Index Policies for the Admission Control and Routing of Impatient Customers to Heterogeneous Service Stations

Citation for published version:
Glazebrook, K, Kirkbride, C & Ouenniche, J 2009, 'Index Policies for the Admission Control and Routing of Impatient Customers to Heterogeneous Service Stations', *Operations Research*, vol. 57, no. 4, pp. 975-989. https://doi.org/10.1287/opre.1080.0632

Digital Object Identifier (DOI):
10.1287/opre.1080.0632

Published In:
Operations Research

OPERATIONS RESEARCH
Vol. 57, No. 4, July–August 2009, pp. 975–989
ISSN 0030-364X | EISSN 1526-5463 | 09 | 5704 | 0975

DOI 10.1287/opre.1080.0632

Index Policies for the Admission Control and Routing of Impatient Customers to Heterogeneous Service Stations

K. D. Glazebrook
Department of Mathematics and Statistics, and Department of Management Science,
Lancaster University, Lancaster, LA1 4YX, United Kingdom, k.glazebrook@lancs.ac.uk

C. Kirkbride
Department of Management Science, Lancaster University, Lancaster, LA1 4YX, United Kingdom,
c.kirkbride@lancaster.ac.uk

J. Ouenniche
Business School, University of Edinburgh, Edinburgh, EH8 9JY, United Kingdom,
jamal.ouenniche@ed.ac.uk

We propose a general Markovian model for the optimal control of admissions and subsequent routing of customers for service provided by a collection of heterogeneous stations. Queue-length information is available to inform all decisions. Admitted customers will abandon the system if required to wait too long for service. The optimisation goal is the maximisation of reward rate earned from service completions, net of the penalties paid whenever admission is denied, and the costs incurred upon every customer loss through impatience. We show that the system is indexable under mild conditions on model parameters and give an explicit construction of an index policy for admission control and routing founded on a proposal of Whittle for restless bandits. We are able to gain insights regarding the strength of performance of the index policy from the nature of solutions to the Lagrangian relaxation used to develop the indices. These insights are strengthened by the development of performance bounds. Although we are able to assert the optimality of the index heuristic in a range of asymptotic regimes, the performance bounds are also able to identify instances where its performance is relatively weak. Numerical studies are used to illustrate and support the theoretical analyses.

Subject classifications: admission control; customer impatience; dynamic programming; index policies; Markov decision processes; monotone policies; restless bandits; routing.
Area of review: Manufacturing, Service, and Supply Chain Operations.
History: Received August 2004; revisions received February 2006, October 2007, June 2008; accepted June 2008. Published online in *Articles in Advance* March 30, 2009.

1. Introduction

Customers arrive seeking service, which may be provided at one of a heterogeneous collection of service stations. At each customer arrival, queue-length information for each station is available to inform a decision concerning whether the customer should be admitted for service at all and, if admitted, at which station she should best be served. However, the customers are impatient or perishable and have a natural lifetime of availability for useful service. For any admitted customer, should that lifetime expire before the customer has been served to completion (or, possibly, before she enters service), then she is lost from the system. Rewards are earned by successful service completions, whereas costs are incurred when customers are either discarded (not admitted) or lost from the system. The goal of optimisation is to control admissions to the system and to route admitted customers in a way that maximises some measure of net reward rate earned.

Gaver and Jacobs (2000) have argued the importance of incorporating customer impatience into service system models. They cite examples of telephone callers placed on hold who are prone to hang up, medical emergency patients who may die while awaiting treatment, and military scenarios in which mobile targets may move out of range while under attack. They also argue that in such contexts a customer's lifetime for service will usually be unknown to the system controller (and possibly also to the customer herself). This is in contrast to situations where tasks are to be scheduled in the face of known hard deadlines. (See Jiang et al. 1996; Lehoczky 1996, 1997a, b; Doytchinov et al. 2001 for examples of the latter.) Consider, for example, a military scenario in which a Blue force is defending a region from attack by an opposing Red force. As Red combatants enter the region and are detected, they are allocated to a member of the Blue force (or, possibly, to a combat group comprising several Blues) for engagement. The presence of a Red in the region gives the Blue defensive force a window of opportunity of unknown duration in which to effect a kill. As soon as Red leaves the region, the opportunity has gone. The goal for the Blue force is the maximisation of its kill rate or some measure of return rate from Reds killed net of the penalties incurred when Reds escape unharmed. The diverse nature of the Blue force may mean, for example, that the speed with which kills can be effected may differ considerably across force members.

Despite the pervasiveness of the phenomenon of customer impatience, there have until recently been few published studies whose primary goal is the optimisation of service system performance and which have attempted to take explicit account of it. Movaghar (1997) has established the optimality of a form of "join the shortest queue" (JSQ) in the context of a finite-capacity queueing system with Markovian dynamics and identical single-server stations. Motivated by applications in call centres, Bassamboo et al. (2005) have discussed routing and admission control in high-volume systems. They give an asymptotic analysis that makes use of a stochastic fluid approximation. Garnett et al. (2002) give an account of the impact of customer impatience on call centre design. Both Glazebrook et al. (2004) and Harrison and Zeevi (2004) have studied the dynamic allocation of service effort in the face of customer losses. Ward and Glynn (2005) develop a heavy-traffic diffusion limit for a $GI/GI/1$ queue with balking or reneging.

Even without the issue of premature customer departures, problems concerning the optimal dynamic routing of customers for service present a formidable challenge to analysis. Probably the strongest line of contributions have been those seeking to establish the optimality of JSQ with respect to a range of performance objectives in the context of systems of identical service stations. Important contributions include those of Winston (1977), Weber (1978), Hordijk and Koole (1990), Menich and Serfozo (1991), and Koole et al. (1999). Armony (2005) discusses the performance of a simple routing rule in a heavy-traffic limit and gives an extensive bibliography of related contributions. Additionally, Krishnan (1987), Tijms (1994), Whittle (1996), and Ansell et al. (2003a) have developed routing policies using dynamic programming policy improvement.

Recently, Whittle (1996) followed by Niño-Mora (2002) have proposed the use of methodologies related to the class of so-called *restless bandit problems* to solve routing problems. This proposal is not without its difficulties because restless bandit problems are now known to be intractable; see Papadimitriou and Tsitsiklis (1999). Whittle's (1988, 1996) approach to the analysis of restless bandits uses a Lagrangian relaxation of the original optimisation problem to develop an index. The latter generalises the classical Gittins index (1979), which yields optimal policies for the class of *multiarmed bandit problems*. Whittle proposes that his index be used to construct a natural class of heuristics for restless bandit problems. Weber and Weiss (1990, 1991) have proved the asymptotic optimality of this heuristic under given conditions, whereas Glazebrook et al. (2002), Ansell et al. (2003b), and Glazebrook et al. (2005) have demonstrated empirically its strong performance in a variety of application contexts. Further, Glazebrook et al. (2002) have discussed the development of bounds on the degree of reward suboptimality of Whittle's index policy.

However, the challenge to successful implementation of Whittle's ideas is substantial. For Whittle's index to be properly defined, the object to which the index attaches (which in the case of routing problems is a service station with a given head count) must pass an *indexability* test. Determination of whether the test is passed is far from straightforward in general. Niño-Mora (2001) has developed sufficient conditions by the use of polyhedral methods that exploit the fact that the system concerned can be shown to satisfy *partial conservation laws*. The Whittle indices themselves emerge from the deployment of an *adaptive greedy algorithm*. All of this can be viewed as a development of the polyhedral analysis of multiarmed bandit problems given by Bertsimas and Niño-Mora (1996). Niño-Mora (2002) further developed these methods with a view to their application to routing problems.

Although acknowledging the power of the polyhedral approach, it is formidably difficult and necessarily indirect. The authors would argue that rather simpler and more direct accounts of indexability are often available. We would further argue, however, that polyhedral methods certainly have an important contribution to make to the *performance analysis* of routing heuristics. In demonstration of these points, we give a simple direct account of a general Markovian model for dynamic admission control/routing to heterogeneous service stations in the face of customer impatience. For these models, indexability is easily established and is close to guaranteed. Only mild conditions on system parameters are needed. The Whittle indices that define the heuristic are direct products of the indexability analysis and, in many cases of interest, can be given explicitly. In some simple cases, the index heuristic modifies a naïve *individually optimal* proposal by taking due account of the impact of routing decisions on system congestion.

Following the indexability analysis, we use the nature of solutions to the Lagrangian relaxation (used to develop the indices) to gain insights regarding the strength of performance of the index heuristic. Although tightness of the Lagrangian relaxation is sufficient for strong performance of the index heuristic, it is not necessary. Stronger performance analyses become available when we deploy polyhedral methods to develop performance bounds for the index heuristic. This is in the spirit of Glazebrook and Niño-Mora (2001) and Glazebrook et al. (2002). Under given conditions, we are able to assert the optimality of the index heuristic in a range of asymptotic regimes. These include simple light and heavy-traffic limits (Examples 4 and 5, respectively, in the paper) and a limit in which the customer arrival rate and the station service rates grow in scale (Example 11). The performance bounds are also able to identify cases where the performance of the index heuristic is relatively weak (Examples 8 and 10).

The remainder of this paper is structured as follows. In §2, we give details of a general class of admission control/routing problems that incorporate customer impatience.

An account of Whittle's proposal in the form in which it will be applied in the paper is given in §3. In §4, we show that the indexability test is passed and give the Whittle indices that determine our heuristic. Section 5 gives a discussion of what insights may be gleaned from the Lagrangian relaxation used in §3 concerning the strength of performance of the index policy. In §6, we present performance bounds for the index heuristic and use them to establish its optimality in a range of asymptotic regimes. Section 7 contains a proof of the heavy-traffic optimality of the index policy. Numerical studies are used to illustrate and support the material in §§5–7. To improve the paper's readability, many of the mathematical details have been placed in an electronic companion, which is part of the online version that can be found at http://or.pubs.informs.org/.

2. The Model

Our problem of admission control/routing impatient customers for service is modelled as a semi-Markov decision process (SMDP) as follows:

(i) The system is observed at all times $t \in \mathbb{R}^+$. Its *state* at t, $\boldsymbol{\mathcal{N}}(t)$, is an M-vector whose mth component is $\mathcal{N}_m(t)$ and is the number of customers, hereafter referred to as the *head count*, waiting at station m (including any in service) at t, $1 \leqslant m \leqslant M$.

(ii) New customers arrive into the system according to a Poisson process of rate λ. Upon arrival, each customer must be either refused admission or routed to one of the M stations for service. Admission and routing decisions are *nonanticipative* and are made in the light of the history of the system (states occupied and actions taken) to date.

(iii) Between arrivals, the state of each station evolves as a continuous-time Markov process. Each station m has an associated sequence of service rates $\{\mu_{m,n}; n \in \mathbb{Z}^+\}$ and loss rates $\{\theta_{m,n}; n \in \mathbb{Z}^+\}$, $1 \leqslant m \leqslant M$. Hence, if the system is currently in state $\boldsymbol{\mathcal{N}} = \{\mathcal{N}_1, \mathcal{N}_2, \ldots, \mathcal{N}_M\}$, then an exponentially distributed amount of time will elapse before the next random event occurs, whether an arrival, a service completion, or a loss from the system. In state $\boldsymbol{\mathcal{N}}$, the rate of this exponential time is given by

$$\lambda + \sum_{m=1}^{M} (\mu_{m,\mathcal{N}_m} + \theta_{m,\mathcal{N}_m}) \equiv \Delta(\boldsymbol{\mathcal{N}}).$$

Further, the next random event will be an arrival, a service completion at station m, or a loss at station m with probabilities $\lambda\{\Delta(\boldsymbol{\mathcal{N}})\}^{-1}$, $\mu_{m,\mathcal{N}_m}\{\Delta(\boldsymbol{\mathcal{N}})\}^{-1}$, and $\theta_{m,\mathcal{N}_m}\{\Delta(\boldsymbol{\mathcal{N}})\}^{-1}$, respectively. Should the event be a service completion at station m, then a positive reward R_m is received, $1 \leqslant m \leqslant M$. A loss (renege) from station m will incur a nonnegative penalty C_m, $1 \leqslant m \leqslant M$. Any refusal to admit to the system (discard) incurs a nonnegative penalty D, where $D \leqslant \max_m C_m$.

Example 1. Suppose that station m has s_m identical servers working in parallel. Service times are independent and identically distributed (i.i.d.) exponential with positive rate μ_m. Further, each admitted customer has a natural lifetime during which she is available for service. These lifetimes are i.i.d. $\exp(\theta_m)$, where $\theta_m > 0$, and are independent of all other aspects of the system. If a customer's lifetime expires before her service is completed, she is lost to the system. Otherwise, her service is completed successfully and a reward is earned. For such a system, we have

$$\mu_{m,n} = \mu_m \min(n, s_m), \quad n \in \mathbb{Z}^+,$$

and

$$\theta_{m,n} = n\theta_m, \quad n \in \mathbb{Z}^+.$$

Example 2. This modifies Example 1 in that now an admitted customer is lost to the system only if her natural lifetime expires before she enters service. Otherwise, she enters service and remains in the system for the service to be completed successfully. We now have

$$\mu_{m,n} = \mu \min(n, s_m), \quad n \in \mathbb{Z}^+,$$

and

$$\theta_{m,n} = \theta_m (n - s_m)^+, \quad n \in \mathbb{Z}^+.$$

Many other examples are plainly possible including those where the renege rate depends upon position in the queue.

(iv) The goal of analysis is to determine a policy for admission control and routing, which will maximise the average net reward earned by the system per unit of time over an infinite horizon or which will come close to doing so. Note that in the absence of losses ($\theta_{m,n} = 0$ for all m, n), it will never be optimal to refuse admission, and any stable routing policy for the system (should one exist) will earn reward rate $R\lambda$ when $R_m = R$, $1 \leqslant m \leqslant M$. Should station rewards differ, then as much use should be made of high reward stations as possible. It is thus the presence of losses that makes this problem challenging.

The policies of prime interest are those in the stationary class $\mathcal{U}$. These make decisions on the basis of the current system state and may be considered as maps from the state space (the set of possible system states) $\mathbb{N}^M$ to the action space $\{1, 2, \ldots, M, *\}$. In the latter, $m \in \{1, 2, \ldots, M\}$ is a decision to route an incoming customer to station m, whereas $*$ is a decision to refuse admission (discard). In the course of our theoretical development in the next section, we shall relax our problem by extending policy class $\mathcal{U}$ to $\mathcal{U}'$. Members of $\mathcal{U}'$ are represented by maps from $\mathbb{N}^M$ to the set of all subsets of $\{1, 2, \ldots, M\}$, denoted $2^{\{1,2,\ldots,M\}}$. Think of $\mathcal{U}'$ as an expanded class of stationary policies that at every arrival epoch may increase by one the head count at any number of service stations. Further, $\mathcal{U}$ may be thought of as the subset of $\mathcal{U}'$ consisting of maps whose range within $2^{\{1,2,\ldots,M\}}$ is restricted to subsets of

$\{1, 2, \ldots, M\}$ whose cardinality is zero (the discard action) or one (a routing action). Generic policies (of whichever class) are denoted u.

We shall impose the following condition on the problem parameters. Note that here and throughout the paper we shall use the terms "increasing" and "decreasing" in their weak sense. We shall add the qualifying "strictly" as required.

Condition (C1). *The sequences* $\{\mu_{m,n}, n \in \mathbb{Z}^+\}$ *and* $\{\theta_{m,n}, n \in \mathbb{Z}^+\}$ *are nonnegative and increasing such that* $\mu_{m,n} + \theta_{m,n}$ *is always strictly positive. Further, for each* m*, there exists* n_m *for which*

$$\mu_{m,n} + \theta_{m,n} > \lambda, \quad n \geqslant n_m. \tag{1}$$

Condition (C1) plainly guarantees stability—indeed, is conservative in guaranteeing the stability of each station when facing the entire arrivals stream. We have opted for (C1) for two reasons. First, it is satisfied by most reasonably modelled systems. Indeed, it is usually the case that $\{\theta_{m,n}, n \in \mathbb{Z}^+\}$ is divergent, as in Examples 1 and 2. Second, relaxation of the problem to the policy class $\mathcal{U}'$ above involves consideration of a class of admission control problems in which each individual service station faces the entire arrival stream. The guarantee of stability that comes from (C1) greatly simplifies this discussion.

(v) The admission control/routing problem outlined in (i)–(iv) could in principle be solved by the methods of stochastic dynamic programming. See, for example, Puterman (1994) and Tijms (1994). However, this is an unrealistic proposition for problems of reasonable size. Neither has it yet proved possible to provide any helpful characterisation of optimal policies other than those that are available from general DP theory (for example, that there must exist an optimal policy in the stationary class $\mathcal{U}$). Hence, we follow Whittle (1996) and Niño-Mora (2002) in proposing the development of *index heuristics* for admission control and routing. In the following sections, we shall describe how a prescription of Whittle (1988) may be deployed to construct such heuristics under quite general conditions.

3. Indexability of Service Stations

Let $u \in \mathcal{U}$ be a stationary policy for admission control and routing. The average reward rate earned under u over an infinite horizon may be written as

$$R^u = \sum_{m=1}^{M} \{R_m \alpha_m(u) - C_m \beta_m(u)\} - D\bar{\lambda}(u), \tag{2}$$

where $\alpha_m(u)$ is the rate of service completions achieved at station m, $\beta_m(u)$ is the rate of losses experienced at station m, and $\bar{\lambda}(u)$ is the rate of discards under u. However, noting that

$$\sum_{m=1}^{M} \{\alpha_m(u) + \beta_m(u)\} + \bar{\lambda}(u) = \lambda,$$

we infer that (2) may be rewritten as

$$R^u = \sum_{m=1}^{M} \{(R_m + D)\alpha_m(u) + (D - C_m)\beta_m(u)\} - D\lambda. \tag{3}$$

We now introduce

$$\gamma_m(u) = \alpha_m(u) + \beta_m(u) \tag{4}$$

as the rate at which customers are admitted to station m under u. Using (4) within (3) gives

$$R^u = \sum_{m=1}^{M} \{(R_m + C_m)\alpha_m(u) + (D - C_m)\gamma_m(u)\} - D\lambda. \tag{5}$$

The inequalities

$$\sum_{m=1}^{M} \gamma_m(u) \leqslant \lambda \iff \sum_{m=1}^{M} \{\lambda - \gamma_m(u)\} \geqslant (M-1)\lambda \tag{6}$$

express the fact that the total rate at which customers are routed to the M stations is bounded above by the arrival rate λ. Our optimization problem may be expressed as

$$R^{\text{opt}} = \sup_{u \in \mathcal{U}} R^u. \tag{7}$$

We now follow Whittle (1988) in relaxing the problem by extending the policy class from $\mathcal{U}$ to $\mathcal{U}'$, while incorporating constraint (6) into the objective in a Lagrangian fashion. We write, for $W \in \mathbb{R}$,

$$R^{\text{opt}}(W) = \sup_{u \in \mathcal{U}'} \left(\sum_{m=1}^{M} \{(R_m + C_m)\alpha_m(u) + (D - C_m)\gamma_m(u)\} - D\lambda + W\left[\sum_{m=1}^{M} \{\lambda - \gamma_m(u)\} - (M-1)\lambda\right] \right). \tag{8}$$

After some algebraic manipulation, (8) yields the more convenient form

$$R^{\text{opt}}(W) = \sup_{u \in \mathcal{U}'} \sum_{m=1}^{M} [(R_m + C_m)\alpha_m(u) + (W - D + C_m)\{\lambda - \gamma_m(u)\}] + \lambda\left\{(D - W)(M-1) - \sum_{m=1}^{M} C_m\right\}. \tag{9}$$

Plainly, from (5)–(8) and the fact that $\mathcal{U} \subseteq \mathcal{U}'$, we conclude that

$$R^{\text{opt}}(W) \geqslant R^{\text{opt}}, \quad W \geqslant 0. \tag{10}$$

However, by construction of the policy class $\mathcal{U}'$, we have a stationwise decomposition of $R^{\text{opt}}(W)$ in (9) and write

$$R^{\text{opt}}(W) = \sum_{m=1}^{M} R_m^{\text{opt}}(W) + \lambda\left\{(D - W)(M-1) - \sum_{m=1}^{M} C_m\right\}, \tag{11}$$

where

$$R_m^{\text{opt}}(W) \equiv \sup_{u \in \mathcal{U}_m'} (R_m + C_m)\alpha_m(u) + (W - D + C_m) \cdot \{\lambda - \gamma_m(u)\}, \quad 1 \leqslant m \leqslant M. \tag{12}$$

In (12), $R_m^{\text{opt}}(W)$ is the value of an optimisation problem *defined in terms of service station m alone*, $1 \leqslant m \leqslant M$. Imagine station m facing the entire incoming stream of arriving customers (with rate λ). The class $\mathcal{U}_m'$ contains stationary policies for determining whether to admit each incoming customer to service at station m (accept) or not (discard). The goal in (12) is to maximise an economic objective with two components. The first, $(R_m + C_m)\alpha_m(u)$, is a reward rate earned from customers served to completion where an income $R_m + C_m$ is generated by each. The second, $(W - D + C_m)\{\lambda - \gamma_m(u)\}$, is a reward rate obtained from payment of an amount $W - D + C_m$ (where W is the Lagrange multiplier) whenever an arriving customer is rejected. Note that Condition (C1) guarantees that the head count at station m remains finite under all policies in $\mathcal{U}_m'$.

Denote the admission control problem for station m outlined in the previous paragraph by $P(m, W)$, $1 \leqslant m \leqslant M$, $W \in \mathbb{R}$. As is often the case with Markov and semi-Markov decision processes under the average reward criterion, there may be a significant issue regarding the nonuniqueness of optimal policies for $P(m, W)$, even within the stationary class. We shall resolve this nonuniqueness in two steps. Firstly, we shall argue that there always exists an optimal policy in the class of *monotone policies*. Hence, in consideration of the $P(m, W)$, $W \in \mathbb{R}$, we may without loss of generality restrict to the monotone class. A monotone policy is such that rejection of incoming customers when the number of customers at the station (head count) is n, say, implies rejection for all greater values of the head count. For the subsequent discussion, we shall use a for the action "accept incoming customer" and b for "discard incoming customer." We use $B(u)$ for the *rejection threshold* for stationary policy u, namely,

$$B(u) = \inf\{n; u(n) = b\}, \tag{13}$$

where $B(u) = \infty$ for any policy that accepts all incoming customers. In (13) and throughout, we use $u(n)$ for the action taken by stationary policy u when the head count is n.

Restriction to the monotone class does not quite resolve the nonuniqueness issue. For any m, there may be a (possibly countably infinite) set of isolated values of W for which there is more than one monotone policy that is optimal for $P(m, W)$. When this is the case, we shall choose the one with the smallest rejection threshold. If we use $\mathcal{U}_m(W)$ for the set of optimal monotone policies for $P(m, W)$, we write

$$\mathcal{B}_m(W) = \min_{u \in \mathcal{U}_m(W)} B(u) \tag{14}$$

for this minimal threshold and $u_m(W)$ for the corresponding monotone policy. We shall also use $u(W)$ for the policy in $\mathcal{U}'$ that operates $u_m(W)$ at each m, $1 \leqslant m \leqslant M$, and which thus achieves $R^{\text{opt}}(W)$.

The question now arises as to whether we can utilise the structure of these policies to develop heuristics for the original routing problem in (7), which is our main concern. To pursue this, we shall require each policy $u_m(W)$, $1 \leqslant m \leqslant M$, to have the following plausible property: As the Lagrange multiplier W increases (along with the payment for rejection $W - D + C_m$), the collection of states in which $u_m(W)$ chooses the discard action b for $P(m, W)$ also increases. Whittle (1988) calls this property *indexability*. It is formalised in Definition 1.

Definition 1. Station m is *indexable* if the rejection threshold $\mathcal{B}_m(\cdot)$: $\mathbb{R} \to \mathbb{N} \cup \{\infty\}$ is decreasing, namely,

$$W > W' \;\Rightarrow\; \mathcal{B}_m(W) \leqslant \mathcal{B}_m(W'), \quad 1 \leqslant m \leqslant M.$$

Should we have indexability for station m, the idea of an index for state (i.e., head count) n as the minimum W such that the optimal rejection threshold falls below n is a natural one.

Definition 2. When station m is indexable, the *Whittle index* for station m in state n is given by

$$W_m(n) = \inf\{W; n \geqslant \mathcal{B}_m(W)\}, \quad 1 \leqslant m \leqslant M.$$

Hence, the Whittle index for station m is the smallest Lagrange multiplier W such that arrivals are rejected when n customers are there. It is natural to interpret $W_m(n) - D + C_m$ as a *fair charge* on station m for admitting (equivalently, a fair compensation for rejecting) an arriving customer when its current head count is n. If all M service stations are indexable, then the relaxation in (8)–(10) will be solved by policy $u(W)$ such that, if a customer arrives at the system at some time t, (copies of) the customer are routed to each station m for which $W_m\{\mathcal{N}_m(t)\} > W$, but *not* to any station for which $W_m\{\mathcal{N}_m(t)\} \leqslant W$. Note that if all the stations have nonpositive indices and $W \geqslant 0$, then $u(W)$ will not route the incoming customer to any station. In these circumstances, $u(W)$ refuses the customer admission.

We now follow Whittle (1988) in arguing that the index-like nature of solutions to the Lagrangian relaxation of (8)–(10) makes it reasonable to propose an *index heuristic* for our original problem in (7), when all stations are indexable. In the event of an arrival at the system at time t, this index heuristic will route the incoming customer to whichever of the stations has the *largest index* (and to any of the index-maximising stations in the event of a tie), provided this index is positive. If all index values are nonpositive, the arriving customer is refused admission.

We now proceed to study the single-station problems $P(m, W)$, $1 \leqslant m \leqslant M$, $W \in \mathbb{R}$, and are able to show that the stations are guaranteed to be indexable. We describe how indices may be computed and give the station indices in closed form in particular cases.

4. Indices for Service Stations

Following §3, we study the single-station admission control problems $P(m, W)$, $1 \leqslant m \leqslant M$, $W \in \mathbb{R}$. In doing so, it will be notationally convenient to drop the station identifier m and refer to the problem $P(W)$, $W \in \mathbb{R}$. Recall that the problem of interest concerns a single service station that is free to admit arriving customers or not. Customers admitted to the station may go on to earn a reward $(R+C)$ upon service completion, should they not be lost from the system before that. A compensating reward $(W-D+C)$ is earned whenever a customer is not admitted to the station. The immediate goal of analysis is to determine when customers should be admitted to the station to maximise the average overall reward rate earned by the system over an infinite horizon. This problem is formulated as an SMDP as follows:

(i) The state of the system at time $t \in \mathbb{R}^+$ is $\mathcal{N}(t)$, the number of customers at the station. New customers arrive at the station according to a Poisson process of rate λ (set equal to the system arrival rate of §2). If $\mathcal{N}(t) = n$, then at t the station is subject to a service rate μ_n and loss rate θ_n. The sequences $\{\mu_n, n \in \mathbb{Z}^+\}$ and $\{\theta_n, n \in \mathbb{Z}^+\}$ satisfy (C1). Hence, if $\mathcal{N}(t) = n$, the first event following t will occur at time $t + A$, where $A \sim \exp(\lambda + \mu_n + \theta_n)$. With probabilities $\lambda(\lambda + \mu_n + \theta_n)^{-1}$, $\mu_n(\lambda + \mu_n + \theta_n)^{-1}$, and $\theta_n(\lambda + \mu_n + \theta_n)^{-1}$ the event concerned will be, respectively, an arrival, a service completion, and a loss to the system. At every service completion, a reward $R + C$ is earned. Losses from the system earn nothing.

(ii) The decision epochs are the times of customer arrivals at the station. Each incoming customer will either be admitted to the system (action a, accept) or not (action b, reject). In the latter case, a compensating reward $W - D + C$ is received. A stationary policy u is a rule for choosing between a and b at each decision epoch in the light of the current state. The immediate goal of analysis will be the determination of a stationary policy to maximise the overall average reward rate earned over an infinite horizon.

Find in the appendix a proof of the fact that there exists a monotone policy that is optimal for $P(W)$. Henceforth, we restrict to the monotone class. In what follows, use of $N \in \mathbb{N} \cup \{\infty\}$ as a policy identifier is taken to indicate the monotone policy u with $B(u) = N$. By standard results, the stationary distribution of the system state (head count) under N, written $\{\Pi_x^N, 0 \leqslant x \leqslant N\}$, is given by

$$\Pi_x^N = \lambda^x \{M(x)\}^{-1} \Pi_0^N, \quad 0 \leqslant x \leqslant N, \tag{15}$$

where

$$M(x) = \prod_{y=1}^{x} (\mu_y + \theta_y)$$

and

$$\Pi_0^N = \left[\sum_{x=0}^{N} \lambda^x \{M(x)\}^{-1}\right]^{-1}. \tag{16}$$

We now write the average reward rate earned under monotone policy N as

$$R^N(W) = (R+C)\mu^N + (W-D+C)\lambda\Pi_N^N, \tag{17}$$

where

$$\mu^N = \sum_{x=1}^{N} \mu_x \Pi_x^N. \tag{18}$$

Please note that it is trivial to show that under Condition (C1) the sequence $\{\Pi_N^N, N \in \mathbb{N}\}$ is strictly decreasing, whereas $\{\mu^N, N \in \mathbb{N}\}$ is increasing and bounded above by λ.

If we write $\bar{R}^{\text{opt}}(W)$ for the optimal value of $P(W)$, then from (17) we have

$$\bar{R}^{\text{opt}}(W) = \max_{N \in \mathbb{N} \cup \{\infty\}} \{(R+C)\mu^N + (W-D+C)\lambda\Pi_N^N\}. \tag{19}$$

Recalling (14), we use $\mathcal{B}(W)$ for the smallest N-value achieving the maximum in (19).

Theorem 1. *The station is indexable.*

Proof. From (19), the function $\bar{R}^{\text{opt}}(W)$ is the upper envelope of a collection of functions that are linear and increasing in W. It follows that $\bar{R}^{\text{opt}}(W)$ is convex and increasing. It is easy to show that there exist W-values, denoted W_1 and W_∞, satisfying

$$\mathcal{B}(W) = \begin{cases} 0, & W \geqslant W_1, \\ \infty, & W < W_\infty. \end{cases} \tag{20}$$

Further, it is straightforward that $\lambda\Pi_{\mathcal{B}(W)}^{\mathcal{B}(W)}$ is the right gradient of $\bar{R}^{\text{opt}}(W)$ for every $W \in \mathbb{R}$. It now follows from the convexity of $\bar{R}^{\text{opt}}(W)$ and the fact that $\{\Pi_N^N, N \in \mathbb{N}\}$ is strictly decreasing that $\mathcal{B}(W)$ must be decreasing in W. We conclude that the station is indexable. □

Having established that the station is indexable, we now proceed to identify index values. There are two cases. One possibility is that $\mathcal{B}(W)$ takes on a finite number of distinct values *only* as W ranges through $\mathbb{R}$. It then follows that $\bar{R}^{\text{opt}}(W)$ is a piecewise-linear function with a finite number of linear pieces. The W-values of the hinges (the points of nondifferentiability of $\bar{R}^{\text{opt}}(W)$ at which the linear pieces meet) correspond to index values. Alternatively, $\mathcal{B}(W)$ may take on a countably infinite number of distinct values. In this case, the W-values of the hinges that connect linear pieces of $\bar{R}^{\text{opt}}(W)$ of positive gradient are index values. These will have a limit point that is the value W_∞ in (20). We now describe the general index structure before proceeding to special cases. Mathematical details are given in the appendix.

Index values are computed by the following algorithm:

Step 1. Set

$$W_1 = D - C + (R+C)\lambda^{-1} \sup_{n \geqslant 1} \{\mu^n (1 - \Pi_n^n)^{-1}\}, \tag{21}$$

with N_1 defined to be the largest maximiser in (21) or ∞ if the supremum is achieved in the limit $n \to \infty$. The value

W_1 is also characterised in (20). It is the index value when the head count at the station is less than N_1. We write

$$W(n) = W_1, \quad n < N_1. \tag{22}$$

If $N_1 = \infty$, stop. Otherwise, go to Step 2.

Step k. For $k \geqslant 2$, set

$$W_k = D - C + (R+C)\lambda^{-1} \cdot \sup_{n \geqslant N_{k-1}+1} \{(\mu^n - \mu^{N_{k-1}})(\Pi_{N_{k-1}}^{N_{k-1}} - \Pi_n^n)^{-1}\}, \tag{23}$$

with N_k defined to be the largest maximiser in (23) or ∞ if the supremum is achieved in the limit $n \to \infty$. The value W_k is the index value when the head count at the station is in the range $N_{k-1} \leqslant n < N_k$. We write

$$W(n) = W_k, \quad N_{k-1} \leqslant n < N_k. \tag{24}$$

If $N_k = \infty$, stop. Otherwise, go to Step $k+1$.

Please note that it is *always* the case that the index is decreasing in the head count. We now proceed to discuss an important special case.

Theorem 2. *If the sequence* $\{(\mu^{n+1} - \mu^n)(\Pi_n^n - \Pi_{n+1}^{n+1})^{-1}, n \in \mathbb{N}\}$ *is decreasing, then the Whittle index,* $W(\cdot): \mathbb{N} \to \mathbb{R}$, *is given as follows*:

$$\begin{aligned} W(n) &= D - C + (R+C)\lambda^{-1}(\mu^{n+1} - \mu^n)(\Pi_n^n - \Pi_{n+1}^{n+1})^{-1} \\ &= D - C + (R+C) \\ &\quad \cdot \left[\mu_{n+1} + \sum_{x=1}^{n} \lambda^x \{M(x)\}^{-1}(\mu_{n+1} - \mu_x)\right] \\ &\quad \times \left[\mu_{n+1} + \theta_{n+1} + \sum_{x=1}^{n} \lambda^x \{M(x)\}^{-1} \right. \\ &\qquad \left. \cdot (\mu_{n+1} + \theta_{n+1} - \mu_x - \theta_x)\right]^{-1}, \quad n \in \mathbb{N}. \end{aligned} \tag{25}$$

Proof. We use the values N_k, $k \geqslant 1$, that arise in (22)–(24) above when computing index values. For convenience, we also write $N_0 = 0$. Under the conditions outlined in the statement of the theorem, it follows that if $N_{k-1} < \infty$, then the maximisations in (21) and (23) are achieved by all n in the range $N_{k-1} + 1 \leqslant n \leqslant N_k$ when $k \geqslant 1$. It follows easily that

$$W_k = D - C + (R+C)\lambda^{-1}(\mu^{n+1} - \mu^n)(\Pi_n^n - \Pi_{n+1}^{n+1})^{-1}, \quad N_{k-1} \leqslant n < N_k, \ k \geqslant 1. \tag{26}$$

The result now follows from the assertions in (22) and (24). Explicit formulae for μ^n, Π_n^n, which use (15), (16), and (18), yield the alternative form of the index in (25). This concludes the proof. □

Theorem 2 raises the additional question of whether simply stated conditions exist that guarantee that $\{(\mu^{n+1} - \mu^n) \cdot (\Pi_n^n - \Pi_{n+1}^{n+1})^{-1}, n \in \mathbb{N}\}$ is indeed a decreasing sequence. The following Conditions (C2) and (C3) are sufficient. A proof of Lemma 3 may be found in the appendix.

Condition (C2). *If* $\mu_n = \mu_{n+1}$ *for any* $n \geqslant 1$, *then* $\mu_n = \mu_{n'}$ *for all* $n' \geqslant n$.

Condition (C3). *The quantity* $(\theta_{n+1} - \theta_n)(\mu_{n+1} - \mu_n)^{-1}$ *is increasing over the range of n for which the denominator is positive.*

Lemma 3. *Under Conditions* (C2) *and* (C3), *the sequence* $\{(\mu^{n+1} - \mu^n)(\Pi_n^n - \Pi_{n+1}^{n+1})^{-1}, n \in \mathbb{N}\}$ *is decreasing.*

Conditions (C2) and (C3) hold in a large number of important cases. In particular, they will be satisfied for any situation in which the sequence $\{\mu_n, n \in \mathbb{N}\}$ is increasing concave with the sequence $\{\theta_n, n \in \mathbb{N}\}$ increasing convex. They are certainly satisfied in the examples described in §2. We now present the form of the index for these cases.

Example 1. If $\mu_n = \mu \min(n, s)$ and $\theta_n = \theta n$, then from Theorem 2 and Lemma 3, the Whittle index is given by

$$W(n) = \begin{cases} D - C + (R+C)\mu(\mu+\theta)^{-1}, & n < s, \\ D - C + (R+C)\mu\left\{\sum_{x=0}^{s-1}(\Pi_0^x)^{-1}\right\} \\ \quad \cdot \left\{\mu\sum_{x=0}^{s-1}(\Pi_0^x)^{-1} + \theta\sum_{x=0}^{n}(\Pi_0^x)^{-1}\right\}^{-1}, & n \geqslant s. \end{cases}$$

In the single-server case $s = 1$, this becomes

$$W(n) = D - C + (R+C)\mu\left[\mu + \theta(n+1) + \sum_{x=1}^{n}\theta\lambda^x(n-x+1) \cdot \left\{\prod_{y=1}^{x}(\mu+\theta y)\right\}^{-1}\right]^{-1}, \quad n \in \mathbb{N}. \tag{27}$$

Example 2. If $\mu_n = \mu \min(n, s)$ and $\theta_n = \theta(n-s)^+$, then from Theorem 2, the Whittle index is given by

$$W(n) = \begin{cases} D + R, & n < s, \\ D - C + (R+C)\mu\left\{\sum_{x=0}^{s-1}(\Pi_0^x)^{-1}\right\} \\ \quad \cdot \left\{\mu\sum_{x=0}^{s-1}(\Pi_0^x)^{-1} + \theta\sum_{x=s}^{n}(\Pi_0^x)^{-1}\right\}^{-1}, & n \geqslant s. \end{cases}$$

In the single-server case $s = 1$, this becomes

$$W(n) = D - C + (R+C)\mu\left[\mu + \theta n + \sum_{x=1}^{n}\theta\lambda^x(n-x+1) \cdot \left\{\prod_{y=1}^{x}(\mu+\theta y)\right\}^{-1}\right]^{-1}, \quad n \in \mathbb{N}. \tag{28}$$

Comment

To further understand the Whittle indices that emerge from the above analysis, focus on Example 1 and the index for the single-server case given in (27). First, note that if an incoming customer is routed to a station with these dynamics when n customers are already present and service is according to FCFS, then the probability that a customer will be served to completion is easily shown to be $\mu\{\mu+\theta(n+1)\}^{-1}$. Recall from the discussion following Definition 2 that in the context of problem $P(W)$, $W(n)-D+C$ may be understood as a *fair charge* on the station for admitting the customer. A naïve proposal would be that this fair charge should simply be the expected return from the admitted customer. The resulting index $\widehat{W}(n)$ would satisfy

$$\widehat{W}(n)-D+C=(R+C)\mu\{\mu+\theta(n+1)\}^{-1},$$

and hence

$$\widehat{W}(n)=D-C+(R+C)\mu\{\mu+\theta(n+1)\}^{-1}, \quad n\in\mathbb{N}. \quad (29)$$

The indices in (27) and (29) are in agreement when $n=0$ and the arriving customer is able to proceed directly to service. However, in general, the charge in (29) has failed to take account of the negative impact of increased congestion resulting from the decision to admit. The quantity in (29) is thus larger than the fair charge in (27) when $n\geqslant 1$. The two indices coincide in a light-traffic limit (i.e., as $\lambda\to 0$) in which the negative impact of a decision to admit on future customers may be safely neglected. A heuristic based on the index in (29) (and constructed as in the concluding paragraph of §3) is called the *individually optimal policy*. This discussion is easily extended to Example 2 and to cases with multiple servers.

5. The Performance of the Index Policy I—Insights from the Lagrangian Relaxation

We have seen from the preceding sections that the Lagrangian relaxation of our admission control/routing problem given in (8) and (9) is easily solved. Recall that when $W\geqslant 0$, $R^{\text{opt}}(W)$ is achieved by any policy in $\mathscr{U}'$, which routes each incoming arrival to those stations whose Whittle index is no less than W. We shall refer to the value of the minimisation problem

$$\min_{W\geqslant 0} R^{\text{opt}}(W) \quad (30)$$

as REL (short for relaxation) in what follows. From (10), we conclude that $\text{REL}\geqslant R^{\text{opt}}$. In this section, we shall discuss what insights can be derived from solutions to the Lagrangian relaxation regarding the quality of performance for the original problem in (7) of the proposed index heuristic, which routes incoming customers to whichever of the stations has the largest index (if positive, and discards if all indices are negative).

We first note from the development in §3 that, should either

(a) $R^{\text{opt}}(0)$ be achieved by a policy in $\mathscr{U}$, or

(b) there exists a positive W-value for which $R^{\text{opt}}(W)$ is achieved by a policy in $\mathscr{U}$ that never discards,

then it must follow that $\text{REL}=R^{\text{opt}}$ and an index policy is optimal for our admission control/routing problem in (7). In Example 3, we give an instance of (b). In Examples 4–7, we describe a range of asymptotic regimes for our model, where we either have (a) (Examples 5 and 6) or (b) (Examples 4 and 7) in the limit considered, suggesting forms of asymptotic optimality for the index policy. In all examples, we shall suppose that station 1 has the largest Whittle index when empty and that $W_1(0)>0$.

Example 3. Suppose that station 1 has stochastic dynamics such that the supremum in (21) is achieved in the limit $n\to\infty$, and hence $N_1=\infty$. This happens, for example, under an assumption (somewhat implausible in practice) that $(\theta_{1,n+1}-\theta_{1,n})(\mu_{1,n+1}-\mu_{1,n})^{-1}$ is decreasing in n. It will then follow from (22) together with the fact that each station index is decreasing in its head count, that

$$W_1(n)=W_1(0)\geqslant \max_{m,\bar{n}}\{W_m(\bar{n})\}, \quad n\in\mathbb{N}, \quad (31)$$

where the maximisation in (31) is over $2\leqslant m\leqslant M$ and $\bar{n}\in\mathbb{N}$. Hence, station 1 *always* has a maximal index irrespective of its head count. It follows that there exists a policy achieving $R^{\text{opt}}\{W_1(0)\}$ that routes each arriving customer to station 1 *only* and which thus is a non-discarding policy belonging to $\mathscr{U}$. It must follow that $\text{REL}=R^{\text{opt}}$ and that the index policy solves our admission control/routing problem in this case.

For clarity and simplicity, Examples 4–7 below will be presented under an assumption that all stations have a single server and a stochastic structure of the kind described in Example 2. Note from (28) that for Example 2, the Whittle index of an empty station is independent of the customer arrival rate λ and of the station's service rate μ.

Example 4 (Light Traffic). Suppose that the arrival rate λ tends to zero, with other model parameters remaining fixed. For very small λ, arriving customers will (under any policy in $\mathscr{U}'$) encounter an empty system with probability close to one. There exists a policy achieving $R^{\text{opt}}\{W_1(0)\}$ that will route a customer arriving at an empty system to one station (station 1) only. It follows that there exists a policy in $\mathscr{U}$ (the index policy) that comes close to achieving $R^{\text{opt}}\{W_1(0)\}$ and, hence, also comes close to achieving R^{opt}. See (b) above. A formal analysis of the asymptotic optimality of the index policy in a light-traffic limit will be given as Theorem 6 in the next section.

Example 5 (Heavy Traffic). Suppose that the arrival rate λ diverges to infinity, with other model parameters remaining fixed. For all sufficiently large λ, it is possible to assert the existence of head count $\widetilde{X}_m$ with the property that the station m index at $\widetilde{X}_m$ (and below) is positive while that at $\widetilde{X}_m+1$ (and above) is negative, $1\leqslant m\leqslant M$. For very large λ, under an optimal policy for the Lagrangian relaxation with $W=0$, almost all arriving customers will *either* encounter a system in which all station indices are negative (head counts are $\widetilde{X}_m + 1$, $1 \leqslant m \leqslant M$) in which case the new arrival will be discarded, *or* one in which exactly one station has positive index (with head count $\widetilde{X}_m$ for the appropriate m) and will receive the customer. The existence of a policy in $\mathcal{U}$ that comes close to achieving $R^{\text{opt}}(0)$ when λ is large follows. It follows from (a) above that the index policy will also come close to achieving R^{opt}. A formal analysis of the asymptotic optimality of the index policy in a heavy-traffic limit will be given in §7.

Example 6 (Slow Service). Suppose that all service rates μ_m tend to zero, with other model parameters remaining fixed. From the form of the indices for Example 2 given in §4, it is easy to see that if $D < C_m$, $1 \leqslant m \leqslant M$, then in this limit station indices are positive if and only if the station is empty. Now consider the system in steady state, evolving under an optimal policy for the Lagrangian relaxation with $W = 0$. When $\max_m \mu_m$ is very small, almost all arriving customers will *either* encounter a system in which all station indices are negative (all head counts are one) in which case any new arrival will be discarded, *or* one in which exactly one station has positive index (and is empty). The existence of a policy in $\mathcal{U}$ that comes close to achieving $R^{\text{opt}}(0)$ when $\max_m \mu_m$ is small and $D < C_m$, $1 \leqslant m \leqslant M$, follows. It follows from (a) that the index policy will come close to achieving R^{opt}.

Example 7 (Fast Service from Station 1). Suppose that service rate μ_1 diverges to infinity, with all other parameters remaining fixed. For very large μ_1, there exists an optimal policy for the Lagrangian relaxation with $W = W_1(0)$ under which almost all arriving customers will find station 1 empty and will be routed there (only). It follows that the index policy comes close to achieving $R^{\text{opt}}\{W_1(0)\}$ and hence R^{opt} when μ_1 is large.

In Table 1, find numerical results relating to some two-station problems, with each station having a single server and stochastic structure of the kind described in Example 1. In all cases, the two stations have common loss rate per customer and common loss penalty, denoted θ and C, respectively. The values of all model parameters may be found in the table. The table gives values of the reward rate from the index policy (column headed *INDEX*) along with the values of R^{opt} (column headed *OPT*) and REL. From the above discussion, it is unsurprising that for those cases in which R^{opt} is close to REL, the index policy is close to optimal. However, please note that the converse is *not* true. There are cases in which REL is considerably larger than R^{opt}, and hence the Lagrangian relaxation is not tight, and yet the index policy remains close to optimal. We explore this general issue further in Example 8, which is discussed in the next section. In the final column of Table 1 (headed π_2), find values, estimated by simulation, of the steady-state probability that the policy achieving REL admits (inadmissibly for the original problem) an incoming customer to *both* stations. As can be seen from the table, the value of this probability is positively associated with the difference $\text{REL} - R^{\text{opt}}$ and is a natural measure of the (lack of) tightness of the Lagrangian relaxation. Moreover, such probabilities are easily estimated from simulations of the relaxed system. This remains true for large problems (i.e., many stations) where direct calculation of R^{opt}, and hence of $\text{REL} - R^{\text{opt}}$, is not possible.

6. The Performance of the Index Policy II—Suboptimality Bounds

We have seen in §5 that examination of the tightness of the Lagrangian relaxation in (8) and (9) can shed light on the quality of performance of the index heuristic. However, the numerical results in Table 1 have demonstrated the limitations of this approach. See also Example 8 below. The index heuristic may perform strongly even when the Lagrangian relaxation is far from tight. We now describe a stronger approach to the evaluation of the index heuristic

Table 1. Average reward rates and measures of the tightness of the Lagrangian relaxation for two-station problems with $R_1 = 1.5$, $\mu_1 = 1.5$, $R_2 = 1$, $\mu_2 = 1$, $C = 1.0$, $D = 0.5$.

λ	θ	INDEX	OPT	REL	π_2	λ	θ	INDEX	OPT	REL	π_2
0.5	0.1	0.6440	0.6440	0.6440	0.0047	2.0	0.1	2.0644	2.0658	2.1704	0.2742
0.5	0.2	0.5629	0.5629	0.5631	0.0134	2.0	0.2	1.7192	1.7210	1.8607	0.2803
0.5	0.3	0.4971	0.4971	0.4975	0.0483	2.0	0.3	1.4587	1.4707	1.5941	0.2658
0.5	0.4	0.4404	0.4404	0.4408	0.0427	2.0	0.4	1.2664	1.2667	1.3781	0.3060
0.5	0.5	0.3906	0.3906	0.3910	0.0380	2.0	0.5	1.0920	1.0934	1.1964	0.3137
1.0	0.1	1.2087	1.2088	1.2121	0.0885	2.5	0.1	2.2853	2.3016	2.4913	0.2888
1.0	0.2	1.0392	1.0392	1.0459	0.1490	2.5	0.2	1.8866	1.9074	2.0948	0.2674
1.0	0.3	0.9047	0.9048	0.9133	0.1343	2.5	0.3	1.6097	1.6157	1.8063	0.2686
1.0	0.4	0.7913	0.7913	0.7997	0.1216	2.5	0.4	1.3730	1.3793	1.5805	0.2786
1.0	0.5	0.6933	0.6933	0.7010	0.1109	2.5	0.5	1.1774	1.1793	1.3750	0.2877
1.5	0.1	1.6850	1.6851	1.7096	0.2451	3.0	0.1	2.2961	2.3446	2.5402	0.2883
1.5	0.2	1.4284	1.4284	1.4712	0.2274	3.0	0.2	1.9315	1.9512	2.1787	0.2883
1.5	0.3	1.2268	1.2268	1.2715	0.2110	3.0	0.3	1.6309	1.6482	1.8575	0.2802
1.5	0.4	1.0599	1.0642	1.1014	0.1956	3.0	0.4	1.3760	1.3982	1.5998	0.2624
1.5	0.5	0.9280	0.9280	0.9643	0.3184	3.0	0.5	1.1759	1.1842	1.3889	0.2647

for cases in which Theorem 2 holds. This comes in the form of *performance guarantees*, bounds on $R^{\text{opt}} - R^{\text{index}}$, the reward rate lost when implementing the index heuristic rather than an optimal policy.

In what follows, we shall use the pair (m, x) to denote "station m with head count x," with $W_m(x)$ for the corresponding index, $1 \leqslant m \leqslant M$, $x \in \mathbb{N}$, and $*$ for the discard option with index zero. We shall say that a policy in $\mathcal{U}$ *makes* (m, x) (respectively, $*$) *active* when it routes an arriving customer to station m when its head count is x (respectively, discards). The bounds we develop will capture the natural idea that a policy for admission control/routing will perform well if the rate at which it activates low-index states is small in comparison with competitor policies.

To develop these ideas further, we renumber the states

$$\left\{\bigcup_{m=1}^{M} \bigcup_{x\in\mathbb{N}} (m, x)\right\} \cup \{*\} \tag{32}$$

in decreasing order of their index values and use $\overline{W}(n)$ for the nth-highest index. If (m, x) and (m, y) have the same index and $x < y$, then (m, x) must come before (m, y) in this ordering. In what follows, each state in (32) may be referred to by the number corresponding to its position in the index ordering. Plainly,

$$\overline{W}(1) = \max\left\{\max_{1\leqslant m\leqslant M} W_m(0), 0\right\}.$$

For all $u \in \mathcal{U}$, $n \in \mathbb{N}$, we shall write $\pi(n, u)$ for the long-run proportion of time for which state n (the nth state in the index ordering) is active under policy u.

We further introduce a matrix $\boldsymbol{\mathcal{A}} \equiv \{\mathcal{A}(n, n'); (n, n') \in \mathbb{N}^2, n \leqslant n'\}$ of constants, each element lying in the range $0 \leqslant \mathcal{A}(n, n') \leqslant 1$, such that

$$\alpha(n, u) = \sum_{n'=n}^{\infty} \mathcal{A}(n, n')\pi(n', u), \quad n \in \mathbb{N},\ u \in \mathcal{U}, \tag{33}$$

furnishes us with an appropriate measure of the rate at which policy u activates states in position n or higher in the index ordering, and whose index is $\overline{W}(n)$ or less. The constants $\mathcal{A}(n, n')$ are as follows: if n' (the n'th-highest index state) is some station m pair (m, x), say, then

$$\mathcal{A}(n, n') = 1 - \lambda(1 - \Pi^{z}_{m,z})(\mu_{m,x+1} + \theta_{m,x+1})^{-1},$$

where z is such that $W_m(z) \leqslant \overline{W}(n) < W_m(z-1)$. Here we use the convention that $W_m(-1) = \infty$. If n' is the discard action $*$, then for all n,

$$\mathcal{A}(n, n') = 1.$$

It is easy to show that for all $u \in \mathcal{U}$, $\alpha(n, u)$ is decreasing in n such that $0 \leqslant \alpha(n, u) \leqslant 1$, $n \in \mathbb{N}$.

The quantities in (33) have been carefully designed such that they, together with the index values $\overline{W}(n)$, $n \in \mathbb{N}$, yield a simple expression for the reward rate R^u, $u \in \mathcal{U}$, namely,

$$R^u = \lambda\left[\overline{W}(1) - D - \sum_{n=2}^{\infty}\{\overline{W}(n-1) - \overline{W}(n)\}\alpha(n, u)\right] \tag{34}$$

$$= -\lambda D + \lambda \sum_{n=1}^{\infty} \overline{W}(n)\{\alpha(n, u) - \alpha(n+1, u)\}. \tag{35}$$

The mathematical ideas underlying the formula in (34) are explained in the online appendix. If we think of the positive quantity $\alpha(n, u) - \alpha(n+1, u)$ somewhat crudely as a measure of the usage made by policy u of the nth-highest index state, then the second term in (35) is an expectation-like quantity that weights these usage measures by the corresponding index values. It is intuitive and suggested by this expression that good policies should attempt to make as much use as possible of high-index states. It seems natural to develop performance guarantees for the index policy that measure its ability to do precisely that.

In what follows, we write $\alpha(n, \text{index})$ and $\alpha(n, \text{opt})$ for the values of $\alpha(n, u)$ when policy u is an index heuristic and an optimal policy, respectively. Further, we use $n(*)$ for the position in the index ordering of the discard option. Our performance guarantees for the index heuristic are given in Theorem 4.

THEOREM 4 (PERFORMANCE GUARANTEES FOR THE INDEX POLICY). *If the hypothesis of Theorem* 2 *holds, then*

$$R^{\text{opt}} - R^{\text{index}} = \lambda \sum_{n=2}^{\infty}\{\overline{W}(n-1) - \overline{W}(n)\} \cdot \{\alpha(n, \text{index}) - \alpha(n, \text{opt})\} \tag{36}$$

$$\leqslant \lambda \sum_{n=2}^{\infty}\{\overline{W}(n-1) - \overline{W}(n)\} \cdot \{\alpha(n, \text{index}) - \inf_{v} \alpha(n, v)\} \tag{37}$$

$$= \lambda \sum_{n=2}^{n(*)}\{\overline{W}(n-1) - \overline{W}(n)\} \cdot \{\alpha(n, \text{index}) - \inf_{v} \alpha(n, v)\}. \tag{38}$$

PROOF. Equation (36) follows immediately from (34), whereas (37) follows from the observations that, for all $n \geqslant 2$,

$$\overline{W}(n-1) \geqslant \overline{W}(n), \qquad \alpha(n, \text{opt}) \geqslant \inf_{v} \alpha(n, v).$$

Finally, because no index heuristic makes use of states $n \geqslant n(*) + 1$, we have that

$$\alpha(n, \text{index}) = \inf_{v} \alpha(n, v) = 0, \quad n \geqslant n(*) + 1,$$

which yields (38). This concludes the proof. □

The reader will find a range of comments on the performance guarantees given in Theorem 4 in the online appendix. In practice, the quantities (36)–(38) will often be dominated by a few initial terms in the summations concerned. In such cases, the bounds are looking for poor usage by the index policy of states that are low (i.e., close to one) in the index ordering.

Example 8 illustrates a range of phenomena related to the narrative of this section and the preceding one. These include (I) relatively poor performance of the index policy related to poor usage of high-index states, (II) cases in which the bounds in (37) and (38) are exact, and (III) strong performance of the index policy, notwithstanding a lack of tightness in the Lagrangian relaxation given in (8) and (9).

Example 8 (High Levels of Impatience). Consider an instance of Example 2 with two single-server stations and a loss rate per waiting customer of θ, which is common to the stations and large enough to guarantee that

$$W_m(n) < 0, \quad n \geqslant 1,\ m = 1, 2.$$

If we further suppose that $R_1 = R_2 = R$, say, then we conclude that

$$W_1(0) = W_2(0) = D + R.$$

In the terminology established in this section prior to Theorem 4, we have $n(*) = 3$,

$$\overline{W}(1) = \overline{W}(2) = D + R, \qquad \overline{W}(3) = 0,$$

zero being the index of the discard option.

Utilising the above index values, we conclude that all index heuristics for admission control/routing will discard an arriving customer when both stations are busy (i.e., nonempty) and will route a customer to the unoccupied station when exactly one is busy. There are two index policies that are stationary and deterministic. These make different choices of station for customers who arrive when the system is empty. Call these policies $1\to 2$ and $2\to 1$ to denote the preference exercised in this latter case. If we write $\pi_d^{1\to 2}$ and $\pi_d^{2\to 1}$ for the long-run proportion of customers discarded under $1\to 2$ and $2\to 1$, respectively, then it is straightforward to show that $\pi_d^{2\to 1} > \pi_d^{1\to 2}$ if and only if $\mu_1 > \mu_2$. The reward rate achieved by policy $1\to 2$ is given by

$$R^{1\to 2} = \lambda R - \lambda(R + D)\pi_d^{1\to 2} \tag{39}$$

and similarly for $R^{2\to 1}$.

In illustration of (I) above, note that the discard probabilities $\pi_d^{1\to 2}$ and $\pi_d^{2\to 1}$ may differ significantly, resulting in a substantial difference in the reward rates associated with the two index policies. For example, for the choices $\mu_1 = 1.737\lambda$, $\mu_2 = 0.263\lambda$, $R = 2D$, the reward rate for $1\to 2$ exceeds that of $2\to 1$ by 10.82%. Moreover, policy $1\to 2$ can be shown to be optimal. The poor performance of the index policy $2\to 1$ is related to the fact that it makes poor usage (in comparison with $1\to 2$) of the two highest index states while making excessive use of the third-highest index state, the discard option.

Concerning the application of the performance bounds of Theorem 4 to this example, suppose that $\mu_1 > \mu_2$, with $1\to 2$ optimal and $2\to 1$ the index policy for evaluation. It is straightforward that

$$\alpha(3, \text{index}) = \pi_d^{2\to 1}, \qquad \alpha(3, \text{opt}) = \inf_v \alpha(3, v) = \pi_d^{1\to 2}.$$

Hence, the bound on the quantity

$$R^{\text{opt}} - R^{\text{index}} = R^{1\to 2} - R^{2\to 1}$$

in (37) is given in this case by

$$\lambda\{\overline{W}(2) - \overline{W}(3)\}\{\alpha(3, \text{index}) - \inf_v \alpha(3, v)\} = \lambda(D + R)(\pi_d^{2\to 1} - \pi_d^{1\to 2})$$

and from (39) is exact. This illustrates (II) above.

In illustration of (III) above, now consider the Lagrangian relaxation for the current setup. Under the condition $\lambda^2 > \mu_1\mu_2$, the minimisation in (30) can be shown to be achieved at $W = 0$. The policy that achieves REL admits an arriving customer to *both* stations when the system is empty and otherwise makes decisions in line with an index policy. The long-run proportion of customers who are routed to both stations is given by

$$\pi_2 = \mu_1\mu_2(2\lambda + \mu_1 + \mu_2) \cdot \{(\lambda + \mu_1)(\lambda + \mu_2)(\lambda + \mu_1 + \mu_2)\}^{-1}. \tag{40}$$

See the discussion of the results in Table 1 at the conclusion of §5. If we set $\mu_1 = \mu_2 = \lambda(1 - \epsilon)$ for some small $\epsilon > 0$, it will then follow that $\pi_d^{1\to 2}$ and $\pi_d^{2\to 1}$ are equal and that both index policies are optimal. However, the fact that π_2 is close to 1/3 indicates that the Lagrangian relaxation is far from tight.

If for much of the time the index heuristic activates a state whose index is close to $\overline{W}(1)$, then it must indeed be close to optimal. This insight relates to Examples 4 and 7 in the preceding section. It is formalised in Corollary 5.

Corollary 5. *If the hypothesis of Theorem 2 holds and, further, there exists $\bar{n} \leqslant n(*)$ for which $\overline{W}(1) - \overline{W}(\bar{n}) \leqslant \delta\overline{W}(1)$ and $\alpha(\bar{n} + 1, \text{index}) \leqslant \epsilon$ for some $\delta \geqslant 0$, $\epsilon > 0$, then*

$$R^{\text{opt}} - R^{\text{index}} \leqslant \lambda\overline{W}(1)(\epsilon + \delta).$$

Proof. Under the hypotheses of the corollary, the expression in (38) is bounded above by

$$\begin{aligned}
&\lambda\left[\sum_{n=2}^{\bar{n}}\{\overline{W}(n-1) - \overline{W}(n)\}\alpha(n, \text{index}) + \sum_{n=\bar{n}+1}^{n(*)}\{\overline{W}(n-1) - \overline{W}(n)\}\alpha(n, \text{index})\right] \\
&\quad\leqslant \lambda[\{\overline{W}(1) - \overline{W}(\bar{n})\} + \overline{W}(\bar{n})\alpha(\bar{n} + 1, \text{index})] \\
&\quad\leqslant \lambda\overline{W}(1)(\epsilon + \delta).
\end{aligned} \tag{41}$$

Note that inequality (41) uses $\bar{W}\{n(*)\} = 0$ and the fact that $\alpha(n, \text{index})$ is decreasing in n and bounded above by one. The result now follows from Theorem 4. □

We can now expand on Example 4 above and develop a light-traffic analysis for all models that fall within the scope of Theorem 2. In Theorem 6, we expand the notations R^u, R^{opt}, etc. to $R^u(\lambda)$, $R^{\text{opt}}(\lambda)$ to emphasise λ-dependence.

Theorem 6 (Light-Traffic Optimality). *If the hypothesis of Theorem 2 holds and $\bar{W}(1) > D$, then the index heuristic is asymptotically optimal in a light-traffic limit in the sense that*

$$\lim_{\lambda\to 0}\{R^{\text{opt}}(\lambda) - R^{\text{index}}(\lambda)\}\{R^{\text{opt}}(\lambda)\}^{-1} = 0$$

Proof. We take $\bar{n} = 1$, $\delta = 0$ in Corollary 5 while noting that $\alpha(2, \text{index})$ is bounded above by the long-run proportion of time that the system is nonempty under the index policy. However, it is plain that the number in the system under an index policy in steady state may be stochastically bounded above by the number present in an $M/M/1$ system in steady state with arrival rate λ and service rate $\min_m(\mu_{m,1} + \theta_{m,1})$, the latter assumed to exceed λ. We thus conclude that

$$\lambda < \min_m(\mu_{m,1} + \theta_{m,1})$$
$$\Longrightarrow \alpha(2, \text{index}) \leqslant \lambda\Big\{\min_m(\mu_{m,1} + \theta_{m,1})\Big\}^{-1}. \quad (42)$$

It now follows from Corollary 5 that

$$R^{\text{opt}}(\lambda) - R^{\text{index}}(\lambda) \leqslant \lambda^2 \bar{W}(1)\Big\{\min_m(\mu_{m,1} + \theta_{m,1})\Big\}^{-1}, \quad (43)$$

where from (25) we see that $\bar{W}(1)$ is λ-independent. If $\bar{W}(1) > D$, then in the limit $\lambda\to 0$, $R^{\text{opt}}(\lambda) = O(\lambda)$ and approaches zero from above. The result now follows from (43). □

Comment

Although the condition $\bar{W}(1) > D$ in Theorem 6 is mild (guaranteeing that a customer entering an empty system can earn a positive net expected return), it may be replaced by the weaker $\bar{W}(1) > 0$ (a customer entering an empty system can earn more than $-D$ and should not be discarded). In the latter event, the form of the result is modified to

$$\lim_{\lambda\to 0}\{R^{\text{opt}}(\lambda) - R^{\text{index}}(\lambda)\}\{R^{\text{opt}}(\lambda) + D\lambda\}^{-1} = 0.$$

The proof is little changed.

Example 7 (Revisited). We can extend Example 7 to all models within the scope of Theorem 2. If the station achieving $\bar{W}(1)$, station 1 say, has associated service rate μ_{11}, which is large, then the index policy will be close to optimal. An account along the lines of the proof of Theorem 6 is straightforward.

Example 9. If the differences between the station rewards R_m are small in comparison to the size of the discard penalty D and, further, customer loss rates are low, then for models within the scope of Theorem 2, it may be that many states (m, x) have index values close to $\bar{W}(1)$. If, further, service rates are sufficient to keep queue lengths small with high probability, then even for small $\delta > 0$, we would expect to find an $\bar{n}$ for which $\bar{W}(1) - \bar{W}(\bar{n}) < \delta\bar{W}(1)$ such that $\alpha(\bar{n}+1, \text{index})$ is small. Corollary 5 then guarantees the strong performance of the index policy.

Example 10. We have already seen in the discussion of Example 8 an instance of poor performance of an index policy being captured by the bounds in Theorem 4. It is a feature of that example that a station achieving $\bar{W}(1)$ has significantly lower service rates than a competitor station. We have found that setups of this kind are prone to generate large values of $\alpha(n, \text{index})$ for small n, yielding large values of the bounds (36)–(38) and correspondingly (relatively) poor performance of the index policy, especially so when customer loss rates are high. It is perhaps also worth pointing out that cases in which μ_1 is small and $\mu_m \gg \mu_1$, $m \neq 1$, are excluded from *both* Examples 6 and 7 above, where service rate configurations guaranteeing strong performance for the index policy are described.

In illustration of Examples 7, 9, and 10 above, find in Table 2 values of $R^{\text{opt}} - R^{\text{index}}$ expressed as a percentage of $R^{\text{opt}} + D\lambda$ for a range of two-station problems. In all cases, the stations have a single server and a stochastic structure of the kind described in Example 2 with a common loss rate per customer θ and common loss penalty C. Other details are given in and above the table. Note that in all cases, it is station 1 that achieves $\bar{W}(1)$. That the index policy should perform relatively poorly when station 1's service rates are low ($\mu_1 = 0.5$) is consistent with Example 10 above, as is the fact that the suboptimalities tend to grow with the loss rate θ. From Examples 7 and 9, we would expect to see the index policy perform strongly when

Table 2. Percentage suboptimality of the index policy for a range of two-station problems with $R_1 = 1.01$, $R_2 = 1$, $\mu_2 = 1$, $C = 1$, $D = 0.5$.

		μ_1					μ_1		
θ	λ	0.5	2.0	5.0	θ	λ	0.5	2.0	5.0
0.05	0.5	0.091	0.000	0.000	0.5	0.5	1.345	0.000	0.000
	1.0	0.299	0.000	0.000		1.0	1.740	0.000	0.000
	2.0	0.338	0.017	0.000		2.0	1.707	0.026	0.000
	5.0	0.015	0.034	0.043		5.0	0.282	0.000	0.040
	10.0	0.004	0.000	0.066		10.0	0.248	0.000	0.000
0.1	0.5	0.306	0.000	0.000	1.0	0.5	1.975	0.000	0.000
	1.0	0.545	0.000	0.000		1.0	2.497	0.000	0.000
	2.0	0.751	0.021	0.001		2.0	1.639	0.005	0.000
	5.0	0.051	0.513	0.058		5.0	0.753	0.000	0.004
	10.0	0.038	0.000	0.040		10.0	0.248	0.000	0.000

Table 3. Percentage suboptimality of the index policy for varying levels of a common loss rate.

(a) $\lambda = 1,\ R_1 = 5,\ \mu_1 = 0.5$									
θ	0.5	0.3	0.1	0.05	0.03	0.01	0.005	0.001	0.0001
Percentage suboptimality	2.946	0.576	0.350	0.440	0.037	0.023	0.024	0.001	0.000
(b) $\lambda = 1,\ R_1 = 5,\ \mu_1 = 0.75$									
θ	0.5	0.3	0.1	0.05	0.03	0.01	0.005	0.001	0.0001
Percentage suboptimality	0.107	0.875	0.000	0.121	0.043	0.032	0.018	0.000	0.000

Note. Other details given in text.

μ_1 is large, particularly for these examples in which the rewards R_1 and R_2 are close together. This is indeed the case.

Example 11 (Low Levels of Impatience). Consider a two-station instance of Example 1 or Example 2 and suppose that there is a loss rate per customer θ that is common to both stations. We consider an asymptotic regime in which $\theta \to 0$ while the remaining stochastic parameters (λ, μ_1, μ_2) stay fixed such that $\lambda < s_1\mu_1 + s_2\mu_2$. Note that if our interest is in studying proportionate suboptimality measures like $(R^{\text{opt}} - R^{\text{index}})(R^{\text{opt}})^{-1}$ or $(R^{\text{opt}} - R^{\text{index}})(R^{\text{opt}} + D\lambda)^{-1}$, the above is equivalent to consideration of an asymptotic regime in which θ remains fixed while the arrival rate (λ) and the service rates (μ_1, μ_2) diverge to infinity in fixed proportion. The latter has echoes of the many-server heavy-traffic limits considered by Halfin and Whitt (1981), among others. An analysis based on Theorem 4 yields the asymptotic optimality of the index policy. Find an account in the online appendix for the Example 2 case when $W_1(0) > W_2(0)$.

By way of illustration, find in Tables 3(a) and 3(b) values of $R^{\text{opt}} - R^{\text{index}}$ expressed as a percentage of $R^{\text{opt}} + D\lambda$ for a range of two-station instances of Example 2 in which $s_m = 1$, $m = 1, 2$ and there is a common loss rate per customer θ and common loss penalty $C = 1$. We also have $R_2 = 1$, $\mu_2 = 1$, and $D = 0.5$. Other parameter values are identified in and above the tables. The results illustrate the asymptotic optimality of the index policy in the limit $\theta \to 0$.

7. The Performance of the Index Policy III—Heavy-Traffic Optimality

Insights from the Lagrangian relaxation expressed in Example 5 above yielded a conjecture of heavy-traffic optimality for the index policy—i.e., optimality in an asymptotic regime in which λ diverges to infinity with other parameters held fixed. To assist in our discussion of when this occurs, we shall express dependence upon λ by using $W_m(n, \lambda)$ as an expanded notation for the Whittle index for station m with head count n and $u_W(\lambda)$ as a shorthand for an index policy. We further denote by $X_m(\lambda)$ the largest head count for which $u_W(\lambda)$ admits customers to station m. Hence,

$$X_m(\lambda) = \begin{cases} -1 & \text{if } W_m(0, \lambda) < 0, \\ \infty & \text{if } W_m(n, \lambda) \geqslant 0,\ n \in \mathbb{N}, \quad \text{and} \\ \max[n;\ W_m(n, \lambda) \geqslant 0] & \text{otherwise.} \end{cases} \tag{44}$$

As in §6, we shall suppose that the hypothesis of Theorem 2 is satisfied. We shall additionally require condition (C4), which is trivially satisfied by Examples 1 and 2.

Condition 4 (C4). *All loss rate sequences $\{\theta_{m,n},\ n \in \mathbb{Z}^+\}$ diverge.*

We introduce the *heavy-traffic indices*

$$\begin{aligned} \widetilde{W}_m(n) &\triangleq \lim_{\lambda\to\infty} W_m(n, \lambda) \\ &= D - C_m + (R_m + C_m)(\mu_{m,n+1} - \mu_{m,n}) \\ &\quad \cdot (\mu_{m,n+1} - \mu_{m,n} + \theta_{m,n+1} - \theta_{m,n})^{-1}, \\ &\qquad n \in \mathbb{N},\ 1 \leqslant m \leqslant M, \end{aligned} \tag{45}$$

as the limiting form of the Whittle indices in the heavy-traffic limit. The form of the limit in (45) is derived from (25). We observe that it is straightforward to show from (25) that, under the hypothesis of Theorem 2,

$$W_m(n, \lambda) \geqslant \widetilde{W}_m(n), \quad n \in \mathbb{N},\ \lambda \in \mathbb{R}^+,\ 1 \leqslant m \leqslant M. \tag{46}$$

The indices $\widetilde{W}_m(n)$ must be decreasing in the head count n and are bounded below by $D - C_m$. Hence, the limits

$$\widetilde{W}_m \triangleq \lim_{n\to\infty} \widetilde{W}_m(n), \quad 1 \leqslant m \leqslant M,$$

must exist. We shall require (C5) to hold.

Condition 5 (C5). $\widetilde{W}_m < 0,\ 1 \leqslant m \leqslant M$.

Under (C5), the heavy-traffic indices will be negative if the head count is large enough. Please note that this is certainly the case for Examples 1 and 2 when $D < C_m$, $1 \leqslant m \leqslant M$. We now write $\widetilde{X}_m$ for the largest head count for which the heavy-traffic index is nonnegative. Under (C5), we must have $\widetilde{X}_m < \infty$, $1 \leqslant m \leqslant M$. Formally, we write

$$\widetilde{X}_m = \begin{cases} -1 & \text{if } \widetilde{W}_m(0) < 0, \text{ and} \\ \max[n;\ \widetilde{W}_m(n) \geqslant 0] & \text{otherwise.} \end{cases} \tag{47}$$

Theorem 7 (Heavy-Traffic Optimality). *If the hypothesis of Theorem 2 holds along with conditions* (C4) *and* (C5), *the index heuristic is asymptotically optimal in a heavy-traffic limit in the sense that*

$$\lim_{\lambda\to\infty}\{R^{\text{opt}}(\lambda)-R^{\text{index}}(\lambda)\}=0.$$

Proof. From (45) and (47), observe that

$$D-C_m+(R_m+C_m)(\mu_{m,n+1}-\mu_{m,n})\cdot(\mu_{m,n+1}-\mu_{m,n}+\theta_{m,n+1}-\theta_{m,n})^{-1}\geqslant 0 \iff 0\leqslant n\leqslant \widetilde{X}_m,\ 1\leqslant m\leqslant M. \tag{48}$$

It follows simply from (48) that the maximum

$$\max_n\{(R_m+D)\mu_{m,n}+(D-C_m)\theta_{m,n}\} \tag{49}$$

is achieved at $n=\widetilde{X}_m+1$, $1\leqslant m\leqslant M$. Further, from (46), (48) and the fact that all indices are decreasing in the head count, we infer the existence of $\Delta<\infty$ for which $\lambda\geqslant\Delta$ implies that

$$W_m(n,\lambda)\geqslant 0 \iff 0\leqslant n\leqslant\widetilde{X}_m,\ 1\leqslant m\leqslant M$$

and hence that

$$X_m(\lambda)=\widetilde{X}_m,\quad \lambda\geqslant\Delta,\ 1\leqslant m\leqslant M. \tag{50}$$

Hence, if $\lambda\geqslant\Delta$, then index policy $u_W(\lambda)$ will admit a customer to the system (specifically, to the station of largest index) and not discard her if and only if at least one station m has queue length no greater than $\widetilde{X}_m$. If station m has queue length $\widetilde{X}_m+1$ (or greater), then further customers will not be routed to it. It must then follow that in the limit $\lambda\to\infty$, $u_W(\lambda)$ holds the queue length at station m fixed at $\widetilde{X}_m+1$, $1\leqslant m\leqslant M$. A fuller discussion of this is given in the appendix.

It immediately follows from (3), (49) and the comments following that, for any stationary policy u and arrival rate $\bar{\lambda}$,

$$\begin{aligned}\lim_{\lambda\to\infty}&\{R^{u_W(\lambda)}(\lambda)+D\lambda\}\\&=\sum_{m=1}^{M}(r_m+D)\mu_{m,\widetilde{X}_m+1}+\sum_{m=1}^{M}(D-C_m)\theta_{m,\widetilde{X}_m+1}\\&=\max_{\underline{n}}\left\{\sum_{m=1}^{M}(r_m+D)\mu_{m,n_m}+\sum_{m=1}^{M}(D-C_m)\theta_{m,n_m}\right\}\\&\geqslant\sum_{\underline{n}}\sum_{m=1}^{M}\{(r_m+D)\mu_{m,n_m}+(D-C_m)\theta_{m,n_m}\}\widetilde{\pi}(\underline{n},u,\bar{\lambda}) \qquad (51)\\&=R^u(\bar{\lambda})+D\bar{\lambda}, \qquad (52)\end{aligned}$$

where in (51), $\widetilde{\pi}(\underline{n},u,\bar{\lambda})$ is the long-run proportion of time the system spends in state $\underline{n}$ under stationary policy u when the arrival rate is $\bar{\lambda}$. It follows immediately from (52) that

$$\lim_{\lambda\to\infty}\{R^{u_W(\lambda)}(\lambda)+D\lambda\}=\lim_{\lambda\to\infty}\{R^{\text{opt}}(\lambda)+D\lambda\}$$

with both limits finite. The result follows. □

Comments

1. The character of the index heuristic $u_W(\lambda)$ changes as the arrival rate moves from near zero (light traffic) to very large values (heavy traffic) in precisely the way required to keep its reward rate performance close to optimal. When $\lambda\cong 0$, $u_W(\lambda)$ acts very much like an *individually optimal* policy that routes each incoming customer to whichever station yields the greatest expected net reward for that individual and discards if all such net rewards are negative. As is explained at the conclusion of §4, this is appropriate in light traffic because longer-term implications of decisions may be safely neglected.

As λ diverges to infinity, the Whittle indices tend to their heavy-traffic versions in (45). Further, the minimum values of the head counts at which the stations refuse admission, namely, $X_m(\lambda)+1$, tend toward the values $\widetilde{X}_m+1$. In the limit, the index policy $u_W(\lambda)$ holds the station head counts fixed at $\widetilde{X}_m+1$, $1\leqslant m\leqslant M$. From the comment around (49), these head counts are exactly the ones at which the maximum net reward rates are earned at each station.

2. A pure reward version ($D=0$, $C_m=0$, $1\leqslant m\leqslant M$) of Theorem 7 may be obtained under the hypothesis of Theorem 2 along with Conditions (C2) and (C3), but with the additional requirement that, for each station m there exists some finite value $\widetilde{\mathscr{X}}_m$ defined by

$$\widetilde{\mathscr{X}}_m=\min\{n;\,\mu_{m,n}=\mu_{m,n+1}\}. \tag{53}$$

This certainly, for example, includes pure reward versions of Examples 1 and 2. The analysis is similar in spirit to the above proof and shows that in the limit $\lambda\to\infty$, $u_W(\lambda)$ keeps at least $\widetilde{\mathscr{X}}_m$ customers at station m almost surely, $1\leqslant m\leqslant M$. Plainly, in such examples net reward rates $R^{\text{opt}}(\lambda)$ and $R^{\text{index}}(\lambda)$ remain positive for all λ.

3. To illustrate the quality of performance of the index policy, most especially with respect to varying values of the customer arrival rate λ, 720 two-station problems were studied. In each case, both stations have a single server and a stochastic structure of the kind described in Example 2 and the requirements of Theorem 7 are met. Further, we suppose there to be a common loss rate per customer θ, common loss penalty C (always set at 1.0) and the discard penalty D is set to be 0.5 for all problems. Station 2 has $R_2=1$ and $\mu_2=1$ throughout. The remaining parameters are set as follows: R_1 is chosen from the set $\{1.01, 1.5, 2, 5\}$, μ_1 from the set $\{0.5, 1, 2, 3, 5\}$, θ from the set $\{0.05, 0.1, 0.2, 0.3, 0.5, 1\}$, and λ from the set $\{0.5, 1, 2, 3, 5, 10\}$. The total number of parameter combinations is $4\times 5\times 6\times 6=720$. Note that the station achieving maximal index $\overline{W}(1)$ is station 1 in every case. For all 720 problems, the net reward rates achieved by an optimal policy for admission control/routing and by the index policy were computed using dynamic programming value iteration. For each problem, the reward rate difference $R^{\text{opt}}-R^{\text{index}}$ was computed and expressed as a percentage

Table 4. Median and maximum percentage suboptimalities for the index heuristic as (R_1, λ) varies.

R_1	λ	0.5	1	2	3	5	10
1.01	Median	0.000	0.000	0.016	0.023	0.042	0.000
	Maximum	1.975	2.497	1.707	3.262	0.753	0.364
1.5	Median	0.000	0.000	0.035	0.041	0.030	0.001
	Maximum	0.026	0.196	1.328	2.127	2.285	1.248
2.0	Median	0.000	0.000	0.039	0.034	0.062	0.007
	Maximum	0.274	0.236	2.276	1.450	0.720	0.191
5.0	Median	0.000	0.000	0.013	0.030	0.094	0.020
	Maximum	0.240	2.946	1.870	1.783	4.053	0.487

of $R^{\text{opt}} + D\lambda$. The 720 problems have been grouped into 24 sets of 30, according to values of the pair (R_1, λ). For each set of 30, the median and maximum percentage suboptimalities for the index policy were computed. They are presented in Table 4. As would be expected from Theorem 6 and comment 1 above, the index heuristic performs well in light traffic, whereas as λ increases to 10, it approaches optimality in line with Theorem 7. The performance of the index policy is remarkably robust, with a worst-case suboptimality of 4.053% in the 720 problems studied.

8. Electronic Companion

An electronic companion to this paper is available as part of the online version that can be found at http://or.journal.informs.org/.

Acknowledgments

The authors acknowledge the support of the Engineering and Physical Sciences Research Council (EPSRC) for this work through the award of grants GR/S45188/01 and EP/E049265/1. They also value the comments of the referees and editors, which have challenged them to strengthen the material and clarify its presentation.

References

Ansell, P. S., K. D. Glazebrook, C. Kirkbride. 2003a. Generalised "join the shortest queue" policies for the dynamic routing of jobs to multiclass queues. *J. Oper. Res. Soc.* **54** 379–389.

Ansell, P. S., K. D. Glazebrook, J. Niño-Mora, M. O'Keeffe. 2003b. Whittle's index policy for a multi-class queueing system with convex holding costs. *Math. Methods Oper. Res.* **57** 21–39.

Armony, M. 2005. Dynamic routing in large-scale service systems with heterogeneous servers. *Queueing Systems* **51** 287–329.

Bassamboo, A., J. M. Harrison, A. Zeevi. 2005. Dynamic routing and admission control in high-volume service systems: Asymptotic analysis via multi-scale fluid limits. *Queueing Systems* **51** 249–285.

Bertsimas, D., J. Niño-Mora. 1996. Conservation laws, extended polymatroids and multi-armed bandit problems: A polyhedral approach to indexable systems. *Math. Oper. Res.* **21** 257–306.

Doytchinov, B., J. Lehoczky, S. Shreve. 2001. Real-time queues in heavy traffic with earliest-deadline-first queue discipline. *Ann. Appl. Probab.* **2** 332–378.

Garnett, O., A. Mandelbaum, M. Reiman. 2002. Designing a call center with impatient customers. *Manufacturing Service Oper. Management* **4** 208–227.

Gaver, D. P., P. A. Jacobs. 2000. Servicing impatient tasks that have uncertain outcomes. Technical report, Naval Postgraduate School, Monterey, CA.

Gittins, J. C. 1979. Bandit processes and dynamic allocation indices (with discussion). *J. Roy. Statist. Soc.* **B41** 148–177.

Glazebrook, K. D., J. Niño-Mora. 2001. Parallel scheduling of multiclass $M/M/1$ queues: Approximate and heavy-traffic optimization of achievable performance. *Oper. Res.* **49** 609–623.

Glazebrook, K. D., H. M. Mitchell, P. S. Ansell. 2005. Index policies for the maintenance of a collection of machines by a set of repairmen. *Eur. J. Oper. Res.* **165** 267–284.

Glazebrook, K. D., J. Niño-Mora, P. S. Ansell. 2002. Index policies for a class of discounted restless bandits. *Adv. Appl. Probab.* **34** 754–774.

Glazebrook, K. D., P. S. Ansell, R. T. Dunn, R. R. Lumley. 2004. On the optimal allocation of service to impatient tasks. *J. Appl. Probab.* **41** 51–72.

Halfin, S., W. Whitt. 1981. Heavy-traffic limits for queues with many exponential servers. *Oper. Res.* **29** 567–587.

Harrison, J. M., A. Zeevi. 2004. Dynamic scheduling of a multi-class queue in the Halfin-Whitt heavy traffic regime. *Oper. Res.* **52** 243–257.

Hordijk, A., G. Koole. 1990. On the optimality of the generalized shortest queue policy. *Probab. Engrg. Inform. Sci.* **4** 477–487.

Jiang, Z., T. G. Lewis, J. Y. Colin. 1996. Scheduling hard real-time constrained periodic tasks on multiple processors. *J. Systems Software* **19** 102–118.

Koole, G., P. D. Sparaggis, D. Towsley. 1999. Minimizing response times and queue lengths in systems of parallel queues. *J. Appl. Probab.* **36** 1185–1193.

Krishnan, K. R. 1987. Joining the right queue: A Markov decision rule. *Proc. 26th IEEE Conf. Decision Control*, San Francisco, 1863–1868.

Lehoczky, J. P. 1996. Real-time queueing theory. *Proc. 17th IEEE Real-Time Systems Sympos.*, Washington, D.C., 186–195.

Lehoczky, J. P. 1997a. Real-time queueing network theory. *Proc. 18th IEEE Real-Time Systems Sympos.*, Los Angeles, 58–67.

Lehoczky, J. P. 1997b. Using real-time queueing theory to control lateness in real-time systems. *Performance Evaluation Rev.* **25** 158–168.

Menich, B., R. Serfozo. 1991. Optimality of routing and servicing in dependent parallel processing systems. *Queueing Systems* **9** 403–418.

Movaghar, A. 1997. Optimal assignment of impatient customers to parallel queues with blocking. *Scientia Iranica* **3** 137–146.

Niño-Mora, J. 2001. Restless bandits, partial conservation laws and indexability. *Adv. Appl. Probab.* **33** 76–98.

Niño-Mora, J. 2002. Dynamic allocation indices for restless projects and queueing admission control: A polyhedral approach. *Math. Programming* **93** 361–413.

Papadimitriou, C. H., J. N. Tsitsiklis. 1999. The complexity of optimal queueing network control. *Math. Oper. Res.* **24** 293–305.

Puterman, M. L. 1994. *Markov Decision Processes*: *Discrete Stochastic Dynamic Programming*. Wiley, New York.

Tijms, H. C. 1994. *Stochastic Models*: *An Algorithmic Approach*. Wiley, Chichester, UK.

Ward, A. R., P. W. Glynn. 2005. A diffusion approximation for a $GI/GI/1$ queue with balking or reneging. *Queueing Systems* **50** 371–400.

Weber, R. R. 1978. On the optimal assignment of customers to parallel queues. *J. Appl. Probab.* **15** 406–413.

Weber, R. R., G. Weiss. 1990. On an index policy for restless bandits. *J. Appl. Probab.* **27** 637–648.

Weber, R. R., G. Weiss. 1991. Addendum to "On an index policy for restless bandits." *Adv. Appl. Probab.* **23** 429–430.

Whittle, P. 1988. Restless bandits: Activity allocation in a changing world. *J. Appl. Probab.* **A25** 287–298.

Whittle, P. 1996. *Optimal Control*: *Basics and Beyond*. Wiley, New York.

Winston, W. 1977. Optimality of the shortest line discipline. *J. Appl. Probab.* **14** 181–189.