Optimal Batch Service of a Polling System under Partial Information

MARK P. VAN OYEN

Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL 60208-3119, USA

DEMOSTHENIS TENEKETZIS

Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109-2122, USA

Abstract: We consider the optimal scheduling of an infinite-capacity batch server in a N-node ring queueing network, where the controller observes only the length of the queue at which the server is located. For a cost criterion that includes linear holding costs, fixed dispatching costs, and linear service rewards, we prove optimality and monotonicity of threshold scheduling policies.

1 Introduction

Large-scale networks such as high-speed communication networks, manu-facturing systems, polling systems, and transportation systems are often charac-terized by workstations that act individually, each possessing an accurate per-ception (perfect knowledge) of its immediate environment but only imperfect knowledge of the overall state of the network. Even when information can be exchanged among stations, there are propagation and processing delays that may render such information partially obsolete; moreover, faults and transmis-sion errors may render the data inaccurate. The optimization problem can thus be expressed as follows: based on partial knowledge of the system state, how should the individual workstations operate so as to maximize some measure of the overall utility for the entire system? While we cannot expect to obtain universal results of that nature, it is possible to isolate certain network aspects for analysis and to formulate optimization problems based on incomplete knowledge of the state. The problem formulated in this paper is motivated by a number of systems including information systems, polling systems, shuttle systems, and automated guided vehicle (AGV) systems with stations arranged in a ring.

Consider a distributed information system with a central database containing information that is generated at various stations. Jobs, that may represent orders or other changes in state that must be tracked, arrive randomly at each

station. The server at the central facility controls the times at which it polls or monitors the stations, but it polls them in a fixed, cyclic order. The time required to poll a station is brief and insensitive to the number of jobs waiting at the station. Thus, service can be modeled by a batch-service mechanism. Holding costs per job are incurred to provide an incentive to serve jobs promptly and thereby maintain an accurate perception of the system. On the other hand, a service fee (penalty) is incurred for the server to connect to a station and serve the jobs there. For generality, a reward earned per job served is also included. The server's objective is to maintain a appropriately up-to-date database for the system, while polling the stations infrequently as appropriate to keep overhead cost low (and to be available for other unmodeled tasks). The server has perfect memory but observes at any instant of time only the number of jobs at the node that is being polled.

Consider next an automated guided vehicle (AGV) or a shuttle in a manufacturing or transportation system. The shuttle transports jobs from one station to the next in a cyclic queueing system. There is no exchange of information amongst the queues. At each instant the shuttle must choose either to dispatch the jobs at the present queue or to wait at the present queue for more jobs to arrive at the system. There is a holding cost per unit time for each job in the system, a reward earned for each job that is served, and a switching cost describing the overhead required to serve a queue and switch to the next. The shuttle (or AGV) can transport all the jobs at a queue to the next node in one unit of time. At each time instant the shuttle knows perfectly the number of jobs in the node where it is waiting and has perfect memory.

A mathematical model and the optimization problem for the systems described above are the following: In discrete time, an infinite capacity batch server visits N queues, each with an i.i.d. arrival process, in a prespecified order. When queue n is the next to be served, the server observes the queue length at n and knows the elapsed time since each of the other queues was last visited. The server can process (dispatch) all the jobs at a queue in one unit of time or wait one or more units of time before serving queue n. Each job has a holding cost per unit time and a reward associated with its completion. A one-unit service time is required to serve the jobs at a queue, and a switching cost is incurred each time the server completes the service of a queue and moves to the next. The batch server controls the service times based on the available information. The objective is to characterize a server dispatching policy that is a function of the available information and minimizes an expected discounted cost due to waiting jobs and the dispatching of the server.

In the case of the information system, the above model allows the server to monitor the current queue length as it waits to serve that queue. In the case of the shuttle system, modeling of realistic systems becomes difficult quickly. The above model assumes that all jobs served at queue n are transported to the next node in the service order $(n \oplus 1)$ and leave the system at node $n \oplus 1$, that the shuttle has infinite capacity, and that the arrival process is memoryless. In some applications, the arrival process may be significantly correlated with the

batch service process and jobs in node n may have destinations other than node $n \oplus 1$.

The above model is related to literature where the server is a shuttle serving one or more terminals. An early treatment of batch service control subject to partial information is found in Ignall and Kolesar [11]. They formulated a two-node shuttle dispatching problem in which the shuttle carries only a single customer, observes only the number of customers at its location, and bases its dispatching decisions only on the queue length of the node where it is located. No attempt was made in [11] to utilize the inferential information about the unobserved terminal. Deb [9] was the first to provide a thorough characterization of an optimal scheduling policy in a *two queue* system with control at both nodes and perfect information (i.e. the lengths of both queues are known to the controller at all times). In Van Oyen and Teneketzis [24], we investigated in discrete time a two queue system where queue length information is exchanged between the nodes subject to a propagation delay. At a given terminal at any time t, the controller's (shuttle dispatcher's) decisions are based on the following information: (1) the shuttle's location at time t (referred to as the present node at t), (2) the queue length of the present node at time t, and (3) a delayed observation of the queue length at the other node at time $t - I$, where I is a nonnegative integer. Under the assumption that shuttle trip lengths exceed the information delay I, we proved that a threshold type policy is optimal: dispatch the shuttle from node n, $n \in \{1, 2\}$, if and only if the queue length at node n exceeds θ^n, a threshold which is a function of the most recent delayed observation of node $m \neq n$. The threshold functions θ^1 and θ^2 are monotone nonincreasing functions of the delayed observation. In addition, in [24] we derived additional conditions for dispatching which further reduce the computational effort required to obtain an optimal threshold policy. Whereas [9] and [24] allowed a general finite capacity batch server, the infinite capacity assumption is shared by a number of previous analyses. A theory for the optimal static control of single-station, infinite-capacity service systems is developed by Stidham [22]. Lee [14] addressed a two node network with imperfect information such as defined in [11], under the assumptions of an infinite capacity shuttle and compound batch arrival processes. As in [11], the *ad hoc* policy considered by Lee did not utilize inferential information regarding the state of the entire network. The issue of partial information was further discussed in Ignall and Kolesar [12], where the problem was simplified as follows: the controller observes only one terminal and no control decisions are exercised at the unobserved terminal (dispatching of the server is automatic upon its arrival at the unobserved terminal). In addition, the batch service capacity was assumed infinite. Ignall and Kolesar did not achieve any characterization of optimal policies for this problem.

To the best of our knowledge, the present paper is the first to present qualitative properties of optimal dynamic control policies for ring-type networks with an infinite capacity batch server under conditions of imperfect information. Through appropriate choice of an information state, we show that the dynamic

programming methodology which proved effective in Deb [9] and Van Oyen and Teneketzis [24] can be extended to ring-type queueing networks with an infinite capacity server and the additional feature of service rewards under conditions of imperfect information. Our contribution lies in developing a framework in which the problem can be analyzed. We determine properties of optimal policies that can be used to reduce the computational effort required to numerically determine an optimal policy. Furthermore, we provide deeper insight into the issues contained in such scheduling problems with imperfect information and can form a basis for approaching more realistic models in a variety of applications.

The remainder of the paper is organized as follows. We precisely formulate our model in Section 2. In Section 3 we characterize an optimal finite-horizon scheduling policy as a threshold-type policy, where the threshold functions are monotone functions of the available observations. We extend the analysis to the infinite-horizon criterion in Section 4. Conclusions are presented in Section 5, and suggestions for further research are made in Section 6.

2 Problem Formulation

We consider the following system in discrete time. A single infinite capacity server provides service according to a predetermined route (or service order) in a network with $N \geq 2$ nodes. Since the route is fixed, we assume that the nodes are labeled 0, 1, 2, ..., $N - 1$ and form a ring such that when the server completes service at node n it goes to node $n \oplus 1$, where $\oplus$ indicates addition modulo N. We assume that tasks (jobs, customers) arrive at each node according to an independent and identically distributed (i.i.d.) sequence of batch sizes, and that the arrival processes are independent of everything else. We denote by $a_n \triangleq (a_n(0), a_n(1), \ldots, a_n(M))$, $M \in \mathbb{Z}^+$, (where $\mathbb{Z}^+$ denotes the nonnegative integers) the probability vector governing the number of arrivals to node n at time t. The server exercises the following control actions: If at time t the server is at node n it has two options: (1) wait at node n ($U_t = 0$) until $t + 1$ when a new decision has to be made; or (2) serve all the tasks waiting at node n (including potential arrivals at t) and move to node $n \oplus 1$ ($U_t = 1$). We assume that if $U_t = 1$, the server takes one unit of time to serve all the tasks at node n and "moves" to queue $n \oplus 1$. For each unit of time a task waits at node n, a holding cost of c_n units is incurred, where $c_n \geq 0$. In addition, a *switching/ dispatching cost* K_n ($K_n \geq 0$) is incurred upon each move/dispatch of the server from node n. On the other hand, a *service reward* of S_n ($S_n \geq 0$) is received for each task served at node n. The server's decision at time t is based on all the observations available up to t and all the decisions (control actions) made up to $t - 1$. The server has imperfect information of the system state. If at time t the

server is at node n, it observes only the number of tasks waiting for service at n (including any arrivals at t); furthermore, at t the server has a probability distribution on the number of tasks waiting at other nodes since it knows the number of time units, e_m, that have elapsed since node $n \ominus m$ was last cleared ($\ominus$ denotes subtraction mod N).

Based on the above assumptions, the objective is to determine a non-anticipative policy g^* that maximizes the total expected discounted reward earned over a horizon $T \in \mathbb{Z}^+ \cup \{\infty\}$ due to service rewards, switching costs and holding costs. Let $X_n(t)$ be the queue length of node n at time t, $m(t)$ be the location of the server at t, and $\mathbb{1}(s)$ denote the indicator function of event s. The server's information state at t can be represented by the vector $(e, x, n) = (e_1, e_2, \ldots, e_{N-1}, x, n)$ where: the last component, n, indicates that the location of the server at t, x denotes the number of tasks present in node n at t, and $e \triangleq (e_1, e_2, \ldots, e_{N-1})$ denotes the number of time units since the server last moved from nodes $n \ominus 1$, $n \ominus 2$, $\ldots$, $n \oplus 1$ respectively (since $n \ominus (N - 1) = n \oplus 1$). Note that: (1) e and z define quantities relative to the present location n, and (2) because the server requires one unit of time to serve all the tasks in one node and to move to the next node, we have $e_1 \geq 1$ and $e_{i+1} \geq 1$ for $i = 1, 2, \ldots, N - 1$. Thus, the information state is an element of the following space: $\mathscr{S} \triangleq \{(e, x, n) \in (\mathbb{Z}^+)^{N-1} \times (\mathbb{Z}^+) \times \{0, 1, \ldots, N - 1\} : i \leq e_i \leq e_{i+1} - 1$ for $i = 1, 2, \ldots, N - 1\}$. Let g be an admissible policy of the server, and let $(e, x, n) \in \mathscr{S}$ be the initial state. Then, the total expected discounted reward over a horizon T using g is given by

$$
\begin{aligned}
W_g^T(e, x, n) = E_g \Bigg\{ &\sum_{t=0}^{T} \beta^t \sum_{m=0}^{N-1} \left[-c_m X_m(t) \mathbb{1}(U_t = 0) \right. \\
&- c_m X_m(t) \mathbb{1}(U_t = 1, n(t) \neq m) \\
&\left. + (S_m X_m(t) - K_m) \mathbb{1}(U_t = 1, n(t) = m) \right] | (e, x, n) \Bigg\}
\end{aligned}
\tag{2.1}
$$

where $0 < \beta \leq 1$ is the discount factor and E_g denotes the expectation under g. Let G be the set of admissible policies. As stated above, the objective is to determine $\inf_{g \in G} W_g^T(e, x, n)$ for any T, (e, x, n). Without any loss of optimality, we restrict attention to the class of policies which are functions of the information state (e, x, n) (see Chapter 6 of Kumar and Varaiya [13]). We begin with the finite horizon problem, and we extend our analysis to the infinite horizon problem afterwards.

3 The Finite Horizon Problem

We analyze the finite horizon case ($T < \infty$) of the problem formulated in Section 2 using stochastic dynamic programming. Assume $0 < \beta \leq 1$, where β is fixed.

Let $V_t(e, x, n)$ denote the maximum expected β-discounted reward-to-go from time t through T conditioned on the information state (e, x, n) at t. The optimality equation is

$$V_t(e, x, n) = \max[h_t(e, x, n), d_t(e, x, n)] , \tag{3.1}$$

where $h_t(e, x, n)$ $(d_t(e, x, n))$ is the maximum expected discounted reward-to-go from time t through T conditioned on state (e, x, n) and the decision to hold (dispatch) at t. Let $X_i(t)$ $(\overline{X}_i(t))$ denote the queue length of node i immediately prior to (respectively, following) the application of U_t. The functions h_t and d_t are given by

$$h_{T+1}(\cdot, \cdot, \cdot) \triangleq 0 , \tag{3.2}$$

$$h_t(e, x, n) = E\left\{ -\sum_{i=1}^{N} c_i X_i(t) + \right.$$

$$\beta V_{t+1}(1 + e_1, 1 + e_2, \ldots, 1 + e_{N-1}, X_n(t + 1), n) | U_t = 0,$$

$$\left. \overline{X}_n(t) = x, \overline{X}_{n \ominus l}(t - e_l) = 0, l = 1, 2, \ldots, N - 1 \right\} \tag{3.3}$$

$$d_{T+1}(\cdot, \cdot, \cdot) \triangleq 0 , \tag{3.4}$$

$$d_t(e, x, n) = S_n x - K_n + E\left\{ -\sum_{i=1}^{N} c_{n \ominus i} X_{n \ominus i}(t) + \right.$$

$$\beta V_{t+1}(1, 1 + e_1, 1 + e_2, \ldots, 1 + e_{N-2}, X_{n \oplus 1}(t + 1), n \oplus 1) | U_t = 1,$$

$$\left. \overline{X}_{n \ominus l}(t - e_l) = 0, l = 1, 2, \ldots, N - 1 \right\} . \tag{3.5}$$

Note that in (3.3) and (3.5), $\overline{X}_{n \ominus l}(t - e_l) = 0$ to indicate a dispatch from $n \ominus l$ at time $t - e_l$.

To avoid nonessential difficulties that may be introduced by a transient period in the evolution of the information state, we assume that at time $t = 0$ operation of the server is in steady state with respect to the available information. That is, we assume that the arrival process is defined for times indexed by $\{\ldots, -2, -1, 0, 1, 2, \ldots\}$ and consider any initial state in $\mathscr{S}$ at $t = 0$. Let $A_j^{e_i}$ denote the probability vector on the number of arrivals to node j in e_i units of time (that is, the e_i-fold convolution of a_j). Thus, for $k \in \{0, 1, \ldots, Me_i\}$, $A_{n \ominus i}^{e_i}(k)$ denotes the probability that $X_{n \ominus i}(t) = k$ given that node n was cleared at $t - e_i$. Letting

α_j denote the mean of a_j (equivalently A_j^1), we find that in the above case, $E\{X_{n\ominus i}(t)\} = \alpha_{n\ominus i}e_i$. We also find it convenient to define the operators $H(e)$ and $D(e)$ to denote the evolution of e during holding and dispatching respectively. For $e = (e_1, e_2, \ldots, e_{N-1})$, let

$$H(e) \triangleq (1 + e_1, 1 + e_2, \ldots, 1 + e_{N-1}) , \tag{3.6}$$

$$D(e) \triangleq (1, 1 + e_1, 1 + e_2, \ldots, 1 + e_{N-2}) . \tag{3.7}$$

We are now able to write (3.3) and (3.5) more explicitly as

$$h_t(e, x, n) = -c_n x - \sum_{i=1}^{N-1} c_{n\ominus i}\alpha_{n\ominus i}e_i$$

$$+ \sum_{k=0}^{M} A_n^1(k)\beta V_{t+1}(H(e), x + k, n) , \tag{3.8}$$

$$d_t(e, x, n) = S_n x - K_n - \sum_{i=1}^{N-1} c_{n\ominus i}\alpha_{n\ominus i}e_i$$

$$+ \sum_{k=0}^{M(1+e_{N-1})} A_{n\oplus 1}^{1+e_{N-1}}(k)\beta V_{t+1}(D(e), k, n \oplus 1) . \tag{3.9}$$

Our analysis rests on the properties of the "derivatives" of V_t, h_t, and d_t with respect to various components of the information state. Let $\delta(j)$ denote a $N - 1$ element vector with a 1 in the jth element and zeros elsewhere. We define for $j \in \{1, 2, \ldots, N - 1\}$, $(e, x, n) \in \mathcal{S}$, and $(e + \delta(j), x, n) \in \mathcal{S}$

$$\Delta_j V_t(e, x, n) = V_t(e + \delta(j), x, n) - V_t(e, x, n) , \tag{3.10}$$

$$\Delta_N V_t(e, x, n) = V_t(e, x + 1, n) - V_t(e, x, n) ; \tag{3.11}$$

Δh, Δd are defined similarly. These differences have clear interpretations. For example, if $N - 1 \geq 2$, $\Delta_2 d_t(e, x, n)$ denotes the incremental expected discounted reward to go (given the decision to dispatch at t) that is associated with an additional unit of elapsed time since the server was last dispatched from node $n \ominus 2$. Note that this implies a comparison of different control sequence histories. Because we focus on characterizing the dispatching decision faced at any given node, it suffices to investigate the relationships among the functions V, h, and d for any two states $(e^1, x^1, n) \in \mathcal{S}$ and $(e^2, x^2, n) \in \mathcal{S}$ at any given time t. The difference between any two such states can be written as a telescoping sum of single-unit incremental state changes. Thus, our analysis defines struc-

tural properties (over the state space $\mathscr{S}$) of the marginal revenue functions (3.10), (3.11), and their Δh, Δd counterparts. Using (3.2), (3.4), (3.8), and (3.9) we find that for $t \in \{1, 2, \ldots, T\}$, $i \in \{1, 2, \ldots, N - 1\}$, $j \in \{1, 2, \ldots, N - 2\}$, and $m \in \{1, 2, \ldots, N\}$

$$\Delta_m h_{T+1}(\cdot, \cdot, \cdot) = 0 \tag{3.12}$$

$$\Delta_m d_{T+1}(\cdot, \cdot, \cdot) = 0 \tag{3.13}$$

$$\Delta_i h_t(e, x, n) = -\alpha_{n \ominus i} c_{n \ominus i} + \sum_{k=0}^{M} A_n^1(k) \beta \Delta_i V_{t+1}(H(e), x + k, n) , \tag{3.14}$$

$$\Delta_N h_t(e, x, n) = -c_n + \sum_{k=0}^{M} A_n^1(k) \beta \Delta_N V_{t+1}(H(e), x + k, n) , \tag{3.15}$$

$$\Delta_j d_t(e, x, n) = -\alpha_{n \ominus j} c_{n \ominus j} + \sum_{k=0}^{M(1+e_{N-1})} A_{n \oplus 1}^{1+e_{N-1}}(k)$$

$$\cdot \beta \Delta_{j+1} V_{t+1}(D(e), k, n \oplus 1) \tag{3.16}$$

$$\Delta_{N-1} d_t(e, x, n) = -\alpha_{n \oplus 1} c_{n \oplus 1} + \sum_{k=0}^{M(1+e_{N-1})} A_{n \oplus 1}^{1+e_{N-1}}(k)$$

$$\cdot \left[\sum_{l=0}^{M} A_{n \oplus 1}^1(l) \beta V_{t+1}(D(e), k + l, n \oplus 1) \right.$$

$$\left. - \beta V_{t+1}(D(e), k, n \oplus 1) \right] \tag{3.17}$$

$$= -\alpha_{n \oplus 1} c_{n \oplus 1} + \sum_{k=0}^{M(1+e_{N-1})} A_{n \oplus 1}^{1+e_{N-1}}(k)$$

$$\cdot \left[\sum_{l=0}^{M} A_{n \oplus 1}^1(l) \sum_{m=0}^{l-1} \beta \Delta_N V_{t+1}(D(e), k + m, n \oplus 1) \right] , \tag{3.18}$$

$$\Delta_N d_t(e, x, n) = S_n , \tag{3.19}$$

where it is assumed in (3.14) that $(e + \delta(i), x, n) \in \mathscr{S}$ and in (3.16) that $(e + \delta(j), x, n) \in \mathscr{S}$.

Our derivation of threshold and monotonicity results rests on two basic results, which we quote.

R1: Let $h(a + 1)$, $h(a)$, $d(a + 1)$, $d(a) \in \mathbb{R}$, and $V(\cdot) \triangleq \max(h(\cdot), d(\cdot))$. If $h(a + 1) - h(a) \leq d(a + 1) - d(a)$, then

$$h(a + 1) - h(a) \leq V(a + 1) - V(a) \leq d(a + 1) - d(a) .$$

R2: Let h and d be functions mapping $\mathbb{Z}^+ \times \mathbb{Z}^+$ to $\mathbb{R}$ and define $V(a, b) = \max(h(a, b), d(a, b))$ for any $(a, b) \in \mathbb{Z}^+ \times \mathbb{Z}^+$. Define $\Delta_1 h(a, b) = h(a + 1, b) - h(a, b)$ and $\Delta_1 d(a, b)$ similarly. If the following properties hold for all $a, b \in \mathbb{Z}^+$:

(i) $\Delta_1 h(a, b + 1) \geq \Delta_1 h(a, b)$
(ii) $\Delta_1 d(a, b + 1) \geq \Delta_1 d(a, b)$
(iii) $\Delta_1 d(a, b) \geq \Delta_1 h(a, b)$
(iv) $d(a, b + 1) - d(a, b) \geq h(a, b + 1) - h(a, b)$;

then $\Delta_1 V(a, b + 1) \geq \Delta_1 V(a, b)$ for all $a, b \in \mathbb{Z}^+$.

Both R1 and R2 follow directly from Lemmas 2.2 and 2.3, respectively, of Van Oyen and Teneketzis [24]. Result R2 presents a sufficient condition for the maximum of two supermodular functions to be supermodular.

We now present Lemma 1, which reveals both the optimality of a threshold-type batch service policy as well as the monotonicity of the threshold functions.

Lemma 1: For any $t \in \{0, 1, \ldots, T\}$, $(e, x, n) \in \mathscr{S}$, and $k, j \in \{1, 2, \ldots, N - 1\}$ such that $k \neq j$:

(i) $\Delta_N d_t(e, x, n) \geq \Delta_N V_t(e, x, n) \geq \Delta_N h_t(e, x, n)$,
(ii) $\Delta_j d_t(e, x, n) \geq \Delta_j V_t(e, x, n) \geq \Delta_j h_t(e, x, n)$,

and the following supermodularity relations hold for the value function:

(iii) $\Delta_j V_t(e, x + 1, n) \geq \Delta_j V_t(e, x, n)$,
(iv) $\Delta_j V_t(e + \delta(k), x, n) \geq \Delta_j V_t(e, x, n)$,
(v) $\Delta_N V_t(e + \delta(j), x, n) \geq \Delta_N V_t(e, x, n)$.

It is assumed in (ii), (iii), (iv), and (v) that e is such that $e_j \leq e_{j+1} - 2$ so $\Delta_j d_t$ can be defined.

Proof: The proof proceeds by induction. The basis for induction ($t = T$) is established as follows. Since $\Delta_N d_T(e, x, n) = S_n \geq -c_n = \Delta_N h_T(e, x, n)$, (i) follows from R1. Similarly, $\Delta_j d_T(e, x, n) = -\alpha_{n \ominus j} c_{n \ominus j} = \Delta_j h_T(e, x, n)$ and R1 yields (ii). We prove (iii) by using R2 as follows. First note that $\Delta_j h_T$ and $\Delta_j d_T$ are constant in x; thus conditions R2(i) and R2(ii) are satisfied. The remaining conditions of R2, namely R2(iii) and R2(iv), are justified by Lemma 1(ii) and Lemma 1(i) respectively at T. Similar arguments using R2 justify properties (iv) and (v) of Lemma 1 at T.

Assume that the result holds at times $t + 1$, $t + 2$, $\ldots$, T. We proceed to complete the induction step. As will be seen, the proof of (i) and (ii) must precede the others. The proofs of (iii), (iv), and (v) are all similar and rest on R2.

Proof of (i): We demonstrate that $\Delta_N d_t(e, x, n) \geq \Delta_N h_t(e, x, n)$, then provide the justification for each step.

$$\Delta_N h_t(e, x, n) = -c_n + \sum_k A_n^1(k)\beta\Delta_N V_{t+1}(H(e), x + k, n)$$

$$\leq -c_n + \sum_k A_n^1(k)\beta\Delta_N d_{t+1}(H(e), x + k, n) \tag{3.20}$$

$$= -c_n + \beta S_n \tag{3.21}$$

$$\leq \Delta_N d_t(e, x, n) \;.$$

We begin using (3.15) in the first step and justify (3.20) by (i) of the induction hypothesis. Equation (3.21) follows by (3.19) and we conclude using (3.19) and the assumptions $c_N \geq 0$, $\beta \leq 1$.

Proof of (ii): We first prove the result when $j \neq N - 1$ (equivalently $n \ominus (N - 1) \neq n \oplus 1$), then the case $j = N - 1$.

Case 1: Suppose $j \neq N - 1$.

$$\Delta_j d_t(e, x, n) = -\alpha_{n\ominus j}c_{n\ominus j} + \sum_k A_{n\oplus 1}^{1+e_{N-1}}(k)\beta\Delta_{j+1} V_{t+1}(D(e), k, n \oplus 1)$$

$$\geq -\alpha_{n\ominus j}c_{n\ominus j} + \sum_k A_{n\oplus 1}^{1+e_{N-1}}(k)\beta\left[-\alpha_{(n+1)\ominus(j+1)}c_{(n+1)\ominus(j+1)} \right.$$

$$\left. + \sum_l A_{n\oplus 1}^1(l)\beta\Delta_{j+1} V_{t+2}(H(D(e)), k + l, n \oplus 1) \right] \tag{3.22}$$

$$= -\alpha_{n\ominus j}c_{n\ominus j} + \beta\left[-\alpha_{n\ominus j}c_{n\ominus j} \right.$$

$$\left. + \sum_m A_{n\oplus 1}^{2+e_{N-1}}(m)\beta\Delta_{j+1} V_{t+2}(H(D(e)), m, n \oplus 1) \right] \tag{3.23}$$

$$\geq -\alpha_{n\ominus j}c_{n\ominus j} + \beta\left[-\alpha_{n\ominus j}c_{n\ominus j} \right.$$

$$\left. + \sum_m A_{n\oplus 1}^{2+e_{N-1}}(m)\beta\Delta_{j+1} V_{t+2}(D(H(e)), m, n \oplus 1) \right] \tag{3.24}$$

$$= -\alpha_{n \ominus j} c_{n \ominus j} + \beta \left[\sum_k A_n^1(k) \Delta_j d_{t+1}(H(e), x + k, n) \right] \tag{3.25}$$

$$\geq -\alpha_{n \ominus j} c_{n \ominus j} + \sum_k A_n^1(k) \beta \Delta_j V_{t+1}(H(e), x + k, n) \tag{3.26}$$

$$= \Delta_j h_t(e, x, n) \; . $$

The first step simply restates (3.16). Equation (3.22) follows from (ii) of the induction hypothesis $(\Delta_{j+1} V_{t+1} \geq \Delta_{j+1} h_{t+1})$ and (3.14). We rewrite the expectation to get (3.23). Noting that $H(D(e)) = D(H(e)) + \delta(1)$, we get (3.24) by (iv) (convexity with respect to e_1 is not required since $j + 1 \geq 2$ under the assumption of Case 1). Equation (3.25) follows from (3.16). We use (ii) of the induction hypothesis to justify (3.26) and conclude using (3.14).

Case 2: Suppose $j = N - 1$.

$$\Delta_{N-1} d_t(e, x, n) = -\alpha_{n \oplus 1} c_{n \oplus 1} + \sum_k A_{n \oplus 1}^{1 + e_{N-1}}(k)$$

$$\cdot \left[\sum_l A_{n \oplus 1}^1(l) \sum_{m=0}^{l-1} \beta \Delta_N V_{t+1}(D(e), k + m, n \oplus 1) \right]$$

$$\geq -\alpha_{n \oplus 1} c_{n \oplus 1} + \beta \sum_k A_{n \oplus 1}^{1 + e_{N-1}}(k) \left[\sum_l A_{n \oplus 1}^1(l) \sum_{m=0}^{l-1} \right. \tag{3.27}$$

$$\left. \cdot \left(-c_{n \oplus 1} + \sum_p A_{n \oplus 1}^1(p) \beta \Delta_N V_{t+2}(H(D(e)), k + m + p, n \oplus 1) \right) \right]$$

$$\geq -\alpha_{n \oplus 1} c_{n \oplus 1} + \beta \left(-\alpha_{n \oplus 1} c_{n \oplus 1} + \sum_i A_{n \oplus 1}^{2 + e_{N-1}}(i) \right.$$

$$\left. \cdot \sum_l A_{n \oplus 1}^1(l) \sum_{m=0}^{l-1} \beta \Delta_N V_{t+2}(D(H(e)), m + i, n \oplus 1) \right) \tag{3.28}$$

$$= -\alpha_{n \oplus 1} c_{n \oplus 1} + \sum_k A_n^1(k) \beta \Delta_{N-1} d_{t+1}(H(e), x + k, n) \tag{3.29}$$

$$\geq \Delta_{N-1} h_t(e, x, n) \; . \tag{3.30}$$

We begin by stating (3.18) and noting $n \ominus (N - 1) = n \oplus 1$. We get (3.27) by using (i) of the induction hypothesis $(\Delta_N V_{t+1} \geq \Delta_N h_{t+1})$ and then applying (3.15). In addition to simplifying terms, we use (v) of the induction hypothesis (since

$H(D(e)) = D(H(e)) + \delta(1))$ to justify the inequality of (3.28). We use (3.18) to get (3.29) (where the expectation with respect to A_n^1 and the argument $x + k$ are both superfluous). Equation (3.30) follows by (ii) of the induction hypothesis and (3.14).

Proof of (iii): We prove (iii) using R2. We begin by proving the supermodularity of h_t and d_t. Using (iii) of the induction hypothesis we get

$$\Delta_j h_t(e, x + 1, n) = -\alpha_{n \ominus j} c_{n \ominus j} + \sum_k A_n^1(k) \beta \Delta_j V_{t+1}(H(e), x + k + 1, n)$$

$$\geq -\alpha_{n \ominus j} c_{n \ominus j} + \sum_k A_n^1(k) \beta \Delta_j V_{t+1}(H(e), x + k, n)$$

$$= \Delta_j h_t(e, x, n) \ .$$

The supermodularity of d_t is trivial. If $j \neq N - 1$, (3.16) shows $\Delta_j d_t(e, x + 1, n) = \Delta_j d_t(e, x, n)$; otherwise (3.18) justifies this conclusion if $j = N - 1$.

Thus, conditions R2(i) and R2(ii) hold. Conditions R2(iii) and R2(iv) hold by properties (ii) and (i), respectively, of Lemma 1 at time t. Consequently, by R2, $\Delta_j V_t(e, x + 1, n) \geq \Delta_j V_t(e, x, n)$.

Proof of (iv): We use R2 to prove that (iv) is valid at t. Property (iv) of the induction hypothesis yields the supermodularity of h_t. We prove the supermodularity of d_t using three cases. First, suppose $j \neq N - 1, k \neq N - 1$.

$$\Delta_j d_t(e + \delta(k), x, n)$$

$$= -\alpha_{n \ominus j} c_{n \ominus j} + \sum_l A_{n \oplus 1}^{1 + e_{N-1}}(l) \beta \Delta_{j+1} V_{t+1}(D(e + \delta(k)), l, n \oplus 1)$$

$$\geq -\alpha_{n \ominus j} c_{n \ominus j} + \sum_l A_{n \oplus 1}^{1 + e_{N-1}}(l) \beta \Delta_{j+1} V_{t+1}(D(e), l, n \oplus 1)$$

$$= \Delta_j d_t(e, x, n) \ .$$

The inequality follows from (iv) at $t + 1$ since $D(e + \delta(k)) = D(e) + \delta(k + 1)$ and $j \neq k$.

Second, suppose $j \neq N - 1, k = N - 1$.

$$\Delta_j d_t(e + \delta(N - 1), x, n)$$

$$= -\alpha_{n \ominus j} c_{n \ominus j} + \sum_l A_{n \oplus 1}^{1 + e_{N-1}}(l) \sum_m A_{n \oplus 1}^1(m) \beta \Delta_{j+1} V_{t+1}(D(e), l + m, n \oplus 1)$$

$$\geq -\alpha_{n-j}c_{n-j} + \sum_l A_{n\oplus 1}^{1+e_{N-1}}(l)\beta\Delta_{j+1}V_{t+1}(D(e), l, n\oplus 1)$$

$$= \Delta_j d_t(e, x, n) \ .$$

Here we have used the fact that $D(e + \delta(N-1)) = D(e)$ and justified the inequality by (iii) of the induction hypothesis.

The final case, $j = N - 1$, $k \neq N - 1$, follows similarly from (3.18) and (v) of the induction hypothesis.

Thus, we have established R2(i) and R2(ii). Both R2(iii) and R2(iv) hold by (ii) of Lemma 1 at t. Part (iv) then follows from result R2.

Proof of (v): The supermodularity of h_t follows from (v) of the induction hypothesis. The supermodularity of d_t follows from $\Delta_N d_t(e + \delta(j), x, n) = S_N = \Delta_N d_t(e, x, n)$. Result R2 yields the supermodularity of V_t, since R2(iii) and R2(iv) follow from properties (i) and (ii) of Lemma 1, respectively, which have already been established at t. □

Statements (i) and (ii) of Lemma 1 define the structural properties of an optimal service policy. Specifically, if it is optimal to dispatch the server at state (e, x, n) at t, (i) indicates that dispatching is also optimal at state (e, x', n), where $x' > x$, since dispatching results in a greater incremental reward than does holding. Similarly, (ii) indicates that the incentive to dispatch is increasing in e_j, the time elapsed since node $n \ominus j$ was last cleared. We proceed to state these implications precisely in the main result, Theorem 1. Since the result follows directly from Lemma 1, we omit a formal proof.

Theorem 1: For $t, e \in (\mathbb{Z}^+)^{N-1}, x \in \mathbb{Z}^+, n \in \{0, 1, \ldots, N - 1\}$ there exists a threshold function, θ_t^n, defined by

$$\theta_t^n(e_1, e_2, \ldots, e_{N-1})$$

$$\triangleq \inf\{z \in \mathbb{Z}^+ : d_t(e_1, e_2, \ldots, e_{N-1}, z, n) > h_t(e_1, e_2, \ldots, e_{N-1}, z, n)\} \ , \quad (3.31)$$

such that the following service policy is optimal: if at t the state is (e, x, n),

$$\text{clear node } n \text{ if and only if } x \geq \theta_t^n(e_1, e_2, \ldots, e_{N-1}) \ . \quad (3.32)$$

Moreover, for any $j \in \{1, 2, \ldots, N - 1\}$ such that $(e + \delta(j), x, n) \in \mathscr{S}$,

$$\theta_t^n(e_1, e_2, \ldots, e_{N-1}) \geq \theta_t^n(e_1, e_2, \ldots, e_{j-1}, e_j + 1, e_{j+1}, \ldots, e_{N-1}) \ . \quad (3.33)$$

Remark: We refer to (3.32) as the threshold property and to (3.33) as the monotonicity of the threshold functions.

As we have shown, the threshold and monotonicity properties rest on the underlying supermodularity of the value function. This was also the case in [24], the two node problem with a finite-capacity server and delayed information. Unlike that problem, however, the infinite batch-service capacity assumed here results in a value function which is convex. This is shown in the following theorem.

Theorem 2: For any $t \in \{0, 1, \ldots, T\}$, $(e, x, n) \in \mathscr{S}$, and $k, j \in \{1, 2, \ldots, N - 1\}$ such that $k \neq j$ the following convexity relations hold for the value function:

(i) $\Delta_N V_t(e, x + 1, n) \geq \Delta_N V_t(e, x, n)$,
(ii) $\Delta_j V_t(e + \delta(j), x, n) \geq \Delta_j V_t(e, x, n)$.

Proof: The theorem holds trivially at T, (see (3.12) and (3.13)) and the induction steps follow.

Proof of (i): We begin by demonstrating the convexity of d_t and h_t in x. That d_t is convex in x is obvious from (3.19). On the other hand, since V_{t+1} is convex in x ((i) of the induction hypothesis) it is clear from (3.8) that h_t is convex in x. Since the maximum of two convex functions is convex, (i) holds at t.

Proof of (ii): The proof is similar to that of (i). This convexity property in e_j can easily be established for h_t using (3.14) and (ii) of the induction hypothesis. The convexity of d_t requires two cases. The first case, $j = N - 1$, is as follows:

$$
\Delta_{N-1} d_t(e + \delta(N - 1), x, n) = -\alpha_{n \oplus 1} c_{n \oplus 1} + \sum_k A_{n \oplus 1}^{1 + e_{N-1}}(k) \left[\sum_p A_{n \oplus 1}^1(p) \right.
$$

$$
\left. \cdot \sum_l A_{n \oplus 1}^1(l) \sum_{m=0}^{l-1} \beta \Delta_N V_{t+1}(D(e), k + p + l, n \oplus 1) \right]
$$

$$
\geq -\alpha_{n \oplus 1} c_{n \oplus 1} + \sum_k A_{n \oplus 1}^{1 + e_{N-1}}(k)
$$

$$
\cdot \left[\sum_l A_{n \oplus 1}^1(l) \sum_{m=0}^{l-1} \beta \Delta_N V_{t+1}(D(e), k + l, n \oplus 1) \right]
$$

$$
= \Delta_{N-1} d_t(e, x, n) \ .
$$

We have used $D(e, \delta(N-1)) = D(e)$ and have justified the inequality by (i) of the induction hypothesis.

The remaining case, $j \neq N-1$, is similar and can be shown using (3.16) and (ii) of the induction hypothesis. □

Properties of the value function such as convexity and supermodularity may be useful in reducing the computational effort required to determine an optimal policy.

4 The Infinite Horizon Problem

We use the technique of value iteration (or successive approximation) to extend the results of Section 3 to the case of an infinite planning horizon: $T = \infty$. We assume the discount factor is fixed and $\beta < 1$. We show that the value function of the T-stage problem converges to the infinite horizon value function as $T \to \infty$. Using this, we prove that an optimal infinite horizon policy is pure Markov, stationary, and inherits the threshold and monotonicity properties stated in Theorem 1 for the finite horizon case.

From (2.1) it is clear that $W_g^T(e, x, n)$ is finite (although unbounded in the state) for any $T < \infty$. We proceed to show that $W_g^\infty(e, x, n)$ is finite.

Lemma 2: For any initial state $(e, x, n) \in \mathscr{S}$ and any $g \in G$, $|W_g^\infty(e, x, n)| < \infty$.

Proof: We begin by defining a bound, $tB(e, x, n)$, on the magnitude of the instantaneous reward at time $t \geq 1$. Let $B(e, x, n)$ be the larger of the maximum rate at which service rewards are earned, say B', or the maximum rate at which holding costs and switching costs are incurred, say B''. Let $y \triangleq \max[x, e_1 M, e_2 M, \ldots, e_{N-1} M]$, $K \triangleq \max_n K_n$, $S \triangleq \max_n S_n$, $c \triangleq \max_n c_n$. With $B' \triangleq S(y + M)$ and $B'' \triangleq K + Nc(y + M)$, let $B(e, x, n) = \max[B', B'']$. Then, for any $g \in G$, noting that $B(e, x, n)$ is a bound for the case $t = 0$ we have

$$|W_g^\infty(e, x, n)| \leq B(e, x, n)\left(1 + \sum_{t=1}^{\infty} t\beta^t\right)$$

$$= B(e, x, n)(1 + \beta(1 - \beta)^{-2}) \ . \qquad \square$$

Define the infinite horizon value function to be $V(e, x, n) = \sup_g W_g^\infty(e, x, n)$. Lemma 2 allows us to directly conclude that $V(e, x, n)$ is finite for any $(e, x, n) \in$

$\mathscr{S}$. More significantly, the bound on instantaneous rewards that was previously obtained in the proof of Lemma 2 allows us to study the infinite horizon value function as a limit of finite horizon value functions. Let $V_0^T(e, x, n)$ denote the value function of the T-horizon problem at time 0.

Lemma 3: For any initial state $(e, x, n) \in \mathscr{S}$ and $t \in \mathbb{Z}^+$,

$$\lim_{T \to \infty} V_t^T(e, x, n) = V(e, x, n) \ . \tag{4.1}$$

Proof: We mimic the proof of Proposition 3.1(i) of Ross [19]. Let g^* denote the optimal infinite horizon policy. Then from (2.1) we have

$$V(e, x, n) = E_{g*}\{\text{return through stage } T\}$$

$$+ E_{g*}\{\text{return over } [T + 1, \infty) | g^* \text{ on } [0, T]\}$$

$$\leq V_0^T(e, x, n) + E_{g*}\{\text{return over } [T + 1, \infty) | g^* \text{ on } [0, T]\} \ . \tag{4.2}$$

On the other hand, policy g^* is by definition superior to a policy which follows the optimal T-stage polity (g^T) through time T, then switches to g^*:

$$V(e, x, n) \geq V_0^T(e, x, n) + E_{g*}\{\text{return over } [T + 1, \infty) | g^T \text{ on } [0, T]\} \ . \tag{4.3}$$

Using (4.2), (4.3), and the bounds on instantaneous reward as defined in the proof of Lemma 2, we get the convergence of V_0^T in T:

$$|V(e, x, n) - V_0^T(e, x, n)| \leq \sum_{t=T+1}^{\infty} t\beta^t B(e, x, n)$$

$$= \beta^T(T(1 - \beta) + 1)B(e, x, n)\beta(1 - \beta)^{-2} \ ,$$

which vanishes in T since $\beta < 1$ and $B(e, x, n)$ is finite. Because the cost criterion and system dynamics do not vary in time, the result holds for any t.   □

Let $d(e, x, n)$ $(h(e, x, n))$ denote the maximum infinite-horizon expected discounted reward conditioned on $U_{t=0} = 1$ $(U_{t=0} = 0)$. Using (3.8), (3.9), Lemma 3, and the fact that only a finite number of states can be reached at time 1 from (e, x, n) at 0, we get

$$\lim_{T \to \infty} h_0^T(e, x, n) = h(e, x, n) \tag{4.4}$$

$$\lim_{T \to \infty} d_0^T(e, x, n) = d(e, x, n) \ . \tag{4.5}$$

Moreover, the optimality equation (3.2) holds in the limit as $T \to \infty$:

$$V(e, x, n) = \max[h(e, x, n), d(e, x, n)] \ . \tag{4.6}$$

As the argument of Theorem 2.2 of Ross [19] indicates, a stationary pure Markov policy that agrees with the optimality equation is optimal. Lemma 3 together with (4.4) and (4.5) imply that V, h, and d inherit all the properties of their finite horizon counterparts as stated in Lemma 1 and Theorem 2. This together with the infinite horizon optimality equation yields Theorem 3, the main result for the infinite horizon problem.

Theorem 3: The following stationary service policy is optimal when $T = \infty$: If in state $(e, x, n) \in \mathcal{S}$ at $t \in \mathbb{Z}^+$,

clear node n at t if and only if $x \geq \theta^n(e_1, e_2, \ldots, e_{N-1})$, where $\tag{4.7}$

$\theta^n(e_1, e_2, \ldots, e_{N-1})$

$$\triangleq \inf\{z \in \mathbb{Z}^+ : d(e_1, e_2, \ldots, e_{N-1}, z, n) > h(e_1, e_2, \ldots, e_{N-1}, z, n)\} \ . \tag{4.8}$$

Moreover, for any $j \in \{1, 2, \ldots, N-1\}$ such that $(e + \delta(j), x, n) \in \mathcal{S}$,

$$\theta^n(e_1, e_2, \ldots, e_{N-1}) \geq \theta^n(e_1, e_2, \ldots, e_{j-1}, e_j + 1, e_{j+1}, \ldots, e_{N-1}) \ . \tag{4.9}$$

Furthermore, the value function is convex; that is, for $k, j \in \{1, 2, \ldots, N-1\}$ such that $k \neq j$,

$$\Delta_N V_t(e_1, e_2, \ldots, e_{N-1}, x + 1, n) \geq \Delta_N V_t(e_1, e_2, \ldots, e_{N-1}, x, n) \ , \tag{4.10}$$

$$\Delta_j V_t(e_1, e_2, \ldots, e_{j-1}, e_j + 1, e_{j+1}, \ldots, e_{N-1}, x, n)$$

$$\geq \Delta_j V_t(e_1, e_2, \ldots, e_{N-1}, x, n) \ . \tag{4.11}$$

Discussion: The threshold property, (4.7) is implied by $\Delta_N d \geq \Delta_N h$. Similarly, the monotonicity property of (4.8) is implied by $\Delta_j d \geq \Delta_j h$. It is interesting to ob-

serve the relationship of the finite horizon threshold to the infinite. Let $\theta_0^n(e, T)$ denote the T-horizon threshold at time 0. If $\lim_{T \to \infty} \theta_0^n(e, T)$ exists, it is easy to show that that this limit is $\theta^n(e)$. The difficulty is that, under certain circumstances, the limit may not exist. Specifically, for states in which two control actions are optimal, the following technical difficulty can arise: Suppose that $d(e, x, n) = h(e, x, n)$ and $\theta^n(e) = x + 1$. The limit of $\theta_0^n(e, T)$ may fail to exist in such a case if, for example, $h_0^T(e, x, n) \geq d_0^T(e, x, n)$ if and only if $T \geq T'$ and T is odd. In this case, $\theta_0^n(e, T)$ equals $x + 1$ for T large and odd, and equals x for T large and even.

5   Conclusions

We modified the problem of Van Oyen and Teneketzis [24] to include a ring network of arbitrary finite size, an infinite capacity server, job service rewards, and a network with no information flow among the nodes. We proved the optimality and monotonicity of threshold policies in both finite and infinite horizon problems. Moreover, the value function was shown to be convex. The dynamic programming equations explicitly indicate the manner in which inferential information is to be used. In this problem (as well as in [24]), the resolution of the dual aspect of control is implicit in the optimal threshold functions. Both models assumed a cyclic service of the queues (which is quite reasonable for some AGV systems and many polling systems). It is our hope that the understanding gained from our contributions will provide a basis for the resolution of similar issues in information networks and other systems.

References

[1] Altman E, Nain P (1992) Closed-loop control with delayed information. Presented at ACM Sigmetrics – Performance '92 Conference, Newport, RI
[2] Artiges D (1994) Optimal routing into two heterogeneous service stations with delayed information. Preprint, INRIA
[3] Barnett A (1973) On operating a shuttle service. Networks 3:305–313
[4] Bertsekas DP (1987) Dynamic programming: Deterministic and stochastic models. Prentice-Hall, Englewood Cliffs, NJ

[5] Beutler FJ, Teneketzis D (1989) Routing in queueing networks under imperfect information: Stochastic dominance and thresholds. Stochastics and Stochastics Reports 26:81–100

[6] Bournas RM, Beutler FJ, Teneketzis D (1993) Optimal flow control allocation policies in communication networks with multiple message classes. IEEE Transactions on Automatic Control 38:390–403

[7] Bournas RM, Beutler FJ, Teneketzis D (1991b) Properties of optimal hop-by-hop allocation policies in networks with multiple transmitters and linear equal holding costs. IEEE Transactions on Automatic Control AC-36:1450–1463

[8] Bournas RM, Beutler FJ, Teneketzis D (1992) Time average and asymptotically optimal flow control policies in networks with multiple transmitters. Annals of Operations Research 35: 327–355

[9] Deb RK (1978) Optimal dispatching of a finite capacity shuttle. Management Science 24: 1362–1372

[10] Dror IM (1978) Shuttle systems with one carrier and two passenger queues. Ph.D. Thesis, Columbia University, New York, NY

[11] Ignall E, Kolesar P (1972) Operating characteristics of a simple shuttle under local dispatching rules. Operations Research 20:1077–1088

[12] Ignall E, Kolesar P (1974) Optimal dispatching of an infinite-capacity shuttle: Control at a single terminal. Operations Research 22:1008–1024

[13] Kumar PR, Varaiya P (1986) Stochastic systems: Estimation, identification, and adaptive control. Prentice-Hall, Englewood Cliffs, NJ

[14] Lee H-S Control policies for queueing and production/inventory systems. Ph.D. Thesis, University of Michigan

[15] Levy H, Sidi M (1990) Polling systems: Applications, modelling, and optimization. IEEE Trans. Commun. 38:1750–1760

[16] Liu Z, Nain P, Towsley D (1992) On optimal polling policies. Queueing Systems (QUESTA) 11:59–83

[17] Rajan R, Agrawal R (1994) Optimal server allocation in homogeneous queueing systems with switching costs. Tech. Rep. ECE-94-2, Department of Electrical and Computer Engineering, University of Wisconsin-Madison

[18] Rosberg Z, Varaiya P, Walrand J (1982) Optimal control of service in tandem queues. IEEE Transactions on Automatic Control AC-27:600–609

[19] Ross S (1983) Introduction to stochastic dynamic programming. Academic Press, New York, NY

[20] Sennot LI (1989) Average cost optimal stationary policies in infinite state markov decision processes with unbounded costs. Operations Research 37:626–633

[21] Sim SH, Templeton JGC (1983) Computational procedures for steady-state characteristics of unscheduled multi-carrier shuttle systems. European J. of Operational Research 12:190–202

[22] Stidham S Jr (1977) Cost models for stochastic clearing systems. Operations Research 25: 100–127

[23] Takagi H (1994) Queueing analysis of polling models: Progress in 1990–1993. In J.H. Dshalalow (Ed.). Frontiers in queueing: Models, methods and problems. CRC Press

[24] Van Oyen MP, Teneketzis D (1992) Optimal scheduling of a finite capacity shuttle under delayed information. Probability in the Engineering and Informational Sciences 6:29–61

[25] Weber RR, Stidham S Jr (1987) Optimal control of service rates in networks of queues. Advances in Applied Probability 19:202–218