

A semi-product-form for the equilibrium state probabilities in a pair of queues with finite batches

P.G. Harrison
Imperial College London
pgh@doc.ic.ac.uk

ABSTRACT

A Markovian network of two queues, with finite size batch Poisson arrivals and departures, is solved approximately, but to arbitrary accuracy, for its equilibrium state probabilities. Below a pair of thresholds on the queue lengths, a modification of the Spectral Expansion Method is used to construct a semi-product-form at all lengths of one queue in a finite lattice strip defined by the threshold of the other queue. No additional special arrival streams are required, for example at empty queues, from which it is already known that a product-form can be constructed. Hence the first exact closed form solution for the equilibrium probabilities in an unmodified Markovian queueing network with batches is obtained, the only constraint being finiteness of the batches. The method is illustrated numerically, first in a tandem network and then in a two-node network with feedback. Simulation results confirm the precision of the method, lying almost on top of the those of the model in plots. Partial batch forwarding and discarding are thereby compared quantitatively.

ACM Reference Format:

P.G. Harrison. 2019. A semi-product-form for the equilibrium state probabilities in a pair of queues with finite batches. In *12th EAI International Conference on Performance Evaluation Methodologies and Tools (VALUE-TOOLS 2019)*, March 12–15, 2019, Palma, Spain. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3306309.3306316>

1 INTRODUCTION

Queueing networks with batch movements are appropriate for modelling the kind of burstiness that has been observed in internet traffic for some years, more recently in cloud networks, which has a degrading effect on network performance. Such models may also be used to provide quantitative analysis of algorithms and schedules that reduce energy consumption by switching off devices when they are not in use and switched on again when they are next required. This switching inherently increases burstiness, not only in power consumption but also in the performance delivered – typically, but not solely, measured by device utilisation, throughput and the response time experienced by users of the devices. Stochastic performance analysis of queues and networks of queues with batches is therefore an important means by which to optimize such systems to find the right balance between energy-saving and performance

according to particular management goals, user requirements and government regulations.

In the next section, the finite batch-queue and batch networks are defined and the basic results on geometric batch-queues and product-form batch networks are summarized, noting the special extra arrival streams to empty queues that are necessary to ensure these special properties [1, 2, 6]. We then obtain, in Section 3, a solution for a single batch-queue without the extra arrival stream and show that the same geometric component dominates the solution at high queue lengths. In Section 4, we conjecture that in a network of two batch queues, the product-form obtained by adding special arrivals to empty queues would similarly dominate when both queue lengths are large enough. An approximation to the equilibrium state-probabilities in a two-node network then follows, using a modification of the Spectral Expansion Method of [7] on two finite lattice strips defined by a threshold on each queue, above which the model is close to product-form. Suitable numerical values for the thresholds are obtained iteratively. Numerics that show an almost perfect match with simulation are provided in Section 5. A quantitative comparison of two variants of the batch transmission protocol between the queues is also given, where partial batches are either forwarded or discarded. Section 6 concludes and considers future potential of the approach.

2 SINGLE BATCH QUEUE

2.1 Definitions and basic results

Our model of batch movements in a single server queue is defined, in line with [5], as follows:

- The state space $\mathcal{S}$ of the queue is the set of non-negative integers;
- Batch arrivals, in this paper of *bounded size* k , $1 \leq k \leq n_a$, are represented by transitions with constant instantaneous rate $a_k : i \rightarrow i + k$ ($i \geq 0$), i.e. from states i to $i + k$;
- *Normal* batch departures, here also of bounded size k , $1 \leq k \leq n_d$, are represented by transitions with constant instantaneous rate $d_k : i + k \rightarrow i$ ($i \geq 0$);
- *Special* batch departures of size k , leading to an empty queue, are represented by transitions with constant instantaneous rate $d_{0k} = \sum_{\ell=k+1}^{n_d} d_\ell : k \rightarrow 0$;
- The ordering of individual tasks in the queue is first come first served (FCFS).

We call this a *batch-queue*. Each batch movement has a corresponding *rate generating function*, defined by $A(z) = \sum_{k=1}^{\infty} a_k z^k$ and $D(z) = \sum_{k=1}^{\infty} d_k z^k$. Thus, $A(1)$ is the total *batch* arrival rate and

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

VALUETOOLS 2019, March 12–15, 2019, Palma, Spain

© 2019 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-6596-3/19/03.

<https://doi.org/10.1145/3306309.3306316>

$A'(1) = \left. \frac{dA}{dz} \right|_{z=1}$ is the total *task* arrival rate in all states; and similarly for $D(z)$ at all positive queue lengths, in view of the definition of the rates of the special departures.

This queue models the effects of *buffering*: tasks are first grouped together into batches prior to input to the queue; then the server independently groups tasks into new batches which are processed and then output. Such queues are used in an “assembly-transfer network” in Chapter 8 of [1], which is appropriate for models of certain manufacturing systems. The interpretation is that batches of size less than or equal to the queue length are “full batches” and the others are “partial batches” (with a size equal to the queue length n which is less than the intended batch size k), which are discarded in the assembly line; these depart at rates d_{0k} ($1 \leq k \leq n_d - 1$).

An additional special arrival stream, in state 0 only, was introduced in [5] with rates chosen to make the equilibrium probability distribution of the length of the queue geometric, whereby product-forms are facilitated in networks by application of the Reversed Compound Agent Theorem (RCAT) [4]. The maximum batch size in the special arrival stream turns out to be $n_a - 1$; hence no special arrivals are needed in the standard $M/M/1$ queue, which already has a geometric distribution at equilibrium. The rate generating function of the special arrivals is determined entirely by $A(z)$ as $A_0(z) = \frac{\rho z A(\rho^{-1}) - A(z)}{1 - \rho z}$. It can be shown that a geometric, equilibrium probability distribution exists, with parameter ρ , if and only if

$$A(\rho^{-1}) + D(\rho) = A(1) + D(1) \quad (1)$$

provided that $A'(1) < D'(1)$ – a stability condition asserting that the total task arrival rate must be less than the total task departure rate at large queue lengths. In this case, it can be shown that there is only one solution to Equation (1).

In this paper networks of two batch-queues are considered, operating in two modes with respect to partial batches. In the first, *discarding* mode, partial batches exit the network whether or not they are destined for the other queue. In the second, non-discarding or *forwarding* mode, partial batches are transmitted to their destination queue. In both cases, the source queue is then left empty.

3 TAIL PROBABILITIES IN THE RAW BATCH QUEUE

We call the batch queue with no special arrivals and the special departures defined above a *raw batch queue*. At equilibrium, when this exists, let the equilibrium probability that the queue has length j be denoted by π_j . The raw batch queue then has balance equations, for $j \geq 1$:

$$(A(1) + D(1))\pi_j = \sum_{i=1}^j a_i \pi_{j-i} + \sum_{i=1}^{n_d} d_i \pi_{j+i} \quad (2)$$

where we assume $a_i = 0$ if $i > n_a$ and $d_i = 0$ if $i > n_d$. Multiplying both sides by z^j and summing from $j = 1$ to ∞ yields, after changing the order of summation in the first term on the right and simplification:

$$[A(1) + D(1) - A(z) - D(z^{-1})](\Pi(z) - \pi_0) = \pi_0 A(z) - \sum_{i=1}^{n_d} \sum_{j=1}^i d_i \pi_j z^{j-i}$$

where the probability generating function $\Pi(z) = \sum_{i=0}^{\infty} \pi_i z^i$. This may be written as:

$$[A(1) + D(1) - A(z) - D(z^{-1})]\Pi(z) = [A(1) + D(1)]\pi_0 - \sum_{i=1}^{n_d} \sum_{j=0}^i d_i \pi_j z^{j-i} \quad (3)$$

For $\Pi(z)$ to be analytic in the unit z -disk, we require the right hand side to be zero whenever the expression $A(1) + D(1) - A(z) - D(z^{-1})$ is zero. We know from Equation (1) that it is zero at $z = 1/\rho$ and we can also see that $z = 1$ is another root. Moreover, we have the following.

PROPOSITION 3.1. *For a stable batch queue with rate parameters defined as above, the expression $A(1) + D(1) - A(x^{-1}) - D(x)$ has roots located in the complex plane as follows:*

- $n_a - 1$ roots lying strictly inside the disk centred at the origin with radius ρ ;
- $n_d - 1$ roots lying strictly outside the unit disk centred at the origin;
- The two roots $x = \rho, 1$.

PROOF. Since $x = 0$ is not a root, the expression in question has the same roots as the polynomial

$$x^{n_a} (A(1) + D(1) - A(x^{-1}) - D(x)) \equiv (A(1) + D(1))x^{n_a} - d_{n_d} x^{n_a+n_d} - \dots - d_1 x^{n_a+1} - a_1 x^{n_a-1} - \dots - a_{n_a}$$

It is shown in [5] that there is only the one real root ρ in the interval $(0, 1)$. Moreover, the gradient of the expression $A(1) + D(1) - A(x^{-1}) - D(x)$ is negative at $x = 1$ in a stable queue, in which $A'(1) < D'(1)$. Therefore the gradient is positive at $x = \rho$. Hence the coefficient of x^{n_a} , namely $A(1) + D(1)$, is greater than the absolute sum of all the other coefficients at $x = \rho^+$ and at $x = 1^-$. There are therefore no roots in the annulus $\rho < |x| < 1$ and the proposition now follows by Rouché’s Theorem. $\square$

We can now return to Equation (3) to find the dominant term in a series expansion solution for the generating function $\Pi(z)$, so establishing the tail behaviour of the equilibrium probabilities.

PROPOSITION 3.2. *A stable batch queue with rate parameters defined as above has geometric equilibrium queue length tail-probabilities with parameter ρ .*

PROOF. Let the roots of the equation $A(1) + D(1) - A(x^{-1}) - D(x) = 0$ be enumerated in ascending order of absolute value $x_1, x_2, \dots, x_{n_a-1}, \rho, 1, x_{n_a+2}, \dots, x_{n_a+n_d}$, according to Proposition 3.1. We assume here that the roots are distinct; the analysis is routinely modified if this is not the case.

Now, the multiplier $[A(1) + D(1) - A(z) - D(z^{-1})]$ on the left hand side of Equation (3) is zero at points $z = x_i^{-1}$ for $i = n_a + n_d, \dots, n_a + 2$ inside the unit z -disk, at $z = 1$, at $z = \rho^{-1}$ and at $z = x_i^{-1}$ for $i = n_a - 1, \dots, 1$. Because the generating function $\Pi(z)$ must be analytic inside the unit z -disk, and also finite at $z = 1$, the right hand side of Equation (3) must vanish when $z = 1, x_{n_a+n_d}^{-1}, \dots, x_{n_a+2}^{-1}$, giving n_d equations in the $n_d + 1$ unknowns $\pi_0, \dots, \pi_{n_d}$ on the right hand side, so that they may each be expressed as a constant multiple of the single unknown variable π_0 (say).

We may write the expression $A(1) + D(1) - A(z) - D(z^{-1})$ as $d_{n_d} z^{-n_d} \prod_{i=1}^{n_a+n_d} (1 - x_i z)$, where $x_{n_a} = \rho$ and $x_{n_a+1} = 1$, the reciprocal of which can be expanded in partial fractions as

$$\frac{z^{n_d}}{d_{n_d}} \prod_{i=n_a+1}^{n_a+n_d} (1 - x_i z)^{-1} \sum_{i=1}^{n_a} \frac{\alpha_i}{1 - x_i z}$$

for some constants α_i . By analyticity of $\Pi(z)$, the product term cancels when multiplied with the right hand side of Equation (3) and we obtain a solution of the form:

$$\Pi(z) = C\pi_0 \sum_{j=0}^{\infty} \rho^j z^j \sum_{i=1}^{n_a} \alpha_i (x_i/\rho)^j$$

so that $\pi_j = C\pi_0 \rho^j \sum_{i=1}^{n_a} \alpha_i (x_i/\rho)^j$ for $j = 0, 1, \dots$. But $x_1 < x_2 < \dots < x_{n_a} = \rho$ and so as $j \rightarrow \infty$, $\pi_j \rightarrow C\pi_0 \alpha_{n_a} \rho^j$. Finally, normalization provides the final equation that determines π_0 and we have the required geometric tail $\pi_j \propto \rho^j$ as $j \rightarrow \infty$. $\square$

4 TWO-NODE NETWORK OF BATCH QUEUES

Consider a network of two *non-discard* queues, where partial batches are transmitted between the nodes rather than departing the network and there are no special arrivals; this model is considered in the asymptotic case in [3]. In this model, the global balance equations are, for *any pair* (j, k) , $j \geq 0, k \geq 0$ (the symbols $\vee$ and $\wedge$ are abbreviations for the maximum and minimum functions):

$$\begin{aligned} \pi_{jk} \left(\sum_{i=1}^{n_{a1}} a_{1i} + \sum_{i=1}^{n_{a2}} a_{2i} + (1 - \delta_{j0}) \sum_{i=1}^{n_{d1}} d_{1i} + (1 - \delta_{k0}) \sum_{i=1}^{n_{d2}} d_{2i} \right) = \\ \sum_{i=1}^{j \wedge n_{a1}} \pi_{j-i, k} a_{1i} + \sum_{i=1}^{k \wedge n_{a2}} \pi_{j, k-i} a_{2i} + \\ p_{12} \sum_{i=1}^{k \wedge n_{d1}} \pi_{j+i, k-i} d_{1i} + (1 - p_{12}) \sum_{i=1}^{n_{d1}} \pi_{j+i, k} d_{1i} + \\ p_{21} \sum_{i=1}^{j \wedge n_{d2}} \pi_{j-i, k+i} d_{2i} + (1 - p_{21}) \sum_{i=1}^{n_{d2}} \pi_{j, k+i} d_{2i} + \\ \delta_{j0} p_{12} \sum_{i=1}^{k \wedge (n_{d1}-1)} \pi_{i, k-i} \sum_{\ell=i+1}^{n_{d1}} d_{1\ell} + \\ \delta_{j0} (1 - p_{12}) \sum_{i=1}^{n_{d1}-1} \pi_{ik} \sum_{\ell=i+1}^{n_{d1}} d_{1\ell} + \\ \delta_{k0} p_{21} \sum_{i=1}^{j \wedge (n_{d2}-1)} \pi_{j-i, i} \sum_{\ell=i+1}^{n_{d2}} d_{2\ell} + \\ \delta_{k0} (1 - p_{21}) \sum_{i=1}^{n_{d2}-1} \pi_{ji} \sum_{\ell=i+1}^{n_{d2}} d_{2\ell} \end{aligned} \quad (4)$$

In the case of two discard queues, the equations simplify a little in that the last four lines on the right hand side combine as

$$\delta_{j0} \sum_{i=1}^{n_{d1}-1} \pi_{ik} \sum_{\ell=i+1}^{n_{d1}} d_{1\ell} + \delta_{k0} \sum_{i=1}^{n_{d2}-1} \pi_{ji} \sum_{\ell=i+1}^{n_{d2}} d_{2\ell}.$$

Now, there can be no special external arrivals coming into (or going out of) a state (j, k) with $j \geq n_{a1} \vee n_{d2}, k \geq n_{a2} \vee n_{d1}$, since the maximum batch sizes in the two special arrival streams at node 1 are $n_{a1} - 1$ and $n_{d2} - 1$; $n_{a2} - 1$ and $n_{d1} - 1$ at node 2. Similarly, no partial batches can be generated in states (j, k) with $j \geq n_{d1}, k \geq n_{d2}$. Hence the balance equations are the same, parameterized by j, k , at all states in the set $\{(j, k) \mid j \geq n_{a1} \vee n_{d2} \vee n_{d1}, k \geq n_{a2} \vee n_{d1} \vee n_{d2}\}$ in both the discard and non-discard models, whether or not there are special batch arrivals.

In line with Section 3, we make the following conjecture:

CONJECTURE 4.1. *When equilibrium exists, the global balance equations of the three models (discard queue with special arrivals, discard queue without special arrivals and non-discard queue) admit a solution that converges to the known (unnormalized) product-form solution of the first of these models at all states (j, k) with $j \geq J$ and $k \geq K$ for large enough $J \geq n_{a1} \vee n_{d2} \vee n_{d1}, K \geq n_{a2} \vee n_{d1} \vee n_{d2}$.*

The rationale is that asymptotically the equilibrium tail distribution will be a product of geometric functions, in line with Proposition 3.1. In Section 4.3 we define a stopping condition for finding suitable values for J, K , but ultimately the numerical result is validated by the fact that it approximates a normalized solution of the balance equations of the underlying Markov process, which must be unique.

Therefore $\pi_{jk} \simeq \pi_{J, K} \rho_1^{j-J} \rho_2^{k-K}$ for $j > J, k > K$ in all of the three models, where ρ_1, ρ_2 are given by the solution to the equations

$$\begin{aligned} A_1(\rho_1^{-1}) + p_{21} D_2(\rho_2 \rho_1^{-1}) + D_1(\rho_1) &= A_1(1) + p_{21} D_2(\rho_2) + D_1(1) \\ A_2(\rho_2^{-1}) + p_{12} D_1(\rho_1 \rho_2^{-1}) + D_2(\rho_2) &= A_2(1) + p_{12} D_1(\rho_1) + D_2(1) \end{aligned} \quad (5)$$

according to [5].

The probabilities π_{jk} for $0 \leq j \leq J, k \geq K$ (green region in Figure 2) and $j \geq J, 0 \leq k \leq K$ (blue region in Figure 2) are determined by generalizing the Spectral Expansion Method of [7], and the remaining probabilities π_{jk} for $0 \leq j < J, 0 \leq k < K$ are found by direct solution of the global balance equations (4).

4.1 Recap on the Spectral Expansion Method

Consider a 2-dimensional Markov process, finite in one dimension, infinite in the other, as depicted in Figure 1. Such a model has typically been used to model queues with Markov modulated Poisson arrivals and exponential service times; see for example [7]. The phase of the modulating process, assumed to vary over a finite set, $\{j \in \mathbb{Z} \mid 0 \leq j \leq J\}$ say, defines the finite dimension and the queue length the infinite dimension. It is assumed that there exists a *threshold*, K say, above which the equilibrium balance equations take the same form in terms of j and k at each level $k \geq K$ of the infinite dimension; i.e. the probability fluxes going out of and coming into state (j, k) have the same expressions in j and k for all $k \geq K, 0 \leq j \leq J$.

Defining the vector $\vec{v}_k = (\pi_{0k}, \pi_{1k}, \dots, \pi_{Jk})$ for $k \geq K$, these balance equation can be rewritten in matrix-vector form

$$\sum_{i=-I_1}^{I_2} \vec{v}_{k-i} Q_i = \vec{0}$$

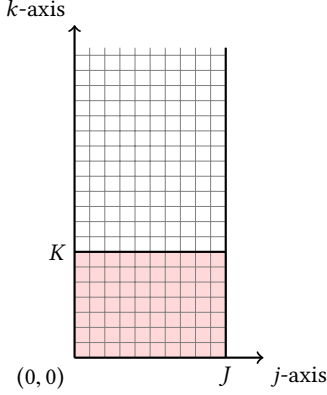


Figure 1: Semi-infinite lattice strip of states and state transition probabilities used in the Spectral Expansion Method. In the red shaded area below the threshold, any transitions are allowed. Above the threshold the balance equations are the same at all levels k .

for certain $J \times J$ matrices Q_i and positive integers I_1, I_2 . The matrix Q_i defines the transition rates between each pair of j -states that yield a change of i in k , i.e. $k \rightarrow k + i$, for $-I_1 \leq i \leq I_2$. Of course, the method relies on the magnitude of such changes in k being bounded, as in queues for example [7].

These balance equations have as a solution $\vec{v}_k = x^k \vec{e}_x$ where x is any solution of the equation

$$\det \left(\sum_{i=-I_1}^{I_2} x^{-i} Q_i \right) = 0 \quad (6)$$

and $\vec{e}_x$ is the corresponding (generalized) eigenvector such that $\vec{e}_x \sum_{i=-I_1}^{I_2} x^{-i} Q_i = \vec{0}$. It can be shown that the general solution of the balance equations above the threshold is then, for $k \geq K$,

$$\vec{v}_k = \sum_{i=1}^{\gamma} c_i x_i^k \vec{e}_i$$

where $\{x_i \mid 1 \leq i \leq \gamma\}$ are all of the roots of Equation (6) with absolute value less than 1 and positive real part, $\vec{e}_i$ is the respective (generalized) left-eigenvector of x_i and c_i is a constant ($1 \leq i \leq \gamma$). The value of γ is such that the number of resulting equations for the equilibrium probabilities up to the threshold K , i.e. $(J+1)(K+1)$, is the same as the number of unknowns, i.e. $(J+1)K + \gamma$, so that $\gamma = J+1$; see [7] for details.

4.2 Modified Spectral Expansion Method in the k - and j -strips.

Consider the strip $\{(j, k) \mid 0 \leq j \leq J, k \geq K + n_{d1}\}$, called the k -strip and shaded green in Figure 2. In the discard model, the rhs

of the global balance equation for state (j, k) simplifies to:

$$\begin{aligned} & \sum_{i=1}^{j \wedge n_{a1}} \pi_{j-i, k} a_{1i} + \sum_{i=1}^{n_{a2}} \pi_{j, k-i} a_{2i} + \\ & p_{12} \sum_{i=1}^{n_{d1}} \pi_{j+i, k-i} d_{1i} + (1 - p_{12}) \sum_{i=1}^{n_{d1}} \pi_{j+i, k} d_{1i} + \\ & p_{21} \sum_{i=1}^{j \wedge n_{d2}} \pi_{j-i, k+i} d_{2i} + (1 - p_{21}) \sum_{i=1}^{n_{d2}} \pi_{j, k+i} d_{2i} + \\ & \delta_{j0} \sum_{i=1}^{n_{d1}-1} \pi_{ik} \sum_{\ell=i+1}^{n_{d1}} d_{1\ell} \end{aligned}$$

with the obvious modification to get the non-discard model with partial batch forwarding. In the third and fourth terms of this expression, the first subscript $j+i$ of π is not restricted to the range $[0, J]$. However, by the above reasoning we can replace any term π_{mn} by $\rho_1^{m-J} \pi_{J, n}$ when $m \geq J$ and $n \geq K$. Consequently, since $k \geq K + n_{d1}$, the right hand side becomes

$$\begin{aligned} & \sum_{i=1}^{j \wedge n_{a1}} \pi_{j-i, k} a_{1i} + \sum_{i=1}^{n_{a2}} \pi_{j, k-i} a_{2i} + \\ & p_{12} \left(\sum_{i=1}^{n_{d1} \wedge J-j} \pi_{j+i, k-i} d_{1i} + \sum_{i=j-j+1}^{n_{d1}} \pi_{J, k-i} \rho_1^{i+j-J} d_{1i} \right) + \\ & (1 - p_{12}) \left(\sum_{i=1}^{n_{d1} \wedge J-j} \pi_{j+i, k} d_{1i} + \sum_{i=j-j+1}^{n_{d1}} \pi_{J, k} \rho_1^{i+j-J} d_{1i} \right) + \\ & p_{21} \sum_{i=1}^{j \wedge n_{d2}} \pi_{j-i, k+i} d_{2i} + (1 - p_{21}) \sum_{i=1}^{n_{d2}} \pi_{j, k+i} d_{2i} + \\ & \delta_{j0} \sum_{i=1}^{n_{d1}-1} \pi_{ik} \sum_{\ell=i+1}^{n_{d1}} d_{1\ell} \quad (7) \end{aligned}$$

Following the Spectral Expansion Method for this strip, we construct matrices, $D^{(1)}(i)$, $p^{(1)}$ and $U^{(1)}(i)$, of transition rates in the first dimension (corresponding to the first subscript j in π_{jk}) that produce increments of $-i$, 0 and $+i$ in the second dimension k , for finite $i > 0$ (because of the assumption of finite batch sizes), respectively. By inspection of each term of the right hand side of the balance equations (7), for $0 \leq h, j \leq J$, we find

$$\begin{aligned} D^{(1)}(i)_{hj} &= p_{21} d_{2, i} \delta_{h, j-i} I_{1 \leq i \leq j \wedge n_{d2}} + (1 - p_{21}) d_{2, i} \delta_{h, j} I_{1 \leq i \leq n_{d2}}; \\ p^{(1)}_{hj} &= a_{1, j-h} I_{1 \leq j-h \leq n_{a1}} + \delta_{j, 0} I_{1 \leq h \leq n_{d1}-1} \sum_{\ell=h+1}^{n_{d1}} d_{1\ell} + \\ & (1 - p_{12}) \left(d_{1, h-j} I_{1 \leq h-j \leq n_{d1} \wedge J-j} + \delta_{h, J} \sum_{i=j-j+1}^{n_{d1}} \rho_1^{i+j-J} d_{1i} \right); \\ U^{(1)}(i)_{hj} &= a_{2i} \delta_{h, j} I_{1 \leq i \leq n_{a2}} + p_{12} \times \\ & \left(d_{1, i} \delta_{h, i+j} I_{1 \leq i \leq n_{d1} \wedge J-j} + \rho_1^{i+j-J} d_{1, i} \delta_{h, J} I_{J-j+1 \leq i \leq n_{d1}} \right). \end{aligned}$$

I is the indicator function, equal to 1 if its subscript is true and 0 if not, and δ is the Kronecker-delta. The corresponding expressions

for the non-discard variant are:

$$\begin{aligned}
D^{(1)}(i)_{hj} &= p_{21}d_{2,i}\delta_{h,j-i}I_{1 \leq i \leq J \wedge n_{d_2}} + (1-p_{21})d_{2,i}\delta_{h,j}I_{1 \leq i \leq n_{d_2}}; \\
P_{hj}^{(1)} &= a_{1,j-h}I_{1 \leq j-h \leq n_{d_1}} + (1-p_{12}) \times \\
&\quad \left(d_{1,h-j}I_{1 \leq h-j \leq n_{d_1} \wedge J-j} + \delta_{h,J} \sum_{i=J-j+1}^{n_{d_1}} \rho_1^{i+j-J} d_{1i} \right. \\
&\quad \left. + \delta_{j,0}I_{1 \leq h \leq n_{d_1}-1} \sum_{\ell=h+1}^{n_{d_1}} d_{1\ell} \right); \\
U^{(1)}(i)_{hj} &= a_{2i}\delta_{h,j}I_{1 \leq i \leq n_{d_2}} + p_{12} \times \\
&\quad \left(d_{1,i}\delta_{h,i+j}I_{1 \leq i \leq n_{d_1} \wedge J-j} + \rho_1^{i+j-J} d_{1,i}\delta_{h,J}I_{J-j+1 \leq i \leq n_{d_1}} \right. \\
&\quad \left. + \delta_{j,0}\delta_{h,i}I_{1 \leq i \leq n_{d_1}-1} \sum_{\ell=i+1}^{n_{d_1}} d_{1\ell} \right).
\end{aligned}$$

We can now use the Spectral Expansion Method described in section 4.1 to solve for the equilibrium probabilities in the k -strip. Let $\vec{v}_k^{(1)} = (\pi_{0k}, \pi_{1k}, \dots, \pi_{Jk})$ be the vector of equilibrium probabilities at level $k \geq K$ and $\Delta^{(1)}$ be the diagonal matrix of total transition rates out of states (j, k) for $j = 0, 1, \dots, J$ at any $k \geq K$, viz. $A_1(1) + A_2(1) + D_2(1) + (1 - \delta_{j,0})D_1(1)$. Then we have:

PROPOSITION 4.2. *The solution for the equilibrium probabilities (when they exist) in the k -strip is, asymptotically,*

$$\vec{v}_k^{(1)} = \sum_{i=1}^{J+1} c_i^{(1)} \vec{e}_i^{(1)} x_i^k \quad (8)$$

where $\{x_i \mid 1 \leq i \leq J+1\}$ are the roots x , with absolute value less than 1 and positive real part, of the equation $\det Q^{(1)}(x) = 0$, in which the matrix

$$Q^{(1)}(x) = (P^{(1)} - \Delta^{(1)}) + \sum_{i=1}^{n_{d_2} \vee n_{d_1}} U^{(1)}(i)x^{-i} + \sum_{i=1}^{n_{d_2}} D^{(1)}(i)x^i;$$

$\vec{e}_i^{(1)}$ is the (generalized) left-eigenvector corresponding to x_i ; and $c_1^{(1)}, c_2^{(1)}, \dots, c_{J+1}^{(1)}$ are constants, as yet unknown.

PROOF. The balance equations on the k -strip can be recast as:

$$\vec{v}_k^{(1)}(P^{(1)} - \Delta^{(1)}) + \sum_{i=1}^{n_{d_2} \vee n_{d_1}} \vec{v}_{k-i}^{(1)} U^{(1)}(i) + \sum_{i=1}^{n_{d_2}} \vec{v}_{k+i}^{(1)} D^{(1)}(i) = 0$$

for $k \geq K$ and the Spectral Expansion Method then yields the required result. It is known that the number of roots inside the unit disk with positive real part is $J+1$. $\square$

We use this solution above the threshold $k = K$, which value is chosen empirically.

Similarly, above the threshold $j = J$ in the j -strip, the asymptotic solution we use is

$$\vec{v}_j^{(2)} = \sum_{i=1}^{K+1} c_i^{(2)} \vec{e}_i^{(2)} y_i^j \quad (9)$$

where $\vec{v}_j^{(2)} = (\pi_{j,0}, \dots, \pi_{j,K})$ and $\{y_i \mid 1 \leq i \leq K+1\}$ are the roots of the equation $\det Q^{(2)}(y) = 0$ with absolute value less than

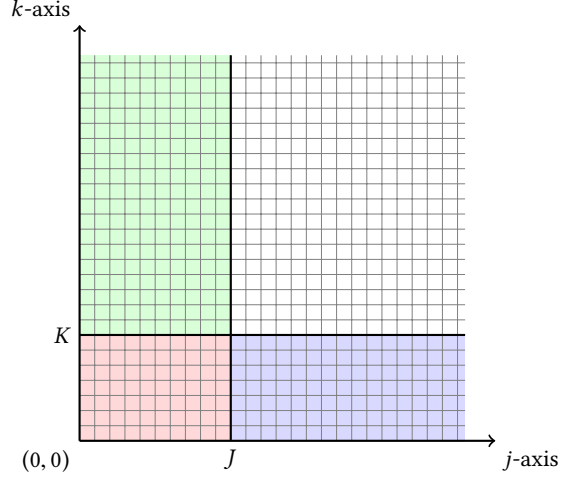


Figure 2: Pair of semi-infinite lattice strips of state transitions used in the modified Spectral Expansion Method for the network of two batch-queues. In the red shaded area, any transitions are allowed; in the green shaded area, the balance equations are the same at all levels k ; in the blue shaded area, the balance equations are the same at all levels j ; in the white area, the balance equations are the same for all states (j, k) .

1 and positive real part. The terms superscripted by a parenthesized (2) are defined symmetrically with the corresponding terms in the k -strip, parenthesized (1), by interchanging 1 with 2 and j with k , keeping i the same.

4.3 Remaining equilibrium probabilities

Referring to Figure 2, the equilibrium probabilities in the green and blue strips are derived by spectral expansion according to the previous section. These expressions include $J + K + 2$ unknown values, $c_1^{(1)}, \dots, c_{J+1}^{(1)}, c_1^{(2)}, \dots, c_{K+1}^{(2)}$. In the pink region $\{(j, k) \mid 0 \leq j \leq J, 0 \leq k \leq K\}$, we use the balance equation (4) directly – any transitions are allowed, so that we can choose the version of the global balance equations that gives a network with or without discards¹. Here there are a further JK variables since the probabilities $\{\pi_{jK} \mid 0 \leq j \leq J\}$ and $\{\pi_{Jk} \mid 0 \leq k \leq K\}$ are given by the constant vectors $\vec{c}^{(1)}$ and $\vec{c}^{(2)}$ in the spectral expansion. There are therefore $JK + J + K + 2 = (J+1)(K+1) + 1$ unknowns and $(J+1)(K+1)$ equations. These yield all the unknowns as a multiple of $\pi_{0,0}$ (say), the value of which comes from normalization.

PROPOSITION 4.3. *Let the solution of the above $(J+1)(K+1)$ equations be $\pi_{j,k} = \tilde{\pi}_{j,k} \pi_{0,0}$, $c_i^{(1)} = \tilde{c}_i^{(1)} \pi_{0,0}$ and $c_i^{(2)} = \tilde{c}_i^{(2)} \pi_{0,0}$.*

¹We may include special arrivals to give a product-form in the discard-case, if required as a check.

Then

$$\pi_{0,0} = \left(\sum_{i=1}^{J+1} \tilde{c}_i^{(1)} E_i^{(1)} x_i^K / (1 - x_i) + \sum_{i=1}^{K+1} \tilde{c}_i^{(2)} E_i^{(2)} y_i^J / (1 - y_i) + \sum_{j=0}^{J-1} \sum_{k=0}^{K-1} \tilde{\pi}_{j,k} + \tilde{\pi}_{J,K} \left(\frac{\rho_1 \rho_2}{(1 - \rho_1)(1 - \rho_2)} - 1 \right) \right)^{-1}$$

where $E_i^{(\cdot)}$ denotes the sum of the components of the vector $\tilde{e}_i^{(\cdot)}$.

PROOF. The normalization equation may be written

$$1 = \sum_{j=0}^{J-1} \sum_{k=0}^{K-1} \pi_{j,k} + \sum_{j=0}^J \sum_{k=K}^{\infty} \pi_{j,k} + \sum_{k=0}^K \sum_{j=J}^{\infty} \pi_{j,k} + \sum_{j=J+1}^{\infty} \sum_{k=K+1}^{\infty} \pi_{j,k} - \pi_{J,K}$$

by partitioning the state space according to Figure 2. The result now follows by substituting the values obtained by solving the above $(J+1)(K+1)$ equations in terms of $\pi_{0,0}$ and using Conjecture 4.1. $\square$

In addition, the point (J, K) lies in both the green and blue regions and so can be expressed in terms of either $\tilde{c}^{(1)}$ or $\tilde{c}^{(2)}$, leading to a single consistency condition. However, at large J, K we expect the equations to be consistent because of our conjecture that asymptotically the model has the correct geometric tail. We use the difference between these evaluations of $\pi_{J,K}$ in either region as a measure of the accuracy of the solution as J and K increase. In our numerical computations we increased J, K until the difference was less than 1 percent or the maximum was less than $10^{-2} \rho_1^J \rho_2^K / ((1 - \rho_1)(1 - \rho_2))$ this quantity being one percent of the total probability remaining in the geometric tail.

The above equations compute all probabilities $\pi_{j,k}$ in terms of $\pi_{0,0}$ (say), which is given by the normalizing equation. We state this as follows.

ALGORITHM 1. Assuming Conjecture 4.1, given an accuracy criterion on $\pi_{J,K}$, the equilibrium probabilities (when they exist) of the raw batch networks defined above can be obtained as follows:

- (1) Set $J = n_{a1} \vee n_{d2} \vee n_{d1}, K = n_{a2} \vee n_{d1} \vee n_{d2}$.
- (2) For $0 \leq j \leq J, k \geq K$, define

$$\tilde{v}_k^{(1)} = \sum_{i=1}^{J+1} c_i^{(1)} \tilde{e}_i^{(1)} x_i^k$$

where $\tilde{e}_i^{(1)}$, x_i are computed numerically by Proposition 4.2 for $1 \leq i \leq J+1$.

- (3) For $0 \leq k \leq K, j \geq J$, define

$$\tilde{v}_j^{(2)} = \sum_{i=1}^{K+1} c_i^{(2)} \tilde{e}_i^{(2)} y_i^j$$

similarly.

- (4) Solve the set of equations formed by $(J+1)(K+1)$ instances of equation (4) with $0 \leq j \leq J, 0 \leq k \leq K$ and $\pi_{j,k}$ substituted by $\sum_{i=1}^{J+1} c_i^{(1)} (\tilde{e}_i^{(1)})_j x_i^k$ for $k \geq K$ and by $\sum_{i=1}^{K+1} c_i^{(2)} (\tilde{e}_i^{(2)})_k y_i^j$ for $j \geq J$, to obtain values for $\tilde{\pi}_{j,k} = \pi_{j,k} / \pi_{0,0}$, $\tilde{c}_i^{(1)} =$

$c_i^{(1)} / \pi_{0,0}$ and $\tilde{c}_i^{(2)} = c_i^{(2)} / \pi_{0,0}$, as in the notation of Proposition 4.3.

- (5) Set $\tilde{\pi}_{0,0}$ to the value of $\pi_{0,0}$ obtained by Proposition 4.3.

Then

$$\pi_{j,k} = \begin{cases} \tilde{\pi}_{0,0} \tilde{\pi}_{j,k} & \text{if } 0 \leq j < J, 0 \leq k < K \\ \tilde{\pi}_{0,0} \sum_{i=1}^{J+1} \tilde{c}_i^{(1)} (\tilde{e}_i^{(1)})_j x_i^k & \text{if } 0 \leq j \leq J, k \geq K \\ \tilde{\pi}_{0,0} \sum_{i=1}^{K+1} \tilde{c}_i^{(2)} (\tilde{e}_i^{(2)})_k y_i^j & \text{if } 0 \leq k \leq K, j \geq J \\ \tilde{\pi}_{0,0} \tilde{\pi}_{J,K} \rho_1^{J-j} \rho_2^{K-k} & \text{otherwise} \end{cases}$$

If the two values computed for $\pi_{J,K}$ are not sufficiently close according to the accuracy criterion, increase J and K and repeat from step 2.

5 NUMERICS

5.1 Tandem network

We applied the method of Section 4 for the cases of both partial batch forwarding and discarding. Referring to Section 2, in the former case, the whole queue is transmitted from node 1 to node 2 when a service completion corresponding to a batch of greater size occurs. In the latter case, the partial batches leave the network. The parameters of the tandem model (with $p_{12} = 1, p_{21} = 0$) are as follows.

- Node 1: Arrival rates 2.5 and 0.6 for batch sizes 1 and 2, respectively, and service rates 1.0 and 2.0, similarly;
Node 2: Arrival rate 2.1 and service rate 7.0 for batch size 1 only (node 2 is M/M/1).

Thus, with partial batch forwarding, when the queue length at node 1 is 1, partial batches of size 1 are transmitted to node 2 at rate 2.0, so that the total single task transmission rate between the nodes becomes 3.0. With discards, the partial batches are lost. These parameters give geometric ratios of 0.824 and 0.842 at nodes 1 and 2 respectively in the product-form model (with special arrivals, here only at node 1).

5.1.1 *Comparison with simulation.* We first compared the probability mass functions computed by the model at each queue for the two scenarios (batch-discarding and batch-forwarding) against the output of two regenerative simulations with over a quarter of a million regenerative cycles, with regeneration state $\{0, 0\}$. Since the curves produced are pretty much coincident, we did not include the confidence bands, which are tight but easily include the model's points. The comparisons can be seen in Figure 3 below.

We see the exceptionally close match between our model and the simulation in all cases, ensured by our choice of J and K .

5.1.2 *Comparison of discarding with forwarding.* The impact of discarding and forwarding on the two queues is shown in Figure 4.

We note that there is virtually no difference between discarding and forwarding on the first queue. This is unsurprising since there is no feedback in this tandem network so that it matters not to node 1 whether its output goes to node 2 or leaves the network.

At the second queue, however, the normally well-behaved M/M/1 queue becomes disrupted at low queue lengths because of the arrivals of partial size-2 batches. With discarding, these are missing altogether when the first queue has fewer than 2 tasks, whereas with forwarding they arrive as single tasks. Hence the load is a

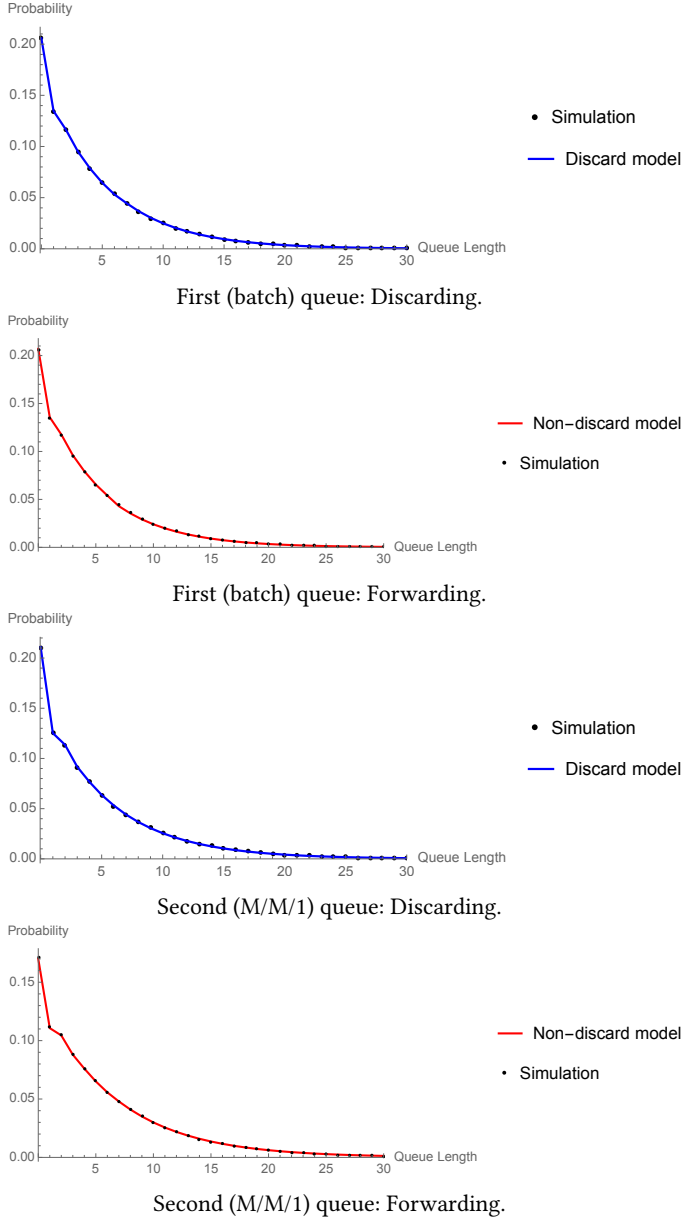


Figure 3: Probability mass functions at the two queues in the tandem network with partial batch discarding and forwarding: comparison with simulation.

little less with discarding, giving the higher probabilities observed at queue lengths up to 5 at node 2.

5.2 General two-node network

To illustrate the use of the model in more general networks, we chose a pair of fully interconnected batch queues with the following parameters.

Node 1: Arrival rates 1.0, 0.6 and 0.3 for batch sizes 1, 2 and 3, respectively, and service rates 1.0, 2.0 and 0.8, similarly;

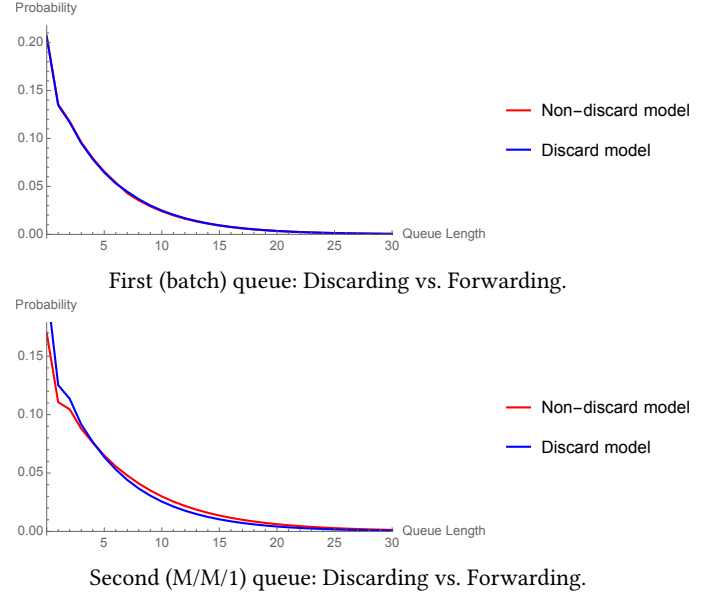


Figure 4: Probability mass functions computed by the model at the two queues in the tandem network: comparison of partial batch discarding and forwarding.

Node 2: Arrival rate 2.0 and 0.2 for batch sizes 1 and 2, respectively, and service rates 2.0, 3.0 and 1.0 for respective batch sizes, 1, 2 and 3.

When special arrivals are added, the server utilizations become $\rho_1 = 0.900$ and $\rho_2 = 0.592$, respectively. We again applied the method of Section 4 for the cases of both partial batch forwarding and discarding.

5.2.1 Comparison with simulation. Proceeding as described for the tandem network above, we obtained the graph shown in Figure 5 comparing the model's predictions against regenerative simulation. Again, agreement is almost perfect for our choice of J and K and confidence bands are not shown.

5.2.2 Comparison of discarding with forwarding. The impact of discarding and forwarding on the two queues in this more general case is shown in Figure 6.

Here, because of the feedback from node 2 to node 1, the difference between discarding and forwarding is significant in both queues. This is particularly noticeable in the first queue since the influx of partial batches from the second node is relatively high, giving a much lower zero-queue probability of about 0.08 compared to 0.2 with discarding. The same applies to queue lengths up to about 9. The jump at queue length 2 is also more pronounced because of the partial batch forwarding from node 2 to node 1. Complementary to this, at the second queue there is little difference, due to its much lower utilization as well as the high utilization at the first queue which leads to fewer discards, the queue length there tending to be longer than the maximum departure batch size.

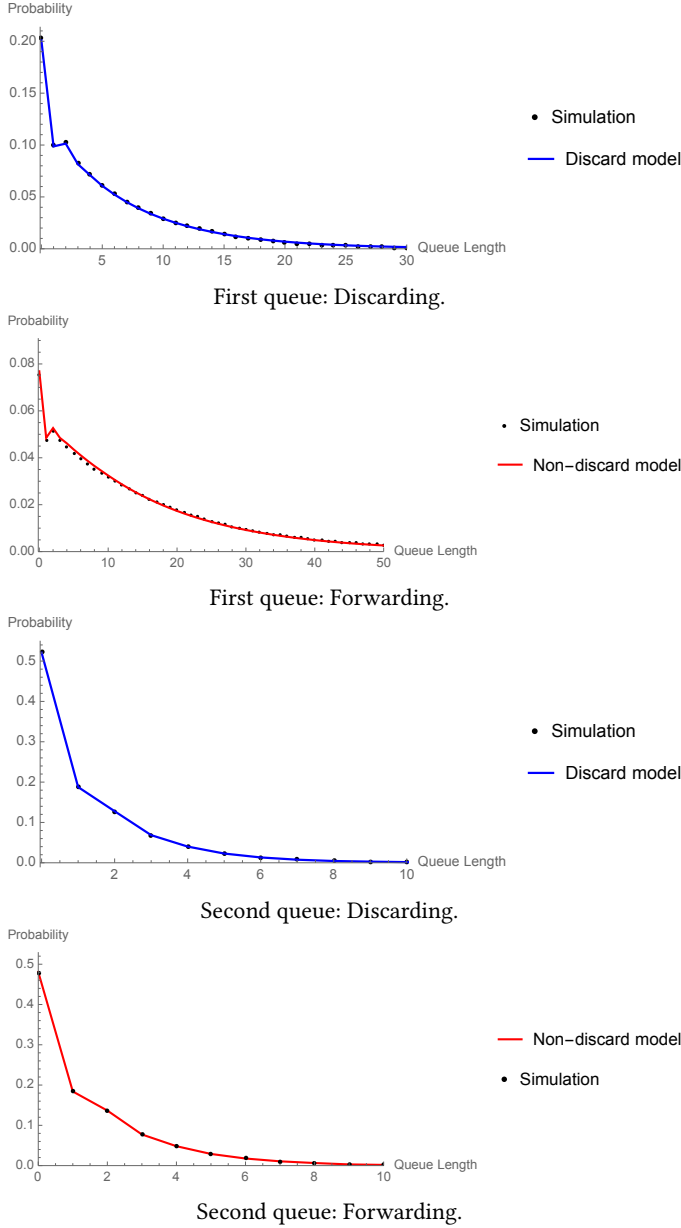


Figure 5: Probability mass functions at the two queues in the general network with partial batch discarding and forwarding: comparison with simulation.

6 CONCLUSION

The main contribution of this paper is the first (to the author's knowledge) solution for the equilibrium state probabilities in an *unmodified* network of queues with batch processing. The method combines existing techniques from product-form solutions and spectral analysis to give a near-exact result in the sense that the error can be reduced arbitrarily at increasing computational cost.

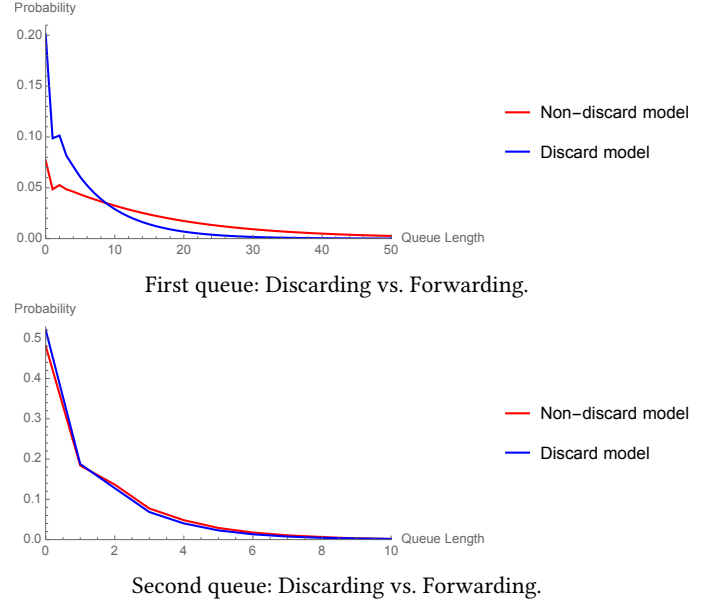


Figure 6: Probability mass functions computed by the model at the two queues in the general network: comparison of partial batch discarding and forwarding.

Numerical results confirm the accuracy, which has also been seen in several other networks not displayed.

Of course, the approach has its limitations:

- (1) Networks are restricted to two nodes with finite batch sizes;
- (2) At high utilizations, large values for the thresholds J and K may be needed, leading to a high-order polynomial equation in the spectral analysis. Methods have been developed to reduce the order by transforming the determinantal equation into a simpler problem in linear algebra with a much more efficient solution [7]. This was not the main point of the present work, but in a software tool such a transformation might be necessary.
- (3) The roots of this same equation may not be distinct, requiring careful management of the eigenvectors corresponding to multiple eigenvalues.
- (4) As J and K become large, the smallest eigenvalues become smaller, approaching zero. This exacerbates the previous problem.

As noted in the introduction, batch-networks are well suited to modelling bursty traffic that occurs in networks of various types, for example router networks and file transfers in storage systems. Moreover, in energy-efficient systems, workload has to be scheduled to introduce *bursts* into the workload, with longer idle periods enabling devices to be switched off more often [8]. These systems are well modelled by batch movements in a network. Product-form approximations have long been known for queueing networks that are modified to include additional arrivals at empty queues, as in [1] for example. Unfortunately, this makes them somewhat limited for analysing energy conservation strategies since at the very time a device is switched off it gets extra arrivals which would require it

to be switched on again! The present work may help to alleviate this problem.

REFERENCES

- [1] X. Chao, M. Miyazawa, and M. Pinedo. *Queueing networks: customers, signals and product form solutions*. Wiley, 1999.
- [2] E. Gelenbe. G-networks with triggered customer movement. *J. Appl. Prob.*, 30:742–748, 1993.
- [3] Peter G. Harrison, Richard A. Hayden, and William J. Knottenbelt. Product-forms in batch networks: Approximation and asymptotics. *Perform. Eval.*, 70(10):822–840, October 2013.
- [4] P.G. Harrison. Turning back time in Markovian process algebra. *Theoretical Computer Science*, 290(3):1947–1986, 2003.
- [5] PG Harrison. Product-form queueing networks with batches. In *European Workshop on Performance Engineering*, pages 250–264. Springer, 2018.
- [6] W. Henderson and P. Taylor. Product form in networks of queues with batch arrivals and batch services. *Queueing Systems*, 6:71–88, 1990.
- [7] I. Mitrani and R. Chakka. Spectral expansion solution for a class of Markov models: Application and comparison with the matrix-geometric method. *Performance Evaluation*, 23:241–260, 1995.
- [8] A.E. Papathanasiou and M.L. Scott. Energy efficiency through burstiness. In *5th IEEE Workshop on Mobile Computing Systems and Applications*, 2003.