Restless Bandits Approach to the Job Scheduling Problem and its Extensions*

Peter Jacko

BCAM — Basque Center for Applied Mathematics
jacko@bcamath.org

Abstract. This survey focuses on the job scheduling problem at a preemptive server, and on its extensions. We present a generic problem formulation within the framework of restless bandits and a unifying approach for designing the Whittle index rule. For the classic job scheduling problem at a preemptive server we show that this index rule is equivalent to the $c\mu$-rule, known to be the optimal solution. We further briefly discuss appealing Whittle index rules recently designed for the extensions of the job scheduling problem with abandonment and with time-varying service rate, which were computationally shown to achieve exceptionally good performance. Finally we list some open problems and possible future research directions.
Keywords: Markov decision process, restless bandits, job scheduling, threshold policies, indexability, index policy, Whittle index, $c\mu$-rule.
AMS 2000 subject classification: Primary 90B36, Secondary 90C40, 90C31, 68M20

1 Introduction

In this paper we present a restless bandits framework for the job scheduling problem and its extensions. The framework provides a powerful modeling setting, together with a unifying approach to the design of index rules proposed by Whittle [26]. This is based on establishing optimality of threshold policies and indexability of particular jobs. The resulting index rule for scheduling of multiple jobs is of high computational interest because it is separable across jobs (or classes of jobs) and often leads to simple appealing formulae that elucidate structural properties of the problem under the hard sample-path constraint of assigning a single server in every period. Index rules are usually significantly easier to implement than more complex cross-depending policies, and optimal

* I am grateful to Urtzi Ayesta who was my collaborator when entering the area of job scheduling and who strongly encouraged me to write this survey. NOTICE: This is the author's version of the work that was published as a book chapter with the following reference: Jacko, P. (2010): Restless Bandits Approach to the Job Scheduling Problem and its Extensions. In A. B. Piunovskiy, editor, Modern Trends in Controlled Stochastic Processes: Theory and Applications, Luniver Press, United Kingdom, pp. 248-267.

solutions are often intractable (i.e., uncomputable in a reasonable time) or too complex to be implementable in real-world systems.

In some systems (with a rather restricted dynamics), index rules provide an elegant optimal solution, such as the $c\mu$-rule for the job scheduling problem [23] and for the multi-class $G/Geo/1$ queue [7], or the Gittins index rule for the multi-armed (classic) bandit problem [11] and for its analogue with Poisson arrivals [25]. In more complex systems, which admit the restless bandit formulation, the Whittle index rule is typically reported to achieve a nearly-optimal performance, both in systems without and with arrivals. Such solution is in many models very appreciated, since the restless bandit problem was proven PSPACE-hard even in its deterministic version [21]. Due to their intractability, these problems are often alternatively addressed in some limiting regimes, such as fluid limit, diffusion limit, heavy traffic, or under the myopic criterion. However, an increasing body of literature on index rules, partly surveyed in this paper and partly in [20], suggests that such alternative solutions are usually systematically outperformed by the Whittle index rule.

2 Restless Bandits Framework for Job Scheduling

In this section we present a discrete-time MDP framework of the multi-armed restless bandit problem as it applies to job scheduling. In particular, a generic job is defined as a special case of the restless bandit. This extends the setting of the multi-armed (classic) bandit problem surveyed in [10], and is a special case of the dynamic and stochastic resource capacity allocation problem introduced in [12].

Consider the time slotted into time epochs $t \in \mathcal{T} := \{0, 1, 2, \dots\}$ at which decisions are made. Time epoch t corresponds to the beginning of time period t. Suppose that there are K jobs labeled $k \in \mathcal{K}' := \{1, 2, \dots, K\}$ awaiting service at time epoch 0 at a preemptive server. At every time epoch, jobs are competing for a server that decides at every time epoch which job to serve during that period. The server can serve at most one job at a time. We will denote the idling option by $k = 0$, and define $\mathcal{K} := \{0\} \cup \mathcal{K}'$. In the following we will present a generic problem formulation under the myopic, discounted, and time-average criterion over an infinite horizon.

2.1 Jobs

Since the capacity of the server is to serve one job at a time, every job can be allocated either zero or one server's capacity units. We denote by $\mathcal{A} := \{0, 1\}$ the *action space*, i.e., the set of allowable levels of capacity allocation. This action space is the same for every job k.

Each job k is defined independently of other jobs as the tuple

$$\left(\mathcal{N}_k, (\boldsymbol{W}_k^a)_{a \in \mathcal{A}}, (\boldsymbol{R}_k^a)_{a \in \mathcal{A}}, (\boldsymbol{P}_k^a)_{a \in \mathcal{A}}\right),$$

where

- $\mathcal{N}_k := \{0\} \cup \mathcal{N}'_k$ is the *state space* composed of state 0 representing completed job and of a finite set of possible states that non-completed job k can occupy;
- $\boldsymbol{W}_k^a := \left(W_{k,n}^a\right)_{n \in \mathcal{N}_k}$, where $W_{k,n}^a$ is the expected one-period capacity consumption, or *work* required by job k at state n if action a is decided at the beginning of a period;
- $\boldsymbol{R}_k^a := \left(R_{k,n}^a\right)_{n \in \mathcal{N}_k}$, where $R_{k,n}^a$ is the expected one-period *reward* earned by job k at state n if action a is decided at the beginning of a period;
- $\boldsymbol{P}_k^a := \left(p_{k,n,m}^a\right)_{n,m \in \mathcal{N}_k}$ is the job-k stationary one-period *state-transition probability matrix* if action a is decided at the beginning of a period, i.e., the (n, m)-element of the matrix, $p_{k,n,m}^a$, is the probability of moving to state m from state n under action a.

This definition of a restless bandit is the most general definition considered in the literature on finite-state-space models. In the discrete-time setting for job scheduling it is, however, natural to restrict the model by the *binary-work* assumption $W_{k,n}^a := a$, i.e., the job consumes all the capacity allocated.

The dynamics of job k is captured by the *state process* $X_k(\cdot)$ and the *action process* $a_k(\cdot)$, which correspond to state $X_k(t) \in \mathcal{N}_k$ and action $a_k(t) \in \mathcal{A}$ at all time epochs $t \in \mathcal{T}$. As a result of deciding action $a_k(t)$ in state $X_k(t)$ at time epoch t, the job k consumes the allocated capacity, earns the reward, and evolves its state for the time epoch $t + 1$. To avoid technical difficulties we will also assume that $R_{k,n}^a$ is bounded.

2.2 Idling Option

It is useful to assume that job $k = 0$ is the *static κ-priced job*, with a single state (therefore static) and obtaining reward κ if being allocated to the server, representing idling of the server. I.e., such a job k is defined by $\mathcal{N}_k := \{0\}, W_{k,0}^a := a, R_{k,0}^a := \kappa a, p_{k,0,0}^a := 1$ for all $a \in \mathcal{A}$. The role of this job is to cut-off capacity allocation to jobs whenever they are priced below κ and could represent also an alternative task, such as battery recharging.

2.3 Unified Optimization Criterion

Before describing the problem we first define an averaging operator that will allow us to discuss the infinite-horizon problem under the traditional β-discounted criterion and the time-average criterion in parallel.

Let the state process $X(\cdot)$, which corresponds to state $X(t)$ at all time epochs $t \in \mathcal{T}$, be adapted to the filtration $\mathcal{I} := \{\mathcal{I}(t) : t \in \mathcal{T}\}$, where $\mathcal{I}(t)$ is the set of information known at time epoch t. Let $\Pi_{\mathcal{I},a}$ be the set of all the policies that for each time epoch t decide (possibly *randomly*) action $a(t)$ based only on the information set $\mathcal{I}(t)$ (i.e., *non-anticipative*). Let $\mathbb{E}_\tau^\pi$ denote the expectation over the state process $X(\cdot)$ and over the action process $a(\cdot)$, conditioned on the information set $\mathcal{I}(\tau)$ and on policy π.

Consider any expected one-period quantity $Q_{X(t)}^{a(t)}$ that depends on state $X(t)$ and on action $a(t)$ at any time epoch t. For any policy $\pi \in \Pi_{\mathcal{I},a}$, any initial time epoch $\tau \in \mathcal{T}$, and any *discount factor* $0 \le \beta \le 1$ we define the infinite-horizon *β-average quantity*[1]

$$\mathbb{B}_\tau^\pi \left[Q_{X(\cdot)}^{a(\cdot)}, \beta, \infty \right] := \lim_{T \to \infty} \frac{\sum_{t=\tau}^{T-1} \beta^{t-\tau} \, \mathbb{E}_\tau^\pi \left[Q_{X(t)}^{a(t)} \right]}{\sum_{t=\tau}^{T-1} \beta^{t-\tau}}. \tag{1}$$

The β-average quantity recovers the traditionally considered quantities in the following three cases:

- *expected time-average quantity* when $\beta = 1$.
- *expected total β-discounted quantity*, scaled by constant $1-\beta$, when $0 < \beta < 1$;
- *myopic quantity* when $\beta = 0$.

Thus, when $\beta = 1$, the problem is formulated under the *time-average criterion*, whereas when $0 < \beta < 1$ the problem is considered under the *β-discounted criterion*. The remaining case when $\beta = 0$ reduces to a static problem and hence is considered in order to define a *myopic policy*. In the following we consider the discount factor β to be fixed and the horizon to be infinite, therefore we omit them in the notation and write briefly $\mathbb{B}_\tau^\pi \left[Q_{X(\cdot)}^{a(\cdot)} \right]$.

2.4 Optimization Problem

We now describe in more detail the problem we consider. Let us define the joint information set as

$$\mathcal{I}(t) := \{X_k(0), a_k(0), X_k(1), \ldots, a_k(t-1), X_k(t) \text{ for each } k \in \mathcal{K}\},$$

so that for each $k \in \mathcal{K}$ the state process $X_k(\cdot)$ is adapted to the joint filtration $\mathcal{I} := \{\mathcal{I}(t) : t \in \mathcal{T}\}$. Therefore also the joint state-process $\boldsymbol{X}(\cdot) := (X_k(\cdot))_{k \in \mathcal{K}}$ is adapted to the filtration $\mathcal{I}$. Let $\Pi_{\mathcal{I},\boldsymbol{a}}$ be the set of all the policies that for each time epoch t decide (possibly *randomized*) joint action $\boldsymbol{a}(t) := (a_k(t))_{k \in \mathcal{K}}$ based only on the information set $\mathcal{I}(t)$ (i.e., *non-anticipative*), i.e., $\Pi_{\mathcal{I},\boldsymbol{a}}$ is the *joint policy space*.

For any discount factor β, the problem is to find a joint policy $\boldsymbol{\pi}$ maximizing the *objective* given by the β-average aggregate reward starting from the initial

[1] For definiteness, we consider $\beta^0 = 1$ for $\beta = 0$.

time epoch 0 subject to the family of *sample path* allocation constraints, i.e.,

$$\max_{\boldsymbol{\pi} \in \Pi_{\mathcal{I},\boldsymbol{a}}} \mathbb{B}_0^{\boldsymbol{\pi}} \left[\sum_{k \in \mathcal{K}} R_{k,X_k(\cdot)}^{a_k(\cdot)} \right] \tag{P}$$

$$\text{subject to } \mathbb{E}_t^{\boldsymbol{\pi}} \left[\sum_{k \in \mathcal{K}} a_k(t) \right] = 1, \text{ for all } t \in \mathcal{T}$$

Note that the constraint could equivalently be expressed by restricting $\Pi_{\mathcal{I},\boldsymbol{a}}$ to policies that at every $t \in \mathcal{T}$ satisfy $\sum_{k \in \mathcal{K}} a_k(t) = 1$ for any possible realization of $\mathcal{I}(t)$.

3 Relaxations and Decomposition of the Problem

Problem (P) is difficult to solve due to the sample path constraint. We will relax that constraint and show that the resulting problem can be decomposed into single-job subproblems, which will be addressed in the next section.

3.1 Relaxations

For notational reasons we will use the fact that $W_{k,X_k(t)}^{a_k(t)} = a_k(t)$ (due to the binary-work assumption) and instead of the allocation constraints in (P) we will consider the sample path *consumption* constraints

$$\mathbb{E}_t^{\boldsymbol{\pi}} \left[\sum_{k \in \mathcal{K}} W_{k,X_k(t)}^{a_k(t)} \right] = 1, \text{ for all } t \in \mathcal{T}. \tag{2}$$

These constraints imply the *epoch-t expected* consumption constraints (evaluated at time epoch 0 instead of t),

$$\mathbb{E}_0^{\boldsymbol{\pi}} \left[\sum_{k \in \mathcal{K}} W_{k,X_k(t)}^{a_k(t)} \right] = 1, \text{ for all } t \in \mathcal{T} \tag{3}$$

requiring that the capacity be fully allocated at every time epoch if conditioned on $\mathcal{I}(0)$. Finally, we may require this constraint to hold only on β-average, as the *β-average capacity consumption constraint*

$$\mathbb{B}_0^{\boldsymbol{\pi}} \left[\sum_{k \in \mathcal{K}} W_{k,X_k(\cdot)}^{a_k(\cdot)} \right] = \mathbb{B}_0^{\boldsymbol{\pi}} [1] . \tag{4}$$

Using $\mathbb{B}_0^{\boldsymbol{\pi}}[1] = 1$, we obtain the following *relaxation* of problem (P),

$$\max_{\boldsymbol{\pi} \in \Pi_{\mathcal{I},\boldsymbol{a}}} \mathbb{B}_0^{\boldsymbol{\pi}} \left[\sum_{k \in \mathcal{K}} R_{k,X_k(\cdot)}^{a_k(\cdot)} \right] \tag{$\mathrm{P^W}$}$$
$$\text{subject to } \mathbb{B}_0^{\boldsymbol{\pi}} \left[\sum_{k \in \mathcal{K}} W_{k,X_k(\cdot)}^{a_k(\cdot)} \right] = 1.$$

Such a relaxation was introduced by Whittle in [26] for the multi-armed restless bandit problem under the time-average criterion, and extended in [16] for the discounted criterion. The above arguments thus provide a proof of the following result.

Proposition 1. *Problem* ($\mathrm{P^W}$) *is a relaxation of problem* (P).

The *Whittle relaxation* ($\mathrm{P^W}$) can be approached by traditional Lagrangian methods, introducing a real-valued Lagrangian parameter, say ν, to dualize the constraint, obtaining thus the following Lagrangian relaxation,

$$\max_{\boldsymbol{\pi} \in \Pi_{\mathcal{I},\boldsymbol{a}}} \mathbb{B}_0^{\boldsymbol{\pi}} \left[\sum_{k \in \mathcal{K}} R_{k,X_k(\cdot)}^{a_k(\cdot)} - \nu \sum_{k \in \mathcal{K}} W_{k,X_k(\cdot)}^{a_k(\cdot)} \right] + \nu. \tag{$\mathrm{P}_\nu^\mathrm{L}$}$$

The Lagrangian parameter ν can be interpreted as a service cost and we emphasize that ($\mathrm{P}_\nu^\mathrm{L}$) is an unconstrained problem. The classic Lagrangian result says the following:

Proposition 2. *For any ν, problem* ($\mathrm{P}_\nu^\mathrm{L}$) *is a relaxation of problem* ($\mathrm{P^W}$)*, and further a relaxation of problem* (P).

Note finally that by the definition of relaxation, ($\mathrm{P}_\nu^\mathrm{L}$) for every ν provides an upper bound for the optimal value of both problem ($\mathrm{P^W}$) and problem (P).

3.2 Decomposition into Single-User Subproblems

We now set out to decompose the optimization problem ($\mathrm{P}_\nu^\mathrm{L}$) as it is standard for Lagrangian relaxations, considering ν as a parameter. Notice that any joint policy $\boldsymbol{\pi} \in \Pi_{\mathcal{I},\boldsymbol{a}}$ defines a set of single-job policies $\widetilde{\pi}_k$ for all $k \in \mathcal{K}$, where $\widetilde{\pi}_k$ is a randomized policy adapted to the *joint* filtration $\mathcal{I}$ and deciding the *job-k* action-process $a_k(\cdot)$. We will write $\widetilde{\pi}_k \in \Pi_{\mathcal{I},a_k}$. We will therefore study the job-k subproblem

$$\max_{\widetilde{\pi}_k \in \Pi_{\mathcal{I},a_k}} \mathbb{B}_0^{\widetilde{\pi}_k} \left[R_{k,X_k(\cdot)}^{a_k(\cdot)} - \nu W_{k,X_k(\cdot)}^{a_k(\cdot)} \right]. \tag{5}$$

4 Solution

In this section we will identify a set of optimal policies $\widetilde{\pi}_k^*$ to (5) for all jobs k, and using them we will construct a joint policy $\boldsymbol{\pi}$ feasible though not necessarily optimal for problem (P).

4.1 Optimal Solution to Single-User Subproblem

In this section we list some known approaches to the solution of problem (5). We will be interested in two related structural properties of the problem, which are *optimality of threshold policies* and *indexability*, formally defined below. These have to be established (often under additional problem-specific assumptions) for every particular model, since neither is true in general.

Definition 1 (Optimality of Threshold Policies). *We say that problem* (5) *is* optimally solvable by threshold policies*, if for every real-valued ν there exists $n \in \mathcal{N}_k \cup \{-1\}$ such that threshold policy serving in states $\mathcal{S}_{N_k - n} := \{m \in \mathcal{N}_k : m > n\}$ and not serving otherwise is optimal for problem* (5)*.*

We note that for practical purposes it may be interesting to have the above property hold even for a subset of values of ν (for instance, $\nu = 0$), but we require it to hold for all real-valued ν in order to show its relationship with indexability. Further, other types of threshold policies may be relevant in some models, for instance $\{m \in \mathcal{N}_k : m < n\}$, but such models can be transformed into our setting by re-labeling the states, so the above definition is of full generality while keeping the notational complexity low.

Although one could define indices in a wide variety of ad hoc ways based on the peculiarities of and researcher's intuition for the model in hand, of our interest will be the index proposed by Whittle in [26], that provides a unifying approach to the design of index rules of a remarkably general modeling validity. Moreover, this index often furnishes a nearly-optimal solution, and typically recovers the optimal index rule if such exists.

Definition 2 (Indexability). *We say that ν-parameter problem* (5) *is* indexable*, if there exist unique values $-\infty \le \nu_{k,n} \le \infty$ for all $n \in \mathcal{N}_k$ such that the following holds for every state $n \in \mathcal{N}_k$:*

1. *if $\nu_{k,n} \ge \nu$, then it is optimal to serve job k in state n, and*
2. *if $\nu_{k,n} \le \nu$, then it is optimal not to serve job k in state n.*

The function $n \mapsto \nu_{k,n}$ is called the (Whittle) index*, and $\nu_{k,n}$'s are called the* (Whittle) index values*.*

We note that this definition generalizes the definitions of indexability in [26] and in [20] (see also the references therein), since here we allow the index values to take also values $-\infty$ and ∞. An immediate consequence of the two definitions is formulated in the following theorem.

Theorem 1. *If problem* (5) *is indexable and the index is non-decreasing, i.e., $\nu_{k,0} \le \nu_{k,1} \le \cdots \le \nu_{k,N_k}$, then problem* (5) *is optimally solvable by threshold policies. Moreover, for a given ν the optimal threshold policy is $\mathcal{S}_{N_k - n^*}$ with $n^* \in \mathcal{N}_k \cup \{-1\}$ such that $\nu_{k,n^*} \le \nu \le \nu_{k,n^*+1}$ (defining $\nu_{k,-1} := -\infty, \nu_{k,N_k+1} := \infty$).*

However, for sufficient conditions of indexability (including conditions for the opposite implication of the above theorem) we will need a more careful analysis, presented next.

4.2 Indexability and Index Evaluation under Discounted Criterion

We will next focus on the case $\beta < 1$, i.e., the problem under the discounted criterion. The results presented in this subsection were mostly introduced in [16, 19] and were also surveyed in [20]. We adapt them as they apply to the framework of this paper.

Problem (5) is a standard stationary MDP problem, for which it is well known that there is an optimal policy which is deterministic (i.e., non-randomized), stationary (i.e., Markovian), and independent of the initial state [22, Chapter 6]. In particular, this implies that there exists an optimal policy based only on the job-k information sets

$$\mathcal{I}_k(t) := \{X_k(0), a_k(0), X_k(1), \ldots, a_k(t-1), X_k(t)\}.$$

Indeed, policy $\widetilde{\pi}_k \in \mathit{\Pi}_{\mathcal{I},a_k}$ that depends on the joint information sets $\mathcal{I}(t)$ can be seen as a randomized policy, since the job-l state-process $X_l(\cdot)$ for $l \neq k$ is not influenced by the job-k action-process $a_k(\cdot)$ prescribed by $\widetilde{\pi}_k$.

Therefore, in order to find an optimal policy to problem (5) it is enough to concentrate on policies $\pi_k \in \mathit{\Pi}_{\mathcal{I}_k,a_k}$ that are stationary. Every such policy can be represented in terms of a *serving set* $\mathcal{S} \subseteq \mathcal{N}_k$, which prescribes to serve whenever the job is in state $n \in \mathcal{S}$ and not to serve whenever the job is in state $n \notin \mathcal{S}$. Thus, an optimal policy to problem (5) can be obtained by solving

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

Notice that (6) is a parametric bi-objective optimization problem and every policy (i.e., serving set) $\mathcal{S}$ is associated with a bi-dimensional point $\mathbb{B}_0^{\mathcal{S}} \left[W_{k,X_k(\cdot)}^{a_k(\cdot)} \right]$, $\mathbb{B}_0^{\mathcal{S}} \left[R_{k,X_k(\cdot)}^{a_k(\cdot)} \right]$. If depicted in a plane with works on the x-axis and rewards on the y-axis, then the optimal policies to (6) lie on the upper boundary of such a region, since the parameter ν gives the slope of the supporting hyperplane (a line in this case) defining an optimum point (i.e., an optimal policy).

We will next analyze scaled quantities $\mathbb{B}_0^{\mathcal{S}} \left[R_{k,X_k(\cdot)}^{a_k(\cdot)} \right] / (1-\beta)$, writing briefly $\mathbb{R}_{k,n}^{\mathcal{S}}$ if the initial state $X_k(0) = n \in \mathcal{N}_k$. Analogously, we will write briefly $\mathbb{W}_{k,n}^{\mathcal{S}}$ and we denote the value function under policy $\mathcal{S}$ by $\mathbb{V}_{k,n}^{\mathcal{S}} := \mathbb{R}_{k,n}^{\mathcal{S}} - \nu \mathbb{W}_{k,n}^{\mathcal{S}}$. These scaled quantities correspond to the usual quantities under the β-discounted criterion.

Let us denote the optimal value function by $\mathbb{V}_{k,n}^{*}$, and an optimal stationary policy by $\mathcal{S}^*$. The Bellman equation for state $n \in \mathcal{N}_k$ is

$$\mathbb{V}_{k,n}^{*} = \max_{a\in\mathcal{A}} \left\{ R_{k,n}^{a} - \nu W_{k,n}^{a} + \beta \sum_{m\in\mathcal{N}_k} p_{k,n,m}^{a} \mathbb{V}_{k,m}^{*} \right\}.$$

Let us denote

$$\mathbb{R}_{k,n}^{\langle a,\mathcal{S}\rangle} := R_{k,n}^{a} + \beta \sum_{m\in\mathcal{N}_k} p_{k,n,m}^{a} \mathbb{R}_{k,m}^{\mathcal{S}}, \quad \mathbb{W}_{k,n}^{\langle a,\mathcal{S}\rangle} := W_{k,n}^{a} + \beta \sum_{m\in\mathcal{N}_k} p_{k,n,m}^{a} \mathbb{W}_{k,m}^{\mathcal{S}}.$$

That is, policy $\langle a, \mathcal{S}\rangle$ is the policy that employs action a in the initial period and then proceeds according to $\mathcal{S}$. We will also consider the rates

$$\nu_{k,n}^{\mathcal{S}} := \frac{\mathbb{R}_{k,n}^{\langle 1,\mathcal{S}\rangle} - \mathbb{R}_{k,n}^{\langle 0,\mathcal{S}\rangle}}{\mathbb{W}_{k,n}^{\langle 1,\mathcal{S}\rangle} - \mathbb{W}_{k,n}^{\langle 0,\mathcal{S}\rangle}}. \tag{7}$$

The following theorem characterizes the index values if problem is indexable.

Theorem 2. *If problem* (5) *is indexable, then either $\nu_{k,n} \in \{-\infty, \infty\}$, or there is a policy $\mathcal{S}$ such that $\nu_{k,n} = \nu_{k,n}^{\mathcal{S}}$.*

Proof. If index value $\nu_{k,n}$ for $n \in \mathcal{N}_k$ is finite, then both serving and not serving is optimal in state n if $\nu = \nu_{k,n}$. The Bellman equation implies

$$R_{k,n}^0 - \nu_{k,n} W_{k,n}^0 + \beta \sum_{m \in \mathcal{N}_k} p_{k,n,m}^0 \mathbb{V}_{k,m}^{\mathcal{S}^*} = R_{k,n}^1 - \nu_{k,n} W_{k,n}^1 + \beta \sum_{m \in \mathcal{N}_k} p_{k,n,m}^1 \mathbb{V}_{k,m}^{\mathcal{S}^*}$$

and using $\mathbb{V}_{k,n}^{\mathcal{S}^*} := \mathbb{R}_{k,n}^{\mathcal{S}^*} - \nu \mathbb{W}_{k,n}^{\mathcal{S}^*}$ we obtain

$$\mathbb{R}_{k,n}^{\langle 0,\mathcal{S}^*\rangle} - \nu_{k,n} \mathbb{W}_{k,n}^{\langle 0,\mathcal{S}^*\rangle} = \mathbb{R}_{k,n}^{\langle 1,\mathcal{S}^*\rangle} - \nu_{k,n} \mathbb{W}_{k,n}^{\langle 1,\mathcal{S}^*\rangle},$$

If index value $\nu_{k,n}$ for $n \in \mathcal{N}_k$ is finite, then $\mathbb{W}_{k,n}^{\langle 1,\mathcal{S}^*\rangle} - \mathbb{W}_{k,n}^{\langle 0,\mathcal{S}^*\rangle} \neq 0$ (otherwise both serving and not serving would be optimal for all values of ν, so $\nu_{k,n}$ would not be unique), therefore we have $\nu_{k,n} = \nu_{k,n}^{\mathcal{S}^*}$. □

Next we present two definitions that, as we will see later, lead to a more specific characterization of the index values, which is of great practical interest for deriving closed-form formulae or a fast algorithmic evaluation of index values. The first definition arises from a parametric linear programming (LP) approach in [19], and the second one from a polyhedral approach based on partial conservation laws (PCL) in [16].

Definition 3 (LP Conditions for Indexability). *We say that problem* (5) *is* LP-indexable with respect to threshold policies $\mathcal{S}_{N_k - n}$ *for $n \in \mathcal{N}_k \cup \{-1\}$, if the following conditions hold:*

1. $\mathbb{W}_{k,n}^{\langle 1,\emptyset\rangle} - \mathbb{W}_{k,n}^{\langle 0,\emptyset\rangle} > 0$ *and* $\mathbb{W}_{k,n}^{\langle 1,\mathcal{N}\rangle} - \mathbb{W}_{k,n}^{\langle 0,\mathcal{N}\rangle} > 0$ *for all $n \in \mathcal{N}_k$;*
2. $\mathbb{W}_{k,n}^{\langle 1,\mathcal{S}_{N_k-n}\rangle} - \mathbb{W}_{k,n}^{\langle 0,\mathcal{S}_{N_k-n}\rangle} > 0$ *and* $\mathbb{W}_{k,n+1}^{\langle 1,\mathcal{S}_{N_k-n}\rangle} - \mathbb{W}_{k,n+1}^{\langle 0,\mathcal{S}_{N_k-n}\rangle} > 0$ *for each $n \in \mathcal{N}_k \setminus \{N_k\}$;*
3. *problem* (5) *is optimally solvable by threshold policies.*

Definition 4 (PCL Conditions for Indexability). *We say that problem* (5) *is* PCL-indexable with respect to threshold policies $\mathcal{S}_{N_k - n}$ *for $n \in \mathcal{N}_k \cup \{-1\}$, if the following conditions hold:*

1. $\mathbb{W}_{k,m}^{\langle 1,\mathcal{S}_{N_k-n}\rangle} - \mathbb{W}_{k,m}^{\langle 0,\mathcal{S}_{N_k-n}\rangle} > 0$ *for each $n \in \mathcal{N}_k \cup \{-1\}\}$ and for all $m \in \mathcal{N}_k$;*
2. $\nu_{k,0}^{\mathcal{S}_{N_k-0}} \le \nu_{k,1}^{\mathcal{S}_{N_k-1}} \le \cdots \le \nu_{k,N_k}^{\mathcal{S}_{N_k-N_k}}$.

The relationship between these two sets of conditions and indexability is given next.

Theorem 3.

1. *If problem* (5) *is PCL-indexable, then it is LP-indexable.*
2. *If problem* (5) *is LP-indexable, then it is indexable, the index is non-decreasing, and the index values satisfy* $\nu_{k,n} = \nu_{k,n}^{\mathcal{S}_{N_k-n}} = \nu_{k,n}^{\mathcal{S}_{N_k-(n-1)}}$.

This theorem shows under what conditions optimality of threshold policies implies indexability. We also note that it can be shown that if a job is indexable with finite index values, then it is LP-indexable with respect to some type of threshold policies (not necessarily of type $\mathcal{S}_{N_k-n}$), hence it provides a sufficient and necessary condition for indexability in this general sense (under finite index values). See [20] for a comprehensive survey on this topic.

4.3 Indexability and Index Evaluation under Time-Average Criterion

It was shown in [17, Section 6.5] that under suitable ergodicity conditions indexability of the problem under the discounted criterion extends to the time-average criterion, and the index values for the time-average criterion can be computed by simply taking the limit $\beta \to 1$ of the index values for the discounted criterion. However, the jobs we consider in this paper are typically not ergodic since state 0 (representing completed job) would typically be absorbing. Such an approach is therefore not theoretically validated for non-ergodic jobs, but the modeling experience (for instance, index values obtained from fluid limit models or by the interchange argument) and the exceptionally good performance of the undiscounted indices in computational experiments suggest its plausibility and relevance.

4.4 Some Useful Results for Establishing Indexability

The results we have presented so far indicate a close relationship (but not equivalence) between optimality of threshold policies and indexability. While threshold policies are important in single-job problems (such as (5)), indexability is useful for designing index rules for multi-job problems (such as (P)). Moreover, the above theorems provide a strong playground for designing efficient algorithms for calculation of both index values and optimal thresholds.

Typically, much of the analysis must be done for each specific model, either studying the Bellman equation and/or work and reward measures defined above in order to establish the structural properties. Usefulness and importance of LP-indexability is elevated by availability of results in a wide variety of models showing optimality of threshold policies and also by existence of models that are indexable but not PCL-indexable (at this moment two such models are known: classic bandits with switching delays, as remarked in [20], and jobs with time-varying service rate analyzed in [4]). On the other hand, PCL-indexability may be a good starting point for models for which no structural results are known, since it may lead the researcher to identify and understand (rather general) conditions under which index monotonicity holds.

Next we list some results we have found useful in analysis of specific models. The first two lemmas are simple exercises, but they are of general validity within the framework of this paper.

Lemma 1. *Condition 1. of PCL-indexability implies conditions 1. and 2. of LP-indexability.*

Lemma 2. *Condition 1. of LP-indexability is satisfied under the binary-work assumption.*

The following result is an adaptation to our setting of the fact that classic bandits are indexable, which was in fact established by introducing the Gittins index [11, 9], and later elucidated by [26] in the framework of restless bandits. Note that its validity under the time-average case was established only under the ergodicity assumption.

Lemma 3 ([26, Proposition 4]). *If $\boldsymbol{P}_k^0 = \boldsymbol{I}$ (where $\boldsymbol{I}$ is an identity matrix), then the job is indexable.*

If the job has only two states, then indexability is guaranteed by the next lemma. Again, its validity under the time-average case was established only under the ergodicity assumption.

Lemma 4 ([16, Section 6.1]). *If $\mathcal{N}_k = \{0, 1\}$, then the job is indexable.*

The following result implies that the myopic index (obtained when $\beta = 0$) always exists, which may be of practical interest if the discounted and the time-average index characterization is of prohibitive complexity that impedes its implementation, or if the problem is not indexable for the desired values of the discount factor.

Lemma 5 ([16, Corollary 5]). *If the discount factor β is small enough, then any job is indexable.*

Finally, in order to prove the monotonicity required in condition 2. of PCL-indexability, in some models it is easier to relate the rates $\nu_{k,n}^{\mathcal{S}}$'s under the same policy, and then to apply the following claim.

Lemma 6 ([17, Proposition 6.4(c)]). *If condition 1. of PCL-indexability holds, then $\nu_{k,n}^{\mathcal{S}_{N_k - n}} \leq \nu_{k,n+1}^{\mathcal{S}_{N_k-(n+1)}}$ is equivalent to $\nu_{k,n}^{\mathcal{S}_{N_k - n}} \leq \nu_{k,n+1}^{\mathcal{S}_{N_k - n}}$.*

4.5 Optimal Solution to Relaxations

The vector of policies $\boldsymbol{\pi}^* := (\widetilde{\pi}^*_k)_{k\in\mathcal{K}}$ identified by the indices of all the jobs is formed by mutually independent single-job optimal policies, therefore this vector is an optimal policy to the Lagrangian relaxation ($\mathrm{P}^{\mathrm{L}}_{\nu}$).

Since a finite-state MDP admits an LP formulation using the standard *state-action frequency* variables (as observed in [16]), strong LP duality implies that there exists ν^* (possibly depending on the joint initial state) such that the Lagrangian relaxation ($\mathrm{P}^{\mathrm{L}}_{\nu^*}$) achieves the same objective value as (P^{W}). Further, if $\nu^* \neq 0$, then LP complementary slackness ensures that the β-average capacity constraint (4) is satisfied by any optimal solution to ($\mathrm{P}^{L}_{\nu^*}$).

4.6 Solution to Original Problem

Since the original problem requires to allocate the server to exactly one option (one of the jobs or idling), then at any time epoch t the index rule prescribes to allocate the server to the job $k^*(t)$ with the highest actual index value, i.e.,

$$k^*(t) := \arg\max_{k\in\mathcal{K}} \nu_{k,X_k(t)}.$$

Under $\beta < 1$, the ties are resolved arbitrarily.

Under the time-average criterion (i.e., $\beta = 1$), we will consider the tie-breaking rule based on the second term of the Laurent expansion of the (first-order) index value $\nu_{k,n}$. This tie-breaking may itself have ties; these are resolved arbitrarily.

Under the assumption *frozen if not served*, i.e., $\boldsymbol{P}^0_k = \boldsymbol{I}$ for all $k \in \mathcal{K}$ (where $\boldsymbol{I}$ is an identity matrix), the above index rules were shown optimal. [23] first considered two-state jobs in the classic job scheduling problem and obtained the $c\mu$-rule, which was later extended to the multi-armed bandit problem under the discounted [11] and the time-average criterion [15, 14], respectively.

If this assumption does not hold, then the above rules are not optimal in general, but are typically reported a nearly-optimal performance [20]. In fact, already Whittle [26] conjectured a form of asymptotic optimality of the (first-order) index rule, which was proved in [24] for symmetric jobs under certain technical conditions. It has been observed in computational experiments for several models that under the time-average criterion the implementation of the second-order index as a tie-breaking rule when the (first-order) index values of different jobs are equal is crucial and may significantly improve system's performance (in comparison to random tie-breaking). For a particular restless bandit problem in continuous time, the tie-breaking based on the Maclaurin expansion was first proposed in [18].

5 Particular Models

In this section we present several models that fall within the framework of section 2 and can be solved by the approach described in section 4. First we give a detailed account of the classic job scheduling problem showing that the Whittle index rule is equivalent to the $c\mu$-rule, known to be optimal. Then we survey recently obtained results on extensions of the classic job scheduling problem that give rise to more complex models, for which, however, the Whittle index rule was computationally confirmed to perform exceptionally well, although not optimal anymore.

It is interesting that the $c\mu$-rule is optimal also when arrivals of new jobs are allowed in the classic job scheduling problem, and for arbitrary arrival processes [7]. One may therefore be prompted to evaluate the performance of the Whittle index rules in more complex models with arrivals. It was actually found in computational experiments for some models (detailed below) that also in such systems its performance is nearly-optimal and typically outperforming solutions derived by other methods or proposed in an ad hoc way.

5.1 Classic Job Scheduling

Consider $K \geq 1$ jobs waiting for service at a server that can serve at most one job at a time. The server is preemptive (i.e., the service of a job can be interrupted at any time epoch even if not completed), so all its capacity is available at every time epoch. We have the action space $\mathcal{A} := \{0, 1\}$, where action 0 means allocating zero capacity (i.e., "not serving"), and action 1 means allocating full capacity (i.e., "serving").

Let $\mu_k > 0$ be the probability that the service of job k is completed within one period (if served) and let $c_k > 0$ be the holding cost per period incurred for job k waiting. The server can also be left idle, denoting this option by $k = 0$, or allocated to a customer with a completed job. Thus, there are $K + 1$ competing options and the task is to decide to which option the server should be allocated. The joint goal is to minimize the aggregate expected holding cost over an infinite horizon.

Thus, the idling option $k = 0$ is defined as the static 0-priced competitor and we define job $1 \leq k \leq K$ with

- state space $\mathcal{N}_k := \{0, 1\}$;
- expected one-period work

$$
\begin{aligned}
W_{k,0}^1 &:= 1, & W_{k,1}^1 &:= 1, \\
W_{k,0}^0 &:= 0, & W_{k,1}^0 &:= 0;
\end{aligned}
$$

- expected one-period reward, i.e., the negative of the holding cost expected to be paid at the next time epoch,

$$
\begin{aligned}
R_{k,0}^1 &:= 0, & R_{k,1}^1 &:= -c_k \cdot (1 - \mu_k) - 0 \cdot \mu_k, \\
R_{k,0}^0 &:= 0, & R_{k,1}^0 &:= -c_k;
\end{aligned}
$$

– one-period state-transition probability matrices

$$\boldsymbol{P}_k^1 := \begin{array}{c} \\ 0 \\ 1 \end{array} \!\!\begin{array}{c} \begin{array}{cc} 0 & 1 \end{array} \\ \begin{pmatrix} 1 & 0 \\ \mu_k & 1-\mu_k \end{pmatrix} \end{array}, \qquad \boldsymbol{P}_k^0 := \begin{array}{c} \\ 0 \\ 1 \end{array} \!\!\begin{array}{c} \begin{array}{cc} 0 & 1 \end{array} \\ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \end{array}.$$

Since two-state jobs are indexable (see the previous section), the only task to do is to evaluate the index. We will do it by showing that this problem is LP-indexable. First, it is not a difficult exercise to establish conditions 1. and 2., and then condition 3. holds because (i) if $\nu > 0$, then it is not optimal to serve the job in state 0, and (ii) if $\nu \leq 0$, then serving in both states is optimal. Therefore, we can use that $\nu_{k,n} = \nu_{k,n}^{\mathcal{S}_{N_k - n}}$, which can be obtained in a straightforward manner giving

$$\nu_{k,1} = \frac{c_k \mu_k}{1-\beta}, \qquad \nu_{k,0} = 0.$$

Under the myopic criterion, we have $\nu_{k,1} = c_k \mu_k$ and $\nu_{k,0} = 0$. Under the time-average criterion, taking the limit $\beta \to 1$ gives $\nu_{k,1} = \infty$, and the second-order (tie-breaking) index for state 1 is $c_k \mu_k$, while it remains zero for state 0. For idling option $k = 0$ we can obtain $\nu_{0,0} = 0$.

Thus, under the myopic criterion, the discounted criterion, and the time-average criterion, the Whittle index rule is equivalent to serving the uncompleted job with highest value $c_k \mu_k$, which is known to be optimal both without arrivals [23] and with arbitrary arrival processes [7].

5.2 Job Scheduling with Abandonment

Suppose now that jobs can abandon the system. Each job k abandons with probability $\theta_k \geq 0$ within one period if it is not served and we have to pay abandonment penalty d_k if this happens. The joint goal is to minimize the aggregate expected holding cost and abandonment penalties over an infinite horizon.

Such jobs can be defined as in the classic problem, except for the following modifications

– expected one-period reward, i.e., the negative of the holding cost and abandonment penalties expected to be paid at the next time epoch,

$$R_{k,0}^1 := 0, \qquad R_{k,1}^1 := -c_k \cdot (1-\mu_k) - 0 \cdot \mu_k,$$
$$R_{k,0}^0 := 0, \qquad R_{k,1}^0 := -c_k \cdot (1-\theta_k) - d_k \cdot \theta_k;$$

– one-period state-transition probability matrices

$$\boldsymbol{P}_k^1 := \begin{array}{c} \\ 0 \\ 1 \end{array} \!\!\begin{array}{c} \begin{array}{cc} 0 & 1 \end{array} \\ \begin{pmatrix} 1 & 0 \\ \mu_k & 1-\mu_k \end{pmatrix} \end{array}, \qquad \boldsymbol{P}_k^0 := \begin{array}{c} \\ 0 \\ 1 \end{array} \!\!\begin{array}{c} \begin{array}{cc} 0 & 1 \end{array} \\ \begin{pmatrix} 1 & 0 \\ \theta_k & 1-\theta_k \end{pmatrix} \end{array}.$$

It was announced in [6] that the Whittle index values under $0 \leq \beta \leq 1$ for this problem are

$$\nu_{k,1} = \frac{c_k(\mu_k - \theta_k) + d_k\theta_k(1 - \beta + \beta\mu_k)}{1 - \beta + \beta\theta_k}, \qquad \nu_{k,0} = 0.$$

This index recovers the $c\mu$ index if the job has no abandonment ($\theta_k = 0$).

We remark that optimality of threshold policies as defined in the previous section is only accomplished if the numerator $c_k(\mu_k - \theta_k) + d_k\theta_k(1-\beta+\beta\mu_k) \geq 0$, which assures the monotonicity of the index values $\nu_{k,0} \leq \nu_{k,1}$. Otherwise, the index value of state 1 is negative, and another type of threshold policies ($\{m \in \mathcal{N}_k : m < n\}$) becomes optimal.

The above index value has a strong impact not only on the order in which jobs are to be served, but also on the ultimate decision whether jobs should ever be served. Since idling is always available as an option and its index value is 0, no job with negative index value for state 1 (i.e., negative numerator) will be served, and so such jobs will eventually leave the system by abandoning. Intuitively, this is beneficial for the system because such a job is more likely to leave by abandoning than by being completed ($\mu_k < \theta_k$) and moreover the abandonment penalty is not too high (relatively to its other parameters).

Under the time-average criterion, any job without abandonment receives absolute priority ($\nu_{k,1} = \infty$) over any job with a positive probability of abandonment. This may seem counterintuitive, because one could believe that it is worth to prefer jobs that are likely to abandon, so that the abandonment penalty is avoided. But notice that if a not abandoning job remains uncompleted forever, then it would accrue (infinite) holding costs that would surely exceed the (finite) abandonment penalties of all the other jobs, and the index rule assures that this cannot happen. Moreover, among uncompleted and not abandoning jobs the rule again chooses according to the $c\mu$-rule.

The scheduling problem with abandonment, in which arrivals of jobs of multiple classes are allowed, is considered intractable, and optimal policy is known only under certain extreme conditions in two-class problems [8]. Interestingly, extensive computational experiments reported in [6] for the two-class problem with Poisson arrivals indicate near-optimality of the above index rule and its systematic dominance over the $c\mu$-rule (which ignores abandonments), and over an alternative $c\mu/\theta$-rule obtained from a fluid limit model in [3] and proved to be asymptotically optimal in a multi-server system. The results further show that the index rule is often equivalent or outperforming the policy obtained by optimal ordering of two classes in the system with a single job of each class (which is, however, a policy not enjoying index values separable across classes, as it depends on the other job's service rate, and therefore is not applicable in systems with more than two classes).

5.3 Job Scheduling with Time-Varying Service Rates

An important job scheduling problem arises in the context of wireless data networks, where available service rate varies in time due to fading. We assume that such varying occurs in an i.i.d. manner between consecutive periods and independently for different jobs, which represents the situation in which the steady-state distribution of the service rate of each job (user) is known. We further suppose that the available service rate for each job becomes known at the beginning of every period so that it can be taken into account for the decision. This problem is known as the channel-aware flow-level scheduling of the downlink wireless data network. Thus, several users are waiting to download finite amounts of data available at the base station, which is in charge of deciding at every period (typically of a very short duration) to which user its data should be sent.

One can then expect that the index rule would no longer be given by static ordering of waiting jobs, but it would dynamically adapt to the actual available service rate. It is suggested by some wireless network standards that the service should be given to the job whose ratio of its actual service rate to its average obtained throughput is highest. Thus, allocation is proposed to be done opportunistically, trying to exploit when the user's available service rate is good with respect to its own statistical behavior. One may expect that due to the very short period durations, the users would not perceive any changes in their downloading rate, while the system may improve its performance with respect to a round-robin scheduling. Performance evaluation of such a policy, known as the Proportionally Fair scheduler, has attracted a considerable research attention in the recent years, especially when approximated by its Markovian variant called the Relatively Best (RB) scheduler, in which the service is given to the users whose actual service rate divided by the mean service rate is highest.

As shown in [4], this problem fits the framework of this paper, and can be naturally generalized to account for holding costs. Let $c_k > 0$ be the holding cost per period incurred for user k waiting while the job is not completed. Suppose that job k can obtain service rates associated to channel conditions defined by a non-empty finite set $\mathcal{N}'_k := \{1, 2, \ldots, N_k\}$ so that condition $n \in \mathcal{N}'_k$ happens in a certain period with probability $q_{k,n}$, having $\sum_{n \in \mathcal{N}'_k} q_{k,n} = 1$. Further, under channel condition n, the probability that the service of job k is completed within one period if being transmitted is $\mu_{k,n}$. Without loss of generality we assume that the channel condition labels are ordered so that $0 \le \mu_{k,1} \le \mu_{k,2} \le \cdots \le \mu_{k,N_k} \le 1$. To ensure that eventually all users leave the system we assume that for every user k, $q_{k,n}\mu_{k,n} \neq 0$ for some channel condition $n \in \mathcal{N}'_k$.

Each job/channel/user k is defined independently of other jobs/channels/users with

- state space $\mathcal{N}_k := \{0\} \cup \mathcal{N}'_k$, where state 0 represents a job already completed, and $\mathcal{N}'_k := \{1, \ldots, N_k\}$ is the set of possible service rates for job k provided the job is uncompleted;
- expected one-period work for any $n \in \mathcal{N}_k$,

$$W^1_{k,n} := 1, \qquad\qquad W^0_{k,n} := 0;$$

– expected one-period reward for any $n \in \mathcal{N}'_k$,

$$R^1_{k,0} := 0, \qquad R^1_{k,n} := -c_k \cdot (1-\mu_{k,n}) - 0 \cdot \mu_{k,n},$$
$$R^0_{k,0} := 0, \qquad R^0_{k,n} := -c_k;$$

– one-period state-transition probability matrix, denoting by $\widetilde{\mu}_{k,n} := 1-\mu_{k,n}$,

$$\boldsymbol{P}^1_k := \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ \vdots \\ N_k \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & \dots & N_k \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 \\ \mu_{k,1} & \widetilde{\mu}_{k,1} q_{k,1} & \dots & \widetilde{\mu}_{k,1} q_{k,N_k} \\ \mu_{k,2} & \widetilde{\mu}_{k,2} q_{k,1} & \dots & \widetilde{\mu}_{k,2} q_{k,N_k} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{k,N_k} & \widetilde{\mu}_{k,N_k} q_{k,1} & \dots & \widetilde{\mu}_{k,N_k} q_{k,N_k} \end{pmatrix} \end{array},$$

$$\boldsymbol{P}^0_k := \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ \vdots \\ N_k \end{array} \begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & \dots & N_k \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & q_{k,1} & q_{k,2} & \dots & q_{k,N_k} \\ 0 & q_{k,1} & q_{k,2} & \dots & q_{k,N_k} \\ 0 & q_{k,1} & q_{k,2} & \ddots & q_{k,N_k} \\ 0 & q_{k,1} & q_{k,2} & \dots & q_{k,N_k} \end{pmatrix} \end{array}.$$

We note that PCL-indexability does not hold for these jobs unless certain restrictive assumptions are made. However, [4] showed that these jobs are LP-indexable and derived the Whittle index under $0 \le \beta \le 1$ for channel condition $n \in \mathcal{N}'_k$ of user k,

$$\nu_{k,n} = \frac{c_k \mu_{k,n}}{(1-\beta) + \beta \sum\limits_{m>n} q_{k,m}(\mu_{k,m} - \mu_{k,n})}, \qquad \nu_{k,0} = 0.$$

Special attention was given to the rule under the time-average criterion, when the problem under $c_k = 1$ corresponds to both the minimization of the mean number of users in the system and the minimization of the mean waiting time. In this case the rule was named the Potential Improvement (PI) rule, as the index relates the available service to the expected potential improvement of the service rate,

$$\nu^{\mathrm{PI}}_{k,n} = \frac{c_k \mu_{k,n}}{\sum\limits_{m>n} q_{k,m}(\mu_{k,m} - \mu_{k,n})} \text{ for } n \neq N_k, \qquad \nu^{\mathrm{PI}}_{k,N_k} = \infty,$$

and the tie-breaking quantity is 0 for $n \neq N_k$, and $c_k \mu_{k,N_k}$. Thus, PI rule results in giving absolute priority to users whose actual service rate is the best possible over all the users with a non-best service rate. Again, among the users with their best service rate we select according to the $c\mu$-rule. When no channel achieves its best quality, PI allocates service to the user with the highest ratio of the actual service rate with respect to the expected potential improvement of the service rate. The absolute priority to the users with their best service rates is the most distinguishing property of PI from the RB rule.

Moreover, it was shown that a sufficient condition for a policy to achieve the maximum stability region in systems with arrivals is that any user with its best service rate will be preferred over any user with a non-best service rate (see, for example [2, 5]), which is satisfied by the PI rule. It was also noted that no other known rule in its full generality achieves the maximum stability region and so each of them fails to be stable in some stabilizable systems. Finally, simulations of a two-class system with Bernoulli arrivals reported in [4] indicate that PI is superior or comparable to the best of the existing scheduling rules under several scenarios, and suggest even a dominance in stochastic sense.

5.4 Job Scheduling with Time-Varying Service and Abandonment Rates

It is not difficult to realize that it is possible to model in our framework also the system with both time-varying service rates and time-varying abandonment that happens with probability $\theta_{k,n}$ in state $n \in \mathcal{N}'_k$. Then, the Whittle index in state $n \in \mathcal{N}'_k$ is

$$\nu_{k,n} = \frac{c_k(\mu_{k,n} - \theta_{k,n}) + d_k \mu_{k,n} \beta \sum_{m \le n} q_{k,m}\theta_{k,m} + d_k \theta_{k,n} \left(1 - \beta + \beta \sum_{m>n} q_{k,m}\mu_{k,m}\right)}{1 - \beta + \beta\theta_{k,n} + \beta \sum_{m>n} q_{k,m}(\mu_{k,m} - \mu_{k,n}) + \beta \sum_{m \le n} q_{k,m}(\theta_{k,m} - \theta_{k,n})}.$$

This rule can be of a great practical interest in wireless systems since currently all the schedulers ignore possible user abandonment, which may lead to wasted resources by allocating server to a user who leaves before its downloading is completed. Further, fully heterogeneous systems, say, including some job classes that are time-varying (but not abandoning) and other that do abandon (but are not time-varying) can be addressed using the above index rule. However, additional assumptions are needed for establishing indexability of such jobs, which is a part of the author's work in progress to be reported in future work.

5.5 Scheduling of Perishable Jobs

We finally remark that the framework of this paper covers also the case of jobs perishable at a fixed deadline time epoch, as studied, for instance, in [13, Chapter 5] in the problem of optimal dynamic promotion of perishable products. That work established PCL-indexability and derived the Whittle index for several variants of perishable jobs.

6 Open Problems

The palette of job scheduling problems arising in applications is extremely rich and in this section we sample only some of them, strongly influenced by the author's research interests, that we believe could be of interest to analyze and design the Whittle index. First, an important extension is to consider a non-Markovian service rate, which can be modeled in our MDP setting by defining the job state as the amount of data required (if the job length is known in advance) or as the amount of service obtained (if the job length is random) and properly defining the transition probability matrices. The result is known for the classic (not changing state when not served) case in continuous state space to be achieved by the Gittins index [10, 1], but the extension to abandonment or time-varying service rate is still open.

Second, important for the applications but often scarce in theoretical work is the notion of finite buffers. For instance, suppose in the wireless system that each user's data to be transmitted is received by the base station at a certain rate and buffered, and whenever the buffer is full, the arriving data is dropped and the user is lost or has to start the transmission anew. One can expect that this setting requires an opportunistic solution with some level of load balancing. Such extensions, however, lead to multidimensional state spaces and are therefore more analytically demanding.

Third, different variants of the wireless system studied above are used in the real-world systems. In some of them, the actual service rate is only known at the end of the period, or is observed after a random delay. Moreover, the service rate may not evolve in an i.i.d manner, but with a certain dependence on the history or on the user's mobility path.

Finally, several theoretical issues remain open, such as validity of the first- and second-order index under the time-average criterion and rigorous justification for the use of the Whittle index rule and its performance evaluation in systems with arrivals (see [13, Chapter 6]) or with multiple servers. Regarding the non-equivalence between optimality of threshold policies and indexability, it would be interesting to find an example of a job solvable by threshold policies but not indexable. To the best of author's knowledge, no such job is known.

References

1. Aalto, S., Ayesta, U. and Righter, R.: On the Gittins index in the $M/G/1$ queue. Queueing Systems **63(1-4)** (2009) 437–458
2. Aalto, S. and Lassila, P.: Flow-level stability and performance of channel-aware priority-based schedulers. In Proceeding of the 6th EURO-NF Conference on Next Generation Internet, 2010
3. Atar, R., Giat, C. and Shimkin, N.: The $c\mu/\theta$ rule for many-server queues with abandonment. Operations Research, to appear (2010)
4. Ayesta, U., Erausquin, M. and Jacko, P.: A modeling framework for optimizing the flow-level scheduling with time-varying channels. Performance Evaluation, to appear (2010)

5. Ayesta, U., Erausquin, M., Jonckheere, M. and Verloop, I. M.: Opportunistic scheduling in wireless systems: Stabilility and asymptotic optimality. July 30 2010
6. Ayesta, U., Jacko, P. and Novak, V.: A nearly-optimal index rule for scheduling of users with abandonment. July 30 2010
7. Buyukkoc, C., Varaiya, P. and Walrand, J.: The $c\mu$ rule revisited. Advances in Applied Probability, **17(1)** (1985) 237–238
8. Down, D. G., Koole, G. and Lewis, M. E.: Dynamic control of a single server system with abandonments. January 23 2010
9. Gittins, J. C.: Bandit processes and dynamic allocation indices. Journal of the Royal Statistical Society, Series B, **41(2)** (1979) 148–177
10. Gittins, J. C.: Multi-Armed Bandit Allocation Indices. J. Wiley & Sons, New York, 1989
11. Gittins, J. C. and Jones, D. M.: A dynamic allocation index for the sequential design of experiments. In J. Gani, editor, Progress in Statistics, North-Holland, Amsterdam, (1974) 241–266
12. Jacko, P.: Adaptive greedy rules for dynamic and stochastic resource capacity allocation problems. Medium for Econometric Applications, **17(4)** (2009) 10–16
13. Jacko, P.: Dynamic Priority Allocation in Restless Bandit Models. Lambert Academic Publishing, 2010
14. Katehakis, M. N. and Rothblum, U. G.: Finite state multi-armed bandit problems: Sensitive-discount, average-reward and average-overtaking optimality. Annals of Applied Probability, **6(3)** (1996) 1024–1034
15. Kelly, F.: Multi-armed bandits with discount factor near one: the Bernoulli case. Annals of Statistics, **9(5)** (1981) 987–1001
16. Niño-Mora, J.: Restless bandits, partial conservation laws and indexability. Advances in Applied Probability, **33(1)** (2001) 76–98
17. Niño-Mora, J.: Dynamic allocation indices for restless projects and queueing admission control: A polyhedral approach. Mathematical Programming, Series A, **93(3)** (2002) 361–413
18. Niño-Mora, J.: Marginal productivity index policies for scheduling a multiclass delay-/loss-sensitive queue. Queueing Systems, **54** (2006) 281–312
19. Niño-Mora, J.: Characterization and computation of restless bandit marginal productivity indices. In Proceedings of the 2nd International Conference on Performance Evaluation Methodologies and Tools. ICST, Brussels, Belgium, October 2007
20. Niño-Mora, J.: Dynamic priority allocation via restless bandit marginal productivity indices. TOP, **15(2)** (2007) 161–198
21. Papadimitriou, C. H. and Tsitsiklis, J. N.: The complexity of optimal queueing network. Mathematics of Operations Research, **24(2)** (1999) 293–305
22. Puterman, M. L.: Markov Decision Processes: Discrete Stochastic Dynamic Programming. John Wiley & Sons, Inc., Hoboken, New Jersey, 2005
23. Smith, W. E.: Various optimizers for single-stage production. Naval Research Logistics Quarterly, **3(1-2)** (1956) 59–66
24. Weber, R. and Weiss, G.: On an index policy for restless bandits. Journal of Applied Probability, **27(3)** (1990) 637–648
25. Whittle, P.: Arm-acquiring bandits. Annals of Probability, **9(2)** (1981) 284–292
26. Whittle, P.: Restless bandits: Activity allocation in a changing world. A Celebration of Applied Probability, J. Gani (Ed.), Journal of Applied Probability, **25A** (1988) 287–298