# The throughput of irreducible closed Markovian queueing networks: Functional bounds, asymptotic loss, efficiency, and the Harrison-Wein conjectures*†

H. Jin‡, J. Ou§, and P. R. Kumar ‡

**Abstract**

Let $N$ be the population of an irreducible closed Markovian queueing network, and denote by $\alpha^u(N)$ the throughput of a scheduling policy $u$. The policy $u$ is said to be *efficient* if $\lim_{N\to+\infty} \alpha^u(N) = \alpha^*$, where $\alpha^*$ is the capacity of a bottleneck station. The quantity $J(u) := \lim_{N\to\infty} \frac{N(\alpha^* - \alpha^u(N))}{\alpha^*}$ is called the *asymptotic loss* of $u$. The policy $u$ is said to be *asymptotically optimal* if $J(u)$ is as small as it can be.

For multi-station irreducible closed Markovian networks, we obtain functional bounds on the throughput which are of the form $\frac{\alpha N}{N+v}$. The coefficients $\alpha$ and $v$ are obtained by solving two linear programs (LPs), which consist of $\frac{L(L+3)}{2}$ variables, where $L$ is the number of buffers in the system. We are thus able to establish efficiency, as well as provide upper and lower bounds on asymptotic loss, through LP procedures.

For balanced systems where all stations are equally loaded, we are able to provide reduced dimensional LPs which consist of just $\frac{S(S+1)}{2}$ variables, where $S$ is the number of stations in the system. We also show that the lower bound on asymptotic loss produced by the reduced dimensional LP can capture interactions between multiple bottlenecks in heavy traffic, and that it always dominates a bound obtainable by extending a conjecture of Harrison and Wein (HW) from two station systems to multi-station systems.

*Please address all correspondence to the third author.

†The research reported here has been supported by the National Science Foundation under Grant Nos. ECS-94-03571, the U.S. Army Research Office under Contract No. DAAL-03-91-G-0182, and the Joint Service Electronics Program under Contract No. N00014-90-J-1270.

‡Department of Electrical and Computer Engineering, and the Coordinated Science Laboratory, University of Illinois, 1308 West Main Street, Urbana, IL 61801.

§Department of Decision Sciences, Faculty of Business Administration, National University of Singapore, Singapore 0511. This work was done while the author was visiting the University of Illinois.

Employing these bounds on two station systems, we prove that a certain policy developed by HW is efficient. This buffer priority policy was conjectured by them to be asymptotically optimal. They also conjectured a finite valued formula for the asymptotic loss of every buffer priority policy. This latter conjecture is not true in its full generality since there exists a certain buffer priority policy for a certain two station system which is not efficient, and consequently has infinite asymptotic loss. However, we provide credence to the conjectures by establishing that for balanced systems, the conjectured asymptotic loss of the HW-policy is a lower bound on the asymptotic loss of all policies, while the conjectured formula for the asymptotic loss applied to the exact opposite of the HW-policy is an upper bound on the asymptotic loss for all non-idling policies. The latter result is established under an additional condition identified by us which guarantees the finiteness of the asymptotic loss, and thus also the efficiency, of every non-idling scheduling policy.

**Key Words:** Queueing networks, closed networks, optimal scheduling, performance evaluation, throughput, efficiency, asymptotic loss, asymptotic optimality.

# 1 Introduction

We address the problem of determining the throughput of irreducible closed Markovian queueing networks. A queueing network is called "closed" if no customers arrive to or depart from it, and there are a certain number of trapped customers, $N$, which circulate endlessly. It is said to be Markovian if all service times are exponentially distributed, and it is said to be irreducible if the routing matrix is so. The throughput of such a network is the rate at which customers circulate. Such closed networks arise in many important applications. In communications networks they arise from the use of window based flow control schemes, see Walrand (1991). In manufacturing systems they arise when part releases follow what is called a "closed-loop" release policy, see Wein (1988). They also arise in several models of computer systems, see Lavenberg (1983).

Except for the limited class of product form networks, little appears to be known about the throughput $\alpha^u(N)$ under a scheduling policy $u$ when the number of trapped customers is $N$ (defined precisely in Section 2). In Kumar and Kumar (1994), lower and upper bounds are obtained for the throughput $\alpha^u(N)$ for each *fixed* $N$. The issue has been raised there of whether for a given scheduling policy the throughput converges to its maximum possible value $\alpha^*$, which is equal to the bottleneck capacity, as the number of trapped customers $N$

is increased to infinity. Let us say that a scheduling policy is *efficient* if this is the case. Sufficient conditions are given in Kumar and Kumar (1994) for establishing efficiency. Also, examples are provided where these sufficient conditions are not met. However, a rigorous proof of inefficiency was not available. In Harrison and Nguyen (1995) it is proved that the closed version of the counterexample for open systems given in Kumar and Seidman (1990) and Lu and Kumar (1991) is not efficient.

Let us call $\overline{J}(u) = \lim\sup_{N\to+\infty} N\frac{(\alpha^*-\alpha^u(N))}{\alpha^*}$ and $\underline{J}(u) = \lim\inf_{N\to+\infty} N\frac{(\alpha^*-\alpha^u(N))}{\alpha^*}$ the upper and lower *asymptotic losses* of a policy $u$. We say that a policy $u$ is *asymptotically optimal* if these numbers are as small as possible. To understand what the asymptotic loss is capturing, one may note that if $\alpha^u(N) = \alpha^*\left(1 - \frac{J(u)}{N} + O\left(\frac{1}{N^2}\right)\right)$, then the asymptotic loss is the constant $J(u)$. While a policy which is asymptotically optimal may not necessarily be optimal for any value of $N$, it may be an acceptably good policy for large or even moderate values of $N$.

For systems consisting of just two stations, Harrison and Wein (HW) (1990) have studied the Reflected Brownian Motion assumed to approximate a "workload imbalance process" in heavy traffic. Based on this they made two contributions. First, they synthesized a certain buffer priority policy, hereafter called the HW-policy and denoted by $\theta_{HW}$, which they conjectured was asymptotically optimal. Second, they conjectured an expression, always finite valued, for the asymptotic loss of any buffer priority policy $\theta$. Chevalier and Wein (1993) have sought to conduct a similar heavy traffic study for closed systems with multiple stations. However a rigorous basis is lacking for any of these results. Indeed the conjectures are not true in their full generality since a certain buffer priority policy for a certain two station system has been proved to be not efficient by Harrison and Nguyen (1995). The asymptotic loss of this policy is therefore $+\infty$, and so the conjectured finite valued formula for the asymptotic loss of every buffer priority policy is false. Concerning the HW-policy, no result has so far been available concerning its efficiency, let alone its asymptotic optimality.

In this paper we obtain the following results:

**(i)** We obtain linear programs (LPs) to obtain bounds on the asymptotic loss for irreducible closed Markovian systems with multiple stations. These LPs, called *functional bound LPs*, provide constants $\underline{\alpha}$, $\underline{v}$ and $\bar{v}$ such that (Theorem 4),

$$\frac{N}{N+\underline{v}}\alpha^* \geq \alpha^u(N) \geq \underline{\alpha}\frac{N}{N+\bar{v}}.$$

As a consequence, a lower bound on asymptotic loss is $\underline{J}(u) \geq \underline{v}$. Also, provided $\underline{\alpha}$ is equal to $\alpha^*$, an upper bound on asymptotic loss is $\overline{J}(u) \leq \bar{v}$.

**(ii)** While the above LPs consist of $\frac{L(L+3)}{2}$ variables where $L$ is the number of buffers in the system, for systems which are balanced, i.e., all stations are equally loaded, we provide reduced dimensional LPs consisting of only $\frac{S(S+1)}{2}$ variables for obtaining functional bounds (Theorems 5 and 7). Here $S$ is the number of stations in the system. It is far smaller than $L$ in many applications of interest. In some cases, the reduced dimensional LP yields the same result as the larger dimensional functional bound LP (Example 2).

**(iii)** We show that the lower bound on asymptotic loss produced for balanced systems by the reduced dimensional LP always dominates a bound obtained by extending to multi-station systems the two station bound conjectured by HW (Theorem 6). We also show by an example that the lower bound on asymptotic loss furnished by the reduced dimensional LP can capture interactions between multiple bottlenecks in heavy traffic (Example 3).

**(iv)** Studying the reduced dimensional LP for balanced systems we identify a condition under which every non-idling scheduling policy is efficient, and provide an upper bound on asymptotic loss (Theorem 7).

**(v)** Turning to two station systems, we show that the HW-policy is efficient and provide an upper bound on its asymptotic loss (Theorem 8).

**(vi)** For balanced two station systems, we prove that no policy can have a lower asymptotic loss than that conjectured by HW for their conjectured asymptotically optimal policy

(Theorem 9).

**(vii)** For balanced two station systems, under the condition identified in (iv), we show that no non-idling policy can have a higher asymptotic loss than the conjectured formula for the asymptotic loss applied to the exact opposite of the conjectured optimal policy (Theorem 10).

Our results for two station systems do not completely prove all of the conjectures made by HW; indeed they are not all true as mentioned earlier. However, we do lend credence to their conjectures for balanced systems by establishing that the minimal and maximal asymptotic losses as per their conjectures are indeed lower and upper bounds respectively; the latter under a condition which guarantees that the asymptotic loss is finite for all non-idling policies. Some such additional condition needs necessarily to be imposed to obtain a finite upper bound on asymptotic loss in view of the counterexample of Harrison and Nguyen (1995). We conjecture that under the additional condition identified by us, all of the conjectures of HW are true for balanced two station systems.

The rest of this paper is organized as follows. In Section 2 we provide a description of the system and define its throughput. In Section 3 we provide the basic LPs for the bounds on the throughput for fixed $N$. In Section 4 we study the infinite population limit of these LPs, and in Section 5 their duals. In Section 6 we provide a fundamental identity and fundamental inequality. In Section 7, we obtain the functional bound LPs by utilizing this fundamental inequality. In Section 8 we obtain the reduced dimensional LP for the functional upper bound for balanced systems, and also show that the asymptotic loss it provides is at least as good as the pairwise extension of the two station bound conjectured by HW to multi-station systems. We also show that the bound can capture interactions between multiple bottleneck stations. In Section 9 we obtain the reduced dimensional LP for the functional lower bound for balanced systems. In Section 10 we describe the conjectures of HW for two station systems. In Section 11 we obtain the functional lower bound for the HW-policy. In Section 12 we show that the conjectured asymptotic loss for the HW-policy is indeed a lower bound on the loss for all policies. Under a certain additional condition we show that the conjectured formula for the asymptotic loss applied to the exact opposite of the HW-policy is indeed an upper bound for all non-idling policies. Section 13 provides some concluding remarks.

## 2 System description and throughput

We consider a system with $S$ stations. There are $L$ buffers labelled $b_1, b_2, \ldots, b_L$. Each buffer is served by one of the stations, as shown in Figure 1. We shall denote by $\sigma(i)$, the particular station which serves $b_i$. We shall use the shorthand "$i \in \sigma$" to denote "$\sigma(i) = \sigma$", i.e., that buffer $b_i$ is served by station $\sigma$. Thus the notation "$i \in \sigma(j)$" means that $b_i$ and $b_j$ are served by the same station. There are $N$ trapped customers in the system. Customers at $b_i$ require an exponentially distributed service time with mean $\frac{1}{\mu_i}$ from station $\sigma(i)$. After completing service at $b_i$, customers move to buffer $b_j$ with probability $p_{ij}$. We will assume that the routing matrix $P = [p_{ij}]$ is irreducible. All service times and routing decisions are independent. We call the resulting system an *irreducible closed Markovian network*.

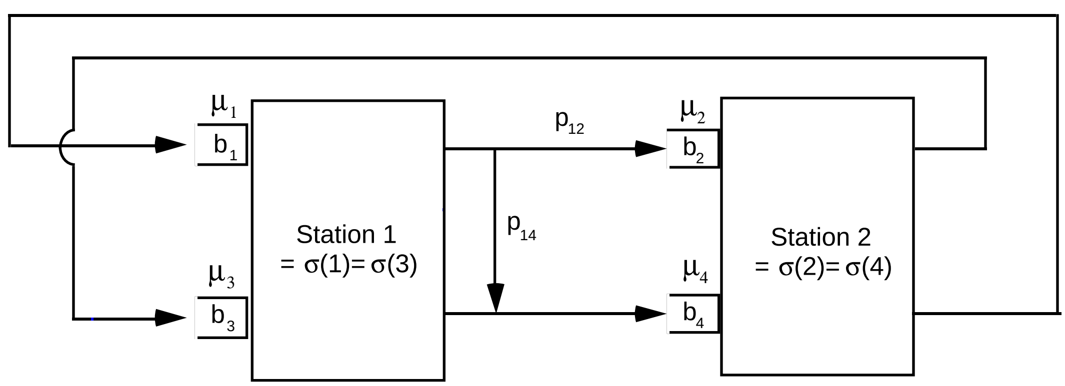


Figure 1: An irreducible closed Markovian queueing network.

If more than one customer is present in its buffers, then a station has to decide which buffer to serve. This decision is made by a *scheduling policy* $u$. Let

$$
\begin{aligned}
w_i(t) &= 1 \quad \text{if station } \sigma(i) \text{ is working on a customer in buffer } b_i \text{ at time } t, \\
&= 0 \quad \text{otherwise.}
\end{aligned}
$$

Then $w(t) := (w_1(t), \ldots, w_L(t))$ denotes the scheduling *decision*. We assume that all stochastic processes are right continuous with left limits. Let $x_i(t)$ denote the number of customers in buffer $b_i$ at time $t$, and set $x(t) := (x_1(t), \ldots, x_L(t))^T$. The stochastic process $\{x(t)\}$ is a finite-state controlled Markov chain. If $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)^T$ is the unit vector with a 1 in the $i$-th place, its transition probabilities are given by

$$\text{Prob } (x(t+h) = x - e_i + e_k \mid x(t) = x, w(t) = w) \quad = \quad \mu_i w_i p_{ik} h + o(h) \text{ for } i \neq k.$$

Throughout this paper we consider a particular long-term average reward criterion $\liminf_{T\to\infty} \frac{1}{T} E \int_0^T c(x(t), w(t))dt$. It is well known from the theory of finite state Markov Decision Process (see Blackwell (1962)) that there exists a *stationary* policy which is optimal. Thus, from now on we restrict attention to stationary scheduling policies, i.e., those where $w(t) = u(x(t))$ is purely a function of the current state $x(t)$, given by a *policy* $u$.

The particular long-term average reward criterion we would like to consider is the "throughput." For *closed re-entrant lines*, which are systems such that $p_{i,i+1} = 1$ for $i = 1, \ldots, L-1$ and $p_{L,1} = 1$ (see Figure 2), the notion of throughput is well defined as the rate at which customers circulate. For such systems, in steady state we have,

$$\begin{aligned} \text{Rate at which customers circulate} \quad &= \quad \mu_i E(w_i(t)) \quad \text{ for all } b_i, \\ &= \quad \mu_i \text{ Prob(Station } \sigma(i) \text{ is busy) } \frac{1/\mu_i}{\sum_{j\in\sigma(i)} 1/\mu_j}, \qquad (1) \\ &= \quad \text{Prob(Station } \sigma \text{ is busy) } (\sum_{j\in\sigma} 1/\mu_j)^{-1} \quad \text{ for all } \sigma. \end{aligned}$$

Thus the throughput is proportional to the probability that any buffer $b_i$ is being served, as well as the probability that any station $\sigma$ is busy. Hence, to within a multiplicative constant, one could take $c(x(t), w(t)) := w_i(t)$ or $c(x(t), w(t)) := \sum_{i\in\sigma} w_i(t)$.

In the general case where the routing is given by an irreducible matrix $P$, we will take as our objective function a "normalized" throughput, which has the convenient property that it is independent of the choice of $b_i$ or of $\sigma$. Let $\pi = (\pi_1, \ldots, \pi_L)$ satisfying

$$\pi = \pi P, \quad \pi_i \geq 0, \text{ and } \sum_{i=1}^{L} \pi_i = 1 \qquad (2)$$

be the unique invariant probability vector associated with $P$. Let

$$\beta_i := E[w_i(t)] \tag{3}$$

be the probability that station $\sigma(i)$ is busy working on a customer in buffer $b_i$ in steady state. Balancing the rate in steady-state at which customers leave and enter $b_j$, we have $\sum_{i \neq j} \beta_i \mu_i p_{ij} = \sum_{i \neq j} \beta_j \mu_j p_{ji}$. Since $\pi$ in (2) is unique, there is an $\alpha$ such that

$$\frac{\beta_i \mu_i}{\pi_i} = \alpha \text{ for all } i. \tag{4}$$

Abusing notation slightly, we shall call $\alpha$ as the *throughput*. It is proportional to the buffer utilizations.

Note that for closed re-entrant lines, $\pi_i \equiv 1/L$, and so $\alpha = L$(Rate at which customers circulate).

To recognize that the throughput $\alpha$ depends both on the stationary policy $u$ as well as the population $N$, we will henceforth denote it as $\alpha^u(N)$.

Let

$$\rho_\sigma := \sum_{i \in \sigma} \frac{\pi_i}{\mu_i} \tag{5}$$

denote the *relative load* on station $\sigma$. We shall say that a system is *balanced* if

$$\rho_\sigma = \rho_{\sigma'} \text{ for all } \sigma, \sigma'. \tag{6}$$

Note that since $\sum_{i \in \sigma} w_i(\tau_n) \leq 1$, from (3) one has $\sum_{i \in \sigma} \beta_i \leq 1$ for every station $\sigma$. Hence, from (4), $\alpha^u(N) \sum_{i \in \sigma} \frac{\pi_i}{\mu_i} \leq 1$ for all $\sigma$. Thus,

$$\alpha^u(N) \leq \alpha^* := \operatorname*{Min}_\sigma \left[ \sum_{i \in \sigma} \frac{\pi_i}{\mu_i} \right]^{-1} = \operatorname*{Min}_\sigma \rho_\sigma^{-1} \text{ for all policies } u \text{ and all } N. \tag{7}$$

Any station $\sigma$ that attains the minimum above, i.e., a $\sigma$ for which $\rho_\sigma = \alpha^*$, is a *bottleneck* station; it constrains the throughput $\alpha^u(N)$ to be no more than $\alpha^*$.

We say that a scheduling policy $u$ is *efficient* if $\lim_{N \to \infty} \alpha^u(N) = \alpha^*$. (More properly, we should say that a *sequence* of scheduling policies $\{u(N)\}$ is efficient if $\alpha^{u(N)}(N)$ has the limit shown. This is because the state spaces of the systems are different for different $N$, and so therefore are the policies $u(N)$. However, we will mainly deal below with buffer priority policies, for which this distinction is only pedantic.)

For the particular reward criterion of maximizing the throughput $\alpha^u(N)$, one can restrict attention to *non-idling* scheduling policies. These are policies where a station is not allowed to stay idle when there is a customer waiting for attention, i.e.,

$$\sum_{i \in \sigma} w_i(t) = 1 \text{ whenever } \sum_{i \in \sigma} x_i(t) \geq 1. \tag{8}$$

We shall pay great attention to the class of *buffer priority policies.* A buffer priority policy is given by a permutation $\theta = \{\theta(1), \theta(2), \ldots, \theta(L)\}$ of $\{1, \ldots, L\}$. Pre-emptive priority is given to $b_i$ over $b_j$ if $\theta(i) < \theta(j)$, i.e.,

$$\begin{aligned} w_j(t) = &\ 1 \quad \text{if } x_j(t) \geq 1, \text{ and } x_i(t) = 0 \text{ for all } i \in \sigma(j) \text{ with } \theta(i) < \theta(j), \\ = &\ 0 \quad \text{otherwise} . \end{aligned}$$

For a buffer priority policy $\theta$, one thus has the relation,

$$w_j(t)x_i(t) = 0 \text{ if } i \in \sigma(j) \text{ and } \theta(i) < \theta(j). \tag{9}$$

Instead of considering the continuous time controlled Markov chain, it is more convenient to consider a discrete-time controlled Markov chain obtained by sampling at certain random times $\{\tau_n\}$. Let us first rescale time so that $\sum_{i=1}^{L} \mu_i = 1$. Attach a clock to each buffer $b_i$, which rings after an exponentially distributed amount of time with mean $\frac{1}{\mu_i}$, and reset every clock whenever one rings. Due to the memoryless property of the exponential service times, if a station is busy working on buffer $b_i$, then the ring of $b_i$ can be regarded as corresponding to an actual service completion. Otherwise we can regard it as a *virtual* service completion which leaves the system state unchanged. We sample the system at the random sequence of times $\{\tau_n : n = 0, 1, 2, \ldots\}$, with $\tau_0 := 0$, at which the clocks ring. This random sampling is called "uniformization," (see Lippman (1975)). The resulting sampled process $\{x(\tau_n) : n = 0, 1, 2, \ldots\}$ has transition probabilities,

$$\text{Prob } (x(\tau_{n+1}) = x - e_i + e_k \mid x(\tau_n), w(\tau_n)) \quad = \quad \mu_i w_i(\tau_n) p_{ik} \text{ for } k \neq i. \tag{10}$$

It is an aperiodic Markov chain, since the system state can remain unchanged after a transition. Due to the uniformization, it has the property that its steady state distribution under any stationary policy is the same as that of the continuous time process $\{x(t) : t \geq 0\}$. In particular, $\beta_i = E[w_i(\tau_n)]$.

Under any buffer priority policy $\theta$, $x(\tau_n)$ has a single closed communicating class. This is because the state $x = (0, 0, \ldots, 0, N, 0, 0, \ldots, 0)$, where the non-zero component corresponds to a buffer with lowest priority at some station $\sigma$ under $\theta$, is reachable from every state.

# 3 Linear programs for bounds on the throughput for fixed N

Consider any non-idling policy. Suppose that the system is started in some steady-state of the Markov chain (there may be more than one closed communicating class in general), and define

$$z_{ij} \quad := \quad E[w_i(\tau_n) x_j(\tau_n)], \tag{11}$$

$$\bar{x}_j \quad := \quad E[x_j(\tau_n)]. \tag{12}$$

Since the policy is non-idling, from (8) we obtain $x_j(\tau_n) = \sum_{i \in \sigma(j)} w_i(\tau_n) x_j(\tau_n),$, and so

$$\bar{x}_j = \sum_{i \in \sigma(j)} z_{ij}. \tag{13}$$

Moreover, for every $\sigma \neq \sigma(j)$, $\sum_{i \in \sigma} w_i(\tau_n) \leq 1$, and so

$$\bar{x}_j \geq \sum_{i \in \sigma} z_{ij} \text{ for all } \sigma \neq \sigma(j). \tag{14}$$

Since the total population is $N$,

$$\sum_{j=1}^{L} \bar{x}_j = N. \tag{15}$$

Since $\sum_{j=1}^{L} x_j(\tau_n) = N$, from (3) we have $E[w_i(\tau_n) \sum_{j=1}^{L} x_j(\tau_n)] = N\beta_i$. From (11) and (4) we thus obtain,

$$\sum_{j=1}^{L} z_{ij} = \frac{N\alpha\pi_i}{\mu_i} \text{ for all } i. \tag{16}$$

From (10), defining $\delta_{ij} = 1$ if $i = j$, and $= 0$ otherwise, we have

$$E[x_i(\tau_{n+1})x_j(\tau_{n+1}) \mid x(\tau_n)] = x_i(\tau_n)x_j(\tau_n)$$
$$+ \textstyle\sum_{\ell,m} \mu_\ell w_\ell(\tau_n)p_{\ell m}[(x_i(\tau_n) - \delta_{\ell i} + \delta_{mi})(x_j(\tau_n) - \delta_{\ell j} + \delta_{mj}) - x_i(\tau_n)x_j(\tau_n)]. \tag{17}$$

Since $E[x_i(\tau_{n+1})x_j(\tau_{n+1})] = E[x_i(\tau_n)x_j(\tau_n)]$ in steady state, due to aperiodicity, we obtain

$$E\sum_{\ell,m} \mu_\ell w_\ell(\tau_n)p_{\ell m}[(x_i(\tau_n) - \delta_{\ell i} + \delta_{mi})(x_j(\tau_n) - \delta_{\ell j} + \delta_{mj}) - x_i(\tau_n)x_j(\tau_n)] = 0 \text{ for all } i \leq j.$$

Using (11, 3, 4) we obtain

$$-\mu_j z_{ji} + \sum_{\ell} \mu_\ell p_{\ell j} z_{\ell i} - \mu_i z_{ij} + \sum_{\ell} \mu_\ell p_{\ell i} z_{\ell j} - \alpha\pi_i p_{ij} - \alpha\pi_j p_{ji} + 2\delta_{ij}\alpha\pi_i = 0 \text{ for } i \leq j. \tag{18}$$

Thus we obtain the following theorem which provides LPs to bound the throughput for fixed $N$. It is the appropriate extension of Theorem 3 of Kumar and Kumar (1994) to closed systems with random routing. Throughout this paper, if $T$ is a LP, we shall denote its *value* by $VT$.

**Theorem 1: Pointwise LP bounds on throughput for fixed $N$.**

**(i)** *Let $\overline{T}(N)$ denote the LP:*

$$\text{Max } \alpha \tag{19}$$

*subject to the non-idling constraints (13, 14), the total population constraint (15), the sampling equalities (16), the equalities (18), and the nonnegativity constraints*

$$z_{ij} \geq 0, \bar{x}_j \geq 0, \alpha \geq 0 \text{ for all } i, j. \tag{20}$$

*Then*

$$\alpha^u(N) \leq V\overline{T}(N) \text{ for all policies } u, \text{ and all } N. \tag{21}$$

**(ii)** *Let $\underline{T}(N)$ denote the LP:*

$$\text{Min } \alpha \tag{22}$$

*subject to the same constraints as $\overline{T}(N)$. Then*

$$\alpha^u(N) \geq V\underline{T}(N) \text{ for all non-idling policies } u, \text{ and all } N. \tag{23}$$

**(iii)** *For a buffer priority policy $\theta$, define the LPs $\overline{T}^\theta(N)$ and $\underline{T}^\theta(N)$ as the LPs $\overline{T}(N)$ and $\underline{T}(N)$ respectively, with the following additional constraints (which follow from (9)),*

$$z_{ji} = 0 \text{ for all } i \in \sigma(j) \text{ with } \theta(i) < \theta(j). \tag{24}$$

*Then*

$$V\underline{T}^\theta(N) \leq \alpha^\theta(N) \leq V\overline{T}^\theta(N) \text{ for all } N. \tag{25}$$

We note that for each $N$ one needs to solve a *separate* LP; hence we shall refer to these as the *pointwise bound LPs*. In Section 7 we will obtain a *single* LP which gives a functional bound, parameterized by $N$.

# 4 The limiting infinite population LPs

Clearly, from (25), a buffer priority policy $\theta$ is efficient if $\lim_{N\to\infty} V\underline{T}^\theta(N) = \alpha^*$. The limit as $N \to +\infty$ of the values of the LPs $\underline{T}^\theta(N)$ is greater than the value of a limiting LP $\underline{T}^\theta(\infty)$, as shown in Theorem 4 of Kumar and Kumar (1994). Thus we obtain the following sufficient condition for efficiency.

**Theorem 2: The limiting infinite population LPs and efficiency.**

**(i)** *Let $\underline{T}^\theta(\infty)$ denote the LP with objective function (22) subject to (13, 14, 20, 24) and*

$$\textstyle\sum_{j=1}^L \bar{x}_j = 1, \tag{26}$$

$$\textstyle\sum_{j=1}^L z_{ij} = \frac{\alpha\pi_i}{\mu_i}, \tag{27}$$

$$-\mu_j z_{ji} + \textstyle\sum_\ell \mu_\ell p_{\ell j} z_{\ell i} - \mu_i z_{ij} + \sum_\ell \mu_\ell p_{\ell i} z_{\ell j} = 0 \text{ for all } i \leq j. \tag{28}$$

*Let $\overline{T}^{\theta}(\infty)$ denote the LP with objective function (19), and subject to the same constraints as $\underline{T}^{\theta}(\infty)$. Then*

$$\begin{aligned} V\underline{T}^{\theta}(\infty) &\le \liminf_{N\to\infty} V\underline{T}^{\theta}(N) \le \lim \inf_{N\to\infty} \alpha^{\theta}(N) \\ &\le \limsup_{N\to\infty} \alpha^{\theta}(N) \le \lim \sup_{N\to\infty} V\overline{T}^{\theta}(N) \le V\overline{T}^{\theta}(\infty) = \alpha^{*}. \end{aligned}$$

**(ii)** *If $V\underline{T}^{\theta}(\infty) = \alpha^{*}$, then the buffer priority scheduling policy $\theta$ is efficient.*

**(iii)** *Define the LPs $\underline{T}(\infty)$ and $\overline{T}(\infty)$ by deleting the buffer priority constraints (24) from the LPs $\underline{T}^{\theta}(\infty)$ and $\overline{T}^{\theta}(\infty)$, respectively. Then*

$$\begin{aligned} V\underline{T}(\infty) &\le \liminf_{N\to\infty} V\underline{T}(N) \le \liminf_{N\to\infty} \alpha^{u}(N) \le \limsup_{N\to\infty} \alpha^{u}(N) \\ &\le \limsup_{N\to\infty} V\overline{T}(N) \le V\overline{T}(\infty) = \alpha^{*} \text{ for all non-idling policies } u. \end{aligned}$$

**(iv)** *If $V\underline{T}(\infty) = \alpha^{*}$, then all non-idling scheduling policies are efficient.*

**Proof:** From any optimal solution $\{z_{ij}, \bar{x}_j, \alpha\}$ of the LP $\underline{T}^{\theta}(N)$, define $z_{ij}(N) := \frac{z_{ij}}{N}, \bar{x}_j(N) := \frac{\bar{x}_j}{N}$ and $\alpha(N) := \alpha$. Clearly $\{z_{ij}(N), \bar{x}_j(N), \alpha(N)\}$ are bounded. Taking limits along a subsequence, we get a feasible solution for $\underline{T}^{\theta}(\infty)$. Thus $V\underline{T}^{\theta}(\infty) \le \liminf_{N\to\infty} V\underline{T}^{\theta}(N)$. The remaining inequalities in (i,iii) follow similarly or from Theorem 1. The results (ii,iv) are obvious.

To show the last equality $V\overline{T}^{\theta}(\infty) = \alpha^{*}$ in (i), simply set $\bar{x}_j = 0$ if either $\sigma(j)$ is not a bottleneck station, or if $b_j$ is *not* the lowest priority buffer at $\sigma(j)$ under $\theta$. Otherwise let $\bar{x}_j$ be any arbitrary nonnegative choice subject only to the restriction (26). Let $z_{ij} := \frac{\alpha^{*}\pi_i}{\mu_i}\bar{x}_j$. It is easy to verify using (2, 7) that this is a feasible solution of $\overline{T}^{\theta}(\infty)$ with value $\alpha^{*}$. Hence $V\overline{T}^{\theta}(\infty) \ge \alpha^{*}$. Clearly $V\overline{T}(\infty) \ge V\overline{T}^{\theta}(\infty) \ge \alpha^{*}$. Thus to prove the last equalities in (i,iii) it suffices to show that $V\overline{T}(\infty) \le \alpha^{*}$.

Summing (27), we have $\sum_{i\in\sigma}\sum_{j=1}^{L} z_{ij} = \alpha\sum_{i\in\sigma}\frac{\pi_i}{\mu_i}$. However from (13, 14, 26), $\sum_{j=1}^{L}\sum_{i\in\sigma} z_{ij} \le \sum_{j=1}^{L}\bar{x}_j = 1$. Hence $\alpha \le \left[\sum_{i\in\sigma}\frac{\pi_i}{\mu_i}\right]^{-1}$ for every $\sigma$. Therefore, by (7), we have $\alpha \leq \alpha^*$. $\square$

## 5 The duals of the infinite population LPs

Throughout, we shall denote the dual of a LP $T$ by $DT$. Any feasible solution of the dual of a primal LP provides a bound on the value (see Murty (1983)). Thus, to utilize Theorem 2.ii, it suffices to show that $D\underline{T}^\theta(\infty)$ has a feasible solution with value $\alpha^*$. This motivates the analysis of the duals of the LPs $\underline{T}^\theta(\infty)$, $\overline{T}^\theta(\infty)$, $\underline{T}(\infty)$ and $\overline{T}(\infty)$. Define $y^+ := \text{Max}(y, 0)$.

**Theorem 3: The duals of the infinite population LPs.**

**(i)** *The dual* $D\underline{T}^\theta(\infty)$ *of* $\underline{T}^\theta(\infty)$ *is the LP:*

$$\text{Max } \alpha$$

*subject to:*

$$\underset{\{i\in\sigma(j) \text{ and } \theta(i)\leq\theta(j)\}}{\text{Max}} \mu_i[\sum_k p_{ik}q_{kj} - q_{ij} - \frac{s_i}{\pi_i}] + \sum_{\sigma\neq\sigma(j)} \underset{i\in\sigma}{\text{Max}} \mu_i[\sum_k p_{ik}q_{kj} - q_{ij} - \frac{s_i}{\pi_i}]^+ \leq -\alpha \text{ for all } j, \tag{29}$$

$$\sum_{i=1}^{L} s_i \leq 1, \tag{30}$$

$$q_{ij} = q_{ji} \text{ for all } i, j. \tag{31}$$

**(ii)** *The dual* $D\overline{T}^\theta(\infty)$ *of* $\overline{T}^\theta(\infty)$ *is the LP:*

$$\text{Min } \alpha$$

*subject to:*

$$\text{Max}_{\{i\in\sigma(j) \text{ and } \theta(i)\leq\theta(j)\}} \mu_i[\textstyle\sum_k p_{ik}q_{kj} - q_{ij} - \frac{s_i}{\pi_i}] + \sum_{\sigma\neq\sigma(j)} \text{Max}_{i\in\sigma} \mu_i[\sum_k p_{ik}q_{kj} - q_{ij} - \frac{s_i}{\pi_i}]^+ \leq \alpha \text{ for all } j, \tag{32}$$

$$\sum_{i=1}^{L} s_i \leq -1, \tag{33}$$

$$q_{ij} = q_{ji} \quad \textit{for all} \quad i, j. \tag{34}$$

**(iii)** *The duals $D\underline{T}(\infty)$ and $D\overline{T}(\infty)$, of $\underline{T}(\infty)$ and $\overline{T}(\infty)$ respectively, are the same as $D\underline{T}^{\theta}(\infty)$ and $D\overline{T}^{\theta}(\infty)$ respectively, except that the condition $\theta(i) \leq \theta(j)$ in the first "max" in (29, 32) is eliminated.*

**(iv)** *If $\{q_{ij}, s_i, \alpha\}$ is a feasible solution of $D\underline{T}^{\theta}(\infty)$ (or $D\underline{T}(\infty)$), then*

$$\liminf_{N \to \infty} \alpha^u(N) \geq \alpha \textit{ for } u = \theta \ \ \textit{(or for all non-idling policies } u\textit{)}.$$

**(v)** $VD\overline{T}^{\theta}(\infty) = VD\overline{T}(\infty) = \alpha^*$.

**Proof:** Consider $\underline{T}^{\theta}(\infty)$. Associate dual variables as follows: $\eta$ with (26), $\frac{\mu_i s_i}{\pi_i}$ with (27), $r_{ij}$ with (28), $p_j$ with (13), $(-w_{\sigma j})$ with (14), and $\gamma_{ij}$ with (24). Then the dual is the following LP, with the associated primal variables shown in parentheses:

$$\text{Max}(-\eta) \tag{35}$$

subject to

$$\eta + p_j - \sum_{\sigma \neq \sigma(j)} w_{\sigma j} \geq 0, \quad (\bar{x}_j) \tag{36}$$

$$-\sum_i s_i \geq -1, \quad (\alpha) \tag{37}$$

$$\begin{aligned} &\textstyle\sum_i \sum_{j=1}^{L} \frac{s_i \mu_i}{\pi_i} 1(i=n, j=m) - \sum_i \sum_{j \geq i} \mu_j r_{ij} 1(j=n, i=m) \\ &+ \textstyle\sum_i \sum_{j \geq i} \sum_\ell \mu_\ell p_{\ell j} r_{ij} 1(\ell = n, i = m) - \sum_i \sum_{j \geq i} \mu_i r_{ij} 1(i=n, j=m) \\ &+ \textstyle\sum_i \sum_{j \geq i} \sum_\ell \mu_\ell p_{\ell i} r_{ij} 1(\ell = n, j = m) - \sum_j p_j \sum_{i \in \sigma(j)} 1(i=n, j=m) \\ &+ \textstyle\sum_j \sum_{\sigma \neq \sigma(j)} \sum_{i \in \sigma} w_{\sigma j}(i=n, j=m) + \gamma_{nm} 1(n \in \sigma(m) \text{ and } \theta(n) > \theta(m)) \geq 0, \quad (z_{nm}) \end{aligned}$$

$$w_{\sigma j} \geq 0 \text{ for } j \notin \sigma. \tag{38}$$

The second to last constraint above simplifies to,

$$\frac{s_n\mu_n}{\pi_n} - \mu_n r_{mn} 1(n \geq m) + \sum_{j\geq m} \mu_n p_{nj} r_{mj}$$
$$-\mu_n r_{nm} 1(m \geq n) + \sum_{i\leq m} \mu_n p_{ni} r_{im} - p_m 1(\sigma(n) = \sigma(m))$$
$$+w_{\sigma(n)m} 1(\sigma(n) \neq \sigma(m)) + \gamma_{nm} 1(n \in \sigma(m) \text{ and } \theta(n) > \theta(m)) \geq 0.$$

Symmetrically extend $r_{ij}$ (used above only for $i \leq j$) by defining $r_{ij} := r_{ji}$ for $i > j$. Then this constraint can be rewritten as,

$$\frac{s_n\mu_n}{\pi_n} - \mu_n r_{nm} - \mu_n r_{nn} 1(n = m) + \sum_j \mu_n p_{nj} r_{mj}$$
$$+\mu_n p_{nm} r_{mm} - p_m 1(\sigma(n) = \sigma(m)) + w_{\sigma(n)m} 1(\sigma(n) \neq \sigma(m))$$
$$+\gamma_{nm} 1(n \in \sigma(m) \text{ and } \theta(n) > \theta(m)) \geq 0.$$

Define

$$\begin{aligned} q_{nm} &:= -r_{nm} \text{ for } n \neq m \\ &:= -\frac{r_{nn}}{2} \text{ for } n = m. \end{aligned}$$

Then the constraint can be rewritten as

$$\frac{s_n\mu_n}{\pi_n} + \mu_n q_{nm} - \sum_j \mu_n p_{nj} q_{mj} - p_m 1(\sigma(n) = \sigma(m)) + w_{\sigma(n)m} 1(\sigma(n) \neq \sigma(m))$$
$$+\gamma_{nm} 1(n \in \sigma(m)) \text{ and } \theta(n) > \theta(m)) \geq 0.$$

This can be decoupled into two separate constraints, one for the case $\sigma(n) = \sigma(m)$, and another for $\sigma(n) \neq \sigma(m)$. Further, since $\gamma_{nm}$ is unconstrained in sign, the inequality is clearly satisfied for $\sigma(m) = \sigma(n)$ and $\theta(n) > \theta(m)$. This gives us,

$$-p_m \geq \operatorname*{Max}_{\{n:n\in\sigma(m) \text{ and } \theta(n)\leq\theta(m)\}} \mu_n \left( \sum_j p_{nj} q_{mj} - q_{nm} - \frac{s_n}{\pi_n} \right), \tag{39}$$

$$w_{\sigma m} \geq \operatorname*{Max}_{n\in\sigma} \mu_n \left( \sum_j p_{nj} q_{mj} - q_{nm} - \frac{s_n}{\pi_n} \right). \tag{40}$$

Renaming $\eta$ as $(-\alpha)$, the constraint (29) follows from (36, 39, 40, 38), and the result (i) is thus established.

The results (ii,iii) follow in a similar way. The result (iv) follows from Theorem 2.i,iii since $VD\underline{T}^{\theta}(\infty) = V\underline{T}^{\theta}(\infty)$ and $VD\underline{T}(\infty) = V\underline{T}(\infty)$. The result (v) follows from the last equalities in Theorem 2.i,iii. $\square$

A mnemonic for the constraints (29, 30) is the following economic interpretation. For simplicity, suppose the system is a re-entrant line. As noted earlier, $\pi_i \equiv \frac{1}{L}$, and $\alpha^* = L$ (Rate at which customers circulate). For each buffer $b_i$, assign a *toll* of \$$s_i$ that a customer has to pay when leaving $b_i$. The total toll paid in a round trip is $\sum s_i = 1$. In addition, when buffer $b_j$ is non-empty, charge a customer \$$q_{ij}$ for leaving buffer $b_i$, and pay the customer \$$q_{kj}$ for reaching $b_k$. Then, the left hand side is the rate of cash flow whenever buffer $b_j$ is nonempty. The inequality (29) then bounds the rate of cash outflow when buffer $b_j$ is non-empty.

# 6 The fundamental identity and inequality

We now obtain a fundamental identity and a fundamental inequality which will allow us in Section 7 to obtain a bound on the throughput for *every* $N$, just by solving a *single* LP. Fix a policy $u$, and a population size $N$. Consider any symmetric matrix $Q$. Then, from (10)

$$\begin{aligned} E\left[x^T(\tau_{n+1})Qx(\tau_{n+1}) \mid x(\tau_n)\right] &= \sum_{i,k} \mu_i w_i(\tau_n) p_{ik} (x(\tau_n) - e_i + e_k)^T Q (x(\tau_n) - e_i + e_k) \\ &\quad + \sum_{i,k} \mu_i (1 - w_i(\tau_n)) p_{ik} x^T(\tau_n) Q x(\tau_n). \end{aligned}$$

Since $E[x^T(\tau_{n+1})Qx(\tau_{n+1})] = E[x^T(\tau_n)Qx(\tau_n)]$ in steady state, due to aperiodicity,

$$E[2\sum_{i,k} \mu_i w_i(\tau_n) p_{ik} (e_k - e_i)^T Q x(\tau_n) + \sum_{i,k} \mu_i w_i(\tau_n) p_{ik} (e_k - e_i)^T Q (e_k - e_i)] = 0.$$

Using $E[w_i(\tau_n)] = \beta_i$, $\beta_i \mu_i = \pi_i \alpha^u(N)$, and $\pi = \pi P$, we obtain

$$E[2\sum_{j,i} \mu_i w_i(\tau_n) \sum_k p_{ik}(q_{kj} - q_{ij}) x_j(\tau_n)] + 2\alpha^u(N) \sum_i \pi_i (q_{ii} - \sum_k p_{ik} q_{ik}) = 0.$$

For any constants $\{s_i\}$, by adding and subtracting $E\sum_{j,i}\mu_i w_i(\tau_n)s_i x_j(\tau_n)/\pi_i = \alpha^u(N)N\sum_i s_i$ (recall $\sum_j x_j(\tau_n)=N$), we obtain the following fundamental identity.

**Lemma 1: The fundamental identity.** *For every policy $u$, and population $N$,*

$$E[\sum_j[\sum_i \mu_i w_i(\tau_n)\sum_k p_{ik}(q_{kj}-q_{ij}-\frac{s_i}{\pi_i})]x_j(\tau_n)]$$
$$+\alpha^u(N)N\sum_i s_i+\alpha^u(N)\sum_{i,j}\pi_i p_{ij}(q_{ii}-q_{ij})=0. \tag{41}$$

Writing $\sum_i=\sum_{i\in\sigma(j)}+\sum_{\sigma\neq\sigma(j)}\sum_{i\in\sigma}$, we obtain from the fundamental identity that

$$E\left[\sum_j[\sum_{i\in\sigma(j)}\mu_i w_i(\tau_n)\sum_k p_{ik}(q_{kj}-q_{ij}-\tfrac{s_i}{\pi_i})+\sum_{\sigma\neq\sigma(j)}\sum_{i\in\sigma}\mu_i w_i(\tau_n)\sum_k p_{ik}(q_{kj}-q_{ij}-\tfrac{s_i}{\pi_i})]x_j(\tau_n)\right]$$
$$+\alpha^u(N)N\sum_i s_i+\alpha^u(N)\sum_{i,j}\pi_i p_{ij}(q_{ii}-q_{ij})=0.$$

From the non-idling property (8) we see that for any constants $\{\gamma_i\}$,

$$\sum_{i\in\sigma(j)}\gamma_i w_i(\tau_n)x_j(\tau_n)\leq \operatorname*{Max}_{i\in\sigma(j)}\gamma_i x_j(\tau_n), \tag{42}$$

$$\sum_{i\in\sigma}\gamma_i w_i(\tau_n)x_j(\tau_n)\leq \operatorname*{Max}_{i\in\sigma}\gamma_i^+ x_j(\tau_n) \text{ for all } \sigma\neq\sigma(j).$$

Note that for a buffer priority policy $\theta$, one can further restrict the maximum in (42) to those $i$'s with $\theta(i)\leq\theta(j)$. Hence from (41) we have obtained the following fundamental inequality.

**Lemma 2: The fundamental inequality.**

**(i)** *For any non-idling policy $u$, and population $N$,*

$$E\sum_j\left[\text{Max}_{i\in\sigma(j)}\mu_i[\sum_k p_{ik}(q_{kj}-q_{ij}-\tfrac{s_i}{\pi_i})]+\sum_{\sigma\neq\sigma(j)}\text{Max}_{i\in\sigma}\mu_i[\sum_k p_{ik}(q_{kj}-q_{ij}-\tfrac{s_i}{\pi_i})]^+\right]x_j(\tau_n)$$
$$+\alpha^u(N)N\sum_i s_i-\alpha^u(N)\sum_{i,j}\pi_i p_{ij}(q_{ji}-q_{ii})\geq 0.$$

**(ii)** *For a buffer priority policy $\theta$, and population $N$,*

$$E\sum_j\left[\text{Max}_{\{i\in\sigma(j)\ and\ \theta(i)\leq\theta(j)\}}\mu_i\sum_k p_{ik}(q_{kj}-q_{ij}-\tfrac{s_i}{\pi_i})\right.$$
$$\left.+\sum_{\sigma\neq\sigma(j)}\text{Max}_{i\in\sigma}[\mu_i\sum_k p_{ik}(q_{kj}-q_{ij}-\tfrac{s_i}{\pi_i})]^+\right]x_j(\tau_n)$$
$$+\alpha^\theta(N)N\sum_i s_i-\alpha^\theta(N)\sum_{i,j}\pi_i p_{ij}(q_{ji}-q_{ii})\geq 0. \tag{43}$$

# 7 Linear programs for functional bounds and asymptotic loss

Define the lower and upper *asymptotic losses* $\underline{J}(u)$ and $\overline{J}(u)$ as in Section 1.

We now obtain two LPs which provide upper and lower *functional bounds*, parameterized by $N$. As a consequence, they also provide bounds on the asymptotic loss.

Consider a buffer priority policy $\theta$, and a population $N$. Let $\{q_{ij}, s_i, \alpha\}$ be any feasible solution to $D\overline{T}^{\theta}(\infty)$. Substituting (32, 33) in the fundamental inequality (43), and using $\sum_j x_j(\tau_n) = N$, we obtain

$$\alpha N - \alpha^{\theta}(N)N - \alpha^{\theta}(N)\sum_{i,j} \pi_i p_{ij}(q_{ji} - q_{ii}) \geq 0. \tag{44}$$

Hence,

$$\alpha^{\theta}(N) \leq \alpha N/[N + \sum_{i,j} \pi_i p_{ij}(q_{ji} - q_{ii})], \tag{45}$$

for all $N$ such that the denominator is positive. Therefore, to obtain the best bound, one should maximize $\sum_{i,j} \pi_i p_{ij}(q_{ji} - q_{ii})$, subject to the same constraints as $D\overline{T}^{\theta}(\infty)$. This gives us the following *functional bound* LPs.

**Theorem 4: The functional bound LPs.**

**(i)** *Let $\overline{FB}^{\theta}(\infty)$ denote the LP:*

$$\text{Max} \sum_{i,j} \pi_i p_{ij}(q_{ji} - q_{ii}), \tag{46}$$

*subject to the constraints (33, 34) and*

$$\underset{\{i\in\sigma(j)\ and\ \theta(i)\leq\theta(j)\}}{\text{Max}} \mu_i[\sum_k p_{ik}(q_{kj} - q_{ij} - \frac{s_i}{\pi_i})] + \sum_{\sigma\neq\sigma(j)} \underset{i\in\sigma}{\text{Max}} \mu_i[\sum_k p_{ik}(q_{kj} - q_{ij} - \frac{s_i}{\pi_i})]^+ \leq \alpha^* \text{ for all } j. \tag{47}$$

*(These are the same constraints as in $D\overline{T}^{\theta}(\infty)$, except that (32) is modified by replacing $\alpha$ by $\alpha^*$ to give (47)). If $v$ is the value of the objective function corresponding to any*

*feasible solution of* $\overline{FB}^{\theta}(\infty)$*, then*

$$\alpha^{\theta}(N) \;\leq\; \frac{N}{N+v}\alpha^{*} \text{ for all } N \text{ such that } N+v \geq 0, \text{ and} \tag{48}$$

$$\alpha^{\theta}(N) \;\leq\; \frac{N}{N+V\overline{FB}^{\theta}(\infty)}\alpha^{*} \text{ for all } N. \tag{49}$$

*Hence* $\underline{J}(\theta) \geq v$ *for (48), and* $\underline{J}(\theta) \geq V\overline{FB}^{\theta}(\infty)$ *for (49).*

**(ii)** *Let* $\overline{FB}(\infty)$ *denote the same LP as* $\overline{FB}^{\theta}(\infty)$*, except that the restriction* $\theta(i) \leq \theta(j)$ *in the first "Max" in (47) is eliminated. If* $v$ *is the value of any feasible solution of* $\overline{FB}(\infty)$*, then*

$$\alpha^{u}(N) \;\leq\; \frac{N}{N+v}\alpha^{*} \text{ for all policies } u, \text{ and all } N \text{ such that } N+v \geq 0, \text{ and} \tag{50}$$

$$\alpha^{u}(N) \;\leq\; \frac{N}{N+V\overline{FB}(\infty)}\alpha^{*} \text{ for all policies } u, \text{ and all } N. \tag{51}$$

*Hence* $\underline{J}(u) \geq v$ *for (50), and* $\underline{J}(u) \geq V\overline{FB}(\infty)$ *for (51).*

**(iii)** *Let* $\underline{FB}^{\theta}(\infty)$ *denote the LP with objective function (46) and subject to the constraints (30, 31) and*

$$\max_{\{i\in\sigma(j) \text{ and } \theta(i)\leq\theta(j)\}} \mu_i[\sum_k p_{ik}(q_{kj} - q_{ij} - \frac{s_i}{\pi_i})] + \sum_{\sigma\neq\sigma(j)} \max_{i\in\sigma} \mu_i[\sum_k p_{ik}(q_{kj} - q_{ij} - \frac{s_i}{\pi_i})]^{+}$$
$$\leq -VD\underline{T}^{\theta}(\infty) \text{ for all } j, \tag{52}$$

*(These are the same constraints as* $D\underline{T}^{\theta}(\infty)$*, except that* $\alpha$ *in (29) is replaced by* $VD\underline{T}^{\theta}(\infty)$*). If* $v$ *is the value of any feasible solution of* $\underline{FB}^{\theta}(\infty)$*, then*

$$\alpha^{\theta}(N) \geq \frac{N}{N-v}VD\underline{T}^{\theta}(\infty) \text{ for all } N, \text{ and} \tag{53}$$

$$\alpha^{\theta}(N) \geq \frac{N}{N-V\underline{FB}^{\theta}(\infty)}VD\underline{T}^{\theta}(\infty) \text{ for all } N. \tag{54}$$

*Hence* $\overline{J}(\theta) \leq -v$ *for (53), and* $\overline{J}(\theta) \leq -V\underline{FB}^{\theta}(\infty)$ *for (54), when* $VD\underline{T}^{\theta}(\infty) = \alpha^{*}$*.*

**(iv)** *Let $\underline{FB}(\infty)$ denote the same LP as $\underline{FB}^{\theta}(\infty)$, except that the restriction $\theta(i) \leq \theta(j)$ in the first "Max" in (52) is eliminated. If $v$ is the value of any feasible solution of $\underline{FB}(\infty)$, then*

$$\alpha^u(N) \geq \frac{N}{N-v} VD\underline{T}(\infty) \text{ for all non-idling policies } u, \text{ and all } N, \text{ and} \tag{55}$$

$$\alpha^u(N) \geq \frac{N}{N - V\underline{FB}(\infty)} VD\underline{T}(\infty) \text{ for all non-idling policies } u, \text{ and all } N. \tag{56}$$

*Hence $\overline{J}(u) \leq -v$ for (55) and $\overline{J}(u) \leq -V\underline{FB}(\infty)$ for (56), when $VD\underline{T}(\infty) = \alpha^*$.*

**Proof:** Recall from Theorem 3.iv that $VD\overline{T}^{\theta}(\infty) = \alpha^*$. Hence $\overline{FB}^{\theta}(\infty)$ is feasible. Hence (48) follows from (45). To show (49) we only need to show that $N + V\overline{FB}^{\theta}(\infty) \geq 0$ for all $N$, for which it is sufficient to show that $V\overline{FB}^{\theta}(\infty) \geq 0$. Let $\sigma^*$ be a particular bottleneck station which attains the minimum in (7), i.e., $\sum_{i:\in\sigma^*} \frac{\pi_i}{\mu_i} = \frac{1}{\alpha^*}$. Now set $s_i = -\alpha^* \frac{\pi_i}{\mu_i}$ for all $i \in \sigma^*$; and $= 0$ otherwise. Also, define $q_{ij} \equiv 0$ for all $i, j$. Then $\{q_{ij}, s_i\}$ satisfies the constraints of $\overline{FB}^{\theta}(\infty)$, and has objective function value of 0. Hence $V\overline{FB}^{\theta}(\infty) \geq 0$. This proves (i). The results in (ii) are similarly proved.

To prove (iii), consider any feasible solution $\{q_{ij}, s_i\}$ of $\underline{FB}^{\theta}(\infty)$ with objective function value $v$. Analogously to (44), we obtain

$$-VD\underline{T}^{\theta}(\infty)N + \alpha^{\theta}(N)(N - v) \geq 0.$$

Since $\alpha^{\theta}(N) \geq 0$, it follows that $N - v \geq 0$ for all $N$. This proves (iii). The result (iv) is similarly established. $\square$

How well do these functionals bounds, which are obtained just by solving a single LP, compare with the pointwise bounds obtained in Theorem 1 by solving a separate LP for each value of the population size $N$? The following example shows that in some cases the single functional bound LP actually solves all the pointwise bound LPs for all values of $N$. Thus it gives a closed form expression for the solutions of all the pointwise bound LPs.

**Example 1**

Consider the system shown in Figure 2.

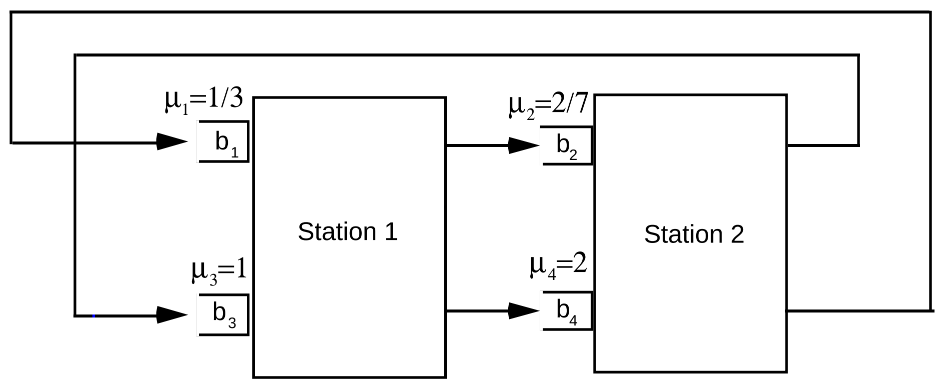


Figure 2: Closed re-entrant line of Example 1.

The functional lower bound obtained for the class of all non-idling policies is

$$\alpha^u(N) \geq \frac{N}{N+1}.$$

This coincides *exactly* with the pointwise bounds obtained for each $N$ by the LPs in Theorem 1. (More precisely, the coincidence is exact for every $N$ for which we have computed the pointwise bound in rational arithmetic, and appears to agree for all other values of $N$ computed by the plotting routine). Thus, in this case, we appear to have explicitly solved for the lower bounds of all the pointwise bound LPs..

For the functional upper bound for the class of all non-idling policies, we obtain

$$\alpha^u(N) \leq \frac{N}{N+\frac{45}{56}}.$$

This again coincides *exactly* with the pointwise bounds obtained for each $N$ in Theorem 1 for $N \geq 10$. However, for $N = 1$, there is a difference. The pointwise bound obtained from Theorem 1 is $\alpha^u(1) \leq \frac{1}{2}$. □

# 8 Reduced dimensional LP for functional upper bound on throughput of balanced systems

In many applications $S$, the number of stations, is far smaller than $L$, the number of buffers in the system. The functional bound LPs of the previous section consist of $\frac{L(L+3)}{2}$ variables. In this section we obtain a reduced dimensional LP consisting of just $\frac{S(S+1)}{2}$ variables for determining a functional upper bound on the throughput of closed balanced multi-station systems. Moreover, we will show that in some cases (Example 2) the reduced dimensional LP actually yields the same result as the original functional bound LP of Section 7. In Section 9 we obtain a similar reduced dimensional LP for obtaining a functional lower bound.

These reduced dimensional LPs are obtained by our discovery of a family of feasible solutions for the original functional bound LPs of Section 7. They involve the *mean remaining work* $W_{\sigma i}$ to be done on a customer in buffer $b_i$ by station $\sigma$, prior to that customer's next departure from buffer $b_L$. (The choice of the reference buffer $b_L$ is arbitrary). The $W_{\sigma i}$'s are the unique solution of the following equations:

$$W_{\sigma L} := \frac{1}{\mu_L} 1(L \in \sigma) \text{ for all } \sigma, \tag{57}$$

$$W_{\sigma i} = \frac{1}{\mu_i} 1(i \in \sigma) + \sum_{j=1}^{L} p_{ij} W_{\sigma j} \text{ for all } i \neq L, \text{ and all } \sigma. \tag{58}$$

The quantities $W_{\sigma i}$ play a central role in Harrison and Wein (1990).

As in the case of open systems treated in Humes, Ou and Kumar (1995), the key idea is to restrict consideration to functions which are quadratic in the mean remaining work $w_\sigma := \sum_i W_{\sigma i} x_i$ for a station $\sigma$, rather than the more general class of functions which are quadratic in the state $x$, i.e., we restrict attention to quadratics of the form $-\alpha^* \sum_{\sigma\sigma'} a_{\sigma\sigma'} w_\sigma w_{\sigma'}$, rather than allowing quadratics of the more general form $\sum_{ij} q_{ij} x_i x_j$. Thus we look for a feasible $Q$ of the form $Q = -\alpha^* W^T A W$, where $W = [W_{\sigma i}]$ and $A = [a_{\sigma\sigma'}]$. This will then result in the *reduced dimensional LP* shown in Theorem 5, with $\frac{S(S+1)}{2}$ variables consisting of the $a_{\sigma\sigma'}$'s, rather than the $\frac{L(L+3)}{2}$ variables consisting of the $q_{ij}$'s and $s_i$'s in the original functional bound LP.

We will also evaluate the quality of this lower bound. We will show in Theorem 6 that it is at least as good as an extension of the lower bound on asymptotic loss conjectured by HW for two station systems, and proved to be true in Theorem 9 of Section 12. This extension of the HW bound for two station systems is obtained as follows. Fix any two separate stations $\sigma \neq \sigma'$. Then set the service times of buffers at all the other stations, except $\sigma$ and $\sigma'$, to zero. Thus the rest of the system is transparent to customers, and it is easy to see that any upper bound on throughput for the resulting two station system consisting of just stations $\sigma$ and $\sigma'$ is also an upper bound on the throughput of the original system. By taking the minimum of the conjectured asymptotic losses obtained by varying $\sigma$ and $\sigma'$, one obtains a conjectured bound for the original multi-station system. The resulting conjectured lower bound on the asymptotic loss, which we shall refer to as the *pairwise bound*, is

$$\alpha^* \operatorname*{Max}_{\sigma \neq \sigma'} \frac{\sum_{ij} \pi_i p_{ij} [(W_{\sigma i} - W_{\sigma' i})^2 - (W_{\sigma i} - W_{\sigma' i})(W_{\sigma j} - W_{\sigma' j})]}{\text{Max}_j (W_{\sigma' j} - W_{\sigma j}) - \text{Min}_j (W_{\sigma' j} - W_{\sigma j})}. \tag{59}$$

The following Theorem exhibits linear constraints on $A = [a_{\sigma\sigma'}]$ which guarantee that the resulting $Q = -\alpha^* W^T A W$ is a feasible solution for balanced systems, and also rewrites the objective function purely in terms of the $a_{\sigma\sigma'}$'s. Thus it obtains the reduced dimensional LP for the functional upper bound on the throughput for balanced systems. The result (ii) below will be used in Theorem 9 to relate the bound to the conjectured asymptotic loss formula of HW.

**Theorem 5: Reduced Dimensional LP for functional upper bound on throughput of balanced systems.** *Consider a balanced system.*

**(i)** *Let $\underline{\nu}$ denote the value of the LP*

$$\text{Max}\, \alpha^* \sum_i \frac{\pi_i}{\mu_i} \sum_\sigma a_{\sigma(i)\sigma} W_{\sigma i} \tag{60}$$

*subject to:*

$$\sum_\sigma a_{\sigma'\sigma} = 0 \text{ for all } \sigma', \tag{61}$$

$$a_{\sigma'\sigma} \;=\; a_{\sigma'\sigma} \text{ for all } \sigma, \sigma', \tag{62}$$

$$\sum_\sigma Min_j \sum_{\sigma'} a_{\sigma\sigma'} W_{\sigma' j} = -1. \tag{63}$$

*Then for every scheduling policy $u$,*

$$\alpha^u(N) \le \alpha^* \frac{N}{N + \underline{\nu}} \text{ for all } N.$$

**(ii)** *Also, under the conditions (61, 62), the objective function can be alternatively written as,*

$$\alpha^* \sum_i \frac{\pi_i}{\mu_i} \sum_\sigma a_{\sigma(i)\sigma} W_{\sigma i} \;=\; -\frac{\alpha^*}{2} \sum_\sigma \sum_{\sigma' \neq \sigma} a_{\sigma\sigma'} \sum_{ij} \pi_i p_{ij} [(W_{\sigma i} - W_{\sigma' i})^2 - (W_{\sigma i} - W_{\sigma' i})(W_{\sigma j} - W_{\sigma' j})].$$

**Proof:** Suppose $A := [a_{\sigma\sigma'}]$ satisfies (61, 62, 63). We will first prove (i) by showing that

$$\begin{aligned} Q &:= -\alpha^* W^T A W, \text{ and} \\ s_i &:= \frac{\alpha^* \pi_i}{\mu_i} \text{Min}_j \sum_{\sigma'} a_{\sigma(i)\sigma'} W_{\sigma' j}, \end{aligned}$$

is a feasible solution for $\overline{FB}(\infty)$, with the value of the objective function (46) equal to (60).

First we show that

$$\sum_j p_{Lj} W_{\sigma j} = \frac{\rho_\sigma}{\pi_L} = \frac{1}{\alpha^* \pi_L}, \tag{64}$$

as follows,

$$\begin{aligned} \pi_L \sum_k p_{Lk} W_{\sigma k} &= \sum_i \pi_i \sum_k p_{ik} W_{\sigma k} - \sum_{i \neq L} \pi_i \sum_k p_{ik} W_{\sigma k} \\ &= \sum_k \pi_k W_{\sigma k} - \sum_{i \neq L} \pi_i \sum_k p_{ik} W_{\sigma k} \text{ (since } \pi_k = \textstyle\sum_i \pi_i p_{ik}) \\ &= \sum_k \pi_k W_{\sigma k} - \sum_{i \neq L} \pi_i \left( -\frac{1}{\mu_i} 1(i \in \sigma) + W_{\sigma i} \right) \text{ (from (58))} \\ &= \pi_L W_{\sigma L} + \sum_i \frac{\pi_i}{\mu_i} 1(i \in \sigma) - \frac{\pi_L}{\mu_L} 1(L \in \sigma) \\ &= \sum_i \frac{\pi_i}{\mu_i} 1(i \in \sigma) \text{ (from (57))} \\ &= \rho_\sigma \text{ (from (5))} \\ &= \frac{1}{\alpha^*} \text{ (from (7), since the system is balanced).} \end{aligned}$$

Hence, from (58,64),

$$\sum_j p_{ij} W_{\sigma' j} = -\frac{1}{\mu_i} 1(i \in \sigma', i \neq L) + W_{\sigma' i} 1(i \neq L) + \frac{1}{\alpha^* \pi_L} 1(i = L). \tag{65}$$

Therefore,

$$\begin{aligned}
\sum_j p_{ij}(W_{\sigma' j} - W_{\sigma' i}) &= (\sum_j p_{ij} W_{\sigma' j}) - W_{\sigma' i} \\
&= -\frac{1}{\mu_i} 1(i \in \sigma', i \neq L) - W_{\sigma' i} 1(i = L) + \frac{1}{\alpha^* \pi_L} 1(i = L) \\
&= -\frac{1}{\mu_i} 1(i \in \sigma') + \frac{1}{\alpha^* \pi_L} 1(i = L) \text{ (from (57))}. \qquad (66)
\end{aligned}$$

Now, noting that $q_{ij} = -\alpha^* \sum_{\sigma\sigma'} a_{\sigma\sigma'} W_{\sigma' i} W_{\sigma j}$, we show that the value of the objective function (46) is equal to (60) as follows,

$$\begin{aligned}
\sum_{ij} \pi_i p_{ij}(q_{ji} - q_{ii}) &= -\alpha^* \sum_{ij} \pi_i p_{ij} \sum_{\sigma\sigma'} a_{\sigma\sigma'}(W_{\sigma i} W_{\sigma' j} - W_{\sigma i} W_{\sigma' i}) \qquad (67) \\
&= -\alpha^* \sum_i \pi_i \sum_{\sigma\sigma'} a_{\sigma\sigma'} W_{\sigma i} (\sum_j p_{ij} W_{\sigma' j} - W_{\sigma' i}) \\
&= -\alpha^* \sum_i \pi_i \sum_{\sigma\sigma'} a_{\sigma\sigma'} W_{\sigma i} (-\frac{1}{\mu_i} 1(i \in \sigma') + \frac{1}{\alpha^* \pi_L} 1(i = L)) \text{ (from (66))} \\
&= \alpha^* \sum_i \pi_i \sum_\sigma \frac{W_{\sigma i}}{\mu_i} a_{\sigma\sigma(i)} - \sum_{\sigma\sigma'} a_{\sigma\sigma'} W_{\sigma L} \\
&= \alpha^* \sum_i \frac{\pi_i}{\mu_i} \sum_\sigma W_{\sigma i} a_{\sigma\sigma(i)} \text{ (from (61))}.
\end{aligned}$$

Now we show that the proposed solution is feasible. Note that from (5,6,7), $\frac{1}{\alpha^*} = \rho_\sigma = \sum_{i \in \sigma} \frac{\pi_i}{\mu_i}$. Hence,

$$\begin{aligned}
\sum_i s_i &= \alpha^* \sum_i \frac{\pi_i}{\mu_i} \mathrm{Min}_j \sum_{\sigma'} a_{\sigma(i)\sigma'} W_{\sigma' j} \\
&= \alpha^* \sum_\sigma \rho_\sigma \mathrm{Min}_j \sum_{\sigma'} a_{\sigma\sigma'} W_{\sigma' j} \\
&= \sum_\sigma \mathrm{Min}_j \sum_{\sigma'} a_{\sigma\sigma'} W_{\sigma' j} \\
&= -1.
\end{aligned}$$

So it only remains to verify (47) with the restriction $\theta(i) \leq \theta(j)$ eliminated in the first "Max" there. For $i \neq L$,

$$\mu_i \left( \sum_k p_{ik} q_{kj} - q_{ij} - \frac{s_i}{\pi_i} \right) = \alpha^* \mu_i \left( -\sum_{\sigma\sigma'} a_{\sigma\sigma'} \sum_k p_{ik} W_{\sigma' k} W_{\sigma j} + \sum_{\sigma\sigma'} a_{\sigma\sigma'} W_{\sigma' i} W_{\sigma j} \right.$$

$$
\begin{aligned}
&- \left(\frac{1}{\mu_i}\text{Min}_k \sum_\sigma a_{\sigma(i)\sigma} W_{\sigma k}\right)\Bigg) \\
&= \alpha^* \mu_i \sum_{\sigma\sigma'} a_{\sigma\sigma'} \left(\frac{1}{\mu_i} 1(i \in \sigma')\right) W_{\sigma j} - \alpha^* \text{Min}_k \sum_\sigma a_{\sigma(i)\sigma} W_{\sigma k} \ \ \text{(from (58))} \\
&= \alpha^* \sum_\sigma a_{\sigma\sigma(i)} W_{\sigma j} - \alpha^* \text{Min}_k \sum_\sigma a_{\sigma\sigma(i)} W_{\sigma k} \\
&\geq 0.
\end{aligned}
$$

For $i = L$,

$$
\begin{aligned}
\mu_L(\sum_k p_{Lk} q_{kj} - q_{Lj} - s_L) &= \alpha^* \mu_L \left( - \sum_{\sigma\sigma'} a_{\sigma\sigma'} \sum_k p_{Lk} W_{\sigma' k} W_{\sigma j} + \sum_{\sigma\sigma'} a_{\sigma\sigma'} W_{\sigma' L} W_{\sigma j} \right. \\
&\quad \left. - \frac{1}{\mu_L} \text{Min}_k \sum_\sigma a_{\sigma(L)\sigma} W_{\sigma k} \right) \\
&= \alpha^* \mu_L \left( - \frac{1}{\alpha^*} \sum_{\sigma\sigma'} \frac{a_{\sigma\sigma'} \rho_{\sigma'} W_{\sigma j}}{\pi_L} + \sum_{\sigma\sigma'} a_{\sigma\sigma'} \frac{1}{\mu_L} 1(L \in \sigma') W_{\sigma j} \right. \\
&\quad \left. - \frac{1}{\mu_L} \text{Min}_k \sum_\sigma a_{\sigma(L)\sigma} W_{\sigma k} \right) \ \ \text{(from (64,57))} \\
&= \alpha^* \mu_L \left( \sum_{\sigma\sigma'} a_{\sigma\sigma'} \frac{1}{\mu_L} 1(L \in \sigma') W_{\sigma j} - \frac{1}{\mu_L} \text{Min}_k \sum_\sigma a_{\sigma(L)\sigma} W_{\sigma k} \right) \ \ \text{(from (61))} \\
&= \alpha^* \left( \sum_\sigma a_{\sigma\sigma(L)} W_{\sigma j} - \text{Min}_k \sum_\sigma a_{\sigma(L)\sigma} W_{\sigma k} \right) \\
&\geq 0.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\text{LHS of (47) (without the restriction } \theta(i) \leq \theta(j)) &= \alpha^* \sum_{\sigma'} \left( \sum_\sigma a_{\sigma\sigma'} W_{\sigma j} - \text{Min}_k \sum_\sigma a_{\sigma\sigma'} W_{\sigma k} \right) \\
&= -\alpha^* \sum_{\sigma'} \text{Min}_k \sum_\sigma a_{\sigma\sigma'} W_{\sigma k} \ \text{(from (61))} \\
&= \alpha^* \ \text{(from (63))}.
\end{aligned}
$$

This proves (i).

To prove (ii), note that if $\sum_{\sigma'} a_{\sigma\sigma'} = 0$ for all $\sigma$, then for any $c_{\sigma\sigma'}$,

$$
\sum_\sigma \sum_{\sigma'} a_{\sigma\sigma'} c_{\sigma\sigma'} = \sum_\sigma \sum_{\sigma' \neq \sigma} a_{\sigma\sigma'} (c_{\sigma\sigma'} - c_{\sigma\sigma}).
$$

Note also that if $a_{\sigma\sigma'} = a_{\sigma'\sigma}$, then for any $d_{\sigma\sigma'}$, one has

$$\begin{aligned}\sum_\sigma \sum_{\sigma' \neq \sigma} a_{\sigma\sigma'} d_{\sigma\sigma'} &= \sum_\sigma \sum_{\sigma' \neq \sigma} a_{\sigma\sigma'} \left( \frac{d_{\sigma\sigma'} + d_{\sigma'\sigma}}{2} \right) \\ &= \sum_\sigma \sum_{\sigma' > \sigma} a_{\sigma\sigma'} (d_{\sigma\sigma'} + d_{\sigma'\sigma}) \text{ (assuming an order on the stations).}\end{aligned}$$

Thus, for $a_{\sigma\sigma'}$ satisfying (61, 62), one has

$$\sum_\sigma \sum_{\sigma'} a_{\sigma\sigma'} c_{\sigma\sigma'} = \sum_\sigma \sum_{\sigma' > \sigma} a_{\sigma\sigma'} (c_{\sigma\sigma'} + c_{\sigma'\sigma} - c_{\sigma\sigma} - c_{\sigma'\sigma'}) \text{ for any } c_{\sigma\sigma'}. \tag{68}$$

Applying this to (67) with $c_{\sigma\sigma'} = W_{\sigma i} W_{\sigma' j} - W_{\sigma i} W_{\sigma' i}$ yields (ii). □

Now we will show that the above reduced dimensional LP always provides a bound at least as good as (59), the conjectured pairwise bound. This will be done by further restricting in the LP of Theorem 5 the sign of $a_{\sigma\sigma'}$ to be negative for $\sigma \neq \sigma'$, and obtaining an explicit feasible, but not necessarily optimal, solution, whose value is (59). Thus the reduced dimensional LP of Theorem 5, which does not feature such a sign restriction, must also provide a bound at least as good as (59). In fact we will present an example (Example 3) where it is strictly better. This will illustrate that our reduced dimensional LP can capture interactions between multiple bottleneck stations.

**Theorem 6: Domination of the pairwise bound for closed balanced systems.** *For a balanced system, consider the LP of Theorem 5 with the additional restriction,*

$$a_{\sigma'\sigma} \leq 0 \qquad \text{for } \sigma \neq \sigma'. \tag{69}$$

*The optimal value of this LP is no less than (59). Hence the value $\underline{\nu}$ of the reduced dimensional LP of Theorem 5 is no less than (59), and its functional throughput bound is at least as good as,*

$$\alpha^u(N) \leq \alpha^* \frac{N}{N + \text{Max}_{\sigma \neq \sigma'} \frac{\alpha^* \sum_{ij} \pi_i p_{ij} \left[ (W_{\sigma i} - W_{\sigma' i})^2 - (W_{\sigma i} - W_{\sigma' i})(W_{\sigma j} - W_{\sigma' j}) \right]}{\text{Max}_j (W_{\sigma' j} - W_{\sigma j}) - \text{Min}_j (W_{\sigma' j} - W_{\sigma j})}}.$$

**Proof.** Using (61), we rewrite the constraint (63) as,

$$\begin{aligned} -1 &= \sum_\sigma \operatorname*{Min}_j \left( \sum_{\sigma'} a_{\sigma\sigma'} W_{\sigma' j} \right) \\ &= \sum_\sigma \operatorname*{Min}_j \left( \sum_{\sigma' \neq \sigma} a_{\sigma\sigma'} [W_{\sigma' j} - W_{\sigma j}] \right). \end{aligned}$$

Let $j(\sigma)$ minimize $\sum_{\sigma' \neq \sigma} a_{\sigma\sigma'} [W_{\sigma' j} - W_{\sigma j}]$ over all $j$. Then using (68) the constraint can be rewritten as,

$$\begin{aligned} -1 &= \sum_\sigma \sum_{\sigma' \neq \sigma} a_{\sigma\sigma'} [W_{\sigma' j(\sigma)} - W_{\sigma j(\sigma)}] \\ &= \sum_\sigma \sum_{\sigma' > \sigma} a_{\sigma\sigma'} [W_{\sigma' j(\sigma)} - W_{\sigma j(\sigma)} + W_{\sigma j(\sigma')} - W_{\sigma' j(\sigma')}]. \end{aligned}$$

Now fix two stations $\sigma \neq \sigma'$, and consider the choice,

$$a_{\sigma'\sigma} = a_{\sigma\sigma'} = -a_{\sigma\sigma} = -a_{\sigma'\sigma'} = \frac{-1}{W_{\sigma' j(\sigma)} - W_{\sigma j(\sigma)} + W_{\sigma j(\sigma')} - W_{\sigma' j(\sigma')}},$$

with all the other elements in $A$ set to zero.

If the denominator above is nonnegative, then this is a feasible solution, since $a_{\sigma\sigma'} \leq 0$, but it is not necessarily optimal. Moreover, since $a_{\sigma\sigma'} \leq 0$, $j(\sigma)$ *maximizes* $(W_{\sigma' j} - W_{\sigma j})$ over all $j$, while $j(\sigma')$ *maximizes* $(W_{\sigma j} - W_{\sigma' j})$ over all $j$. Hence

$$\begin{aligned} a_{\sigma\sigma'} &= \frac{-1}{\mathrm{Max}_j(W_{\sigma' j} - W_{\sigma j}) + \mathrm{Max}_j(W_{\sigma j} - W_{\sigma' j})} \\ &= \frac{-1}{\mathrm{Max}_j(W_{\sigma' j} - W_{\sigma j}) - \mathrm{Min}_j(W_{\sigma' j} - W_{\sigma j})}, \end{aligned}$$

which is indeed non-positive. Hence this choice of $A$ is feasible.

From Theorem 5.ii, the value of the objective function corresponding to this choice is

$$\frac{\alpha^* \sum_{ij} \pi_i p_{ij} \left[ (W_{\sigma i} - W_{\sigma' i})^2 - (W_{\sigma i} - W_{\sigma' i})(W_{\sigma j} - W_{\sigma' j}) \right]}{\mathrm{Max}_j(W_{\sigma' j} - W_{\sigma j}) - \mathrm{Min}_j(W_{\sigma' j} - W_{\sigma j})}.$$

Choosing that pair $\sigma \neq \sigma'$ which maximizes this expression gives the best among these feasible solutions, and is hence also feasible, proving the theorem. $\square$

In the following example this pairwise bound (59) is in fact optimal for the functional upper bound LP. Thus in this example at least we have explicitly solved $\overline{FB}(\infty)$.

**Example 2: The pairwise bound is optimal for $\overline{FB}(\infty)$**

Consider the system shown in Figure 3.

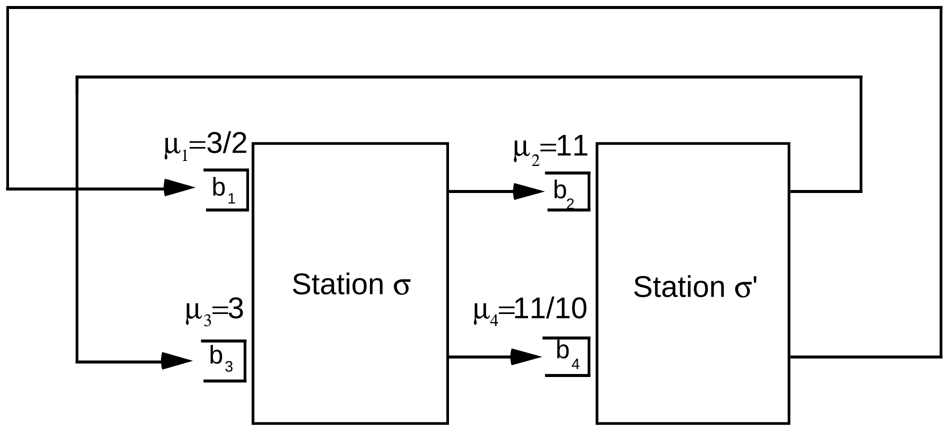


Figure 3: Closed re-entrant line of Example 2.

It has $L = 4$, $\alpha^* = 4$, $\pi_i \equiv \frac{1}{4}$, $p_{12} = p_{23} = p_{34} = p_{41} = 1$, $W_{\sigma'1} - W_{\sigma 1} = 0$, $W_{\sigma'2} - W_{\sigma 2} = 1 - \frac{1}{\mu_3}$, $W_{\sigma'3} - W_{\sigma 3} = \frac{1}{\mu_4} - \frac{1}{\mu_3}$, $W_{\sigma'4} - W_{\sigma 4} = \frac{1}{\mu_4}$. Thus,

$$\underline{\nu} \geq \frac{\alpha^* \sum_{ij} \pi_i p_{ij} \left[ (W_{\sigma i} - W_{\sigma' i})^2 - (W_{\sigma i} - W_{\sigma' i})(W_{\sigma j} - W_{\sigma' j}) \right]}{\mathrm{Max}_j (W_{\sigma' j} - W_{\sigma j}) - \mathrm{Min}_j (W_{\sigma' j} - W_{\sigma j})} = \frac{757}{990}.$$

This is *exactly* the value of the functional upper bound LP also. The resulting functional throughput bound is $\alpha^u(N) \leq \frac{4N}{N + 757/990}$. □

However, the pairwise bound (59) only captures the interactions between *pairs* of bottleneck stations. In the following three station example, the reduced dimensional LP of Theorem 5 without the sign restriction does capture interactions between *all* the three bottleneck stations.

**Example 3: Interaction between multiple bottleneck stations and the optimal sign-indefinite solution**

Consider the re-entrant line of Figure 4.

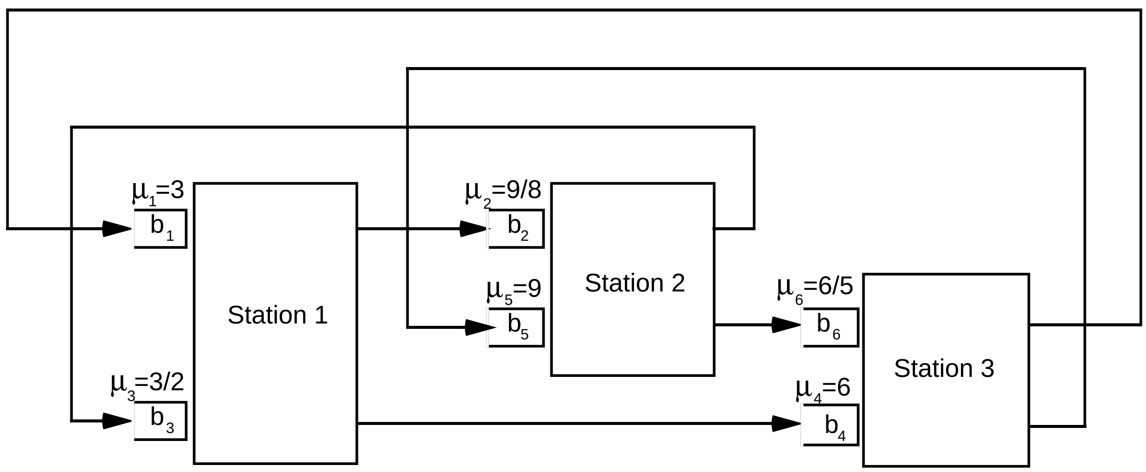


Figure 4: Closed re-entrant line of Example 3.

The explicit feasible solution identified in Theorem 6 gives $\underline{\nu} = 247/288$. It captures only the interaction between the two stations 2 and 3. It is not even optimal in the class of all matrices $A$ with non-positive off diagonal elements. This is basically due to the non-commutation of the two extremum operations in (59), which suggests, as in zero-sum game theory, that a "mixed" strategy is optimal. The best bound in the class of symmetric matrices $A$ with non-positive off-diagonal terms is $\underline{\nu} = 289/324$, which comes from,

$$A_{\text{Non-Positive Off-Diagonal Terms}} = \begin{bmatrix} 4/3 & -1 & -1/3 \\ -1 & 1 & 0 \\ -1/3 & 0 & 1/3 \end{bmatrix}.$$

However, the reduced dimensional LP of Theorem 5, by allowing *sign-indefinite* entries in $A$, obtains the even better bound $\underline{\nu} = 2065/2304$, which comes from,

$$A_{\text{Sign-Indefinite}} = \begin{bmatrix} 3/2 & -9/8 & -3/8 \\ -9/8 & 63/64 & 9/64 \\ -3/8 & 9/64 & 15/64 \end{bmatrix}.$$

It clearly results from the interactions between *all* the three bottleneck stations. □

## 9 Reduced dimensional LP for functional lower bound on throughput of balanced systems: A sufficient condition for efficiency of all policies

Now we address the problem of obtaining a reduced dimensional LP for obtaining a *lower* functional bound on the throughput of closed balanced multi-station systems. We obtain a sufficient condition under which we prove that all non-idling scheduling policies are efficient, and provide a lower bound on throughput given by a reduced dimensional LP.

Very interestingly, we will show later in Theorem 10 of Section 12 that when specialized to two station systems this lower bound shows that all non-idling policies have asymptotic loss no more than the formula for asymptotic loss conjectured by HW applied to the exact opposite of their conjectured asymptotically optimal policy, lending further credence to their conjectures.

**Theorem 7: Sufficient condition for efficiency of all non-idling policies and functional lower bound on throughput of balanced systems.** *Consider the following LP for a balanced system,*

$$\text{Min } -\alpha^* \sum_i \frac{\pi_i}{\mu_i} \sum_\sigma a_{\sigma(i)\sigma} W_{\sigma i},$$

*subject to,*

$$\sum_{\sigma'} a_{\sigma'\sigma} = 0 \text{ for all } \sigma, \tag{70}$$

$$\sum_\sigma Max_{j \notin \sigma} \sum_{\sigma'} a_{\sigma\sigma'} W_{\sigma' j} = -1, \tag{71}$$

$$a_{\sigma'\sigma} = a_{\sigma\sigma'} \text{ for all } \sigma, \sigma'.$$

*(i) If the LP is feasible, than all non-idling policies are efficient.*

*(ii) Moreover, if $\bar{\nu}$ is the value of the LP, then*

$$\alpha^u(N) \geq \alpha^* \frac{N}{N+\bar{\nu}} \text{ under any non-idling scheduling policy } u.$$

**Proof:** Define

$$Q := \alpha^* W^T A W, \text{ and}$$
$$s_i := -\frac{\alpha^* \pi_i}{\mu_i} \mathrm{Max}_{k \notin \sigma(i)} \sum_\sigma a_{\sigma(i)\sigma} W_{\sigma k} \text{ for all } i.$$

We will show that this solution is feasible for $\underline{FB}(\infty)$. Clearly,

$$\mu_i \left( \sum_k p_{ik} q_{kj} - q_{ij} - \frac{s_i}{\pi_i} \right) = \alpha^* \left( -\sum_\sigma a_{\sigma(i)\sigma} W_{\sigma j} + \mathrm{Max}_{k \notin \sigma(i)} \sum_\sigma a_{\sigma(i)\sigma} W_{\sigma k} \right).$$

Hence, if $i \notin \sigma(j)$, then $j \notin \sigma(i)$, and so

$$\mu_i \left( \sum_k p_{ik} q_{kj} - q_{ij} - s_i \right) \geq 0 \text{ for } i \notin \sigma(j).$$

Thus

$$\begin{aligned} \text{LHS of (52) (without the restriction } \theta(i) \leq \theta(j)) &= \alpha^* \sum_{\sigma'} \left[ -\sum_\sigma a_{\sigma'\sigma} W_{\sigma j} + \mathrm{Max}_{k \notin \sigma'} \sum_\sigma a_{\sigma'\sigma} W_{\sigma k} \right] \\ &= \alpha^* \sum_{\sigma'} \mathrm{Max}_{k \notin \sigma'} \sum_\sigma a_{\sigma'\sigma} W_{\sigma k} \text{ (using (70))} \\ &= -\alpha^* \text{ (from (71))}. \end{aligned}$$

Moreover $\sum_i s_i = 1$, since $\sum_{i \in \sigma} \frac{\alpha^* \pi_i}{\mu_i} = 1$ for all $\sigma$. □

# 10 Two station systems: The HW conjectures

In recent years, there has been much interest in studying queueing networks by analyzing Brownian networks intended to approximate their behavior in heavy traffic. The greatest success, in the authors' opinion, has been in the application of this approach to the study of closed queueing networks with *two* stations. By studying the conjectured Reflected Brownian Motion arising from a "workload imbalance process," Harrison and Wein (1990) made two contributions. First they synthesized a buffer priority policy, which they conjectured had the best throughput for large $N$. Second, they obtained a simple formula for approximating the throughput of any buffer priority policy for large $N$. Their solutions are elegant, and provide an understanding of the relationship between the workload imbalance process and the idleness forced on stations, which leads to a loss in the throughput. We now provide a more precise statement of how we interpret their conjectures.

Let $\sigma_1$ and $\sigma_2$ denote the two stations in the system. Without loss of generality, we suppose from now on that $\rho_{\sigma_1} \le \rho_{\sigma_2}$. Define the *normalized load* $\bar{\rho}_\sigma := \rho_\sigma / \operatorname{Max}_{\sigma'} \rho_{\sigma'}$. The *HW index* $\eta_i$ of buffer $b_i$ is,

$$\begin{aligned} \eta_i &:= \bar{\rho}_{\sigma_2} W_{\sigma_1,i} - \bar{\rho}_{\sigma_1} W_{\sigma_2,i} \text{ for } i \in \sigma_1, \\ &:= \bar{\rho}_{\sigma_1} W_{\sigma_2,i} - \bar{\rho}_{\sigma_2} W_{\sigma_1,i} \text{ for } i \in \sigma_2. \end{aligned}$$

Let $\theta_{HW} = \{\theta_{HW}(1), \ldots, \theta_{HW}(L)\}$ be a buffer priority policy which gives priority to buffers with lower values of the index $\eta_i$, i.e., $\theta_{HW}(i) < \theta_{HW}(j)$ only if $\eta_i \le \eta_j$ (breaking ties arbitrarily). We call $\theta_{HW}$ the *HW-policy.*

**Conjecture 1: Asymptotic optimality of $\theta_{HW}$.** HW conjecture that the buffer priority policy $\theta_{HW}$ is asymptotically optimal (we should mention that this is our interpretation of the gist of HW):

$$\limsup_{N\to\infty} N[\alpha^* - \alpha^{\theta_{HW}}(N)] \le \liminf_{N\to\infty} N[\alpha^* - \alpha^u(N)] \text{ for all policies } u.$$

Implicit in the above conjecture is the assumption that $\theta_{HW}$ is efficient:

$$\lim_{N\to\infty} \alpha^{\theta_{HW}}(N) = \alpha^*.$$

**Conjecture 2: Asymptotic loss of buffer priority policies**. Our interpretation of the gist of HW on the asymptotic loss of any buffer priority policy $\theta$ is the following:

$$\begin{aligned} \lim_{N\to\infty} N[\alpha^* - \alpha^\theta(N)] &= \frac{\sigma^2}{2(d(\theta) - c(\theta))} \text{ if the system is balanced,} \\ &= \frac{\rho_{\sigma_1} - \rho_{\sigma_2}}{\exp\left(\frac{2(\rho_{\sigma_1} - \rho_{\sigma_2})(d(\theta) - c(\theta))}{\sigma^2}\right) - 1} \text{ if the system is unbalanced.} \end{aligned}$$

Above,

$$\sigma^2 := \sum_{i,j} \frac{\pi_i p_{ij}}{\rho_{\sigma(i)}} \left[ (\bar{\rho}_{\sigma_2} W_{\sigma_1,j} - \bar{\rho}_{\sigma_1} W_{\sigma_2,j}) - (\bar{\rho}_{\sigma_2} W_{\sigma_1,i} - \bar{\rho}_{\sigma_1} W_{\sigma_2,i}) \right]^2.$$

is the conjectured variance of the Reflected Brownian Motion which is the presumed limit of their "workload imbalance process." (We omit the calculations involved in reducing the expression given by HW to that above). For a *balanced* system, this expression simplifies to the following formula (the reader should compare this to Theorem 5.ii):

$$\sigma^2 = 2\alpha^* \sum_{i,j} \pi_i p_{ij} [(W_{\sigma_1,i} - W_{\sigma_2,i})^2 - (W_{\sigma_1,i} - W_{\sigma_2,i})(W_{\sigma_1,j} - W_{\sigma_2,j})]. \tag{72}$$

For a *balanced re-entrant line*, simple calculations (using (79) below) give $\sigma^2 = \frac{2\alpha^*}{L} \sum_{i=1}^{L} \frac{\eta_i}{\mu_i}$. The quantities $c(\theta)$ and $d(\theta)$ are given by,

$$c(\theta) := \text{Min}\left\{\bar{\rho}_{\sigma_2} W_{\sigma_1,i_1(\theta)} - \bar{\rho}_{\sigma_1} W_{\sigma_2,i_1(\theta)}, \bar{\rho}_{\sigma_2} W_{\sigma_1,i_2(\theta)} - \bar{\rho}_{\sigma_1} W_{\sigma_2,i_2(\theta)}\right\}, \tag{73}$$

$$d(\theta) := \text{Max}\left\{\bar{\rho}_{\sigma_2} W_{\sigma_1,i_1(\theta)} - \bar{\rho}_{\sigma_1} W_{\sigma_2,i_1(\theta)}, \bar{\rho}_{\sigma_2} W_{\sigma_1,i_2(\theta)} - \bar{\rho}_{\sigma_1} W_{\sigma_2,i_2(\theta)}\right\}. \tag{74}$$

where,

$$b_{i_n(\theta)} := \text{ lowest priority buffer at } \sigma_n. \tag{75}$$

The work of Harrison and Nguyen (1995) shows that the closed version of the example in Lu and Kumar (1991) is *not* efficient. Thus the above conjecture is not true for all buffer priority policies in general.

In Sections 11 and 12 we show the following results:

**(i)** The HW-policy is efficient for all two station systems.

**(ii)** $\alpha^{\theta_{HW}}(N) \geq \frac{N}{N+\alpha^* \sum_{i,j} \pi_i p_{ij} \left[\left(\bar{\rho}_{\sigma_2} W_{\sigma_1,i} - \bar{\rho}_{\sigma_1} W_{\sigma_2,i}\right) - \text{Min}\left\{\bar{\rho}_{\sigma_2} W_{\sigma_1,i} - \bar{\rho}_{\sigma_1} W_{\sigma_2,i}, \bar{\rho}_{\sigma_2} W_{\sigma_1,j} - \bar{\rho}_{\sigma_1} W_{\sigma_2,j}\right\}\right]} \alpha^*$ for *all* two station systems.

**(iii)** $\alpha^{\theta_{HW}}(N) \geq \frac{N}{N+1} \alpha^*$ for all *balanced* re-entrant lines.

**(iv)** For a balanced re-entrant line, no system can be proved to have an asymptotic loss less than 1 through our approach.

**(v)** For balanced systems, for all policies $u$,

$$\alpha^u(N) \leq \frac{N}{N + \frac{\sigma^2}{2(d(\theta_{HW}) - c(\theta_{HW}))}} \alpha^*.$$

Hence no policy can have an asymptotic loss better than that conjectured for the HW-policy.

**(vi)** If

$$\operatorname*{Min}_{i \in \sigma_1}(W_{\sigma_1,i} - W_{\sigma_2,i}) > \operatorname*{Max}_{i \in \sigma_2}(W_{\sigma_1,i} - W_{\sigma_2,i}), \tag{76}$$

then *all* non-idling scheduling policies are efficient. Also, for every such policy $u$,

$$\alpha^u(N) \geq \frac{N}{N + \frac{\sigma^2}{2(d(\theta_{Anti-HW}) - c(\theta_{Anti-HW}))}} \alpha^*,$$

where $\theta_{Anti-HW}$ is the exact *opposite* of the HW-policy. So no non-idling policy can have a worse asymptotic loss than the conjectured formula of HW applied to the opposite of the HW-policy.

These results do not completely prove all the results conjectured by HW (indeed they are not all true), but do give credibility to them. We conjecture that under the additional condition (76), missing in their work, all their conjectures do hold for balanced systems.

# 11 Functional lower bound for HW-policy

In this section we focus on $\theta_{HW}$. We show that it is efficient, obtain a functional lower bound on its throughput, and thus also an upper bound on its asymptotic loss. Moreover, for balanced closed re-entrant lines, we show that no lower functional bound than that for $\theta_{HW}$ can be established for any policy, if one uses the approach developed here. All this is done by merely exhibiting a particular feasible solution for $\underline{FB}^{\theta_{HW}}(\infty)$ of Theorem 4.

**Theorem 8: Efficiency, upper bound on asymptotic loss, and functional lower bound for $\theta_{HW}$.** *Consider a system with two stations.*

**(i)** *$\theta_{HW}$ is efficient.*

**(ii)** $\alpha^{\theta_{HW}}(N) \geq \frac{N}{N+\sum_{i,j}\pi_i p_{ij}\alpha^*\left[\left(\bar{\rho}_{\sigma_2}W_{\sigma_1,i}-\bar{\rho}_{\sigma_1}W_{\sigma_2,i}\right)-\text{Min}\left\{\bar{\rho}_{\sigma_2}W_{\sigma_1,i}-\bar{\rho}_{\sigma_1}W_{\sigma_2,i},\bar{\rho}_{\sigma_2}W_{\sigma_1,j}-\bar{\rho}_{\sigma_1}W_{\sigma_2,j}\right\}\right]}\alpha^*$ *for all* $N$*. Hence*

$$\overline{J}(\theta_{HW}) \leq \sum_{i,j}\pi_i p_{ij}\alpha^*\left[(\bar{\rho}_{\sigma_2}W_{\sigma_1,i}-\bar{\rho}_{\sigma_1}W_{\sigma_2,i}) - \text{Min}\left\{\bar{\rho}_{\sigma_2}W_{\sigma_1,i}-\bar{\rho}_{\sigma_1}W_{\sigma_2,i},\bar{\rho}_{\sigma_2}W_{\sigma_1,j}-\bar{\rho}_{\sigma_1}W_{\sigma_2,j}\right\}\right].$$

**(iii)** *For a balanced re-entrant line,* $\alpha^{\theta_{HW}}(N) \geq \frac{N}{N+1}\alpha^*$ *for all* $N$*. Hence* $\overline{J}(\theta_{HW}) \leq 1$.

**(iv)** *For a balanced re-entrant line,* $\underline{FB}(\infty)$ *and* $\underline{FB}^\theta(\infty)$ *cannot have values of* $\sum_{i,j}\pi_i p_{ij}(q_{ii}-q_{ji})$ *less than 1, if* $VD\underline{T}(\infty)=\alpha^*$ *or* $VD\underline{T}^\theta(\infty)=\alpha^*$*, respectively.*

**Proof:** Define

$$\begin{aligned} s_i &:= \frac{\alpha^*\pi_i}{\mu_i} \text{ for } i\in\sigma_2, \\ &:= 0 \text{ for } i\in\sigma_1, \end{aligned} \tag{77}$$

$$q_{ij} := \alpha^*\,\text{Min}\left\{\bar{\rho}_{\sigma_2}W_{\sigma_1,i}-\bar{\rho}_{\sigma_1}W_{\sigma_2,i},\bar{\rho}_{\sigma_2}W_{\sigma_1,j}-\bar{\rho}_{\sigma_1}W_{\sigma_2,j}\right\}. \tag{78}$$

This solution $\{q_{ij}, s_i\}$ clearly satisfies (30, 31). Thus if we can verify that (29) is satisfied with $\alpha=\alpha^*$, then (i) is true, and since (52) is satisfied, it is a feasible solution for $\underline{FB}^{\theta_{HW}}(\infty)$.

The following equalities from (65) are used throughout the following proofs:

$$\begin{aligned} \sum_k p_{ik}W_{\sigma_n,k} &= W_{\sigma_n,i} \text{ if } i\neq L \text{ and } \sigma(i)\neq\sigma_n, \\ &= -\frac{1}{\mu_i}+W_{\sigma_n,i} \text{ if } i\neq L \text{ and } i\in\sigma_n, \\ &= \frac{\rho_{\sigma_n}}{\pi_L} \text{ if } i=L. \end{aligned} \tag{79}$$

Now note that,

$$\begin{aligned} &\sum_k p_{ik}\,\text{Min}\left\{\bar{\rho}_{\sigma_2}W_{\sigma_1,k}-\bar{\rho}_{\sigma_1}W_{\sigma_2,k},\bar{\rho}_{\sigma_2}m_{\sigma_1,j}-\bar{\rho}_{\sigma_1}W_{\sigma_2,j}\right\} \\ &\leq \text{Min}\{\sum_k p_{ik}(\bar{\rho}_{\sigma_2}W_{\sigma_1,k}-\bar{\rho}_{\sigma_1}W_{\sigma_2,k}),\bar{\rho}_{\sigma_2}m_{\sigma_1,j}-\bar{\rho}_{\sigma_1}W_{\sigma_2,j}\}. \end{aligned}$$

Thus we obtain from the definition of $q_{ij}$ in (78) that

$$\begin{aligned} \sum_k p_{ik}\left(q_{kj}-q_{ij}-\frac{s_i}{\pi_i}\right) \leq\ & \alpha^*\,\text{Min}\left\{\sum_k p_{ik}(\bar{\rho}_{\sigma_2}m_{\sigma_1,k}-\bar{\rho}_{\sigma_1}W_{\sigma_2,k}),\bar{\rho}_{\sigma_2}W_{\sigma_1,j}-\bar{\rho}_{\sigma_2}m_{\sigma_2,j}\right\} \\ & -\alpha^*\,\text{Min}\left\{\bar{\rho}_{\sigma_2}W_{\sigma_1,i}-\bar{\rho}_{\sigma_1}m_{\sigma_2,i},\bar{\rho}_{\sigma_2}W_{\sigma_1,j}-\bar{\rho}_{\sigma_1}W_{\sigma_2,j}\right\}-\frac{s_i}{\pi_i}. \end{aligned} \tag{80}$$

Using (79), the definition of $s_i$ in (77), the assumption $\bar{\rho}_{\sigma_1} \leq \bar{\rho}_{\sigma_2} = 1$, and considering the various possibilities for $i$ and $j$, we upper bound the RHS of (80) as follows.

**Case i:** $j \in \sigma_1$, $i \in \sigma_1$, $i = L$, and $\theta_{HW}(i) \leq \theta_{HW}(j)$.

$$\begin{aligned}
\text{RHS of (80)} &= \alpha^* \operatorname{Min}\left\{-\frac{\bar{\rho}_{\sigma_2}}{\mu_i} + \bar{\rho}_{\sigma_2} W_{\sigma_1,i} - \bar{\rho}_{\sigma_2} W_{\sigma_2,i}, \bar{\rho}_{\sigma_2} m_{\sigma_1,j} - \bar{\rho}_{\sigma_1} W_{\sigma_2,j}\right\} \\
&\quad -\alpha^* \operatorname{Min}\left\{\bar{\rho}_{\sigma_2} W_{\sigma_1,j} - \bar{\rho}_{\sigma_1} W_{\sigma_2,i}, \bar{\rho}_{\sigma_2} W_{\sigma_1,j} - \bar{\rho}_{\sigma_1} W_{\sigma_2,j}\right\} \quad (\text{since } i \neq L, i \in \sigma_1) \\
&= -\alpha^* \frac{\bar{\rho}_{\sigma_2}}{\mu_i} \qquad (\text{since } \theta_{HW}(i) \leq \theta_{HW}(j)) \\
&= -\frac{\alpha^*}{\mu_i} \qquad (\text{since } \bar{\rho}_{\sigma_2} = 1).
\end{aligned}$$

**Case ii:** $j \in \sigma_1$, $i \in \sigma_1$, $i = L$, and $\theta_{HW}(i) \leq (j)$.

$$\begin{aligned}
\text{RHS of (80)} &= \alpha^* \operatorname{Min}\left\{0, \bar{\rho}_{\sigma_2} W_{\sigma_1,j} - \bar{\rho}_{\sigma_1} W_{\sigma_2,j}\right\} \\
&\quad -\alpha^* \operatorname{Min}\left\{\frac{\bar{\rho}_{\sigma_2}}{\mu_L}, \bar{\rho}_{\sigma_2} W_{\sigma_1,j} - \bar{\rho}_{\sigma_1} m_{\sigma_2,j}\right\} \quad (\text{since } i = L \text{ and } i \in \sigma_1) \\
&= -\frac{\alpha^*}{\mu_L} \qquad (\text{since } \theta_{HW}(i) \leq \theta_{HW}(j) \text{ and } \bar{\rho}_{\sigma_2} = 1).
\end{aligned}$$

**Case iii:** $j \in \sigma_1$, $i \in \sigma_2$, and $i \neq L$.

$$\begin{aligned}
\text{RHS of (80)} &= \alpha^* \operatorname{Min}\left\{\frac{\bar{\rho}_{\sigma_1}}{\mu_i} + \bar{\rho}_{\sigma_2} W_{\sigma_1,i} - \bar{\rho}_{\sigma_2} W_{\sigma_2,i}, \bar{\rho}_{\sigma_2} m_{\sigma_1,j} - \bar{\rho}_{\sigma_1} W_{\sigma_2,j}\right\} \\
&\quad -\alpha^* \operatorname{Min}\left\{\bar{\rho}_{\sigma_2} W_{\sigma_1,j} - \bar{\rho}_{\sigma_1} W_{\sigma_2,i}, \bar{\rho}_{\sigma_2} W_{\sigma_1,j} - \bar{\rho}_{\sigma_1} W_{\sigma_2,j}\right\} - \frac{\alpha^*}{\mu_i} \\
&\qquad (\text{since } i \in \sigma_2 \text{ and } i \neq L) \\
&\leq -\alpha^* \frac{\bar{\rho}_{\sigma_1}}{\mu_i} - \frac{\alpha^*}{\mu_i} \\
&\leq 0 \qquad (\text{since } \bar{\rho}_{\sigma_1} \leq 1).
\end{aligned}$$

**Case iv:** $j \in \sigma_1$, $i \in \sigma_2$, and $i = L$.

$$\text{RHS of (80)} = \alpha^* \operatorname{Min}\left\{0, \bar{\rho}_{\sigma_2} W_{\sigma_1,j} - \bar{\rho}_{\sigma_1} m_{\sigma_2,j}\right\}$$

$$
\begin{aligned}
& -\alpha^* \operatorname{Min}\left\{-\frac{\bar{\rho}_{\sigma_1}}{\mu_L}, \bar{\rho}_{\sigma_2} W_{\sigma_1,j} - \bar{\rho}_{\sigma_1} W_{\sigma_2,j}\right\} \\
& -\frac{\alpha^*}{\mu_i} \qquad \text{(since } i = L \text{ and } i \in \sigma_2\text{)} \\
\leq \;& \frac{\alpha^* \bar{\rho}_{\sigma_1}}{\mu_i} - \frac{\alpha^*}{\mu_i} \\
\leq \;& 0 \qquad \text{(since } \bar{\rho}_{\sigma_1} \leq 1\text{).}
\end{aligned}
$$

**Case v:** $j \in \sigma_2$, $i \in \sigma_1$, and $i \neq L$.

$$
\begin{aligned}
\text{RHS of (80)} \;=\;& \alpha^* \operatorname{Min}\left\{-\frac{\bar{\rho}_{\sigma_2}}{\mu_i} + \bar{\rho}_{\sigma_2} m_{\sigma_1,i} - \bar{\rho}_{\sigma_1} W_{\sigma_2,i}, \bar{\rho}_{\sigma_2} W_{\sigma_1,j} - \bar{\rho}_{\sigma_1} m_{\sigma_2,j}\right\} \\
& -\alpha^* \operatorname{Min}\left\{\bar{\rho}_{\sigma_2} W_{\sigma_1,i} - \bar{\rho}_{\sigma_1} m_{\sigma_2,i}, \bar{\rho}_{\sigma_2} W_{\sigma_1,j} - \bar{\rho}_{\sigma_1}, W_{\sigma_2,j}\right\} \\
& \text{(since } i \neq L \text{ and } i \in \sigma_1\text{)} \\
\leq \;& 0.
\end{aligned}
$$

**Case vi:** $j \in \sigma_2$, $i \in \sigma_1$, and $i = L$.

$$
\begin{aligned}
\text{RHS of (80)} \;=\;& \alpha^* \operatorname{Min}\left\{0, \bar{\rho}_{\sigma_2} W_{\sigma_1,j} - \bar{\rho}_{\sigma_1} m_{\sigma_2,j}\right\} \\
& -\alpha^* \operatorname{Min}\left\{\frac{\bar{\rho}_{\sigma_2}}{\mu_L}, \bar{\rho}_{\sigma_2} W_{\sigma_1,j} - \bar{\rho}_{\sigma_1} W_{\sigma_2,j}\right\} \text{(since } i = L \text{ and } i \in \sigma_1\text{)} \\
\leq \;& 0.
\end{aligned}
$$

**Case vii:** $j \in \sigma_2$, $i \in \sigma_2$, $i \neq L$ and $\theta_{HW}(i) \leq \theta_{HW}(j)$.

$$
\begin{aligned}
\text{RHS of (80)} \;=\;& \alpha^* \operatorname{Min}\left\{\frac{\bar{\rho}_{\sigma_1}}{\mu_i} + \bar{\rho}_{\sigma_2} W_{\sigma_1,i} - \bar{\rho}_{\sigma_1} W_{\sigma_2,i}, \bar{\rho}_{\sigma_2} m_{\sigma_1,j} - \bar{\rho}_{\sigma_1} W_{\sigma_2,j}\right\} \\
& -\alpha^* \operatorname{Min}\left\{\bar{\rho}_{\sigma_2} W_{\sigma_1,i} - \bar{\rho}_{\sigma_1} W_{\sigma_2,i}, \bar{\rho}_{\sigma_2} W_{\sigma_1,j} - \bar{\rho}_{\sigma_1} W_{\sigma_2,j}\right\} \\
& -\frac{\alpha^*}{\mu_i} \qquad \text{(since } i \neq L \text{ and } i \in \sigma_2\text{)} \\
=\;& -\frac{\alpha^*}{\mu_i} \qquad \text{(since } \theta_{HW}(i) \leq \theta_{HW}(j)\text{).}
\end{aligned}
$$

**Case viii:** $j \in \sigma_2$, $i \in \sigma_2$, $i = L$ and $\theta_{HW}(i) \leq \theta_{HW}(j)$.

$$
\text{RHS of (80)} \;=\; \alpha^* \operatorname{Min}\left\{0, \bar{\rho}_{\sigma_2} W_{\sigma_1,j} - \bar{\rho}_{\sigma_1} m_{\sigma_2,j}\right\}
$$

$$
\begin{aligned}
& -\alpha^* \text{Min}\left\{-\frac{\bar{\rho}_{\sigma_1}}{\mu_L}, \bar{\rho}_{\sigma_2} W_{\sigma_1,j} - \bar{\rho}_{\sigma_1} W_{\sigma_2,j}\right\} \\
& -\frac{\alpha^*}{\mu_i} \qquad (\text{since } i = L \text{ and } i \in \sigma_2) \\
= & -\frac{\alpha^*}{\mu_i} \qquad (\text{since } \theta_{HW}(i) \le \theta_{HW}(j)).
\end{aligned}
$$

Thus, we see that in all cases,

$$
\underset{\{i\in\sigma(j) \text{ and } \theta_{HW}(i)\le\theta_{HW}(j)\}}{\text{Max}} \mu_i[\sum_k p_{ik}(q_{kj} - q_{ij} - \frac{s_i}{\pi_i})] + \underset{i\notin\sigma(j)}{\text{Max}} \mu_i[\sum_k p_{ik}(q_{kj} - q_{ij} - \frac{s_i}{\pi_i})]^+ = -\alpha^*.
$$

Hence we see from Theorem 3.i and Theorem 2.i that $VD\underline{T}^{\theta_{HW}}(\infty) = \alpha^*$, proving claim (i). To obtain the functional lower bound on the throughput $\alpha^{\theta_{HW}}(N)$ in (ii), we simply invoke (53) of Theorem 4.iii.

To show (iii), note that for a re-entrant line, $\pi_i = \frac{1}{L}$, $p_{i,i+1} = 1$ for $1 \le i \le L-1$, and $p_{L,1} = 1$. Hence $v = \sum_{i,j} \pi_i p_{ij}(q_{ji} - q_{ii}) = \frac{1}{L}\sum_i (q_{i+1,i} - q_{ii})$. For $i \in \sigma_1$,

$$
\begin{aligned}
q_{i+1,i} &= \alpha^* \text{Min}\left\{W_{\sigma_1,i+1} - m_{\sigma_2,i+1}, W_{\sigma_{1,i}} - W_{\sigma_{2,i}}\right\} \qquad (\text{since} \bar{\rho}_{\sigma_1} = \bar{\rho}_{\sigma_2} = 1) \\
&= \alpha^* \text{Min}\{-\frac{1}{\mu_i} + W_{\sigma_1,i} - W_{\sigma_2,i}, +W_{\sigma_1,i} - W_{\sigma_2,i}\} \qquad (\text{since } i \in \sigma_1) \\
&= -\frac{\alpha^*}{\mu_i} + q_{ii}.
\end{aligned}
$$

For $i \in \sigma_2$, similarly $q_{i+1,i} = q_{ii}$. Hence

$$
\begin{aligned}
v &= -\frac{\alpha^*}{L}\sum_{i\in\sigma_1}\frac{1}{\mu_i} \\
&= -\alpha^* \sum_{i\in\sigma_1}\frac{\pi_i}{\mu_i} \quad \left(\text{since } \pi_i = \frac{1}{L}\right) \\
&= -1 \qquad (\text{from (7) since the system is balanced}).
\end{aligned}
$$

To show (iv) note that for any feasible solution $\{q_{ij}, s_i\}$ to $\underline{FB}(\infty)$ or $\underline{FB}^\theta(\infty)$, one would have $\mu_i(q_{i+1,i} - q_{ii} - Ls_i) \le -\alpha^*$ (we interpret the subscript $L+1$ as 1). Hence $\sum_{i=1}^L (q_{i+1,i} - q_{ii}) \le -\sum_{i=1}^L \frac{\alpha^*}{\mu_i} + L\sum_{i=1}^L s_i = -L\sum_{i=1}^L \frac{\alpha^*\pi_i}{\mu_i} + L\sum_{i=1}^L s_i = -2L + L = -L$. So $\frac{1}{L}\sum_{i=1}^L (q_{ii} - q_{i+1,i}) \ge 1$. □

## 12 The HW and Anti-HW lower and upper bounds

We will now show that the conjectured asymptotic loss of the HW-policy is a lower bound for all policies in balanced systems. We will also show that the conjectured formula for the asymptotic loss applied to the exact *opposite* of the HW-policy is an upper bound for all non-idling policies in balanced systems, provided that a certain additional condition holds. Under this condition all non-idling policies are thus in particular also efficient.

The first of these results is obtained by merely showing that the bound of Theorem 6 for multi-station systems is indeed the conjectured asymptotic loss of the HW-policy when it is specialized to two station systems.

**Theorem 9: Functional upper bound and lower bound on asymptotic loss for all policies in balanced systems.** *Consider a balanced two station system.*

**(i)** $\alpha^u(N) \leq \frac{N}{N+\frac{\sigma^2}{2(d(\theta_{HW})-c(\theta_{HW}))}}\alpha^*$ *for all policies $u$, and all $N$.*

**(ii)** $\underline{J}(u) \geq \frac{\sigma^2}{2(d(\theta_{HW})-c(\theta_{HW}))}$ *for all policies $u$. Thus,* no *policy has lower asymptotic loss than the conjectured asymptotic loss of the HW-policy.*

**Proof:** From Theorem 6 specialized to to station systems, we see that $\alpha^u(N) \leq \frac{N}{N+\frac{\sigma^2}{2(d-c)}}\alpha^*$, where $\sigma^2$ is as defined in (72),

$$c := \operatorname*{Min}_i(W_{\sigma_1,i} - W_{\sigma_2,i}), \text{ and } d := \operatorname*{Max}_i(W_{\sigma_1,i} - W_{\sigma_2,i}).$$

The proof is therefore completed by showing that $c = c(\theta_{HW})$ and $d = d(\theta_{HW})$. From the definition of $\theta_{HW}$, and (73,74,75), it follows that all we need to show is that the minimum of $(W_{\sigma_1,i} - W_{\sigma_2,i})$ is attained at some $i \in \sigma_2$, and the maximum at some $i \in \sigma_1$. For this, we use (79,6). For $i \neq L$ and $i \in \sigma_1$, we have $\sum_k p_{ik}(W_{\sigma_1,k} - W_{\sigma_2,k}) = -\frac{1}{\mu_i} + (W_{\sigma_1,i} - W_{\sigma_2,i})$. Hence the minimum is either attained at $i = L$ or at some $i \in \sigma_2$. If $i = L$ with $i \in \sigma_1$ attains the minimum, then $(W_{\sigma_{1,L}} - W_{\sigma_{2,L}}) = \frac{1}{\mu_L} > 0$, but $\sum_k p_{Lk}(W_{\sigma_1,k} - W_{\sigma_2,k}) = 0$, which is a contradiction. Thus the minimum is attained at some $i \in \sigma_2$. Similarly, the maximum is attained at some $i \in \sigma_1$. $\square$

Now we turn to the issue of establishing an upper bound on asymptotic loss for all non-idling policies in balanced systems. From the arguments made by HW in favor of $\theta_{HW}$ one expects that the non-idling policy with the *worst* performance is that which uses the exact *opposite* ordering to $\theta_{HW}$. Let us call such a buffer priority policy which gives preference to buffers with larger values of the index $\eta_i$ as $\theta_{Anti-HW}$. We now identify a sufficient condition under which *all* non-idling policies indeed have a lower asymptotic loss than the conjectured formula for asymptotic loss applied to $\theta_{Anti-HW}$ (and are thus also all efficient). This is more evidence in favor of the conjectures of HW.

This result also is obtained by applying the functional lower bound of Theorem 7 to two station systems. We simply identify a sufficient condition under which the LP of Theorem 7 is feasible, with an explicit feasible solution whose value is equal to the conjectured formula for the asymptotic loss applied to the $\theta_{Anti-HW}$.

**Theorem 10: Sufficient condition for efficiency, functional lower bound, and upper bound on asymptotic loss, for all non-idling policies.** *Consider a balanced two station system. Suppose*

$$\operatorname*{Min}_{j \notin \sigma_2}(W_{\sigma_1 j} - W_{\sigma_2 j}) > \operatorname*{Max}_{j \notin \sigma_1}(W_{\sigma_1 j} - W_{\sigma_2 j}). \tag{81}$$

*Then,*

**(i)** *All non-idling policies $u$ are efficient.*

**(ii)** $\alpha^u(N) \geq \frac{N}{N + \frac{\sigma^2}{2(d(\theta_{Anti-HW}) - c(\theta_{Anti-HW}))}} \alpha^*$ *for all non-idling policies $u$, and all $N$. Hence* $\overline{J}(u) \leq \frac{\sigma^2}{2(d(\theta_{Anti-HW}) - c(\theta_{Anti-HW}))}$.

**Proof:** Consider the LP of Theorem 7 for balanced multi-station systems. We will impose the additional restriction $a_{\sigma\sigma'} \leq 0$ for $\sigma \neq \sigma'$. The proof proceeds in the same way as Theorem 6 followed from Theorem 5 under this additional restriction.

Suppose that the following set is nonempty:

$$\Delta := \left\{ (\sigma, \sigma') : \sigma \neq \sigma' \text{ and } \underset{j \notin \sigma'}{\text{Min}}(W_{\sigma j} - W_{\sigma' j}) > \underset{j \notin \sigma}{\text{Max}}(W_{\sigma j} - W_{\sigma' j}) \right\}.$$

Pick $(\bar{\sigma}, \bar{\sigma}') \in \Delta$. Then set

$$\begin{aligned} a_{\overline{\sigma}\overline{\sigma}'} &= \frac{-1}{\underset{j \in \overline{\sigma}'}{\text{Min}}(W_{\sigma j} - W_{\sigma' j}) - \underset{j \in \overline{\sigma}}{\text{Max}}(W_{\overline{\sigma}' j} - W_{j \in \overline{\sigma}})}, \\ a_{\overline{\sigma}\overline{\sigma}'} &= 0 \text{ if } (\sigma, \sigma') \neq (\overline{\sigma}, \overline{\sigma}') \text{ and } (\sigma, \sigma') \neq (\overline{\sigma}', \overline{\sigma}). \end{aligned}$$

This solution is feasible for the LP in Theorem 7. Hence we obtain that $\alpha^u(N) \geq \alpha^* \frac{N}{N+\widetilde{\nu}}$,, where

$$\widetilde{\nu} := \underset{(\sigma,\sigma') \in \Delta}{\text{Min}} \frac{\sum_{ij} \pi_i p_{ij} \left[ (W_{\sigma i} - W_{\sigma' i})^2 - (W_{\sigma i} - W_{\sigma' i})(W_{\sigma j} - W_{\sigma' j}) \right]}{\underset{j \notin \sigma'}{\text{Min}}(W_{\sigma j} - W_{\sigma' j}) - \underset{j \notin \sigma}{\text{Max}}(W_{\sigma j} - W_{\sigma' j})}.$$

Specializing this to two station systems using $\sigma = \sigma_1$ and $\sigma' = \sigma_2$, and noting that the denominator above is equal to $(d(\theta_{Anti-HW}) - c(\theta_{Anti-HW}))$ under the assumed sufficient condition, proves the result. □

**Example 4**

Consider the system of Figure 5.

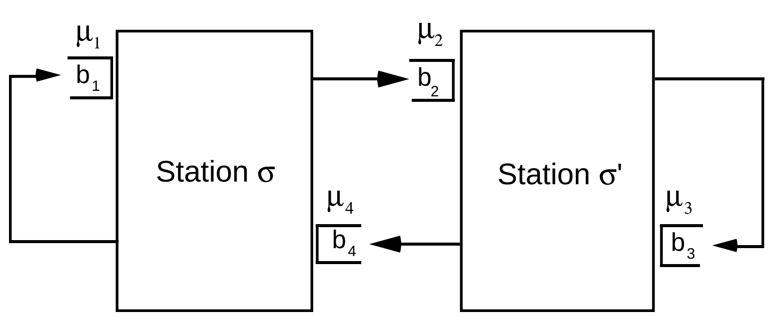


Figure 5: Closed inefficient system of Example 4.

Suppose

$$\frac{1}{\mu_2} + \frac{1}{\mu_4} > 1. \tag{82}$$

In Kumar and Seidman (1990) it was shown that in the open version of this system with deterministic service times, even when the capacity conditions $\frac{1}{\mu_1} + \frac{1}{\mu_4} < 1$ and $\frac{1}{\mu_2} + \frac{1}{\mu_3} < 1$ hold, the throughput was shown to be limited to $[\frac{1}{\mu_2} + \frac{1}{\mu_4}]^{-1}$ for all clearing policies. The same system with the longest processing time buffer priority policy $\theta_{LPT}$, which gives priority to buffers $b_2$ and $b_4$, was proved to be unstable in Lu and Kumar (1991) Recently, Harrison and Nguyen (1995) have proved that for the closed Markovian version shown in Figure 5, the policy $\theta_{LPT}$ is inefficient.

Let $\mu_F \geq \mu_S$, and suppose that $\mu_1 = \mu_3 =: \mu_F$ and $\mu_2 = \mu_4 =: \mu_S$. Then

$$\begin{aligned} \operatorname*{Min}_{i \in \sigma}(W_{\sigma i} - W_{\ sigma' i}) &= \operatorname{Min}\left\{0, \frac{1}{\mu_S}\right\} = 0, \\ \operatorname*{Max}_{i \in \ sigma'}(W_{\sigma i} - W_{\ sigma' i}) &= \operatorname{Max}\left\{\frac{-1}{\mu_F}, \frac{1}{\mu_S} - \frac{1}{\mu_F}\right\} = \frac{1}{\mu_S} - \frac{1}{\mu_F} \geq 0. \end{aligned}$$

Hence (81) does *not* hold, and so it is possible that some non-idling policy is inefficient. A natural candidate to examine is $\theta_{Anti-HW}$. This gives preference at $\sigma$ to $b_4$, and at $\ sigma'$ at $b_2$, which is the same policy as $\theta_{LPT}$. Consider the normalization $\frac{1}{\mu_F} + \frac{1}{\mu_S} = 1$, and thus $\alpha^* = 4$ from (7). As $\mu_F \to +\infty$, from numerical computation of the LPs, we find that

$$\lim_{N \to \infty} V\underline{T}^{\theta_{Anti-HW}}(N) = V\underline{T}^{\theta_{Anti-HW}}(\infty) = 2.$$

This lower bound is in consonance with the value $[\frac{1}{\mu_2} + \frac{1}{\mu_4}]^{-1}$ for the throughput in the open case. It has been shown to be precisely equal to the throughput in Harrison and N-guyen (1995). □

# 13 Concluding remarks

Many interesting issues arise. It would be useful to show that, as in Example 1, $\lim_{N \to \infty} V\underline{T}(N) = V\underline{T}(\infty)$, and that nothing is therefore lost by studying only $V\underline{T}(\infty)$.

This is related to the issue of whether the set of solutions to the dual is compact, see Murty (1983). It would also be useful if one could investigate and exploit the phenomenon of state space collapse in heavy traffic.

From our analysis it appears that the missing condition in the heavy traffic work of HW is (81). We conjecture that if this additional condition is imposed, then the HW conjectures are all valid, and that there is in fact a heavy traffic limit as proposed by them, at least for balanced Markovian systems. We are also led naturally to the conjecture that whenever a closed balanced two station system admits an inefficient non-idling policy, then the Anti-HW-policy is inefficient.

Finally, we note that functional bound results can also be obtained for closed systems which are *not* irreducible; see Ginsberg and Kumar (1996) The novel issue that then arises is the competition for service between multiple routes, a phenomenon of interest in communication networks with window based protocols such as TCP/IP, where different origin-destination pairs share the same communication network.

**Acknowledgments** The authors thank all those involved in the reviewing process for their efforts.

# References

[1] J. Walrand. *Communication Networks–A first course.* Aksen Associates, Boston, 1991.

[2] L. M. Wein. Scheduling semiconductor wafer fabrication. *IEEE Transactions on Semiconductor Manufacturing*, 1(3):115–130, August 1988.

[3] Stephen Lavenberg, Editor. *Computer Performance Modeling Handbook.* Academic Press, New York, 1983.

[4] S. Kumar and P. R. Kumar. Performance bounds for queueing networks and scheduling policies. *IEEE Transactions on Automatic Control*, AC-39:1600–1611, August 1994.

[5] J. M. Harrison and V. Nguyen. Some badly behaved closed queueing networks. In Frank P. Kelly and Ruth Williams, editors, *Stochastic Networks*, volume 71, pages 117–124. Springer–Verlag, New York, NY, 1995.

[6] P. R. Kumar and T. I. Seidman. Dynamic instabilities and stabilization methods in distributed real-time scheduling of manufacturing systems. *IEEE Transactions on Automatic Control*, AC-35(3):289–298, March 1990.

[7] S. H. Lu and P. R. Kumar. Distributed scheduling based on due dates and buffer priorities. *IEEE Transactions on Automatic Control*, AC-36(12):1406–1416, December 1991.

[8] J. M. Harrison and L. M. Wein. Scheduling networks of queues: Heavy traffic analysis of a two-station closed network. *Operations Research*, 38(6):1052–1064, 1990.

[9] P. B. Chevalier and L. M. Wein. Scheduling networks of queues: Heavy traffic analysis of a multi–station closed network. *Operations Research*, 41(4):743–757, July–August 1993.

[10] D. Blackwell. Discrete dynamic programming. *Ann. of Math. Statist.*, 33:719–726, 1962.

[11] S. Lippman. Applying a new device in the optimization of exponential queueing systems. *Operations Research*, 23:687–710, 1975.

[12] K. G. Murty. *Linear Programming*. John Wiley and Sons, New York, NY, 1983.

[13] C. Humes, Jr., J. Ou, and P. R. Kumar. The delay of open Markovian queueing networks: Uniform functional bounds, heavy traffic pole multiplicities, and stability. Submitted to Mathematics of Operations Research, July 1995.

[14] M. Ginsberg and P. R. Kumar. Throughput bounds for closed networks with competing routes. In preparation. 1996.