Opportunistic Schedulers for Optimal Scheduling of Flows in Wireless Systems with ARQ Feedback

Peter Jacko*, Sofía S. Villar†

*BCAM—Basque Center for Applied Mathematics, Mazarredo 14, 48009 Bilbao (Basque Country), Spain
†Department of Statistics, Universidad Carlos III de Madrid, Avda. Universidad 30, 28911 Leganés (Madrid), Spain

Abstract—**In this paper we study three opportunistic schedulers for the problem of optimal multi-class flow-level scheduling in wireless downlink and uplink systems. For user channels we employ the Gilbert-Elliot model of good and bad channel condition with flow-level interpretation, and assume an automatic repeat query (ARQ) feedback, so that channel state information is available at the end of the slot only if the user was scheduled. The problem is essentially a Partially-Observable Markov Decision Process with a sample-path resource constraint. Given its complexity, we study two naïve schedulers: the myopic rule and the belief-state rule. Further, realizing that the problem fits the multi-armed restless bandit framework, we consider the relaxation of the problem which instead of serving a given number of flows on sample-path allows for serving that number of flows only in expectation, and derive an optimal Whittle index policy in closed form. We further discuss the interpretation of the resulting novel Whittle-index-based heuristic scheduler and evaluate its performance against the two naïve schedulers in simulations under the time-average criterion. According to the Whittle-index-based scheduler, the users whose last channel feedback gave good condition and those not served yet receive an absolute priority over those whose last channel feedback gave bad condition, which extends to this setting the property of channel-aware schedulers that are known to be maximally stable. In addition, we obtain tie-breaking index values for setting priorities among users in each of the two groups. In case of a single user class, the scheduler becomes independent of the problem parameters and equivalent to both the myopic and belief-state scheduler, and has a simple universal structure which can be represented by three first-in-first-out priority lists.**

Keywords: opportunistic scheduling, flow-level scheduling, ARQ feedback, Partially-Observable Markov Decision Process, restless bandits, Whittle index, myopic policy

I. Introduction

In this paper we design and characterize in closed form a novel opportunistic scheduler and two naïve schedulers for the problem of optimal multi-class flow-level scheduling. The problem is motivated by the necessity of designing an implementable scheduler for wireless downlink and uplink systems, in which users arrive randomly and depart upon service completion. It is believed that good schedulers are those that are opportunistic, since Knopp and Humblet (1995) showed that time-varying transmission conditions such as those in wireless systems can be exploited by *opportunistic scheduling* to enhance the system capacity. However, the understanding of the right way of opportunism in systems with heterogeneous users is non-trivial and it is not known at the moment what an optimal scheduler is, due to the inherent complexity of the problem that is likely to be PSPACE-hard (Ayesta et al., 2010).

Flow-level scheduling in systems like the CDMA 1xEV-DO (cf. Bender et al., 2000) has therefore been analyzed by ad-hoc approaches, simulations, approximate techniques or under restrictive assumptions (e.g., the *time-scale separation* principle or single-class systems) in order to design simple schedulers (Sadiq and de Veciana, 2010; Aalto et al., 2011; Ayesta et al., 2010; Jacko, 2011), and in the asymptotic regimes in order to establish optimality or maximal stability of such schedulers (Borst, 2005; Bonald et al., 2009; Aalto and Lassila, 2010; Ayesta et al., 2011; Ouyang et al., 2011a). However, all these papers addressed systems under the often unrealistic assumption that the channel condition of each user in the system is known at the beginning of each slot (so called *channel-aware* scheduling).

In this paper we take a step forward towards removing this assumption and consider a system with an automatic repeat query (ARQ) feedback, so that channel state information is available *at the end* of the slot *only if* the user was scheduled. We assume existence of feedback because of the necessity of reliability on data transmission (see also Ouyang et al. (2011b)). This feature adds yet another level of complexity to the problem, and we take a Bayesian approach to handle it. We focus on a downlink system, but the results hold also for any centralized uplink system with synchronized users. That is, at the beginning of each slot, the scheduler decides which users are allowed to receive data and the base station consequently transmits data via the corresponding channels to the scheduled users. We address the optimal scheduling problem with Markovian channel condition evolution. For simplicity we focus on channels with only two possible conditions, thus each channel is modeled by a Gilbert-Elliot model (Gilbert, 1960). The channel model assumes good (high transmission rate) and bad (low transmission rate) channel condition with flow-level interpretation, i.e., with a different probability to depart from the system. The scheduler is further allowed to do rate adaptation, i.e., to decide whether to transmit at the high or at the low transmission rate. The problem is described in Section II in more detail.

Mathematically, we formulate the problem of optimal user scheduling in a system with fixed server capacity and with

Emails: jacko@bcamath.org, svillar@est-econ.uc3m.es.
In Proceedings of ITC24, September 2012.

time-varying service rate. Such an optimization problem becomes analytically tractable if the fixed-capacity constraint is relaxed so that the server capacity allocation must be satisfied only on average (see, e.g., Whittle, 1988; Knopp and Humblet, 1995; Jacko, 2010b). This approach has become popular for scheduling problems in wireless networks (Zhao et al., 2008; Niño-Mora, 2008; Ahmad et al., 2009; Ayesta et al., 2010; Jacko, 2011; Ouyang et al., 2011b), mainly because it allows to design well-performing schedulers by formulating such problems within the multi-armed restless bandit problem framework, and performing a problem decomposition into single-user subproblems (Whittle, 1988). Moreover, in the single-class case it allows to develop simple optimal schedulers in some cases (Zhao et al., 2008; Ahmad et al., 2009; Liu et al., 2011; Wang et al., 2012).

Most of this literature, however, considers a packet-level system, in which a fixed number of users (that never depart) with their packet queues must be served, or in which a secondary user tries to transmit data over a fixed number of channels. Such models are fundamentally different from our flow-level model, in which user flows randomly arrive and depart once their transmission is completed, addressed within the restless bandit framework in Ayesta et al. (2010); Jacko (2011).

A. Main Results

We develop in Section III a novel model with the following characteristics:

- flow level (finite flow sizes);
- partial channel observability via ARQ feedback;
- non-zero transmission rate in bad state;
- rate adaptation.

The formulation is done within the framework of Partially Observable Markov Decision Processes (POMDP), relying on the Bayesian updates from slot to slot, taking into account ARQ feedback whenever available. The objective is to minimize the total expected time-average holding costs, which covers also minimization of the average expected waiting time as a special case. After employing the Lagrangian relaxation approach proposed by Whittle (1988) for the multi-armed restless bandit problem, the problem is decomposed into single-user parametric problems (where one must pay for service). Under the assumption of *indexability*, which is often tedious to check analytically (but which we conjecture to hold in this case), we derive in Section IV closed-form expressions of the Whittle index, which is shown to imply optimality of threshold policies.

Section V is dedicated to the design of a novel scheduler based on the derived Whittle index. We also define two naïve schedulers: the myopic rule and the belief-state rule. In the case of single-class system (i.e., flows having the same all the parameters and differing only by the available historical information), all these schedulers become equivalent, independent of the problem parameters, and have a simple universal structure which can be represented by three first-in-first-out priority lists of uncompleted flows:

(i) in the high priority list there are all the users served in the previous slot whose feedback gave good condition;
(ii) in the medium priority list there are all the users with no known feedback;
(iii) in the low priority list there are all the users whose last feedback gave bad condition.

Note that this scheduler has the *stay-on-a-winner* property, in the sense that every time a flow is scheduled and observes condition G, then it is served also in the next slot. This is an interesting and desirable property of our scheduler, which essentially means to be *opportunistic*. As a consequence, it can never happen that a flow whose last observed state was G has been starving at all. Further, newly arrived users (belonging to the medium priority list) are served as early as there are not enough high-priority users, which is desirable for quick refreshing of their channel condition information. On the other hand, due to the first-in-first-out order within each list, the newly arrived flows and the flows whose last observed state was B are served according to the longest-starved-first rule, which is desirable for fairness considerations.

We finally evaluate and discuss the performance of the proposed schedulers in Section VI under simple scenarios corresponding to the formulation described in Section III. Proofs are omitted due to space restrictions, and will also appear in the full version of this paper.

II. Problem Description

We consider the problem of designing a flow scheduling policy for a time-slotted system such as the CDMA 1xEV-DO (Bender et al., 2000), in which the available transmission rate of each user is time-varying. 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}). Flows $k = 1, 2, \dots$ 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 flow k is not completed. The channel for transmission of flow k (or shortly channel k) can take two conditions from a set $\mathcal{N}'_k := \{\mathrm{B}, \mathrm{G}\}$. The transmission channel 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 flow 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 flow k can be served with transmission/service rate $s_{k,n}$ (in bits per second), which is assumed to be a multiple of $1/\varepsilon$, or any lower. However, higher transmission rate than $s_{k,n}$ results in undecodable data leading to an outage. 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 flow has already been completed, then no transmission occurs. The base station is assumed to be preemptive (i.e., the service of a flow 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 users it should be allocated during that slot.

The goal is to minimize the expected aggregate holding cost over an infinite horizon 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 uncompleted flows in the system.

Let the flow 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. Ayesta et al. (2010); Jacko (2011) showed how the probability of departure of users with geometric sized flows can be computed respectively in an exact and in an approximate way when the expected flow size (in bits) of a user is much larger than the amount of bits that can be served in one slot. The departure probability of a flow 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 $\mu_{k,n} \approx \dfrac{\varepsilon s_{k,n}}{\mathbb{E}[B_k]}$.

A. Observability of Channel Conditions

This problem under full observability of channel conditions was studied in Jacko (2011). We assume in this paper that not all of the user channel conditions are known at every slot. Nevertheless, it is assumed that the scheduler knows for each user k whether she is in the system (in either good or bad condition) or not (the service has already been completed). If in the system and scheduled, then the scheduler must decide whether to transmit the data at rate $s_{k,G}$ or $s_{k,B}$.

There are several implementation variants of the feedback after the scheduled data are transmitted. In this paper we consider an automatic repeat query (ARQ) feedback, which is, in general, based on acknowledgements and timeouts. In particular, we assume that all scheduled users send back at the end of the slot their true condition (B, G) during the slot or whether they have received the last bit of the flow (*). However, no information is obtained from users that were not scheduled in the slot.

The decision of what rate to use for transmitting can also be made in several ways. However, incorporating rate adaptation into the problem creates an additional dimension and analysis very quickly becomes intractable. We therefore assume that this decision is taken so that the one-slot probability of completing the flow is maximized, given the current probability of being in the good condition.

III. POMDP Formulation

In this section we present a POMDP formulation of the discrete-time flow 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 allows for fitting the problem in the multi-armed restless bandit framework. This trick thus leads to designing a well-founded scheduling rule, which we then propose to be used in systems with arrivals. The restless bandits approach to POMDP problems appeared also in Jacko and Niño-Mora (2008); Niño-Mora and Villar (2011).

Consider K flows labeled by $k \in \mathcal{K}$ waiting for service at a base station that can serve M flows 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 flow scheduling as described in Jacko (2010b).

Recall that we consider the time slotted into epochs $t \in \mathcal{T}$ at which decisions can be made.

A. Flows, 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, if transmitted at rate $s_{k,n}$, then the probability that the service of flow k is completed within one period if being served is $\mu_{k,n}$. Note that we have $0 \leq \mu_{k,\mathrm{B}} \leq \mu_{k,\mathrm{G}} \leq 1$.

Letting $\theta_k := \mu_{k,B}/\mu_{k,G}$, the base station decides to transmit at rate $s_{k,G}$ if a scheduled user k is in belief state (probability of being in good condition) $x > \theta_k$, and at rate $s_{k,B}$ if a scheduled user k is in belief state $x \leq \theta_k$. In addition, if user k is scheduled when being in belief state x, then at the end of the same period the base station receives feedback

$$o_{k,x} := \begin{cases} G, & \text{with probability } (1-\mu_{k,B})x, \text{ if } x \leq \theta_k; \\ B, & \text{with probability } (1-\mu_{k,B})(1-x), \text{ if } x \leq \theta_k; \\ *, & \text{with probability } \mu_{k,B}, \text{ if } x \leq \theta_k; \\ G, & \text{with probability } (1-\mu_{k,G})x, \text{ if } x > \theta_k; \\ B, & \text{with probability } (1-x), \text{ if } x > \theta_k; \\ *, & \text{with probability } x \cdot \mu_{k,G}, \text{ if } x > \theta_k; \end{cases}$$

These expressions should be obvious to the reader, just note in the case $x > \theta_k$ (transmitting at rate $s_{k,G}$) that if the channel condition is B, then the data is not decoded by the user, and therefore the user cannot depart, so the probability to receive B as feedback is simply the probability of being in state B, which is $(1-x)$.

Each flow-channel-user triple k is defined independently of other flow-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

- $\mathcal{N}_k := \{*\} \cup \mathcal{X}_k$ is the *state space*, where state $*$ represents a flow already completed, and $\mathcal{X}_k := [0, 1]$ is the set of *belief states* (posterior probabilities) that channel k is in condition G provided the flow 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 $x \in \mathcal{X}_k$,

$$R_{k,*}^a := 0, \qquad R_{k,x}^0 := -c_k,$$
$$R_{k,x}^1 := -c_k \cdot (1 - \mathbb{P}[o_{k,x} = *]);$$

- $\boldsymbol{P}_k^a := \left(p_{k,n,m}^a\right)_{n,m\in\mathcal{N}_k}$ is the user-k stationary one-period *state-transition probability matrix* (note that given an initial belief state, the number of accessible belief states is countable, as seen below) 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, for any $x \in \mathcal{X}_k$,

$$p_{k,*,*}^a := 1, \quad p_{k,x,y}^0 := 1, \text{ if } y = xq_{k,G,G} + (1-x)q_{k,B,G},$$

and

$$p_{k,x,y}^1 := \begin{cases} \mathbb{P}[o_{k,x} = G], & \text{if } y = q_{k,G,G}; \\ \mathbb{P}[o_{k,x} = B], & \text{if } y = q_{k,B,G}; \\ \mathbb{P}[o_{k,x} = *], & \text{if } y = *; \end{cases}$$

The dynamics of user k is thus captured by the *state process* $N_k(\cdot)$ and the *action process* $a_k(\cdot)$, which correspond to state $N_k(t) \in \mathcal{N}_k$ 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 $N_k(t)$ at time epoch t, the user k consumes the allocated capacity, earns the reward, and evolves its state for the time epoch $t+1$. If $N_k(t) \in \mathcal{X}_k$ (i.e., the flow k is uncompleted by time t), then the belief state evolution can be summarized as follows

$$N_k(t+1) = \begin{cases} N_k(t)\, q_{k,G,G} + (1 - N_k(t))q_{k,B,G}, & \\ \qquad\qquad \text{w.p. } 1, \text{ if } a_k(t) = 0; & \\ q_{k,G,G}, \quad \text{w.p. } \mathbb{P}[o_{k,x} = G], \text{ if } a_k(t) = 1; & \\ q_{k,B,G}, \quad \text{w.p. } \mathbb{P}[o_{k,x} = B], \text{ if } a_k(t) = 1; & \\ *, \quad \text{w.p. } \mathbb{P}[o_{k,x} = *], \text{ if } a_k(t) = 1; & \end{cases}$$

B. 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 $\boldsymbol{a}(\cdot) := (a_k(\cdot))_{k\in\mathcal{K}}$, i.e., $\Pi_{\boldsymbol{X},\boldsymbol{a}}$ is the *joint policy space*. Let $\mathbb{E}_\tau^{\boldsymbol{\pi}}$ denote the expectation over the state process $\boldsymbol{X}(\cdot)$ and over the action process $\boldsymbol{a}(\cdot)$, conditioned on the state-process history $\boldsymbol{X}(0), \boldsymbol{X}(1), \ldots, \boldsymbol{X}(\tau)$ and on policy $\boldsymbol{\pi}$.

The problem is to find a joint policy $\boldsymbol{\pi}$ maximizing the *objective* given by the time-average aggregate reward starting from the initial time epoch 0 subject to the family of *sample path* allocation constraints, i.e.,

$$\max_{\boldsymbol{\pi}\in\Pi_{\boldsymbol{X},\boldsymbol{a}}} \lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}_0^{\pi}\left[\sum_{k\in\mathcal{K}} R_{k,X_k(t)}^{a_k(t)}\right] \tag{P}$$
$$\text{subject to } \mathbb{E}_t^{\boldsymbol{\pi}}\left[\sum_{k\in\mathcal{K}} a_k(t)\right] = M, \text{ for all } t \in \mathcal{T}$$

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

IV. Solution

Problem (P) can be relaxed by requiring to serve M flows per slot only *on time-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. 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

$$\max_{\widetilde{\pi}_k\in\Pi_{\boldsymbol{X},a_k}} \lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}_0^{\widetilde{\pi}_k}\left[R_{k,X_k(t)}^{a_k(t)} - \nu W_{k,X_k(t)}^{a_k(t)}\right] \tag{1}$$

The main idea of our approach is to identify a set of optimal policies $\widetilde{\pi}_k^*$ for (1) for each $k \in \mathcal{K}$, and using them to construct a joint *heuristic* policy $\boldsymbol{\pi}$, feasible though not necessarily optimal for problem (P).

A. Optimal Solution to Single-User Subproblem

In certain cases, problem (1) can be optimally solved by assigning a set of index values $\nu_{k,n}$ to each state $n \in \mathcal{N}_k$ (Niño-Mora, 2007; Jacko, 2010a). If this is the case, the problem is called *indexable*. In the following we conjecture that (1) is indexable and characterize the index values.

Given the real-world system by which the current model is motivated, it is natural to assume that $q_{k,G,G} - q_{k,B,G} > 0$ for all k. A channel with this property is often called the *positively autocorrelated* channel, or channel with (positive) memory. Note that in the iid (or uncorrelated) case ($q_{k,B,G} = q_{k,G,G}$), the belief state never changes (the channel has "no memory"). We will therefore present here an approach under this assumption, i.e., $\rho_k := q_{k,G,G} - q_{k,B,G} > 0$. Other cases can be treated analogously, and are omitted due to space restrictions.

Note that the interval $[q_{k,B,G}; q_{k,G,G}]$ can be conveniently written as $[q_{k,B,G}; q_{k,B,G} + \rho_k]$. It is straightforward to prove that interval $[q_{k,B,G}; q_{k,B,G} + \rho_k]$ is an absorbing subset of states as long as the flow is uncompleted.

Proposition 1: Under the assumption of positive autocorrelation, if the user belief state $X_k(t) \in [q_{k,B,G}; q_{k,B,G} + \rho_k]$, then the user belief state $X_k(t+1)$ representing the posterior probability that the channel is in good condition, also belongs in the interval $[q_{k,B,G}; q_{k,B,G} + \rho_k]$.

This result is helpful, because if the initial state of any flow belongs to this interval, then its belief state never leaves it unless it is completed. We therefore define a reduced state space $\widetilde{\mathcal{N}}_k := \{*\} \cup \widetilde{\mathcal{X}}_k$, where state $*$ represents a flow already completed, and $\widetilde{\mathcal{X}}_k := [q_{k,B,G}; q_{k,B,G} + \rho_k]$. Therefore, in the case in which the initial state vector is unknown, it is enough to obtain index values for belief states in $\widetilde{\mathcal{X}}_k$, given that a reasonable estimate for it is its steady state value, which lies within this interval. We believe the following statement be true.

Conjecture 1 (Indexability): Under the assumption of positive autocorrelation, problem (1) of user k is *indexable*, i.e., there exist unique values $-\infty \leq \nu_{k,n} \leq \infty$ for all $n \in \widetilde{\mathcal{N}}_k$ such that the following holds for every state $n \in \widetilde{\mathcal{N}}_k$:

(i) if $\nu_{k,n} \geq \nu$, then it is optimal to serve flow k in state n, and

(ii) if $\nu_{k,n} \leq \nu$, then it is optimal not to serve flow 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*. An intuitive notion of indexability was introduced in Whittle (1988), who realized that not always such an index-based solution exists. The definition given here is more general and follows Jacko (2010b).

In the following we shall use the notation to represent the passive (i.e., if not serving) dynamics

$$\phi_k(x) := q_{k,B,G} + \rho_k x. \tag{2}$$

As a consequence the fixed point of the passive dynamics (i.e., the belief state of a flow that is never served) is also in the interval $[q_{k,B,G}; q_{k,G,G}]$. In fact, such a fixed point is nothing but the steady-state probability of being in the good condition under matrix $\boldsymbol{Q}_k$,

$$q_{k,G}^{\text{SS}} = \frac{q_{k,B,G}}{1 + q_{k,B,G} - q_{k,G,G}} = \frac{q_{k,B,G}}{1 - \rho_k}, \tag{3}$$

assuming $q_{k,B,G} > 0$ so that the steady-state distribution exists and is positive for condition G.

We present the solution for problem (1) under the time-average criterion, which is in fact obtained in the limit from the solution under the discounted criterion, which however requires more complicated formulae and is less relevant in practice (omitted here due to space restrictions). Moreover, we restrict our attention only to the case $q_{k,B,G} > \theta$ due to space restrictions (otherwise there are more cases to deal with). This covers the self-interesting case of ON-OFF channels ($\theta_k = 0$) and the cases when $\mu_{k,B}$ is small enough with respect to $\mu_{k,G}$, in particular, if $\mu_{k,B} < q_{k,B,G}\mu_{k,G}$.

We will need some more notation. In case $q_{k,B,G} \leq x < q_{k,G}^{\text{SS}}$, let T_k (as a function of x) be an integer such that $\phi_k^{T_k}(q_{k,B,G}) > x$ and $\phi_k^{T_k - 1}(q_{k,B,G}) \leq x$, where $\phi_k^t(y)$ is the t-th functional power of $\phi_k(y)$. Such a T_k exists, is unique and positive, because (2) implies that $\phi_k^t(q_{k,B,G}) = q_{k,B,G}\frac{1-\rho_k^{t+1}}{1-\rho_k}$, which is an increasing function of t with limit $q_{k,G}^{\text{SS}} > x$. We remark that T_k can be interpreted as the *starvation age* (i.e., the number of periods without serving) of the flow in the system since the last B feedback (due to which the belief state was set to $q_{k,B,G}$).

Let further $\phi_{k,G}^*$ be a weighted harmonic mean of the one-slot probability of moving to the good condition of a flow with starvation age T_k and the steady-state probability of being in good condition under such dynamics

$$\phi_{k,G}^* := \frac{1}{\dfrac{\mu_{k,G}}{\phi_k^{T_k}(q_{k,B,G})} + \dfrac{1 - \mu_{k,G}}{\phi_{k,G}^{\text{SS}}}}, \tag{4}$$

where $\phi_{k,G}^{\text{SS}}$ is the steady-state probability for condition G of matrix $\boldsymbol{Q}_k$ with $\phi_k^{T_k}(q_{k,B,G})$ instead of $q_{k,B,G}$, and can be easily shown to be

$$\phi_{k,G}^{\text{SS}} = \frac{\phi_k^{T_k}(q_{k,B,G})}{1 + \phi_k^{T_k}(q_{k,B,G}) - q_{k,G,G}}. \tag{5}$$

We are now ready to characterize the index values in closed form, which is the main theoretical result of this paper.

Theorem 4.1: Under Conjecture 1 (in case $\mu_B < q_{k,B,G}\mu_G$), the index values under the time-average criterion are as follows. For a belief state $x \in \widetilde{\mathcal{X}}_k$,

$$\nu_{k,x} = \begin{cases} +\infty, & \text{if } q_{k,G}^{\text{SS}} \leq x \leq q_{k,G,G} \\ c_k \left\{ \phi_{k,G}^* \dfrac{1 - (1 - \mu_{k,G}) q_{k,G,G}}{(1 - \rho_k)(q_{k,G}^{\text{SS}} - x)[1 - (1 - \mu_{k,G})\phi_{k,G}^*]} \right. \\ \quad \left. + (T_k(x) + 1) \left[\dfrac{(1 - \mu_{k,G}) q_{k,G,G}}{1 - (1 - \mu_{k,G})\phi_{k,G}^*} - 1 \right] \right\}, \\ & \text{if } q_{k,B,G} \leq x < q_{k,G}^{\text{SS}} \end{cases}$$

and $\nu_{k,*} = 0$.

The index values possess the following properties.

Proposition 2: The index value $\nu_{k,q_{k,B,G}}$ is non-negative.

Proposition 3: The index values $\nu_{k,x}$ are increasing in x, and $\nu_{k,x} \to +\infty$ as $x \to q_{k,G}^{\text{SS}}$.

Because of this monotonicity of the index values we moreover obtain that policies of threshold type are optimal.

Proposition 4 (Optimality of Threshold Policies): Under Conjecture 1 (in case $\mu_B < q_{k,B,G}\mu_G$), for every real-valued ν there exists $z \in \widetilde{\mathcal{X}}_k$ such that threshold policy serving in states $\mathcal{S}_z := \{x \in \widetilde{\mathcal{X}}_k : x > z\}$ and not serving otherwise is optimal for problem (1).

B. Optimal Solution to Lagrangian Relaxation

The vector of policies $\boldsymbol{\pi}^* := (\widetilde{\pi}_k^*)_{k \in \mathcal{K}}$ identified in Conjecture 1 and Theorem 4.1 is formed by mutually independent single-user optimal policies, therefore this vector is an optimal policy to the Lagrangian relaxation of the original problem.

Note that for a given value of ν, this policy essentially means to serve everyone whose last feedback was G (with infinite index value), and moreover those whose last feedback was B and have been starving long enough (with respect to their parameters). Everyone will eventually be served because the index value during starvation increases to infinity.

V. New Opportunistic Schedulers

A. Whittle-Index-Based Scheduler

Since the original scheduling problem requires to allocate the base station to exactly M flows at every slot t, Whittle (1988) proposed to employ at every slot action 1 for the M flows with highest actual index values $\nu_{k,N_k(t)}$, and conjectured that such a heuristic be asymptotically optimal as both the number of served flows M and the overall number of flows K grow to infinity in a fixed proportion. This was shown true under certain technical conditions in Weber and Weiss (1990). In this section we extend these ideas to design a scheduler for the problem with arrivals.

First, note that for all the flows available at time slot $t = 0$, and for all the flows arriving later, there is a natural initial state: the steady-state $q^{\mathrm{SS}}_{k,G}$. This is because of the assumption that information is only received after the flow has been scheduled, and the initial belief state should be the one of not having been scheduled for an infinitely long time, which is indeed $q^{\mathrm{SS}}_{k,G}$. The index value of any newly arrived flow remains $+\infty$ until it is scheduled for the first time.

Further, since there may be too many flows with index value $+\infty$, we will need a tie-breaking rule to set priorities among such flows. Following Ayesta et al. (2010), we choose among such flows the one with highest second-order term of the Laurent expansion of the discounted index value, which, for $q^{\mathrm{SS}}_{G} \le x \le q_{G,G}$, is

$$\nu^{(2)}_{k,x} = \frac{c_k \mu_{k,G} x}{1 - (1 - \mu_{k,G})(q_{k,G,G} - x)},$$

It is easy to prove that such a quantity is increasing.

Proposition 5: The tie-breaking index value $\nu^{(2)}_{k,x}$ is increasing in x.

Finally, based on the above arguments and results under the time-average criterion presented in the previous section, we propose a new scheduler which serves users as follows: at every slot t,

(i) serve up to M flows whose last observed condition was G or have never been served, giving higher priority to those with higher value $\nu^{(2)}_{k,x}$;

(ii) if there are less than M flows whose last observed condition was G or have never been served, then in addition to those, serve also flows whose last observed condition was B, giving higher priority to those with higher value $\nu_{k,x}$;

In addition, break ties arbitrarily.

The proposed opportunistic scheduler gives absolute priority to flows whose last observed channel condition was G. Flows whose last observed channel condition was B are served only if there are not enough absolute priority flows. This may be perhaps surprising, but it is in the same vein as the results in Jacko (2011) for a fully observable system (where channel conditions are known for all the users at the beginning of every slot), where Whittle index opportunistic scheduler gives absolute priority to users in G over those in B, and schedulers with such a property were proved maximally stable in Ayesta et al. (2011).

In addition to maximal stability in the fully-observable system, the use of $c_k \mu_{k,G}$ for tie-breaking was shown to lead to fluid-optimality in Ayesta et al. (2011). In our partially-observable model the expression is a bit more involved, but with an interesting interpretation of the *expected* departure probability $c_k \mu_{k,G} x$ divided by the unit complement of the *potential improvement of the belief state*, since $q_{k,G,G} - x$ is the belief improvement due to feedback, multiplied by the probability of not departing from the system $1 - \mu_{k,G}$. In fact, the tie-breaking value of a flow that was served in the previous slot and obtained feedback G is $c_k \mu_{k,G} x = c_k \mu_{k,G} q_{k,G,G}$, which is nothing but the expected reduction in the holding costs.

B. Scheduler for Single-Class System

In case of flows belonging all to the same class (i.e., having the same all the parameters and differing only by the current state), the scheduler simplifies significantly. Suppose that we start from initial conditions as described earlier in this section.

Recall that the tie-breaking in (i) implies that the users whose last observed state was G are served according to shortest-starved-first, due to Proposition 5. Then the proposed scheduler will have the *stay-on-a-winner* property, in the sense that every time a flow is scheduled and observes condition G, then it is served also in the next slot, because there cannot ever be more than M flows in such a situation. This is an interesting and desirable property of our scheduler, which essentially means to be *opportunistic*. As a consequence, it can never happen that a flow whose last observed state was G has been starving at all. Moreover, due to Proposition 3 the flows whose last observed state was B are served according to longest-starved-first.

The scheduler for single-class systems can be written algorithmically as follows: at every slot t,

(i) serve all the flows that were scheduled in the previous slot and whose observed condition was G;

(ii) if there are less than M flows whose observed condition in the previous slot was G, then in addition to those, serve also (in total up to M) flows that have never been served;

(iii) if there are less than M flows whose observed condition in the previous slot was G or have never been served, then in addition to those, serve also (in total up to M) flows whose last observed condition was B, giving higher priority to those that have been starving longer;

In addition, break ties arbitrarily.

C. Naïve Schedulers

It is further natural and interesting from the implementation point of view to study much simpler schedulers, such as those we define next.

a) Myopic Scheduler: The myopic scheduler serves at every slot M flows with highest immediate cost reduction (breaking ties arbitrarily), defined by

$$\nu_{k,x}^{\text{myopic}} := c_k \mu_{k,G} x.$$

The myopic scheduler is also of theoretical importance, since it becomes optimal in the multi-class setting if feedback brings no relevant information (i.e., when $\rho_k = 0$ and so $x = q_{k,G}^{\text{SS}}$ always). Notice that in that case, the departure probability of a class-k flow is $\mu_{k,G} q_{k,G}^{\text{SS}}$ and so it recovers the classical job scheduling setting, in which the $c\mu$-rule is optimal (where μ is the departure probability) (Buyukkoc et al., 1985).

Proposition 6: The myopic scheduler is optimal under arbitrary arrivals if $\rho_k = 0$ for all flow classes k.

We suspect that the suboptimality decreases as ρ_k (for a fixed k) approaches zero, because this value is close to the myopic index if $\rho_k \approx 0$. Therefore, we will be interested in studying its performance especially when ρ_k is relatively large.

We will also be interested in the following scheduler, which is independent of all the problem parameters except for $\boldsymbol{Q}_k$.

b) Belief-State Scheduler: The belief-state scheduler serves at every slot M flows with highest immediate probability of being in G (breaking ties arbitrarily), defined by

$$\nu_{k,x}^{\text{belief}} := x,$$

Note finally that both the myopic scheduler and the belief-state scheduler become *equivalent* to the Whittle-index-based scheduler in the single-class case. This is an interesting feature observed also in packet-level models.

VI. Experimental Study

In order to compare the performance of the three schedulers, we have performed simulations in a variety of scenarios with heterogeneous flows. We would like to emphasize that in a vast majority of scenarios we have tested, the three schedulers performed equivalently (the obtained objective values did not differ significantly at 95% level of confidence). In the two scenarios reported below the belief-state scheduler is not significantly different from the Whittle-index-based scheduler, therefore the belief-state scheduler is omitted in the figures. For comparison we however include a scheduler that at every slot randomly chooses which uncompleted flow to serve.

See Figure 1(a) for illustration of such an effect in scenario 1 with the following parameters: $K = 2, M = 1, c_k = 1, q_{1,G,G} = 0.9, q_{1,B,G} = 0.1, q_{2,G,G} = 0.8, q_{2,B,G} = 0.2, \mu_{2,G} = 0.08$, and varying $\mu_{1,G}$. Thus, decreasing the expected size of one flow significantly decreases the total expected holding cost (since the other flow is much smaller), improving over the random scheduler around 10%.

It is well known in channel-aware systems that myopic schedulers are suboptimal if the class parameters are such that one class persistently gets absolute priority over some other class. In such a situation, the users of the latter may be delayed service and accumulate, which may lead to unstability. Such a situation is reported in Scenario 2 Figure 1(b) with the following parameters: $K = 2, M = 1, c_k = 1, q_{2,G,G} = 0.7, q_{2,B,G} = 0.2, \mu_{1,G} = 0.02, \mu_{2,G} = 0.08$, and varying $q_{1,G,G}$ and $q_{1,B,G}$ so that the steady-state probability remains the same, $q_{1,G}^{\text{SS}} = 0.4$. Indeed, we can observe a statistically significant superiority of the Whittle-index-based scheduler over the myopic scheduler, of $2 - 6\%$ in the mean relative performance, and a notable improvement over the random scheduler of around 15%.

Note that these differences are relatively big, taking into account the simple scenarios of only two flows presented here. We expect to see even larger differences in simulations with arrivals (which will be reported in the full paper version).

VII. Conclusion

The results of this paper provide us with several novel conclusions. It is interesting that there are many states with infinite-valued index; to the best of our knowledge, this is first such result. Another striking finding is the exceptional performance of the belief-state scheduler, which is moreover independent of the job size parameters. Such a policy was proposed also in Niño-Mora and Villar (2011), but usually has been ignored by the researchers.

On the other hand, it is desirable to study optimality of the proposed schedulers. We may be able to prove optimality in the single-class setting, following the similar ideas as in Liu et al. (2011). However, general optimality of the Whittle-index-based scheduler in the multi-class setting is unlikely.

VIII. Acknowledgements

The authors would like to thank to Urtzi Ayesta, Sem Borst, Vincenzo Mancuso, Richard Weber for fruitful and encouraging discussions on this topic, and the four anonymous referees for their suggestions. Research partially supported by grant MTM2010-17405 (Ministerio de Ciencia e Innovación, Spain) and grant PI2010-2 (Department of Education and Research, Basque Government).

References

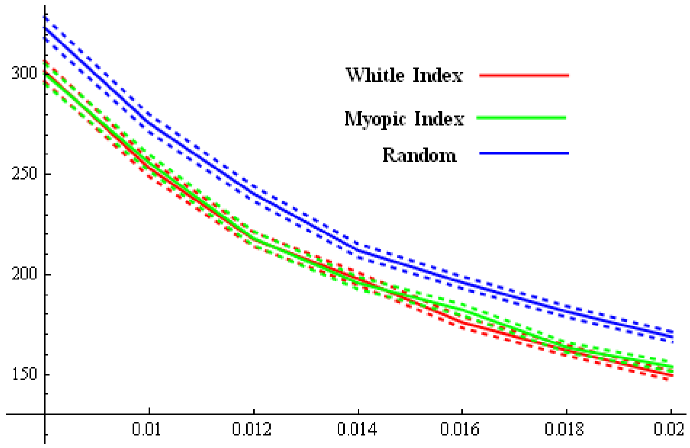
Aalto, S. and Lassila, P. (2010). Flow-level stability and performance of channel-aware priority-based schedulers. In *Proceeding of NGI 2010 (6th EURO-NF Conference on Next Generation Internet)*.

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

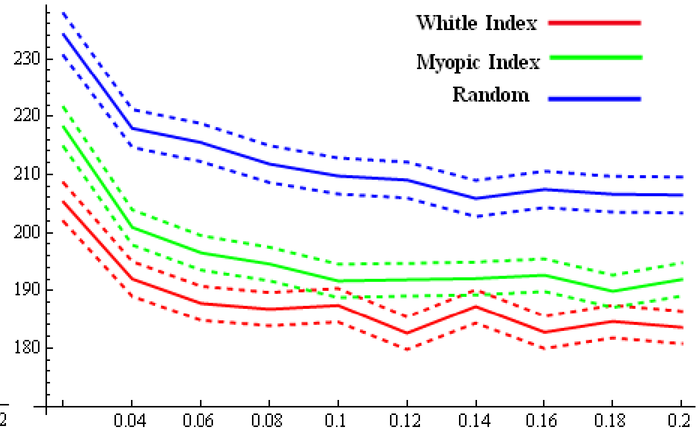
Ahmad, S. H. A., Liu, M., Javidi, T., Zhao, Q., and Krishnamachari, B. (2009). Optimality of myopic sensing in multichannel opportunistic access. *IEEE Transactions on Information Theory*, 55(9):4040–4050.

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: Stability and asymptotic optimality. arXiv:1101.5794v1.



(a) Scenario 1: $\mu_{1,G} = [0.008, 0.01, 0.012, 0.014, 0.016, 0.018, 0.02]$.



(b) Scenario 2: $q_{1,B,G} = [0.02, 0.04, 0.06, 0.08, 0.1, 0.12, 0.14, 0.16, 0.18, 0.2]$.

Fig. 1. Performance in terms of the total expected holding cost of three schedulers in simulations. Each scheduler is represented by the mean and the 95% confidence interval obtained from 10^4 simulation runs.

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., 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.

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

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

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.

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

Jacko, P. and Niño-Mora, J. (2008). Marginal productivity index policies for problems of admission control and routing to parallel queues with delay. Universidad Carlos III de Madrid, Working paper 08-72.

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.

Liu, K., Weber, R., and Zhao, Q. (2011). Indexability and Whittle index for restless bandit problems involving reset processes. In *Proceedings of CDC 2011*.

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

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

Niño-Mora, J. (2008). An index policy for dynamic fading-channel allocation to heterogeneous mobile users with partial observations. In *Next Generation Internet (NGI) Networks*, pages 231 – 238.

Niño-Mora, J. and Villar, S. S. (2011). Sensor scheduling for hunting elusive hiding targets via Whittle's restless bandit index policy. In *5th International Conference on Network Games, Control and Optimization (NetGCoOp), 2011*.

Ouyang, W., Eryilmaz, A., and Shroff, N. B. (2011a). Asymptotically optimal downlink scheduling over Markovian fading channels. *arXiv:1108.3768v2 [cs.NI]*.

Ouyang, W., Murugesan, S., Eryilmaz, A., and Shroff, N. B. (2011b). Exploiting channel memory for joint estimation and scheduling in downlink networks. *arXiv:1009.3959v6*.

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

Wang, K., Chen, L., Liu, Q., and Al Agha, K. (2012). On optimality of myopic sensing policy with imperfect sensing in multi-channel opportunistic access. *arXiv:1202.0477v1*.

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.

Zhao, Q., Krishnamachari, B., and Liu, K. (2008). On myopic sensing for multi-channel opportunistic access: Structure, optimality, and performance. *IEEE Transactions on Wireless Communications*, 7(12):5431–5440.