. The state transition matrix $\boldsymbol{M}_i$ of channel i ($i \in \mathcal{N}$) is given as follows

$$\boldsymbol{M}_i = \begin{bmatrix} p_{11}^{(i)} & 1-p_{11}^{(i)} \\ p_{01}^{(i)} & 1-p_{01}^{(i)} \end{bmatrix}$$

We assume that for any i, PTx i transmits data to PRx i over channel i at each slot. At the end of each slot, PRx i sends an acknowledgement (ACK) to the corresponding PTx i on channel i if the packet is successfully decoded. The absence of an acknowledgement (denoted as NACK) signifies that the 'outage' event happened on channel i at slot t. We define 'outage' as data-packet decoding failure at PRx i, and denote the probability of the outage event

$$O_s(i) \triangleq \Pr(\text{decoding failure} \mid \text{the state of channel}_i \text{ is}),\ s \in \{1, 0\}$$

where $0 \le O_1(i) < O_0(i) \le 1,\ \forall\, i \in \mathcal{N}$.

An SU, equipped with k ($1 \le k < N$) receiving and transmitting antennas (denoted as STx and SRx), can transmit data packets on k channels opportunistically as long as the interference that it generates to PUs is limited. To exploit instantaneous transmission opportunities, the SU probes k primary channels by overhearing the primary feedback signals so as to learn the CSI at primary receivers before deciding whether transmit data on the probed channels.

Specifically, when probing a primary channel, the SU exploits ACK/NACK packet to estimate the CSI of the primary channel. Throughout our analysis we assume that the SU can perfectly overhear the ACK/NACK packet on channel i once it decides to probe channel i. This is a reasonable assumption as the ACK/NACK packets are usually transmitted in a more robust way, that is, at lower data rate. We leave the generic case of imperfect overhearing for future investigation.

2.2 RMAB formulation

- State space: Let $\boldsymbol{S}(t) \triangleq [S_1(t), \ldots, S_N(t)]$ be the CSI vector where $S_i(t) \in \{0, 1\}$ denote the 'bad', 'good' state of channel i at slot t.
- Partially observable CSI: Owing to the constraint of energy and time, the SU can only probe k channels each slot, and accordingly, obtain partial CSI, that is, the CSI vector $S(t)$ is only partially observable to the SU.
- Action space: Let $\mathcal{A}(t)$ be the set of channels probed by the SU at slot t where $\mathcal{A}(t) \subset \mathcal{N} \triangleq \{1, 2, \ldots, N\}$ and $|\mathcal{A}(t)| = k < N$.
- Observation pace: Let $\mathcal{K}(t) \triangleq \{K_{\sigma_1}(t), \ldots, K_{\sigma_k}(t)\}$ be the observation set where $K_{\sigma_i}(t) \in \{0, 1\}$ denotes the observation status NACK, ACK of the probed channel σ_i ($\sigma_i \in \mathcal{A}(t)$) in slot t.
- Information state: A sufficient statistics is a N -dimension belief vector $\Omega(t) \triangleq [\omega_1(t), \ldots, \omega_N(t)]$, where $\omega_i(t)$ ($i \in \mathcal{N}$) is the posterior probability that the state of channel i is good given all the past observations and actions of the SU.
- Probability transmission: Given $\Omega(t)$ and $\mathcal{A}(t)$, the belief vector $\Omega(t+1)$ can be updated recursively through feedback observation $\mathcal{K}(t)$ according to the following Bayes rule (1)

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

where, the operators $\phi_i(\cdot)$,$\varphi_i(\cdot)$, $\tau_i(\cdot)$, $f_i^1(\cdot)$ and $f_i^0(\cdot)$ are defined as follows

$$\phi_i(x) \triangleq \frac{(1-O_1(i))x}{f_i^1(x)} = \frac{(1-O_1(i))x}{(1-O_1(i))x + (1-O_0(i))(1-x)}$$

$$\varphi_i(x) \triangleq \frac{O_1(i)x}{f_i^0(x)} = \frac{O_1(i)x}{O_1(i)x + O_0(i)(1-x)}$$

$$\tau_i(x) \triangleq (p_{11}^{(i)} - p_{01}^{(i)})x + p_{01}^{i}$$

$\phi_i(x)$($\varphi_i(x)$) represents the probability of successful decoding (decoding failure) with $S_i(t) = 1$ and that of successful decoding (decoding failure), respectively; $\tau_i(x)$ is the Markovian evolving rule.

A channel probing policy π is composed of a serials of mappings $\pi = [\pi_1, \ldots, \pi_T]$ where π_t maps the belief vector $\Omega(t)$ to the action $\mathcal{A}(t)$ at each slot t: that is, $\pi_t : \Omega(t) \mapsto \mathcal{A}(t),\ |\mathcal{A}(t)| = k$.

Therefore the SU's optimisation problem $\boldsymbol{P}$ is to find the optimal policy π^* which maximises the expected accumulated discounted reward over a finite time horizon

$$\boldsymbol{P}:\ \pi^* = \arg\max_{\pi} \mathbb{E}_{\pi}\left[\sum_{t=1}^{T} \beta^{t-1} R(\pi_t(\Omega(t)))|\Omega(1)\right] \tag{2}$$

where (1) $R(\pi_t(\Omega(t)))$ is the reward in slot t under the policy π_t with the initial belief vector $\Omega(1)$ [If no information on the initial system state is available, each entry of $\Omega(1)$ can be set to the stationary distribution $\omega_i(1) = p_{01}^{(i)}/(1 + p_{01}^{(i)} - p_{11}^{(i)})$, and (2) $0 \le \beta \le 1$ is the discount factor characterising the feature that future reward is less valuable than immediate reward.

For the ease of analysis, $\boldsymbol{P}$ can be rewritten as the dynamic programming formulation **DP**

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

$$V_t(\Omega(t)) = \max_{\mathcal{A}(t)} \mathbb{E}\Bigg[R(\pi_t(\Omega(t))) + \beta \sum_{\mathcal{E} \subseteq \mathcal{A}(t)} \prod_{i \in \mathcal{E}} f_i^1(\omega_i(t)) \prod_{j \in \mathcal{A}(t)}^{j \notin \mathcal{E}} f_j^0(\omega_i(t)) V_{t+1}(\Omega(t+1))\Bigg]$$

where (1) $V_t(\Omega(t))$ is the value function corresponding to the maximal expected reward from time slot t to T, and (2) $\Omega(t+1)$ follows the evolution (1) when the channels in the subset $\mathcal{E}$ are

IET Commun., 2015, Vol. 9, Iss. 7, pp. 1017–1025

observed in 'good' state while the channels in $\mathcal{A}(t)\backslash\mathcal{E}$ are observed in 'bad' state.

2.3 Myopic policy

Theoretically, the optimal policy of $\boldsymbol{P}$ can be obtained by solving the above dynamic programming **DP**. It is infeasible, however, because of the impact of the current action on the future reward and the unaccountable space of the belief vector, thus obtaining the optimal policy directly from the above recursive equations is computationally prohibitive. Hence, a natural alternative is to seek a simple myopic policy maximising the immediate reward defined as follows:

Definition 1: Let $F(\Omega_A(t)) \triangleq \mathbb{E}[R(\pi_t(\Omega(t)))]$ denote the expected immediate reward obtained at slot t under the policy π_t with $\Omega_A(t) \triangleq \{\omega_i(t): i \in \mathcal{A}(t)\}$, then the myopic policy is to probe the k channels that maximises $F(\Omega_A(t))$, that is, $\bar{\mathcal{A}}(t) = \text{argmax}_{\mathcal{A}(t)\subseteq\mathcal{N}} F(\Omega_A(t))$.

To make our analysis more generic, we focus on a class of reward functions $F(\Omega_A(t))$, termed as regular functions defined in [18]. More specifically, the expected immediate reward function $F(\Omega_A(t))$ is assumed to be symmetrical, monotonically non-decreasing and decomposable [18]. Under this assumption, the myopic policy is to choose the k channels with the largest belief value. However, the myopic policy only reflects the 'exploitation' in the two conflicting factors: 'exploitation' and 'exploration', thus it is not clear whether it would not be optimal or not. Hence, it is of significant importance to justify the optimality of the myopic policy, which is exactly the focus of the following sections.

3 Value function and its structure

In this section, we introduce the 'pseudo value function' [1] and then give a number of auxiliary lemmas. For the ease of presentation, we first state the notations and parameters employed in the following analysis.

3.1 Notations

(1) $\hat{\omega}_{-i}(t) \triangleq \{\hat{\omega}_j(t): j \in \mathcal{A}(t), j \neq i\}$, $\Omega_{-I}(t) \triangleq \Omega(t)\backslash\{\omega_I(t)\}$.
(2) $\Delta_{\max} \triangleq \max\ \{F(1, \hat{\omega}_{-i}(t)) - F(0, \hat{\omega}_{-i}(t)): \hat{\omega}_{-i}(t) \in [0, 1]^{k-1}, 1 \leq t \leq T\}$.
(3) $\Delta_{\min} \triangleq \min\{F(1, \hat{\omega}_{-i}(t)) - F(0, \hat{\omega}_{-i}(t)): \hat{\omega}_{-i}(t) \in [0, 1]^{k-1}, 1 \leq t \leq T\}$.
(4) $\delta \triangleq \max_{i\in\mathcal{N}} |p_{11}^{(i)} - p_{01}^{(i)}|$.

Note $\omega(t)$ $(\Omega(t))$ and $\omega(\Omega)$ will be interchangeably used without ambiguity.

Lemma 1: $f_i^1(\omega)$ is monotonically increasing in ω while $f_i^0(\omega)$ monotonically decreasing in ω.

Proof: The lemma holds since $f_i^1(\omega) = (1 - O_1(i))\omega + (1 - O_0(i))(1 - \omega)$ and $f_i^0(\omega) = 1 - f_i^1(\omega)$. □

Lemma 2: The following properties of $\tau_i(\omega_i(t))$ hold:

(1) If $p_{01}^{(i)} < p_{11}^{(i)}$, $\tau_i(\omega_i(t))$ is monotonically increasing in $\omega_i(t)$ and $p_{01}^{(i)} \leq \tau_i(\omega_i(t)) \leq p_{11}^{(i)}$, $\forall 0 \leq \omega_i(t) \leq 1$.

(2) If $p_{01}^{(i)} > p_{11}^{(i)}$, $\tau_i(\omega_i(t))$ is monotonically decreasing in $\omega_i(t)$ and $p_{11}^{(i)} \leq \tau_i(\omega_i(t)) \leq p_{01}^{(i)}$, $\forall 0 \leq \omega_i(t) \leq 1$.

Proof: Lemma 2 holds from $\tau_i(\omega_i(t)) = (p_{11}^{(i)} - p_{01}^{(i)})\omega_i(t) + p_{01}^{(i)}$. □

3.2 Pseudo value function

In this part, we introduce two pseudo value functions in the recursive form. The objective of introducing AVF is to conveniently analyse the performance of myopic policy, while that of introducing adjugate auxiliary value function (AAVF) is to decompose belief vector. Specifically, the optimal problem $\boldsymbol{P}$ depends on initial belief vector and Bayes rule. Given Bayes rule (1), $\boldsymbol{P}$ depends on the initial belief vector $\Omega(1)$ and further $\Omega(t)$ in the decision-making process. However, the belief vector $\Omega(t)$ influences both probing policy (i.e. π_t: $\Omega(t) \mapsto \mathcal{A}(t)$) and the reward value (i.e. AVF), which makes the analysis difficult because of the tight coupling of policy and reward value. Thus, the AAVF is elaborately designed to decompose the belief vector into two parts of which one reflects probing policy and the other the value, and then we can study the probing policy without focusing on the reward value to some extent.

Definition 2: [Auxiliary value function (AVF) and AAVF] The AVF: $W_t(\Omega(t))$ and AAVF $\hat{W}_t(\Omega(t); \hat{\Omega}(t))$ $(1 \leq t \leq T,\ t+1 \leq r \leq T)$ are defined as follows

$$\text{AVF}\begin{cases} W_T(\Omega(T)) = F(\Omega_{\bar{\mathcal{A}}}(T)) \\ W_r(\Omega(r)) = F(\Omega_{\bar{\mathcal{A}}}(r)) + \beta \sum\limits_{\mathcal{E}\subseteq\bar{\mathcal{A}}(r)} C^{\mathcal{E}}_{\bar{\mathcal{A}}(r)} W_{r+1}(\Omega_{\mathcal{E}}(r+1)) \\ W_t^{\mathcal{A}}(\Omega(t)) = F(\Omega_{\mathcal{A}(t)}(t)) + \beta \underbrace{\sum\limits_{\mathcal{E}\subseteq\mathcal{A}(t)} C^{\mathcal{E}}_{\mathcal{A}(t)} W_{t+1}(\Omega_{\mathcal{E}}(t+1))}_{\Gamma^{\mathcal{A}}(\Omega(t))} \end{cases} \tag{3}$$

(see (4))

where

(1) $\widehat{C}^{\mathcal{E}}_{\mathcal{M}} \triangleq \prod_{i\in\mathcal{E}} f_i^1(\hat{\omega}_i(t)) \prod_{j\in\mathcal{M}\backslash\mathcal{E}} f_j^0(\hat{\omega}_j(t))$ denotes the expected probability that the channels in $\mathcal{E}$ are observed in '<u>good</u>' state whereas those in $\mathcal{M}\backslash\mathcal{E}$ are 'bad';
(2) $\Omega_{\mathcal{E}}(t+1)$ and $\Omega_{\mathcal{E}}(r+1)$ are generated by $\langle\Omega(t), \mathcal{A}(t), \mathcal{E}\rangle$ and $\langle\Omega(r), \bar{\mathcal{A}}(r), \mathcal{E}\rangle$, respectively, according to (1) and then sorted by belief value.
(3) $\hat{\Omega}_{\mathcal{E}}(t+1)$ and $\hat{\Omega}_{\mathcal{E}}(r+1)$ are generated by $\langle\hat{\Omega}(t), \mathcal{A}(t), \mathcal{E}\rangle$ and $\langle\hat{\Omega}(r), \bar{\mathcal{A}}(r), \mathcal{E}\rangle$, respectively, according to (1), and the order of channel index keeps consistent with that of $\Omega_{\mathcal{E}}(t+1)$ and $\Omega_{\mathcal{E}}(r+1)$, respectively.
(4) $\overline{\mathcal{A}}(r)$ and $\mathcal{A}(t)$ of AAVF are the same with that of AVF.
(5) If $\hat{\Omega}(t) = \Omega(t)$, then AAVF degenerates into AVF.

3.3 Structure of value function

This part gives some critical structure properties of AVF and AAVF, that is, 'symmetry' and 'decomposability', which were proved in our previous work [1] and recaptured here for completeness.

$$\text{AAVF}\begin{cases} \hat{W}_T(\Omega(T); \hat{\Omega}(T)) = F(\hat{\Omega}_{\bar{\mathcal{A}}}(T)) \\ \hat{W}_r(\Omega(r); \hat{\Omega}(r)) = F(\hat{\Omega}_{\bar{\mathcal{A}}}(r)) + \beta \sum\limits_{\mathcal{E}\subseteq\bar{\mathcal{A}}(r)} \widehat{C}^{\mathcal{E}}_{\bar{\mathcal{A}}(r)} \hat{W}_{r+1}(\Omega_{\mathcal{E}}(r+1); \hat{\Omega}_{\mathcal{E}}(r+1)) \\ \hat{W}_t^{\mathcal{A}}(\Omega(t); \hat{\Omega}(t)) = F(\hat{\Omega}_{\mathcal{A}(t)}(t)) + \beta \sum\limits_{\mathcal{E}\subseteq\mathcal{A}(t)} \widehat{C}^{\mathcal{E}}_{\mathcal{A}(t)} \hat{W}_{t+1}(\Omega_{\mathcal{E}}(t+1); \hat{\Omega}_{\mathcal{E}}(t+1)) \end{cases} \tag{4}$$

Lemma 3: $W_t(\Omega(t))$ is symmetrical in ω_i, ω_j for any $i, j \in \mathcal{A}(t)$ or $i, j \notin \mathcal{A}(t)$ for all $t = 1, 2, \ldots, T$, that is,

$$W_t(\omega_1, \ldots, \omega_i, \ldots, \omega_j, \ldots, \omega_N) = W_t(\omega_1, \ldots, \omega_j, \ldots, \omega_i, \ldots, \omega_N)$$

Lemma 4: $\hat{W}_t^{\mathcal{A}}(\Omega; \hat{\Omega}(t))$ is decomposable for all $t = 1, 2, \ldots, T$, that is,

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

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

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

Corollary 1: For any belief vector $\boldsymbol{\Omega}$, it holds that $\forall l, m \in \mathcal{N}$, $t = 1, 2, \ldots, T$

$$\hat{W}_t^{\mathcal{A}}(\Omega; \hat{\Omega}_0) - \hat{W}_t^{\mathcal{A}}(\Omega; \hat{\Omega}_1) = (\hat{\omega}_l - \hat{\omega}_m)[W_t^{\mathcal{A}}(\Omega; \hat{\Omega}_2) - W_t^{\mathcal{A}}(\Omega; \hat{\Omega}_3)]$$

where

$$\hat{\Omega}_0 = (\hat{\omega}_1, \ldots, \hat{\omega}_l, \ldots, \hat{\omega}_m, \ldots, \hat{\omega}_N)$$
$$\hat{\Omega}_1 = (\hat{\omega}_1, \ldots, \hat{\omega}_m, \ldots, \hat{\omega}_l, \ldots, \hat{\omega}_N)$$
$$\hat{\Omega}_2 = (\hat{\omega}_1, \ldots, 1, \ldots, 0, \ldots, \hat{\omega}_N)$$
$$\hat{\Omega}_3 = (\hat{\omega}_1, \ldots, 0, \ldots, 1, \ldots, \hat{\omega}_N)$$

Remark 1: Lemma 3 implies that the reward generated by AVF remains the same against any channel permutation within the probed channels and within the non-probed channels. That is, given channel i, $j \in \mathcal{A}(t)$, probing channel i first and then channel j will generate same reward with the case of probing channel j first and then channel i. Further, given $\mathcal{A}(t)$, the same reward will be attained no matter what order is adopted to probe these channels in $\mathcal{A}(t)$.

Remark 2: Lemma 4 states that given the probing policy, the reward attained from AAVF can be decomposed into two terms with deterministic realisations 0 and 1 in any channel of the value belief vector. Mathematically, the value function is 'linear' or 'piecewise linear', and accordingly, it can be written as the combination of two value functions in two endpoints (0 and 1).

4 Analysis on optimality of myopic policy

In this section, we sequentially derive sufficient conditions to guarantee the optimality of the myopic policy for four different cases: NCOC, EC, PCEC and NCEC . Next, for a special scenario, we conduct a comparative study on the obtained optimality results, which clearly shows the tradeoff between sufficient conditions and channel characteristics.

4.1 Negatively correlated homogeneous channels

In this part, we consider the case of negatively correlated homogenous channels. That is, the following holds:

(1) $\forall i, p_{11}^{(i)} = p_{11}, p_{01}^{(i)} = p_{01}, O_1(i) = O_1, O_0(i) = O_0$ (homogeneous).
(2) $p_{11} < p_{01}$ (negatively correlated).

For the ease of analysis, we assume

(3) $\forall i, p_{11} \le \hat{\omega}_i \le p_{01}$.

In [15], the authors shows the myopic policy is not optimal by a counterexample, and since then research on the performance of myopic policy is circumvented in the existing works involved in the optimality of myopic policy under the context of RMAB. However, in this paper, we prove that the myopic policy is optimal only imposing a weak condition on the initial belief vector $\Omega(1)$, that is, $\forall i$, $p_{11} \le \hat{\omega}_i \le p_{01}$. In fact, the weak condition will be automatically satisfied from the second slot since the belief value would enter into $[p_{11}, p_{01}]$ according to Lemma 2.

Through analysing the structure of myopic policy and belief vector [21], we find the proof for positively correlated homogenous model [1] can be slightly modified to fit into the negatively correlated homogenous model. Hence, we first give the following structure of belief vector, and then points out the nuance in the proof of optimality of myopic policy.

Theorem 1 (Structure of myopic policy): If $O_1/O_0 \le p_{11}(1 - p_{01})/[p_{01}(1 - p_{11})]$, 'we have the following channel order rules at the end of each slot'.

(1) The initial channel ordering $\boldsymbol{Q}(1)$ is determined by the initial belief vector

$$\omega_{\sigma_1}(1) \ge \cdots \ge \omega_{\sigma_N}(1) \Rightarrow \boldsymbol{Q}(1) = (\sigma_1, \sigma_2, \ldots, \sigma_N),$$

(2) The channels over which ACKs are observed will be moved to the end of the queue, and the channels over which NACKs are observed will stay at the head of the queue while reversing the order of other channels.

Proof: Assume $\boldsymbol{Q}(t) = (\sigma_1, \ldots, \sigma_N)$ at slot t, we thus have $p_{01} \ge \omega_{\sigma_1}(t) \ge \cdots \ge \omega_{\sigma_N}(t) \ge p_{11}$. If ACK is observed over channel σ_1, then $\omega_{\sigma_1}(t+1) = \tau(\phi(\omega_{\sigma_1}(t))) \le \tau(\omega_{\sigma_1}(t)) \le \cdots \le \tau(\omega_{\sigma_N}(t))$ by Lemma 2, and thus $\boldsymbol{Q}(t+1) = (\sigma_N, \ldots, \sigma_1)$ according to the descending order of ω. If NACK is observed over channel σ_1, then $\omega_{\sigma_1}(t+1) = \tau(\varphi(\omega_{\sigma_1}(t))) \ge \tau(p_{11}) \ge \tau(\omega_{\sigma_N}(t)) \ge \cdots \ge \tau(\omega_{\sigma_2}(t))$, and further $\boldsymbol{Q}(t+1) = (\sigma_1, \sigma_N, \ldots, \sigma_2)$. □

Remark 3: Assume $\boldsymbol{Q}(t) = (\sigma_1, \ldots, \sigma_N)$ at slot t where $\omega_{\sigma_1}(t) \ge \cdots \ge \omega_{\sigma_N}(t)$. When ACK and NACK are observed over channel σ_1, respectively, the structure of $\boldsymbol{Q}(t+1)$ is stated in the following table. Meanwhile, $\boldsymbol{Q}(t+1)$ for PCOC is also listed for the purpose of comparison. As shown in Table 1, $\boldsymbol{Q}(t+1)$ shows the reverse order in two cases. It is the reverse order which preserves three kinds of exchange operation in Lemma 5–7. Thus, Lemma 5–7 still hold by exchanging p_{11} and p_{01}.

Following the similar induction in [1], we have the following three lemmas and theorem.

Lemma 5: Given that (1) F is regular, (2) $O_1/O_0 \le p_{11}(1 - p_{01})/p_{01}(1 - p_{11})$, (3) $\beta \le \Delta_{\min}/[\Delta_{\max}((1 - O_1/O_0)(1 - p_{11}) + \delta O_1/1 - \delta(1 - O_1))]$, if $\hat{\omega}_l \ge \hat{\omega}_m$, it holds that for $1 \le t \le T$

$$W_t(\hat{\Omega}_0) \ge W_t(\hat{\Omega}_1)$$

where, $\hat{\Omega}_0 = (\hat{\omega}_1, \ldots, \hat{\omega}_l, \ldots, \hat{\omega}_m, \ldots, \hat{\omega}_N)$, $\hat{\Omega}_1 = (\hat{\omega}_1, \ldots, \hat{\omega}_m, \ldots, \hat{\omega}_l, \ldots, \hat{\omega}_N)$.

Table 1 Structure of Myopic Policy ($\boldsymbol{Q}(t) = (\sigma_1, \ldots, \sigma_N)$)

Obs. σ_1	Positively correlated	Negatively correlated
ACK	$\boldsymbol{Q}(t+1) = (\sigma_1, \sigma_2, \ldots, \sigma_N)$	$\boldsymbol{Q}(t+1) = (\sigma_N, \ldots, \sigma_2, \sigma_1)$
NACK	$\boldsymbol{Q}(t+1) = (\sigma_2, \ldots, \sigma_N, \sigma_1)$	$\boldsymbol{Q}(t+1) = (\sigma_1, \sigma_N, \ldots, \sigma_2)$

IET Commun., 2015, Vol. 9, Iss. 7, pp. 1017–1025

Lemma 6: Given that (1) F is regular, (2) $O_1/O_0 \le p_{11}(1-p_{01})/p_{01}(1-p_{11})$, (3) $\beta \le \Delta_{\min}/[\Delta_{\max}((1-O_1/O_0)(1-p_{11})+\delta O_1/1-\delta(1-O_1))]$, if $\hat{\omega}_l \ge \hat{\omega}_m$, it holds that for $1 \le t \le T$

$$\hat{W}_t(\dot{\Omega};\hat{\Omega}_0) - \hat{W}_t(\dot{\Omega};\hat{\Omega}_1) \le \frac{1-p_{11}}{O_0}\Delta_{\max}$$

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

Lemma 7: Given that (1) F is regular, (2) $O_1/O_0 \le p_{11}(1-p_{01})/p_{01}(1-p_{11})$, (3) $\beta \le \Delta_{\min}/[\Delta_{\max}((1-O_1/O_0)(1-p_{11})+\delta O_1/1-\delta(1-O_1))]$, if $\hat{\omega}_l \ge \hat{\omega}_m$, it holds that for $1 \le t \le T$

$$\hat{W}_t(\dot{\Omega};\hat{\Omega}_0) - \hat{W}_t(\dot{\Omega};\hat{\Omega}_1) \le \delta\Delta_{\max}\frac{1-[\beta\delta(1-O_1)]^{T-t+1}}{1-\beta\delta(1-O_1)}$$

where, $\hat{\Omega}_0 = (\hat{\omega}_1, \hat{\omega}_2, \ldots, \hat{\omega}_{N-1}, \hat{\omega}_N)$, $\hat{\Omega}_1 = (\hat{\omega}_N, \hat{\omega}_2, \ldots, \hat{\omega}_{N-1}, \hat{\omega}_1)$

Theorem 2: If $p_{11} \le \omega_i(1) \le p_{01}$ for $1 \le i \le N$, the myopic policy is optimal if (1) $F(\Omega)$ is regular; (2) $O_1/O_0 \le p_{11}(1-p_{01})/p_{01}(1-p_{11})$; (3) $\beta \le \Delta_{\min}/[\Delta_{\max}((1-O_1/O_0)(1-p_{11})+\delta O_1/1-\delta(1-O_1))]$.

Remark 4: Theorem 2 gives the sufficient conditions to justify the optimality of the myopic policy, that is, probing those best channels, for the NCOC. More importantly, this theorem counter proves the intuition that the myopic policy is not optimal for negatively correlated case.

4.2 Heterogeneous channels

In this part, we consider the case of the EC which implicitly includes two cases: positively correlated channels, negatively correlated channels.

We start by showing the following important lemma (Lemma 8) and then establish the sufficient condition to guarantee the optimality of the myopic policy. In Lemma 8, we consider two belief vectors $\Omega a_l = (\Omega a_{-l}, \omega_l)$ and $\Omega'_l = (\Omega_{-l}, \omega_l')$ that differ only in one element $\omega_l \le \omega_l'$. Let $\mathcal{A}$ and $\mathcal{A}'$ denote the largest k elements in Ω_l and Ω'_l, respectively, [Without ambiguity, $\mathcal{A}(t)$ and $\mathcal{A}$ would be used interchangeably. The tie, if there exists, is resolved according to the increasing order of channel index], Lemma 8 gives the lower bound and the upper bound on $W_t^{\mathcal{A}'}(\Omega'_l) - W_t^{\mathcal{A}}(\Omega_l)$.

Lemma 8: Given

(1) $\Omega_l = (\Omega_{-l}, \omega_l)$, $\Omega'_l = (\Omega_{-l}, \omega_l')$, $\omega_l \le \omega_l'$.
(2) $\Delta_{\min} \ge \Delta_{\max}\sum_{i=1}^{T-t}\beta^i\delta^i$,

we have for $1 \le t \le T$

- if $l \in \mathcal{A}'$ and $l \in \mathcal{A}$

$$(\omega_l' - \omega_l)\left(\Delta_{\min} - \Delta_{\max}\sum_{i=1}^{T-t}\beta^i\delta^i\right) \le W_t^{\mathcal{A}'}(\Omega'_l) - W_t^{\mathcal{A}}(\Omega_l) \le (\omega_l' - \omega_l)\Delta_{\max}\left(1 + \sum_{i=1}^{T-t}\beta^i\delta^i\right)$$

- if $l \notin \mathcal{A}'$ and $l \notin \mathcal{A}$,

$$\left|W_t^{\mathcal{A}'}(\Omega'_l) - W_t^{\mathcal{A}}(\Omega_l)\right| \le (\omega_l' - \omega_l)\Delta_{\max}\sum_{i=1}^{T-t}\beta^i\delta^i$$

- if $l \in \mathcal{A}'$ and $l \notin \mathcal{A}$,

$$\left|W_t^{\mathcal{A}'}(\Omega'_l) - W_t^{\mathcal{A}}(\Omega_l)\right| \le (\omega_l' - \omega_l)\Delta_{\max}\left(1 + \sum_{i=1}^{T-t}\beta^i\delta^i\right)$$

Proof: The proof is given in A. □

In the following lemma, we consider $\mathcal{A}_l$ and $\mathcal{A}_m$ differing in one element, that is, $\mathcal{A}_l$, $\{l\} = \mathcal{A}_m$, $\{m\}$, $l \in \mathcal{A}_l$ and $m \in \mathcal{A}_m$ and $\omega_l > \omega_m$ and establish sufficient condition such that $W_t^{\mathcal{A}_l}(\Omega) > W_t^{\mathcal{A}_m}(\Omega)$.

Lemma 9: Given $m \in \mathcal{A}_m$, $l \in \mathcal{A}_l$, $\omega_l > \omega_m$ and $\mathcal{A}_l$, $\{l\} = \mathcal{A}_m$, $\{m\}$, if $\Delta_{\min} \ge 2\Delta_{\max}\sum_{i=1}^{T-1}\beta^i\delta^i$, then $W_t^{\mathcal{A}_l}(\Omega) > W_t^{\mathcal{A}_m}(\Omega)$.

Proof: Let Ω' denote the set of channel belief values with $\omega_l' > \omega_m$ and $\omega_i' > \omega_i$ for $\forall i \ne l$ and $i \in \mathcal{N}$, then $W_t^{\mathcal{A}_l}(\Omega') = W_t^{\mathcal{A}_m}(\Omega')$. By Lemma 8, we have

$$\begin{aligned} &W_t^{\mathcal{A}_l}(\Omega) - W_t^{\mathcal{A}_m}(\Omega) \\ &= [W_t^{\mathcal{A}_l}(\Omega) - W_t^{\mathcal{A}_l}(\Omega')] - [W_t^{\mathcal{A}_m}(\Omega) - W_t^{\mathcal{A}_m}(\Omega')] \\ &\ge (\omega_l' - \omega_l)(\Delta_{\min} - \Delta_{\max}\sum_{i=1}^{T-t}\beta^i\delta^i) - (\omega_l' - \omega_l)\Delta_{\max}\sum_{i=1}^{T-t}\beta^i\delta^i \\ &\quad \times (\omega_l' - \omega_l)(\Delta_{\min} - 2\Delta_{\max}\sum_{i=1}^{T-t}\beta^i\delta^i) \ge 0 \end{aligned}$$

□

Based on Lemma 9, we have the following theorem which states the optimal condition of the myopic policy.

Theorem 3: The myopic policy is optimal if $\sum_{i=1}^{T-1}\beta^i\delta^i \le (\Delta_{\min}/2\Delta_{\max})$, specifically, if $T \to \infty$, $\beta\delta/(1-\beta\delta) \le \Delta_{\min}/(2\Delta_{\max})$.

Proof: When T 1/2 ∞, we prove the theorem by backward induction. The theorem holds trivially for T. Assume that it holds for $T-1$, …, $t+1$, that is, the optimal accessing policy is to sense the best channel from time slot $t+1$ to T. We now show that it holds for t. □

Suppose, by contradiction, that given the belief vector $\Omega \triangleq \{\omega_{i_1}, \ldots, \omega_{i_N}\}$ and $\omega_1 < \omega_2 < \cdots < \omega_N$, the optimal policy is to probe the best channels from time slot $t+1$ to T and thus, at slot t, to probe channels $\mathcal{A}(t) = \{i_1, \ldots, i_k\} \ne \overline{\mathcal{A}}(t) = \{1, \ldots, k\}$, given that the latter, $\overline{\mathcal{A}}(t)$, includes the best k channels in terms of belief value at slot t. There must exist i_m and i_l at slot t such that $m \le k < l$ and $\omega_{i_m} < \omega_{i_k} \le \omega_{i_l}$. It then follows from Lemma 9 that $W_t^{\{i_1,\ldots,i_k\}}(\Omega) < W_t^{\{i_1,\ldots,i_{m-1},i_l,i_{m+1},\ldots,i_k\}}(\Omega)$, which contradicts with the assumption that the latter is the optimal policy. This contradiction completes our proof.

When $T \to \infty$, the proof follows straightforwardly by noticing that $\sum_{i=1}^{\infty}x^i = x/(1-x)$ for any $x \in (0, 1)$. □

4.3 Positively correlated heterogeneous channels

In this part, we consider the PCEC , that is, $\forall i, p_{11}^{(i)} > p_{01}^{(i)}$. Following the similar induction as Lemma 8, we have the following lemma.

Lemma 10: Given

(1) $\Omega_l = (\Omega_{-l}, \omega_l)$, $\Omega_l' = (\Omega_{-l}, \omega_l')$, $\omega_l \le \omega_l'$;
(2) $\forall i$, $p_{11}^{(i)} > p_{01}^{(i)}$,
we have for $1 \le t \le T$

- if $l \in \mathcal{A}'$ and $l \in \mathcal{A}$

$$(\omega_l' - \omega_l)\Delta_{\min} \le W_t^{\mathcal{A}'}(\Omega_l') - W_t^{\mathcal{A}}(\Omega_l) \le (\omega_l' - \omega_l)\Delta_{\max}\left(1 + \sum_{i=1}^{T-t}\beta^i\delta^i\right)$$

- if $l \notin \mathcal{A}'$ and $l \notin \mathcal{A}$

$$0 \le W_t^{\mathcal{A}'}(\Omega_l') - W_t^{\mathcal{A}}(\Omega_l) \le (\omega_l' - \omega_l)\Delta_{\max}\sum_{i=1}^{T-t}\beta^i\delta^i$$

- if $l \in \mathcal{A}'$ and $l \notin \mathcal{A}$

$$0 \le W_t^{\mathcal{A}'}(\Omega_l') - W_t^{\mathcal{A}}(\Omega_l) \le (\omega_l' - \omega_l)\Delta_{\max}\left(1 + \sum_{i=1}^{T-t}\beta^i\delta^i\right)$$

Based on Lemma 10, it is easy to obtain the following sufficient conditions for the optimality of the myopic policy.

Theorem 4: The myopic policy is optimal if $\sum_{i=1}^{T-1}\beta^i\delta^i \le \Delta_{\min}/\Delta_{\max}$, specifically, if $T \to \infty$, $\beta\delta/(1-\beta\delta) \le \Delta_{\min}/\Delta_{\max}$.

Remark 5: The sufficient conditions for PCEC in Theorem 4 is looser than those for EC in Theorem 3, which reflects the fact that the channel model of the latter covers that of the former with loosing part of optimality.

4.4 Negatively correlated heterogeneous channels

In this part, we consider the PCEC , that is, $p_{11}^{(i)} < p_{01}^{(i)}$ for $\forall i$. Following the similar induction as Lemma 8, we have the following lemma.

Lemma 11: Given

(1) $\Omega_l = (\Omega_{-l}, \omega_l)$, $\Omega_l{}' = (\Omega_{-l}, \omega_l{}')$, $\omega_l \le \omega_l{}'$;
(2) $\forall i$, $p_{11}^{(i)} < p_{01}^{(i)}$;
(3) $\Delta_{\min} \ge \Delta_{\max}\sum_{i=1}^{T-t}\beta^i\delta^i$ for $1 \le t \le T$,

we have

- if $l \in \mathcal{A}'$ and $l \in \mathcal{A}$

$$(\omega_l' - \omega_l)\left(\Delta_{\min} - \Delta_{\max}\sum_{i=1}^{T-t}\beta^i\delta^i\right) \le W_t^{\mathcal{A}'}(\Omega_l') - W_t^{\mathcal{A}}(\Omega_l) \le (\omega_l' - \omega_l)\Delta_{\max}\left(1 + \sum_{i=1}^{T-t}\beta^i\delta^i\right)$$

- if $l \notin \mathcal{A}'$ and $l \notin \mathcal{A}$

$$-(\omega_l' - \omega_l)\Delta_{\max}\sum_{i=1}^{T-t}\beta^i\delta^i \le W_t^{\mathcal{A}'}(\Omega_l') - W_t^{\mathcal{A}}(\Omega_l) \le (\omega_l' - \omega_l)\Delta_{\max}\left(-\beta\delta + \sum_{i=1}^{T-t}\beta^i\delta^i\right)$$

- if $l \in \mathcal{A}'$ and $l \notin \mathcal{A}$

$$\left|W_t^{\mathcal{A}'}(\Omega_l') - W_t^{\mathcal{A}}(\Omega_l)\right| \le (\omega_l' - \omega_l)\Delta_{\max}\left(1 + \sum_{i=1}^{T-t}\beta^i\delta^i\right)$$

Table 2 Optimal conditions of myopic policy

	Homogeneous	heterogeneous
$p_{11}^{(i)} > p_{01}^{(i)}$	$\forall\delta$, $0 \le \beta \le 1$	$\beta\delta \le 1/2$
$p_{11}^{(i)} < p_{01}^{(i)}$	$\forall\delta$, $0 \le \beta \le 1$	$\beta\delta \le \sqrt{2} - 1$
$p_{11}^{(i)} \ne p_{01}^{(i)}$		$\beta\delta \le 1/3$

Based on Lemma 11, it is easy to obtain the following sufficient conditions for the optimality of the myopic policy.

Theorem 5: The myopic policy is optimal if $2\sum_{i=1}^{T-1}\beta^i\delta^i - \beta\delta \le \Delta_{\min}/\Delta_{\max}$, specifically, if $T \to \infty$, $\beta\delta(1+\beta\delta)/(1-\beta\delta) \le \Delta_{\min}/\Delta_{\max}$.

Remark 6: The sufficient conditions for NCEC in Theorem 5 is looser than those for EC in Theorem 3, which reflects the fact that the channel model of the latter covers that of the former with loosing part of optimality.

4.5 Discussion

To illustrate the application of the obtained result, we study a concrete underlay CR system where the SU can transmit at rate r_1 if the channel probed is observed in the good state and $r_0 (r_0 \le r_1)$ for the bad state. In this scenario, the utility function can be formulated as

$$F(\Omega_{\mathcal{A}}) = \sum_{i \in \mathcal{A}}[r_1 \cdot \omega_i + r_0 \cdot (1 - \omega_i)]$$

thus, $\Delta_{\min} = \Delta_{\max} = r_1 - r_0$.

According to Theorem 2–5 and Theorem 1 of [1], we have the sufficient conditions to guarantee the optimality of the myopic policy, which are stated in the following Table 2.

5 Conclusion

We have investigated the optimality of the myopic policy for the RMAB problem arisen in the field of underlay CR systems, and obtained the sufficient conditions to guarantee the optimality of the myopic policy for four different cases. The obtained results, combined with the optimality results in [1], constitutes a complete paradigm regarding how to optimally choose channels to access in a underlay cognitive radio system. As future work, a natural direction is to study whether the proposed sufficient conditions are necessary. If not, we need to derive much better sufficient conditions to guarantee the optimality of the myopic policy. Another direction we are pursuing is to investigate the RMAB problem with multiple players with potentially conflicts among them and to study the structure and the optimality of the myopic policy in that context.

6 Acknowledgments

The authors thank the editor and the anonymous referee for their valuable comments and suggestions that improved the clarity and quality of this manuscript. This work was supported by the National Natural Science Foundation of China under Grant 61303027/61373024, the China Postdoctoral Science Foundation 2013M531753/2014T70748, the Fundamental Research Funds for the Central Universities (WUT:2014-IV-067) and the Agence Nationale de la Recherche (ANR) under grant Green-Dysan (ANR-12-IS03).

7 References

1 Wang, K., Chen, L., Liu, Q.: 'Opportunistic spectrum access by exploiting primary user feedbacks in underlay cognitive radio systems: An optimality analysis', *IEEE J. Sel. Top. Signal Process.*, 2013, **7**, (5), pp. 869–882
2 Mitola, J., Maguire, G.Q.: 'Cognitive radios: making software radios more personal', *IEEE Pers. Commun.*, 1999, **6**, pp. 13–18
3 Haykin, S.: 'Cognitive radio: Brain-empowered wireless communications', *IEEE J. Sel. Areas Commun.*, 2005, **23**, (2), pp. 201–220
4 Etkin, R., Parekh, A., Tse, D.: 'Spectrum sharing for unlicensed bands', *IEEE J. Sel. Areas Commun.*, 2007, **25**, pp. 517–528
5 Xing, Y., Chandramouli, R., Mangold, S., Shankar, S.N.: 'Dynamic spectrum access in open spectrum wireless networks', *IEEE J. Sel. Areas Commun.*, 2006, **24**, pp. 626–637
6 Federal Communications Commission: Spectrum policy task force report (etdocket no.02-135). November 2002
7 Ghasemi, A., Sousa, E.S.: 'Fundamental limits of spectrum sharing in fading environments', *IEEE Trans. Wirel. Commun.*, 2007, **6**, pp. 649–658
8 Zhang, R., Liang, Y.-C.: 'Exploiting multi-antennas for opportunistic spectrum sharing in cognitive radio networks', *IEEE J. Sel. Top. Signal Process.*, 2008, **2**, pp. 88–102
9 Papadimitriou, C.H., Tsitsiklis, J.N.: 'The complexity of optimal queueing network control', *Math. Oper. Res.*, 1999, **24**, (2), pp. 293–305
10 Wang, K., Chen, L., Al Agha, K., Liu, Q.: 'On optimality of myopic policy in opportunistic spectrum access: the case of sensing multiple channels and accessing one channel', *IEEE Wirel. Commun. Lett.*, 2012, **1**, (5), pp. 452–455
11 Guha, S., Munagala, K.: 'Approximation algorithms for partial-information based stochastic control with markovian rewards'. Proc. IEEE Symp. on Foundations of Computer Science (FOCS), Providence, RI, October 2007
12 Guha, S., Munagala, K.: 'Approximation algorithms for restless bandit problems'. Proc. ACM-SIAM Symp. on Discrete Algorithms (SODA), New York, January 2009
13 Bertsimas, D., Nino-Mora, J.E.: 'Restless bandits, linear programming relaxations, and a primal-dual heuristic', *Oper. Res.*, 2000, **48**, (1), pp. 80–90
14 Zhao, Q., Krishnamachari, B., Liu, K.: 'On myopic sensing for multi-channel opportunistic access: Structure, optimality, and performance', *IEEE Trans. Wirel. Commun.*, 2008, **7**, (3), pp. 5413–5440
15 Ahmand, S., Liu, M., Javidi, T., zhao, Q., Krishnamachari, B.: 'Optimality of myopic sensing in multichannel opportunistic access', *IEEE Trans. Inf. Theory*, 2009, **55**, (9), pp. 4040–4050
16 Ahmad, S., Liu, M.: 'Multi-channel opportunistic access: a case of restless bandits with multiple plays'. Allerton Conf., Monticello, Il, September–October 2009
17 Liu, K., Zhao, Q., Krishnamachari, B.: 'Dynamic multichannel access with imperfect channel state detection', *IEEE Trans. Signal Process*, 2010, **58**, (5), pp. 2795–2807
18 Wang, K., Chen, L.: 'On optimality of myopic policy for restless multi-armed bandit problem: An axiomatic approach', *IEEE Trans. Signal Process*, 2012, **60**, (1), pp. 300–309
19 Wang, K., Liu, Q., Chen, L.: 'On optimality of greedy policy for a class of standard reward function of restless multi-armed bandit problem', *IET Signal Process.*, 2011, **6**, (6), pp. 584–593
20 Lapiccirella, F.E., Liu, K., Ding, Z.: 'Multi-channel opportunistic access based on primary arq messages overhearing'. Proc. of IEEE ICC 2011, Kyoto, June 2011
21 Wang, K., Liu, Q., Lau, F.C.M.: 'Multi-channel opportunistic access by overhearing primary ARQ messages', *IEEE Trans. Veh. Technol.*, 2013, **62**, (7), pp. 3486–3492
22 Wang, K., Chen, L., Liu, Q.: 'On optimality of myopic policy for restless multi-armed bandit problem with non i.i.d. arms and imperfect detection', *IEEE Trans. Veh. Technol.*, 2014, **63**, (5), pp. 2478–2483

8 Appendix: Proof of Lemma 8

We prove the lemma by backward induction. For slot T, we have

(1) For $l \in \mathcal{A}'$, $l \in \mathcal{A}$, it holds that $W_T^{\mathcal{A}'}(\Omega_l') - W_T^{\mathcal{A}}(\Omega_l) = (r_1 - r_0)(\omega_l' - \omega_l)$;
(2) For $l \notin \mathcal{A}'$, $l \notin \mathcal{A}$, it holds that $W_T^{\mathcal{A}'}(\Omega_l') - W_T^{\mathcal{A}}(\Omega_l) = 0$;
(3) For $l \in \mathcal{A}'$, $l \notin \mathcal{A}$, it exists at least one channel m such that $\omega'_l \geq \omega_m \geq \omega_l$. It then holds that $0 \leq W_T^{\mathcal{A}'}(\Omega_l') - W_T^{\mathcal{A}}(\Omega_l) \leq \Delta_{\max}(\omega_l' - \omega_l)$.

Therefore Lemma 8 holds for slot T.

Assume that Lemma 8 holds for $T-1, \ldots, t+1$, then we prove the lemma for slot t.

We first prove the first case: $l \in \mathcal{A}'$, $l \in \mathcal{A}$. By rewriting $\Gamma(\Omega(t))$ in (3) and developing $\omega_l(t+1)$ in $\Omega(t+1)$, we have:

$$\Gamma^{\mathcal{A}'}(\Omega_l') = \sum_{\mathcal{E} \subseteq \mathcal{A}(t),\, \{l\}} C^{\mathcal{E}}_{\mathcal{A},\, \{l\}} [f_l^1(\omega_l'(t)) W_{t+1}(\Omega_{-l}, \tau_l(\phi_l(\omega_l'(t)))) + f_l^0(\omega_l'(t)) W_{t+1}(\Omega_{-l}, \tau_l(\varphi_l(\omega_l'(t))))] \quad (5)$$

$$\Gamma^{\mathcal{A}}(\Omega_l) = \sum_{\mathcal{E} \subseteq \mathcal{A}(t),\, \{l\}} C^{\mathcal{E}}_{\mathcal{A},\, \{l\}} [f_l^1(\omega_l(t)) W_{t+1}(\Omega_{-l}, \tau_l(\phi_l(\omega_l(t)))) + f_l^0(\omega_l(t)) W_{t+1}(\Omega_{-l}, \tau_l(\varphi_l(\omega_l(t))))] \quad (6)$$

Furthermore, we have (see (7))

where $f_l^0(\omega_l) \geq f_l^0(\omega_l')$ from Lemma 1.

Next, we analyse the first term $W_{t+1}(\Omega_{-l}, \tau_l(\phi_l(\omega_l'))) - W_{t+1}(\Omega_{-l}, \tau_l(\phi_l(\omega_l)))$ of RHS of (7) through three cases

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

$$0 \leq \left| W_{t+1}(\Omega_{-l}, \tau_l(\phi_l(\omega_l'(t)))) - W_{t+1}(\Omega_{-l}, \tau_l(\phi_l(\omega_l(t)))) \right| \leq \left| \tau_l(\phi_l(\omega_l'(t))) - \tau_l(\phi_l(\omega_l(t))) \right| \Delta_{\max} \sum_{i=0}^{T-t-1} \beta^i \delta^i$$

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

$$0 \leq \left| W_{t+1}(\Omega_{-l}, \tau_l(\phi_l(\omega_l'))) - W_{t+1}(\Omega_{-l}, \tau_l(\phi_l(\omega_l))) \right| \leq \left| \tau_l(\phi_l(\omega_l')) - \tau_l(\phi_l(\omega_l)) \right| \Delta_{\max} \sum_{i=1}^{T-t-1} \beta^i \delta^i$$

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

$$0 \leq \left| W_{t+1}(\Omega_{-l}, \tau_l(\phi_l(\omega_l'))) - W_{t+1}(\Omega_{-l}, \tau_l(\phi_l(\omega_l))) \right| \leq \left| \tau_l(\phi_l(\omega_l')) - \tau_l(\phi_l(\omega_l)) \right| \Delta_{\max} \sum_{i=0}^{T-t-1} \beta^i \delta^i$$

Combining Case 1–3, we obtain the bounds of the first term of (7) as follows

$$0 \leq \left| W_{t+1}(\Omega_{-l}, \tau_l(\phi_l(\omega_l'))) - W_{t+1}(\Omega_{-l}, \tau_l(\phi_l(\omega_l))) \right| \leq \left| \tau_l(\phi_l(\omega_l')) - \tau_l(\phi_l(\omega_l)) \right| \Delta_{\max} \sum_{i=0}^{T-t-1} \beta^i \delta^i \quad (8)$$

Further, we can obtain the bounds of the second and third terms of

$$\Gamma^{\mathcal{A}'}(\Omega_l') - \Gamma^{\mathcal{A}}(\Omega_l) = \sum_{\mathcal{E} \subseteq \mathcal{A}(t),\, \{l\}} C^{\mathcal{E}}_{\mathcal{A},\, \{l\}} \Big\{ f_l^1(\omega_l) [W_{t+1}(\Omega_{-l}, \tau_l(\phi_l(\omega_l'))) - W_{t+1}(\Omega_{-l}, \tau_l(\phi_l(\omega_l)))] + (f_l^0(\omega_l) - f_l^0(\omega_l')) \cdot [W_{t+1}(\Omega_{-l}, \tau_l(\phi_l(\omega_l'))) - W_{t+1}(\Omega_{-l}, \tau_l(\varphi_l(\omega_l)))] + f_l^0(\omega_l') [W_{t+1}(\Omega_{-l}, \tau_l(\varphi_l(\omega_l'))) - W_{t+1}(\Omega_{-l}, \tau_l(\varphi_l(\omega_l)))] \Big\} \quad (7)$$

RHS of (7) by the similar induction as follows

$$0 \le \left|W_{t+1}(\Omega_{-l}, \tau_l(\phi_l(\omega_l'))) - W_{t+1}(\Omega_{-l}, \tau_l(\varphi_l(\omega_l)))\right| \le \left|\tau_l(\phi_l(\omega_l')) - \tau_l(\varphi_l(\omega_l))\right|\Delta_{\max}\sum_{i=0}^{T-t-1}\beta^i\delta^i \quad (9)$$

$$0 \le \left|W_{t+1}(\Omega_{-l}, \tau_l(\varphi_l(\omega_l'))) - W_{t+1}(\Omega_{-l}, \tau_l(\varphi_l(\omega_l)))\right| \le \left|\tau_l(\varphi_l(\omega_l')) - \tau_l(\varphi_l(\omega_l))\right|\Delta_{\max}\sum_{i=0}^{T-t-1}\beta^i\delta^i \quad (10)$$

Therefore combining (7)–(10) and $|p_{11}^{(l)} - p_{01}^{(l)}| \le \delta$, we have

$$\begin{aligned}\left|W_t^{\mathcal{A}'}(\Omega_l') - W_t^{\mathcal{A}}(\Omega_l)\right| &\le \left|\Delta_{\max}(\omega_l' - \omega_l)\right| + \left|\beta(\Gamma^{\mathcal{A}'}(\Omega_l') - \Gamma^{\mathcal{A}}(\Omega_l))\right| \\ &\le \Delta_{\max}(\omega_l' - \omega_l) + \beta\Delta_{\max}(\omega_l' - \omega_l)\delta\sum_{i=0}^{T-t-1}\beta^i\delta^i \\ &= (\omega_l' - \omega_l)\Delta_{\max}\sum_{i=0}^{T-t}\beta^i\delta^i\end{aligned}$$

and

$$\begin{aligned}\left|W_t^{\mathcal{A}'}(\Omega_l') - W_t^{\mathcal{A}}(\Omega_l)\right| &\ge \left|\Delta_{\min}(\omega_l' - \omega_l)\right| - \left|\beta(\Gamma^{\mathcal{A}'}(\Omega_l') - \Gamma^{\mathcal{A}}(\Omega_l))\right| \\ &\ge \Delta_{\min}(\omega_l' - \omega_l) - \beta\Delta_{\max}(\omega_l' - \omega_l)\delta\sum_{i=0}^{T-t-1}\beta^i\delta^i \\ &= (\omega_l' - \omega_l)\left(\Delta_{\min} - \Delta_{\max}\sum_{i=1}^{T-t}\beta^i\delta^i\right)\end{aligned}$$

To the end, we complete the proof of the first part, $l \in \mathcal{A}'$, $l \in \mathcal{A}$, of Lemma 8.

Secondly, we prove the second case $l \notin \mathcal{A}'$, $l \notin \mathcal{A}$. In this case, $\mathcal{A}'(t) = \mathcal{A}(t)$. Assuming $k \in \mathcal{A}(t)$, we have:

$$\Gamma^{\mathcal{A}'}(\Omega_l') = \sum_{\mathcal{E}\subseteq\mathcal{A}(t),\{k\}} C^{\mathcal{E}}_{\mathcal{A},\{k\}}\Big[f_k^1(\omega_k(t))W_{t+1}(\Omega_{-k}', \tau_k(\phi_k(\omega_k(t)))) + f_k^0(\omega_k(t))W_{t+1}(\Omega_{-k}', \tau_k(\varphi_k(\omega_k(t))))\Big] \quad (11)$$

$$\Gamma^{\mathcal{A}}(\Omega_l) = \sum_{\mathcal{E}\subseteq\mathcal{A}(t),\{k\}} C^{\mathcal{E}}_{\mathcal{A},\{k\}}\Big[f_k^1(\omega_k(t))W_{t+1}(\Omega_{-k}, \tau_k(\phi_k(\omega_k(t)))) + f_k^0(\omega_k(t))W_{t+1}(\Omega_{-k}, \tau_k(\varphi_k(\omega_k(t))))\Big] \quad (12)$$

Thus (see (13))

$$\Gamma^{\mathcal{A}'}(\Omega_l') - \Gamma^{\mathcal{A}}(\Omega_l) = \sum_{\mathcal{E}\subseteq\mathcal{A}(t),\{k\}} C^{\mathcal{E}}_{\mathcal{A},\{k\}}\Big[f_k^1(\omega_k)[W_{t+1}(\Omega_{-k}', \tau_k(\phi_k(\omega_k))) - W_{t+1}(\Omega_{-k}, \tau_k(\phi_k(\omega_k)))] + f_k^0(\omega_k)[W_{t+1}(\Omega_{-k}', \tau_k(\varphi_k(\omega_k))) - W_{t+1}(\Omega_{-k}, \tau_k(\varphi_k(\omega_k)))]\Big] \quad (13)$$

For the first term of RHS of (13), if channel l is never chosen for $W_{t+1}(\Omega'_{-k}, \tau_k(\phi_k(\omega_k)))$ and $W_{t+1}(\Omega_{-k}, \tau_k(\phi_k(\omega_k)))$ from the slot $t+1$ to the end of time horizon of interest T, that is to say, $l \notin \mathcal{A}'(r)$ and $l \notin \mathcal{A}(r)$ for $t+1 \le r \le T$, it is easy to know $W_{t+1}(\Omega'_{-k}, \tau_k(\phi_k(\omega_k))) - W_{t+1}(\Omega_{-k}, \tau_k(\phi_k(\omega_k))) = 0$. Otherwise, it exists t^0 $(t+1 \le t^0 \le T)$ such that the following three cases hold.

Case 1: $l \notin \mathcal{A}'(r)$ and $l \notin \mathcal{A}(r)$ for $t \le r \le t^0 - 1$ while $l \in \mathcal{A}'(t^0)$ and $l \in \mathcal{A}(t^0)$;
Case 2: $l \notin \mathcal{A}'(r)$ and $l \notin \mathcal{A}(r)$ for $t \le r \le t^0 - 1$ while $l \notin \mathcal{A}'(t^0)$ and $l \in \mathcal{A}(t^0)$;
Case 3: $l \notin \mathcal{A}'(r)$ and $l \notin \mathcal{A}(r)$ for $t \le r \le t^0 - 1$ while $l \in \mathcal{A}'(t^0)$ and $l \notin \mathcal{A}(t^0)$.

For Case 1, according to the hypothesis ($l \in \mathcal{A}'$ and $l \in \mathcal{A}$), we have

$$\begin{aligned}&\left|W_{t^0}(\Omega_l'(t^0)) - W_{t^0}(\Omega_l(t^0))\right| \\ &\le \Delta_{\max}\left|\omega_l'(t^0) - \omega_l(t^0)\right|\sum_{i=0}^{T-t^0}\beta^i\delta^i \\ &= \Delta_{\max}\left|(p_{11}^{(l)} - p_{01}^{(l)})^{t^0-t}(\omega_l'(t) - \omega_l(t))\right|\sum_{i=0}^{T-t^0}\beta^i\delta^i\end{aligned}$$

Considering that $|(p_{11}^{(l)} - p_{01}^{(l)})^{t^0-t}|$ and $\sum_{i=0}^{T-t^0}\beta^i\delta^i$ are decreasing with t^0 ($t^0 \ge t+1$), thus,

$$\begin{aligned}&\left|W_{t+1}(\Omega_l'(t+1)) - W_{t+1}(\Omega_l(t+1))\right| \\ &\le (r_1 - r_0)\left|p_{11}^{(l)} - p_{01}^{(l)}\right|(\omega_l'(t) - \omega_l(t))\sum_{i=0}^{T-t-1}\beta^i\delta^i\end{aligned}$$

For Case 2–3, by the induction hypothesis ($l \in \mathcal{A}'$, $l \notin \mathcal{A}$ or $l \in \mathcal{A}$, $l \notin \mathcal{A}'$), we have the similar results with Case 1.

Combing the results of the three cases, we obtain

$$\begin{aligned}&\left|W_{t+1}(\Omega'_{-k}, \tau_k(\phi_k(\omega_k))) - W_{t+1}(\Omega_{-k}, \tau_k(\phi_k(\omega_k)))\right| \\ &\le \Delta_{\max}\left|p_{11}^{(l)} - p_{01}^{(l)}\right|(\omega_l'(t) - \omega_l(t))\sum_{i=0}^{T-t-1}\beta^i\delta^i\end{aligned}$$

For the second term of RHS of (13), we can obtain the similar result.

Combing the bounds of the above two terms of RHS of (13), we have

$$\begin{aligned}\left|W_t^{\mathcal{A}'}(\Omega_l') - W_t^{\mathcal{A}}(\Omega_l)\right| &= \left|\beta(\Gamma^{\mathcal{A}'}(\Omega_l') - \Gamma^{\mathcal{A}}(\Omega_l))\right| \\ &\le \Delta_{\max}(\omega_l'(t) - \omega_l(t))\sum_{i=1}^{T-t}\beta^i\delta^i\end{aligned}$$

which completes the proof of Lemma 8 when $l \notin \mathcal{A}'$ and $l \notin \mathcal{A}$.

Finally, we prove the third case $l \in \mathcal{A}'(t)$ and $l \notin \mathcal{A}(t)$, then it exists at least one channel, denoted as ω_m, such that $\omega'_l \ge \omega_m \ge \omega_l$. We have

$$\begin{aligned}&W_t^{\mathcal{A}'}(\Omega_l'(t)) - W_t^{\mathcal{A}}(\Omega_l(t)) \\ &= W_t^{\mathcal{A}'}(\omega_1, \ldots, \omega_l', \ldots, \omega_N) - W_t^{\mathcal{A}}(\omega_1, \ldots, \omega_l, \ldots, \omega_N) \\ &= W_t^{\mathcal{A}'}(\omega_1, \ldots, \omega_l', \ldots, \omega_N) - W_t^{\mathcal{A}'}(\omega_1, \ldots, \omega_l = \omega_m, \ldots, \omega_N) \\ &\quad + W_t^{\mathcal{A}}(\omega_1, \ldots, \omega_l = \omega_m, \ldots, \omega_N) - W_t^{\mathcal{A}}(\omega_1, \ldots, \omega_l, \ldots, \omega_N)\end{aligned} \quad (14)$$

According to the induction hypothesis ($l \in \mathcal{A}'$ and $l \in \mathcal{A}$), the first

IET Commun., 2015, Vol. 9, Iss. 7, pp. 1017–1025

term of the RHS of (14) can be bounded as follows

$$
\begin{aligned}
0 &\le (\omega_l'(t) - \omega_m(t))\left(\Delta_{\min} - \Delta_{\max}\sum_{i=1}^{T-t}\beta^i\delta^i\right)\\
&\le W_t^{\mathcal{A}'}(\omega_1, \ldots, \omega_l', \ldots, \omega_N) - W_t^{\mathcal{A}'}(\omega_1, \ldots, \omega_l = \omega_m, \ldots, \omega_N)\\
&\le (\omega_l'(t) - \omega_m(t))\Delta_{\max}\sum_{i=0}^{T-t}\beta^i\delta^i
\end{aligned}
\tag{15}
$$

Meanwhile, the second term of the RHS of (14) is inducted by hypothesis ($l \notin \mathcal{A}'$ and $l \notin \mathcal{A}$):

$$
\begin{aligned}
&-(\omega_m(t) - \omega_l(t))\Delta_{\max}\sum_{i=1}^{T-t}\beta^i\delta^i\\
&\le W_t^{\mathcal{A}}(\omega_1, \ldots, \omega_l = \omega_m, \ldots, \omega_N) - W_t^{\mathcal{A}}(\omega_1, \ldots, \omega_l, \ldots, \omega_N)\\
&\le (\omega_m(t) - \omega_l(t))\Delta_{\max}\sum_{i=1}^{T-t}\beta^i\delta^i
\end{aligned}
\tag{16}
$$

Therefore we have, combining (14), (15) and (16)

$$
\begin{aligned}
&(\omega_l'(t) - \omega_m(t))\Delta_{min} - (\omega_l'(t) - \omega_l(t))\Delta_{\max}\sum_{i=1}^{T-t}\beta^i\delta^i\\
&\le W_t^{\mathcal{A}'}(\Omega_l'(t)) - W_t^{\mathcal{A}}(\Omega_l(t))\\
&\le (\omega_l'(t) - \omega_l(t))\Delta_{\max}\sum_{i=0}^{T-t}\beta^i\delta^i - (\omega_m(t) - \omega_l(t))\Delta_{\max}\sum_{i=1}^{T-t}\beta^i\delta^i
\end{aligned}
$$

further

$$
\left|W_t^{\mathcal{A}'}(\Omega_l'(t)) - W_t^{\mathcal{A}}(\Omega_l(t))\right| \le (\omega_l'(t) - \omega_l(t))\Delta_{\max}\sum_{i=0}^{T-t}\beta^i\delta^i
$$

Thus, we complete the proof of the third part, $l \in \mathcal{A}'(t)$ and $l \notin \mathcal{A}(t)$, of Lemma 8.

To the end, Lemma 8 is concluded.