Dynamic Control of Birth-and-Death Restless Bandits: Application to Resource-Allocation Problems

Maialen Larrañaga, Urtzi Ayesta, and Ina Maria Verloop

***Abstract*— We develop a unifying framework to obtain efficient index policies for restless multi-armed bandit problems with birth-and-death state evolution. This is a broad class of stochastic resource allocation problems whose objective is to determine efficient policies to share resources among competing projects. In a seminal work, Whittle developed a methodology to derive well-performing (Whittle's) index policies that are obtained by solving a relaxed version of the original problem. Our first main contribution is the derivation of a closed-form expression for Whittle's index as a function of the steady-state probabilities. In some particular cases, qualitative insights can be obtained from its expression; nevertheless, it requires several technical conditions to be verified. We, therefore, formulate a fluid version of the relaxed optimization problem, and in our second main contribution, we develop a fluid index policy. The latter does provide qualitative insights and it is equivalent to Whittle's index policy in the light-traffic regime. The applicability of our approach is illustrated by two important problems: optimal class selection and optimal load balancing. Allowing state-dependent capacities, we can model important phenomena, e.g., power-aware server-farms and opportunistic scheduling in wireless systems. Whittle's index and our fluid index policy show remarkably good performance in numerical simulations.**

***Index Terms*— Whittle's index, restless bandits, birth-and-death processes, fluid approximation, Lagrangian relaxation, index policies.**

Manuscript received July 30, 2015; accepted March 28, 2016; approved by IEEE/ACM TRANSACTIONS ON NETWORKING Editor A. Eryilmaz. Date of publication June 30, 2016; date of current version December 15, 2016. This work was supported by the Agence Nationale de la Recherche through the Project ANR JCJC RACON. The work of M. Larrañaga was supported by a research grant within the Foundation Airbus Group. A shorter version of this paper was presented at the IEEE INFOCOM 2015 [1].

M. Larrañaga is with the Centre National de la Recherche Scientifique (CNRS), Laboratoire des Signaux et Systèmes, CentraleSupélec, Gif-sur-Yvette 91190, France (e-mail: maialen.larranaga@supelec.fr).

U. Ayesta is with the Centre National de la Recherche Scientifique, Laboratory for Analysis and Architecture of Systems, 31400 Toulouse, France, and also with IKERBASQUE, Basque Foundation for Science, 48011 Bilbao, Spain; Universidad del Pais Vasco, Euskal Herriko Unibertsitatea (UPV/EHU), University of the Basque Country, 20018 Donostia, Spain; and the Laboratory for Analysis and Architecture of Systems, Institut National Polytechnique, Université de Toulouse, 31400 Toulouse, France (e-mail: urtzi@laas.fr).

I. M. Verloop is with the Centre National de la Recherche Scientifique, Institut de Recherche en Informatique de Toulouse, 31071 Toulouse, France, and also with the Laboratory for Analysis and Architecture of Systems, Institut National Polytechnique, Université de Toulouse, 31400 Toulouse, France (e-mail: verloop@irit.fr).

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org, provided by the authors. It consists of an appendix containing most of the proofs of this paper in a 4.3-MB PDF.

Digital Object Identifier 10.1109/TNET.2016.2562564

I. INTRODUCTION

OUR OBJECTIVE is to develop a unifying framework to obtain well performing policies for stochastic resource allocation problems. The model we consider is rather general, and aims at capturing the fundamental decision making problem arising in resource allocation problems among competing projects. Two broad classes of problems that fall inside our framework are that of optimal class selection and optimal load balancing for heterogeneous servers. We allow both state-dependent arrivals and state-dependent capacities. The latter can model important phenomena such as speed scaling in power-aware systems or fading in wireless channels, where the capacity scales with the number of users.

An optimal policy will in general be a complex function of all the input parameters and the number of competing projects. In practice such problems can be solved only for very specific instances. In some cases, a so-called *index policy* is optimal, that is, the solution to the stochastic control problem is characterized by an *index*, which depends on the state of the project, that determines which action is optimal to take. Optimality of index policies has enjoyed a great popularity. The solution to a complex control problem that, a priori, might depend on the entire state space, turns out to have a strikingly simple structure. A classical example is a multi-class single-server queue with linear holding costs where it is known that the celebrated $c\mu$-rule is optimal with respect to average holding cost, that is, to serve the classes in decreasing order of priority according to the product $c_k\mu_k$, where c_k is the holding cost per class-k customer, and μ_k^{-1} is the mean service requirement of class-k customers, [2]. The simple structure of the optimal policy vanishes however in the presence of, e.g., convex costs, servers with state-dependent capacities and/or impatient customers [3]–[6]. Another classical result that can be seen as an index policy is the optimality of Shortest-Remaining-Processing-Time (SRPT), where the index of each customer is given by its remaining service time [7].

Both examples fit the general context of Multi-Armed Bandit Problems (MABP). A MABP is a particular case of a Markov Decision Process: at every decision epoch the scheduler needs to select one *bandit*, and an associated reward is accrued. The state of this selected bandit evolves stochastically, while the state of all other bandits remains *frozen*. The scheduler knows the state of all bandits, the rewards in every state, and the transition probabilities, and aims at maximizing

the total average reward. In a ground-breaking result Gittins showed that the optimal policy that solves a MABP is an index rule, nowadays commonly referred to as Gittins' index policy [8]. Thus, for each bandit, one calculates Gittins' index, which depends only on its own current state and stochastic evolution. The optimal policy activates in each decision epoch the bandit with highest current index.

Despite its generality, in multiple cases of practical interest the problem cannot be cast as a MABP. In a seminal work [9], Whittle introduced the so-called Restless Bandit Problem (RBP), a generalization of the standard MABP. In a RBP all bandits in the system incur a cost. The scheduler selects a number of bandits to be made active, and all bandits might evolve over time according to a stochastic kernel that depends on whether the bandit is made active. The objective is to determine a control policy that optimizes the average performance criterion. RBP provides a powerful modeling framework, but its solution has in general a complex structure that might depend on the entire state-space description. Whittle considered a relaxed version of the problem (where the restriction on the number of *active* bandits needs to be respected on average only, and not in every decision epoch), and showed that the solution to the relaxed problem is of index type, referred to as *Whittle's index*. Whittle then defined a heuristic for the original problem, referred to as Whittle's index policy, where in every decision epoch the bandit with highest Whittle index is selected. It has been shown that the Whittle index policy performs strikingly well, see [10] for a discussion, and is asymptotically optimal under certain conditions, see [11], [12]. The latter explains the importance given in the literature to the calculation of Whittle's index. In addition to resource allocation problems, Whittle's index has been applied in a wide variety of cases, including website morphing and pharmaceutical trials, [8, Ch. 6]. In the last years many researchers have applied Whittle's index approach for the opportunistic scheduling in wireless networks, see [13], [14]. The recent survey paper [15] is a good reference on the application of index policies in scheduling.

In order to calculate Whittle's index there are two main difficulties: first, one needs to establish a technical property known as *indexability*, and second, the calculation of the Whittle index itself might be involved or even infeasible.

In this paper we focus on deriving efficient index policies for a RBP in the particular case where each bandit can be modeled as a birth-and-death stochastic process. The birth-and-death process is a special case of a continuous-time Markov process where the state transitions are of only two types: "births", which increase the state variable by one and "deaths", which decrease the state by one. Birth-and-death processes have many applications in demography, queueing theory, performance engineering, epidemiology and biology. We refer to [16] as one of the first works developing an index policy for a problem with birth-and-death dynamics.

In our first main contribution, we derive a sufficient condition for the indexability property to hold and we derive a closed-form expression for Whittle's index. We show that Whittle's index can be expressed as a function of the steady-state probabilities and it can thus numerically be calculated. In some particular cases this expression allows us to obtain qualitative insights, however, this is not the case in a general setting. We therefore formulate a fluid version of the relaxed optimization problem, where the objective is *bias optimality*, i.e., to determine the policy that minimizes the cost of bringing the fluid to its equilibrium. Our approach is motivated by the pioneering work where fluid control models were used to approximate stochastic optimization problems, see Avram *et al.* [17] and Weiss [18]. We give a closed-form expression for the fluid index, which provides full insights into the effect of the parameters. The advantage of the fluid approach lies in its relatively simple expressions compared to the stochastic one, and in the fact that one does not need to verify for indexability or optimality of threshold policies. In addition, it can be established that both policies are equivalent in a light-traffic regime.

We illustrate the applicability of our approach with two classes of resource-sharing problems: optimal class selection and optimal load balancing in heterogeneous servers. In both cases we allow for general holding cost functions and state-dependent capacities and arrivals. As representative examples we consider (i) scheduling in a multi-class opportunistic downlink channel, (ii) load balancing in a power-aware server farm and (iii) scheduling a make-to-stock queue with perishable items. Numerical experiments show that for all three examples the Whittle index policy and the fluid index policy are nearly optimal.

In summary the main contributions of this paper are:

- Unifying approach to obtain Whittle's index policy for birth-and-death bandits under average cost criterion.
- Development of a fluid-based approach to derive a novel index policy, based on the fluid index, yielding a simple closed-form expression.
- Study of two examples of practical interest: opportunistic scheduling in downlink channels and load balancing in power-aware server farms.

The paper is organized as follows. In Section II we present the birth-and-death restless bandit model and its optimization framework. In Section III we present the relaxation of the original problem and derive Whittle's index, and in Section IV we derive the fluid index. In Section V we prove the equivalence of Whittle's index and fluid index in a light-traffic regime. Finally, in Section VI the performance of Whittle's index policy and the fluid index policy is numerically evaluated.

II. Model Description and Preliminaries

We consider a stochastic resource allocation problem with K on-going projects or bandits. Let $N_k(t) \in \{0, 1, \ldots\}$ denote the state of bandit k at time t, $k = 1, \ldots, K$. Decision epochs are defined as the moments when a bandit changes state. At each decision epoch, the controller can choose for each bandit between two actions: action $a = 0$, that is, making the bandit passive, or action $a = 1$, that is, making the bandit active, with the restriction that at any moment in time at most $M < K$ bandits can be made active. Throughout this paper we consider bandits that are modeled as a continuous time birth-and-death process, that is, when bandit k is in state n_k, it changes the state after an exponentially distributed amount

of time, and can go either to state $(n_k - 1)^+$ or state $n_k + 1$. The transition rates for bandit k depend only on n_k (and not on the state of the other bandits). More precisely, when N_k denotes the state of bandit $k = 1, \ldots, K$, the transition rates of the vector $\vec{N} = (N_1, \ldots, N_K)$ are

$$\begin{cases} \vec{N} \to \vec{N} + \vec{e}_k & \text{with transition rate } b_k^a(N_k), \\ \vec{N} \to \vec{N} - \vec{e}_k & \text{with transition rate } d_k^a(N_k), \end{cases} \tag{1}$$

where $\vec{e}_k$ is a K-dimensional vector with all zeros except for the k-th component which is equal to 1, and $d_k^a(0) = 0$.

We note that the transitions of a bandit depend on the action chosen. In particular, the state of bandits can evolve both when being active and passive. In the literature this is commonly known as the RBP, which we briefly described in Section I, see [8]. We note that, given the action taken in state $\vec{N}$, the dynamics of each bandit is independent of the others, see (1).

A policy ϕ decides which bandit is made active. Because of the Markov property, we can focus on policies that base their decisions only on the current state of the bandits. For a given policy ϕ, $N_k^\phi(t)$ denotes the state of bandit k at time t and $\vec{N}^\phi(t) = (N_1^\phi(t), \ldots, N_K^\phi(t))$. Let $S_k^\phi(\vec{N}^\phi(t)) \in \{0, 1\}$ represent whether or not bandit k is made active at time t under policy ϕ. At most M out of the K bandits can be made active, or equivalently, at least $K - M$ bandits have to be passive. Hence, we have the constraint

$$\sum_{k=1}^{K} (1 - S_k^\phi(\vec{N})) \geq K - M. \tag{2}$$

Let us denote by $\mathcal{U}$ the set of policies that satisfy Equation (2) and make the system ergodic. Throughout the paper we assume that $b_k^a(\cdot), d_k^a(\cdot)$ are such that $\mathcal{U} \neq \emptyset$. For bandit k, let $C_k(n, a)$ denote the cost per unit of time when in state n and it is either passive (action $a = 0$) or active (action $a = 1$). We assume that $C_k(n, a)$ is bounded by a polynomial of finite degree.

A. Examples

Our main motivation to study this problem comes from resource allocation problems arising in multi-class multi-server environments. Assuming there are K classes of users, each class is represented by a bandit, and the state N_k of bandit k represents the number of class-k users in the system. Furthermore, $b_k^a(N_k)$ and $d_k^a(N_k)$ denote the arrival and departure rate, respectively. Having a state-dependent departure rate allows us to model important phenomena such as power-aware server farms and user impatience in which users may leave the system before finishing service. In the former the departure rate will be proportional to the speed-scaling term $(N_k)^\alpha$, see [19], and in the latter the departure rate will include a term $\theta_k N_k$, where θ_k is the abandonment rate of class-k users, see [20]–[22]. To illustrate the applicability of our framework, we now present two broad classes of problems that fall inside the framework presented. Both examples are further developed in Section VI.

The first class of problems concerns the multi-class setting of Figure 1. The objective is to determine which M classes

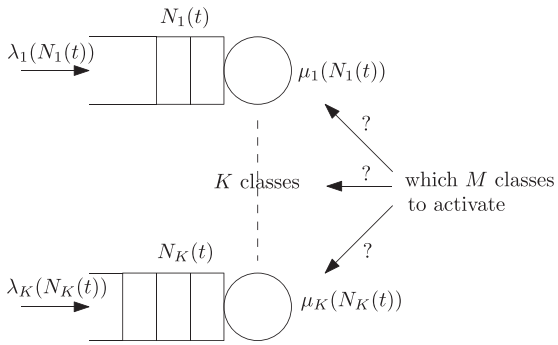

Fig. 1. A multi-class system where M classes can be simultaneously served.

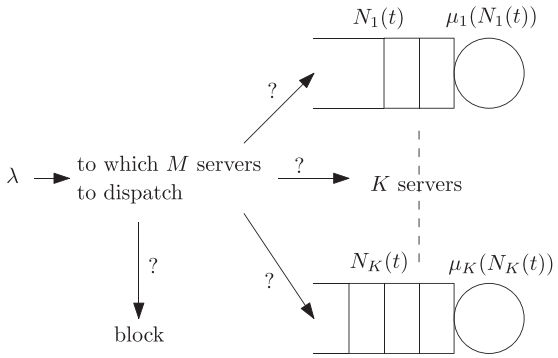

Fig. 2. Load balancing in a multi-server system.

to be made simultaneously active. Hence, the transition rates are as follows: $b_k^a(N_k) = \lambda_k(N_k)$ and $d_k^a(N_k) = \mu_k(N_k)a$, where $a = 1$ in case class k is served. We allow the arrival and departure rate of each class to depend on its queue length. In Section VI we use this model to study optimal scheduling in a wireless downlink problem where, as a consequence of opportunistic scheduling, the capacity increases with the number of users, see [23]. We further note that when $M = 1$ and $\mu_k(N_k) = \mu_k$, this model captures the classical single-server multi-class queue.

The second class of problems is the load balancing problem, see Figure 2, where new arrivals must be dispatched to K heterogeneous servers, or must be blocked. We allow an arrival to be dispatched to at most M servers (simultaneously), where $M = 1$ is the typical value for load-balancing problems. Hence, the transition rates are as follows: $b_k^a(N_k) = \lambda a$ and $d_k^a(N_k) = \mu_k(N_k)$, where $a = 1$ in case an arrival is routed to server k. In Section VI we investigate (i) how to optimally dispatch users in a power-aware server farm, where the capacity of servers follows a speed-scaling rule and (ii) how to optimally produce perishable items in a make-to-stock production system.

B. Optimal Control

The objective of this paper is to find scheduling policies $\phi \in \mathcal{U}$ that minimize the average-cost criteria

$$\mathcal{C}^\phi := \limsup_{T \to \infty} \sum_{k=1}^{K} \frac{1}{T} \mathbb{E}\left(\int_0^T C_k(N_k^\phi(t), S_k^\phi(\vec{N}^\phi(t))) \, \mathrm{d}t \right). \tag{3}$$

The above problem can be seen as a particular case of a Markov Decision Process (MDP), see Puterman [24] for a comprehensive treatment of MDP's. For problem (3), it is known that if there exist g and $V(\cdot)$ that satisfy the Dynamic Programming equation

$$g = \min_{\vec{s},s.t.\ \sum_k s_k \leq M} \left(\sum_{k=1}^{K} \Big[C_k(n_k, s_k) + b_k^{s_k}(n_k) V(\vec{n} + e_k) + d_k^{s_k}(n_k) V(\vec{n} - e_k) - (d_k^{s_k}(n_k) + b_k^{s_k}(n_k)) V(\vec{n}) \Big] \right), \tag{4}$$

a stationary policy that realizes the minimum in (4) is optimal, [24]. Here, $g = \min_\phi \mathcal{C}^\phi$ and $V(\vec{n})$ is the value function. The latter captures the difference in cost between starting in state $\vec{n}$ and an arbitrary reference state. In general, an optimal policy for (3) (or equivalently (4)) cannot be found, and structural results are only available for particular cases. Numerically, optimal policies can be found using Value Iteration or Policy Improvement algorithms. However, the curse of dimensionality renders infeasible to find a solution even for very small instances of the problem.

For certain examples it is possible to explicitly solve (4) and to characterize an optimal stochastic control. An important class of problems for which this is possible is known as the MABP problem, which we introduced in Section I. In this case only one bandit can be made active ($M = 1$) and only the active bandit can change state, that is, $b_k^0(n_k) = d_k^0(n_k) = 0$ and $b_k^1(n_k) \geq 0$, $d_k^1(n_k) \geq 0$. An optimal solution of (3) has a simple structure, known as Gittins' index policy, see [8]. In brief, there exist functions $G_k(n_k)$, depending only on the parameters of bandit k, such that an optimal policy in state $\vec{n}$ prescribes to serve the bandit having currently the highest index $G_k(n_k)$. However, for the restless bandit context $(b_k^0(n_k), d_k^0(n_k) \geq 0)$, as considered in this paper, finding optimal policies is typically out of reach. In the next section we will describe the methodology, introduced by Whittle [9], to derive approximate solutions to (3).

III. Lagrangian Relaxation and Whittle's Index Policy

The solution to (3) under constraint (2) cannot be solved in general. Following Whittle [9], a very fruitful approach has been to study the relaxed problem in which the constraint on the number of active bandits must be satisfied on *average*, and not in every decision epoch, that is,

$$\limsup_{T\to\infty} \frac{1}{T} \mathbb{E} \left(\int_0^T \sum_{k=1}^{K} (1 - S_k^\phi(\vec{N}^\phi(t))) \mathrm{d}t \right) \geq K - M. \tag{5}$$

Note that the set of policies that make the relaxed problem ergodic include $\mathcal{U} \neq \emptyset$, *i.e.*, the set of ergodic policies for the original problem. The objective of the relaxed problem is hence to determine a policy that solves (3) under constraint (5). An optimal policy for the relaxed problem, which turns out to be of index type, then serves as heuristic for the original optimization problem.

The relaxed problem can be solved by considering the following unconstrained problem: find a policy ϕ that minimizes

$$\limsup_{T\to\infty} \frac{1}{T} \mathbb{E} \left(\int_0^T \left(\sum_{k=1}^{K} C_k(N_k^\phi(t), S_k^\phi(\vec{N}^\phi(t))) + W(K - M - \sum_{k=1}^{K} (1 - S_k^\phi(\vec{N}^\phi(t)))) \right) \mathrm{d}t \right), \tag{6}$$

where W is the Lagrange multiplier. The key observation made by Whittle is that problem (6) can be decomposed into K subproblems, one for each different bandit k, that is, minimize:

$$\mathcal{C}_k^\phi := \limsup_{T\to\infty} \frac{1}{T} \mathbb{E} \left(\int_0^T \Big(C_k(N_k^\phi(t), S_k^\phi(N_k^\phi(t))) - W(1 - S_k^\phi(N_k^\phi(t))) \Big) \mathrm{d}t \right). \tag{7}$$

The solution to (6) is obtained by combining the solution to the K separate optimization problems (7). Under a stationarity assumption, we can invoke ergodicity to show that (7) is equivalent to minimizing

$$\mathbb{E}(C_k(N_k^\phi, S_k^\phi(N_k^\phi))) - W\ \mathbb{E}(\mathbf{1}_{S_k^\phi(N_k^\phi)=0}), \tag{8}$$

where N_k^ϕ is distributed as the stationary distribution of the state of bandit k under policy ϕ. Observe that the multiplier W can be interpreted as *subsidy* for passivity.

Problem (7) (equivalently (8)) is an MDP as well and an optimal policy is the solution of the Dynamic Programming equation

$$g_k = \min \Big(C_k(n,1) + b_k^1(n)\Delta V(n) - d_k^1(n)\Delta V(n-1), C_k(n,0) - W + b_k^0(n)\Delta V(n) - d_k^0(n)\Delta V(n-1) \Big), \tag{9}$$

with $g_k = \min_\phi \mathcal{C}_k^\phi$ the minimum cost under an optimal policy, and $\Delta V(n) = V(n+1) - V(n)$.

A. Indexability and Whittle's Index

Indexability is the property that allows us to develop a heuristic for the original problem. This property requires to establish that as the Lagrange multiplier, or equivalently the subsidy for passivity, W, increases, the collection of states in which the optimal action is *passive* increases. It was first introduced by Whittle [9] and we formalize it in the following definition.

Definition 1: A bandit is *indexable* if the set of states in which *passive* is an optimal action in (7) (denoted by $D_k(W)$) increases in W, that is, $W' < W \ \Rightarrow D_k(W') \subseteq D_k(W)$.

If indexability is satisfied, Whittle's index in state N_k is defined as follows:

Definition 2: When a bandit is indexable, *Whittle's index* in state N_k is defined as the smallest value for the subsidy such that actions passive and active are equally attractive for bandit k in state N_k. The Whittle index is denoted by $W_k(N_k)$.

Given that the indexability property holds, Whittle established in [9] that the solution to the relaxed control problem (6) will be to activate all bandits that are in a state n_k such

that their Whittle's index exceeds the subsidy for passivity, i.e., $W_k(n_k) > W$. In particular, a standard Lagrangian argument shows that there exists a value $W = W^*$, for which the constraint (5) is binding, i.e., an optimal policy ϕ that solves Problem (6) for $W = W^*$ will on average activate (at most) M bandits.

B. Threshold Policies

For certain problems, it can be established that the structure of an optimal solution of problem (7) is of threshold type. That is, optimality of a monotone policy can be shown: there is a threshold $n_k(W)$ such that when bandit k is in a state $n_k \leq n_k(W)$, then action a is optimal, and otherwise action a' is optimal, $a, a' \in \{0, 1\}$ and $a \neq a'$. In this section we describe the Whittle index in case an optimal structure is of threshold type. Before doing so, we further discuss optimality of threshold policies.

We let policy $\phi = n$ denote a threshold policy with threshold n, and we refer to it as 0-1 type if $a = 0$ and $a' = 1$, and 1-0 type if $a = 1$ and $a' = 0$. Optimality of a threshold policy for a relaxed optimization problem has been proved for example in [2], [20], and [21]. Further examples can be found in [8, Sec. 6.5]. In the next proposition we give sufficient conditions for an optimal policy that solves problem (7) to be of threshold type. The proof can be found in the supplementary.

Proposition 1: Assume

$$b_k^a(N_k) = \lambda_k(N_k), \quad \text{and } d_k^a(N_k) = \mu_k(N_k)a.$$

Then there exists an $n_k \in \{-1, 0, 1, \ldots\}$ such that a 0-1 type of threshold policy, with threshold n_k, optimally solves problem (7).

If instead,

$$b_k^a(N_k) = \lambda_k(N_k)a, \quad \text{and } d_k^a(N_k) = \mu_k(N_k),$$

then there exists an $n_k \in \{-1, 0, 1, \ldots\}$ such that a 1-0 type of threshold policy, with threshold n_k, optimally solves problem (7).

From the statement of Proposition 1 we see that an optimal policy that solves the example depicted in Figure 1 will be of 0-1 type, whereas an optimal policy that solves the example depicted in Figure 2 will be of 1-0 type.

The Whittle index is the smallest value of W such that we are indifferent of the action taken in state n. In case optimality of threshold policies can be established, one is indifferent of the action taken in state n if the performance under threshold policies $n-1$ and n are equal. That is, W is such that

$$\begin{aligned}&\mathbb{E}(C_k(N_k^n, S_k^n(N_k^n))) - W\mathbb{E}(\mathbf{1}_{S_k^n(N_k^n)=0}) \\ &\quad = \mathbb{E}(C_k(N_k^{n-1}, S_k^{n-1}(N_k^{n-1}))) - W\mathbb{E}(\mathbf{1}_{S_k^{n-1}(N_k^{n-1})=0}),\end{aligned}$$

which after some algebra writes

$$W = \frac{\mathbb{E}(C_k(N_k^n, S_k^n(N_k^n))) - \mathbb{E}(C_k(N_k^{n-1}, S_k^{n-1}(N_k^{n-1})))}{\mathbb{E}(\mathbf{1}_{S_k^n(N_k^n)=0}) - \mathbb{E}(\mathbf{1}_{S_k^{n-1}(N_k^{n-1})=0})}.$$

Under the assumption that bandit k evolves in a birth-and-death fashion, the latter can be expressed as a function of the steady-state probabilities for bandit k under threshold policy n, i.e., $\pi_k^n(\cdot)$. In particular, $\mathbb{E}(\mathbf{1}_{S_k^n(N_k^n)=0}) = \sum_{m=0}^n \pi_k^n(m)$ when n is a 0-1 type of threshold policy and $\mathbb{E}(\mathbf{1}_{S_k^n(N_k^n)=0}) = 1 - \sum_{m=0}^n \pi_k^n(m)$ when n is a 1-0 type of threshold policy.

In the next proposition we state Whittle's index as a function of the steady-state probabilities and show that indexability is satisfied under a condition on the steady-state probabilities.

Proposition 2: Assume an optimal solution of (7) is of threshold type, and $\sum_{i=0}^n \pi_k^n(i)$ is strictly increasing in n, with $\pi_k^n(m)$ the steady-state probability for bandit k of being in state m under threshold policy n. Then, bandit k is indexable.

If the structure of an optimal solution of problem (7) is of 0-1 type, then, in case

$$\frac{\mathbb{E}(C_k(N_k^n, S_k^n(N_k^n))) - \mathbb{E}(C_k(N_k^{n-1}, S_k^{n-1}(N_k^{n-1})))}{\sum_{m=0}^n \pi_k^n(m) - \sum_{m=0}^{n-1} \pi_k^{n-1}(m)}, \quad (10)$$

is non-decreasing in n, Whittle's index $W_k(n_k)$ is given by (10) and is hence non-decreasing. Similarly, if the structure of an optimal solution of problem (7) is of 1-0 type, then, in case (10) is non-decreasing in n, $-W_k(n_k)$ is given by (10) and hence Whittle's index is non-increasing.

To the best of our knowledge, it has not been previously reported that when the bandits evolve in a birth-and-death fashion one can obtain an explicit expression of Whittle's index. For the particular case of a multi-class single-server abandonment queue, this was derived in [25].

Remark 1: In the case in which (10) is non-monotone in n, Whittle's index can no longer be derived by equating the average cost of two consecutive threshold policies. Instead, Whittle's index can be obtained following an algorithm as described for instance in [21] and [25] for particular examples. In this algorithm, comparing threshold policy -1 to the appropriate threshold policy $n_0 > -1$, we compute $W(n_0)$. Whittle's index for all thresholds $m \in \{0, 1, \ldots, n_0\}$, i.e., $W(m)$, equals $W(n_0)$. The correct choice of n_0 is the result of an optimization problem. The index in state $n_0 + 1$, is then computed by comparing threshold policy n_0 to an appropriate threshold policy $n_1 > n_0$. Then $W(m) = W(n_1)$ for all $m \in \{n_0 + 1, \ldots, n_1\}$. The algorithm stops when in an iteration i, n_i equals ∞.

Remark 2 ($\sum_{m=0}^n \pi^n(m)$ Being Strictly Increasing): Having $\sum_{m=0}^n \pi^n(m)$ to be strictly increasing makes sure that in Equation (10) we do not divide by 0. This assumption may exclude models of interest though. There are however tricks to overcome this assumption. We explain the latter using the example of the M/M/1 queue: Consider the M/M/1 queue under threshold policy n with 0-1 structure. This system is equivalent to the classical FIFO M/M/1 queue with n permanent customers. It is known that the probability that the stationary process is in state n is then $1 - \rho$. Hence, $\sum_{m=0}^n \pi^n(m) = 1 - \rho$ for all n. The subsidy obtained on average is therefore $W(1 - \rho)$, which is *independent* of the threshold policy n. This implies that the subsidy does not allow us to discriminate between different thresholds. In [8, Sec. 6.5] this is circumvented by looking at the discounted cost problem, instead of to the average cost criteria and scaling the immediate cost.

C. Whittle's Index Policy

In this section we describe how the optimal solution to the relaxed optimization problem is used to obtain a heuristic for the original model. The optimal solution to the relaxed problem, that is, activate all bandits that are in a state n_k such that $W_k(n_k) > W$, might be unfeasible for the original model where at most M bandits can be served at a time. Hence, Whittle [9] proposed the following heuristic, which is referred to as Whittle's index policy. In Section VI we discuss Whitte's index policy for several applications.

Definition 3 (Whittle's Index Policy): Assume at time t we are in state $\vec{N}(t) = \vec{n}$. The Whittle index policy activates the M bandits having currently the highest *non-negative* Whittle's index value $W_k(n_k)$.

Note that in case all bandits are in a state such that their Whittle's index is negative, all bandits are kept passive. The latter is a direct consequence of the relaxed optimization problem: when the Whittle index is negative for a bandit in state $\tilde{n}$, this means that it is made active only if $W < W_k(\tilde{n}) < 0$, that is, when a *cost* is payed for being passive.

In general it can be hard to verify whether an optimal solution is of threshold type, and whether (10) is non-decreasing in n. Both are needed in order to define Whittle's index using the formula given by Proposition 2. In addition, Whittle's index depends on the steady-state probabilities and hence, in many cases, does not provide qualitative insights on the behavior of the index policy.

In the next section we therefore develop a fluid approximation of (7) in order to derive a fluid index, which provides insights and can serve as a heuristic for the original stochastic problem.

IV. Fluid Version of Relaxed Optimization Problem

In this section we will solve the fluid version of the relaxed optimization problem (7), that is, we only take into account the average behavior of the system. As opposed to the stochastic relaxed problem, for the fluid version we *do* obtain an insightful expression for the so-called *fluid index*.

In Section IV-A we describe the fluid dynamics and the fluid version of the relaxed optimization problem. In Section IV-B we give a solution of the relaxed fluid model and the fluid index. In Section IV-C we define the fluid index policy, which serves as a heuristic for the original problem.

A. Fluid Model and Bias Optimality

We approximate the stochastic relaxed optimization problem as presented in Section III by a deterministic fluid model, where bandit k has a continuous state space $[0, \infty)$ instead of a discrete state space $\{0, 1, \ldots\}$. The fluid dynamics will be defined by only taking into account the mean dynamics of the stochastic process.

Let $m_k(t) \in [0, \infty)$ be the state of bandit k and $s_k(t) \in \{0, 1\}$ the control parameter. Let u denote a fluid control that determines $s_k^u(t)$, that is, whether bandit k is active or not. We use the following compact notation for the drift under action a:

$$f_k^a(m_k) := b_k^a(m_k) - d_k^a(m_k), \quad a = 0, 1,$$

with $m_k \geq 0$, where for non-integer values of m_k the functions b_k^0, d_k^0, b_k^1 and d_k^1 are defined such that they are continuous. We further assume $f_k^a(m_k)$ to be non-increasing in m_k for $a \in \{0, 1\}$. The fluid dynamics under control u can then be written as follows:

$$\frac{\mathrm{d}m_k^u(t)}{\mathrm{d}t} = (1 - s_k^u(t)) f_k^0(m_k^u(t)) + s_k^u(t) f_k^1(m_k^u(t)), \quad (11)$$

where the control u is such that $m_k^u(t) \geq 0$ for all t.

At time t, we define the cost for the fluid version of the relaxed problem (7) as

$$C_k(m_k(t), s_k(t)) := (1 - s_k(t)) C_k(m_k(t), 0) + s_k(t) C(m_k(t), 1),$$

where in non-integer values for m_k the function $C_k(m_k, a)$ is defined such that it is continuous in m_k.

An *equilibrium point* $(\bar{m}_k, \bar{s}_k)$ of the fluid dynamics is such that $\frac{\mathrm{d}m_k(t)}{\mathrm{d}t} = 0$, that is,

$$(1 - \bar{s}_k) f_k^0(\bar{m}_k) + \bar{s}_k f_k^1(\bar{m}_k) = 0,$$

or equivalently, $\bar{s}_k = f_k^0(\bar{m}_k)/(f_k^0(\bar{m}_k) - f_k^1(\bar{m}_k))$, with $\bar{s}_k \in [0, 1]$. That is, in equilibrium, a fraction of time $\bar{s}_k$ $(1 - \bar{s}_k)$ the action $a = 1$ $(a = 0)$ is chosen. Define $\bar{s}_k(\bar{m}_k) := f_k^0(\bar{m}_k)/(f_k^0(\bar{m}_k) - f_k^1(\bar{m}_k))$ and we assume throughout this paper $\bar{s}_k(\bar{m}_k)$ to be strictly monotone in $\bar{m}_k$. A discussion on the latter assumption can be found in Remark 4.

As stated in 7, in the stochastic model we aim to minimize for a given bandit the unconstrained optimization problem, that is, we minimize the time-average of the cost minus the subsidy obtained. In equilibrium, $\bar{s}_k$ is the average amount of time the system is active, hence, the fluid version of (7) will be to find the equilibrium point that minimizes the cost at equilibrium $EC_k(\bar{s}_k, W)$, where

$$EC_k(\bar{s}_k, W) := (1 - \bar{s}_k) C_k(\bar{m}_k, 0) + \bar{s}_k C(\bar{m}_k, 1) - W(1 - \bar{s}_k).$$

We denote by (m_k^*, s_k^*) an optimal equilibrium point and define the optimal equilibrium cost under subsidy W as

$$EC_k^*(W) := (1 - s_k^*)(C_k(m_k^*, 0) - W) + s_k^* C_k(m_k^*, 1). \quad (12)$$

Since the time-average criteria might be attained by several controls, see [24, Ch. 8], we are interested in controls that are *bias-optimal*. That is, among all controls that reach the optimal equilibrium point, a bias-optimal control is the one that minimizes the cost to get to this equilibrium point. Hence, our aim is to find the control u that minimizes the total bias cost, that is, the cost and subsidy obtained over time minus the optimal cost in equilibrium, denoted as

$$J_k^u(m_k(0), W) := \int_0^\infty \big(C_k(m_k(t), s_k^u(t)) - W(1 - s_k^u(t)) - EC_k^*(W) \big) \mathrm{d}t. \quad (13)$$

We define $J_k(m_k(0), W) := \min_u J_k^u(m_k(0), W)$.

The theory of optimal control shows that a sufficient condition in order for a control to be bias optimal is to solve the Hamilton-Jacobi-Bellman (HJB) equation, [24]:

$$EC_k^*(W) = \min\Big(C_k(m_k,1) + f_k^1(m_k)\partial J_k(m_k,W)/\partial m_k, \\ C_k(m_k,0) - W + f_k^0(m_k)\partial J_k(m_k,W)/\partial m_k\Big). \tag{14}$$

Then, for a given state m_k, an optimal action in that state is given by a minimizer of the right-hand-side in (14).

The main advantage of our approach is that (14) can be solved in general, see Proposition 3, while solving (7) (or equivalently (9)) requires to establish that an optimal policy for the relaxed problem is of threshold structure.

Remark 3: An alternative route to obtain (13) is to consider the total discounted cost criterion

$$\mathcal{C}^\phi(\beta) := \sum_{k=1}^K \mathbb{E}\left(\int_0^\infty \mathrm{e}^{-\beta t} C_k(N_k^\phi(t), S_k^\phi(\vec{N}^\phi(t)))\,\mathrm{d}t\right),$$

with $\beta > 0$ a discount factor, and to consider its fluid version. We then get a deterministic control problem under the total discounted cost criterion which is difficult to solve in general. As in Section III, we relax the service constraint and allow that the total discounted number of bandits active is $M/(1-\beta)$ or lower. For a single bandit, the objective of the relaxed fluid problem with discounted cost is then to find a control u that minimizes

$$J_k^{u,\beta}(m_k(0),W) := \int_0^\infty \mathrm{e}^{-\beta t}(\big(C_k(m_k(t), s_k^u(t)) \\ - W(1-s_k^u(t)))\mathrm{d}t.$$

The HJB equation in this case is given by

$$\beta J_k^\beta(m_k,W) \\ = \min_s(C_k(m_k,1) + \beta f_k^1(m_k)\partial J_k^\beta(m_k,W)/\partial m_k, \\ C_k(m_k,0) - W + \beta f_k^0(m_k)\partial J_k^\beta(m_k,W)/\partial m_k), \tag{15}$$

see [24, Ch. 10], where $J_k^\beta(m_k,W) = \min_u J_k^{u,\beta}(m_k,W)$. We now note that as $\beta \to 1$, $\beta J_k^\beta(m_k,W) \to EC_k^*(W)$, see [24, Corollary 8.2.5], and we thus obtain that (15) converges to (14).

Remark 4: As highlighted in Remark 2 the assumption on $\sum_{m=0}^n \pi_k^n(m)$ being strictly increasing (required for indexability of Whittle's index) excludes certain models. In the fluid context some of the models are excluded from our analysis due to the assumption that $\bar{s}_k(\bar{m}_k)$ is strictly monotone in $\bar{m}_k$. Let us consider an M/M/1 queue with controlled arrivals, where arrivals are exponentially distributed with rate $\tilde{\lambda}_k a$, $a = 0, 1$ and departures follow a Poisson process with rate $\tilde{\lambda}_k < \tilde{\mu}_k$. Then $(1-\bar{s}_k)f_k^0(m_k) + \bar{s}_k f_k^1(m_k) = -\tilde{\lambda}_k + \bar{s}_k\tilde{\mu}_k$. The latter equals 0 when $\bar{s}_k = \tilde{\lambda}_k/\tilde{\mu}_k$. Hence, when $\bar{s}_k = \tilde{\lambda}_k/\tilde{\mu}_k$, any m_k is an equilibrium point. The latter implies that the fraction of time the system is passive is the same no matter the equilibrium considered. Hence, the subsidy for passivity does not discriminate between states. In Section VI we use this setting to model a make-to-stock problem, and we overcome the issue raised above by considering stocked items to be perishable.

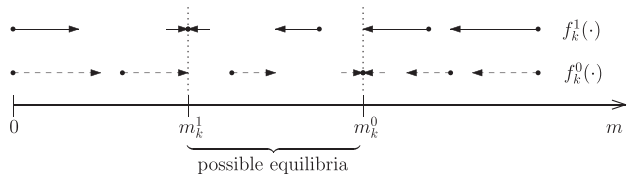

Fig. 3. Representation of fluid equilibria and drift functions when $m_k^1 < m_k^0$.

B. Optimal Fluid Control and Fluid Index

In this section we derive an optimal solution for the relaxed fluid problem (13) for a given bandit. This solution is described by a fluid index function, which allows a simple closed-form expression. Based on the fluid index we define in Section IV-C a heuristic for the original stochastic model, which we will show in Section VI to perform nearly optimal.

In order to define the fluid index, we need the following notation: we denote by m_k^i the value of m_k such that $f_k^i(m_k) = 0$, $i = 0, 1$. We adopt the convention that $m_k^i = \infty$ in case $f_k^i(m_k) > 0$ for all $m_k \geq 0$, and that $m_k^i = 0$ in case $f_k^i(m_k) < 0$ for all $m_k \geq 0$, that is, $m_k^i \in [0, \infty)$. The structure of the fluid index will depend on how m_k^1 and m_k^0 are ordered. In Figure 3 we show the drifts in case $m_k^1 < m_k^0$. The shape of the fluid index depends on whether the state m_k is such that $m_k < m_k^1$, $m_k \in [m_k^1, m_k^0]$, or $m_k > m_k^0$. In the first case, both drifts $f_k^0(m_k)$ and $f_k^1(m_k)$ are positive, in the second case the drifts are bidirectional, while in the third case the drifts are both negative.

We have assumed throughout the paper that $f_k^a(\cdot)$ is non-increasing for $a = 0, 1$ and $\bar{s}_k(\bar{m}_k)$ strictly monotone in $\bar{m}_k$. In order to define the fluid index policy, we will provide a definition and we will make the following additional assumptions on the drifts.

Definition 4: If $m_k^0 > m_k^1$, we define $\bar{a} := 1$ and if $m_k^1 \geq m_k^0$ we define $\bar{a} := 0$.

Assumption 1: We assume that:

- $f_k^a(m_k)$ is differentiable on $[m^{\bar{a}}, m^{1-\bar{a}}]$ for $a = 0, 1$.
- $f_k^a(m_k)$ is convex on m_k for $a = 0, 1$.
- $f_k^{\bar{a}}(m_k) - f_k^{\bar{a}}(\bar{m}_k) \geq (\leq) f_k^{1-\bar{a}}(m_k) - f_k^{1-\bar{a}}(\bar{m}_k)$, for all $m_k \leq (\geq)\bar{m}_k$, with $\bar{m}_k \in [m_k^{\bar{a}}, m_k^{1-\bar{a}}]$.
- $1 - \bar{a} + (2\bar{a}-1)\bar{s}_k(\bar{m}_k)$ is convex in $\bar{m}_k \in [m_k^{\bar{a}}, m_k^{1-\bar{a}}]$.

The hypotheses in Assumption 1 are easily verified for particular problems. This is done in the three examples considered in Section VI.

In the following proposition we give the expression for the fluid index and state an optimal solution of the fluid problem (13).

Proposition 3: Let Assumption 1 hold. Assume $C_k(m_k, a)$, $a = 0, 1$, is differentiable for $m_k \in [m^{\bar{a}}, m^{1-\bar{a}}]$, convex and non-decreasing for all m_k and $a = 0, 1$.

We assume $C_k(m_k, \bar{a}) - C_k(\bar{m}_k, \bar{a}) \leq C_k(m_k, 1-\bar{a}) - C_k(\bar{m}_k, 1-\bar{a})$. We define

$$w_k^{(a)}(m_k) := (f_k^1(m_k) - f_k^0(m_k))\frac{C_k(m_k,a) - C_k(m_k^a,a)}{f_k^a(m_k)}, \\ \text{for } a = 0, 1,$$

$$w_k^{(2)}(m_k) \\ := \frac{(f_k^1(m_k) - f_k^0(m_k))(f_k^0(m_k)\frac{\mathrm{d}C_k(m_k,1)}{\mathrm{d}m_k} - f_k^1(m_k)\frac{\mathrm{d}C_k(m_k,0)}{\mathrm{d}m_k})}{f_k^0(m_k)\frac{\mathrm{d}f_k^1(m_k)}{\mathrm{d}m_k} - f_k^1(m_k)\frac{\mathrm{d}f_k^0(m_k)}{\mathrm{d}m_k}},$$

and assume $(2\bar{a}-1)w_k^{(i)}(m_k)$, $i=0,1,2$, is non-decreasing for all m_k. Then an optimal solution of (13) is $s_k(t)=1$ if $W \leq w_k(m_k)$ and $s_k(t)=0$ if $W > w_k(m_k)$, where $w_k(m_k)$ is a continuous function defined as follows

$$w_k(m_k) := C_k(m_k,0) - C_k(m_k,1) + \begin{cases} w_k^{(\bar{a})}(m_k) & \text{if } m_k < m_k^{\bar{a}}, \\ w_k^{(2)}(m_k) & \text{if } m_k \in [m_k^{\bar{a}}, m_k^{1-\bar{a}}], \\ w_k^{(1-\bar{a})}(m_k) & \text{if } m_k > m_k^{1-\bar{a}}. \end{cases}$$

We will refer to the function $w_k(\cdot)$ as the *fluid index*.

The fluid index enables for the following interpretation. The function $EC_k(\bar{s}_k, W)$ is convex in $\bar{s}_k$ and $\partial EC_k(\bar{s}_k, W)/\partial \bar{s}_k = 0$ if and only if $W = C_k(m_k,0) - C_k(m_k,1) + w_k^{(2)}(m_k)$. That is, when $m_k \in [m_k^{\bar{a}}, m_k^{1-\bar{a}}]$, the fluid index is the value of W that minimizes the cost at equilibrium m_k. These calculations can be found in the proof of Proposition 3.

We observe from Proposition 3 that monotonicity of $w_k(m_k)$ in m_k implies that threshold policies are optimal for problem (13): in the case $m_k^1 < m_k^0$ ($m_k^1 > m_k^0$), non-decreasingness (non-increasingness) of $w_k(\cdot)$ implies that a threshold policy of structure 0-1 (1-0) is optimal, that is, it is optimal to be passive if and only if $m_k \leq m_k'(W)$ ($m_k \geq m_k'(W)$), with $m_k'(W)$ such that $w_k(m_k'(W)) = W$. This as opposed to the stochastic model, where optimality of threshold policies needs to be verified independently and might be difficult to derive.

Monotonicity of $w_k(m_k)$ is a simple property to verify. This represents an advantage with respect to the stochastic model, since in general optimality of threshold policies for birth-and-death stochastic bandits and indexability are difficult to establish. In Section VI we will show the monotonicity of the fluid index to be satisfied for three examples. The next lemma states sufficient conditions for $w_k(\cdot)$ to be monotone.

Lemma 1: Assume $C_k(m_k,1) = C_k(m_k,0)$ and $\frac{\mathrm{d}f_k^1(m_k)}{\mathrm{d}m_k} = \frac{\mathrm{d}f_k^0(m_k)}{\mathrm{d}m_k}$. Let $C_k(m_k,1)$ be non-decreasing in m_k and let $C_k(m_k,1)$ and $f_k^1(m_k)$ be polynomials of degree $P > 0$ and $\alpha \geq 0$ ($P > \alpha$), respectively. Then the function $(2\bar{a}-1)w_k(m_k)$ is non-decreasing if $f_k^{\bar{a}}(m_k) - f_k^{1-\bar{a}}(m_k) < 0$.

Proof 1: The proof follows after substituting $C_k(m_k,1) = C_k(m_k,0)$ and $\frac{\mathrm{d}f_k^1(m_k)}{\mathrm{d}m_k} = \frac{\mathrm{d}f_k^0(m_k)}{\mathrm{d}m_k}$ in the expressions of Proposition 3, and using that $f_k^i(\cdot)$ is non-increasing.

In Section III we defined the indexability property that allowed us to use the index values as a heuristics for the original problem. For the fluid model we use the same definition, that is, the fluid bandit is *indexable* if the collection of states in which the optimal action is passive increases as W increases. This property follows for the fluid model directly from the fact that $D_k(W) = \{m_k : W > w_k(m_k)\}$, see Definition 1. This as opposed to the stochastic model, for which indexability needs to be verified independently.

The generality of our approach is illustrated by the fact that when applied to classical problems, it retrieves well-known index policies. For instance, for a multi-class queue with user impatience and linear holding cost, our fluid index reduces to the $c\mu/\theta$-rule (introduced and asymptotic fluid optimality established in [20]). The $c\mu$-rule can also be retrieved from the latter by multiplying the fluid index and the abandonment rate θ, and letting $\theta \to 0$, [25].

C. Fluid Index Policy

The property of indexability allows us to define a heuristic for (3) based on the fluid index $w_k(\cdot)$ as obtained for the fluid version of the relaxed problem.

Definition 5 (Fluid Index Policy): Assume at time t we are in state $\vec{N}(t) = \vec{n}$. The fluid index policy prescribes to serve the M bandits having currently the highest non-negative fluid index $w_k(n_k)$.

In Section VI we will present numerical simulations that show that the performance of our fluid index policy is in fact nearly optimal. In addition, we numerically compare the fluid index with Whittle's index for the stochastic model.

V. Equivalence of Stochastic Index and Fluid Index in Light-Traffic Regime

In this section we consider the Whittle index as given in Proposition 2 and the fluid index proposed in Proposition 3 and prove that they become equivalent in a light-traffic regime. Recall that the birth transition rates are given by $b_k^a(\cdot)$ for $a = 0,1$. We will assume throughout this section that $b_k^a(m_k) = \lambda\gamma_k$ for all m_k, in problems where 0-1 type of threshold policies are optimal, and $b_k^a(m_k) = \lambda\gamma_k a$ for all m_k, in problems where 1-0 type of threshold policies are optimal. Here $\sum_{k=1}^K \gamma_k = 1$, and λ represents the intensity of the input to the system. We consider the light-traffic regime $\lambda \to 0$, that is, the birth rate is close to 0.

We first compute the Whittle index in light traffic in Proposition 4. The proof can be found in the supplementary.

Proposition 4: Let $W_k(\cdot)$ be as given in (10). Then $W_k(n_k) = W_k^{LT}(n_k) + o(1)$ as $\lambda \downarrow 0$, where

$$W_k^{LT}(n_k) := C_k(n_k,0) - C_k(n_k,1) + (d_k^1(n_k) - d_k^0(n_k))\frac{C_k(n_k,\bar{a}) - C_k(0,\bar{a})}{d_k^{\bar{a}}(n_k)},$$

and $\bar{a} = 0$ if the threshold that solves (8) is of 0-1 type and $\bar{a} = 1$ if the threshold that solves (8) is of 1-0 type.

Throughout this section we assume $d_k^a(m_k) > 0$ for all $m_k > 0$ and $a = 0,1$, which implies $m_k^0 \downarrow 0$ and $m_k^1 \downarrow 0$ as $\lambda \downarrow 0$. Hence, the fluid index in the light-traffic regime is given by $w_k(m_k) = C_k(m_k,0) - C_k(m_k,1) + w_k^{(\bar{a})}(m_k)$. In the next proposition we establish the equivalence of the Whittle index and the fluid index in the light-traffic regime. The proof can be found in the supplementary.

Proposition 5: Assume $d_k^a(m_k) > 0$ for all $m_k > 0$ and $a = 0,1$. Let $W_k(\cdot)$ be given as in (10). Assume an optimal solution for (13) to be the threshold policy n_k. Then

$$\lim_{\lambda \downarrow 0} W_k(m_k) = \lim_{\lambda \downarrow 0} w_k(m_k),$$

for all integer m_k.

VI. Case Studies

In this section we evaluate both the stochastic and fluid index policies for three examples of birth-and-death bandits. The main advantage of the index policies is that they are easily implementable and are applicable to many different resource allocation problems. The objective is to show how these policies apply to the following three decision making problems: (i) opportunistic scheduling in a wireless downlink channel, (ii) optimal blocking/routing in a power-aware server farm, and (iii) production of perishable items in a make-to-stock queue. Example (i) belongs to the class of problems depicted in Figure 1 and Examples (ii) and (iii) belongs to the class of problems depicted in Figure 2. In all cases, we compare the performance of Whittle's index policy (10) and the fluid index policy, as given in Proposition 3, against the optimal policy, which is computed using the Value Iteration approach, see [24]. Our overall conclusion is that the performance of the Whittle and the fluid index policies is nearly optimal.

A. Opportunistic Scheduling in a Wireless Downlink

In this section we consider a wireless downlink channel shared by K classes of users. Class-k users arrive according to a Poisson process of rate λ_k and their service requirement is exponentially distributed with mean $1/\tilde{\mu}_k$. At any moment in time, the base station can send data to at most one of the users present in the system. We assume the channel quality of a class-k user to be independent of the other users and that it can be modeled with a uniform random variable G_k on $[0, \gamma_k)$. As a consequence of opportunistic scheduling, the capacity when serving class k is the maximum of N_k i.i.d. random variables $G_{k,1}, \ldots, G_{k,N_k}$, distributed as G_k, see [23]. Hence, the expected capacity is given by $\mathbb{E}(\max(G_{k,1}, \ldots, G_{k,N_k})) = \gamma_k N_k(t)/(N_k(t)+1)$. We therefore take as departure rate $\mu_k(N_k) = \mu_k N_k/(N_k+1)$, where $\mu_k := \tilde{\mu}_k\gamma_k$. This Markov decision process is characterized by the following transition rates:

$$b_k^a(n_k) = \lambda_k, \quad \text{and } d_k^a(n_k) = \mu_k \frac{n_k}{n_k+1} a, \tag{16}$$

where $a = 1$ ($a = 0$) stands for serving (not serving) class k, see Figure 1. To make sure that a stationary policy exists that makes the system stable we assume $\rho := \sum_{k=1}^K \lambda_k/\mu_k < 1$.

The objective is to minimize the average holding cost, where $C_k(N_k, a)$ is the holding cost when having N_k class-k users in the system. Note that $C_k(N_k, a) = C_k(N_k)$ represents holding costs for users in the *system*, while $C_k(N_k, a) = C_k((N_k - a)^+)$ represents holding costs for users in the *queue*.

In this setting an optimal policy of the relaxed optimization problem (7) is of threshold type with 0-1 structure. The latter follows directly from Proposition 1.

The steady-state probabilities (obtained using the standard formula for a birth-and-death process) of class k under threshold policy n_k are given by

$$\pi_k^{n_k}(m_k) = 0, \quad \forall m_k \le n_k - 1,$$
$$\pi_k^{n_k}(m_k) = \left(\frac{\lambda_k}{\mu_k}\right)^{m_k - n_k} \frac{m_k+1}{n_k+1} \pi^{n_k}(n_k), \quad \forall m_k \ge n_k + 1,$$
$$\pi_k^{n_k}(n_k) = 1/\left(1 + \frac{1}{n_k+1} \sum_{i=1}^{\infty} \left(\frac{\lambda_k}{\mu_k}\right)^i (n_k + 1 + i)\right).$$

We now check that the function $\sum_{i=0}^n \pi_k^n(i)$ is strictly increasing in n, that is, $\pi_k^n(n) \le \pi_k^{n+1}(n+1)$. To do so observe that after some algebra $\pi_k^n(n) \le \pi_k^{n+1}(n+1)$ simplifies to $(\frac{1}{n+1} - \frac{1}{n+2}) \sum_{i=1}^{\infty} (\frac{\lambda_k}{\mu_k})^i i \ge 0$, which is satisfied for all n. From Proposition 2 we have now that the problem is indexable and that Whittle's index is given by (10) in case (10) is non-decreasing. Equation (10) can be numerically computed and verified to be non-decreasing. However, it does not help to provide insights into the properties of Whittle's index policy. This is the main motivation to derive the fluid index, which has a tractable closed-form expression.

The fluid dynamics are given by $\frac{\mathrm{d}m_k(t)}{\mathrm{d}t} = \lambda_k - \mu_k \frac{m_k}{m_k+1} s_k(t)$, where $s_k(t) \in \{0, 1\}$ ($s_k(t) = 1$ if station k is activated). Hence, $m_k^0 = \infty$ and $m_k^1 = \lambda_k/(\mu_k - \lambda_k)$, that is, the equilibrium points satisfy $\bar{m}_k \in [m_k^1, \infty)$. From Proposition 3 we can now derive the fluid index, which describes the policy that minimizes the bias-optimal criteria as given in (13).

We now present the two main results of this section. In Proposition (6) we derive the fluid index and in Proposition 6 we characterize the optimal solution for problem (13) with transition rates (16).

Proposition 6: Assume $C_k(m_k, a)$ is differentiable, convex and non-decreasing in m_k. In addition, assume $C_k(m_k, 0) - C_k(m_k, 1)$ to be convex non-decreasing in m_k. Then, the fluid index is non-decreasing and given by:

$$w_k(m_k) = C_k(m_k, 0) - C_k(m_k, 1) + \begin{cases} w_k^{(1)}(m_k) & \text{if } m_k < \lambda_k/(\mu_k - \lambda_k), \\ w_k^{(2)}(m_k) & \text{if } \lambda_k/(\mu_k - \lambda_k) \le m_k, \end{cases} \tag{17}$$

where

$$w_k^{(1)}(m_k) = \mu_k m_k \frac{C_k((\lambda_k/(\mu_k - \lambda_k), 1) - C_k(m_k, 1)}{\lambda_k - (\mu_k - \lambda_k) m_k},$$
$$w_k^{(2)}(m_k) = m_k(m_k+1)\left(\frac{\mathrm{d}C_k(m_k, 1)}{\mathrm{d}m_k} - \frac{\mathrm{d}C_k(m_k, 0)}{\mathrm{d}m_k}\right) + \frac{m_k^2 \mu_k}{\lambda_k} \frac{\mathrm{d}C_k(m_k, 0)}{\mathrm{d}m_k}.$$

Proof 2: Equation (17) being non-decreasing follows from observing that for any convex non-decreasing function $C_k(m_k, 1)$, for $m_k \le m_k'$, the function $\frac{C_k(m_k', 1) - C_k(m_k, 1)}{m_k' - m_k}$, is non-decreasing in m_k and from the fact that $C_k(m_k, 0) - C_k(m_k, 1)$ is convex and non-decreasing implies that $\frac{\mathrm{d}_k C_k(m_k, 0)}{\mathrm{d}m_k} - \frac{\mathrm{d}_k C_k(m_k, 1)}{\mathrm{d}m_k} \ge 0$ and is non-decreasing. Equation (17) now follows from Proposition 3 together with Proposition 7.

In the following proposition we present a bias-optimal solution for the fluid problem (13).

Proposition 7: Assume $C_k(m_k, a)$ is differentiable, convex and non-decreasing in m_k. In addition, assume $C_k(m_k, 0) - C_k(m_k, 1)$ to be convex non-decreasing in m_k. An optimal solution for problem (13) with transitions rates (16) is: $s_k(t) = 1$ if $W \le w_k(m_k)$ and $s_k(t) = 0$ if $W > w_k(m_k)$, where $w_k(m_k)$ is as defined in Proposition 6.

Proof 3: The proof follows by verifying Assumption 1. We have $f_k^a(m_k) = \lambda_k - \mu_k \frac{m_k}{m_k+1} a$, for $a \in \{0, 1\}$. Differentiability of $f_k^a(m_k)$ follows directly, also monotonicity of

TABLE I
EXAMPLE 1: RELATIVE SUB OPTIMALITY GAP IN %

ρ	0.1	0.2	0.3	0.4
Whittle index policy	0.20289	1.16215	2.54794	3.54934
Fluid index policy	0.20289	1.16215	2.55440	3.54936
ρ	0.5	0.6	0.7	0.8
Whittle index policy	3.52057	2.54793	1.56715	0.66077
Fluid index policy	3.52098	2.55439	1.60799	0.75140

$\bar{s}_k(\bar{m}_k)$ for $\bar{m}_k$ in $[m_k^1, m_k^0]$ and convexity of $f_k^0(m_k)$. The function $f_k^1(m_k)$ satisfies

$$\frac{\mathrm{d}^2 f_k^1(m_k)}{\mathrm{d}m_k^2} = \mu_k \frac{2}{(m_k+1)^3} \geq 0,$$

for all $m_k \geq 0$, and it is hence convex in m_k.

We have, $\bar{s}_k(\bar{m}_k) = \lambda_k(\bar{m}_k+1)/\mu_k\bar{m}_k$, hence

$$\frac{\mathrm{d}^2}{\mathrm{d}m_k^2}\left(\bar{s}_k(\bar{m}_k)\right) = 2\frac{\lambda_k}{\mu_k \bar{m}_k^3} \geq 0,$$

that is, $\bar{s}_k(\bar{m}_k)$ is convex for $\bar{m}_k \in [m_k^1, m_k^0]$.

The inequality $f_k^1(m_k) - f_k^1(\bar{m}_k) \geq (\leq) f_k^0(m_k) - f_k^0(\bar{m}_k)$ for all $m_k \leq (\geq)\bar{m}_k$ and $\bar{m}_k \in [m_k^1, m_k^0]$, simplifies to $\mu_k(\frac{\bar{m}_k}{\bar{m}_k+1} - \frac{m_k}{m_k+1}) \geq (\leq) 0$ for all $m_k \leq (\geq)\bar{m}_k$ and $\bar{m}_k \in [m_k^1, m_k^0]$, which is satisfied due to $\frac{m_k}{m_k+1}$ being a non-decreasing function in m_k.

Then from Propositions 3 and 6, an optimal solution for problem (13) is $s_k(t) = 1$ if $W \leq w_k(m_k)$ and $s_k(t) = 0$ if $W > w_k(m_k)$.

The fluid index being non-decreasing implies that the fluid index policy as defined in Section IV-C will give more importance to a class to be served as its queue length grows.

Having a closed-form expression for the fluid index as given in (17) gives us insights on the behavior of the system with respect to the parameters involved. For the sake of clarity assume linear cost of type $C_k(m,0) = C_k(m,1) = c_k m$, and $\lambda_k = \lambda\delta_k$. Then $w_k(m_k) = c_k m_k/(1-\rho_k)$ for $m_k < \lambda_k/(\mu_k - \lambda_k)$, and $w_k(m_k) = c_k m_k^2/\rho_k$ otherwise. Hence, as $\lambda \downarrow 0$, in states very close to the origin priority is given according to $c_k m_k/(1-\rho_k)$ and otherwise according to $c_k\mu_k m_k^2/\delta_k$.

In the example below we will numerically evaluate the performance of the two index policies against the optimal policy.

Example 1: Let us assume two classes of users with $\mu_1 = 16$, $\mu_2 = 27$, and $\lambda_1/\mu_1 = \rho/2$, $\lambda_2/\mu_2 = \rho/2$. We further assume that the cost function is given by $C_k(n,a) = c_k(n-a)^2 + b_k(n-a)$ for $k \in \{1,2\}$, with $b_1 = 0.1, b_2 = 1, c_1 = 2$ and $c_2 = 1.5$, that is, quadratic holding cost for the number of users waiting to be served. We compute the relative error of the Whittle index policy as well as the relative error of the fluid index policy with respect to the optimal policy, see Table I. We observe that both index policies perform nearly optimal across all loads.

In Figure 4 we depict the actions taken under the optimal policy, Whittle's index policy, and the fluid index policy, for $\rho = 0.5$. The three policies are characterized by the three switching curves as depicted in the figure. Below the curve class 1 is served and above the curve class 2 is served. We observe that the two switching curves corresponding to the fluid index policy and the Whittle index policy coincide in almost the entire state space, and capture the qualitative structure of the optimal policy.

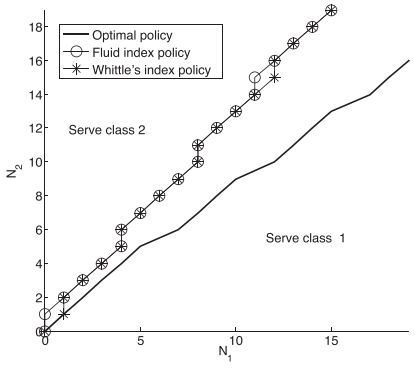

Fig. 4. Switching curves under the optimal policy, Whittle's index policy and fluid index policy.

B. Routing/Blocking in a Power-Aware Server-Farm

We consider a server farm with K heterogeneous service stations each having one server, see Figure 2. Users arrive to the system following a Poisson process of rate λ. An arriving user is either routed to one of the stations or is blocked. The service capacity of the power-aware servers follows a speed-scaling rule [19] in order to balance between power consumption and server capacity. We assume that when in state N_k, the service capacity is $c(N_k) := N_k^\alpha$, with $0 < \alpha < 1$. The service requirement of a user in station k is exponentially distributed with mean $1/\mu_k$. Hence, the departure rate is $\mu_k(N_k) = \mu_k N_k^\alpha$.

Each time a user is blocked for service a penalty D is paid, hence, implying blocking cost to occur at rate λD. A common model for the power consumption is $c(N_k)^{1/\alpha}$, hence, we have that the power consumed in state N_k is equal to N_k. We therefore take as cost $C_k(N_k, a) = C_k(N_k) + \beta_k N_k + D\lambda(1-a)$, where $C_k(N_k)$ represents the holding cost of having users in server k and $\beta_k \geq 0$ controls the relative cost of power consumption. We assume $C_k(N_k)$ to be convex. The aim is to find the optimal blocking/routing policy in order to minimize the sum of the average holding cost, power consumption and penalty for blocking users. An optimal load balancing policy must strike the right balance between dispatching a user to a server with a large queue length (which implies a large increase in holding cost, due to convexity, but a high service rate), dispatching to a server with a small queue length (which implies a small increase in holding cost but a small service rate), and blocking a user (which implies a blocking cost, however no additional holding cost is incurred). This is a very complex optimization problem. We will see that the two index policies as described in this paper are able to perform close to optimal.

The Markov chain has the following transitions:

$$b_k^a(m_k) = \lambda a, \text{ and } d_k^a(m_k) = \mu_k m_k^\alpha, \tag{18}$$

where $a = 0$ ($a = 1$) stands for blocking (accepting) a user in server k. We note that, due to the speed-scaling rule, the system is always stable. We first determine the fluid index policy for this model. The fluid dynamics is given by $\frac{\mathrm{d}m_k(t)}{\mathrm{d}t} = \lambda s_k(t) - \mu_k m_k^\alpha$, with $s_k(t) \in \{0,1\}$. We have $m_k^0 = 0$, and $m_k^1 = (\lambda/\mu_k)^{1/\alpha}$, that is, the equilibrium points are in the interval $\bar{m}_k \in [0, m_k^1]$.

Next we present the two main results of this section. In Proposition 8 we derive the fluid index and in Proposition 9 we characterize the optimal solution for problem (13) for transition rates (18). The proof for Proposition 8 follows from Proposition 3 and Lemma 1.

Proposition 8: Assume $C_k(m_k)$ to be a polynomial of degree P with $P > \alpha$. Then, the fluid index is non-increasing and given by:

$$w_k(m_k) = D\lambda + \begin{cases} w_k^{(2)}(m_k) & \text{if } 0 \le m_k \le (\lambda/\mu_k)^{\alpha^{-1}}, \\ w_k^{(1)}(m_k) & \text{if } (\lambda/\mu_k)^{\alpha^{-1}} < m_k, \end{cases}$$

where

$$w_k^{(2)}(m_k) = -\frac{\lambda\alpha^{-1}m_k}{\mu_k m_k^\alpha}\frac{\mathrm{d}\tilde{C}_k(m_k)}{\mathrm{d}m_k}$$

$$w_k^{(1)}(m_k) = -\lambda\frac{(\tilde{C}_k((\lambda/\mu_k)^{\alpha^{-1}}) - \tilde{C}_k(m_k))}{\lambda - \mu_k \min(T, m_k^\alpha)},$$

with $\tilde{C}_k(m_k) = C_k(m_k) + \beta_k m_k$.

In the following proposition we present an optimal solution for problem (13).

Proposition 9: Assume $C_k(m_k)$ to be a polynomial of degree P with $P > \alpha$. An optimal solution for problem (13) with transition rates (18) is: $s_k(t) = 1$ if $W \le w_k(m_k)$ and $s_k(t) = 0$ if $W > w_k(m_k)$, where $w_k(m_k)$ is as given in Proposition 8.

Proof 4: The proof follows by verifying Assumption 1. We have $f_k^a(m_k) = \lambda a - \mu_k m_k^\alpha$, for $a \in \{0,1\}$ and $\alpha < 1$. Differentiability of $f_k^a(m_k)$ follows directly as well as monotonicity of $\bar{s}_k(\bar{m}_k)$ in $[m_k^0, m_k^1]$. The function $f_k^a(m_k)$ satisfies

$$\frac{\mathrm{d}^2 f_k^a(m_k)}{\mathrm{d}m_k^2} = -\alpha(\alpha-1)\mu_k m_k^{\alpha-2} \ge 0,$$

for all $m_k \ge 0$, due to the assumption $\alpha < 1$. Hence, $f_k^a(m_k)$ is convex in m_k for $a = 0, 1$.

We have $1 - \bar{s}_k(\bar{m}_k) = 1 - \mu_k m_k^\alpha/\lambda$, hence

$$\frac{\mathrm{d}^2}{\mathrm{d}m_k^2}\left(1 - \bar{s}_k(\bar{m}_k)\right) = -\frac{\mu_k\alpha(\alpha-1)\bar{m}_k^{\alpha-2}}{\lambda} \ge 0,$$

for all $\bar{m}_k \in [m_k^1, m_k^0]$, since $\alpha < 1$. Hence, $1 - \bar{s}_k(\bar{m}_k)$ is convex in $\bar{m}_k$.

The inequality $f_k^1(m_k) - f_k^1(\bar{m}_k) \le (\ge) f_k^0(m_k) - f_k^0(\bar{m}_k)$ simplifies to $\mu_k(\bar{m}_k^\alpha - m_k^\alpha) \le (\ge)\mu_k(\bar{m}_k^\alpha - m_k^\alpha)$, which is satisfied for all m_k and $\bar{m}_k$.

Then from Proposition 3 and 8, an optimal solution for problem (13) is $s_k(t) = 1$ if $W \le w_k(m_k)$ and $s_k(t) = 0$ if $W > w_k(m_k)$.

The fluid index being non-increasing implies that the fluid index policy will prefer to route to servers having a relatively small queue length. Since the fluid index policy only routes to servers with a positive fluid index, there is an $\overline{N}_k$ such that when $N_k \ge \overline{N}_k$, no users will be routed to this server k.

As in the previous section we use Proposition 8 to obtain interesting insights for particular cases. For the sake of clarity assume linear cost of type $C_k(m) = c_k m$. Then, as $\lambda \uparrow \infty$, $w_k(m_k)$ will be given by $D\lambda + w_k^{(2)}(m_k)$, and $w_k^{(2)}(m_k) = -\lambda c_k \frac{m_k^{1-\alpha}}{\mu_k\alpha}$, hence priority will be given according to $c_k\frac{m_k^{1-\alpha}}{\mu_k\alpha}$.

For the model under study 1-0 type of threshold policies are an optimal solution of the relaxed optimization problem (7). The proof follows from Proposition 1.

The steady-state probabilities (obtained using the standard formula for a birth-and-death process) of class k under threshold policy n_k are given by

$$\pi_k^{n_k}(m_k) = 0, \quad \forall m_k \ge n_k + 2,$$

$$\pi_k^{n_k}(m_k) = \frac{\lambda^{m_k}}{\mu_k^{m_k}\prod_{i=1}^{m_k} i^\alpha}\pi^{n_k}(0), \quad \forall m_k \le n_k + 1,$$

$$\pi_k^{n_k}(0) = \left(\sum_{m_k=0}^{n_k+1}\frac{\lambda^{m_k}}{\mu_k^{m_k}\prod_{i=1}^{m_k} i^\alpha}\right)^{-1}.$$

We now check that the function $\sum_{i=0}^n \pi_k^n(i)$ is strictly increasing in n, or equivalently, $\pi_k^n(n+1)$ is strictly decreasing in n. We then want to prove $\pi_k^n(n+1) \le \pi^{n-1}(n)$ which after some algebra simplifies to

$$\pi^{n-1}(0) \ge \frac{\lambda}{\mu(n+1)^\alpha}\pi^n(0)$$

$$\Leftrightarrow \sum_{m_k=0}^{n+1}\frac{\lambda^{m_k}}{\mu_k^{m_k}\prod_{i=1}^{m_k} i^\alpha} \ge \frac{\lambda}{\mu(n+1)^\alpha}\sum_{m_k=0}^{n}\frac{\lambda^{m_k}}{\mu_k^{m_k}\prod_{i=1}^{m_k} i^\alpha}$$

$$\Leftrightarrow \sum_{m_k=1}^{n}\frac{\lambda^{m_k}}{\mu_k^{m_k}\prod_{i=1}^{m_k-1} i^\alpha}\left(\frac{1}{m_k^\alpha} - \frac{1}{(n+1)^\alpha}\right) + 1 \ge 0,$$

the last inequality is satisfied due to $1/m_k^\alpha - 1/(n+1)^\alpha \ge 0$ for all $m_k \in \{1, \ldots, n\}$. From Proposition 2 we have now that the problem is indexable and that Whittle's index is given by (10) in case (10) is non-decreasing. Equation (10) can be numerically computed and verified to be non-decreasing. Note that the optimal structure of the fluid version of the relaxed optimization problem is also of 1-0 structure (since the fluid index is non-increasing).

We now present an example to evaluate the performance of both index policies.

Example 2: In this example we assume 2 classes of users which arrive at rate $\lambda = 18$. We set the speed scaling parameter at $\alpha = 1/2$. The cost function is such that $C_k(m_k, a) = C_k(m_k) + \beta_k m_k + D\lambda a$, and we assume $C_k(m_k) = c_k m_k^2$ where $c_1 = c_2 = 2$, and $\beta_1 = 3, \beta_2 = 5$. We further assume that the cost for blocking users is $D = 25$. The service rates μ_1, μ_2 are such that $\mu_1 = \mu_2 = 2\lambda/\rho$. We set $M = 1$, that is, a customer can be routed to at most one server. We observe in Table II that the performance for the Whittle index policy as well as for the fluid index policy for various values of ρ is nearly optimal. Moreover, in Figure 5 we illustrate the optimal strategy together with the Whittle index policy for $\rho = 2.5$.

TABLE II
EXAMPLE 2: RELATIVE SUB OPTIMALITY GAP IN %

ρ	0.1	0.3	0.5
Fluid index	0.08704×10^{-7}	0.16036×10^{-7}	0.13968×10^{-7}
Whittle's index	0.08704×10^{-7}	0.16036×10^{-7}	0.13968×10^{-7}
ρ	0.7	0.9	1.1
Fluid index	0.06279×10^{-7}	0.08210×10^{-7}	0.06124×10^{-7}
Whittle's index	0.06279×10^{-7}	0.08210×10^{-7}	0.06124×10^{-7}
ρ	1.5	2	2.5
Fluid index	0.01872×10^{-7}	0.06099×10^{-7}	0.10921×10^{-7}
Whittle's index	0.01872×10^{-7}	0.06099×10^{-7}	0.07110×10^{-7}

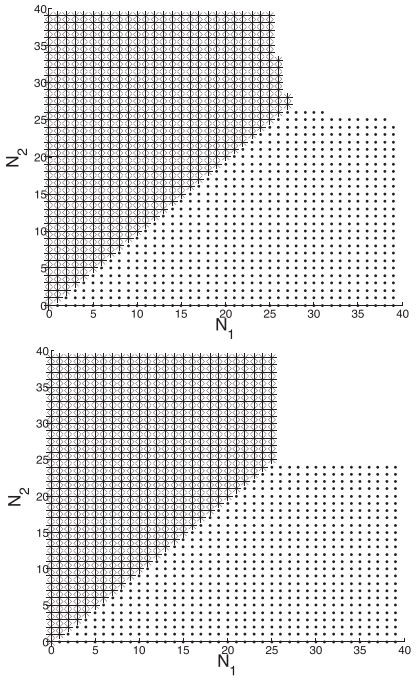

Fig. 5. In the area with "*" (".") class 1 (class 2) is prioritized and in the white area users are blocked. Above: Optimal policy. Below: Whittle index policy and the fluid index policy.

The fluid index policy coincides with the strategy given by Whittle's index policy and captures the qualitative structure of the optimal policy.

C. Production of Perishable Items in a Make-to-Stock Queue

We consider a single production machine which produces K different classes of items. Demands for a class-k item arrive to the system following a Poisson process of rate λ_k. The machine can only produce one item at a time, and the production time is exponentially distributed with mean $1/\mu_k$. The items, once produced, are stocked until either a user requests the item or the item perishes. We assume that perishing is exponentially distributed with mean $1/\theta_k$. The machine chooses whether to produce an item or whether to idle. In case the machine chooses to produce an item, a decision on which class-k item to produce has to be made. This problem belongs to the class of problems depicted in Figure 2. This model, without the possibility of items to perish, was considered in [26]. We denote by $N_k(t)$ the number of class-k items stocked in the inventory. The Markov decision problem is characterized by the following transition rates:

$$b_k^a(m_k) = \mu_k a \quad \text{and } d_k^a(m_k) = \lambda_k + \theta_k m_k, \tag{19}$$

where $a = 1 (a = 0)$ stands for producing (not producing) the items of that class. We note that, due to the abandonments, the system is always stable. If a demand for a class-k item arrived when $N_k = 0$, *i.e.*, no class-k items are left in the stock, then the sale is lost. The latter incurs a per lost cost $D_k > 0$. Every time a class-k item perishes a per item cost, δ_k, is paid. The system incurs a cost $c_k(N_k, a)$ per unit of time N_k class-k items are stocked, $a = 0, 1$. Hence, the cost per unit of time incurred for class k is

$$C_k(N_k, a) = c_k(N_k, a) + \delta_k \theta_k N_k + \lambda_k D_k \mathbb{1}_{\{N_k=0\}}.$$

The objective of the stochastic model is to minimize the average cost, *i.e.*, $\limsup_{T\to\infty} \frac{1}{T}\mathbb{E}\int_0^T C_k(N_k^\phi(t), S_k^\phi(t))\mathrm{d}t$, which is equivalent to

$$\limsup_{T\to\infty} \frac{1}{T}\mathbb{E}\int_0^T (\tilde{C}_k(N_k^\phi(t), S_k^\phi(t)) + \lambda_k D_k \pi_k^\phi(0))\mathrm{d}t,$$

with $\tilde{C}_k(m, a) = c_k(m, a) + \delta_k \theta_k m$.

In this setting an optimal policy of the relaxed optimization problem (7) is of threshold type with 1-0 structure. The proof follows from Proposition 1.

The steady-state-probabilities $\pi_k^{n_k}(m_k)$ under threshold n_k are given as follows,

$$\pi_k^{n_k}(m_k) = \frac{\mu_k^{m_k}}{\prod_{i=1}^{m_k}(\lambda_k + \theta i)}\pi^{n_k}(0), \quad \text{for all } m_k \le n_k + 1,$$
$$\pi_k^{n_k}(m_k) = 0, \quad \text{for all } m_k \ge n_k + 2,$$

where $\pi_k^{n_k}(0) = \left(\sum_{m=0}^{n_k+1} \frac{\mu_k^m}{\prod_{i=1}^m(\lambda_k+\theta_k i)}\right)^{-1}$. We now check that $\pi^n(n+1)$ is strictly decreasing in n as required for indexability in Proposition 2. We therefore need to prove that $\pi^n(n+1) \le \pi^{n-1}(n)$ for all $n \ge 0$, that is,

$$\frac{\mu_k^{n+1}}{\prod_{i=1}^{n+1}(\lambda_k + \theta_k i)}\pi_k^n(0) \le \frac{\mu_k^n}{\prod_{i=1}^{n}(\lambda_k + \theta_k i)}\pi_k^{n-1}(0)$$
$$\Leftrightarrow \frac{\mu_k}{\lambda_k + \theta_k(n+1)}\sum_{m=0}^{n}\frac{\mu_k^m}{\prod_{i=1}^m(\lambda_k+\theta_k i)}$$
$$\le \sum_{m=0}^{n+1}\frac{\mu_k^m}{\prod_{i=1}^m(\lambda_k+\theta_k i)}$$
$$\Leftrightarrow \sum_{m=1}^{n}\frac{\mu_k^m}{\prod_{i=1}^{m-1}(\lambda_k+\theta_k i)}\left(\frac{1}{\lambda_k+\theta_k m} - \frac{1}{\lambda_k+\theta_k(n+1)}\right) + 1 \ge 0.$$

The latter inequality is satisfied due to $(\frac{1}{\lambda_k+\theta_k m} - \frac{1}{\lambda_k+\theta_k(n+1)}) \ge 0$ for all $m \le n$. Hence, from Proposition 2 we have that Whittle's index is given by (10) in case (10) is non-increasing. Equation (10) can be numerically computed and verified to be non-increasing.

The fluid dynamics are given by $\frac{\mathrm{d}m_k(t)}{\mathrm{d}t} = \mu_k s_k(t) - \lambda_k - \theta_k m_k$ for all $m_k \ge 0$. We will assume $\lambda_k < \mu_k$ and then by the convention assumed in Section IV-B, *i.e.,* $m_k^a = 0$ in case $f_k^a(m_k) < 0$ for all $m_k \ge 0$ and $a = 0, 1$, we have that $m_k^0 = \max\{0, -\lambda_k/\theta_k\} = 0$ and $m_k^1 = \max\{\frac{\mu_k-\lambda_k}{\theta_k}, 0\} = \frac{\mu_k-\lambda_k}{\theta_k}$. In the stochastic system once the server decides to idle no higher states are visited and therefore, under threshold policy n_k, the average cost corresponding to

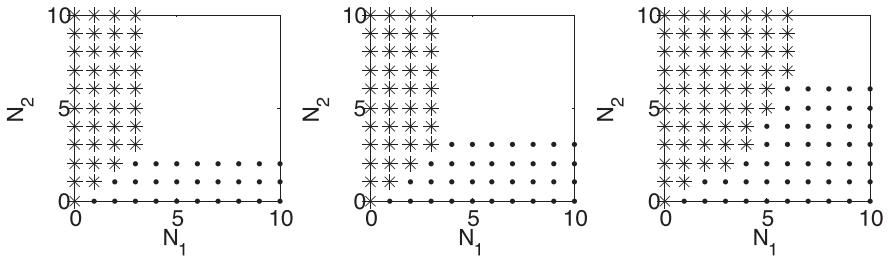

Fig. 6. In the area with "*" (".") class 1 (class 2) is prioritized and in the white area users are blocked. Left: Optimal policy. Middle: Whittle index policy and, right: the fluid index policy.

lost sales equals $D_k\lambda_k\pi_k^{n_k}(0)$. This is a cost that is paid on average, and therefore, the bias optimality criterion of the fluid model cannot capture it, since the fluid system under threshold policies never reaches 0 unless that is the equilibrium point. Instead, we incorporate the sale lost cost per unit of time the passive action is chosen, that is, in the fluid model we consider as cost $c_k(m_k,a)+\theta_k\delta_k m_k+\lambda_k D_k(1-a)$.

The fluid index is derived in the following proposition.

Proposition 10: Assume $c_k(m_k,a)$ is convex differentiable and non-decreasing, and $c_k(m_k,1)-c_k(m_k,0)$ is convex non-decreasing. Then, the fluid index is non-increasing and given by:

$$w_k(m_k)=c_k(m_k,0)-c_k(m_k,1)+D_k\lambda_k+\begin{cases}w_k^{(2)}(m_k), & \text{if } m_k\le(\mu_k-\lambda_k)/\theta_k\\ w_k^{(1)}(m_k), & \text{if } m_k>(\mu_k-\lambda_k)/\theta_k,\end{cases}\tag{20}$$

where

$$w_k^{(2)}(m_k)=\left(-m_k-\frac{\lambda_k}{\theta_k}\right)\left(\frac{\mathrm{d}c_k(m_k,1)}{\mathrm{d}m_k}-\frac{\mathrm{d}c_k(m_k,0)}{\mathrm{d}m_k}\right)-\frac{\mu_k}{\theta_k}\frac{\mathrm{d}c_k(m_k,0)}{\mathrm{d}m_k},$$

$$w_k^{(1)}(m_k)=-\frac{\mu_k}{\theta_k}\left(\frac{c_k(m_k,1)-c_k((\mu_k-\lambda_k)/\theta_k,1)}{m_k-(\mu_k-\lambda_k)/\theta_k}\right)-\mu_k\delta_k.$$

Proof 5: Equation (20) follows from Proposition 3 and Lemma 11. The index being non-decreasing follows from observing that for any convex non-decreasing function $c_k(m_k,1)$, for $m_k\le m_k'$, the function $\frac{c_k(m_k',1)-c_k(m_k,1)}{m_k'-m_k}$, is non-decreasing in m_k. Also from the fact that $c_k(m_k,1)-c_k(m_k,0)$ being convex and non-decreasing implies that $\frac{\mathrm{d}_k c_k(m_k,1)}{\mathrm{d}m_k}-\frac{\mathrm{d}_k c_k(m_k,0)}{\mathrm{d}m_k}\ge 0$ and is non-decreasing.

We now present an optimal solution of problem (13)

Proposition 11: Assume $c_k(m_k,a)$ is convex differentiable and non-decreasing, and $c_k(m_k,1)-c_k(m_k,0)$ is convex non-decreasing. An optimal solution for problem (13) with transitions rates (19) is: $s_k(t)=1$ if $W\le w_k(m_k)$ and $s_k(t)=0$ if $W>w_k(m_k)$, where $w_k(m_k)$ is as given in Proposition 10.

Proof 6: The proof follows by verifying Assumption 1. We have $f_k^a(m_k)=\mu_k a-\lambda_k-\theta_k m_k$, for $a\in\{0,1\}$. Differentiability of $f_k^a(m_k)$ follows directly as well as monotonicity of $\bar{s}_k(\bar{m}_k)$ in $[m_k^0,m_k^1]$. The function $f_k^a(m_k)$ satisfies

$$\frac{\mathrm{d}^2 f_k^a(m_k)}{\mathrm{d}m_k^2}=0,$$

for all $m_k\ge 0$. Hence, $f_k^a(m_k)$ is convex in m_k for $a=0,1$.

We have $1-\bar{s}_k(\bar{m}_k)=1-(\lambda_k+\theta_k m_k)/\mu_k$, hence

$$\frac{\mathrm{d}^2}{\mathrm{d}m_k^2}\left(1-\bar{s}_k(\bar{m}_k)\right)=0,$$

for all $\bar{m}_k\in[m_k^1,m_k^0]$. Hence, $1-\bar{s}_k(\bar{m}_k)$ is convex in $\bar{m}_k$.

The inequality $f_k^1(m_k)-f_k^1(\bar{m}_k)\le(\ge)f_k^0(m_k)-f_k^0(\bar{m}_k)$ simplifies to $\bar{m}_k-m_k\le(\ge)\bar{m}_k-m_k$, which is satisfied for all m_k and $\bar{m}_k$.

Then from Proposition 3 and Proposition 8, an optimal solution for problem (13) is $s_k(t)=1$ if $W\le w_k(m_k)$ and $s_k(t)=0$ if $W>w_k(m_k)$.

The fluid index being non-increasing implies that the more class-k items are in stock, less is the priority to produce a class-k item. If the fluid index is negative for all the classes then the fluid index policy prescribes not to produce any item, and hence the machine idles.

We evaluate the performance of the index policies in the following example.

TABLE III
EXAMPLE 3: RELATIVE SUB OPTIMALITY GAP IN %

ρ	0.1	1	1.5	2
Whittle index policy	1.204×10^{-4}	0.0036	0.0099	0.0285
Fluid index policy	0.0037	0.0038	0.0164	0.0369
ρ	2.5	3	3.5	4
Whittle index policy	0.0676	0.1321	0.2242	4.0901
Fluid index policy	0.0605	0.0807	0.0808	0.3246

Example 3: In this example we assume two classes of items which are produced at rate $\mu_1=4$ and $\mu_2=5$, in case the machine decides to produce. We see $c_k(m_k,a)=c_k m_k^2+b_k m_k$ where $c_1=1,c_2=2$ and $b_1=b_2=1$.. We further assume, $\theta_1=2,\theta_2=2.5$, and that the cost for perishing items is $\delta_1=0.5,\delta_2=3$ and cost per lost sale $D_1=20$, $D_2=14$. Demands for items arrive at rates $\lambda_1=3.5$, $\lambda_2=4.8$. We set $M=1$, that is, only one machine can produce the items. In Figure 6 we illustrate the optimal policy together with the Whittle index and the fluid index policies for $\rho=2$. The Whittle index policy and fluid index policy capture the qualitative structure of the optimal solution. Although in this example the fluid index policy prescribes higher production than the optimal and the Whittle index policies (see Figure 6 (right)), we see in Table III that the suboptimality gap of both heuristics is very small.

VII. Conclusions and Further Research

In the two main contributions of the paper we have (i) derived a closed-form expression for Whittle's index for a birth-and-death process, and (ii) developed a fluid framework to derive fluid index policies. The Whittle index is given in a compact expression and it can be numerically computed, however it requires to establish optimality of threshold policies, and in addition, it does not provide qualitative insights into the index policy. On the other hand, the fluid index is much simpler to calculate, does not require to verify for optimality of threshold policies, and it *does* provide qualitative insights.

The numerical examples have shown that the fluid index policy and the Whittle index policy have a very similar performance. An interesting problem would be to mathematically obtain bounds on the performance of the fluid index policy compared to Whittle's index policy. The latter is known to be asymptotically optimal as the number of bandits that can be simultaneously made active grows proportionally with the population of bandits, see [11], [12], [27], [28].

VIII. Acknowledgment

Part of the work was carried out in CNRS, Institut de Recherche en Informatique de Toulouse, and also with CNRS, Laboratory for Analysis and Architecture of Systems, Toulouse, France.

References

[1] M. Larrañaga, U. Ayesta, and I. M. Verloop, "Stochastic and fluid index policies for resource allocation problems," in *Proc. IEEE INFOCOM*, Apr./May 2015, pp. 1230–1238.

[2] C. Buyukkoc, P. Varaya, and J. Walrand, "The $c\mu$ rule revisited," *Adv. Appl. Probab.*, vol. 17, no. 1, pp. 237–238, 1985.

[3] P. S. Ansell, K. D. Glazebrook, J. Niño-Mora, and M. O'Keeffe, "Whittle's index policy for a multi-class queueing system with convex holding costs," *Math. Methods Oper. Res.*, vol. 57, no. 1, pp. 21–39, Apr. 2003.

[4] C. F. Bispo, "The single-server scheduling problem with convex costs," *Queueing Syst.*, vol. 73, no. 3, pp. 261–294, Mar. 2013.

[5] J. M. George and J. M. Harrison, "Dynamic control of a queue with adjustable service rate," *Oper. Res.*, vol. 49, no. 5, pp. 720–731, Sep./Oct. 2001.

[6] M. Larrañaga, U. Ayesta, and I. M. Verloop, "Dynamic fluid-based scheduling in a multi-class abandonment queue," *Perform. Eval.*, vol. 70, no. 10, pp. 841–858, Oct. 2013.

[7] L. E. Schrage and L. W. Miller, "The queue M/G/1 with the shortest remaining processing time discipline," *Oper. Res.*, vol. 14, no. 4, pp. 670–684, Aug. 1966.

[8] J. Gittins, K. Glazebrook, and R. Weber, *Multi-Armed Bandit Allocation Indices*. New York, NY, USA: Wiley, 2011.

[9] P. Whittle, "Restless bandits: Activity allocation in a changing world," *J. Appl. Probab.*, vol. 25, pp. 287–298, Jan. 1988.

[10] J. Niño-Mora, "Dynamic priority allocation via restless bandit marginal productivity indices," *Top*, vol. 15, no. 2, pp. 161–198, 2007.

[11] R. R. Weber and G. Weiss, "On an index policy for restless bandits," *J. Appl. Probab.*, vol. 27, no. 3, pp. 637–648, Sep. 1990.

[12] I. M. Verloop, "Asymptotically optimal priority policies for indexable and non-indexable restless bandits," *Ann. Appl. Probab.*, to be published.

[13] S. Aalto, P. Lassila, and P. Osti, "Whittle index approach to size-aware scheduling with time-varying channels," in *Proc. ACM SIGMETRICS*, 2015, pp. 57–69.

[14] I. Taboada, U. Ayesta, P. Jacko, and F. Liberal, "Opportunistic scheduling of flows with general size distribution in wireless time-varying channels," in *Proc. 26th Int. Teletraffic Congr.*, Sep. 2014, pp. 1–9.

[15] K. D. Glazebrook, D. J. Hodge, C. Kirkbride, and R. J. Minty, "Stochastic scheduling: A short history of index policies and new approaches to index generation for dynamic resource allocation," *J. Scheduling*, vol. 17, no. 5, pp. 407–425, Oct. 2014.

[16] K. D. Glazebrook, C. Kirkbride, and D. Ruiz-Hernandez, "Spinning plates and squad systems: Policies for bi-directional restless bandits," *Adv. Appl. Probab.*, vol. 38, no. 1, pp. 95–115, Mar. 2006.

[17] F. Avram, D. Bertsimas, and M. Ricard, "Optimization of multiclass queueing networks: A linear control approach," in *Stochastic Networks; Proceedings of the IMA*, New York, NY, USA: Springer, 1994, vol. 71, pp. 199–234.

[18] G. Weiss, "On optimal draining of re-entrant fluid lines," in *Stochastic Networks*, F. P. Kelly and R. Williams, Eds. Springer, 1995, pp. 91–103.

[19] A. Wierman, L. L. H. Andrew, and A. Tang, "Power-aware speed scaling in processor sharing systems," in *Proc. IEEE INFOCOM*, Apr. 2009, pp. 2007–2015.

[20] R. Atar, C. Giat, and N. Shimkin, "The $c\mu/\theta$ rule for many-server queues with abandonment," *Oper. Res.*, vol. 58, no. 5, pp. 1427–1439, Aug. 2010.

[21] K. D. Glazebrook, C. Kirkbride, and J. Ouenniche, "Index policies for the admission control and routing of impatient customers to heterogeneous service stations," *Oper. Res.*, vol. 57, no. 4, pp. 975–989, Mar. 2009.

[22] M. Larrañaga, U. Ayesta, and I. M. Verloop, "Index policies for a multi-class queue with convex holding cost and abandonments," in *Proc. ACM SIGMETRICS*, 2014, pp. 125–137.

[23] S. Borst, "User-level performance of channel-aware scheduling algorithms in wireless data networks," *IEEE/ACM Trans. Netw.*, vol. 13, no. 3, pp. 636–647, Jun. 2005.

[24] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. New York, NY, USA: Wiley, 2005.

[25] M. Larrañaga, U. Ayesta, and I. M. Verloop, "Asymptotically optimal index policies for an abandonment queue with convex holding cost," *Queueing Syst.*, vol. 81, nos. 2–3, pp. 99–169, Nov. 2015.

[26] M. H. Veatch and L. M. Wein, "Scheduling a make-to-stock queue: Index policies and hedging points," *Oper. Res.*, vol. 44, no. 4, pp. 634–647, 1996.

[27] W. Ouyang, A. Eryilmaz, and N. B. Shroff, "Asymptotically optimal downlink scheduling over Markovian fading channels," in *Proc. IEEE INFOCOM*, Mar. 2012, pp. 1224–1232.

[28] B. Ji, G. Gupta, M. Sharma, X. Lin, and N. B. Shroff, "Achieving optimal throughput and near-optimal asymptotic delay performance in multichannel wireless networks with low complexity: A practical greedy scheduling policy," *IEEE/ACM Trans. Netw.*, vol. 23, no. 3, pp. 880–893, Jun. 2015.

[29] D. P. Bertsekas, *Dynamic Programming and Optimal Control*. Belmont, MA, USA: Athena Scientific, 2005.

Maialen Larrañaga, received the master's degree in mathematics from the University of the Basque Country in 2012, and the Ph.D. degree in computer science from the Institute National Polytechnique (INP) de Toulouse and University of the Basque Country in 2015. The thesis was with CNRS, LAAS, and INP-ENSEEIHT under the supervision of Prof. Urtzi Ayesta (CNRS, LAAS) and Ina Maria Verloop (CNRS, IRIT) with a Fondation AIRBUS Group scholarship. Her project was developed in BCAM in collaboration with INGETEAM S.A. Since 2015, she holds a post-doctoral position with Laboratoire des Signaux et Systèmes, CentraleSupeléc.

Urtzi Ayesta, received the B.S. and M.S. degrees in telecomunication engineering from Nafarroako Unibertsitate Publikoa-Universidad Publica de Navarra, Spain, the M.S. degree in electrical engineering from Columbia University, USA, and the Ph.D. degree from Université de Nice-Sophia Antipolis, France. His Ph.D. research work was the research laboratories of INRIA Sophia-Antipolis and France Telecom R&D. He is currently a CNRS Researcher with LAAS, Toulouse, France, and he also holds an Adjunct Lecturer position (part-time appointment funded by Ikerbasque) with the Computer Science Faculty, University of the Basque Country.

Ina Maria Verloop, received the M.Sc. degree in mathematics from Utrecht University, The Netherlands, in 2005, and the Ph.D. degree from the Mathematics and Computer Science Department, Eindhoven University of Technology, in 2009. Her Ph.D. research was carried with the Probability, Networks and Algorithms Department, Center for Mathematics and Computer Science, Amsterdam, The Netherlands. From 2009 to 2011, she held a post-doctoral position with the Basque Center for Applied Mathematics, Derio, Spain. Since 2011, she has been a CNRS Researcher, IRIT, Toulouse, France. Her research interests are in scheduling, queueing theory, and stochastic optimization and their application to the performance analysis and optimization of communication networks.