Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

Closed Reentrant Queueing Networks Under Affine Index Policies: Throughput Bounds, Examples and Asymptotic Loss

James R. Morrison * **Joshua P. Juen** **

* *Department of Industrial Engineering, KAIST, Daejeon 305-701, Republic of Korea (Tel: 82-42-869-3127; e-mail: james.morrison@kaist.edu).*
** *Department of Engineering and Technology, Central Michigan University, Mount Pleasant, MI 48859 USA (e-mail: juen1jp@cmich.edu).*

Abstract: We extend linear programming performance evaluation methods to closed reentrant queueing networks. The approach automatically generates the parameters for a surrogate of the differential cost function and enables us to obtain bounds on the system throughput at reduced computational cost than exact solution methodologies. A comparison study of the bounds with the actual performance for tractable examples is conducted. The results show that the bounds can be quite good, in particular for unbalanced networks. For the closed version of a well known unstable network, we investigate the performance of the bounds and explore the asymptotic loss of the system.

1. INTRODUCTION

Closed queueing networks can be used to model many modern systems including manufacturing systems with a fixed number of transportation units (such as lot transport in an automated semiconductor wafer fabrication facility), internet communication protocols, the UNIX operating system and manufacturing under CONWIP (constant work in progress) release policies. However, outside of a limited class of queueing networks possessing the so-called product form equilibrium probability distribution, few networks can be analyzed explicitly.

The performance evaluation and optimization of queueing networks is complicated by the curse of dimensionality. As the size of the problem increases, the computational requirements for optimization grow exponentially (see Papadimitriou and Tsitsiklis [1999]). To address the difficulties associated with analysis, performance bound techniques such as those of Kumar and Kumar [1994] and Bertsimas, et al. [1994] were developed. Approximate dynamic programming methods via the construction of surrogates for the differential cost function (de Farias and Van Roy [2003]) strive to overcome the complexity. Further work to extend and refine these approaches has been conducted, see, for example, Morrison and Kumar [1999], Morrison and Kumar [2002] and Veatch [2005].

Work along the lines of those above and directed to closed reentrant queueing networks has been conducted in Jin et al. [1997] and Morrison and Kumar [2001]. There, buffer priority policies and the class of all nonidling polices were studied. Employing the results of Morrison and Kumar [2002], we extend these previous approaches to closed reentrant queueing networks operating under *affine index policies* with throughput as the cost function of interest. Our results demonstrate that the bounds can be surprisingly tight or quite loose. For comparison purposes in tractable closed networks, we also explicitly solve for the performance.

In addition, for an example closed network (with an unstable mode) we investigate the rate at which the throughput converges to its limiting value. It is shown that this rate depends upon the number of bottlenecks in the network and that the concept of asymptotic loss (see Jin et al. [1997]) should be generalized.

The paper is organized as follows. In Section 2, we describe closed reentrant queueing networks and recall results relevant to our subsequent development. In Section 3, the performance bounds for closed queueing networks under affine index policies are obtained. The character of the bounds is studied in Section 4. Section 5 investigates the asymptotic loss of the closed analog of an unstable network. Concluding remarks are presented in Section 6.

2. SYSTEM DESCRIPTION

Closed reentrant queueing networks consist of E stations and L buffers. Customers await service in the buffers, labeled $b_1, \ldots, b_L$, and receive service from the stations, labeled $\sigma_1, \ldots, \sigma_E$. Only one customer may receive service from a station at a given instant and the service is preempt-resume, that is, a customer may be interrupted during service and return to service with no service time lost. The service time for a customer in buffer b_i is exponentially distributed with mean μ_i. All service times are independent. Each station caters to a distinct subset of the buffers; let $\sigma(i)$ denote the server that provides service to customers in buffer b_i.

The feature that gives such a network the closed moniker is the restriction that there are N trapped customers who

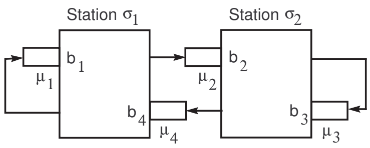

Fig. 1. A closed reentrant queueing network

circulate within the network. No customers exit and none enter the system. Such a restriction is a way to model systems where each time a customer exits, another is released into the system (or if there are a fixed number of palates upon which customers travel). The network is termed reentrant as we assume that the customers are routed from one buffer to the next in deterministic fashion following the receipt of service. A customer after receiving service from station $\sigma(i)$ while in buffer b_i next moves to buffer b_{i+1}, unless $i = L$ in which case the customer returns to buffer b_1. Further, assume that all processes are right-continuous with left-limits. Such a network is depicted in Fig. 1.

Let $x_i(t)$ denote the number of customers in buffer b_i (including any receiving service from that buffer) at time t and let $x(t) := (x_1(t), \ldots, x_L(t))^T$ be the vector of buffer lengths. Define $w_i(t) = 1$ if a customer from buffer b_i is receiving service at time t and $w_i(t) = 0$, otherwise. Let $w(t) := (w_1(t), \ldots, w_L(t))^T$. A station may only provide service to one customer at a time so that $\sum_{\{i:\sigma(i)=\sigma\}} w_i(t) \leq 1$, for all $\sigma \in \{\sigma_1, \ldots, \sigma_E\}$. Let us further assume that the control policy dictating $w(t)$ is dependent only upon the customer locations $x(t)$.

Employing uniformization, that is, sampling time at the instants τ_n when either a virtual or real service completion occurs (see, Lippman [1975]), $x(n) := x(\tau_n)$ is a discrete-time time-homogeneous Markov chain with finite state space. Let $p_{x,y}$ denote the one-step transition probability from state x to state y. The chain is aperiodic since it can remain in any given state for an arbitrary length of time (by firing only virtual transitions). When considering steady-state performance, we can suppress the dependence upon n and simply denote the state as $x \in \mathcal{Z}_+^L$. We thus hereafter write $w(x)$ to denote the control action in state x. We let $\mathcal{S} \subseteq \mathcal{Z}_+^L$ denote the state space. Including possibly transient states $\mathcal{S} = \{x \in \mathcal{Z}_+^L : e^T x = N\}$, where $e = (1, \ldots, 1)^T$ is the L-length vector of ones.

To assess the performance of the system, it behooves us to choose a performance criterion. The common performance objective for a closed queueing network is throughput (which one expects to increase with the number of trapped customers N). Let us define the throughput under a scheduling policy u (prescribing the control actions $w(x)$) with initial condition $x(0)$ as

$$\alpha_N^u := \liminf_{T\to+\infty} \frac{D_j(T)}{T},$$

where $D_j(T)$ is the number of customers to have received service while in buffer b_j in the time interval $[0, T)$. More precisely, as the policy may depend upon N, one must specify a sequence of scheduling policies and initial conditions. In steady state, the rate at which customers exit each buffer must be the same so that it does not matter which $D_j(T)$ we investigate – they all yield the same throughput. As such, we define the following cost function $c(x)$ on the state space

$$c(x) := \frac{1}{L}\sum_{i=1}^{L} \mu_i w_i(x). \tag{1}$$

The cost function merely calculates the total rate at which customers depart from all buffers and takes the arithmetic average. The maximum achievable throughput for a closed reentrant queueing network α^* is given as

$$\alpha^* = \min_{\sigma}\left\{\frac{1}{\sum_{\{j:\sigma(j)=\sigma\}} \frac{1}{\mu_j}}\right\}.$$

This value is dictated by the station (or stations) which require the longest to work on a customer and we have $\alpha_N^u \leq \alpha^*$, for all policies u and populations N. A station is termed a *bottleneck* if it achieves the minimum in the definition of α^*. A network is termed balanced if all stations are bottlenecks.

While the fact that the state space is finite ensures that one could find an equilibrium probability distribution (and hence measures of performance), the computational complexity of this activity grows dramatically with L or N. Further, one cannot analytically deduce the limiting infinite population throughput by this approach. Hence, we turn our attention to performance bounds and restrict attention to affine index policies as defined in Morrison and Kumar [2002].

Definition: Affine Index Policies. Assign to each buffer b_i an index η_i such that $\eta_i := k_i + \sum_j m_j^i x_j$, where k_i and m_j^i are given constants. The buffer b_i with highest index η_i, from among those non-empty buffers at a station, is given preempt-resume service at that station. Ties may be broken arbitrarily.

The subsequent lemma is well known and provides the framework for the bounds.

Lemma 1: Average cost inequality bounds. Consider a Markov chain with countable state space $\mathcal{S}$ as above, deterministic initial condition $x(0)$ and cost function $c(x)$ such that $E|c(x(k))| < +\infty$ for every k. If the function $W : \mathcal{S} \to \mathcal{R}$ and the constant $J \in \mathcal{R}$ satisfy

$$J + W(x) \leq c(x) + \sum_{y\in S} p_{x,y} W(y),\ \forall x \in \mathcal{S}, \tag{2}$$

and $\lim_{N\to\infty}[EW(x(N))/N] \leq 0$ then

$$J \leq \liminf_{N\to\infty} \frac{1}{N}\sum_{k=0}^{N-1} E[c(x(k))]. \tag{3}$$

Similarly for an upper bound.

The following lemma has been extracted from Morrison and Kumar [2002] and will serve in our extension of the bounds to our system (it is a key element employed in the proof of their Theorem 2.2).

Lemma 2: Constraints on Polyhedra. Consider a non-empty polyhedron P in $\mathcal{R}_+^L$ given as $P = \{x : Ax \geq b, x \geq 0\}$. Let h denote the number of linear constraints characterizing the polyhedron (not including the nonnegativity constraints). The real numbers J and c and the vector $r \in \mathcal{R}^L$ satisfy the inequality

$$J \le c + r^T x, \text{ for all } x \in P$$

if and only if there is a vector $y \in \mathcal{R}_+^h$ such that

$$A^T y \le r \text{ componentwise, and } b^T y \ge J - c.$$

These lemmas are employed in the subsequent section.

3. PERFORMANCE BOUNDS

Here we extend the work of Morrison and Kumar [2002] to allow for the form of our cost function. Following their development, we first partition the state space into points contained in polyhedra characterized by the absence or presence of customers in the buffers. Let $\Phi = \{0,1\}^L$ denote the set of L-length vectors with each element either a 0 or a 1. Let $\phi = (\phi_1, \dots, \phi_L)^T \in \Phi$ denote such a vector. Let $\mathcal{S}^\phi$ denote the subset of the state space $\mathcal{S}$ such that $x \in \mathcal{S}^\phi$ if and only if $x_i = 0$ for $\phi_i = 0$ and $x_i \ge 1$ for $\phi_i = 1$. Clearly, $S^\phi \subseteq \mathcal{S} \subseteq \mathcal{Z}_+^L$ and $\cup_\phi \mathcal{S}^\phi = \mathcal{S}$.

We further partition each $\mathcal{S}^\phi$ into points contained in polyhedra essentially characterized by which of the non-empty buffers are receiving service. Let Ω denote the set of all E-length vectors $\omega = (\omega_1, \dots, \omega_E)^T$ with values $\omega_j \in \{0\} \cup \{i : \sigma(i) = \sigma_j\}$. We interpret the vector ω as a list of the buffers receiving service at each station, with idle stations indicated by $\omega_j = 0$. (Note that our non-idling control policy ensures that if a station has at least one customer in any of its constituent buffers, it will provide service.) Let $\psi = (\phi, \omega)$ be a composite index and $\mathcal{S}^\psi = \mathcal{S}^{(\phi,\omega)}$ denote the set of points

$$\begin{aligned}\mathcal{S}^\psi := \{x \in \mathcal{S}^\phi : \eta_{\omega_j}(x) \ge \eta_\ell(x), \forall \ell \in \sigma_j \text{ with } \phi_\ell = 1,\\ \forall j \text{ with } \omega_j \neq 0\}.\end{aligned}$$

Let Ψ denote the set of ψ for which $\mathcal{S}^\psi$ is non-empty. Note that because we did not specify how ties were to be broken, the $\mathcal{S}^\psi$ may not be disjoint (some states may be in common on the boundary of such regions), however, $\cup_\psi \mathcal{S}^\psi = \mathcal{S}$. It is straightforward to demonstrate the following lemma by noting that the affine index policy imposes linear constraints on the buffer levels when particular buffers receive service.

Lemma 3: The states in $\mathcal{S}^\psi$ are contained in a polyhedron. For $\psi \in \Psi$, the states in $\mathcal{S}^\psi$ are contained in a polyhedron, call it P^ψ. The polyhedron is given as $P^\psi = \{x \in \mathcal{R}_+^L : A^\psi x \ge b^\psi\}$ where A^ψ and b^ψ characterize the linear constraints

$$\begin{aligned}
x_i &\ge 1 \text{ if } \phi_i = 1,\\
x_i &= 0 \text{ if } \phi_i = 0,\\
\sum_{i=1}^{L} (m_i^{\omega_j} - m_i^\ell) x_i &\ge k_\ell - k_{\omega_j},\\
&\qquad \forall \ell \in \sigma_j \text{ with } \phi_\ell = 1,\\
&\qquad \forall j \text{ with } \omega_j \neq 0,\\
e^T x &= N.
\end{aligned}$$

Recall that e is the L-length vector of ones, so that the last is the restriction that the customer population is fixed. Further, $\mathcal{S} \subset \cup_\psi \mathcal{S}^\psi$.

Hereafter we extend the development of Morrison and Kumar [2002] to enable bounds for our cost function. The distinction is that the cost function is not linear in the state x, but rather is essentially a constant on the polyhedra of Lemma 3 (also, the polyhedra are restricted to a simplex by the population constraint). Let the set $I(\omega)$ denote the set of the values of the nonzero elements of ω, that is, the set of working buffers.

Lemma 4: The average cost inequality on polyhedra characterized by the working buffers. Consider a simple quadratic surrogate for the differential cost function, that is let $W(x) = p^T x + (1/2) x^T Q x$, and recall the form of our cost function (1). For $x \in \mathcal{S}^\psi$, excepting possibly those x for which a tie in the affine index occurs, the average cost inequality of (2) has the form

$$J \le d^\psi + {r^\psi}^T x,$$

where

$$\begin{aligned}
d^\psi &:= \sum_{i \in I(\omega)} \left\{ \mu_i [p^T + \frac{1}{2}(e_{i+1} - e_i)^T Q](e_{i+1} - e_i) + \frac{\mu_i}{L} \right\}\\
r^\psi &:= \sum_{i \in I(\omega)} \mu_i Q (e_{i+1} - e_i).
\end{aligned}$$

Here, e_i is the L-length vector of zeros with a 1 in the i-th position. e_{L+1} is to be interpreted as e_1.

We thus arrive at the following theorem for throughput bounds in a closed reentrant queueing network under an affine index policy.

Theorem 1: Linear programming performance bound. Consider a closed reentrant queueing network under an affine index policy with deterministic initial condition $x(0)$ and the cost function as in (1). Let (α, p, Q, and vectors y^ψ) denote the decision variables in the linear program $\underline{T}$ given as

$$\text{Max } \alpha$$

subject to

$$\begin{aligned}
{A^\psi}^T y^\psi &\le r^\psi,\\
{b^\psi}^T y^\psi &\ge \alpha - d^\psi,\\
y^\psi &\ge 0,
\end{aligned}$$

for all $\psi \in \Psi$. The value of the linear program $V\underline{T}$ is a lower bound on the throughput as in (3). Similarly for an upper bound.

Proof: Ensuring that the average cost inequality holds in each $\mathcal{S}^\psi$ will ensure that it holds on $x \in \mathcal{S}$, since $\mathcal{S} \subset \cup_\psi \mathcal{S}^\psi$. (Note that the polyhedra P^ψ are subsets of $\mathcal{R}_+^L$ rather than $\mathcal{Z}_+^L$, so that we will thus ensure (2) on more than just the points $x \in \mathcal{S}$.) Since the average cost inequality has a bilinear form by Lemma 4, we can use Lemma 2 to reduce it to a linear program as stated in the Theorem. The only concern is that Lemma 4 excludes states for which a tie in the affine index may occur. This is accounted for since, by considering all $\psi \in \Psi$, we are

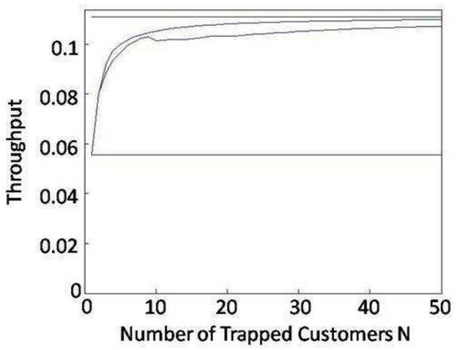

Fig. 2. Performance bounds for the network of Example 2.

in fact requiring the average cost inequality to hold for all possible ways in which ties may be settled (so long as they are stationary). □

4. PERFORMANCE BOUND EXAMPLES

In this section we consider several networks and control policies to serve as examples upon which to test our bounds. We find that the bounds are of varying quality. In the throughput examples, the upper horizontal line depicts the maximum possible throughput α^* (defined in Section 2). The lower horizontal line depicts the least possible throughput (obtained when $N = 1$ with the same value $\alpha_1^u = (\sum_i (1/\mu_i))^{-1}$ for any non-idling policy u).

Example 1: A single-station network. In this case, there is always a customer available to work on and so for any customer population N and non-idling policy u, $\alpha_N^u = \alpha^*$. For single-station example networks studied, the upper and lower bounds of Theorem 1 coincide (except at a few values of N where numerics may be to blame). □

Example 2: A two-station network under an affine index policy. Consider the network of Figure 1. Let $\mu_1 = 1/3$, $\mu_2 = 1/6$, $\mu_3 = 1/3$ and $\mu_4 = 1/6$. This network is balanced with $\alpha^* = 1/9$. Consider the affine index policy $\eta_1(x) = 10 + 3x_1$, $\eta_2(x) = 5 + x_2$, $\eta_3(x) = 10 + 3x_3$ and $\eta_4(x) = 5 + x_4$. The performance bounds are depicted in Figure 2. Recall that the upper horizontal line depicts α^*, while the lower horizontal line depicts the least possible throughput (for $N = 1$). The bounds lie between these lines. Note that they are fairly tight and that the lower bound is non-monotonic. □

Example 3: The same two-station network under the LBFS policy. Consider the balanced network of Example 2, with affine index policy $\eta_1(x) = 4$, $\eta_2(x) = 2$, $\eta_3(x) = 3$ and $\eta_4(x) = 1$. This policy is a closed variant of the static priority last buffer first served (LBFS) policy. This policy imposes the restriction that buffers b_1 and b_3 can never have more than one customer in steady state. The performance bounds are depicted in Figure 3 and are to be interpreted as in Example 2. The bounds are very tight. Interestingly, the upper bounds for Examples 2 and 3 are almost identical, though the lower bounds for Example 3 are superior. Thus, though this does not prove superiority of the LBFS policy in this case, it is guaranteed to have performance in a much tighter region. Note also that since this policy is a static priority policy, the bounds are the same as those of previous work for that class. □

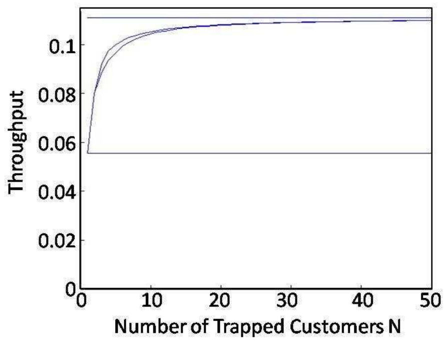

Fig. 3. Performance bounds for the network of Example 3.

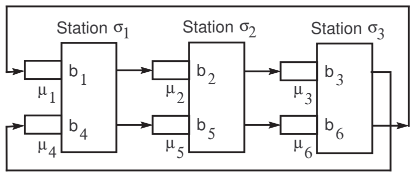

Fig. 4. The three-station network of Example 4.

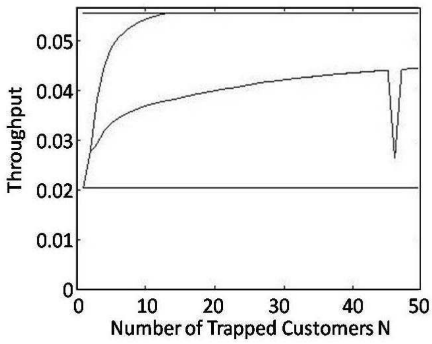

Fig. 5. Performance bounds for the network of Example 4.

Example 4: A three-station network under an affine index policy. Consider the network of Figure 4. Let $\mu_1 = 1/12$, $\mu_2 = 1/12$, $\mu_3 = 1/12$, $\mu_4 = 1/6$, $\mu_5 = 1/4$ and $\mu_6 = 1/3$ so that $\alpha^* = 1/18$. Station σ_1 is the bottleneck (but it is not very dominant). Consider the affine index policy $\eta_1(x) = 14x_1$, $\eta_2(x) = 7x_2$, $\eta_3(x) = 2.5x_3$, $\eta_4(x) = x_4$, $\eta_5(x) = 3.5x_5$ and $\eta_6(x) = 1.5x_4$. The performance bounds are depicted in Figure 5. The quality of the bounds is poor. Note the dip in the lower bound (we assume this is due to numerical issues). □

Example 5: A heavily unbalanced three-station network under an affine index policy. Consider the network and control policy of Example 4. Let $\mu_1 = 1/14$, $\mu_2 = 1/7$, $\mu_3 = 2/7$, $\mu_4 = 1/14$, $\mu_5 = 1/7$ and $\mu_6 = 2/7$ so that $\alpha^* = 1/28$. Station σ_1 is the very dominant bottleneck. The performance bounds are depicted in Figure 6. The quality of the bounds is much better, however there are a few spikes that we attribute to numerical issues. □

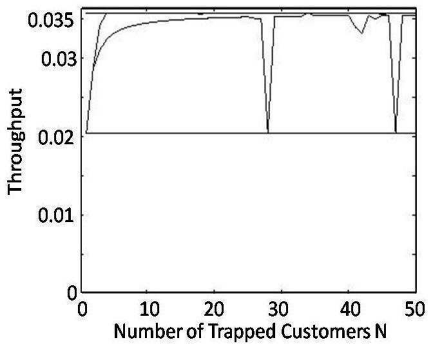

Fig. 6. Performance bounds for the network of Example 5.

5. PERFORMANCE EVALUATION OF THE CLOSED ANALOG OF AN UNSTABLE NETWORK

We next consider the performance evaluation of the network depicted in Figure 1 operating under the static priority (buffer priority) policy with $\eta_1(x) = 1$, $\eta_2(x) = 3$, $\eta_3(x) = 2$ and $\eta_4(x) = 4$. Under this policy, the network may not achieve the bottleneck throughput, since after some finite time both b_2 and b_4 cannot have customers simultaneously. As a consequence, buffers b_2 and b_4 cannot be in service simultaneously. For this reason, these two buffers behave as if they are served by a single *virtual station*. The throughput is thus restricted as

$$\alpha_N^u \le \min\left\{\frac{1}{\frac{1}{\mu_1}+\frac{1}{\mu_4}}, \frac{1}{\frac{1}{\mu_2}+\frac{1}{\mu_3}}, \frac{1}{\frac{1}{\mu_2}+\frac{1}{\mu_4}}\right\}.$$

Results leading to the conclusion above were obtained in Dai and Vande Vate [2000] and Dai et al. [2004] for open reentrant queueing networks and subsequently in Morrison and Kumar [1998] for closed reentrant queueing networks.

Here we study the performance of this network for various values of the process rates. The bounds (which are equivalent to those of Morrison and Kumar [2001] since the policy is a static priority) are compared with the exact performance (calculated via the solution of the equilibrium probability equations). In addition, we numerically explore the rate at which the throughput converges to the bottleneck rate α^*. That is, we study the asymptotic loss ν, typically defined as

$$\nu := \lim_{N\to\infty} N\Big(1 - \frac{\alpha_N^u}{\alpha^*}\Big).$$

If the asymptotic loss ν is finite, then one may consider that $\alpha_N^u \approx \alpha^*(1-\nu/N)$. If the asymptotic loss $\nu = 0$, then the throughput converges at a faster rate to α^*. For a single bottleneck, we expect that $\alpha_N^u \approx \alpha^*(1-\beta^N)$, where β is a constant. If the asymptotic loss converges to infinity rather than a finite constant, then the throughput converges more slowly. We conjecture that this is possible when there are more bottlenecks than actual stations, that is, when a virtual station has the same throughput restriction as the real stations. The form for the throughput we conjecture is $\alpha_N^u = \alpha^*(1-\gamma/\sqrt{N})$. This is analogous to a form proposed in Humes et al. [1997] for open queueing networks in which the mean number of customers has the form $K/(1-\rho)^2$, where ρ is the system loading. To our knowledge, no queueing networks of the type studied here have been discovered that possess such forms. However, our numerical studies suggest just such a behavior for the system of Figure 1 with three bottlenecks (two real and one virtual).

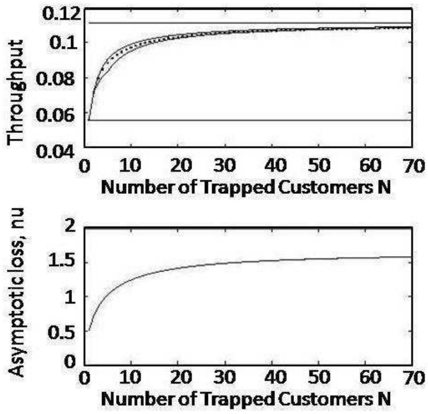

Fig. 7. Performance of the network of Example 6.

Example 6: Two real bottlenecks. Consider the network of Figure 1 with $\mu_1 = 1/6$, $\mu_2 = 1/3$, $\mu_3 = 1/6$ and $\mu_4 = 1/3$ and affine index policy stated previously. Both stations are bottlenecks and $\alpha^* = 1/9$. The throughput bounds (solid lines) and the exact performance (dots) are depicted in the top of Figure 7. The asymptotic loss (calculated for each N) is depicted in the lower half of Figure 7 and appears to converge to about the value $\nu = 1.6$. □

Example 7: One real bottleneck. Consider the network of Figure 1 with $\mu_1 = 1/6$, $\mu_2 = 1/3$, $\mu_3 = 1/3$ and $\mu_4 = 1/6$ and affine index policy stated previously. Station σ_1 is the bottleneck and $\alpha^* = 1/12$. The throughput bounds (solid lines) and the exact performance (dots) are depicted in the top of Figure 8. The bounds are not as good in this case (perhaps suggesting that the quadratic surrogate for the differential cost function could be improved by terms of another form). The asymptotic loss converges to zero, indicating (as expected) that the throughput converges faster than $\alpha_N^u \approx \alpha^*(1-\nu/N)$. Instead we consider the form $\alpha_N^u \approx \alpha^*(1-\beta^N)$ and calculate the *exponential loss* β as

$$\beta := \lim_{N\to\infty} \Big(1 - \frac{\alpha_N^u}{\alpha^*}\Big)^{1/N}.$$

The exponential loss (calculated for each N) as is depicted in the lower half of Figure 8 and appears to converge to about the value $\beta = 0.84$. □

Example 8: One virtual and two real bottlenecks. Consider the network of Figure 1 with $\mu_1 = 1/4$, $\mu_2 = 1/4$, $\mu_3 = 1/4$ and $\mu_4 = 1/4$ and affine index policy stated previously. Station σ_1, σ_2 and the pair $\{b_2, b_4\}$ are bottlenecks and $\alpha^* = 1/8$. The throughput bounds (solid lines) and the exact performance (dots) are depicted in the top of Figure 9. The bounds are not very good. The asymptotic loss appears to diverge, indicating that the throughput may converge more slowly than $\alpha_N^u \approx \alpha^*(1-\nu/N)$. Instead we consider the form $\alpha_N^u \approx \alpha^*(1-\gamma/\sqrt{N})$ and calculate the *square root loss* γ as

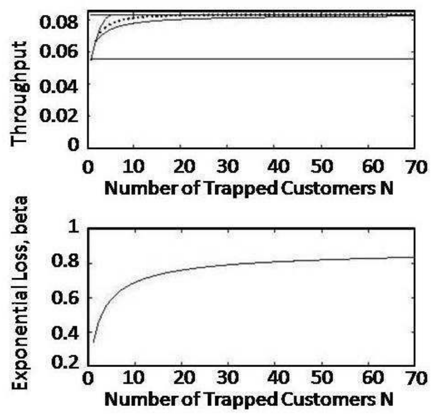

Fig. 8. Performance of the network of Example 7.

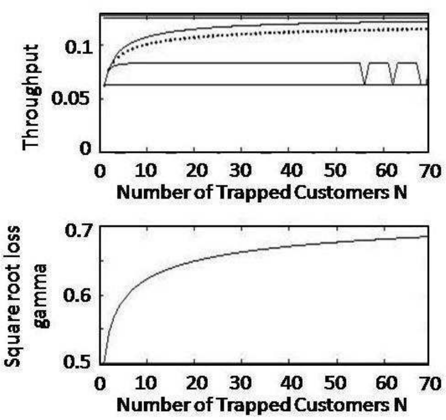

Fig. 9. Performance of the network of Example 8.

$$\gamma := \lim_{N\to\infty} \sqrt{N}\Big(1 - \frac{\alpha_N^u}{\alpha^*}\Big).$$

The square root loss (calculated for each N) is depicted in the lower half of Figure 9 and appears to converge to a value $\gamma < 0.70$. □

6. CONCLUDING REMARKS

For closed reentrant queueing networks operating under affine index policies, we have extended recent work to enable performance bounds via linear programming computation. The bounds have been shown to be quite good in some instances and quite poor in others. In addition to exploring the quality of the bounds, we have investigated the throughput performance of the closed analog of an unstable open network. It was shown via computation of the equilibrium probability distributions for fixed customer populations that the rate at which the throughput converges to the bottleneck throughput depends upon the number of bottlenecks in the system. For a single bottleneck, exponential convergence was observed. For two bottlenecks, the expected behavior $\alpha_N^u \approx \alpha^*(1-\nu/N)$ was observed. Finally for three bottlenecks in a two station network (one additional virtual bottleneck), the throughput appeared to have the form $\alpha_N^u \approx \alpha^*(1-\nu/\sqrt{N})$.

Further work could be conducted in several directions. First, the surrogate for the differential cost function could be improved. That is, is there a better prototype for the closed network than quadratic (or piecewise quadratic)? Can an exponential loss or square root loss formulation be developed to bound the values of β and γ along the lines of previous work for the traditional asymptotic loss?

REFERENCES

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

D. Bertsimas, I. C. Paschalidis and J. N. Tsitsiklis. Optimization of multiclass queueing networks: Polyhedral and nonlinear characterizations of achievable performance. *Annals of Applied Probability*, 4:43–75, 1994.

D. P. de Farias and B. Van Roy. The linear programming approach to approximate dynamic programming. *Operations Research*, 51:850–865, 2003.

C. H. Papadimitriou and J. N. Tsitsiklis. The complexity of optimal queueing network control. *Mathematics of Operations Research*, 24:293–305, 1999.

J. R. Morrison and P. R. Kumar. New linear program performance bounds for queueing networks. *Journal of Optimization Theory and Applications*, 100:575–597, 1999.

M. H. Veatch. Approximate dynamic programming for networks: Fluid models and constraint reduction. Submitted for publication, 2005. Available at http://faculty.gordon.edu/ns/mc/mike_veatch/

H. Jin, J. Ou and P. R. Kumar. The throughput of irreducible closed Markovian queueing networks: Functional bounds, asymptotic loss, efficiency, and the Harrison-Wein conjectures. *Mathematics of Operations Research*, 22:886–920, 1997.

J. R. Morrison and P. R. Kumar. New linear program performance bounds for closed queueing networks. *Discrete Event Dynamic Systems*, 11:291–317, 2001.

James R. Morrison and P. R. Kumar. Linear programming performance bounds for Markov chains with polyhedrally translation invariant transition probabilities and applications to unreliable manufacturing systems and enhanced wafer fab models. *Proceedings of the ASME International Mechanical Engineering Congress and Exposition (IMECE)*, IMECE2002-39274, New Orleans, LA, November 2002.

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

J. G. Dai and J. H. Vande Vate. The Stability of Two-Station Multi-Type Fluid Networks. *Operations Research*, 48:721–744, 2000.

J. G. Dai, J. J. Hasenbein and J. H. Vande Vate. Stability and instability of a two-station queueing network. *Annals of Applied Probability*, 14:326–377, 2004.

J. R. Morrison and P. R. Kumar. On the guaranteed throughput and efficiency of closed reentrant lines. *Queueing Systems Theory and Applications*, 28:33–54, 1998.

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. *Mathematics of Operations Research*, 22:921–954, 1997.