# Opportunistic Scheduling Revisited Using Restless Bandits: Indexability and Index Policy

Kehao Wang
Wuhan University of Technology
Wuhan, China
kehao.wang@whut.edu.cn

Jihong Yu, Lin Chen
Université de Paris-Sud
Orsay, France
Email: chen@lri.fr, jihongy@sfu.ca

Moe Win
Massachusetts Institute of Technology
Cambridge, USA
Email: moewin@mit.edu

***Abstract*—We revisit the opportunistic scheduling problem where a server opportunistically serves multiple classes of users under time varying multi-state Markovian channels. The aim of the server is to find an optimal policy minimizing the average waiting cost of users. Mathematically, the problem can be cast to a restless bandit one, and a pivot to solve restless bandit by index policy is to establish indexability. Despite the theoretical and practical importance of the index policy, the indexability is still open for the opportunistic scheduling in the heterogeneous multi-state channel case. To fill this gap, we mathematically propose a set of sufficient conditions on channel state transition matrix under which the indexability is guaranteed, and consequently, the index policy is feasible. We further develop a simplified procedure to compute the index by reducing the complexity from more than quadratic to linear. Our work consists of a small step toward solving the opportunistic scheduling problem in its generic form involving multi-state Markovian channels and multi-class users.**

***Index Terms*—Restless bandit; indexability; stochastic scheduling; performance evaluation**

## I. Introduction

We revisit the following opportunistic scheduling problem involving a base station, also referred to as server, and different classes of users with heterogeneous demands. The system operates on time varying multi-state Markovian channels. Each channel at different states and different classes has different transmission rate, i.e., the evolution of channels is Markovian and class-dependent. For users connected to (or entered) the system but not served immediately, their waiting costs increase with time. In such opportunistic scheduling scenario, a central problem is how to exploit the server's capacity to serve the users. This problem can be formalized to the problem of designing an optimum opportunistic scheduling policy minimizing the average waiting cost.

The above opportunistic scheduling problem is fundamental in many classical and emerging wireless communication systems such as mobile cellular systems including 4G LTE and the emerging 5G, heterogeneous networks (HetNet). A central problem is to design efficient schedulers exploiting the capacity of the base station to serve heterogeneous users opportunistically.

### *A. State of the art*

Due to its fundamental importance, the opportunistic scheduling problem has attracted a large body of research on channel-aware schedulers addressing one or more system performance metrics in terms of throughput, fairness, and stability [1–16, 18–20, 24–26].

The seminal work in [19] showed that the system capacity can be improved by opportunistically serving users with maximal transmission rate. Such schedulers are called c$\mu$-rule or *MaxRate* scheduler. In fact, the *MaxRate* scheduler was myopically throughput optimal, i.e., maximizing the current slot transmission rate but ignoring the impact of the current scheduling on the future throughput, and consequently, was shown to perform bad in system stability from the long-term viewpoint. For instance, the number of waiting users in the system explodes with the increase of system load. Meanwhile, the *MaxRate* scheduler does not fairly schedule those users with lower transmission rate.

To balance system throughput and fairness, the Proportionally Fair (PF) scheduler was proposed and implemented in CDMA 1xEV-DO system of 3G cellular networks [7]. Technically, the PF scheduler maximizes the logarithmic throughput of the system rather than traditional throughput, and as a result, provides better fairness [20]. In [10], the authors approximated the PF by the relatively best (RB) scheduler, and analyzed flow-level stability of the PF scheduler. Actually, the RB scheduler gives priority to users according to their ratio of the current transmission rate to the mean transmission rate. Naturally, it is fair to users by taking future evolution into account. Consequently, it can provide a minimal throughput to the users with low accessible transmission rates, at the price of being not maximally stable at flow-level [1].

Other schedulers, e.g., *scored based (SB)* [8], *proportionally best (PB)*, and *potentially improvement (PI)*, belong to the family of the best-condition schedulers. These schedulers give priority to users according to their respective evaluated channel condition, and accordingly, do not have a direct association with transmission rate. They are not myopically throughput-optimal, but rather have a good performance in the long-term. They are maximally stable [6, 18], but they do not consider fairness.

The above schedulers all assume independent and identically distributed channels. For the more challenging scenario with Markovian channels, there exist some works on homogeneous channels [3–5] and heterogeneous channels [13, 14]. Under the Markovian channel model, the opportunistic scheduling problem can be mathematically cast to a restless multiarmed bandit (RMAB) [27]. The RMAB is of fundamental importance in stochastic decision theory due to its generic nature, and has application in numerous engineering problems. The central problem in investigating and solving an instance of RMAB is to establish its *indexability*. Once the indexability is established, an index policy can be constructed by assigning an index for each state of each arm to measure the reward of activating the arm at that state. The policy thus consists of simply activating those arms with the largest indices.

In the context of opportunistic scheduling which is the topic of our study, the authors [5] considered a flow-level scheduling problem with time-homogeneous channel state transition where the probability of being in a state is fixed for any time slot no matter how long the system evolves. For the same channel model, the authors [3] considered the opportunistic scheduling problem under the assumption of traffic size following the Pascal distribution. In [3–5], the indexability was first proved, and then the similar closed form Whittle index was obtained [27]. For heterogeneous channels, the authors of [13, 14] considered a generic flow-level scheduling problem with heterogeneous channel state transition, but carried out their work based on a conjecture that the problem is indexable. As a result, they can only verify the indexability of the proposed policy for very specific scenarios by numerical test before computing the policy index. *The indexability of the opportunistic scheduling for the heterogeneous multi-state Markovian channels, despite its theoretical and practical importance, is still open today.*

### B. Main Results and Contributions

To bridge the above theoretical gap, we investigate the indexability of the heterogeneous channel case formulated in [13, 14], and mathematically construct a set of sufficient conditions on channel state transition matrix under which the indexability is guaranteed and consequently the index policy is feasible. Our work thus consists of a small step towards solving the opportunistic scheduling problem in its generic form involving multi-state Markovian channels. As a desirable feature, the indexability conditions established in this work only depend on channel state transition matrix without imposing constraints on other user-dependent parameters such as service rate and waiting cost.

**Notation:** Throughout the paper, boldface lower and upper case letters represent column vectors and matrices, respectively. $(\cdot)^{\mathsf{T}}$ represents the transpose of a matrix or a vector. $(\cdot)^{-1}$ represents the inverse of a matrix. $\mathbf{e}_i$ denotes an $N$-dimensional column vector with 1 in the $i$-th element and 0 in others. $\mathbf{E}_N$ denotes the $N \times N$ identity matrix. $\mathbf{1}_N$ denotes an $N$-dimensional column vector with 1 in all elements. $\mathbf{0}_N$ denotes an $N$-dimensional column vector with 0 in all elements. $\mathbf{1}_N^k$ denotes the $N$-dimensional column vector with 1 in the first $k$ elements and 0 in the remaining $N-k$ elements.

## II. System Model

As mentioned in the induction, we consider a wireless communication system where a server schedules jobs of heterogeneous users. The system operates in a time-slotted fashion where $\tau$ denotes the slot duration and $t \in \mathcal{T} := \{0, 1, \cdots\}$ denote the slot index.

### A. Job, channel, and user models

Suppose that there are $K$ classes of users, $k \in \mathcal{K} := \{1, 2, \cdots, K\}$. Each user of class $k$ is uniquely associated with a job of class $k$ which he requests from the server, and with a dedicated wireless channel of class $k$ through which the transmission is done.

*User arrivals*. For each class $k \in \mathcal{K}$, the number of arriving users in class $k$ during the time slot $t \in \mathcal{T}$, denoted as $A_k(t)$, forms an i.i.d. arrival process $\{A_k(t)\}_{t\in\mathcal{T}}$ with generic element $A_k$ and mean $\xi_k := \mathbb{E}_0\{A_k\} < 1$, where $\mathbb{E}_0[\cdot]$ denotes the expectation conditioned on information available at time epoch 0. The arrivals are assumed to be mutually independent across user classes.

*Job sizes*. The job (or flow) size $b_k$ of class $k$ users in bits is geometrically distributed with mean $\mathbb{E}\{b_k\} < 1$ for class $k \in \mathcal{K}$.

*Channel condition*. For each user, the channel condition varies from slot to slot, independently of all other users. For each class $k$ user, the set of discretized channel conditions is denoted by the finite set $\mathcal{N}_k' := \{1, 2, \cdots, N_k\}$. The channel conditions typically correspond to modulation and coding schemes (MCSs) of the user's transmission technology. The probability that the channel condition of a class $k$ user in $n \in \mathcal{N}_k'$ is denoted by $q_{k,n} > 0$, where $\sum_{n\in\mathcal{N}_k'} q_{k,n} = 1$.

*Channel condition evolution*. We assume that at each slot, the channel condition of each user in the system evolves according to a class-dependent Markov chain. Thus, for each user of class $k \in \mathcal{K}$, we can define a Markov chain with state space $\mathcal{N}_k'$. We further define $q_{k,n,m} := \mathbb{P}(Z_k(t+1) = m|Z_k(t) = n)$, where $Z_k(t)$ denotes the channel condition of a class $k$ user at time $t$. The class $k$ channel condition transition probability matrix is thus defined as:

$$\mathbf{Q}^{(k)} := \begin{bmatrix} q_{k,1,1} & q_{k,1,2} & \cdots & q_{k,1,N_k} \\ q_{k,2,1} & q_{k,2,2} & \cdots & q_{k,2,N_k} \\ \vdots & \vdots & \ddots & \vdots \\ q_{k,N_k,1} & q_{k,N_k,2} & \cdots & q_{k,N_k,N_k} \end{bmatrix},$$

where $\sum_{m\in\mathcal{N}_k'} q_{k,n,m} = 1$ for every $n \in \mathcal{N}_k'$.

*Transmission rates*. When a class $k$ user is in channel condition $n \in \mathcal{N}_k'$, he can receive data at transmission rate $s_{k,n}$, i.e., his job is served at rate $s_{k,n}$. We assume that the higher the label of the channel condition, the higher the transmission rate, i.e., $0 \leqslant s_{k,1} < s_{k,2} < \cdots < s_{k,N_k}$.

*Waiting costs*. For every user of class $k$, the system operator accrues waiting cost $c_k$ ($c_k > 0$) at the end of every slot if her job is uncompleted.

### B. Server model

The server is assumed to have *full knowledge* of the system parameters. We investigate the case where the server can serve one user each slot. However, our analysis can be straightforwardly generalized to the case where multiple users can be served each slot. At the beginning of every slot, the server observes the actual channel conditions of all users in the system, and decides which user to serve during the slot. We assume that the server is preemptive, i.e, at any time it can interrupt the service of a user whose job is not yet completed. The server is also allowed to stay idle, and note that it is not work-conserving because of the time varying transmission rate. We denote by $\mu_{k,n}$ the departure probability that the job is completed within the current time slot when the server serves a class $k$ user in channel condition $n \in \mathcal{N}_k'$. Note that the departure probabilities are increasing in the channel condition, i.e., $0 \leqslant \mu_{k,1} < \cdots < \mu_{k,N_k} \leqslant 1$, because the transmission rates satisfy $0 \leqslant s_{k,1} < s_{k,2} < \cdots < s_{k,N_k}$.

### C. Opportunistic scheduling problem

In the opportunistic scheduling model formulated here, a central problem is how to exploit the server's capacity to serve users. This problem can be formalized to the problem of designing an optimum opportunistic scheduling policy minimizing the average waiting cost.

## III. Restless Bandit Formulation and Analysis

In this section we analyze the scheduling policy by casting it to a RMAB problem. For the ease of analysis, we investigate the discounted waiting costs by introducing a discount factor $0 \leqslant \beta < 1$. The time-average case is a special case where $\beta \to 1$ basically.

### A. Job-channel-user bandit

We denote by $\mathcal{A}_k := \{0, 1\}$ the action space of user $k$ where action 1 means serving the user and 0 not serving him.

Every job-channel-user couple of class $k$ is characterized by the tuple $\big(\mathcal{N}_k, (\mathbf{W}_k^a)_{a\in\mathcal{A}}, (\mathbf{R}_k^a)_{a\in\mathcal{A}}, (\mathbf{P}_k^a)_{a\in\mathcal{A}}\big)$, where

(1) $\mathcal{N}_k := \{0\} \cup \mathcal{N}_k'$ is the user state space, where state 0 indicates that the job is completed, and state $n \in \mathcal{N}_k'$ indicates that the current channel condition is $n$ and the job is uncompleted;
(2) $\mathbf{W}_k^a := (W_{k,n}^a)_{n\in\mathcal{N}_k}$, where $W_{k,n}^a$ is the expected one-slot capacity consumption, or work required by a user at state $n$ if action $a$ is chosen. Specifically, for every state $n \in \mathcal{N}_k$, $W_{k,n}^1 = 1$ and $W_{k,n}^0 = 0$;
(3) $\mathbf{R}_k^a := (R_{k,n}^a)_{n\in\mathcal{N}_k}$, where $R_{k,n}^a$ is the expected one-slot reward earned by a user at state $n$ if action $a$ is selected. Specifically, for every state $n \in \mathcal{N}_k'$, it is the negative of the expected waiting cost, $R_{k,0}^a = 0$, $R_{k,n}^1 = -\bar{\mu}_{k,n} c_k$ where $\bar{\mu}_{k,n} = 1 - \mu_{k,n}$, and $R_{k,n}^0 = -c_k$.
(4) $\mathbf{P}_k^a := (p_{k,n,m}^a)_{n,m\in\mathcal{N}_k}$, where $p_{k,n,m}^a$ is the probability for a user evolving from state $n$ to state $m$ if action $a$ is selected. The one-slot transition probability matrices for action 0 and 1 are as below:

$$\mathbf{P}_k^0 = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & q_{k,1,1} & \cdots & q_{k,1,N_k} \\ 0 & q_{k,2,1} & \cdots & q_{k,2,N_k} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & q_{k,N_k,1} & \cdots & q_{k,N_k,N_k} \end{bmatrix},$$

$$\mathbf{P}_k^1 = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \mu_{k,1} & \bar{\mu}_{k,1} q_{k,1,1} & \cdots & \bar{\mu}_{k,1} q_{k,1,N_k} \\ \mu_{k,2} & \bar{\mu}_{k,2} q_{k,2,1} & \cdots & \bar{\mu}_{k,2} q_{k,2,N_k} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{k,N_k} & \bar{\mu}_{k,N_k} q_{k,N_k,1} & \cdots & \bar{\mu}_{k,N_k} q_{k,N_k,N_k} \end{bmatrix}.$$

The dynamics of user $j$ of class $k$ are captured by the state process $X_k(\cdot)$ and the action process $a_j(\cdot)$, which correspond to state $X_j(t) \in \mathcal{N}_k$ and action $a_j(t) \in \mathcal{A}_k$ at any slot $t$.

### B. Restless bandit formulation and opportunistic scheduling

Let $\Pi_{X,a}^t$ denote the set of all the policies composed of actions $a(0), a(1) \cdots, a(t)$, where $a(t)$ is determined by the state history $X(0), X(1), \cdots, X(t)$ and the action history $a(0), a(1) \cdots, a(t-1)$, i.e.,

$$\begin{aligned}\Pi_{X,a}^t :&= \Big\{a(i)\Big| a(i) = \phi\big(X^{0:i}, a^{0:i-1}\big), i = 0, 1 \cdots, t\Big\} \\ &\stackrel{(e)}{=} \Big\{a(i)\Big| a(i) = \phi\big(X(i)\big), i = 0, 1 \cdots, t\Big\},\end{aligned}$$

where, $\phi$ is a mapping $\phi : (X^{0:i}, a^{0:i-1}) \to a(i)$, $X^{0:i} := (X(0), \cdots, X(i))$ and $a^{0:i-1} := (a(0), \cdots, a(i-1))$, and (e) is due to the Markovian feature.

Let $\Pi_{\mathbf{X},\mathbf{a}}^t$ denote the space of randomized and non-anticipative policies depending on the joint state-process $\mathbf{X} := (X_k(\cdot))_{k\in\mathcal{K}}$ and the joint action-process $\mathbf{a}(\cdot) := (a_k(\cdot))_{k\in\mathcal{K}}$, i.e., $\Pi_{\mathbf{X},\mathbf{a}}^t = \bigcup_{k\in\mathcal{K}} \Pi_{X_k,a_k}^t$ is the joint policy space.

Let $\mathbb{E}_\tau^\pi$ denote the expectation over the future states $X(\cdot)$ and the action process $a(\cdot)$, conditioned on past states $X(0)$, $X(1)$, $\cdots$, $X(\tau)$ and the policy $\pi \in \Pi_{X,a}^\tau$.

Consider any expected one-slot quantity $Q_{X(t)}^{a(t)}$ that depends on state $X(t)$ and action $a(t)$ at any time slot $t$. For any policy $\pi \in \Pi_{X,a}^\infty$ and any discount factor $0 \leqslant \beta < 1$, we define the infinite horizon average quantity as

$$\mathbb{B}_0^\pi\left\{Q_{X(\cdot)}^{a(\cdot)}, \beta, \infty\right\} := \lim_{T\to\infty} \frac{\sum_{t=0}^{T-1} \beta^t \mathbb{E}_t^\pi\left\{Q_{X(t)}^{a(t)}\right\}}{\sum_{t=0}^{T-1} \beta^t}. \quad (1)$$

In the following we consider the discount factor $\beta$ to be fixed and the horizon to be infinite, therefore we omit them in $\mathbb{B}_0^\pi\left\{Q_{X(\cdot)}^{a(\cdot)}, \beta, \infty\right\}$ and write briefly $\mathbb{B}_0^\pi\left\{Q_{X(\cdot)}^{a(\cdot)}\right\}$.

We are now ready to formulate the opportunistic scheduling problem faced by the server as below.

**Problem 1** (Optimum Opportunistic Scheduling)**.** For any discount factor $\beta$, the optimum opportunistic scheduling problem is to find a joint policy $\pi = (\pi_1, \cdots, \pi_K) \in \Pi_{\mathbf{X},\mathbf{a}}^\infty$

maximizing the total discounted reward (i.e., minimizing the total discounted cost), mathematically defined as below.

$$\text{(P):} \qquad \max_{\pi \in \Pi_{\mathbf{X},\mathbf{a}}} \mathbb{B}_0^{\pi} \left\{ \sum_{k \in \mathcal{K}} R_{k,X_k(\cdot)}^{a_k(\cdot)} \right\} \tag{2}$$

$$\text{s.t.} \qquad \sum_{k \in \mathcal{K}} a_k(t) = 1, \quad t = 0, 1, \cdots . \tag{3}$$

By relaxing constraint (3) and using Lagrange method [27], we have the following subproblem for each class $k \in \mathcal{K}$:

$$\text{(SP):} \qquad \max_{\pi_k \in \Pi_{X_k, a_k}} \mathbb{B}_0^{\pi_k} \left\{ R_{k,X_k(\cdot)}^{a_k(\cdot)} - \nu W_{k,X_k(\cdot)}^{a_k(\cdot)} \right\}. \tag{4}$$

Hence, our goal is to find the optimal policies $\pi_k^*$ for the subproblem (SP) $k \in \mathcal{K}$, and then construct a feasible joint policy $\pi = (\pi_1^*, \cdots, \pi_K^*)$ for the problem (P). In the following, we thus focus on the subproblem (SP) and drop the subscript $k$.

## IV. Indexability Analysis

In this section, we first make a special assumption about the channel state transition matrix, based on which we obtain the threshold structure of optimal scheduling strategy. We then establish the indexability of the optimum opportunistic scheduling problem under the posed assumption.

### A. Transition Matrices and Threshold Structure

**Definition 1** ([22]). Let $\Pi(N) := \{(x_1, \cdots, x_N) : \sum_{i=1}^N x_i = 1, x_1, \cdots, x_N \geqslant 0\}$. For $\mathbf{x}_1$, $\mathbf{x}_2 \in \Pi(N)$, then $\mathbf{x}_1$ first–order stochastically dominates $\mathbf{x}_2$, denoted as $\mathbf{x}_1 \geqslant_s \mathbf{x}_2$, if the following exists for $j = 1, 2, \cdots, N$,

$$\sum_{i=j}^{N} \mathbf{x}_{1i} \geqslant \sum_{i=j}^{N} \mathbf{x}_{2i}.$$

**Assumption 1.** The channel transition matrix[1] $\mathbf{Q}$ can be written as

$$\mathbf{Q} = \mathbf{O}_0 + \epsilon_1 \mathbf{O}_1 + \epsilon_2 \mathbf{O}_2 + \cdots + \epsilon_{2N-2} \mathbf{O}_{2N-2},$$

where $\mathbf{q} := [q_1, q_2, \cdots, q_N]^\mathsf{T}$, $\mathbf{O}_j$ is defined in (5) (see the top of the next page), $\epsilon_j$ and $\lambda$ are negative real numbers satisfying the following inequalities:

1) $\epsilon_j \leqslant 0$ for all $j$ $(1 \leqslant j \leqslant 2N-2)$,
2) $\lambda \leqslant \epsilon_j + \epsilon_{2N-1-j}$ for all $j$ $(1 \leqslant j \leqslant N)$.

Basically, the assumption implies that

1) Any two adjacent rows (i.e., $\mathbf{Q}_i, \mathbf{Q}_{i+1}$) of matrix $\mathbf{Q}$ differ in only two adjacent positions (i.e, $i, i+1$).
2) The first row vector of $\mathbf{Q}$ first-order stochastically dominates the second row, the second one dominates the third one, and so on, i.e., $\mathbf{Q}_1 \geqslant_s \mathbf{Q}_2 \geqslant_s \cdots \geqslant_s \mathbf{Q}_N$.

Assumption 1 leads to the following lemma on the threshold structure of optimum scheduling policy.

**Lemma 1** (Threshold structure). Under Assumption 1, for every real-valued $\nu$, there exists $n \in \mathcal{N} \cup \{-1\}$ such that the optimum scheduling policy only schedules transmission in channel states $\delta_{N-n} := \{m \in \mathcal{N} : m > n\}$.

*Proof.* Please see the technical report on arxiv. □

### B. Indexability Proof

To circumvent the long proof of indexability of Whittle index, we establish the indexability result by checking the LP-indexability condition [21].

For $\pi_k \in \Pi_{\pi_k, a_k}$, we introduce the concept of serving set, $\delta$ $(\delta \subseteq \mathcal{N}_k)$, such that the user is served if $n \in \delta$ and not served if $n \notin \delta$. By slightly introducing ambiguity, $\delta$ can also be regarded as a policy of serving the set $\delta$.

Thus, the subproblem (4) can be transformed to

$$\max_{\delta \in \mathcal{N}_k} \mathbb{B}_0^{\delta} \left\{ R_{k,X_k(\cdot)}^{a_k(\cdot)} - \nu W_{k,X_k(\cdot)}^{a_k(\cdot)} \right\}. \tag{6}$$

For further analysis, we define

$$\mathbb{R}_n^{\delta} := \frac{\mathbb{B}_0^{\delta} \left\{ R_{k,X_k(\cdot)}^{a_k(\cdot)} \right\}}{1-\beta}, \quad \mathbb{W}_n^{\delta} := \frac{\mathbb{B}_0^{\delta} \left\{ W_{k,X_k(\cdot)}^{a_k(\cdot)} \right\}}{1-\beta}. \tag{7}$$

By Lemma 1, if there exists price $\nu_n$ for $n \in \mathcal{N}'$ such that both transmitting and not transmitting are optimal for $\nu = \nu_n$, then there exists a set, $\delta^*$, such that both including state $n$ in $\delta^*$ and not including state $n$ lead to the same reward, i.e.,

$$\mathbb{R}_n^{\delta^* \cup \{n\}} - \nu_n \mathbb{W}_n^{\delta^* \cup \{n\}} = \mathbb{R}_n^{\delta^* \setminus \{n\}} - \nu_n \mathbb{W}_n^{\delta^* \setminus \{n\}}. \tag{8}$$

A straightforward consequence is that changing the action only in the initial period must also lead to the same reward, i.e.,

$$\mathbb{R}_n^{\langle 0, \delta^* \rangle} - \nu_n \mathbb{W}_n^{\langle 0, \delta^* \rangle} = \mathbb{R}_n^{\langle 1, \delta^* \rangle} - \nu_n \mathbb{W}_n^{\langle 1, \delta^* \rangle}, \tag{9}$$

where $\langle a, \delta^* \rangle$ is the policy that employs action $a$ in the initial period and then proceeds according to $\delta^*$.

Then, if $\mathbb{W}_n^{\langle 1, \delta^* \rangle} - \mathbb{W}_n^{\langle 0, \delta^* \rangle} \neq 0$, we have

$$\nu_n = \frac{\mathbb{R}_n^{\langle 1, \delta^* \rangle} - \mathbb{R}_n^{\langle 0, \delta^* \rangle}}{\mathbb{W}_n^{\langle 1, \delta^* \rangle} - \mathbb{W}_n^{\langle 0, \delta^* \rangle}}. \tag{10}$$

We further define

$$\nu_n^{\delta} := \frac{\mathbb{R}_n^{\langle 1, \delta \rangle} - \mathbb{R}_n^{\langle 0, \delta \rangle}}{\mathbb{W}_n^{\langle 1, \delta \rangle} - \mathbb{W}_n^{\langle 0, \delta \rangle}}. \tag{11}$$

**Definition 2** ([21]). Problem (4) is LP-indexable with price

$$\nu_n = \nu_n^{\delta_{N-n}} = \frac{\mathbb{R}_n^{\langle 1, \delta_{N-n} \rangle} - \mathbb{R}_n^{\langle 0, \delta_{N-n} \rangle}}{\mathbb{W}_n^{\langle 1, \delta_{N-n} \rangle} - \mathbb{W}_n^{\langle 0, \delta_{N-n} \rangle}}, \tag{12}$$

if the following conditions hold:

(1) $\forall\, n \in \mathcal{N}$, $\mathbb{W}_n^{\langle 1, \emptyset \rangle} - \mathbb{W}_n^{\langle 0, \emptyset \rangle} > 0$, $\mathbb{W}_n^{\langle 1, \mathcal{N} \rangle} - \mathbb{W}_n^{\langle 0, \mathcal{N} \rangle} > 0$ ;
(2) $\forall\, n \in \mathcal{N} \setminus \{N\}$, $\mathbb{W}_n^{\langle 1, \delta_{N-n} \rangle} - \mathbb{W}_n^{\langle 0, \delta_{N-n} \rangle} > 0$ and $\mathbb{W}_{n+1}^{\langle 1, \delta_{N-n} \rangle} - \mathbb{W}_{n+1}^{\langle 0, \delta_{N-n} \rangle} > 0$ ;
(3) For each real value $\nu$ there exists $n \in \mathcal{N} \cup \{-1\}$ such that the serving set $\delta_{N-n}$ is optimal.

In order to check the LP-indexability, we first characterize the four quantities in (12) under $\delta_{N-n}$ for any $n$.

[1] Note that we drop the subscript class index $k$ for conciseness.

$$
\mathbf{O}_j := \begin{cases} \mathbf{q}(\mathbf{1}_N)^{\mathsf{T}} + \lambda \mathbf{E}_N, & \text{if } j = 0, \\ [\underbrace{\mathbf{0}_N, \cdots, \mathbf{0}_N}_{N-j-1}, -\mathbf{1}_N^{N-j}, \mathbf{1}_N^{N-j}, \underbrace{\mathbf{0}_N, \cdots, \mathbf{0}_N}_{j-1}], & \text{if } 1 \leqslant j \leqslant N-1, \\ [\underbrace{\mathbf{0}_N, \cdots, \mathbf{0}_N}_{j-N}, \mathbf{1}_N - \mathbf{1}_N^{j-N+1}, \mathbf{1}_N^{j-N+1} - \mathbf{1}_N, \underbrace{\mathbf{0}_N, \cdots, \mathbf{0}_N}_{2N-2-j}], & \text{if } N \leqslant j \leqslant 2N-2. \end{cases} \tag{5}
$$

Based on balance equations, when $n$ is not chosen in the initial slot, we have the following in the matrix form

$$
(\mathbf{E}_N - \beta \mathbf{M}_0) \cdot \mathbf{r}_0 = \mathbf{c}_0, \tag{13}
$$

where,

$$
\begin{aligned}
&\mathbf{M}_0 = [\mathbf{Q}_1^{\mathsf{T}}, \cdots, \mathbf{Q}_n^{\mathsf{T}}, \quad \mathbf{Q}_{n+1}^{\mathsf{T}} \bar{\mu}_{n+1}, \cdots, \mathbf{Q}_N^{\mathsf{T}} \bar{\mu}_N]^{\mathsf{T}}, \\
&\mathbf{c}_0 = [-c, \cdots, -c, \quad -c, \quad -c\bar{\mu}_{n+1}, \cdots, -c\bar{\mu}_N]^{\mathsf{T}}, \\
&\mathbf{r}_0 = [\mathbb{R}_1^{\langle 0, \delta_{N-n} \rangle}, \cdots, \mathbb{R}_n^{\langle 0, \delta_{N-n} \rangle}, \mathbb{R}_{n+1}^{\langle 1, \delta_{N-n} \rangle}, \cdots, \mathbb{R}_N^{\langle 1, \delta_{N-n} \rangle}]^{\mathsf{T}}.
\end{aligned}
$$

Similarly, when $n$ is chosen in the initial slot, we have the following

$$
(\mathbf{E}_N - \beta \mathbf{M}_1) \cdot \mathbf{r}_1 = \mathbf{c}_1, \tag{14}
$$

where,

$$
\begin{aligned}
&\mathbf{M}_1 = [\mathbf{Q}_1^{\mathsf{T}}, \cdots, \mathbf{Q}_{n-1}^{\mathsf{T}}, \mathbf{Q}_n^{\mathsf{T}} \bar{\mu}_n, \cdots, \mathbf{Q}_N^{\mathsf{T}} \bar{\mu}_N]^{\mathsf{T}}, \\
&\mathbf{c}_1 = [-c, \cdots, -c, \quad -c\bar{\mu}_n, -c\bar{\mu}_{n+1}, \cdots, -c\bar{\mu}_N]^{\mathsf{T}}, \\
&\mathbf{r}_1 = [\mathbb{R}_1^{\langle 0, \delta_{N-n} \rangle}, \cdots, \mathbb{R}_{n-1}^{\langle 0, \delta_{N-n} \rangle}, \mathbb{R}_n^{\langle 1, \delta_{N-n} \rangle}, \cdots, \mathbb{R}_N^{\langle 1, \delta_{N-n} \rangle}]^{\mathsf{T}}.
\end{aligned}
$$

Thus, from (17) and (18), we can obtain

$$
\mathbb{R}_n^{\langle 0, \delta_{N-n} \rangle} = \mathbf{e}_n^{\mathsf{T}} (\mathbf{E}_N - \beta \mathbf{M}_0)^{-1} \mathbf{c}_0, \tag{15}
$$

$$
\mathbb{R}_n^{\langle 1, \delta_{N-n} \rangle} = \mathbf{e}_n^{\mathsf{T}} (\mathbf{E}_N - \beta \mathbf{M}_1)^{-1} \mathbf{c}_1. \tag{16}
$$

Similarly, replacing $\mathbf{c}_0$ and $\mathbf{c}_1$ by $\mathbf{1}_N - \mathbf{1}_N^n$ and $\mathbf{1}_N - \mathbf{1}_N^{n-1}$ from (13) and (14), respectively, we have

$$
(\mathbf{E}_N - \beta \mathbf{M}_0) \cdot \mathbf{w}_0 = \mathbf{1}_N - \mathbf{1}_N^n, \tag{17}
$$

$$
(\mathbf{E}_N - \beta \mathbf{M}_1) \cdot \mathbf{w}_1 = \mathbf{1}_N - \mathbf{1}_N^{n-1}, \tag{18}
$$

where,

$$
\begin{aligned}
&\mathbf{w}_0 = [\mathbb{W}_1^{\langle 0, \delta_{N-n} \rangle}, \cdots, \mathbb{W}_n^{\langle 0, \delta_{N-n} \rangle}, \mathbb{W}_{n+1}^{\langle 1, \delta_{N-n} \rangle}, \cdots, \mathbb{W}_N^{\langle 1, \delta_{N-n} \rangle}]^{\mathsf{T}}, \\
&\mathbf{w}_1 = [\mathbb{W}_1^{\langle 0, \delta_{N-n} \rangle}, \cdots, \mathbb{W}_{n-1}^{\langle 0, \delta_{N-n} \rangle}, \mathbb{W}_n^{\langle 1, \delta_{N-n} \rangle}, \cdots, \mathbb{W}_N^{\langle 1, \delta_{N-n} \rangle}]^{\mathsf{T}}.
\end{aligned}
$$

Further,

$$
\mathbb{W}_n^{\langle 0, \delta_{N-n} \rangle} = \mathbf{e}_n^{\mathsf{T}} (\mathbf{E}_N - \beta \mathbf{M}_0)^{-1} (\mathbf{1}_N - \mathbf{1}_N^n), \tag{19}
$$

$$
\mathbb{W}_n^{\langle 1, \delta_{N-n} \rangle} = \mathbf{e}_n^{\mathsf{T}} (\mathbf{E}_N - \beta \mathbf{M}_1)^{-1} (\mathbf{1}_N - \mathbf{1}_N^{n-1}). \tag{20}
$$

After obtaining the four critical quantities, we now check the LP-indexability condition.

**Lemma 2.** Under Assumption 1, for any $n \in \mathcal{N} \setminus \{N\}$, we have

(1) $\mathbb{W}_n^{\langle 1, \delta_{N-n} \rangle} > \mathbb{W}_n^{\langle 0, \delta_{N-n} \rangle}$,
(2) $\mathbb{W}_{n+1}^{\langle 1, \delta_{N-n} \rangle} > \mathbb{W}_{n+1}^{\langle 0, \delta_{N-n} \rangle}$.

*Proof.* Please see the technical report on arxiv.. □

**Lemma 3.** Under Assumption 1, Problem (4) is LP-indexable with price $\nu_n$ in (12).

*Proof.* According to Definition 2, we prove the indexability by checking the three conditions.

(1) Obviously, $\mathbb{W}_n^{\langle 0, \emptyset \rangle} = 0$, $\mathbb{W}_n^{\langle 1, \emptyset \rangle} \geqslant 1$, and $\mathbb{W}_n^{\langle 1, \mathcal{N} \rangle} = \frac{1}{1-\beta}$. For any $\delta$, $\mathbb{W}_n^{\delta} \leqslant \frac{1}{1-\beta}$, and further $\mathbb{W}_n^{\langle 0, \mathcal{N} \rangle} < \frac{1}{1-\beta}$.
(2) The second condition is proved in Lemma 2.
(3) The third condition is proved in Lemma 1.

Therefore, the LP-indexability is proved. □

Following Lemma 3, the following theorem states our main result on the indexability of Problem (4).

**Theorem 1** (indexability)**.** Under Assumption 1, we have

(1) if $\nu \leqslant \nu_n$, then it is optimal to serve the user in state $n$ ($n \in \mathcal{N}'$);
(2) if $\nu \geqslant \nu_n$, then it is optimal not to serve the user in state $n$ ($n \in \mathcal{N}'$)

Taking into account the time complexity $O(N^{2.4})$ of computing matrix determinant [17], the time complexity to calculate $\nu_n$ is $O(N^{2.4})$ according to (12). In the following section, we present a procedure to compute the index $\nu_n$, based on the structure of transition matrix $\mathbf{Q}$, with complexity $O(N)$.

## V. Scheduling Policy

After obtaining the index for each subproblem, we can construct the joint scheduling policy for the original problem. The optimum scheduling policy is to serve the user in $k^*(t)$ with the highest actual price, i.e.,

$$
k^*(t) := \begin{cases} \operatorname{argmax}_{k \in \mathcal{K}} \left[ \nu_{k, X_k(t)} \right], & \text{if } \nu_{k, X_k(t)} < \infty \\ \operatorname{argmax}_{k \in \mathcal{K}} \left[ \lim_{\beta \to 1} (1-\beta) \nu_{k, X_k(t)} \right], & \text{if } \nu_{k, X_k(t)} \to \infty. \end{cases} \tag{21}
$$

Actually, $\nu_{k, X_k(t)} < \infty$ always holds if $0 \leqslant \beta < 1$. It happens $\nu_{k, X_k(t)} \to \infty$ only when $\beta = 1$ and $X_k(t) = N_k$, corresponding to the average case. Therefore, the second item, $c_k \mu_k$, of Laurent expansion of $\nu_{k, X_k(t)}$ would be taken as the index in the case of $\beta = 1$. In particular, the marginal productivity index (MPI) is proposed in Algorithm 1.

The MPI scheduler always serves user currently with the best condition, and is one of the best-condition schedulers, which has stability property in Markovian setting shown in [18].

**Algorithm 1** MPI scheduler

**for** $t \in \mathcal{T}$
  $C \leftarrow$ number of system users in $N_k$ $(k \in \mathcal{K})$
**if** $C \geqslant 1$ **then**
  Serve one user in $N_k$ with $\max\{c_k \mu_k\}$ $(k \in \mathcal{K})$
  (breaking ties randomly)
**else**
  Serve the user $k^*(t)$ with highest index value
  (breaking ties randomly)
**end if**
**end for**

**Theorem 2.** The MPI scheduler with one server is maximally stable under arbitrary arrivals.

## VI. Numerical Simulation

In this section, we compare the proposed MPI scheduler with the following policies

- the $c\mu$ rule, $\nu_{k,n}^{c\mu} = c_k \mu_{k,n}$,
- the RB rule, $\nu_{k,n}^{\text{RB}} = \frac{c_k \mu_{k,n}}{\sum_{m=1}^{N_k} q_{k,m}^{\text{SS}} \mu_{k,m}}$,
- the PB rule, $\nu_{k,n}^{\text{PB}} = \frac{c_k \mu_{k,n}}{\mu_{k,N_k}}$,
- the SB rule, $\nu_{k,n}^{\text{SB}} = c_k \sum_{m=1}^{N_k} q_{k,m}^{\text{SS}}$,

where $q_{k,m}^{\text{SS}}$ is the stationary probability of state $m$ of an user of class $k$.

Specifically, we only consider the case with at most one user served at each time slot. If there are more than one user having the highest index value, we uniformly choose one of them. Also, we only consider two classes of users for a clear comparison in the performance difference of policies.

We let $\tau = 1.67$ msec for each slot for practical application [23]. The arrival probability for one new user of class $k$ is characterized by $\xi_k = \rho_k \mu_{k,N_k}$. For comparison, we consider transmission rate $s_{k,n}$ adopted in [23], and job size $\mathbb{E}_0[b_k] = 0.5$Mb for HTML, $\mathbb{E}_0[b_k] = 5$Mb for PDF, and $\mathbb{E}_0[b_k] = 50$Mb for MP3 . In this case, the departure probability is determined by $\mu_{k,n} = \tau s_{k,n} / \mathbb{E}_0[b_k]$. For simplicity, we assume $\rho_1 = \rho_2$, and the system load $\rho = \rho_1 + \rho_2$ is considered to vary from 0.3 to 1 for better presentation. The initial channel condition of a new user at the moment of entering system is assumed to be determined by the stationary probability vector, i.e., with probability $q_{k,m}^{\text{SS}}$ in state $m$ for a new user of class $k$. The parameter setting for the following scenarios is stated in **Table I**.

In this case, the users are divided into two different classes in which each user requires a job of expected size of 0.5Mb and has the same waiting cost $c_1 = c_2 = 1$. The channel state transition matrix is same, shown in **Table I**. Meanwhile, the second class of users has a better transmission rate than the first class. Our goal is to minimize the time-average number of user waiting for service in the system.

Under this setting, the two policies (SB and MPI) can be shown to bring about the same scheduling rule. Thus, **Figure 1** shows that the time-average waiting cost varies with system load $\rho$ for four policies, and the average number of users in system varies with time slots . Obviously, we observe that the behaviour of all policies is quite similar except RB. **Figure 1** clearly shows that MPI, PB and $c\mu$ perform better than RB since the average increase of users is 3.6 users per second for MPI, PB and $c\mu$ while 6.5 users per second for RB. The RB policy performs well at low loads but clearly has problems with stability since the policy appears to become unstable close to 0.85 at which point the time-average waiting cost begins rising very steeply. This indicates that the policy is not correctly balancing the utilization of both size information and rate information.

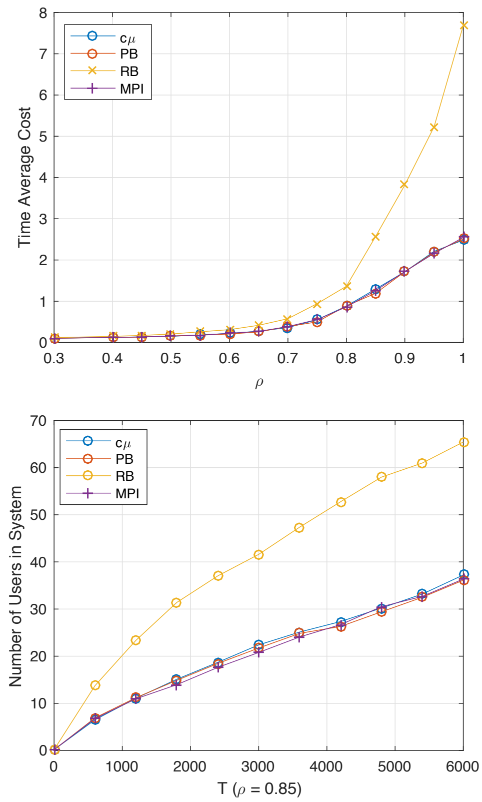


Fig. 1. Scenario 1 [U]: Time average waiting cost as a function of $\rho$; [D]: Average number of users in the system as a function of time.

## VII. Conclusion

In this paper, we have investigated the opportunistic scheduling problem involving multi-class multi-state time-varying Markovian channels. Generally, the problem can be formulated as a restless bandit problem. To the best of our knowledge, the existing work only established index policy for two-state channel process, and derived some limited results on multi-state time-varying channels under an assumption of indexability as a prerequisite. To fill this gap, for the class of state transition matrices characterized by our proposed sufficient condition, we have proved the indexability of the index policy. Simulation results show that the proposed index policy is effective in scheduling multiple class multi-state channels.

TABLE I
PARAMETERS ADOPTED IN SIMULATION.

| No. | $s_{k,n}$ (Mb/s) | $(c_1, c_2)$ | Channel Transition Matrices | Job Size (Mb) |
|---|---|---|---|---|
| 1 | $\begin{matrix} 8.4 & 16.8 & 25.2 \\ 26.88 & 44.688 & 80.64 \end{matrix}$ | $(1, 1)$ | $\begin{pmatrix} 0.00 & 0.80 & 0.20 \\ 0.30 & 0.50 & 0.20 \\ 0.30 & 0.60 & 0.10 \end{pmatrix}, \begin{pmatrix} 0.00 & 0.80 & 0.20 \\ 0.30 & 0.50 & 0.20 \\ 0.30 & 0.60 & 0.10 \end{pmatrix}$ | $0.5, 0.5$ |

## REFERENCES


[1] S. Aalto and P. Lassila. Flow-level stability and performance of channel-aware priority-based schedulers. In *6th EURO-NGI Conference on Next Generation Internet*, pages 1–8, Paris, France, 2010.

[2] S. Aalto, P. Lassila, and P. Osti. On the optimal trade-off between sprt and opportunistic scheduling. In *Proc. ACM Sigmetrics*, San Jose, CA, June 2011.

[3] 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.

[4] S. Aalto, P. Lassila, and P. Osti. Whittle index approach to size-aware scheduling for time-varying channels with multiple states. *Queueing Systems*, 83:195–225, 2016.

[5] U. Ayesta, M. Erausquin, and P. Jacko. A modeling framework for optimizing the flow-level scheduling with time-vayring channels. *Performance Evaluation*, 67:1014–1029, 2010.

[6] U. Ayesta, M. Erausquin, M. Jonckheere, and I. M. Verloop. Scheduling in a random environment: Stability and asymptotic optimality. *IEEE/ACM Trans. Netw.*, 21(1):258–271, Jan. 2013.

[7] P. Bender, P. Black, M. Grob, R. Padovani, N. Sindhushayana, and A. Viterbi. CDMA/HDR: a bandwidth-efficient high-speed wireless data service for nomadic users. *IEEE Commun. Mag.*, 38(7):70–77, Aug. 2000.

[8] T. Bonald. A score-based opportunistic scheduler for fading radio channels. In *Proc. of European Wireless*, pages 283–292, 2004.

[9] T. Bonald, S. Borst, N. Hegde, M. Jonckheere, and A. Proutiere. Flow-level performance and capacity of wireless networks with user mobility. *Queueing Systems*, 63:131–164, 2009.

[10] S. Borst. User-level performance of channel-aware scheduling algorithms in wireless data networks. *IEEE/ACM Trans. Netw.*, 13(3):636–647, Mar. 2005.

[11] S. Borst and M. Jonckheere. Flow-level stability of channel-aware scheduling algorithms. In *Proc. of WiOpt*, 2006.

[12] C. Buyukkoc, P. Varaiya, and J. Walrand. The $c\mu$ rule revisited. *Advances in Applied Probability*, 17(1):237–238, 1985.

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

[14] F. Cecchi and P. Jacko. Nearly-optimal scheduling of users with markovian time-varying transmission rates. *Performance Evaluation*, 99-100:16–36, 2016.

[15] P. Jacko. Value of information in optimal flow-level scheduling of users with markovian time-varying channels. *Performance Evaluation*, 68(11):1022–1036, 2011.

[16] P. Jacko, E. Morozov, L. Potakhina, and I. M. Verloop. Maximal flow-level stability of best-rate schedulers in heterogeneous wireless systems. *Trans. Emerg. Telecommun. Technol.*, page doi:10.1002/ett.2930, 2015.

[17] E. Kaltofen and G. Villard. On the complexity of computing determinants. *Computational Complexity*, 13(3-4):91–130, 2005.

[18] J. Kim, B. Kimb, J. Kimb, and Y. H. Baec. Stability of flow-level scheduling with markovian time-varying channels. *Performance Evaluation*, 70(2):148–159, 2013.

[19] R. Knopp and P. Humblet. Information capacity and power control in single-cell multiuser communications. In *Proc. IEEE Int. Conf. Commun.*, pages 331–335, Seattle, WA, Mar. 1995.

[20] H. J. Kushner and P. A. Whiting. Convergence of proportional fair sharing algorithms under general conditions. *IEEE Trans. Wireless Commun.*, 3(4):1250–1259, Aug. 2004.

[21] J. N. Mora. Characterization and computation of restless bandit marginal productivity indices. In *Proc. of ICST*, 2007.

[22] A. Mller and D. Stoyan. *Comparison Methods for Stochastic Models and Risk*. John Wiley & Sons Ltd., New York, NY, 2002.

[23] S. Sesia, I. Toufik, and M. Baker. *LTE–The UMTS Long Term Evolution: From Theory to Practice*. John Wiley & Sons Ltd., New York, NY, 2011.

[24] I. Taboada, P. Jacko, U. Ayesta, and F. Liberal. Opportunistic scheduling of flows with general size distribution in wireless time-varying channels. In *Proc. of ITC-26*, 2014.

[25] I. Taboada, F. Liberal, and P. Jacko. An opportunistic and non-anticipating size-aware scheduling proposal for mean holding cost minimization in time-varying channels. *Performance Evaluation*, 79:90–103, 2014.

[26] P. van de Ven, S. Borst, and S. Shneer. Instability of maxweight scheduling algorithms. In *Proc. IEEE Conf. on Computer Commun.*, pages 1701–1709, 2009.

[27] P. Whittle. Restless bandits: activity allocation in a changing world. *Journal of Applied Probability*, 25A:287–298, 1988.