A Maclaurin-series expansion approach to multiple paired queues

Eline De Cuypere, Koen De Turck, Dieter Fiems∗

Ghent University, Department of Telecommunications and Information Processing

Abstract

Motivated by kitting processes in assembly systems, we consider a Markovian queueing system with \( K \) paired finite-capacity buffers. Pairing means that departures from the buffers are synchronised and that service is interrupted if any of the buffers is empty. To cope with the inherent state-space explosion problem, we propose an approximate numerical algorithm which calculates the first \( L \) coefficients of the Maclaurin series expansion of the steady-state probability vector in \( O(KLM) \) operations, \( M \) being the size of the state space.

1. Introduction

We consider a system of \( K \) queues, each queue having finite capacity. Let \( C_i \) denote the capacity of the \( i \)th queue. Moreover, for each of the queues, customers arrive in accordance with an independent Poisson process, let \( \lambda_i > 0 \) denote the arrival rate in queue \( i \). Departures from the different queues are paired which means that there are simultaneous departures from all queues with rate \( \mu \) as long as all queues are non-empty. If one of the queues is empty, there are no departures.

The queueing system at hand is motivated by kitting processes in assembly systems. A kitting process collects the necessary parts for a given end product in a container prior to assembly. While conceptually simple, kitting comes with many advantages. Kitting clearly mitigates storage space requirements at the assembly station since no part inventories need to be kept there. Moreover, parts are placed in proper positions in the container such that assembly time reductions can be realised [3, 9]. A kitting process is obviously related to a paired queueing system: the inventories of the different parts that go into the kit correspond to the service time and kitting is blocked if the kit consists of parts from the different buffers, generally distributed interarrival times at the different buffers are synchronised, which leads to entirely different dynamics. However, in fork-join queueing systems both arrivals and departures in the different buffers are synchronised, which leads to entirely different dynamics.

2. Maclaurin-series expansion

As arrivals in the different queues are modelled by Poisson processes and the service time distribution is exponential, the state of the system is described by a vector \( i \in \mathcal{C} \) whose \( k \)th element corresponds to the queue size of the \( k \)th buffer. Here \( \mathcal{C} = \mathcal{C}_1 \times \ldots \times \mathcal{C}_K \) denotes the state space of this continuous-time Markov chain (CTMC), with \( \mathcal{C}_k = \{0, 1, \ldots, C_k\} \) being the set of possible levels of queue

∗Corresponding author (Dieter.Fiems@UGent.be), Ghent University, Department of Telecommunications and Information Processing St-Pietersnieuwstraat 41, 9000 Gent, Belgium.
where comparing the terms in the terms in
\[ \pi(i_1, i_2, \ldots, i_K) \left( \mu \prod_{\ell=1}^{K} \mathbb{1}_{\{i_\ell > 0\}} + \sum_{\ell=1}^{K} \mathbb{1}_{\{i_\ell < C_\ell\}} \lambda_\ell \right) = \]
\[ \pi(i_1 + 1, i_2 + 1, \ldots, i_K + 1) \mu \prod_{\ell=1}^{K} \mathbb{1}_{\{i_\ell < C_\ell\}} + \sum_{\ell=1}^{K} \sum_{k=0}^{\infty} \pi_0(i_\ell, i_{\ell-1}, i_{\ell-1} - 1, i_{\ell+1}, \ldots, i_K) \lambda_\ell \mathbb{1}_{\{i_\ell > 0\}}, \quad (1) \]
for all \( i = (i_1, i_2, \ldots, i_K) \in \mathcal{C} \) and where \( \mathbb{1}_x \) is the indicator function which equals one if \( x \) is true and equals zero otherwise. While the former system of equations is easily solved if there are only a few queues with low capacity, otherwise. While the former system of equations is easily solved if there are only a few queues with low capacity, the size of the state space explodes for even a moderate number of queues and reasonable queue capacities and a direct solution is computationally infeasible.

To mitigate this state space explosion problem, we rely on a Miacrion series expansion in \( \mu \). It is shown in the appendix that \( \pi(i) \) is analytic in \( \mu = 0 \) and therefore admits the representation,
\[ \pi(i) = \sum_{n=0}^{\infty} \pi_n(i) \mu^n, \]
for \( 0 \leq \mu < \mu_0 \) and for \( i \in \mathcal{C} \). Here \( \mu_0 \) is a non-negative value for which a lower bound is provided in the appendix.

Substituting the former expression in the balance equations yields,
\[ \sum_{n=0}^{\infty} \pi_n(i_1, i_2, \ldots, i_K) \mu^n \times \left( \mu \prod_{\ell=1}^{K} \mathbb{1}_{\{i_\ell > 0\}} + \sum_{\ell=1}^{K} \mathbb{1}_{\{i_\ell < C_\ell\}} \lambda_\ell \right) = \]
\[ \sum_{n=0}^{\infty} \pi_n(i_1 + 1, i_2 + 1, \ldots, i_K + 1) \mu^{n+1} \prod_{\ell=1}^{K} \mathbb{1}_{\{i_\ell < C_\ell\}} + \sum_{n=0}^{\infty} \sum_{\ell=1}^{K} \pi_0(i_\ell, i_{\ell-1}, i_{\ell-1} - 1, i_{\ell+1}, \ldots, i_K) \lambda_\ell \mu^n \mathbb{1}_{\{i_\ell > 0\}}. \]

For \( i \in \mathcal{C}^* = \mathcal{C} \setminus \{c\} \), with \( c = [C_1, C_2, \ldots, C_K] \), comparing the terms in \( \mu^0 \) on both sides of the former equation yields,
\[ \pi_0(i_1, i_2, \ldots, i_K) = 0, \]
whereas comparing the terms in \( \mu^n \) for \( n > 0 \) gives,
\[ \pi_n(i_1, i_2, \ldots, i_K) = \frac{1}{\sum_{\ell=1}^{K} \mathbb{1}_{\{i_\ell < C_\ell\}} \lambda_\ell} \times \]
\[ \left( \mathbb{1}_{\{n > 0\}} \pi_{n-1}(i_1 + 1, i_2 + 1, \ldots, i_K + 1) \prod_{\ell=1}^{K} \mathbb{1}_{\{i_\ell < C_\ell\}} \right) \]
\[ + \sum_{\ell=1}^{K} \pi_n(i_1, \ldots, i_{\ell-1}, i_\ell - 1, i_{\ell+1}, \ldots, i_K) \lambda_\ell \mathbb{1}_{\{i_\ell > 0\}} \]
\[ - \mathbb{1}_{\{n > 0\}} \pi_{n-1}(i_1, i_2, \ldots, i_K) \prod_{\ell=1}^{K} \mathbb{1}_{\{i_\ell > 0\}} \right). \quad (2) \]

For \( i = c \), such a comparison does not yield an expression for \( \pi_n(i) \). To determine the remaining unknown, we invoke the normalisation condition:
\[ \sum_{i \in \mathcal{C}} \pi_0(i) = 1, \quad \sum_{i \in \mathcal{C}} \pi_n(i) = 0. \]

Solving for \( \pi_n(c) \) then yields,
\[ \pi_0(c) = 1, \quad \pi_n(c) = - \sum_{i \in \mathcal{C}} \pi_n(i) \]

Once the series expansion of the steady state distribution has been obtained, the expansions of various performance measures directly follow. Let \( X \sim \pi \), then for a performance measure \( J = E[f(X)] \) we have,
\[ J = \sum_{i \in \mathcal{C}} f(i) \pi(i) = \sum_{i \in \mathcal{C}} \sum_{n=0}^{\infty} f(i) \pi_n(i) \mu^n = \sum_{n=0}^{\infty} \sum_{i \in \mathcal{C}} f(i) \pi_n(i) \mu^n = \sum_{n=0}^{\infty} J_n \mu^n, \quad (3) \]
for \( 0 \leq \mu < \mu_0 \) with,
\[ J_n = \sum_{i \in \mathcal{C}} f(i) \pi_n(i). \]

The interchange of the sumations is justified by the finiteness of \( \mathcal{C} \) and the convergence of \( \sum_{n=0}^{\infty} \pi_n(i) \mu^n \) for all \( i \in \mathcal{C} \). As such, any term \( J_n \) in the expansion of a performance measure \( J \) can be calculated from the corresponding vector \( \pi_n \) of the expansion of the steady-state vector. Performance measures of interest include amongst others the 6th order moment of the queue content of the \( k \)th queue \( f(i) = f(i_k) \), the blocking probability \( f(i) = 1 - \prod_{j=1}^{K} \mathbb{1}_{\{i_j > 0\}} \) and the throughput \( f(i) = \mu \prod_{j=1}^{K} \mathbb{1}_{\{i_j > 0\}} \).

**Computational complexity.** From (2), calculation of \( \pi_n(i) \) takes at most \( K + 2 \) additions and one division (assuming the rate sums are known). Hence, the computational complexity of calculating \( \pi_n \) is \( O(KM) \), with \( M = |\mathcal{C}| \) the size of the state space. Having the same complexity for every additional term in the expansion, calculating the first \( L \) coefficients then has complexity \( O(KML) \).

As the size of the state space is very large, limited memory consumption is equally important. To limit memory consumption to the size of storing only one steady-state vector one can proceed as follows. Assuming one is mainly interested in the expansion of a number of performance measures, note that once the \( r \)th term of the expansion of the steady state vector is determined, the corresponding
terms in the expansions of various performance measures can be determined as well; see (3). Hence, there is no need to keep track of previous terms of the expansion of steady-state probabilities unless they are required for further calculations of coefficients of steady state probabilities. From (2) one sees that \( \pi_n(i) \) is expressed in terms of \( \pi_{n-1}(j) \), with \( j \) larger then \( i \) (lexicographically). This means that the coefficients of the vector \( \pi_{n-1} \) can be overwritten progressively during the calculation of \( \pi_n \) and memory for only one vector of size \( M \) is needed.

3. Numerical results

To illustrate our series expansion approach, we now assess its accuracy by means of some numerical examples. First, consider a system with \( K = 5 \) paired queues, each queue having capacity \( C = 10 \). Moreover, the arrival intensity at each queue is equal to \( \lambda = 1 \). Hence, the paired queueing system is symmetric and performance measures are equal for all queues. Figures 1(a) and 1(b) depict the mean queue content and the blocking probability in a queue versus the service rate \( \mu \), respectively. As the state-space \( C \) is finite, the deviation matrix of the CTMC with generator matrix \( Q_0 \) exists, with state \( \pi^{(0)}(c) = 1 \) and \( \pi^{(0)}(i) = 0 \) for \( i \in C \).

Let \( D_0 \) be the deviation matrix of the CTMC with generator matrix \( Q_0 \).

\[
D_0 = \int_0^\infty (P_0(t) - \Pi_0) dt .
\]

Here the family \( \{P_0(t) = \exp(Q_0t), t \geq 0\} \) is the Markov semigroup of the CTMC, and \( \Pi_0 = \lim_{t \to \infty} P_0(t) = 1' \pi^{(0)}, 1' \) being a column vector of ones. As the state-space \( C \) is finite, the deviation matrix is well defined. Moreover, the deviation matrix satisfies \( D_01' = 0 \) — the row sums are zero — and,

\[
D_0Q_0 = Q_0D_0 = \Pi_0 - I .
\]

Theorem 1. The solution \( \pi^{(\mu)} \) of the CTMC adheres to the following power series expansion,

\[
\pi^{(\mu)} = \sum_{k=0}^{\infty} \left( \pi^{(0)}(Q_1D_0)^k \right) \mu^k ,
\]

for \( 0 \leq \mu < \mu_0, \mu_0^{-1} \) being the spectral radius of \( Q_1D_0 \). Moreover, \( \mu_0 \) is bounded from below by \( \mu_*^0 \) and \( \mu_*^1 \),

\[
\mu_*^0 = 2 \int_0^\infty \left( 1 - \prod_{k=1}^{K} F(t; C_k, \lambda_k) \right) dt \geq \left( \sum_{k=1}^{K} \frac{C_k}{\lambda_k} \right)^{-1} = \mu_*^1 ,
\]

with \( F \) being the Erlang distribution,

\[
F(t; C_k, \lambda_k) = 1 - \sum_{n=0}^{C_k-1} \frac{1}{n!} \lambda_k^t \lambda_k^n .
\]

Appendix: Convergence of the power series

We now justify the series expansion. The basic ideas in this section date back to the seminal work of Schweitzer [11]. The series expansion is validated by explicitly constructing such an expansion. We first introduce some additional notation and the basic notion of the deviation matrix of a CTMC.

Let \( \pi^{(\mu)} \) denote the steady state solution \( \{\pi(i)\}_{i \in C} \) of the balance equations (1). We have made the dependence of \( \pi^{(\mu)} \) on \( \mu \) explicit for ease of notation. The balance equations can then be written in matrix notation as follows,

\[
\pi^{(\mu)}Q^{(\mu)} = \pi^{(\mu)}(Q_0 + \mu Q_1) = 0 ,
\]

where \( Q^{(\mu)} \) is the \([C] \times [C]\) generator matrix of the CTMC and where \( Q_0 \) and \( Q_1 \) are known matrices that do not depend on \( \mu \). In view of the system assumptions it is readily seen that \( Q^{(0)} = Q_0 \) only has one recurrent state, i.e. \( c \) (the full state) is recurrent and all the others are transient. Therefore, the stationary vector \( \pi^{(0)} \) exists, with state \( \pi^{(0)}(c) = 1 \) and \( \pi^{(0)}(i) = 0 \) for \( i \in C \).

Let \( D_0 \) be the deviation matrix of the CTMC with generator matrix \( Q_0 \),

\[
D_0 = \int_0^\infty (P_0(t) - \Pi_0) dt .
\]

Here the family \( \{P_0(t) = \exp(Q_0t), t \geq 0\} \) is the Markov semigroup of the CTMC, and \( \Pi_0 = \lim_{t \to \infty} P_0(t) = 1' \pi^{(0)} \), \( 1' \) being a column vector of ones. As the state-space \( C \) is finite, the deviation matrix is well defined. Moreover, the deviation matrix satisfies \( D_01' = 0 \) — the row sums are zero — and,

\[
D_0Q_0 = Q_0D_0 = \Pi_0 - I .
\]

Theorem 1. The solution \( \pi^{(\mu)} \) of the CTMC adheres to the following power series expansion,

\[
\pi^{(\mu)} = \sum_{k=0}^{\infty} \left( \pi^{(0)}(Q_1D_0)^k \right) \mu^k ,
\]

for \( 0 \leq \mu < \mu_0, \mu_0^{-1} \) being the spectral radius of \( Q_1D_0 \). Moreover, \( \mu_0 \) is bounded from below by \( \mu_*^0 \) and \( \mu_*^1 \),

\[
\mu_*^0 = 2 \int_0^\infty \left( 1 - \prod_{k=1}^{K} F(t; C_k, \lambda_k) \right) dt \geq \left( \sum_{k=1}^{K} \frac{C_k}{\lambda_k} \right)^{-1} = \mu_*^1 ,
\]

with \( F \) being the Erlang distribution,

\[
F(t; C_k, \lambda_k) = 1 - \sum_{n=0}^{C_k-1} \frac{1}{n!} \lambda_k^t \lambda_k^n .
\]
Figure 1: Mean queue content and blocking probability for a symmetric paired queueing system.

Figure 2: Mean (a) and variance (b) of the queue content of an asymmetric paired queueing system.
Proof. Multiplying (4) by $D_0$ and invoking (6) yields,

$$\pi^{(\mu)}(Q_0 + \mu Q_1)D_0 = \pi^{(\mu)}(I_1 - I) + \pi^{(\mu)}\mu Q_1 D_0 = 0.$$  

Moreover, we have $\pi^{(\mu)}\Pi_0 = \pi^{(\mu)}\Pi^{(0)} = \pi^{(0)}$, such that,

$$\pi^{(\mu)}(I - \mu Q_1 D_0) = \pi^{(0)}.$$  

The spectral radius of $\mu Q_1 D_0$ is $\mu/\mu_0$. Hence for $\mu < \mu_0$, $(I - \mu Q_1 D_0)$ is invertible and the Neumann series converges to the inverse,

$$\sum_{k=0}^{\infty}(\mu Q_1 D_0)^k = (I - \mu Q_1 D_0)^{-1}.$$  

Combining the previous expressions immediately yields the series expansion (7).

As all elements but the last column of $\Pi_0$ are zero, only the last column of $D_0$ may contain negative values; see (5). Moreover, the row sums of $D_0$ are zero, hence the last column is equal in absolute value to the sum of the other columns. The entries in the last column of $D_0$ have the following interpretation,

$$[D_0]_{ic} = -\int_0^\infty(1 - [P_0(t)]_{ic})dt = -E[T_i],$$

where $T_i$ is a random variable denoting the time it takes to reach the full state $c$ from state $i$ (assuming no departures). This interpretation shows that $\gamma = E[T_0] \geq E[T_i]$ for all $i \in C$ where $0$ denotes the empty state.

The time to fill up the $i$th queue is Erlang distributed with $C_i$ stages and rate $\lambda_i$ and the time to fill up all queues is the maximum of $K$ Erlang distributed random variables. Therefore, the cumulative distribution of $T_0$ is the product of $K$ Erlang distributions and $\gamma$ is calculated by integrating this distribution,

$$\gamma = \int_0^\infty \left(1 - \prod_{k=1}^K F(t; C_k, \lambda_k)\right)dt.$$  

Moreover, the maximum of $K$ non-negative random variables is bounded from above by the sum of these random variables, which yields the following crude upper bound for $\gamma$,

$$\gamma \leq \sum_{k=1}^K \frac{C_k}{\lambda_k},$$

the $k$th term in the sum on the right-hand side corresponding to the mean time to fill up the $k$th queue.

As the row sums of $Q_1$ are zero ($Q_1^{\mu}$ is a generator matrix for every $\mu$), we have $Q_1\Pi_0 = 0$. Moreover, for any induced matrix norm, we have $\|Q_1 D_0\| \geq \mu_0$. Therefore, we find,

$$\mu_0^{-1} \leq \|Q_1 D_0\| = \|Q_1(D_0 + \gamma \Pi_0)\| \leq \|Q_1\|\|D_0 + \gamma \Pi_0\|.$$  

Particularly using the maximum absolute row sum norm, we have $\|Q_1\| = 2; \|Q_1\|_\infty = -1$ if all queues are non-empty in state $i$ and $0$ if this not the case such that the corresponding row sums equal 2 and 0 respectively. In view of the definition of $\gamma$, one easily verifies that the matrix $D_0 + \gamma \Pi_0$ has no negative entries. Recalling that $D_0$ has zero row sums, this shows that all row sums of $D_0 + \gamma \Pi_0$ equal $\gamma$: $\|D_0 + \gamma \Pi_0\| = \gamma$ and,

$$\frac{1}{\mu_0} \leq 2\gamma \leq \frac{1}{\mu_0^2},$$

which proves the lower bound $\mu_0^0$ for $\mu_0$. The lower bound $\mu_1^*$ follows from $\mu_0^{-1} \leq 2\gamma$ and the crude bound (8) for $\gamma$.

To illustrate Theorem 1, figure 3 depicts $\mu_0$, the spectral radius of $Q_1 D_0$ and the lower bounds $\mu_0^0$ and $\mu_1^*$ for a system with $K = 3$ paired queues, each queue having a varying capacity from 2 to 10. As the figure shows, the bounds are much smaller than the convergence radius. It should be noted that both bounds are easy-to-derive but also rather loose bounds on the convergence radius. The bounds above can be made tighter by (1) not relying on the submultiplicative property of the matrix norm; (2) a matrix norm which is more adapted to this model. Both these approaches quickly lead to lengthy calculations and we consider them to be outside of the scope of the paper.

Acknowledgements

The second author is a post-doctoral fellow with the Research Foundation, Flanders. This research has been funded by the Interuniversity Attraction Poles Programme initiated by the Belgian Science Policy Office.

References


