Opportunistic Multichannel Access with Imperfect Observation: A Fixed Point Analysis on Indexability and Index-based Policy

Kehao Wang*§, Lin Chen†, Jihong Yu‡, Moe Win§

*Wuhan University of Technology, Wuhan, China

†Université de Paris-Sud XI, Orsay, France

‡Department of Computer Science, Simon Fraser University, Canada

§Laboratory of Information & Decision Systems, Massachusetts Institute of Technology, Cambridge, USA

kehao.wang@whut.edu.cn, lin.chen@lri.fr, jihongy@sfu.ca, moewin@mit.edu

***Abstract*—We consider the multichannel opportunistic access problem, in which a user decides, at each time slot, which channel to access among multiple Gilbert-Elliot channels in order to maximize his aggregated utility (e.g., the expected transmission throughput) given that the observation of channel state is error-prone. The problem can be cast into a restless multiarmed bandit problem which is proved to be PSPACE-Hard. An alternative approach, given the problem hardness, is to look for simple channel access policies. Whittle index policy is a very popular heuristic for restless bandits, which is provably optimal asymptotically and has good empirical performance. In the case of imperfect observation, the traditional approach of computing the Whittle index policy cannot be applied because the channel state belief evolution is no more linear, thus rendering the indexability of our problem open. In this paper, we mathematically establish the indexability and establish the closed-form Whittle-index, based on which index policy can be constructed. The major technique in our analysis is a fixed point based approach which enable us to divide the belief information space into a series of regions and then establish a set of periodic structures of the underlying nonlinear dynamic evolving system, based on which we devise the linearization scheme for each region to establish indexability and compute the Whittle index for each region.**

***Index Terms*—Restless bandit; Whittle index; indexability; fixed point; nonlinear operator**

I. INTRODUCTION

A. Background

We consider an opportunistic multichannel communication system with heterogenous Gilbert-Elliot channels [3], in which a user is limited to sense and transmit only on one channel each time due to limitation on sensing capability[1]. Given that channel sensing in practice is not perfect, the fundamental optimization problem we address in this paper is how the user exploits the imperfect sensing results and the stochastic properties of channels to maximize its utility (e.g., expected throughput) by switching among channels opportunistically.

B. Related Works

The opportunistic channel access can be cast into a restless multiarmed bandit (RMAB) problem, which is proved to be PSPACE-hard [1]. To the best of our knowledge, very few results are reported on the structure of the optimal policy of a generic RMAB due to its high complexity.

The myopic strategy, due to its simple and tractable structure, has recently attracted extensive research attention. It essentially consists of sensing the channels that maximize the expected immediate reward while ignoring the impact of the current decision on future reward. Along this research thrust, the optimality of the myopic policy is partially established for the homogeneous Gilbert-Elliot channel case under perfect sensing [1]. In [2], the authors studied the case of heterogeneous channels and derived a set of closed-form sufficient conditions to guarantee the optimality of the myopic policy. In [1], the authors proposed a sufficient condition framework for the optimality of myopic policy. In [1], the authors gave the sufficient conditions for multi-state channels. For the imperfect sensing Gilbert-Elliot channels, [8] proved the optimality of the myopic policy for the specifical case of two channels. In [2, 2, 2], the authors derived closed-form condition to guarantee the optimality of the myopic policy for arbitrary number of channels.

Generally speaking, the structure of the optimum access policy is only characterized for a subset of parameter space under which the myopic policy is proved optimum. Beyond this parameter space, we need to turn to a more generic policy, Whittle index policy, introduced by P. Whittle in [2]. The Whittle index policy has been a very popular heuristic for restless bandit, which, while suboptimal in general, is provably optimal in asymptotic sense [2, 2] and has good empirical performance. The Whittle index policy and its variants have been studied extensively in engineering applications, e.g., sensor scheduling [1, 1], multi-UAV coordination [1], crawling web content [9], channel allocation in wireless networks [1, 7], and job scheduling [1, 2]. More comprehensive treatments of indexable restless bandits can be found in [4–6].

C. Our Work and Contributions

The central pivot in the Whittle index policy analysis is to establish the indexability of the problem and compute the

[1]The technical analysis in this paper can be extended to address the case where a user is allowed to sense a fixed number of channels.

corresponding index. In our problem, for a subset of specific scenarios characterized by the corresponding parameter spaces (e.g., [2, 2]), the Whittle index policy degenerates to the myopic policy. However, beyond those scenarios, the structure of the index-based policy is still open, which is the focus of this paper (cf. **Table I**).

TABLE I
SUMMARY OF RELATED WORK AND THIS PAPER

parameter domain	policy	optimality
$p_{11} \geqslant p_{01}, \epsilon \leqslant \frac{p_{01}(1-p_{11})}{p_{11}(1-p_{01})}$	myopic	optimal [2]
$p_{11} \leqslant p_{01}, \epsilon \leqslant \frac{p_{11}(1-p_{01})}{p_{01}(1-p_{11})}$	myopic	optimal [2]
$p_{01}^{(i)} \geqslant p_{11}^{(i)},\ \epsilon_i \leqslant \frac{(1-p_{11}^{(i)})\cdot p_{01}^{(i)}}{(1-p_{01}^{(i)})\cdot p_{11}^{(i)}}$	index	this paper

The major technical challenge to establish the indexability in our problem comes from the imperfect sensing, where the false alarm rate is involved in the propagation of belief information and makes the value function no longer linear as in existing studies. As a result, the traditional approach of computing the Whittle index cannot be used in this context. To the best of our knowledge, there does not exist a closed-form Whittle index for the nonlinear case; only numerical simulation is conducted under a strict assumption on the indexability [1].

To address the challenge caused by nonlinearity, we investigate the fixed points of belief evolution function (which is non-linear), based on which we establish a set of periodic structures of the resulting dynamic system. We then use the derived properties to linearize the value function by a piecewise approach to prove the Whittle indexability and derive the closed-form Whittle index. Our results in this paper thus solves the multi-channel opportunistic scheduling problem under imperfect channel sensing by establishing its indexability and constructing the corresponding index policy. Due to the generality of the problem, our results can be applied in a wide range of engineering applications where the underlying optimization problems can be cast into restless bandits with imperfect sensing of bandit states. Therefore, the terminology and analysis in this paper should be understood generically.

II. System Model

We consider a time-slotted multi-channel opportunistic communication system, in which a user is able to access a set $\mathcal{N}$ of N independent channels, each characterized by a Markov chain of two states, *good* (1) and *bad* (0). The channel state transition matrix $\mathbf{P}^{(i)}$ for channel i ($i \in \mathcal{N}$) is given as follows

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

We assume that channels go through state transition at the beginning of each slot t. The system operates in a synchronously time slotted fashion with the time slot indexed by t ($t = 0, 1, \cdots$).

Due to hardware constraints and energy cost, the user is allowed to sense only one of the N channels at each slot t. We assume that the user makes the channel selection decision at the beginning of each slot after the channel state transition. Once a channel is selected, the user detects the channel state $S_i(t)$, which can be considered as a binary hypothesis test:

$$\mathcal{H}_0 : S_i(t) = 1 \text{ (good)} \quad vs. \quad \mathcal{H}_1 : S_i(t) = 0 \text{ (bad)}.$$

The performance of channel i state detection is characterized by the probability of false alarm ϵ_i and the probability of miss detection δ_i:

$$\begin{aligned} \epsilon_i &:= \mathrm{P}\{\text{decide } \mathcal{H}_1 \mid \mathcal{H}_0 \text{ is true }\}, \\ \delta_i &:= \mathrm{P}\{\text{decide } \mathcal{H}_0 \mid \mathcal{H}_1 \text{ is true }\}. \end{aligned}$$

Based on the imperfect detection outcome in slot t, the user determines whether to access channel i for transmission. We denote the action on channel n made by the user at slot t by $a_n(t)$, i.e.,

$$a_n(t) = \begin{cases} 1, & \text{if channel } n \text{ is chosen in slot } t, \\ 0, & \text{if channel } n \text{ is not chosen in slot } t. \end{cases}$$

Thus, $\sum_{n=1}^{N} a_n(t) = 1$ for all t, indicating that exactly one channel is chosen in each slot.

Since failed transmissions may occur, acknowledgements (ACKs) are necessary to ensure guaranteed delivery. Specifically, when the receiver successfully receives a packet from a channel, it sends an acknowledgement to the transmitter over the same channel at the end of slot. Otherwise, the receiver does nothing, i.e., a NAK is defined as the absence of an ACK, which occurs when the transmitter did not transmit over this channel or transmitted but the channel is busy in this slot. We assume that acknowledgements are received without error since acknowledgements are always transmitted over idle channels.

Obviously, by imperfectly sensing only one of N channels, the user cannot observe the state information of the whole system. Hence, the user has to infer the channel states from its decision history and observation history so as to make its future decision. To this end, we define the *channel state belief vector* (hereinafter referred to as *belief vector* for briefness) $\mathbf{w}(t) \triangleq \{\omega_i(t), i \in \mathcal{N}\}$, where $0 \le \omega_i(t) \le 1$ is the conditional probability that channel i is in state good (i.e., $S_i(t) = 1$) conditioned on the decision history and observation history.

To ensure that the user and its intended receiver tune to the same channels in each slot, channel selections should be based on common observation: $K(t) \in \{0 \text{ (NAK)}, 1 \text{ (ACK)}\}$ in each slot rather than the detection outcome at the transmitter.

Given the sensing action $\{a_i(t)\}_{i\in\mathcal{N}}$ and the observation $K(t)$, the belief vector in $t+1$ slot can be updated recursively using Bayes Rule as shown in (1):

$$\omega_i(t+1) = \begin{cases} p_{11}^{(i)}, & a_i(t) = 1, K(t) = 1 \\ \Gamma_i(\omega_i(t)), & a_i(t) = 1, K(t) = 0 \\ \mathcal{T}_i(\omega_i(t)), & a_i(t) = 0 \end{cases} \qquad (1)$$

where,

$$\mathcal{T}_i(\omega_i(t)) := \omega_i(t)p_{11}^{(i)} + (1-\omega_i(t))p_{01}^{(i)}, \quad (2)$$

$$\varphi_i(\omega_i(t)) := \frac{\epsilon_i \omega_i(t)}{1-(1-\epsilon_i)\omega_i(t)}, \quad (3)$$

$$\Gamma_i(\omega_i(t)) := \mathcal{T}_i(\varphi_i(\omega_i(t))). \quad (4)$$

We would like to emphasize that the sensing error introduces technical complications in the system dynamics (i.e., $\varphi_i(\omega_i(t))$) due to its nonlinearity. Therefore, the analysis methods and results [1, 1, 2] in the perfect sensing case where the belief evolution is linear cannot be applied to the scenario with sensing error.

III. Problem Formulation

In this section, we formulate the optimisation problem of opportunistic multichannel access faced by the user. Mathematically, let $\pi = \{\pi(t)\}_{t\geqslant 0}$ denote the sensing policy, with $\pi(t)$ defined as a mapping from the belief vector $\mathbf{w}(t)$ to the action of sensing one channel in each slot t:

$$\pi(t): \ \mathbf{w}(t) \to \{1, 2, \cdots, N\}, \ t = 0, 1, \cdots. \quad (5)$$

Let

$$a_n^{\pi}(t) = \begin{cases} 1, & \text{if channel } n \text{ is chosen under } \pi(t), \\ 0, & \text{if channel } n \text{ is not chosen under } \pi(t). \end{cases} \quad (6)$$

Let $\Pi_n := \{a_n^{\pi}(t): \ t \geqslant 0\}$ be policy space on channel n under the sensing policy π, then $\Pi = \cup_{n=1}^{N} \Pi_n$ is the joint policy space.

We are interested in the user's optimization problem to find the optimal sensing policy π^* that maximizes the expected total discounted reward over an infinite horizon. The following gives the formal definition of the optimal sensing problem:

$$\textbf{(OP):} \quad \max_{\pi\in\Pi} \mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t \sum_{n=1}^{N} \left(a_n^{\pi}(t)(1-\epsilon_n)\omega_n(t)\right)\right] \quad (7)$$

$$\text{s.t.} \quad \sum_{n=1}^{N} a_n^{\pi}(t) = 1, \quad t = 0, 1, \cdots, \infty. \quad (8)$$

In the following, we decompose Problem **(OP)** into N similar subproblems by relaxing the constraint (8),

$$\textbf{(SP):} \ \max_{\pi_n \in \Pi_n} \mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t \left(a_n^{\pi_n}(t)(1-\epsilon_n)\omega_n(t) + \nu(1 - a_n^{\pi_n}(t))\right)\right]. \quad (9)$$

To solve the original optimisation problem **(OP)**, we first seek the optimal policy π_n^* for subproblem n ($n \in \mathcal{N}$), and then construct a feasibly approximation policy $\pi = (\pi_1^*, \pi_2^*, \cdots, \pi_N^*)$ for the original problem **(P)**.

IV. Technical Preliminary: Indexability and Whittle Index

Let $V_{\beta,\nu}(\omega)$ be the value function corresponding to the subproblem (9), which denotes the maximum discounted reward accrued from a single-armed bandit process with subsidy ν when the initial belief state is $\omega := \{\omega_i(0)\}_{i\in\mathcal{N}}$.

Considering the two possible actions in each slot, we have

$$V_{\beta,\nu}(\omega) = \max\Big\{V_{\beta,\nu}(\omega; a=0), V_{\beta,\nu}(\omega; a=1)\Big\}, \quad (10)$$

where

$$V_{\beta,\nu}(\omega; a=0) = \nu + \beta V_{\beta,\nu}(\mathcal{T}(\omega)),$$

$$V_{\beta,\nu}(\omega; a=1) = (1-\epsilon)\omega + \beta\Big[(1-\epsilon)\omega V_{\beta,\nu}(p_{11}) + (1-(1-\epsilon)\omega)V_{\beta,\nu}(\Gamma(\omega))\Big].$$

$V_{\beta,\nu}(\omega; a=1)$ denotes the reward obtained by taking action a in the first slot following by the optimal policy in future slots, and $V_{\beta,\nu}(\omega; a=0)$ denotes the sum of the subsidy ν obtained in the current slot under the passive action ($a=0$) and the total discounted future reward $\beta V_{\beta,\nu}(\mathcal{T}(\omega))$.

Remark. In an infinite time horizon, a decision should be made at each slot, and the different decision leads to different evolution of belief information ω. Thus, in the following, we call (10) a dynamic system without introducing ambiguity.

Remark. We would like to point out that $V_{\beta,\nu}(\Gamma(\omega))$ (specifically $\varphi(\omega)$) brings about the nonlinear belief update of the dynamic system (10), and leads to the complicated characteristics of the Whittle index.

The optimal action a^* for the belief state ω under subsidy ν is given by

$$a^* = \begin{cases} 1, & \text{if } V_{\beta,\nu}(\omega; a=1) > V_{\beta,\nu}(\omega; a=0) \\ 0, & \text{otherwise.} \end{cases} \quad (11)$$

We define the passive set $\mathcal{P}(\nu)$ under subsidy ν as

$$\mathcal{P}(\nu) := \Big\{\omega : V_{\beta,\nu}(\omega; a=1) \leqslant V_{\beta,\nu}(\omega; a=0)\Big\}. \quad (12)$$

We next introduce some definitions related to the indexability of our problem.

Definition 1 (Indexability)**.** Problem (9) is indexable if the passive set $\mathcal{P}(\nu)$ of the corresponding single-armed bandit process with subsidy ν monotonically increases from $\emptyset$ to the whole state space $[0, 1]$ as ν increases from $-\infty$ to $+\infty$.

Under the indexability condition, Whittle index is defined as follows:

Definition 2 (Whittle index [2])**.** If Problem (9) is indexable, its Whittle index $W(\omega)$ of the state ω is the infimum subsidy ν such that it is optimal to make the arm passive at ω. Equivalently, Whittle index is the infimum subsidy ν that makes the passive and active actions equally rewarding

$$W(\omega) = \inf\Big\{\nu : V_{\beta,\nu}(\omega; a=1) \leqslant V_{\beta,\nu}(\omega; a=0)\Big\}. \quad (13)$$

Definition 3 (Threshold Policy)**.** Given a certain ν, there exists ω^* $(0 \leqslant \omega^* \leqslant 1)$ such that $V_{\beta,\nu}(\omega^*;1) = V_{\beta,\nu}(\omega^*;0)$. The threshold policy is defined as follows

1) $a^* = 1$ for any ω $(\omega^* < \omega \leqslant 1$ $)$ while $a^* = 0$ for any ω $(0 \leqslant \omega < \omega^*$ $)$, or
2) $a^* = 0$ for any ω $(\omega^* < \omega \leqslant 1$ $)$ while $a^* = 1$ for any ω $(0 \leqslant \omega < \omega^*$ $)$.

Definition 4. Problem (9) is CMI-indexable if the subsidy ν computed by the threshold policy is a continuous and monotonically increasing (CMI) function of ω.

V. Summary of Main Results

In this section, we summarize the main results of our paper. The detailed analysis and proofs of the results will be presented in later sections.

Our central result is the establishment of the CMI-indexability of the opportunistic multichannel access problem, as stated in the theorem below.

Theorem 1. Given $\epsilon_i \leqslant \frac{(1-\max\{p_{11}^{(i)},p_{01}^{(i)}\})\cdot\min\{p_{11}^{(i)},p_{01}^{(i)}\}}{(1-\min\{p_{11}^{(i)},p_{01}^{(i)}\})\cdot\max\{p_{11}^{(i)},p_{01}^{(i)}\}}$ $(\forall i \in \mathcal{N})$, Problem (9) is CMI-indexable.

To prove the indexability, we need to prove the continuity and increasing monotonicity of ν in ω. Thus, we first derive the closed form ν. We then can easily show that ν is continuous and monotonically increasing in ω.

Given the indexability result, we proceed to derive the Whittle index in the following theorem.

Theorem 2. The Whittle index $W_\beta(\omega)$ for channel i is given as follow.

1) The case of negatively correlated channels, i.e., $p_{11}^{(i)} \leqslant p_{01}^{(i)}$: See (20).
2) The case of positively correlated channels, i.e., $p_{11}^{(i)} \geqslant p_{01}^{(i)}$: See (21).

The following corollary bridges our results with existing body of works on myopic policy by showing that in a particular case with stochastically identical channels, the Whittle index-based policy we derive degenerates to the myopic policy.

Corollary 1. $W_\beta(\omega)$ is a monotonically non-decreasing function of ω. As a consequence, the Whittle index policy is equivalent to the myopic (or greedy) policy for the considered RMAB with stochastically identical channels.

For the case of optimizing average reward, i.e., $\beta = 1$, we derive the Whittle index $W(\omega) = \lim_{\beta\to 1} W_\beta(\omega)$ as follows

Theorem 3. The Whittle index $W(\omega)$ for channel i is given as follow.

1) The case of negatively correlated channels, i.e., $p_{11}^{(i)} \leqslant p_{01}^{(i)}$: See (22).
2) The case of positively correlated channels, i.e., $p_{11}^{(i)} \geqslant p_{01}^{(i)}$: See (23).

Based on the Whittle index, we can construct the index-based access policy for Problem (**OP**): the user chooses the channel $i^* = \mathrm{argmax}_{i\in\mathcal{N}} W_\beta(\omega_i)$ for the discounted case and $i^* = \mathrm{argmax}_{i\in\mathcal{N}} W(\omega_i)$ for the case of optimizing the average reward.

The main challenges in obtaining the indexability result in our problem comes from the nonlinear operator $\Gamma_i(\cdot)$, summarized as below:

1) The nonlinear operator $\Gamma_i(\cdot)$ brings about nonlinear propagation of belief information in the evolution of the dynamic system.
2) The value function $V_{\beta,\nu}(\omega)$ is also nonlinear and intractable to compute due to the nonlinearity of $\Gamma_i(\cdot)$.

To address the above challenges, we analyze the fixed points of the operators $\mathcal{T}_i$, Γ_i, as well as their combinations, and divide the belief information space into a series of regions using the fixed points. We then establish a set of periodic structures of the underlying nonlinear dynamic evolving system, based on which we further devise the linearization scheme for each region.

VI. Fixed Point Analysis

In this section, we derive the fixed points of the mappings $\mathcal{T}_i(\cdot)$ and $\Gamma_i(\cdot)$ and their structural properties. To make our analysis concise, we omit the channel index i.

Lemma 1 (Fixed point of $\mathcal{T}(\cdot)$: $p_{01} \leq p_{11}$)**.** Consider the case $p_{01} \leqslant p_{11}$, the following structural properties of $\mathcal{T}(\omega(t))$ hold:

(1) $\mathcal{T}(\omega(t))$ is monotonically increasing in $\omega(t)$;
(2) $p_{01} \leq \mathcal{T}(\omega(t)) \leq p_{11}$, $\forall$ $0 \leq \omega(t) \leq 1$;
(3) $\mathcal{T}^k(\omega(t)) = \mathcal{T}(\mathcal{T}^{k-1}(\omega(t)))$ monotonically converges to $\omega_0 := \frac{p_{01}}{1-(p_{11}-p_{01})}$ as $k \to \infty$.

Proof. Noticing that $\mathcal{T}(\omega(t))$ can be written as $\mathcal{T}(\omega(t)) = (p_{11} - p_{01})\omega(t) + p_{01}$, Lemma 1 holds straightforwardly. □

Lemma 2 (Fixed point of $\mathcal{T}(\cdot)$: $p_{01} > p_{11}$)**.** Consider the case $p_{01} > p_{11}$. Denote $\mathcal{T}^0(\omega) = \omega$ and $\mathcal{T}^k(\omega) = \mathcal{T}(\mathcal{T}^{k-1}(\omega))$, then $\mathcal{T}^{2k}(\omega)$ and $\mathcal{T}^{2k+1}(\omega)$ $(\omega \in [p_{11}, p_{01}])$ converge, from opposite directions, to $\omega_0 := \frac{p_{01}}{1-(p_{11}-p_{01})}$ as $k \to \infty$. In particular, we have

(1) $\mathcal{T}^k(\omega) > \omega$ if $p_{11} \leqslant \omega < \omega_0$;
(2) $\mathcal{T}^k(\omega_0) = \omega_0$;
(3) $\mathcal{T}^k(\omega) \leqslant \omega$ if $\omega_0 \leqslant \omega < p_{01}$.

Proof. It is easy to obtain the lemma, noticing $\mathcal{T}(\omega) = (p_{11} - p_{01})\omega + p_{01}$ and $-1 < p_{11} - p_{01} < 0$. □

Lemma 3. When $\epsilon \leqslant \frac{(1-\max\{p_{11},p_{01}\})\cdot\min\{p_{11},p_{01}\}}{(1-\min\{p_{11},p_{01}\})\cdot\max\{p_{11},p_{01}\}}$, then

(1) $\varphi(\omega(t))$ monotonically increases with $\omega(t)$;
(2) $\varphi(\omega(t)) \leqslant \min\{p_{11}, p_{01}\}$, $\forall \min\{p_{11}, p_{01}\} \leqslant \omega(t) \leqslant \max\{p_{11}, p_{01}\}$;
(3) $\varphi(0) = 0, \varphi(1) = 1$.

Proof. According to (3) and (4), it is easy to obtain the results. □

Lemma 4. Given $p_{01} > p_{11}$. Let $\Gamma(\omega) = \mathcal{T}(\varphi(\omega))$, there exists $\bar{\omega}_0 \in [\mathcal{T}(p_{11}), p_{01}]$ such that

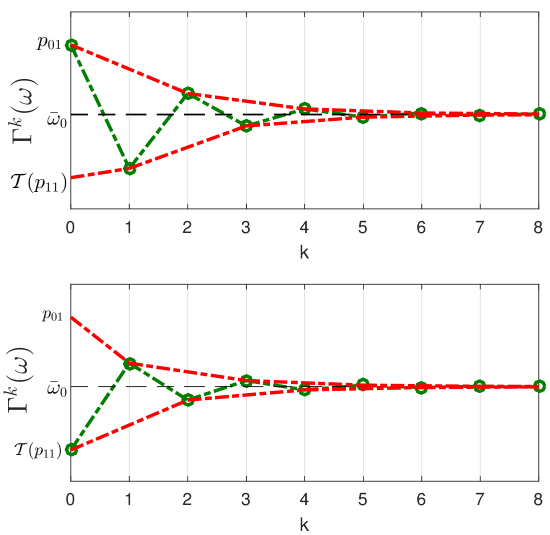

Fig. 1. $\Gamma^k(\omega)$ evolution as k ($p_{11} < p_{01}$). [U]: $\bar{\omega}_0 \leqslant \omega \leqslant p_{01}$; [D]: $\mathcal{T}(p_{11}) \leqslant \omega \leqslant \bar{\omega}_0$. Red line indicates the envelop and greed line indicates the evolution as k.

(1) $\Gamma(\omega) > \omega$, if $\mathcal{T}(p_{11}) \leqslant \omega < \bar{\omega}_0$;
(2) $\Gamma(\bar{\omega}_0) = \bar{\omega}_0$;
(3) $\Gamma(\omega) < \omega$, if $\bar{\omega}_0 \leqslant \omega < p_{01}$.

Lemma 5 (Fixed point of $\Gamma(\cdot)$: $p_{01} > p_{11}$)**.** Let $\Gamma^0(\omega) = \omega$ and $\Gamma^k(\omega) = \Gamma(\Gamma^{k-1}(\omega))$, $\Gamma^{2k}(\omega)$ and $\Gamma^{2k+1}(\omega)$ ($\omega \in [\mathcal{T}(p_{11}), p_{01}]$) converge, from opposite directions, to $\bar{\omega}_0$ as $k \to \infty$ (see **Figure 1**). In particular, we have

(1) $\Gamma^k(\omega) \leqslant \omega$ if $\bar{\omega}_0 \leqslant \omega < p_{01}$;
(2) $\Gamma^k(\bar{\omega}_0) = \bar{\omega}_0$;
(3) $\Gamma^k(\omega) > \omega$ if $\mathcal{T}(p_{11}) \leqslant \omega < \bar{\omega}_0$.

VII. Threshold Policy and Adjoint Dynamic System

In this section, we first express the value function by threshold policy, and then introduce an adjoint dynamic system to facilitate the analysis on nonlinear dynamics.

A. Threshold Policy

Let $L(\omega, \omega')$ be the minimum amount of time required for a passive arm to transit across ω' starting from ω, i.e.,

$$L(\omega, \omega') \triangleq \min\Big\{k : \mathcal{T}^k(\omega) > \omega'\Big\}. \tag{14}$$

According to Lemma 1, we have for the case of $p_{11} \geqslant p_{01}$

$$L(\omega, \omega') = \begin{cases} 0, & \text{if } \omega > \omega' \\ \left\lfloor \log_{p_{11}-p_{01}}^{\frac{\omega_0-\omega'}{\omega_0-\omega}} \right\rfloor + 1, & \text{if } \omega \leqslant \omega' < \omega_0 \\ \infty, & \text{if } \omega \leqslant \omega',\ \omega' \geqslant \omega_0\ , \end{cases} \tag{15}$$

and, for the case of $p_{11} < p_{01}$

$$L(\omega, \omega') = \begin{cases} 0, & \text{if } \omega > \omega' \\ 1, & \text{if } \omega \leqslant \omega' \text{ and } \mathcal{T}(\omega) > \omega' \\ \infty, & \text{if } \omega \leqslant \omega' \text{ and } \mathcal{T}(\omega) \leqslant \omega'. \end{cases} \tag{16}$$

Under the threshold policy, the arm will be activated if its belief state crosses a certain threshold ω'. In other words, starting from an arbitrary belief state ω, the first active action on the arm is taken after $L(\omega, \omega')$ slots.

Based on the structure of threshold policy, $V_{\beta,\nu}(\omega)$ can be characterized in terms of $V_{\beta,\nu}(\mathcal{T}^{t_0-1}(\omega); a = 1)$ for some $t_0 \in \{1, 2, \cdots, \infty\}$, wher $t_0 = L(\omega, \omega^*) + 1$ is the slot when the belief ω reaches the threshold ω^* for the first time. Specially, in the first $L(\omega, \omega^*)$ slots, the subsidy ν is obtained in each slot. In slot $t_0 = L(\omega, \omega^*)+1$, the belief state reaches the threshold ω^* and the arm is activated. The total reward thereafter is $V_{\beta,\nu}(\mathcal{T}^{L(\omega,\omega^*)}(\omega); a = 1)$. Taking into account β, we thus have

$$\begin{aligned} V_{\beta,\nu}(\omega) =& \frac{1-\beta^{L(\omega,\omega^*)}}{1-\beta}\nu \\ &+ \beta^{L(\omega,\omega^*)} V_{\beta,\nu}(\mathcal{T}^{L(\omega,\omega^*)}(\omega); a = 1). \end{aligned} \tag{17}$$

B. Adjoint Dynamic System

In the dynamic system (10), the belief information ω represents two kinds of information:

- policy information, i.e., action a depends on ω;
- value information, i.e., the reward value of the dynamic system (or value function) depends on ω.

To better characterize the dynamic evolution of (10), we separate the two roles of ω by mathematically letting ω only represent the value while introducing $\lfloor \omega \rceil$ to indicate information used to make an action (corresponding to the policy).

Specifically, we introduce the following adjoint dynamic system

$$V_{\beta,\nu}(\omega; \lfloor \omega \rceil) = \max\Big\{V_{\beta,\nu}(\omega; \lfloor \omega, 0 \rceil), V_{\beta,\nu}(\omega; \lfloor \omega, 1 \rceil)\Big\}, \tag{18}$$

where,

$$\begin{aligned} &V_{\beta,\nu}(\omega; \lfloor \omega, 0 \rceil) = \nu + \beta V_{\beta,\nu}(\mathcal{T}(\omega); \lfloor \mathcal{T}(\omega) \rceil), \\ &V_{\beta,\nu}(\omega; \lfloor \omega, 1 \rceil) = (1-\epsilon)\omega + \beta\big[(1-\epsilon)\omega V_{\beta,\nu}(p_{11}; \lfloor p_{11} \rceil) \\ &\qquad\qquad + (1-(1-\epsilon)\omega) V_{\beta,\nu}(\Gamma(\omega)); \lfloor \Gamma(\omega) \rceil)\big]. \end{aligned}$$

where, $\lfloor \omega, a \rceil$ represents making action a ($a = 0, 1$) given the policy information ω.

Proposition 1. Given ν, $V_{\beta,\nu}(\omega; a = 1)$ and $V_{\beta,\nu}(\omega; a = 0)$ are piecewise linear and convex in ω.

Proof. We prove the proposition by induction. In slot T, we have $V^T_{\beta,\nu}(\omega; a = 0) = \nu$ and $V^T_{\beta,\nu}(\omega; a = 1) = (1-\epsilon)\omega$, which follows $V^T_{\beta,\nu}(\omega) = \max\{V^T_{\beta,\nu}(\omega; a = 0), V^T_{\beta,\nu}(\omega, a = 1)\}$ is piecewise linear and convex in ω.

Assume $V^{t+1}_{\beta,\nu}(\omega; a = 1)$ and $V^{t+1}_{\beta,\nu}(\omega; a = 0)$ are piecewise linear and convex in ω, it is easy to show that both

$V^t_{\beta,\nu}(\omega; a=1)$ and $V^t_{\beta,\nu}(\omega; a=0)$ are piecewise linear and convex in ω according to Eq. (10). Letting $T \nearrow \infty$, we prove the proposition. $\square$

Lemma 6. $V_{\beta,\nu}(\omega; \lfloor\omega, 1\rceil)$ is decomposable in ω, i.e.,

$$\begin{aligned} V_{\beta,\nu}(\omega; \lfloor\omega, 1\rceil) &= (1-\epsilon)\omega + \beta[(1-\epsilon)\omega V_{\beta,\nu}(p_{11}; \lfloor p_{11}\rceil) \\ &\quad + \epsilon\omega V_{\beta,\nu}(p_{11}; \lfloor\Gamma(\omega)\rceil) + (1-\omega)V_{\beta,\nu}(p_{01}; \lfloor\Gamma(\omega)\rceil)]. \end{aligned}$$

Proof.

$$\begin{aligned} & V_{\beta,\nu}(\omega; \lfloor\omega, 1\rceil) \\ &= (1-\epsilon)\omega + \beta[(1-\epsilon)\omega V_{\beta,\nu}(p_{11}; \lfloor p_{11}\rceil) \\ &\qquad + (1-(1-\epsilon)\omega)V_{\beta,\nu}(\Gamma(\omega); \lfloor\Gamma(\omega)\rceil)] \\ &\overset{(a)}{=} (1-\epsilon)\omega + \beta[(1-\epsilon)\omega V_{\beta,\nu}(p_{11}; \lfloor p_{11}\rceil) \\ &\qquad + (1-\omega)(1-(1-\epsilon)0)V_{\beta,\nu}(\mathcal{T}(0); \lfloor\Gamma(\omega)\rceil) \\ &\qquad + \omega(1-(1-\epsilon)1)V_{\beta,\nu}(\mathcal{T}(1); \lfloor\Gamma(\omega)\rceil)] \\ &= (1-\epsilon)\omega + \beta[(1-\epsilon)\omega V_{\beta,\nu}(p_{11}; \lfloor p_{11}\rceil) \\ &\qquad + (1-\epsilon)(1-\omega)V_{\beta,\nu}(p_{01}; \lfloor\Gamma(\omega)\rceil) \\ &\qquad + \epsilon\omega V_{\beta,\nu}(p_{11}; \lfloor\Gamma(\omega)\rceil)], \end{aligned} \tag{19}$$

where, (a) is due to Proposition 1. $\square$

Remark. In (19), for $V_{\beta,\nu}(p_{11}; \lfloor p_{11}\rceil)$ and $V_{\beta,\nu}(p_{11}; \lfloor\Gamma(\omega)\rceil)$, we can see that though they have the same value information p_{11}, they have different policy information, i.e., $\lfloor p_{11}\rceil$ and $\lfloor\Gamma(\omega)\rceil$. Hence, $V_{\beta,\nu}(p_{11}; \lfloor p_{11}\rceil) \neq V_{\beta,\nu}(p_{11}; \lfloor\Gamma(\omega)\rceil)$ except that both $\lfloor p_{11}\rceil$ and $\lfloor\Gamma(\omega)\rceil$ can lead a same action policy for the dynamic system.

VIII. Linearization of Value Function

In this section, we focus on the linearization of value function $V_{\beta,\nu}(\omega; \lfloor\omega, 1\rceil)$ for the case of negatively correlated channels, i.e., $p^{(i)}_{11} < p^{(i)}_{01}$, which serves as the basis to compute the Whittle index. As for the case of $p^{(i)}_{11} \geqslant p^{(i)}_{01}$, we omit the technical details for the limited space. Again, we consider one channel by dropping channel index i.

In many practical systems, the initial belief ω is set to ω_0 [8]. It can then be checked that $\min\{p_{01}, p_{11}\} \leqslant \omega \leqslant \max\{p_{01}, p_{11}\}$. Moreover, even the initial belief does not fall in $[\min\{p_{01}, p_{11}\}, \max\{p_{01}, p_{11}\}]$, all the belief values are bounded in the interval from the second slot following Lemma 1. Hence the following results can be extended by treating the first slot separately from the future slots. Therefore, we assume $\min\{p_{01}, p_{11}\} \leqslant \omega \leqslant \max\{p_{01}, p_{11}\}$ in the first slot in our analysis.

We divide the region $[p_{11}, p_{01}]$ into four subregions using the two fixed points ω_0 and $\bar{\omega}_0$:

$$[p_{11}, p_{01}] = [p_{11}, \omega_0) \cup [\omega_0, \mathcal{T}(p_{11})) \cup [\mathcal{T}(p_{11}), \bar{\omega}_0) \cup [\bar{\omega}_0, p_{01}].$$

In the following, we derive the linearized value function for these subregions, respectively.

A. Region $\big[p_{11}, \omega_0\big) \cup \big[\omega_0, \mathcal{T}(p_{11})\big)$

Proposition 2. If $p_{11} \leqslant \omega^* < \mathcal{T}(p_{11})$, it holds that $L(\mathcal{T}(\varphi(\omega)), \omega^*) = 0$ for any $\omega \in [p_{11}, p_{01}]$, .

Proof. In the case of $p_{11} < p_{01}$, $\varphi(\omega)$ monotonically increase with ω while $\mathcal{T}(\omega)$ monotonically decreases with ω. Thus, $\mathcal{T}(\varphi(\omega)) \geqslant \mathcal{T}(p_{11}) > \omega^*$ for $\omega \in [p_{11}, p_{01}]$ when $0 \leqslant \epsilon \leqslant \frac{p_{11}(1-p_{01})}{p_{01}(1-p_{11})}$. Therefore, $L(\mathcal{T}(\varphi(\omega)), \omega^*) = 0$. $\square$

Lemma 7. When $p_{11} \leqslant \omega^* < \mathcal{T}(p_{11})$, for any $\omega \in [p_{11}, p_{01}]$, the following holds

$$\begin{aligned} V_{\beta,\nu}(\omega; \lfloor\omega, 1\rceil) &= (1-\epsilon)\omega + \beta[(1-\epsilon)\omega V_{\beta,\nu}(p_{11}; \lfloor p_{11}\rceil) \\ &\quad + \epsilon\omega V_{\beta,\nu}(p_{11}; \lfloor\Gamma(\omega)\rceil) + (1-\omega)V_{\beta,\nu}(p_{01}; \lfloor\Gamma(\omega)\rceil)], \end{aligned}$$

where

$$\begin{aligned} & V_{\beta,\nu}(p_{11}; \lfloor p_{11}\rceil) \\ &= V_{\beta,\nu}(p_{11}; \lfloor p_{11}, 0\rceil) \\ &= \nu + \beta V_{\beta,\nu}(\mathcal{T}(p_{11}); \lfloor\mathcal{T}(p_{11})\rceil) \\ &= \nu + \beta(1-\epsilon)\mathcal{T}(p_{11}) \\ &\qquad + \beta^2[(1-\epsilon)\mathcal{T}(p_{11})V_{\beta,\nu}(p_{11}; \lfloor p_{11}\rceil) \\ &\qquad + (1-(1-\epsilon)\mathcal{T}(p_{11}))V_{\beta,\nu}(\Gamma(\mathcal{T}(p_{11})); \lfloor\Gamma(\mathcal{T}(p_{11}))\rceil)] \\ &\overset{(e1)}{=} \nu + \beta(1-\epsilon)\mathcal{T}(p_{11}) \\ &\qquad + \beta^2[(1-\epsilon)\mathcal{T}(p_{11})V_{\beta,\nu}(p_{11}; \lfloor p_{11}\rceil) \\ &\qquad + (1-(1-\epsilon)\mathcal{T}(p_{11}))V_{\beta,\nu}(\Gamma(\mathcal{T}(p_{11})); \lfloor\Gamma(\omega)\rceil)] \\ &\overset{(e2)}{=} \nu + \beta(1-\epsilon)\mathcal{T}(p_{11}) \\ &\qquad + \beta^2[(1-\epsilon)\mathcal{T}(p_{11})V_{\beta,\nu}(p_{11}; \lfloor p_{11}\rceil) \\ &\qquad + \epsilon\mathcal{T}(p_{11})V_{\beta,\nu}(p_{11}; \lfloor\Gamma(\omega)\rceil) \\ &\qquad + (1-\mathcal{T}(p_{11}))V_{\beta,\nu}(p_{01}; \lfloor\Gamma(\omega)\rceil)], \end{aligned}$$

$$\begin{aligned} & V_{\beta,\nu}(p_{11}; \lfloor\Gamma(\omega)\rceil) \\ &= V_{\beta,\nu}(p_{11}; \lfloor\Gamma(\omega), 1\rceil) \\ &\overset{(e3)}{=} (1-\epsilon)p_{11} + \beta[(1-\epsilon)p_{11}V_{\beta,\nu}(p_{11}; \lfloor p_{11}\rceil) \\ &\qquad + \epsilon p_{11}V_{\beta,\nu}(p_{11}; \lfloor\Gamma(\omega)\rceil) \\ &\qquad + (1-p_{11})V_{\beta,\nu}(p_{01}; \lfloor\Gamma(\omega)\rceil)], \end{aligned}$$

$$\begin{aligned} & V_{\beta,\nu}(p_{01}; \lfloor\Gamma(\omega)\rceil) \\ &= V_{\beta,\nu}(p_{01}; \lfloor\Gamma(\omega), 1\rceil) \\ &\overset{(e4)}{=} (1-\epsilon)p_{01} + \beta[(1-\epsilon)p_{01}V_{\beta,\nu}(p_{11}; \lfloor p_{11}\rceil) \\ &\qquad + \epsilon p_{01}V_{\beta,\nu}(p_{11}; \lfloor\Gamma(\omega)\rceil) \\ &\qquad + (1-p_{01})V_{\beta,\nu}(p_{01}; \lfloor\Gamma(\omega)\rceil)]. \end{aligned}$$

Proof. (e1) follows Proposition 2, (e2), (e3) and (e4) follow Lemma 6. $\square$

B. Region: $\big[\mathcal{T}(p_{11}), p_{01}\big)$

Based on Lemma 5, we have the following important corollary.

Corollary 2. When $\mathcal{T}(p_{11}) \leqslant \omega^* < p_{01}$, we have

(1) When $\mathcal{T}(p_{11}) \leqslant \omega^* < \bar{\omega}_0$, the first crossing time of the non-linear belief part $\Gamma^i(\omega^*)$ $(i = 1, 2, \cdots)$ will be 0 in the evolving process; that is, $L(\Gamma^i(\omega^*), \omega^*) = 0$;
(2) When $\bar{\omega}_0 \leqslant \omega^* < p_{01}$, the first crossing time of the non-linear belief part $\mathcal{T}^i(\Gamma(\omega^*))$ $(i = 0, 1, 2, \cdots)$ will be ∞ in the evolving process; that is, $L(\mathcal{T}^i(\Gamma(\omega^*)), \omega^*) = \infty$.

Proof. (1) By Lemma 5, we have that $\Gamma^i(\omega^*) > \omega^*$ when $\mathcal{T}(p_{11}) \leqslant \omega^* < \bar{\omega}_0$, and furthermore, $L(\Gamma^i(\omega^*), \omega^*) = 0$. (2) By Lemma 5, $\omega_0 < \Gamma(\omega^*) \leqslant \omega^*$ when $\bar{\omega}_0 \leqslant \omega^* < p_{01}$. Furthermore, by Lemma 2, we have $\mathcal{T}^i(\Gamma(\omega^*)) \leqslant \omega^*$, which means $L(\mathcal{T}^i(\Gamma(\omega^*)), \omega^*) = \infty$. □

Corollary 3. When $\mathcal{T}(p_{11}) \leqslant \omega^* < \bar{\omega}_0$, we have

$$\begin{aligned}
&V_{\beta,\nu}(\omega^*; \lfloor \omega^*, 1 \rceil) \\
&= (1-\epsilon)\omega^* + \beta[(1-\epsilon)\omega^* V_{\beta,\nu}(p_{11}, \lfloor p_{11} \rceil) \\
&\qquad + (1 - (1-\epsilon)\omega^*) V_{\beta,\nu}(\Gamma(\omega^*); \lfloor \Gamma(\omega^*) \rceil)] \\
&= (1-\epsilon)\omega^* + \beta[(1-\epsilon)\omega^* V_{\beta,\nu}(p_{11}; \lfloor p_{11} \rceil) \\
&\qquad + \epsilon\omega^* V_{\beta,\nu}(p_{11}; \lfloor \Gamma(\omega^*) \rceil) \\
&\qquad + (1-\omega^*) V_{\beta,\nu}(p_{01}; \lfloor \Gamma(\omega^*) \rceil)],
\end{aligned}$$

where,

$$V_{\beta,\nu}(p_{11}; \lfloor p_{11} \rceil) = \frac{\nu}{1-\beta},$$

$$\begin{aligned}
&V_{\beta,\nu}(p_{11}; \lfloor \Gamma(\omega^*) \rceil) \\
&= V_{\beta,\nu}(p_{11}; \lfloor \Gamma(\omega^*), 1 \rceil) \\
&= (1-\epsilon)p_{11} + \beta[(1-\epsilon)p_{11} V_{\beta,\nu}(p_{11}; \lfloor p_{11} \rceil) \\
&\qquad + (1 - (1-\epsilon)p_{11}) V_{\beta,\nu}(\Gamma(p_{11}); \lfloor \Gamma^2(\omega^*) \rceil)] \\
&= (1-\epsilon)p_{11} + \beta[(1-\epsilon)p_{11} V_{\beta,\nu}(p_{11}; \lfloor p_{11} \rceil) \\
&\qquad + (1 - (1-\epsilon)p_{11}) V_{\beta,\nu}(\Gamma(p_{11}); \lfloor \Gamma(\omega^*) \rceil)] \\
&= (1-\epsilon)p_{11} + \beta[(1-\epsilon)p_{11} V_{\beta,\nu}(p_{11}; \lfloor p_{11} \rceil) \\
&\qquad + \epsilon p_{11} V_{\beta,\nu}(p_{11}; \lfloor \Gamma(\omega^*) \rceil) \\
&\qquad + (1-p_{11}) V_{\beta,\nu}(p_{01}; \lfloor \Gamma(\omega^*) \rceil)],
\end{aligned}$$

$$\begin{aligned}
&V_{\beta,\nu}(p_{01}; \lfloor \Gamma(\omega^*) \rceil) \\
&= (1-\epsilon)p_{01} + \beta[(1-\epsilon)p_{01} V_{\beta,\nu}(p_{11}; \lfloor p_{11} \rceil) \\
&\qquad + \epsilon p_{01} V_{\beta,\nu}(p_{11}; \lfloor \Gamma(\omega^*) \rceil) \\
&\qquad + (1-p_{01}) V_{\beta,\nu}(p_{01}; \lfloor \Gamma(\omega^*) \rceil)].
\end{aligned}$$

C. Region: $\lfloor \bar{\omega}_0, p_{01})$

When $\bar{\omega}_0 \leqslant \omega^* < p_{01}$, we have $L(\mathcal{T}(\varphi(\omega^*)), \omega^*) = \infty$ by Corollary 2. Thus,

$$\begin{aligned}
&V_{\beta,\nu}(\omega^*; \lfloor \omega^*, 1 \rceil) \\
&= (1-\epsilon)\omega^* + \beta[(1-\epsilon)\omega^* V_{\beta,\nu}(p_{11}; \lfloor p_{11} \rceil) \\
&\qquad + (1 - (1-\epsilon)\omega^*) V_{\beta,\nu}(\Gamma(\omega^*); \lfloor \Gamma(\omega^*) \rceil)] \\
&= (1-\epsilon)\omega^* + \beta[(1-\epsilon)\omega^* V_{\beta,\nu}(p_{11}; \lfloor p_{11} \rceil) \\
&\qquad + (1 - (1-\epsilon)\omega^*) \frac{\nu}{1-\beta}].
\end{aligned}$$

IX. Numerical Study

In this section, we evaluate the performance of the Whittle index policy by comparing with the myopic policy (sensing the best channel in terms of belief value).

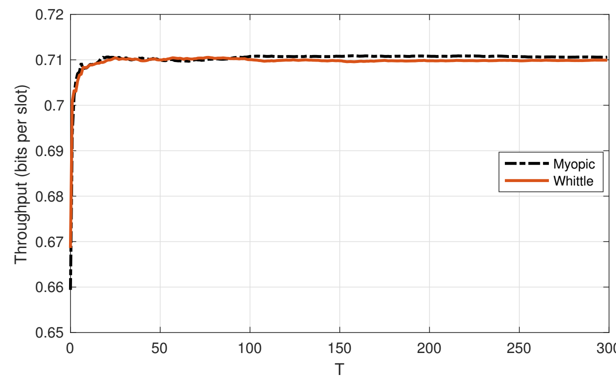

Fig. 2. $N = 3, \beta = 1, \epsilon_i = 0.01$, $\{(p_{01}^{(i)}, p_{11}^{(i)})\}_{i=1}^{3} = \{(0.3, 0.7), (0.4, 0.8), (0.5, 0.7)\}$.

From Figure 2, we observe that the Whittle index policy has almost the same performance with the myopic policy.

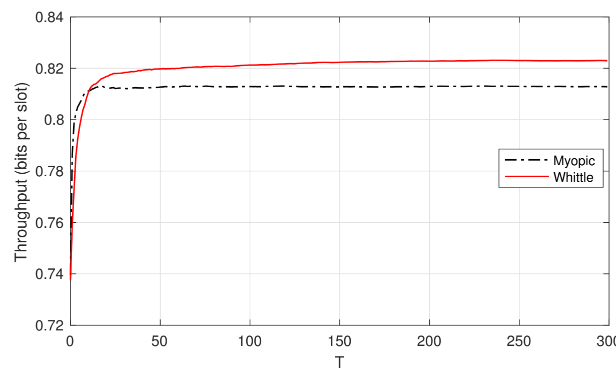

Fig. 3. $N = 10, \beta = 1, \epsilon_i = 0.01$, $\{(p_{01}^{(i)}, p_{11}^{(i)})\}_{i=1}^{10} = \{(0.3, 0.9), (0.8, 0.1), (0.3, 0.8), (0.1, 0.9), (0.9, 0.1), (0.4, 0.8), (0.5, 0.3), (0.3, 0.3), (0.3, 0.6), (0.8, 0.1)\}$.

From Figure 3, we can see that the Whittle index policy performs a little worse than the myopic policy when $T \leqslant 18$, while after that threshold time, performs better. This can be easily explained as follow: the myopic policy performs better in the initial period since it only exploits information to maximize utility but ignores exploring information for future decision. However, the Whittle index considers the balance between exploitation and exploration, so it performs better after the initial period.

X. Conclusion

In this paper, we study the Whittle index policy for multi-channel opportunistic access problem with imperfect observation. The traditional approach of computing the Whittle index policy cannot be applied because the channel state belief evolution is no more linear, thus rendering the indexability of our problem open. To bridge the gap, we mathematically establish the indexability and establish the closed-form Whittle-index, based on which index policy can be constructed. The major technique is our analysis is a fixed point based approach which enable us to divide the belief information space into a series of regions and then establish a set of periodic structures of the underlying nonlinear dynamic evolving system, based on which we devise the linearization scheme for each region to establish indexability and compute the Whittle index for each region.

XI. Acknowledgments

This work is supported by national natural science foundation of China (No. 61672395) and China Postdoctoral Science Foundation (No. 2013M531753, No. 2014T70748).

References

[1] S. Aalto, P. Lassila, and P. Osti. Whittle index approach to size-aware scheduling with time-varying channels. In *Proc. ACM Sigmetrics*, Portland, OR, June 2015.

[2] F. Cecchi and P. Jacko. Scheduling of users with markovian time-variyng transmission rates. In *Proc. ACM Sigmetrics*, Pittsburgh, PA, June 2013.

[3] E. N. Gilbert. Capacity of a Burst-Noise Channel. *Bell System Technical Journal*, 39(5):1253–1265, Sep 1960.

[4] J. Gittins, K. Glazebrook, and R. Weber. *Multi-armed Bandit Allocation Indices*. John Wiley & Sons Ltd., New York, NY, 2nd edition edition, 2011.

[5] D. R. Hernandez. *Indexable Restless Bandits*. VDM Verlag, 2008.

[6] P. Jacko. *Dynamic Priority Allocation in Restless Bandit Models*. LAP Lambert Academic Publishing, 2010.

[7] K. Liu and Q. Zhao. Indexability of Restless Bandit Problems and Optimality of Whittle Index for Dynamic Multichannel Access. *IEEE Trans. Inf. Theory*, 56(11):5547–5567, Nov. 200.

[8] K. Liu, and Q. Zhao, and B. Krishnamachari. Dynamic Multichannel Access With Imperfect Channel State Detection. *IEEE Trans. Signal Process.*, 58(5):2795–2807, May 2010.

[9] V. S. B. Konstantin E. Avrachenkov. Whittle index policy for crawling ephemeral content. In *Proc. of 2015 IEEE 54th Annual Conference on Decision and Control (CDC)*, 2015.

[10] Y. Liu, M. Liu, and S. H. A. Ahmad. Sufficient Conditions on the Optimality of Myopic Sensing in Opportunistic Channel Access: A Unifying Framework. *IEEE Trans. Inf. Theory*, 60(8):4922–4940, Aug. 2014.

[11] J. Nino-Mora and S. S. Villar. Sensor scheduling for hunting elusive hiding targets via whittle's restless bandit index policy. In *Proc. of NetGCoop*, pages 1–8, Paris, France, Oct 2011.

[12] J. L. Ny, M. Dahleh, and E. Feron. Multi-UAV Dynamic Routing with Partial Observations Using Restless Bandit Allocation Indices. In *Proc. of American Control Conf.*, pages 4220–4225, Seattle, WA, June 2008.

[13] W. Ouyang, S. Murugesan, A. Eryilmaz, and N. B. Shroff. Exploiting channel memory for joint estimation and scheduling in downlink networks–a whittle's indexability analysis. *IEEE Trans. Inf. Theory*, 4(61):1702–1719, Apr. 2015.

[14] Y. Ouyang and D. Teneketzis. On the Optimality of Myopic Sensing in Multi-state resources. *IEEE Trans. Inf. Theory*, 60(1):681–696, Jan. 2014.

[15] C. H. Papadimitriou and J. N. Tsitsiklis. The complexity of optimal queueing network control. *Math. Oper. Res.*, 24(2):293–305, 1999.

[16] P. R. Singh, Xueying Guo. Index policies for optimal mean-variance trade-off of inter-delivery times in real-time sensor networks. In *Proc. of 2015 IEEE Conference on Computer Communications (INFOCOM)*, 2015.

[17] S. Ahmad and M. Liu. Multi-channel opportunistic access: A case of restless bandits with multiple players. In *Proc. Allerton Conf. Commun. Control Comput*, pages 1361–1368, Oct. 2009.

[18] S. Ahmad, and M. Liu, and T. Javidi, and Q. zhao and B. Krishnamachari. Optimality of myopic sensing in multichannel opportunistic access. *IEEE Trans. Inf. Theory*, 55(9):4040–4050, Sep. 2009.

[19] T. J. S. H. Ahmad, M. Liu, Q. Zhao, and B. Krishnamachari. Optimality of myopic sensing in multichannel opportunistic access. *IEEE Trans. Inf. Theory*, 55(9):4040–4050, 2009.

[20] I. M. Verloop. Asymptotically Optimal Priority Policies for Indexable and Non-indexable Restless Bandits. *Annals of Applied Probability*, 26(4):1947–1995, Sep. 2016.

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

[22] K. Wang, L. Chen, and Q. Liu. pportunistic spectrum access by exploiting primary user feedbacks in underlay cognitive radio systems: An optimality analysis. *IEEE J. Sel. Topics Signal Process.*, 7(5):869–882, 2013.

[23] K. Wang, L. Chen, Q. Liu, and K. A. Agha. On optimality of myopic sensing policy with imperfect sensing in multi-channel opportunistic access. *IEEE Trans. Commun.*, 61(9):3854–3862, 2013.

[24] K. Wang, Q. Liu, F. Li, L. Chen, and X. Ma. Myopic policy for opportunistic access in cognitive radio networks by exploiting primary user feedbacks. *IET Commun.*, 9(7):1017–1025, 2015.

[25] R. R. Weber and G. Weiss. On an Index Policy for Restless Bandits. *J. Appl. Probab.*, 27(3):637–648, Sep. 1990.

[26] P. Whittle. Restless Bandits: Activity Allocation in a Changing World. *J. Appl. Probab.*, 24:287–298, 1988.

$$W_\beta(\omega)=\begin{cases}\frac{(1-\epsilon_i)(\beta p_{01}^{(i)}+(\omega-\beta\mathcal{T}_i(\omega)))}{1+\beta(p_{01}^{(i)}-\epsilon_i p_{11}^{(i)})-\beta^2(1-\epsilon_i)\mathcal{T}_i(p_{11}^{(i)})-\beta(1-\epsilon_i)(\omega-\mathcal{T}_i(\omega))}, & \text{if } p_{11}^{(i)}\leqslant\omega<\omega_0^{(i)}\\ \frac{(1-\epsilon_i)(\beta p_{01}^{(i)}+(1-\beta)\omega)}{1+\beta(p_{01}^{(i)}-\epsilon_i p_{11}^{(i)})-\beta^2(1-\epsilon_i)\mathcal{T}_i(p_{11}^{(i)})-\beta(1-\beta)(1-\epsilon_i)\omega}, & \text{if } \omega_0^{(i)}<\omega<\mathcal{T}_i(p_{11}^{(i)})\\ \frac{(1-\epsilon_i)(\beta p_{01}^{(i)}+(1-\beta)\omega)}{1+\beta(p_{01}^{(i)}-\epsilon_i p_{11}^{(i)})-\beta(1-\epsilon_i)\omega}, & \text{if } \mathcal{T}_i(p_{11}^{(i)})\leqslant\omega<\bar{\omega}_0^{(i)}\\ (1-\epsilon_i)\omega, & \text{if } \bar{\omega}_0^{(i)}\leqslant\omega\leqslant p_{01}^{(i)}.\end{cases} \tag{20}$$

$$W_\beta(\omega)=\begin{cases}(1-\epsilon_i)\omega, & \text{if } p_{01}^{(i)}\leqslant\omega\leqslant\underline{\omega}_0^{1,(i)}\\ W_\beta(\underline{\omega}_0^{n,(i)})+(\omega-\underline{\omega}_0^{n,(i)})\frac{W_\beta(\overline{\omega}_0^{n,(i)})-W_\beta(\underline{\omega}_0^{n,(i)})}{\overline{\omega}_0^{n,(i)}-\underline{\omega}_0^{n,(i)}}, & \text{if } \underline{\omega}_0^{n,(i)}<\omega<\overline{\omega}_0^{n,(i)} \text{ and } n=1,2,\cdots\\ \frac{(1-\epsilon_i)(1-\beta^{n+1})(\omega-\beta\mathcal{T}_i(\omega))+C_6}{C_0(\omega-\beta\mathcal{T}_i(\omega))+C_7}, & \text{if } \overline{\omega}_0^{n,(i)}\leqslant\omega<\mathcal{T}_i^n(\varphi(p_{11}^{(i)})),\ n=1,2,\cdots\\ \frac{(1-\epsilon_i)(1-\beta^{n+1})(\omega-\beta\mathcal{T}_i(\omega))+C_6}{(1-\epsilon_i)(\beta-\beta^{n+1})(\omega-\beta\mathcal{T}_i(\omega))+C_9}, & \text{if } \mathcal{T}_i^n(\varphi(p_{11}^{(i)}))\leqslant\omega<\underline{\omega}_0^{n+1,(i)},\ n=1,2,\cdots\\ \frac{(1-\epsilon_i)\omega}{1-\beta(1-\epsilon_i)(p_{11}^{(i)}-\omega)}, & \text{if } \omega_0^{(i)}\leqslant\omega\leqslant p_{11}^{(i)}\end{cases} \tag{21}$$

$$W(\omega)=\begin{cases}\frac{(1-\epsilon_i)(p_{01}^{(i)}+\omega-\mathcal{T}_i(\omega))}{1+p_{01}^{(i)}-\epsilon_i p_{11}^{(i)}-(1-\epsilon_i)\mathcal{T}_i(p_{11}^{(i)})-(1-\epsilon_i)(\omega-\mathcal{T}_i(\omega))}, & \text{if } p_{11}^{(i)}\leqslant\omega<\omega_0^{(i)}\\ \frac{(1-\epsilon_i)p_{01}^{(i)}}{1+p_{01}^{(i)}-\epsilon_i p_{11}^{(i)}-(1-\epsilon_i)\mathcal{T}_i(p_{11}^{(i)})}, & \text{if } \omega_0^{(i)}\leqslant\omega<\mathcal{T}_i(p_{11}^{(i)})\\ \frac{(1-\epsilon_i)p_{01}^{(i)}}{1+p_{01}^{(i)}-\epsilon_i p_{11}^{(i)}-(1-\epsilon_i)\omega}, & \text{if } \mathcal{T}_i(p_{11}^{(i)})\leqslant\omega<\bar{\omega}_0^{(i)}\\ (1-\epsilon_i)\omega, & \text{if } \bar{\omega}_0^{(i)}\leqslant\omega\leqslant p_{01}^{(i)}.\end{cases} \tag{22}$$

$$W(\omega)=\begin{cases}(1-\epsilon_i)\omega, & \text{if } p_{01}^{(i)}\leqslant\omega\leqslant\underline{\omega}_0^{1,(i)}\\ W(\underline{\omega}_0^{n,(i)})+(\omega-\underline{\omega}_0^{n,(i)})\frac{W(\overline{\omega}_0^{n,(i)})-W(\underline{\omega}_0^{n,(i)})}{\overline{\omega}_0^{n,(i)}-\underline{\omega}_0^{n,(i)}}, & \text{if } \underline{\omega}_0^{n,(i)}<\omega<\overline{\omega}_0^{n,(i)} \text{ and } n=1,2,\cdots\\ \frac{(1-\epsilon_i)(n+1)(\mathcal{T}_i(\omega)-\omega)-(1-\epsilon)\mathcal{T}_i^n(p_{01}^{(i)})}{(1-\epsilon_i)(n+1-(1-\epsilon)p_{11}^{(i)})(\mathcal{T}_i(\omega)-\omega)+C_7'}, & \text{if } \overline{\omega}_0^{n,(i)}\leqslant\omega<\mathcal{T}_i^n(\varphi(p_{11}^{(i)})),\ n=1,2,\cdots\\ \frac{(1-\epsilon)(n+1)(\mathcal{T}_i(\omega)-\omega)-(1-\epsilon)\mathcal{T}_i^n(p_{01}^{(i)})}{(1-\epsilon)[n(\mathcal{T}_i(\omega)-\omega)+p_{11}^{(i)}]-1-\mathcal{T}_i^n(p_{01}^{(i)})+\epsilon\mathcal{T}_i^n(p_{11}^{(i)})}, & \text{if } \mathcal{T}_i^n(\varphi(p_{11}^{(i)}))\leqslant\omega<\underline{\omega}_0^{n+1,(i)},\ n=1,2,\cdots\\ \frac{(1-\epsilon_i)\omega}{1-(1-\epsilon_i)(p_{11}^{(i)}-\omega)}, & \text{if } \omega_0^{(i)}\leqslant\omega\leqslant p_{11}^{(i)}\end{cases} \tag{23}$$

where,

$$\begin{aligned}C_0=&(1-\epsilon_i)\beta[1-\beta^n p_{11}^{(i)}(1-\epsilon_i)-\beta^{n+1}(1-(1-\epsilon_i)p_{11}^{(i)})],\\ C_6=&(1-\epsilon_i)(1-\beta)\beta^{n+1}\mathcal{T}_i^n(p_{01}^{(i)}),\\ C_7=&-\epsilon_i(1-\beta)\beta^{n+1}(1-\beta(1-\epsilon_i)p_{11}^{(i)})\mathcal{T}_i^n(p_{11}^{(i)})+(1-\beta)\beta^{n+1}[1+\beta(1-\epsilon_i)(1-p_{11}^{(i)})]\mathcal{T}_i^n(p_{01}^{(i)})\\ &-\epsilon_i(1-\epsilon_i)(1-\beta)\beta^{n+1}p_{11}^{(i)}\mathcal{T}_i^{n-1}(p_{11}^{(i)})-(1-\epsilon_i)(1-\beta)\beta^{n+1}(1-p_{11}^{(i)})\mathcal{T}_i^{n-1}(p_{01}^{(i)})+(1-\beta)[1-\beta(1-\epsilon_i)p_{11}^{(i)}],\\ C_7'=&\epsilon_i[1-(1-\epsilon_i)p_{11}^{(i)}]\mathcal{T}_i^n(p_{11}^{(i)})-[1+(1-\epsilon_i)(1-p_{11}^{(i)})]\mathcal{T}_i^n(p_{01}^{(i)})\\ &+\epsilon_i(1-\epsilon_i)p_{11}^{(i)}\mathcal{T}_i^{n-1}(p_{11}^{(i)})+(1-\epsilon_i)(1-p_{11}^{(i)})\mathcal{T}_i^{n-1}(p_{01}^{(i)})+(1-\epsilon_i)p_{11}^{(i)}-1,\\ C_9=&(1-\beta)[1-\beta(1-\epsilon_i)p_{11}^{(i)}]+(1-\beta)\beta^{n+1}[\mathcal{T}_i^n(p_{01}^{(i)})-\epsilon_i\mathcal{T}_i^n(p_{11}^{(i)})].\end{aligned}$$