Stochastic and fluid index policies for resource allocation problems

M. Larrañaga[1,2,5], U. Ayesta[2,3,4,5], I.M. Verloop[1,5]
[1]CNRS, IRIT, 2 rue C. Carmichel, F-31071 Toulouse, France.
[2]CNRS, LAAS, 7 avenue du colonel Roche, F-31400 Toulouse, France
[3]IKERBASQUE, Basque Foundation for Science, 48011 Bilbao, Spain
[4]UPV/EHU, University of the Basque Country, 20018 Donostia, Spain
[5]Univ. de Toulouse, INP, LAAS, F-31400 Toulouse, France

***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. It can be efficiently calculated, however, it requires several technical conditions to be verified, and in addition, it does not provide qualitative insights into Whittle's index. 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 is close to Whittle's index. 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. Numerical simulations show that Whittle's index and our fluid index policy are both nearly optimal.**

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, 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, [1]. 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 [2], [3], [4], [5]. 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 [6].

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 [7]. 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 [8], Whittle introduced the so-called Restless BP (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

The PhD fellowship of Maialen Larrañaga is funded by a research grant of the Foundation Airbus Group (http://fondation.airbus-group.com/).

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 [9] for a discussion, and is asymptotically optimal under certain conditions, [10]. 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 opportunistic spectrum access, website morphing and pharmaceutical trials, [7, Chapter 6]. The recent survey paper [11] 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 RMABP 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.

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. However it does not allow to obtain qualitative insights. 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. [12] and Weiss [13]. 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.

We illustrate the applicability of our approach with two important 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 and (ii) load balancing in a power-aware server farm. Numerical experiments show that for both 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 the performance of Whittle's index policy and the fluid index policy is numerically evaluated. Due to lack of space, some proofs are omitted. They may be found in the full version of the paper [14].

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

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

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 *restless bandit problem*, see [7]. 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

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

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 $a=1$).

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 [15], 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 [16], [17], [18]. 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 V.

The first class of problems concerns the multi-class setting of Figure 1. The objective is to determine which M classes 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 V 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 [19]. We further note that when $M=1$ and $\mu_k(N_k)=\mu_k$, this model captures the classical single-server multi-class queue.

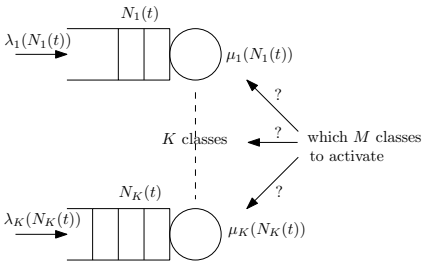

Fig. 1. A multi-class system where M classes can be simultaneously served

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 V we investigate how to optimally dispatch users in a power-aware server farm, where the capacity of servers follows a speed-scaling rule.

B. Optimal control

The objective of this paper is to find scheduling policies

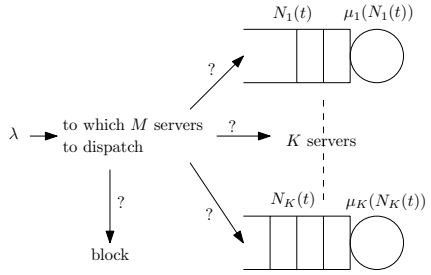

Fig. 2. Load balancing in a multi-server system

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 [20] 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

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

a stationary policy that realizes the minimum in (4) is optimal, [20]. 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 instances. Numerically, optimal policies can be found using Value Iteration or Policy Improvement algorithms. However, the curse of dimensionality renders infeasible to find the solution even for very small instances of the problem.

For certain examples it is possible to explicitly solve (4) and to characterize the optimal stochastic control. An important class of problems for which this is possible is known as the *multi-armed bandit* problem. 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 $b_k^1(n_k)\ge 0$, $d_k^1(n_k)\ge 0$. The optimal solution of (3) has a simple structure, known as Gittins' index policy, see [7]. In brief, there exist functions $G_k(n_k)$, depending only on the parameters of bandit k, such that the 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 $(b_k^0(n_k), d_k^0(n_k)\ge 0)$, 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 [8], 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 [8], 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,

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

The objective of the relaxed problem is hence to determine the 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

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

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:

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

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

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

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) is an MDP as well and the optimal policy is the solution of the Dynamic Programming equation

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

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 [8] 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 an optimal policy for (7) is indifferent of the action in state N_k. The Whittle's index is denoted by $W_k(N_k)$.

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., the 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 $a\neq a'$. 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], [17], [18]. Further examples can be found in [7, Section 6.5].

In the next proposition we show that when optimality of threshold policies can be established, then indexability is satisfied under a condition on the steady-state probabilities of threshold policies. In addition, we derive a closed-form expression for Whittle's index, which is expressed as a function of these steady-state probabilities.

Proposition 1: 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.

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 [8] proposed the following heuristic, which is referred to as Whittle's index policy. In Section V 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, see Proposition 1. 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 the 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 $f_k^i(m_k)$ to be non-increasing in m_k for $i \in \{0, 1\}$. 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 $C_k(m_k(t), s_k(t)) = (1 - s_k(t))C_k(m_k(t), 0) + s_k(t)C(m_k(t), 1) - W(1 - s_k(t))$, 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, $(1 - \bar{s}_k) f_k^0(\bar{m}_k) + \bar{s}_k f_k^1(\bar{m}_k) = 0$, with $\bar{s}_k \in [0, 1]$. That is, in equilibrium, a fraction of time $\bar{s}$ $(1 - \bar{s})$ the action $a = 1$ $(a = 0)$ is chosen.

In the stochastic model we aim to minimize for a given bandit the relaxed optimization problem, that is, we minimize the time-average of the cost minus the subsidy obtained, as stated in (7). 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 $(1 - \bar{s}_k)C_k(\bar{m}_k, 0) + \bar{s}_k C(\bar{m}_k, 1) - W(1 - \bar{s}_k)$. 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 will be attained by several controls, see [20, Chapter 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, [20]:

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

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 2, while solving (7) (or equivalently (9)) requires to establish that an optimal policy for the relaxed problem is of threshold structure.

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

$W(1-s_k^u(t)))\mathrm{d}t$. Hence, an optimal control for the relaxed fluid discounted control problem is the solution of

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

see [20, Chapter 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 [20, Corollary 8.2.5], and we thus obtain that (15) converges to (14).

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 V to perform nearly optimal.

In order to give the statement of 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, and that $m_k^i=0$ in case $f_k^i(m_k)<0$ for all m_k. 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. In the following proposition we give the expression for the fluid index and state an optimal solution of the fluid problem (13).

Proposition 2: Assume $C_k(m_k,a)$ and $f_k^a(m_k)$, $a=0,1$, are differentiable in m_k on $[\min(m_k^0,m_k^1),\max(m_k^0,m_k^1)]$ and $f_k^0(m_k)/(f_k^0(m_k)-f_k^1(m_k))$ convex in m_k on $[\min(m_k^0,m_k^1),\max(m_k^0,m_k^1)]$. We define

$$w_k^{(1)}(m_k) = (f_k^1(m_k)-f_k^0(m_k))\frac{C_k(m_k,1)-C_k(m_k^1,1)}{f_k^1(m_k)},$$
$$w_k^{(2)}(m_k) = \frac{(f_k^1(m_k)-f_k^0(m_k))(f_k^0(m_k)\frac{\mathrm{d}C_k(m_k,1)}{\mathrm{d}m_k}-f_k^1(m_k)\frac{\mathrm{d}C_k(m_k,0)}{\mathrm{d}m_k})}{f_k^0(m_k)\frac{\mathrm{d}f_k^1(m_k)}{\mathrm{d}m_k}-f_k^1(m_k)\frac{\mathrm{d}f_k^0(m_k)}{\mathrm{d}m_k}},$$
$$w_k^{(3)}(m_k) = (f_k^1(m_k)-f_k^0(m_k))\frac{C_k(m_k,0)-C_k(m_k^0,0)}{f_k^0(m_k)}.$$

If $m_k^0>m_k^1$, we define the continuous function

$$w_k(m_k) = C_k(m_k,0)-C_k(m_k,1) + \begin{cases} w_k^{(1)}(m_k) & \text{if } m_k<m_k^1,\\ w_k^{(2)}(m_k) & \text{if } m_k\in[m_k^1,m_k^0],\\ w_k^{(3)}(m_k) & \text{if } m_k>m_k^0.\end{cases}$$

If $\mathrm{d}C_k(m,0)/\mathrm{d}m \geq \mathrm{d}C_k(m,1)/\mathrm{d}m$, and $w_k^{(i)}(m_k)$, $i=1,2,3$, 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 $W>w_k(m_k)$.

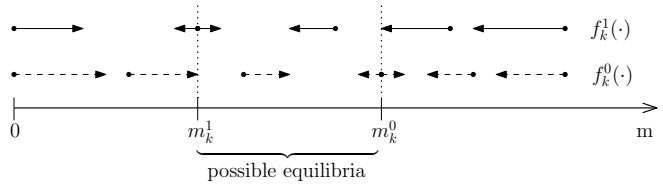

Fig. 3. Representation of fluid equilibria and drift functions when $m_k^1<m_k^0$.

If $m_k^1>m_k^0$, we define the continuous function

$$w_k(m_k) = C_k(m_k,0)-C_k(m_k,1) + \begin{cases} w_k^{(3)}(m_k) & \text{if } m_k<m_k^0,\\ w_k^{(2)}(m_k) & \text{if } m_k\in[m_k^0,m_k^1],\\ w_k^{(1)}(m_k) & \text{if } m_k>m_k^1.\end{cases}$$

If $\mathrm{d}C_k(m,1)/\mathrm{d}m \geq \mathrm{d}C_k(m,0)/\mathrm{d}m$, and $w_k^{(i)}(m_k)$, $i=1,2,3$, is non-increasing 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 $W>w_k(m_k)$.

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

We observe from Proposition 2 that monotonicity of $w_k(m)$ in m 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.

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 $D_k(W)=\{m_k : W\geq w_k(m_k)\}$, see Definition 1. This as opposed to the stochastic model, for which indexability needs to be verified independently.

Monotonicity of $w_k(m)$ is a simple property to verify. This represents a huge 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 V we will show the monotonicity of the fluid index to be satisfied for two 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 $\alpha\geq 0$, respectively. Then,

- when $m_k^1<m_k^0$, the fluid index $w_k(m_k)$ is non-decreasing for all m_k, if $f_k^1(m_k)-f_k^0(m_k)<0$ for all m_k and $\alpha<P$,
- when $m_k^1>m_k^0$, the fluid index $w_k()$ is non-increasing for all m_k, if $f_k^1(m_k)-f_k^0(m_k)>0$ for all m_k and $\alpha<P$.

Proof: 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 2, and using that $f_k^i(\cdot)$ is non-increasing. ■

Remark 2: The generality of our approach is illustrated by the fact that when applied to classical problems, it retrieves well-known index policies. For instance, it can be verified that in the case of a multi-class queue with linear holding costs our fluid index becomes the optimal $c\mu$-rule, while for convex holding costs it coincides with the Generalized $c\mu$-rule (introduced and heavy-traffic optimality established in [21]). 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 [16]).

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 4 (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 V 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. Case studies

In this section we evaluate both the stochastic and fluid index policies for birth-and-death bandits. The main advantage of these 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 two decision making problems: (i) opportunistic scheduling in a wireless downlink channel, which belongs to the class of problems depicted in Figure 1, and (ii) optimal blocking/routing in a power-aware server farm, which belongs to the class of problems depicted in Figure 2. In both cases, we compare the performance of Whittle's index policy (10) and the fluid index policy, as given in Proposition 2, against the optimal policy, which is computed using the Value Iteration approach, see [20]. 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 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 [19]. 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: $b_k^a(n_k) = \lambda_k$, and $d_k^a(n_k) = \mu_k \frac{n_k}{n_k+1} a$, where $a=1$ ($a=0$) stands for serving (not serving) class k, see Figure 1. In order for the system to be 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*.

Assuming an optimal solution of the relaxed optimization problem (7) is of threshold type 0-1, the Whittle index as given in (10) can be numerically computed, where 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 $\pi_k^{n_k}(m_k) = 0$, $\forall m_k \le n_k - 1$, $\pi_k^{n_k}(m_k) = \left(\frac{\lambda_k}{\mu_k}\right)^{m_k-n_k} \frac{m_k+1}{n_k+1}\pi^{n_k}(n_k)$, $\forall m_k \ge n_k+1$, and $\pi_k^{n_k}(n_k) = 1/\left(1+\frac{1}{n_k+1}\sum_{i=1}^{\infty}\left(\frac{\lambda_k}{\mu_k}\right)^i (n_k+1+i)\right)$. It can be checked that $\sum_{i=0}^n \pi_k^n(i)$ is strictly increasing in n as required for indexability, see Proposition 1.

Besides the fact that the threshold structure still needs to be established, the expression in (10) for the Whittle index does not help to obtain 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 is 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 2 we can now derive the fluid index, which describes the policy that minimizes the bias-optimal criteria as given in (13).

Proposition 3: Assume $C_k(m_k,a)$ is convex and non-decreasing, and $C_k(m_k,0) - C_k(m_k,1)$ and $\frac{\mathrm{d}C_k(m_k,1)}{\mathrm{d}m_k} - \frac{\mathrm{d}C_k(m_k,0)}{\mathrm{d}m_k}$ are non-decreasing. Then, the fluid index is non-decreasing and given by:

$$w_k(m_k) = C_k(m_k,0) - C_k(m_k,1) + \begin{cases} w_k^{(1)}(m_k) & \text{if } m_k < \lambda_k/(\mu_k-\lambda_k), \\ w_k^{(2)}(m_k) & \text{if } \lambda_k/(\mu_k-\lambda_k) \le m_k, \end{cases} \quad (16)$$

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

Proof: Equation (16) follows from Proposition 2. Non-decreasingness 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. ■

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

Having a closed-form expression for the fluid index as given in (16) gives us insights on the behavior of the system with respect to the parameters involved. For the sake of clarity

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

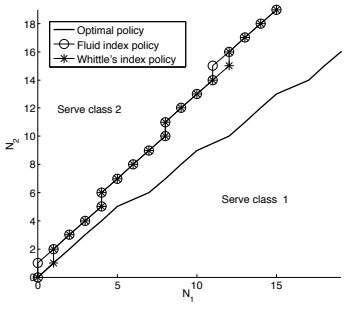

Fig. 4. Switching curves under the optimal policy, Whittle's index policy and fluid index policy.

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 close to the origin priority is given according to $c_k m_k/(1-\rho_k)$ and far from it according to $c_k\mu_k m_k^2/\delta_k$.

Example 1. Let us assume 2 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 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.

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 [15] in order to balance between power consumption and server capacity. We assume that when in state N_k, the service capacity is $c(N_k) := \min(T, N_k^\alpha)$, with $\alpha > 0$, where $T > 0$ represents the maximum capacity of the server. The service requirement of a user in station k is exponentially distributed with mean $1/\mu_k$. Hence, the departure rate is $\mu_k(N_k) = \mu_k \cdot \min(T, N_k^\alpha)$.

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 $\min(T, N_k^\alpha)^{1/\alpha}$. We therefore take as cost $C_k(N_k, a) = C_k(N_k) + \beta_k \min(T, N_k^\alpha)^{1/\alpha} + D\lambda(1-a)$, 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 $C_k(N_k, a)$ 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: $b_k^a(m_k) = \lambda a$, and $d_k^a(m_k) = \mu_k \min(T, m_k^\alpha)$, where $a = 0$ ($a = 1$) stands for blocking (accepting) a user in server k. We first determine the fluid index policy for this model. The fluid dynamics is given by $\frac{\mathrm{d}m_k(t)}{\mathrm{d}t} = \lambda s_k(t) - \mu_k \min(T, m_k^\alpha)$, with $s_k(t) \in \{0,1\}$. In case $T > \lambda/\mu_k$, we have $m_k^0 = 0$, and $m_k^1 = (\lambda/\mu_k)^{1/\alpha}$, that is, the equilibrium points are in the interval $\bar{m}_k \in [0, m_k^1]$. Hence, taking $T > \lambda/\mu_k$ we focus on the interesting case where in equilibrium not the full capacity is used, that is, speed scaling plays a role. We derive the fluid index in the following proposition, that follows from Proposition 2 and Lemma 1.

Proposition 4: Assume $T > \lambda/\mu_k$ and let $C_k(m_k)$ be a polynomial of degree P with $P > \alpha$. Then, the fluid index is non-increasing and given by:

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

where $w_k^{(2)}(m_k) = -\frac{\lambda\alpha^{-1}m_k}{\mu_k m_k^\alpha}\frac{\mathrm{d}\tilde{C}_k(m_k)}{\mathrm{d}m_k}$ and $w_k^{(1)}(m_k) = -\lambda\frac{(\tilde{C}_k((\lambda/\mu_k)^{\alpha^{-1}}) - \tilde{C}_k(m_k))}{\lambda - \mu_k \min(T, m_k^\alpha)}$, with $\tilde{C}_k(m_k) = C_k(m_k) + \beta_k \min(T, m_k^\alpha)^{1/\alpha}$.

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 $N_k \geq \overline{N}_k$, no users will be routed to this server k.

As in the previous section we use Proposition 4 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 $c_k \frac{m_k^{1-\alpha}}{\mu_k \alpha}$.

Note that the optimal structure of the fluid version of the relaxed optimization problem is of 1-0 structure (since the fluid index is non-increasing). We therefore assume that an optimal solution of the stochastic relaxed optimization problem (7) is

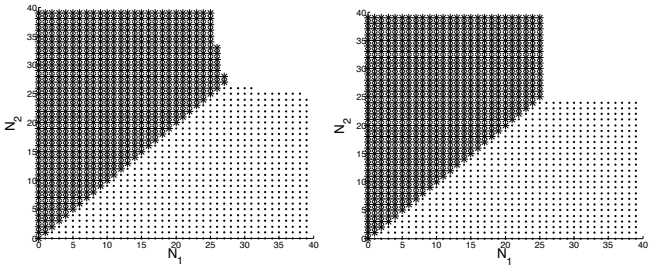

Fig. 5. In the area with "." ("*") class 1 (class 2) is prioritized and in the white area users are blocked. Left: Optimal policy. Right: Whittle index policy and the fluid index policy.

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}

of threshold type 1-0 as well, and hence Whittle's index can be numerically computed, as explained in Proposition 1.

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 $\alpha = 1/2$ and $T = \infty$. 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 $\rho = 2.5$. The fluid index policy coincides with the strategy given by Whittle's index policy and captures the qualitative structure of the optimal policy.

VI. 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 [22], [10], [23].

REFERENCES

[1] C. Buyukkoc, P. Varaya, and J. Walrand, "The $c\mu$ rule revisited," *Adv. Appl. Prob.*, vol. 17, pp. 237–238, 1985.

[2] P. S. Ansell, K. D. Glazebrook, J. Nino-Mora, and M. O'Keeffe, "Whittle's index policy for a multi-class queueing system with convex holding costs," *Mathematical Methods of Operations Research*, vol. 57, no. 1, pp. 21–39, 2003.

[3] C. Bispo, "The single-server scheduling problem with convex costs," *Queueing Systems*, vol. 73, pp. 261–294, 2013.

[4] J. George and J. Harrison, "Dynamic control of a queue with adjustable service rate," *Operations Research*, vol. 49, no. 5, pp. 720–731, 2001.

[5] M. Larrañaga, U. Ayesta, and I. M. Verloop, "Dynamic fluid-based scheduling in a multi-class abandonment queue," *Perf. Eval., Proc. of IFIP Performance 2013.*, vol. 70, no. 10, pp. 841–858, 2013.

[6] L. Schrage and L. Miller, "The queue M/G/1 with the shortest remaining processing time discipline," *Oper. Res.*, vol. 14, pp. 670–684, 1966.

[7] J. Gittins, K. Glazebrook, and R. Weber, *Multi-armed Bandit Allocation Indices*. Wiley, 2011.

[8] P. Whittle, "Restless bandits: Activity allocation in a changing world," *Journal of Applied Probability*, vol. 25, pp. 287–298, 1988.

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

[10] R. Weber and G. Weiss, "On an index policy for restless bandits," *Journal of Applied Probability*, vol. 27, pp. 637–648, 1990.

[11] K. Glazebrook, D. Hodge, C. Kirkbride, and R. Minty, "Stochastic scheduling: A short history of index policies and new approaches to index generation for dynamic resource allocation," *Journal of Scheduling*, pp. 1–19, 2013.

[12] F. Avram, D. Bertsimas, and M. Richard, *Stochastic networks; proceedings of the IMA*, ch. Optimization of multiclass queuing networks: a linear control approach, pp. 199–234. 1994.

[13] G. Weiss, "On optimal draining of reentrant fluid lines," *Stochastic Networks, eds. F.P. Kelly and R.J. Williams*, pp. 91–103, 1995.

[14] M. Larrañaga, U. Ayesta, and I. M. Verloop, "Stochastic and fluid index policies for resource allocation problems," *LAAS report 15001*.

[15] A. Wierman, L. Andrew, and A. Tang, "Power-aware speed scaling in processor sharing systems," in *Proceedings of IEEE INFOCOM*, 2009.

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

[17] K. Glazebrook, C. Kirkbride, and J. Ouenniche, "Index policies for the admission control and routing of impatient customers to heterogeneous servcie stations," *Operations Research*, vol. 57, pp. 975–989, 2009.

[18] M. Larrañaga, U. Ayesta, and I. M. Verloop, "Index policies for a multi-class queue with convex holding cost and abandonment," in *Proceedings of ACM SIGMETRICS*, 2014.

[19] S. C. Borst, "User-level performance of channel-aware scheduling algorithms in wireless data networks," *IEEE/ACM Transactions on Networking*, vol. 13, no. 3, pp. 636–647, 2005.

[20] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, 2005.

[21] J. A. van Mieghem, "Dynamic scheduling with convex delay costs: The generalized $c\mu$ rule," *The Annals of Applied Probability*, vol. 5, no. 3, pp. 808–833, 1995.

[22] W. Ouyang, A. Eryilmaz, and N. Shroff, "Asymptotically optimal downlink scheduling over Markovian fading channels," *Proceedings of IEEE INFOCOM*, pp. 1–9, 2012.

[23] I. M. Verloop, "Asymptotic optimal control of multi-class restless bandits." HAL Report, available at http://hal.archives-ouvertes.fr/hal-00743781, Aug. 2014.