Value of Information in Optimal Flow-Level Scheduling of Users with Markovian Time-Varying Channels *

Peter Jacko
BCAM, Spain

July 19, 2011

Abstract

In this paper we design, characterize in closed-form, and evaluate a new index rule for Markovian time-varying channels, which gives rise to a simple opportunistic scheduling rule for flow-level scheduling in wireless downlink systems. For user channels we employ the Gilbert-Elliot model with a flow-level interpretation: the channel condition follows a general two-state Markov chain with distinct probabilities of finishing the flow transmission. The index value of the bad channel condition takes into account both the one-period and the steady-state potential improvement of the service completion probability, while the good channel condition gets an absolute priority with the $c\mu$-index (well-known to be throughput-optimal) as the tie-breaking rule. Our computational study confirms near-optimality of the proposed rule in most of the instances, and suggests that information about the channels steady state is often enough to achieve near-optimality.

Keywords: Markov Decision Process, dynamic programming, opportunistic scheduling, $c\mu$-rule, restless bandits, Whittle index

1 Introduction

In this paper we study the end-user experience and the overall performance of systems providing service via time-varying channels. Such systems are now ubiquitous in communications due to the recent technological advances in wireless communication, including local area networks, cellular data networks, cognitive radio systems, and satellite communication. The development of future technological standards builds on the same paradigm, since time-varying channels can be exploited by *opportunistic scheduling* to enhance the system capacity, as shown in Knopp and Humblet (1995). These systems are typically locally centralized, i.e., they involve a network of central elements (e.g., routers, base stations, satellites), which continuously collect information and adaptively decide the bandwidth allocation to a subset of users demanding service.

The main implementation differences in the above-mentioned systems and their variants imply different amounts of information available for the scheduler. The possibility of exploitation of available information is essentially transformed into different overall performance and end-user experience from the use of the service. Since information gathering requires higher implementation costs and may have other non-desirable effects (such as delays), it is important to understand how much (in terms of performance) can be gained in such systems if more information is available. In other words, we need to study how much is lost if some information is unobservable with respect to the best policy implementable in fully-observable systems, which is the objective of this paper.

*The author would like to thank to Samuli Aalto, Urtzi Ayesta, Thomas Bonald, Sem Borst, Martin Erausquin, Matthieu Jonckheere, and Ina Maria (Maaike) Verloop for fruitful and encouraging discussions on this topic. Research partially supported by grant MTM2010-17405 (Ministerio de Ciencia e Innovación, Spain) and grant PI2010-2 (Department of Education and Research, Basque Government). NOTICE: This is the author's version of the work that was published as a journal paper with the following reference: Jacko, P. (2011): Value of Information in Optimal Flow-Level Scheduling of Users with Markovian Time-Varying Channels. Performance Evaluation ??, pp. ???-???.

The above question can be formulated mathematically as the problem of optimal user scheduling in a system with fixed bandwidth capacity and with time-varying service rate. The focus is on users with finite-size jobs that randomly arrive and leave once the service of their job is completed. Scheduling of such users is known as *flow-level* scheduling, and is understood as extremely difficult to solve for optimality due to the fixed bandwidth capacity constraint (Ayesta et al., 2010). Research has therefore mainly focused on characterizing the stability region and identifying maximally stable policies (Borst, 2005; Bonald et al., 2009; Aalto and Lassila, 2010; Ayesta et al., 2011) and on partial characterization of optimal policies, mainly under the restrictive *time-scale separation* principle or for single-user-class systems (Sadiq and de Veciana, 2010; Aalto et al., 2011). Nevertheless, the scheduling optimization problem becomes analytically tractable if the fixed-capacity constraint is relaxed so that the bandwidth allocation must be satisfied only on average (see, e.g., Whittle, 1988; Knopp and Humblet, 1995; Jacko, 2010b).

We will focus on a downlink system, although both downlink and uplink transmission can be covered by the same model as long as scheduling is centralized (typically at the base station). That is, at the beginning of each slot, the scheduler decides which users are allowed to send/receive data. If a scheduled user is waiting for data available at the base station (downlink), then the scheduler allows the base station to transmit data via her channel to the scheduled user. If a scheduled user wants to send data (uplink), then the scheduler informs the user at the beginning of the slot that the channel is available. If synchronization of users is possible, then the uplink case can be implemented by giving the users the possibility to *sense* the availability of their channel during a subpart of the slot and the scheduler to send a noise during the sensing interval so that the user perceives that the channel is not available. This modeling aspect does not differ from *packet-level* models (see, e.g., Liu et al., 2003, for a survey of packet-level models). By packet-level models we understand scheduling of a fixed population of permanent users wanting to send to or receive from a base station randomly arriving packets backlogged in user-dedicated queues.

Although several modeling aspects of packet-level models and flow-level models are analogous, the consideration of finite-length jobs makes the good scheduling policies completely different. The MaxWeight scheduler proposed and proved maximally stable for a packet-level model with ON/OFF channels in Tassiulas and Ephremides (1993) may not be maximally stable in flow-level models (van de Ven et al., 2009). Another well-known scheduler, the *Proportional Fair* (PF) scheduler of the Qualcomm High Data Rate (HDR) system 1xEV-DO (Bender et al., 2000), was shown to maximize the aggregate logarithmic utility of obtained throughput in a packet-level model (Kushner and Whiting, 2004). The Markovian variant of PF, so-called Relatively Best scheduler, is roughly equivalent to PF and maximally stable under certain user symmetry conditions in flow-level systems (Borst, 2005), but may not be maximally stable in general (Aalto and Lassila, 2010; Ayesta et al., 2010, 2011).

The above considerations have raised interest in designing and studying scheduling algorithms which are maximally stable at flow-level (Bonald, 2004b; Sadiq and de Veciana, 2009; Aalto and Lassila, 2010; Ayesta et al., 2010, 2011; Liu et al., 2011). Sadiq and de Veciana (2009) proposed a modification of the MaxWeight scheduler, which at every slot serves the user with highest product of the time since arrival and the actual transmission rate, with ties broken in favor of older users. The model in Liu et al. (2011) tries to balance permanent users (with randomly arriving packets to the queue) and flows (that have finite-size jobs to transmit). If there are only flows in the system, then priority is given to a user that is in its best channel condition or whose job can be completed in one slot (with oldest-first or randomized tie-breaking among such users), otherwise an arbitrary user is scheduled. These schedulers therefore require the knowledge about the time since arrival or the remaining job size of each user.

More related to our paper are schedulers that rely on stochastic properties of user channels and job sizes, which are more important from implementation point of view. For systems, in which the possible channel conditions are discrete, existing patents (Chaponniere et al., 2002; Bonald, 2004a; Ayesta and Jacko, 2010) propose to keep track of historical observations of the channel conditions of every user, which (if the history window is large enough), provide an accurate estimate of the steady-state distribution of channel conditions. Possible gains from the *Score-Based* (SB) scheduler were described in Bonald (2004b); SB scheduler serves at each slot the user for which the ranking of the current channel condition over possible channel conditions is best. Aalto and Lassila (2010) proposed the so-called *Proportionally Best* (PB) scheduler, which at each slot serves a user whose ratio of the current channel condition with respect to her best possible condition is highest. They further concluded that the shortest-remaining-processing-time (SRPT) tie-breaking instead of randomized tie-breaking does not necessarily improve performance.

The *Potential Improvement* (PI) scheduler for i.i.d. channel condition evolution was designed in Ayesta et al. (2010) based on optimally solving a relaxation of the problem. PI serves at each slot the user with highest ratio of the current transmission rate with respect to the expected potential improvement in the service rate; if more than one user is in her best channel condition, then the user with shortest expected job size is served (among such users).

Using the fluid-limit approach, Ayesta et al. (2011) proved for systems with i.i.d. channel condition evolution that all the schedulers that serve a user in her best channel condition as long there is such a user are maximally stable. Moreover, the authors proved that if a scheduler in addition serves the user with shortest expected job size as the tie-breaking rule, then it is fluid-optimal. Note that SB, PB, and PI satisfy the condition for maximal stability, but only PI is fluid-optimal. Of course, both SB and PB could be easily modified by implementing the expected-job-size tie-breaking in order to become fluid-optimal.

It is likely that i.i.d. channel condition evolution is usually unrealistic, but on the other extreme, existing patents rely exclusively on an estimate of the steady-state distribution of the channel condition in a certain time window, and design schedulers that take decisions based on such an information. For instance, Ayesta and Jacko (2010) proposed the PI scheduler to be employed in systems with Markovian channels by defining the probabilities $q_{k,n}$ from the estimate of the steady-state distribution of the channel condition. However, the performance of PI scheduler in Markovian systems has not been studied yet and it is not at all clear whether the results obtained for i.i.d. evolution remain valid if channel condition evolution is Markovian, for which only an estimate of the steady-state distribution is known instead of the real transition probabilities.

In this paper we approach the downlink optimal scheduling problem with Markovian channel condition evolution, which is fundamentally more complex than the case of i.i.d. channel condition evolution. For transparency, we focus on channels with only two possible conditions, so each channel is modeled by a Gilbert-Elliot model (Gilbert, 1960). The problem is described in Section 2 in more detail.

The main contributions of the paper are as follows:

(i) a new flow-level MDP model for scheduling of users transmitting downlink/uplink via channels with Markovian dynamics in Section 3,

(ii) an optimal index policy to a relaxation of the above model in Section 4; the policy shows that while for infinite-size jobs there is no advantage of knowing the state transition probabilities over the steady-state probabilities, such information becomes extremely profitable as the job size becomes smaller.

(iii) a new simple and intuitive opportunistic scheduling rule based on the above index policy in Section 5, which can be interpreted as a generalized Potential Improvement rule and recovers optimal or asymptotically optimal policies in a list of special cases,

(iv) a numerical study in Section 6 comparing a variety of opportunistic scheduling rules with an optimal policy, providing insights for resolving the trade-off between opportunistic scheduling and maximum throughput policies in systems with time-varying service rate and suggesting that steady-state information is often enough to achieve near-optimality.

2 Problem Description

We consider the job scheduling problem in a time-slotted system such as the CDMA 1xEV-DO (Bender et al., 2000), in which the available service rate of each user fluctuates. Slots are denoted by $t \in \mathcal{T} := \{0, 1, 2, \dots\}$ and slot duration is denoted by ε (in seconds, typically of order 10^{-3}). Jobs $k = 1, 2, \dots$ appear randomly from users that are within the transmission distance from a base station, which can serve M users at every slot in parallel. Let $c_k > 0$ be the holding cost per slot incurred for user waiting while the transmission of job k is not completed. The channel for transmission of job k (or shortly channel k) can take two quality conditions from a set $\mathcal{N}'_k := \{\mathrm{B}, \mathrm{G}\}$. The transmission quality condition of each channel evolves randomly and independently of the other channels and of the decisions of the base station. The channel conditions evolution for job k is Markovian with one-slot transition probabilities $q_{k,n,m}$ to move from condition n to condition m, satisfying $q_{k,n,\mathrm{B}} + q_{k,n,\mathrm{G}} = 1$ for all $n \in \mathcal{N}'_k$. The

channel-k condition transitions will be denoted by

$$\boldsymbol{Q}_k = \begin{array}{c} \\ \mathrm{B} \\ \mathrm{G} \end{array} \begin{array}{c} \begin{array}{cc} \mathrm{B} & \mathrm{G} \end{array} \\ \begin{pmatrix} q_{k,\mathrm{B},\mathrm{B}} & q_{k,\mathrm{B},\mathrm{G}} \\ q_{k,\mathrm{G},\mathrm{B}} & q_{k,\mathrm{G},\mathrm{G}} \end{pmatrix} \end{array}$$

If channel k is in condition n, then job k is served with transmission/service rate $s_{k,n}$ (in bits per second), which is assumed to be a multiple of $1/\varepsilon$. Without loss of generality we assume that the channel condition labels are ordered so that $0 \leq s_{k,\mathrm{B}} \leq s_{k,\mathrm{G}}$. So, "B" can be interpreted as bad channel condition, while "G" can be interpreted as good channel condition.

If the base station is allocated to a user whose job has already been completed, then no transmission occurs. The base station is assumed to be preemptive (i.e., the service of a job can be interrupted at the beginning of any slot even if not completed). Thus, the base station decides at the beginning of every period to which user it should be allocated during that slot.

The goal is to minimize the expected aggregate holding cost over an infinite horizon under the discounted criterion (with a discount factor $0 \leq \beta < 1$) and under the time-average criterion. The problem with $c_k = 1$ for all k corresponds to minimization of the mean waiting time and minimization of the mean number of jobs in the system.

Ayesta et al. (2010) showed that the probability of departure of users with jobs with geometric size can be computed in an approximate way when the expected job size (in bits) of a user is much larger than the amount of bits that can be served in one slot. Next we give an exact probability of departure that leads to the same approximation.

Lemma 1. *Let the job size of user k be a geometrically distributed random variable denoted by B_k (in bits), and let $\mathbb{E}[B_k]$ denote its expectation. Then the departure probability of a job k in channel condition n is $\mu_{k,n} = \min\{1, 1-(1-1/\mathbb{E}[B_k])^{\varepsilon s_{k,n}}\}$, which can be approximated if $\varepsilon s_{k,n}/\mathbb{E}[B_k] \approx 0$ by $\mu_{k,n} \approx \dfrac{\varepsilon s_{k,n}}{\mathbb{E}[B_k]}$. We remark that the approximation is exact for $\varepsilon s_{k,n} = 1$.*

3 MDP Formulation

In this section we present an MDP formulation of the discrete-time job sequencing problem (without arrivals), in which we allow for time-varying departure probability as described in the previous section. Ignoring arrivals makes the problem analytically tractable and leads to designing a well-founded scheduling rule which we then propose to be used in systems with arrivals.

Consider K jobs labeled by $k \in \mathcal{K}$ waiting for service at a base station that can serve M jobs at a time by transmitting a data flow through a dedicated channel to the corresponding user. The setting fits the *multi-armed restless bandit problem* Whittle (1988); Niño-Mora (2001), which can be adapted to job scheduling as described in Jacko (2010b).

Consider the time slotted into epochs $t \in \mathcal{T} := \{0, 1, 2, \dots\}$ at which decisions can be made.

3.1 Jobs, Channels, and Users

Every user k can be allocated either zero capacity of the base station or be one of the M users served. We denote by $\mathcal{A} := \{0, 1\}$ the *action space*, i.e., the set of allowable levels of capacity allocation. Here, action 0 means allocating zero capacity (i.e., "not serving"), and action 1 means allocating one capacity (i.e., "serving"). This action space is the same for every user k. Further, under channel condition n, the probability that the service of job k is completed within one period if being served is $\mu_{k,n}$. According to Lemma 1 we have $0 \leq \mu_{k,\mathrm{B}} \leq \mu_{k,\mathrm{G}} \leq 1$.

Each job-channel-user triple k is defined independently of other job-channel-user triples 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*, where state 0 represents a job already completed, and $\mathcal{N}_k' := \{\mathrm{B}, \mathrm{G}\}$ is the set of possible quality conditions of channel k provided the job is uncompleted;

- $\boldsymbol{W}_k^a := \left(W_{k,n}^a\right)_{n\in\mathcal{N}_k}$, where $W_{k,n}^a$ is the (expected) one-period capacity consumption, or *work* required by user k at state n if action a is decided at the beginning of a period; in particular, for any $n \in \mathcal{N}_k$, $W_{k,n}^1 := 1, W_{k,n}^0 := 0$;

- $\boldsymbol{R}_k^a := \left(R_{k,n}^a\right)_{n\in\mathcal{N}_k}$, where $R_{k,n}^a$ is the expected one-period *reward* earned by user k at state n if action a is decided at the beginning of a period; in particular, for any $n \in \mathcal{N}_k'$,

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

- $\boldsymbol{P}_k^a := \left(p_{k,n,m}^a\right)_{n,m\in\mathcal{N}_k}$ is the user-k stationary one-period *state-transition probability matrix* if action a is decided at the beginning of a period, i.e., $p_{k,n,m}^a$ is the probability of moving to state m from state n under action a; in particular,

$$\boldsymbol{P}_k^1 := \begin{array}{c} \\ 0 \\ \mathrm{B} \\ \mathrm{G} \end{array}\!\!\begin{array}{c} \begin{array}{ccc} 0 & \mathrm{B} & \mathrm{G} \end{array} \\ \begin{pmatrix} 1 & 0 & 0 \\ \mu_{k,\mathrm{B}} & (1-\mu_{k,\mathrm{B}})q_{k,\mathrm{B},\mathrm{B}} & (1-\mu_{k,\mathrm{B}})q_{k,\mathrm{B},\mathrm{G}} \\ \mu_{k,\mathrm{G}} & (1-\mu_{k,\mathrm{G}})q_{k,\mathrm{G},\mathrm{B}} & (1-\mu_{k,\mathrm{G}})q_{k,\mathrm{G},\mathrm{G}} \end{pmatrix} \end{array}, \quad \boldsymbol{P}_k^0 := \begin{array}{c} \\ 0 \\ \mathrm{B} \\ \mathrm{G} \end{array}\!\!\begin{array}{c} \begin{array}{ccc} 0 & \mathrm{B} & \mathrm{G} \end{array} \\ \begin{pmatrix} 1 & 0 & 0 \\ 0 & q_{k,\mathrm{B},\mathrm{B}} & q_{k,\mathrm{B},\mathrm{G}} \\ 0 & q_{k,\mathrm{G},\mathrm{B}} & q_{k,\mathrm{G},\mathrm{G}} \end{pmatrix} \end{array}.$$

The dynamics of user k is thus 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 user k consumes the allocated capacity, earns the reward, and evolves its state for the time epoch $t+1$.

3.2 A 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 $\Pi_{X,a}$ be the set of all the policies that for each time epoch t decide (possibly *randomized*) action $a(t)$ based only on the state-process history $X(0), X(1), \ldots, X(t)$ and on the action-process history $a(0), a(1), \ldots, a(t-1)$ (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 state-process history $X(0), X(1), \ldots, X(\tau)$ 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 $\pi \in \Pi_{X,a}$, 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]

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

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* (scaled by constant $1-\beta$). The remaining case when $\beta = 0$ 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 $\mathbb{B}_\tau^\pi\left[Q_{X(\cdot)}^{a(\cdot)}\right]$.

3.3 Optimization Problem

Now we can define the optimization problem. Let $\Pi_{\boldsymbol{X},\boldsymbol{a}}$ be the space of randomized and non-anticipative policies depending on the joint state-process $\boldsymbol{X}(\cdot) := (X_k(\cdot))_{k\in\mathcal{K}}$ and deciding the joint action-process

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

$\boldsymbol{a}(\cdot) := (a_k(\cdot))_{k\in\mathcal{K}}$, i.e., $\Pi_{\boldsymbol{X},\boldsymbol{a}}$ 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 time epoch 0 subject to the family of *sample path* allocation constraints, i.e.,

$$\max_{\boldsymbol{\pi}\in\Pi_{\boldsymbol{X},\boldsymbol{a}}} \mathbb{B}_0^{\boldsymbol{\pi}}\left[\sum_{k\in\mathcal{K}} R_{k,X_k(\cdot)}^{a_k(\cdot)}\right] \tag{P}$$
$$\text{subject to } \mathbb{E}_t^{\boldsymbol{\pi}}\left[\sum_{k\in\mathcal{K}} a_k(t)\right] = M, \text{ for all } t\in\mathcal{T}$$

Note that the constraint could equivalently be expressed by restricting $\Pi_{\boldsymbol{X},\boldsymbol{a}}$ to policies satisfying $\sum_{k\in\mathcal{K}} a_k(t) = 1$ for any possible joint state-process history $\boldsymbol{X}(0), \boldsymbol{X}(1), \ldots, \boldsymbol{X}(t)$, for all $t\in\mathcal{T}$.

4 Solution

Problem (P) can be relaxed by requiring to serve M jobs per slot only *on β-average* as proposed in Whittle (1988), which is further approached by incorporating a Lagrangian multiplier ν and can be decomposed into a parameterized optimization problem below (for more details see Jacko (2009)). Notice that any joint policy $\boldsymbol{\pi}\in\Pi_{\boldsymbol{X},\boldsymbol{a}}$ defines a set of single-user policies $\widetilde{\pi}_k$ for all $k\in\mathcal{K}$, where $\widetilde{\pi}_k$ is a randomized and non-anticipative policy depending on the *joint* state-process $\boldsymbol{X}(\cdot)$ and deciding the *user-k* action-process $a_k(\cdot)$. We will write $\widetilde{\pi}_k\in\Pi_{\boldsymbol{X},a_k}$. We will therefore study the user-k subproblem

$$\max_{\widetilde{\pi}_k\in\Pi_{\boldsymbol{X},a_k}} \mathbb{B}_0^{\widetilde{\pi}_k}\left[R_{k,X_k(\cdot)}^{a_k(\cdot)} - \nu W_{k,X_k(\cdot)}^{a_k(\cdot)}\right]. \tag{2}$$

The main idea of our approach is to identify a set of optimal policies $\widetilde{\pi}_k^*$ for (2) for each $k\in\mathcal{K}$, and using them to construct a joint policy $\boldsymbol{\pi}$, feasible though not necessarily optimal for problem (P).

4.1 Optimal Solution to Single-User Subproblem under the Gilbert-Elliot Channel Model

In certain cases, problem (2) can be optimally solved by assigning a set of index values $\nu_{k,n}$ to each state $n\in\mathcal{N}_k$ (Niño-Mora, 2007b; Jacko, 2010a). If this is the case, the problem is so-called *indexable*. In the following we prove that jobs with Gilbert-Elliot channel are indexable and we characterize the index values.

For the Gilbert-Elliot model, assuming $q_{k,B,G} > 0$ so that the steady-state distribution exists and is positive for condition G, let us define the following weighted harmonic mean of the one-period channel condition transition probability and of the steady-state channel condition probability,

$$q_{k,B,G}^* := \frac{1}{\frac{1-\beta(1-\mu_{k,G})}{q_{k,B,G}} + \frac{\beta(1-\mu_{k,G})}{q_{k,G}^{\mathrm{SS}}}}, \tag{3}$$

where the steady-state probability of matrix $\boldsymbol{Q}_k$ for condition G can be easily shown to be

$$q_{k,G}^{\mathrm{SS}} = \frac{q_{k,B,G}}{1+q_{k,B,G}-q_{k,G,G}}. \tag{4}$$

Let us denote the index values for user k by

$$\nu_{k,G}^* := \frac{c_k\mu_{k,G}}{(1-\beta)}, \qquad \nu_{k,B}^* := \frac{c_k\mu_{k,B}}{(1-\beta)+\beta q_{k,B,G}^*(\mu_{k,G}-\mu_{k,B})}, \qquad \nu_{k,0}^* := 0. \tag{5}$$

The following are the main theoretical results of this paper.

Proposition 1 (Optimality of Threshold Policies)**.** *For every real-valued ν there exists $n \in \mathcal{N}_k \cup \{-1\}$ such that threshold policy serving in states $\mathcal{S}_{N-n} := \{m \in \mathcal{N}_k : m > n\}$ and not serving otherwise is optimal for problem (2).*

Proposition 2 (Indexability)**.** *The following holds for problem (2) of user k:*

(i) if $\nu \leq \nu^*_{k,n}$, then it is optimal to serve under user's k channel condition $n \in \mathcal{N}'_k$;

(ii) if $\nu \geq \nu^*_{k,n}$, then it is optimal not to serve under user's k channel condition $n \in \mathcal{N}'_k$;

(iii) if $\nu \leq \nu^*_{k,0}$, then it is optimal to serve when the job k is already completed (i.e., when $n = 0$);

(iv) if $\nu \geq \nu^*_{k,0}$, then it is optimal not to serve when the job k is already completed (i.e., when $n = 0$);

4.2 Optimal Solution to Lagrangian Relaxation

The vector of policies $\boldsymbol{\pi}^* := (\widetilde{\pi}^*_k)_{k \in \mathcal{K}}$ identified in Proposition 2 is formed by mutually independent single-user optimal policies, therefore this vector is an optimal policy to the Lagrangian relaxation of the original problem.

5 New Opportunistic Scheduling Rule

Since the original scheduling problem requires to allocate the base station to exactly M jobs, at every slot t we propose to allocate the base station to the M jobs with the highest actual index value $\nu^*_{k,X_k(t)}$. In the following we discuss this rule under the time-average criterion, in which the index value simplify and the rule can be interpreted as a generalized Potential Improvement rule.

5.1 Generalized Potential Improvement Rule

Under the time-average criterion ($\beta = 1$), the index values in (5) simplify to the following values, which we term the *Generalized Potential Improvement* (PI*) index:

$$\nu^*_{k,G} := \infty, \qquad \nu^*_{k,B} := \frac{c_k \mu_{k,B}}{q^*_{k,B,G}(\mu_{k,G} - \mu_{k,B})}, \qquad \nu^*_{k,0} := 0,$$

Due to the infinite index value of channels in good condition, in addition we propose the following tie-breaking rule: If more than M channels are in good condition, then allocate the base station to the M jobs with the highest actual second-order index value

$$\lim_{\beta \to 1} (1-\beta)\nu^*_{k,X_k(t)} = \begin{cases} c_k \mu_{k,G}, & \text{if } X_k(t) = \mathrm{G}, \\ 0, & \text{otherwise.} \end{cases}$$

Such a tie-breaking is based on the second term of the Laurent expansion of $\nu^*_{k,n}$ at $\beta = 1$, so it induces the same policy as the discounted rule for β close enough to 1. This tie-breaking may itself have ties; these are resolved arbitrarily.

The scheduling rule using the above index values and tie-breaking quantities will be shortly called the *PI*-rule*, as it is a generalization of the PI-rule proposed in Ayesta et al. (2010) for i.i.d. time-varying channels. Of course, under i.i.d. time-varying channels (i.e., when $q_{k,G,G} = q_{k,B,G}$) the two rules are identical since $q^*_{k,B,G} = q_{k,B,G}$ (which in fact holds for any β).

The PI*-rule can be rewritten algorithmically in order to prescribe which user to serve at every slot as follows:

(i) pick M users with highest value $c_k \mu_{k,G}$ from among the users in condition G;

(ii) if $M' < M$ users are in condition G, then pick (in addition to those in condition G) $M - M'$ users in condition B with highest value $\nu^*_{k,B}$

Thus, the PI*-rule results in giving absolute priority to users whose actual channel condition is good (condition G), i.e., the base station can serve a user with bad channel quality (condition B) only if there are not enough users with good channel quality. Further, if there are several users with absolute priority, PI* prescribes to serve users according to the classic $c\mu$-rule, where μ is the instantaneous job completion probability (under the good channel condition), which is known to be throughput-optimal if the users are always in the good state.

On the other hand, if there not enough users with channels in good condition, then PI* allocates the base station also to the users with the highest ratio of the actual departure probability (multiplied by the holding cost) with respect to a factor that can be interpreted as the potential improvement of the departure probability, where the improvement is measured by a weighted harmonic mean of the one-period probability and the steady-state probability of moving to condition G.

5.2 Performance of PI* Rule in Systems with Arrivals

Although the new opportunistic rule was derived in a model without arrivals, in the following we give a list of conditions under which the PI* rule is optimal or asymptotically optimal in systems with arrivals. Some of the results apply to multi-class system, in which the jobs with the same characteristics (except for the actual channel condition) are grouped into K classes.

Of special importance will be the $c\mu$-rule, which has been proved optimal in single-server scheduling of jobs with geometric job sizes in case of no channel variation Buyukkoc et al. (1985):

$c\mu$-rule: *Serve the non-empty class with the highest value $c_k\mu_k$.*

As a consequence, we have the following two results.

Proposition 3. *If $q_{k,B,G} = q_{k,G,B} = 0$ for all k (so that no time-variation is present) and $M = 1$, then the PI* rule is optimal under arbitrary arrivals. Note that it is equivalent to the $c\mu$-rule, in which μ_k is the job completion probability under the initial channel condition.*

Proposition 4. *If $\beta = 0$ and $M = 1$, then the PI* rule is optimal under arbitrary arrivals. Note that it is equivalent to the $c\mu$-rule, in which μ_k is the job completion probability under the actual channel condition.*

Lott and Teneketzis (2000) gave sufficient conditions for optimality of the $c\mu$-rule in queueing systems with $M \geq 1$ servers (which cannot serve more than one user from the same queue) and with ON/OFF channels (in which $\mu_{k,B} = 0$). In that setting, the rule serves M non-empty and connected (i.e., in condition G) queues with highest index value $c_k\mu_{k,G}$. These results can be adapted to PI* as follows.

Proposition 5. *If channels of all the users belonging to the same class are perfectly correlated, $\mu_{k,B} = 0$ for all k, and at most one job of every class can be served (but number of classes served $M \geq 1$), then the PI* rule is optimal under arbitrary arrivals, if there is a labeling of classes such that*

$$c_k\mu_{k,G}\frac{1-\beta}{1-\beta+\beta\mu_{k,G}} \geq c_l\mu_{l,G}, \text{ for all } 1 \leq k < l \leq K. \tag{6}$$

Moreover, if the arrivals are Bernoulli with probability λ_k of an arrival for every class k, then the condition can be relaxed to

$$c_k\mu_{k,G}\frac{1-\beta}{1-\beta+\beta\mu_{k,G}(1-\lambda_k)} \geq c_l\mu_{l,G}, \text{ for all } 1 \leq k < l \leq K. \tag{7}$$

Notice that the first optimality condition is never satisfied for $\beta = 1$, which corresponds to the time-average criterion. Notice also that if there is exactly one job arrival to each class ($\lambda_k = 1$ for all k), then the second condition is satisfied for any β. We finally remark that these optimality results hold also for the finite-horizon optimization problem.

Ayesta et al. (2011) showed that all the so-called *best rate priority* rules are maximally stable and fluid-optimal in systems with i.i.d. channel evolution and capacity $M = 1$ job in service. Such rules must give priority to users in the good condition, and to choose among them accordingly to the $c\mu$-rule; both these conditions are satisfied by PI*.

Proposition 6. *If $q_{k,B,G} = q_{k,G,G}$ for all k (i.i.d. channel evolution) and $M = 1$, then the PI* rule is maximally stable and fluid-optimal under the time-average criterion and arrival processes that for each class are i.i.d. with finite second moment.*

We believe that this claim of Ayesta et al. (2011) can be extended to the case of Markovian channel evolution and $M \geq 1$, but becomes technically much more involved. We therefore formulate the following conjecture.

Conjecture 1. *The PI* rule is maximally stable and fluid-optimal under the time-average criterion and arrival processes that for each class are i.i.d. with finite second moment. Moreover, the stability region is the same as in the i.i.d. channel evolution case.*

Whittle (1988) conjectured for the restless bandit problem (with a fixed population of K bandits) that the index rule derived by the Lagrangian relaxation becomes asymptotically optimal as both the capacity M and the number of bandits K grow to a fixed proportion. This conjecture was proved true under certain conditions that are hard to check in Weber and Weiss (1990). As we have mentioned earlier, our model without arrivals is a special case of the restless bandit problem. Although it is not clear whether these sufficient conditions are obeyed by our model, and although we admit arrivals of new jobs to the system, we believe that the original conjecture of Whittle (1988) may remain valid with a slight modification. Such a result would be of special importance, because it guarantees scalability of the PI* rule.

Conjecture 2. *If both the capacity M and the arrival rates of each class grow in fixed proportion, then the PI* rule approaches optimality under arrival processes that for each class are i.i.d. with finite second moment.*

6 Experimental Study

In order to study performance of the PI* rule in systems not covered by theoretical results stated in the previous section, we consider a two-class system with Bernoulli user arrivals and single-user capacity $M = 1$. Every arriving job belongs to one of $K = 2$ classes which may differ in all the parameters as in the previous discussion. A new job arrives to the system with probability λ per slot. Given a user arrival, she belongs to class k and finds herself in channel condition n with probability $\lambda_{k,n}/\lambda$.

In order to keep the state space finite and assure stability of the system (i.e., finiteness of the value function), we consider a system with blocking of new arrivals of a class whenever there are 10 uncompleted jobs of the respective class. Nevertheless, for comparison we also report the value of the parameter

$$\varrho^* := \frac{\lambda_{1,B} + \lambda_{1,G}}{\mu_{1,G}} + \frac{\lambda_{2,B} + \lambda_{2,G}}{\mu_{2,G}},$$

which was shown in Ayesta et al. (2011) to characterize the stability region in (non-blocking) systems with i.i.d. channel evolution by $\varrho^* < 1$.

We have performed experiments in many scenarios and selected six, which we believe illustrate typical pattern. The parameters set in the six scenarios reported here are summarized in Table 1. We have compared PI*, PI-SS (a variant of PI* in which $q^*_{k,B,G}$ is replaced by the steady-state probability $q^{\mathrm{SS}}_{k,G}$), and PI1 (a variant of PI* in which $q^*_{k,B,G}$ is replaced by the one-slot probability $q_{k,B,G}$), both with $c\mu$ tie-breaking and with randomized tie-breaking in condition G. Moreover, we have included the SB rule (Bonald, 2004b) modified to use $c\mu$ tie-breaking.

Scenario 1. Varying both q_{1BG} and q_{1GB} for class 1 means changing the character of its channel from slow-fading to fast-fading. We keep $\varrho^* = 0.5$. We can see no significant effect of such a change in the channel on performance. More interestingly, all rules with $c\mu$ tie-breaking are optimal (except for the first point), while all the rules with randomized tie-breaking perform significantly worse with relative suboptimality gap between 5% and 8%.

Parameters	μ_{1B}	μ_{1G}	μ_{2B}	μ_{2G}	q_{1BG}	q_{1GG}	q_{2BG}	q_{2GG}	$\lambda_{1B} = \lambda_{1G}$	$\lambda_{2B} = \lambda_{2G}$	$c_1 = c_2$
Scenario 1	0.001	0.01	0.1	0.2	$[0.1, 0.9]$	$1 - q_{1BG}$	0.1	0.4	0.002	0.01	1
Scenario 2	$[0.05, 0.25]$	1.00	0.0002	0.1	0.99999	0.00001	0.01	0.01	0.1	0.3	1
Scenario 3	0.001	0.01	0.1	0.2	0.2	$[0.1, 0.9]$	0.1	0.4	0.002	0.01	1
Scenario 4	$\alpha 0.001$	$\alpha 0.01$	0.1	0.2	0.2	0.84	0.1	0.4	0.002	0.01	1
Scenario 5	0.0072	0.009	$\alpha 0.1$	$\alpha 0.2$	0.2	0.8	0.1	0.9	0.002	0.01	1
Scenario 6	0.0072	0.009	$\alpha 0.1$	$\alpha 0.2$	0.2	0.84	0.1	0.4	0.002	0.01	1

Table 1: Parameters set in the experimental study.

Scenario 2. In Scenario 2, jobs of class 1 are completed within one slot if the channel condition is good ($\mu_{1G} = 1$), but good condition is extremely unlikely. Note that $\mu_{1G} = 1$ implies that PI* is the same as PI1. We vary μ_{1B}, that is, the quality of transmission in the bad condition. We have $\varrho^* = 6.2$, i.e, the system is overloaded and therefore new arrivals are often blocked. It is important to note that all the rules achieve excellent performance well below 1%. The jumps in performance are caused by changes in the priority of service under bad channel conditions in all the PI variants, while SB does not change. It is interesting to realize that PI-SS with $c\mu$ tie-breaking is the best rule, as it is the first in switching, later followed by PI*. For $\mu_{1B} \geq 0.17$, SB performs inferior than all the PI variants, even those with randomized tie-breaking.

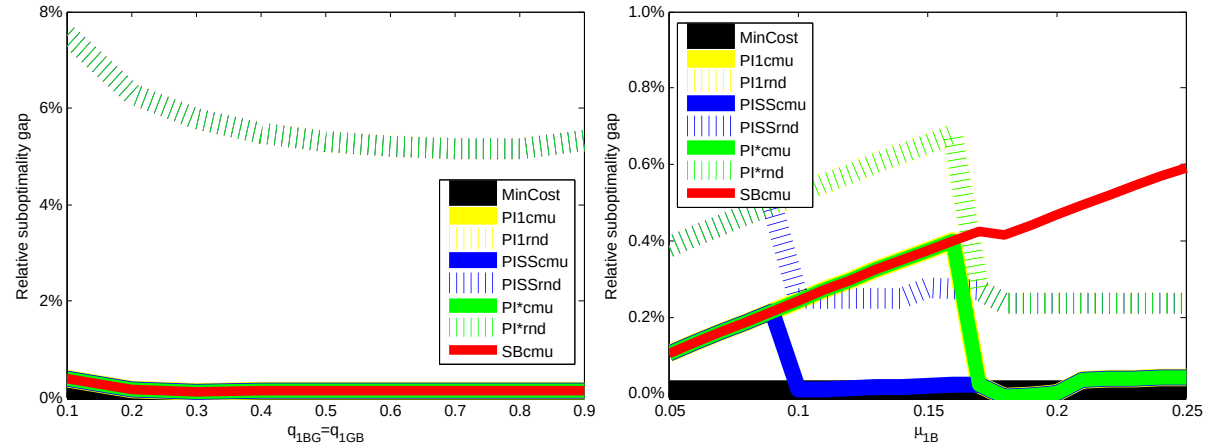
Scenario 3. In this scenario we vary the probability for class 1 of maintaining the good channel. We keep $\varrho^* = 0.5$. All the rules with $c\mu$ tie-breaking have equivalent performance below 3% (in a subinterval even optimal), and all the rules with randomized tie-breaking have also equivalent performance, which varies significantly. In particular, when the probability of maintaining the good channel increases (i.e., q_{qGB} in Figure 1(c) approaches zero), randomized rules deteriorate significantly. On the other hand, the randomized rules outperform slightly the $c\mu$ rules in the other extreme.

Scenario 4. Scenario 4 presented in Figure 1(d) shows an effect of decreasing expected job size of class 1, which is approximated by multiplying the completion probabilities in both channel conditions by the same factor. Note that ϱ^* varies from 4.1 to 0.3, and it crosses the stability region at $\alpha = 0.44$. Again, all the rules with $c\mu$ tie-breaking have equivalent performance and all the rules with randomized tie-breaking have also equivalent performance, but significantly worse. Except for the "unstable" values $\alpha \leq 0.44$, the $c\mu$ rules remain below 1% gap.

Scenarios 5 and 6. These two scenarios show that for small intervals of parameters the suboptimality gap of the rules with $c\mu$ tie-breaking can be large in the "unstable" setting. We emphasize that we have not observed such an effect in other scenarios, so it is likely to be very uncommon. In Scenario 6, the same parameter is varied as in Scenario 5, but on an interval 10 times larger. In both scenarios, ϱ^* starts with a value above 1 and ends with a value below 1, and it crosses the stability region at $\alpha = 0.18$. In Scenario 5, the rules with randomized tie-breaking significantly outperform the other rules, which achieve the relative suboptimality gap of over 80%. In Scenario 6, the best performing rules are SB and PI*, which are identical except for interval [0.1,0.3], in which PI* is better, and [0.3,0.6], in which SB is better. In Scenario 6 we can observe that PI-SS is significantly worse than PI* and SB, achieving the relative suboptimality gap of about 10% while PI* and SB are below 2%.

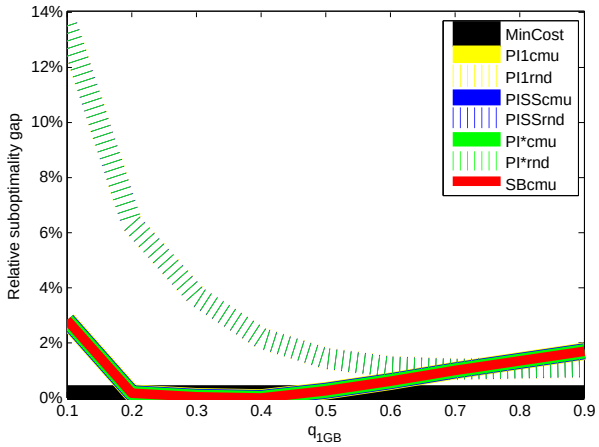
7 Conclusion

Based on a solid mathematical ground of Whittle's and Lagrangian relaxation, we have designed in this paper a new opportunistic scheduling rule for Markovian time-varying channels. It was by no means obvious that the index for Markovian channels would be an analogy of the Potential Improvement rule, derived in Ayesta et al. (2010) for i.i.d. channel evolution. An excellent nearly-optimal performance of this new PI* rule in computational experiments and comparison to its variants gives insights about the value of information to be taken into account in order to design schedulers with satisfactory performance in wireless networks. Moreover, it suggests a relatively simple answer for how the trade-off between the

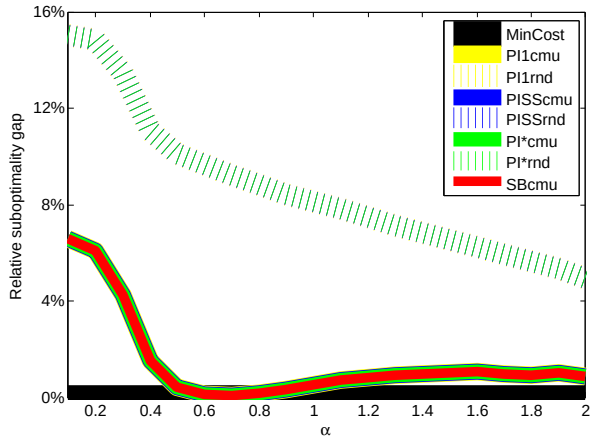


(a) Scenario 1

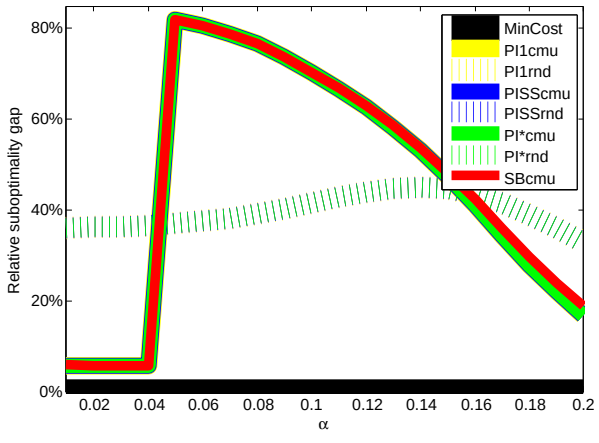
(b) Scenario 2



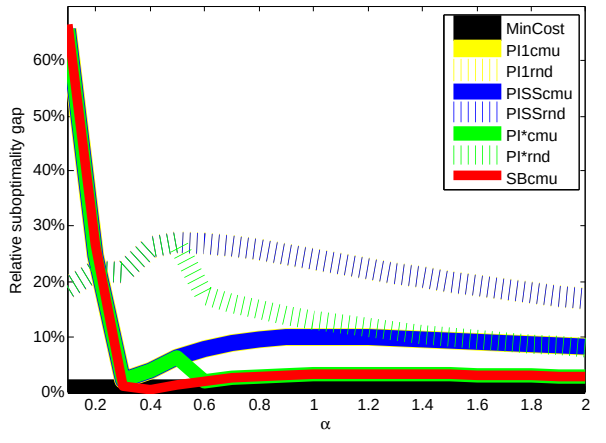
(c) Scenario 3



(d) Scenario 4



(e) Scenario 5



(f) Scenario 6

Figure 1: Relative suboptimality gap as a function of varying parameters according to Table 1.

short-job prioritization and opportunistic gains (studied also in Sadiq and de Veciana, 2010; Aalto et al., 2011, for deterministic job sizes) should be resolved in time-varying systems.

Notice that for *short jobs*, i.e., those satisfying $\mu_{k,G} = 1$, PI* is the same as PI1, since $q^*_{k,B,G} = q_{k,B,G}$. On the other hand, for *long jobs*, i.e., those satisfying $\mu_{k,G} \approx 0$, PI* is approximately the same as PI-SS. We have seen in our experiments that the differences in performance between PI, PI*, PI1 and SB are negligible as long as the $c\mu$-rule is taken for tie-breaking in the good state. Inferiority of SB was observed

only in systems with short jobs (Scenario 2), but otherwise the comparable performance of SB is rather surprising. It seems that the effect of the tie-breaking rule in the good channel condition is much more important than the effect of the choice under the bad conditions. However, it is not clear whether such an excellent performance of SB would propagate to systems with more than two channel conditions. In that case the importance of priorities in the non-best conditions should dramatically increase.

We believe that the basic principles of giving an absolute priority to users in the best channel condition, and comparing the others according to a ratio of their current transmission rate and the potential improvement in their transmission rate extend also to systems with a more complex functionality. Those include especially non-geometric job sizes and partially observable channel conditions. Note that in the case of general job sizes, but non-varying channels, our approach leads to the Gittins index rule, which is optimal (Gittins, 1989). Theoretical analysis of indexability in time-varying setting may, however, be prohibitively complicated in order to derive closed-form index values or fast implementable algorithms for their computation.

References

Aalto, S. and Lassila, P. (2010). Flow-level stability and performance of channel-aware priority-based schedulers. In *Proceeding of NGI 2010 (6th EURO-NF Conference on Next Generation Internet)*.

Aalto, S., Penttinen, A., Lassila, P., and Osti, P. (2011). On the optimal trade-off between SRPT and opportunistic scheduling. In *Proceedings of Sigmetrics*.

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

Ayesta, U., Erausquin, M., Jonckheere, M., and Verloop, I. M. (2011). Scheduling in a random environment: Stabilility and asymptotic optimality. arXiv:1101.5794v1.

Ayesta, U. and Jacko, P. (2010). Method for selecting a transmission channel within a time division multiple access (TDMA) communications system. EU Patent Application.

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

Bonald, T. (2004a). Procédé de sélection de canal de transmission dans un protocole d'accès multiple à répartition dans le temps et système de communication mettant en oeuvre un tel procédé. EU Patent.

Bonald, T. (2004b). A score-based opportunistic scheduler for fading radio channels. In *Proceedings of European Wireless*, pages 283–292.

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

Borst, S. (2005). User-level performance of channel-aware scheduling algorithms in wireless data networks. *IEEE/ACM Transactions on Networking*, 13(3):636–647.

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

Chaponniere, E. F., Black, P. J., Holtzman, J. M., and Tse, D. N. C. (2002). Transmitter directed code division multiple access system using path diversity to equitably maximize throughput. US Patent.

Gilbert, E. N. (1960). Capacity of a burst-noise channel. *Bell Systems Technical Journal*, 39:1253–1266.

Gittins, J. C. (1989). *Multi-Armed Bandit Allocation Indices*. J. Wiley & Sons, New York.

Jacko, P. (2009). Adaptive greedy rules for dynamic and stochastic resource capacity allocation problems. *Medium for Econometric Applications*, 17(4):10–16. Available online at http://www.met-online.nl. Invited paper.

Jacko, P. (2010a). *Dynamic Priority Allocation in Restless Bandit Models*. Lambert Academic Publishing. Invited book.

Jacko, P. (2010b). Restless bandits approach to the job scheduling problem and its extensions. In Piunovskiy, A. B., editor, *Modern Trends in Controlled Stochastic Processes: Theory and Applications*, pages 248–267. Luniver Press, United Kingdom.

Knopp, R. and Humblet, P. (1995). Information capacity and power control in single-cell multiuser communications. In *Proceedings of IEEE International Conference on Communications*, pages 331–335.

Kushner, H. and Whiting, P. (2004). Convergence of proportional-fair sharing algorithms under general conditions. *IEEE Transactions on Wireless Communications*, 3:1250–1259.

Liu, S., Ying, L., and Srikant, R. (2011). Throughput-optimal opportunistic scheduling in the presence of flow-level dynamics. *IEEE/ACM Transactions on Networking*, PP(99):1.

Liu, X., Chong, E. K. P., and Shroff, N. B. (2003). Optimal opportunistic scheduling in wireless networks. In *Proceedings of IEEE 58th Vehicular Technology Conference*, pages 1417–1421.

Lott, C. and Teneketzis, D. (2000). On the optimality of an index rule in multichannel allocation for single-hop mobile networks with multiple service classes. *Probability in the Engineering and Informational Sciences*, 14:259–297.

Niño-Mora, J. (2001). Restless bandits, partial conservation laws and indexability. *Advances in Applied Probability*, 33(1):76–98.

Niño-Mora, J. (2007a). 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.

Niño-Mora, J. (2007b). Dynamic priority allocation via restless bandit marginal productivity indices. *TOP*, 15(2):161–198.

Puterman, M. L. (2005). *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, Inc., Hoboken, New Jersey.

Sadiq, B. and de Veciana, G. (2009). Throughput optimality of delay-driven Maxweight scheduler for a wireless system with flow dynamics. In *Proceedings of Annual Allerton Conference on Communication, Control and Computing*.

Sadiq, B. and de Veciana, G. (2010). Balancing SRPT prioritization vs opportunistic gain in wireless systems with flow dynamics. In *Proceedings of ITC-22*.

Tassiulas, L. and Ephremides, A. (1993). Dynamic server allocation to parallel queues with randomly varying connectivity. *IEEE Transactions on Information Theory*, 39(2):466–478.

van de Ven, P., Borst, S., and Shneer, S. (2009). Instability of MaxWeight scheduling algorithms. In *Proceedings of IEEE Infocom*, pages 1701–1709.

Weber, R. and Weiss, G. (1990). On an index policy for restless bandits. *Journal of Applied Probability*, 27(3):637–648.

Whittle, P. (1988). Restless bandits: Activity allocation in a changing world. *A Celebration of Applied Probability, J. Gani (Ed.), Journal of Applied Probability*, 25A:287–298.

A Auxiliary Material and Proofs

A.1 Proof of Lemma 1

Let $\Delta := \varepsilon s_{k,n}$ denote the amount of bits transferred in one slot in channel condition n. Then the probability that a user leaves the system if served in channel condition n is $\mathbb{P}[b \leq B_k \leq b + \Delta | B_k > b] = \mathbb{P}[B_k \leq \Delta]$, which does not depend on the attained service b due to the memoryless property of geometric distribution.

If $\mathbb{E}[B_k] = 1/(1-\alpha)$, then given that the job size is at least b bits, it is at least $b+1$ bits with probability $\alpha = 1 - 1/\mathbb{E}[B_k]$. Therefore, $\mathbb{P}[B_k > \Delta] = \alpha^{\Delta}$, so $\mathbb{P}[B_k \leq \Delta] = 1 - \alpha^{\Delta} = 1 - (1 - 1/\mathbb{E}[B_k])^{\Delta}$. Moreover, $\mathbb{P}[B_k \leq \Delta] \to \Delta/\mathbb{E}[B_k]$ as $\Delta/\mathbb{E}[B_k] \to 0$.

A.2 Work-Reward Analysis

In order to prove Proposition 1 and Proposition 2, we will focus on the case $\beta < 1$, i.e., the problem under the discounted criterion. Problem (2) 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 (Puterman, 2005, Chapter 6). In particular, this implies that there exists an optimal policy which only depends on the user-k state-process $X_k(\cdot)$. Indeed, policy $\widetilde{\pi}_k \in \Pi_{\boldsymbol{X},a_k}$ that depends on the joint state-process $\boldsymbol{X}(\cdot)$ can be seen as a randomized policy, since the user-l state-process $X_l(\cdot)$ for $l \neq k$ is not influenced by the user-k action-process $a_k(\cdot)$ prescribed by $\widetilde{\pi}_k$.

Therefore, in order to find an optimal policy to problem (2) it is enough to concentrate on stationary policies $\pi_k \in \Pi_{X_k,a_k}$. Every such policy can be represented in terms of a *serving set* $\mathcal{S} \subseteq \mathcal{N}_k$, which prescribes to allocate the base station's service (i.e., to serve) whenever the user is in state $n \in \mathcal{S}$ and not to serve whenever the user is in state $n \notin \mathcal{S}$. Thus, an optimal policy to problem (2) can be obtained by solving

$$\max_{\mathcal{S} \subseteq \mathcal{N}_k} \mathbb{B}_0^{\mathcal{S}} \left[R_{k,X_k(\cdot)}^{a_k(\cdot)} \right] - \nu \, \mathbb{B}_0^{\mathcal{S}} \left[W_{k,X_k(\cdot)}^{a_k(\cdot)} \right]. \tag{8}$$

Notice that (8) 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 (8) 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 $\mathbb{R}_n^{\mathcal{S}}$ if the initial state $X_k(0) = n \in \mathcal{N}_k$. Analogously, we will write briefly $\mathbb{W}_n^{\mathcal{S}}$ and the value function under policy $\mathcal{S}$ we denote by $\mathbb{V}_n^{\mathcal{S}} := \mathbb{R}_n^{\mathcal{S}} - \nu \mathbb{W}_n^{\mathcal{S}}$. These scaled quantities correspond to the usual quantities under the β-discounted criterion. Optimality of threshold policies and indexability under the time-average criterion is obtained in the limit $\beta \to 1$. In the rest of this section we will omit the user subscript $k \leq K-1$ to simplify the notation.

A.3 Proof of Proposition 1

Let us denote the optimal value function by $\mathbb{V}_n^*$. An inspection of the Bellman equation leads to the following lemma, which establishes optimality of threshold policies in Proposition 1.

Lemma 2.

(i) Suppose that $\nu > 0$ holds. Then, if it is optimal to serve in state 0 (resp. B), then it is optimal to serve in state B (resp. G).

(ii) Suppose that $\nu \leq 0$. Then it is optimal to serve in any state $n \in \mathcal{N}$.

Proof. The Bellman equation for state $n \in \mathcal{N}$ is $\mathbb{V}_n^* = \max_{a \in \mathcal{A}} \left\{ R_n^a - \nu W_n^a + \beta \sum_{m \in \mathcal{N}} p_{n,m}^a \mathbb{V}_m^* \right\}$. After

plugging the definitions of the action-dependent parameters for state $n \in \mathcal{N}'$, we obtain

$$\begin{aligned}\mathbb{V}_n^* &= \max\{-c(1-\mu_n) - \nu + \beta\left[(1-\mu_n)q_{n,B}\mathbb{V}_B^* + (1-\mu_n)q_{n,G}\mathbb{V}_G^* + \mu_n\mathbb{V}_0^*\right] \ ; \\ &\quad -c + \beta\left[q_{n,B}\mathbb{V}_B^* + q_{n,G}\mathbb{V}_G^*\right]\} \\ &= -c + \beta\left[q_{n,B}\mathbb{V}_B^* + q_{n,G}\mathbb{V}_G^*\right] + \max\left\{-\nu + \mu_n\left(c + \beta\mathbb{V}_0^* - \beta q_{n,B}\mathbb{V}_B^* - \beta q_{n,G}\mathbb{V}_G^*\right) \ ; 0\right\},\end{aligned}$$

where the first term in the curly brackets corresponds to action 1 and the second one to action 0. Serving (i.e., action 1) is optimal in state $n \in \mathcal{N}'$, if the first term is greater than or equal to the second term.

Analogously for state 0, using the Bellman equation it is straightforward to obtain that action 1 is optimal in state 0 iff $\nu \leq 0$ and action 0 is optimal in state 0 iff $\nu \geq 0$, so

$$\mathbb{V}_0^* = \max\{-\nu + \beta\mathbb{V}_0^* \ ; \beta\mathbb{V}_0^*\} = \begin{cases} -\dfrac{\nu}{1-\beta}, & \text{if } \nu \leq 0, \\ 0, & \text{if } \nu \geq 0. \end{cases}$$

(i) If $\nu > 0$, we have $\mathbb{V}_0^* = 0$, and so we can simplify the Bellman equation for states $n \in \mathcal{N}'$ to

$$\mathbb{V}_n^* = -c + \beta\left[q_{n,B}\mathbb{V}_B^* + q_{n,G}\mathbb{V}_G^*\right] + \max\left\{-\nu + \mu_n\left(c - \beta q_{n,B}\mathbb{V}_B^* - \beta q_{n,G}\mathbb{V}_G^*\right) \ ; 0\right\}.$$

Next we show that if it is optimal to serve in state B, then it is optimal to serve in state G. In order to obtain a contradiction, let us assume that it is optimal to serve at state B, and at the same time it is not optimal to serve at state G. Then the Bellman equations reduce to

$$\begin{aligned}\mathbb{V}_B^* &= -c + \beta\left[(1-q_{B,G})\mathbb{V}_B^* + q_{B,G}\mathbb{V}_G^*\right] - \nu + \mu_B\left(c - \beta(1-q_{B,G})\mathbb{V}_B^* - \beta q_{B,G}\mathbb{V}_G^*\right), \\ \mathbb{V}_G^* &= -c + \beta\left[(1-q_{G,G})\mathbb{V}_B^* + q_{G,G}\mathbb{V}_G^*\right],\end{aligned}$$

which can be rewritten in the form of a system of two equations with two unknowns

$$\begin{aligned}\left(1 - \beta(1-\mu_B)(1-q_{B,G})\right)\mathbb{V}_B^* - \beta(1-\mu_B)q_{B,G}\mathbb{V}_G^* &= -c(1-\mu_B) - \nu, \\ -\beta(1-q_{G,G})\mathbb{V}_B^* + \left(1 - \beta q_{G,G}\right)\mathbb{V}_G^* &= -c.\end{aligned}$$

In order to solve this system of linear equations, we continue by defining the following weighted harmonic mean,

$$q_{B,G}^{\maltese} := \frac{1}{\dfrac{1-\beta}{q_{B,G}} + \dfrac{\beta}{q_G^{\mathrm{SS}}}}, \tag{9}$$

where q_G^{SS} is defined in (4). Then the solution to the above system of equations is

$$\mathbb{V}_B^* = -\frac{c(1-\mu_B) + \nu(1-q_{B,G}^{\maltese})}{(1-\beta)(1-\mu_B) + \mu_B(1-q_{B,G}^{\maltese})}, \tag{10}$$

$$\mathbb{V}_G^* = -\frac{c(1-\mu_B) + c\mu_B\frac{q_{B,G}^{\maltese}}{\beta q_{B,G}} + \beta\nu(1-q_{G,G})\frac{q_{B,G}^{\maltese}}{\beta q_{B,G}}}{(1-\beta)(1-\mu_B) + \mu_B(1-q_{B,G}^{\maltese})}. \tag{11}$$

It is easy to check that the denominator in these two expressions is strictly positive, which is useful in the following manipulations.

Since it is optimal to serve at state B, we have that

$$-\nu + \mu_B\left(c - \beta(1-q_{B,G})\mathbb{V}_B^* - \beta q_{B,G}\mathbb{V}_G^*\right) \geq 0,$$

which is a contradiction if $\mu_B = 0$. In the remaining case, after plugging (10) and (11), it can be shown in a straightforward way to be equivalent to $\nu(1-\beta) \leq c\mu_B$. On the other hand, since it is

not optimal to serve at state G, we have that

$$-\nu + \mu_G \left(c - \beta(1 - q_{G,G})\mathbb{V}_B^* - \beta q_{G,G}\mathbb{V}_G^*\right) < 0,$$

which, after plugging (10) and (11), can be shown to be equivalent to

$$\nu \left\{ (1-\beta)(1-\mu_B) + (\mu_B - \mu_G)(1 - q_{B,G}^{\maltese}) + (1-\beta)\mu_G \frac{q_{B,G}^{\maltese}}{\beta q_{B,G}} \right\} > c\mu_G(1-\mu_B) + c\mu_G\mu_B \frac{q_{B,G}^{\maltese}}{\beta q_{B,G}}. \tag{12}$$

Using $\mu_G \geq \mu_B$ and $c\mu_B \geq \nu(1-\beta)$ (that we have obtained from the first inequality), we have that

$$c\mu_G(1-\mu_B) \geq \nu(1-\beta)(1-\mu_B), \qquad c\mu_G\mu_B \frac{q_{B,G}^{\maltese}}{\beta q_{B,G}} \geq \nu(1-\beta)\mu_G \frac{q_{B,G}^{\maltese}}{\beta q_{B,G}},$$

which combining with (12) implies

$$\begin{aligned} \nu \left\{ (1-\beta)(1-\mu_B) + (\mu_B - \mu_G)(1 - q_{B,G}^{\maltese}) + (1-\beta)\mu_G \frac{q_{B,G}^{\maltese}}{\beta q_{B,G}} \right\} \\ > \nu \left\{ (1-\beta)(1-\mu_B) + (1-\beta)\mu_G \frac{q_{B,G}^{\maltese}}{\beta q_{B,G}} \right\}, \end{aligned}$$

which is equivalent to

$$(\mu_B - \mu_G)(1 - q_{B,G}^{\maltese}) < 0.$$

This is a contradiction with $\mu_G \geq \mu_B$, since $q_{B,G}^{\maltese} \leq \max\{q_{B,G}, q_G^{\mathrm{SS}}\} \leq 1$ due to the properties of weighted harmonic mean.

Therefore the statement holds under $\nu > 0$ for all $n \in \mathcal{N} \setminus \{G\}$.

(ii) If $\nu \leq 0$, then serving is optimal in state 0. Notice that the one-period net reward $R_n^a - \nu W_n^a \leq -\nu$ for any state $n \in \mathcal{N}$ and any action $a \in \mathcal{A}$. Hence $\mathbb{V}_n^* \leq -\nu/(1-\beta) = \mathbb{V}_0^*$ for any $n \in \mathcal{N}'$, and therefore (using $c > 0$) also $c + \beta\mathbb{V}_0^* - \beta q_{n,B}\mathbb{V}_B^* - \beta q_{n,G}\mathbb{V}_G^* > 0$, and finally, $-\nu + \mu_n(c + \beta\mathbb{V}_0^* - \beta q_{n,B}\mathbb{V}_B^* - \beta q_{n,G}\mathbb{V}_G^*) \geq 0$ for any state $n \in \mathcal{N}'$. That is, serving is optimal in any state $n \in \mathcal{N}$.

□

A.4 Proof of Proposition 2

In order to prove the existence of optimal index values ν_n in terms of properties (i) and (ii) of Proposition 2, we will establish validity of a sufficient condition called *LP-indexability* introduced in (Niño-Mora, 2007a, Definition 5.3), which will be stated after defining some necessary concepts. The analysis in the following paragraphs also shows how to evaluate such index values provided they exist. An immediate result are the *balance equations* given in the following lemma.

Lemma 3. *For all states $n \in \mathcal{N}$ we have $\mathbb{W}_n^{\mathcal{N}} = 1/(1-\beta)$, and under any policy $0 \notin \mathcal{S}$ we have*

$$\mathbb{R}_n^{\mathcal{S}} = \begin{cases} -c(1-\mu_n) + (1-\mu_n)\beta \sum_{m \in \mathcal{N}'} q_{n,m}\mathbb{R}_m^{\mathcal{S}}, & \text{if } 0 \neq n \in \mathcal{S}, \\ -c + \beta \sum_{m \in \mathcal{N}'} q_{n,m}\mathbb{R}_m^{\mathcal{S}}, & \text{if } 0 \neq n \notin \mathcal{S}, \\ 0, & \text{if } n = 0. \end{cases}$$

$$\mathbb{W}_n^{\mathcal{S}} = \begin{cases} 1 + (1-\mu_n)\beta \sum_{m \in \mathcal{N}'} q_{n,m} \mathbb{W}_m^{\mathcal{S}}, & \text{if } 0 \neq n \in \mathcal{S}, \\ \beta \sum_{m \in \mathcal{N}'} q_{n,m} \mathbb{W}_m^{\mathcal{S}}, & \text{if } 0 \neq n \notin \mathcal{S}, \\ 0, & \text{if } n = 0. \end{cases}$$

Proof. Directly from the definition of β-average reward and work, respectively, we have

$$\mathbb{R}_n^{\mathcal{S}} = R_n^{n \in \mathcal{S}} + \beta \sum_{m \in \mathcal{N}} p_{n,m}^{n \in \mathcal{S}} \mathbb{R}_m^{\mathcal{S}}, \qquad \mathbb{W}_n^{\mathcal{S}} = W_n^{n \in \mathcal{S}} + \beta \sum_{m \in \mathcal{N}} p_{n,m}^{n \in \mathcal{S}} \mathbb{W}_m^{\mathcal{S}},$$

where $n \in \mathcal{S}$ equals 1 if true and 0 otherwise. Substituting the values of $R_n^{n \in \mathcal{S}}$, $W_n^{n \in \mathcal{S}}$ and $p_{n,m}^{n \in \mathcal{S}}$ given in the definition of the job-channel-user triple, and simplifying, results in the above characterization. □

If index value ν_n for $n \in \mathcal{N}'$ with the desired properties (i) and (ii) (and index value ν_0 for $n = 0$ with the desired properties (iii) and (iv)) stated in Proposition 2 exists, then both serving and not serving is optimal if $\nu = \nu_n$. This means that there is a policy, say $\mathcal{S}^*$, such that both including state n in $\mathcal{S}^*$ and not including it lead to the same objective value, i.e.,

$$\mathbb{R}_n^{\mathcal{S}^* \cup \{n\}} - \nu_n \mathbb{W}_n^{\mathcal{S}^* \cup \{n\}} = \mathbb{R}_n^{\mathcal{S}^* \setminus \{n\}} - \nu_n \mathbb{W}_n^{\mathcal{S}^* \setminus \{n\}}.$$

A straightforward consequence of this and of the balance equation is that changing the action only in the initial period must also lead to the same objective values, i.e.,

$$\mathbb{R}_n^{\langle 0, \mathcal{S}^* \rangle} - \nu_n \mathbb{W}_n^{\langle 0, \mathcal{S}^* \rangle} = \mathbb{R}_n^{\langle 1, \mathcal{S}^* \rangle} - \nu_n \mathbb{W}_n^{\langle 1, \mathcal{S}^* \rangle},$$

where policy $\langle a, \mathcal{S}^* \rangle$ is the policy that employs action a in the initial period and then proceeds according to $\mathcal{S}^*$. Then, whenever $\mathbb{W}_n^{\langle 1, \mathcal{S}^* \rangle} - \mathbb{W}_n^{\langle 0, \mathcal{S}^* \rangle} \neq 0$, we have

$$\nu_n = \frac{\mathbb{R}_n^{\langle 1, \mathcal{S}^* \rangle} - \mathbb{R}_n^{\langle 0, \mathcal{S}^* \rangle}}{\mathbb{W}_n^{\langle 1, \mathcal{S}^* \rangle} - \mathbb{W}_n^{\langle 0, \mathcal{S}^* \rangle}}, \tag{13}$$

We will therefore study ν_n under all policies $\mathcal{S}$, defined as

$$\nu_n^{\mathcal{S}} := \frac{\mathbb{R}_n^{\langle 1, \mathcal{S} \rangle} - \mathbb{R}_n^{\langle 0, \mathcal{S} \rangle}}{\mathbb{W}_n^{\langle 1, \mathcal{S} \rangle} - \mathbb{W}_n^{\langle 0, \mathcal{S} \rangle}}. \tag{14}$$

From the balance equations we can obtain the following characterization of these quantities.

Lemma 4. *For any state $n \in \mathcal{N}'$ under any policy $0 \notin \mathcal{S}$ we have*

$$\nu_n^{\mathcal{S}} = \mu_n \frac{c - \beta \sum_{m \in \mathcal{N}'} q_{n,m} \mathbb{R}_m^{\mathcal{S}}}{1 - \mu_n \beta \sum_{m \in \mathcal{N}'} q_{n,m} \mathbb{W}_m^{\mathcal{S}}}, \qquad \nu_0^{\mathcal{S}} = 0. \tag{15}$$

Proof. Using the characterization of $\mathbb{R}_n^{\langle a, \mathcal{S} \rangle}$ for $a \in \mathcal{A}$ from Lemma 3, we obtain that

$$\mathbb{R}_n^{\langle 1, \mathcal{S} \rangle} - \mathbb{R}_n^{\langle 0, \mathcal{S} \rangle} = \mu_n \left(c - \beta \sum_{m \in \mathcal{N}'} q_{n,m} \mathbb{R}_m^{\mathcal{S}} \right).$$

Similarly, we obtain that

$$\mathbb{W}_n^{\langle 1, \mathcal{S} \rangle} - \mathbb{W}_n^{\langle 0, \mathcal{S} \rangle} = 1 - \mu_n \beta \sum_{m \in \mathcal{N}'} q_{n,m} \mathbb{W}_m^{\mathcal{S}}. \tag{16}$$

Then, the definition in (14) yields the stated characterization. Finally, $\nu_0^{\mathcal{S}} = 0$ is obtained trivially. □

Since our objective is to find the optimal index values ν_n in terms of properties (i) and (ii) of Proposition 2, we postulate that

$$\text{For all } n \in \mathcal{N} : \nu_n = \nu_n^{\mathcal{S}_{N-n}} \text{ for } \mathcal{S}_{N-n} := \{m \in \mathcal{N} : m > n\}. \tag{17}$$

We will verify this postulate using a sufficient condition LP-indexability (Niño-Mora, 2007a, Definition 5.3), which in our problem can be simplified to the following.

Definition 1. *Problem (2) is LP-indexable with index values ν_n given in (17), if the following conditions hold:*

(i) $\mathbb{W}_n^{\langle 1,\emptyset\rangle} - \mathbb{W}_n^{\langle 0,\emptyset\rangle} > 0$ *and* $\mathbb{W}_n^{\langle 1,\mathcal{N}\rangle} - \mathbb{W}_n^{\langle 0,\mathcal{N}\rangle} > 0$ *for all* $n \in \mathcal{N}$*;*

(ii) $\mathbb{W}_n^{\langle 1,\mathcal{S}_{N-n}\rangle} - \mathbb{W}_n^{\langle 0,\mathcal{S}_{N-n}\rangle} > 0$ *and* $\mathbb{W}_{n+1}^{\langle 1,\mathcal{S}_{N-n}\rangle} - \mathbb{W}_{n+1}^{\langle 0,\mathcal{S}_{N-n}\rangle} > 0$ *for each* $n \in \mathcal{N} \setminus \{N\}$*;*

(iii) For every real-valued ν there exists $n \in \mathcal{N} \cup \{-1\}$ such that the serving set $\mathcal{S}_{N-n}$ is optimal.

We will first need to characterize the above quantities under $\mathcal{S}_{N-n}$ for any $n \in \mathcal{N}$. Let us denote by

$$\overline{\mathbb{W}}_l^{\mathcal{S}_{N-n}} := \beta \sum_{m \in \mathcal{N}'} q_{l,m} \mathbb{W}_m^{\mathcal{S}_{N-n}}.$$

According to the balance equations in Lemma 3, we have

$$\begin{aligned} \mathbb{W}_l^{\mathcal{S}_{N-n}} &= 1 + (1-\mu_l)\overline{\mathbb{W}}_l^{\mathcal{S}_{N-n}}, && \text{if } l \in \mathcal{S}_{N-n}, \\ \mathbb{W}_l^{\mathcal{S}_{N-n}} &= \overline{\mathbb{W}}_l^{\mathcal{S}_{N-n}}, && \text{if } l \notin \mathcal{S}_{N-n}. \end{aligned}$$

So, we need to solve the following system of linear equations:

$$\overline{\mathbb{W}}_j^{\mathcal{S}_{N-n}} - \beta \sum_{l \in \mathcal{S}_{N-n}} q_{j,l}(1-\mu_l)\overline{\mathbb{W}}_l^{\mathcal{S}_{N-n}} - \beta \sum_{l \notin \mathcal{S}_{N-n}} q_{j,l}\overline{\mathbb{W}}_l^{\mathcal{S}_{N-n}} = \beta \sum_{l \in \mathcal{S}_{N-n}} q_{j,l}, \text{ for all } j \in \mathcal{N}'. \tag{18}$$

Lemma 5. *For the Gilbert-Elliot model, (18) has the following solution:*

(i) for $n = 2$, so that the policy is $\mathcal{S}_{N-n} = \emptyset$, we have

$$\overline{\mathbb{W}}_l^{\mathcal{S}_{N-n}} = 0 \text{ for all } l \in \mathcal{N}'.$$

(ii) for $n = 1$, so that the policy is $\mathcal{S}_{N-n} = \{G\}$, we have

$$\overline{\mathbb{W}}_B^{\mathcal{S}_{N-n}} = \frac{\beta}{\dfrac{1-\beta}{q^*_{B,G}} + \beta\mu_G}, \qquad \overline{\mathbb{W}}_G^{\mathcal{S}_{N-n}} = \frac{(1-\beta)\beta q_{G,G} + \beta^2 q_{B,G}}{q_{B,G}\left\{\dfrac{1-\beta}{q^*_{B,G}} + \beta\mu_G\right\}}.$$

(iii) for $n = 0$, so that the policy is $\mathcal{S}_{N-n} = \{B, G\}$, we have

$$\overline{\mathbb{W}}_B^{\mathcal{S}_{N-n}} = \frac{\dfrac{\beta}{q^*_{B,G}}}{\dfrac{1-\beta(1-\mu_B)}{q^*_{B,G}} + \beta(\mu_G - \mu_B)}.$$

Analogously for rewards we need to solve the following system:

$$\overline{\mathbb{R}}_j^{\mathcal{S}_{N-n}} - \beta \sum_{l \in \mathcal{S}_{N-n}} q_{j,l}(1-\mu_l)\overline{\mathbb{R}}_l^{\mathcal{S}_{N-n}} - \beta \sum_{l \notin \mathcal{S}_{N-n}} q_{j,l}\overline{\mathbb{R}}_l^{\mathcal{S}_{N-n}} = -\beta c + \beta c \sum_{l \in \mathcal{S}_{N-n}} q_{j,l}\mu_l, \text{ for all } j \in \mathcal{N}'. \tag{19}$$

Lemma 6. *For the Gilbert-Elliot model, (19) has the following solution:*

(i) for $n = 2$, so that the policy is $\mathcal{S}_{N-n} = \emptyset$, we have

$$\overline{\mathbb{R}}_l^{\mathcal{S}_{N-n}} = \frac{-\beta c}{1-\beta} \text{ for all } l \in \mathcal{N}'.$$

(ii) for $n = 1$, so that the policy is $\mathcal{S}_{N-n} = \{G\}$, we have

$$\overline{\mathbb{R}}_B^{\mathcal{S}_{N-n}} = \frac{-\beta c\left(\frac{1}{q^*_{B,G}} - \mu_G\right)}{\frac{1-\beta}{q^*_{B,G}} + \beta\mu_G}, \qquad \overline{\mathbb{R}}_G^{\mathcal{S}_{N-n}} = \frac{-\beta c\left(\frac{q_{B,G}}{q^*_{B,G}} - q_{G,G}\mu_G\right)}{q_{B,G}\left\{\frac{1-\beta}{q^*_{B,G}} + \beta\mu_G\right\}}.$$

After plugging the expressions from the last two lemmas into (15), we have the following characterization.

Lemma 7. *For the Gilbert-Elliot model, we have*

$$\nu_G^{\emptyset} = \frac{c\mu_G}{1-\beta}, \qquad \nu_B^{\emptyset} = \frac{c\mu_B}{1-\beta}, \tag{20}$$

$$\nu_G^{\{G\}} = \frac{c\mu_G}{1-\beta}, \qquad \nu_B^{\{G\}} = \frac{c\mu_B}{1-\beta+\beta q^*_{B,G}(\mu_G-\mu_B)}. \tag{21}$$

Next we establish that the LP-indexability holds.

Lemma 8. *For the Gilbert-Elliot model, problem (2) is LP-indexable with index values ν_n given in (17).*

Proof.

(i) It is straightforward to obtain from Lemma 3 that $\mathbb{W}_n^{\langle 1,\emptyset\rangle} - \mathbb{W}_n^{\langle 0,\emptyset\rangle} = \mathbb{W}_n^{\langle 1,\mathcal{N}\rangle} - \mathbb{W}_n^{\langle 0,\mathcal{N}\rangle} = 1$ for all $n \in \mathcal{N}$.

(ii) We need to prove positivity of four differences. First, we have $\mathbb{W}_0^{\langle 1,\{B,G\}\rangle} - \mathbb{W}_0^{\langle 0,\{B,G\}\rangle} = 1 > 0$, because state 0 is absorbing, so action 0 will be applied all the time under policy $\{B, G\}$. Further, according to (16) the remaining three inequalities are equivalent to proving $\mu_m \overline{\mathbb{W}}_m^{\mathcal{S}_{N-n}} < 1$ for $(n, m) \in \{(0, B), (B, B), (B, G)\}$. Using Lemma 5 we have

$$\mu_B \overline{\mathbb{W}}_B^{\mathcal{S}_{N-0}} = \frac{\beta\mu_B}{(1-\beta)+\beta q^*_{B,G}(\mu_G-\mu_B)+\beta\mu_B} < 1, \qquad \mu_B \overline{\mathbb{W}}_B^{\mathcal{S}_{N-1}} = \frac{\beta\mu_B}{\frac{1-\beta}{q^*_{B,G}}+\beta\mu_G} < 1,$$

where we have used properties $\mu_G \geq \mu_B$ and $\beta < 1$.

Finally, case $(n, m) = (B, G)$ requires expanding $q^*_{B,G}$ and becomes more tedious, but otherwise it is straightforward to prove that $\mu_G \overline{\mathbb{W}}_G^{\mathcal{S}_{N-1}} < 1$.

(iii) This is established in Proposition 1.

□

Notice that expressions (5) are obtained from (20) by Definition 1 using $\nu_n = \nu_n^{\mathcal{S}_{N-n}}$. Therefore, since LP-indexability is a sufficient condition for the properties (i) and (ii) (and (iii) and (iv)) of Proposition 2, we conclude its proof.