Queueing Systems 38, 205–212, 2001

Approximations for Multi-Class Departure Processes

RENÉ CALDENTEY rcaldent@mit.edu
Department of Industrial Engineering, University of Chile, and
Sloan School of Management, MIT, Cambridge, MA 02139, USA

Received 10 June 2000; Revised 28 November 2000

Abstract. The exact analysis of a network of queues with multiple products is, in general, prohibited because of the non-renewal structure of the arrival and departure processes. Two-moment approximations (decomposition methods, Whitt [9]) have been successfully used to study these systems. The performance of these methods, however, strongly depends on the quality of the approximations used to compute the squared coefficient variation (CV) of the different streams of products.

In this paper, an approximation method for computing the squared coefficient of variation of the departure stream from a multi-class queueing system is presented. In particular, we generalize the results of Bitran and Tirupati [3] and Whitt [11] related to the interference effect.

Keywords: multi-class queueing networks, parametric-decomposition approximations, departure processes

Consider the following situation:

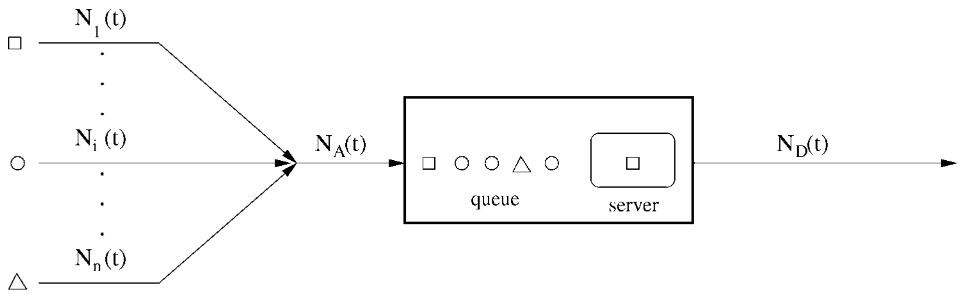

Figure 1.

A single server queueing station serves, according to a FIFO policy, a family of n different products. The arrival process of product i ($i = 1, \ldots, n$) is renewal and it is characterized by the counting process $\{N_i(t)\colon t \geqslant 0\}$. The arrival processes are independent across products. In addition, we define

- X_i: the random variable that describes the inter-arrival interval for product i.
- $\lambda_i = (E[X_i])^{-1}$: the mean arrival rate of product i.
- C_i: the squared coefficient of variation of the inter-arrival interval of product i.
- $\Phi_i(s)$: the Laplace–Stieljes transform of the cdf of X_i.

From the point of view of the server, the arrival process to the station is $\{N^A(t): t \geqslant 0\}$, where $N^A(t) = \sum_{i=1}^n N_i(t)$. Except for very particular cases (e.g., the $N_i(t)$ are Poisson processes), neither the arrival nor the departure streams to the station are renewal processes. If the interest is to model the behavior of a network of queues with multiple products where each station of the network looks like the one in figure 1, then the non-renewal structure of the arrival and departure processes makes the exact analysis of the system almost intractable.

For that reason, approximate methods have been developed to evaluate different performance measures of the system. One of the most successful approaches has been the decomposition method. In simple terms, this method approximates the different streams of products on the network by renewal processes, and then analyzes each station in isolation (the reader is referred to [2,8,9] for more details about the decomposition method). The quality of the results produced by this method is, of course, strongly related to the quality of the approximations made to represent the non-renewal processes by renewal processes. Therefore, most of the effort required by the decomposition approach is in the characterization of the different streams of products within the system.

The performance measures for a GI/G/1 queue are based on the first two moments of the arrival process and service distribution [6]. For this reason, mean and coefficient of variation are the main statistics that researchers have considered when developing the decomposition method.

For the particular case of figure 1, we are interested in the behavior of the departure stream, and, in particular, in the point processes that characterize the departure of each type of product. In terms of the aggregate departure stream, there are good approximations for the mean and CV of the inter-departure interval (see [8,10]). For example,

$$\lambda^D = \sum_{i=1}^{n} \lambda_i, \tag{1}$$

$$C^D = \rho^2 C^S + \left(1 - \rho^2\right) C^A, \tag{2}$$

where λ^D is the total departure rate, C^D, C^S and C^A are the CV of the departure, service and arrival processes, respectively. And ρ is the traffic intensity of the station.

A first approach to estimate from (1) and (2) the parameters for each product is to assume that each departure has probability $p_i \equiv \lambda_i/\lambda^D$ $(i = 1, \dots, n)$ of being an i product. Thus, the output stream of product i is characterized by [9]

$$\lambda_i^D = \lambda_i, \tag{3}$$

$$C_i^D = p_i C^D + 1 - p_i. \tag{4}$$

Relation (3) is of course exact. However (4) is in general an approximation which is based on a Markovian routing structure and it does not fully incorporate the structure of the departure process which is strongly related to the arrival process.

An important refinement to (4) was first proposed by Bitran and Tirupati [3]. The main idea is to incorporate the *interference* that other products have on a particular product. For example, consider figure 2.

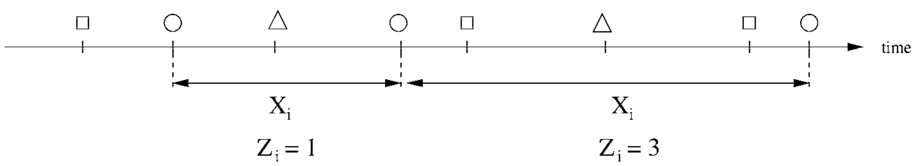

Figure 2.

In figure 2, X_i is the inter-departure interval for product i (circles). Z_i represents the number of products different than i that depart within the interval X_i. Therefore, if we approximate the aggregate departure process by a renewal process with inter-departure interval X^D, we have the following identity:

$$X_i = \sum_{k=1}^{Z_i+1} X_k^D, \tag{5}$$

where $\{X_k^D\}_{k\geqslant 1}$ is a sequence of i.i.d. random variables. Defining $W_i = Z_i + 1$, the improvement to (4) is [3]

$$C_i^D = p_i C^D + C_{W_i}, \tag{6}$$

where C_{W_i} is the CV of W_i. In their paper, Bitran and Tirupati [3] discuss the difficulty of evaluating C_{W_i} and proposed three types of approximations based on special distributions (Poisson and Erlang). Whitt [11], extended the work of Bitran and Tirupati, approximating C_i^D by considering specific renewal processes. In particular, Whitt used batch-Poisson and batch-deterministic processes to extend Bitran and Tirupati's results.

In what follows, we extend those results to the general case. Let us first decompose $Z_i \equiv \sum_{j\neq i} Z_{ji}$, where Z_{ji} is the number of departure of product j during an inter-departure interval of product i. Since the server uses a FIFO discipline, Z_{ji} also represents the number of arrivals of product j during an inter-arrival interval of product i. In addition, the arrival streams of the different products are independent renewal processes, thus an arbitrary arrival time of product i represents a random incidence time for the other arrival processes. Therefore, conditioned on $X_i = t$, $Z_{ji}(t)$ is the counting process associated with the corresponding equilibrium process of the arrivals of product j (see [5] for more details about the equilibrium process). Let $K_j(n,t) \equiv Pr(Z_{ji}(t) = n)$, then the Laplace transform of $K_j(n,t)$ is (see [4] for details on the derivation of (7))

$$K_j(z,s) \equiv \int_0^\infty \mathrm{e}^{-st}\left(\sum_{n=0}^\infty z^n K_j(n,t)\right)\mathrm{d}t = \frac{1}{s} + \frac{\lambda_j(1-\Phi_j(s))(z-1)}{s^2(1-z\Phi_j(s))}. \tag{7}$$

Taking the first derivative of $K_j(z,s)$ with respect to z and evaluating the result at $z = 1$, we obtain the Laplace transform for the mean of $Z_{ji}(t)$, i.e.,

$$L\bigl(E\bigl[Z_{ji}(t)\bigr]\bigr) = \left(\frac{\partial K_j(z,s)}{\partial z}\right)_{z=1} = \left(\frac{\lambda_j(1-\Phi_j(s))^2}{s^2(1-z\Phi_j(s))^2}\right)_{z=1} = \frac{\lambda_j}{s^2}. \tag{8}$$

Thus, inverting the transform we get

$$E\big[Z_{ji}(t)\big] = \lambda_j t. \tag{9}$$

If we now take the second derivative of $K_j(z, s)$ with respect to z, we have that

$$L\Big(E\big[\big(Z_{ji}(t)\big)^2 - Z_{ji}(t)\big]\Big) = \left(\frac{\partial^2 K_j(z, s)}{\partial z^2}\right)_{z=1} = \frac{2\lambda_j \Phi_j(s)}{s^2(1 - \Phi_j(s))}. \tag{10}$$

Let

$$f_j(t) \equiv E\big[\big(Z_{ji}(t)\big)^2 - Z_{ji}(t)\big] = L^{-1}\left(\frac{2\lambda_j \Phi_j(s)}{s^2(1 - \Phi_j(s))}\right).$$

Proposition 1. If the arrival processes for the different products are mutually independent, then the coefficient of variation of W_i is

$$C_{W_i} = (1 - p_i)\big[(1 - p_i)C_i + p_i\big] - (C_i + 1)\sum_{j \neq i} p_j^2 + p_i^2 \sum_{j \neq i} E\big[f_j(X_i)\big]. \tag{11}$$

In addition, let $F_{kj}(\cdot)$ be the k-fold convolution of the distribution of X_j, and let Y_{kj} be a random variable with cdf $F_{kj}(t)$. Then,

$$f_j(t) = 2\lambda_j \sum_{k=1}^{\infty}\left(\int_0^t F_{kj}(z)\,\mathrm{d}z\right) = 2\lambda_j \sum_{k=1}^{\infty} E\big[(t - Y_{kj})^+\big]. \tag{12}$$

Proof. See the appendix. □

Relation (11) is exact. However, computing $f_j(\cdot)$ from (12) is not always a simple task. In the general case, numerical methods can be used to (i) solve the inversion problem in (10) (see [1]) and then apply (11), or (ii) estimate $f_j(t)$ from (12) through simulation. In addition, a simple bound can be found by applying Jensen's inequality to (12), i.e.,

$$E\big[(t - Y_{kj})^+\big] \geqslant \left(t - \frac{k}{\lambda_j}\right)^+.$$

We conclude this note by analyzing four special cases that have received special attention in the literature:

Case 1. Poisson arrivals

Clearly, one the most classical assumptions in the queueing literature is that arrival processes are Poisson. Under this assumption, the X_j are exponentially distributed with rate λ_j and the inversion problem can be solved in closed form. In fact,

$$f_j(t) = L^{-1}\left(\frac{2\lambda_j \Phi_j(s)}{s^2(1 - \Phi_j(s))}\right) = L^{-1}\left(\frac{2\lambda_j^2}{s^3}\right) = \lambda_j^2 t^2. \tag{13}$$

Thus, replacing (13) in (11) and after some straightforward manipulations we obtain

$$C_{W_i} = (1 - p_i)\big[(1 - p_i)C_i + p_i\big]. \tag{14}$$

This condition was first derived in [3] as part of their first approximation. In fact, combining (6) and (14), we recover relation (11) in [3]

$$C_i^D = p_i C^D + (1 - p_i)\big[(1 - p_i)C_i + p_i\big].$$

Case 2. Asymptotic result

Another important case arises when each of the input streams has a small intensity when compared to the aggregate stream. In this situation, we can use the asymptotic approximation (see [4] for details)

$$L^{-1}\left(\frac{\Phi_j(s)}{s(1 - \Phi_j(s))}\right) \approx \lambda_j t + \frac{C_j - 1}{2}.$$

Thus, combining this relation and the property $sL(f(t)) = L(f'(t)) + f(0)$, we end up with

$$f_j(t) = \lambda_j^2 t^2 + \lambda_j(C_j - 1)t. \tag{15}$$

Finally, replacing (15) in (11) we get

$$C_{W_i} = (1 - p_i)^2 C_i + p_i \sum_{j \neq i} p_j C_j. \tag{16}$$

If we aggregate all the classes different than i in a single class i_-, then the previous relation is equivalent to

$$C_{W_i} = (1 - p_i)^2 C_i + (1 - p_i)p_i C_{i_-}.$$

This is exactly equation (4) in [11] and it was obtained using batch-Poisson and batch-deterministic renewal processes.

Both previous cases exploit some asymptotic result. In what follows, we present two cases that have a different inspiration. Based on the decomposition method, we know that product streams are usually characterized by their first two moments. Moreover, depending on the value of the squared coefficient of variation, a sum or a mixture of two exponential random variables can be used to fit the first two moments (see [8]). The sum of two exponentials is used when the CV is lower than 1, and the mixture of two exponentials is used when the CV is greater than 1.

Case 3. Sum of two exponentials

Let $X_j = Y_{1j} + Y_{2j}$, where Y_{rj} is an exponential random variable with rate $\mu_{rj},\ r = 1, 2$. Then, $\Phi_j(s)$ is

$$\Phi_j(s) = \left(\frac{\mu_{1j}}{\mu_{1j} + s}\right)\left(\frac{\mu_{2j}}{\mu_{2j} + s}\right).$$

The function $f_j(t)$ can be computed solving

$$f_j(t) = 2\lambda_j L^{-1}\left(\frac{\mu_{1j}\mu_{2j}}{s^3(s + \mu_{1j} + \mu_{2j})}\right).$$

In this case, the inversion problem can be obtained in closed form. In fact, after some algebra we get

$$f_j(t) = 2\lambda_j\left(\frac{\mu_{1j}\mu_{2j}}{\mu_{1j} + \mu_{2j}}\right)\left[\frac{t^2}{2} - \frac{t}{\mu_{1j} + \mu_{2j}} + \frac{1 - \mathrm{e}^{-(\mu_{1j}+\mu_{2j})t}}{(\mu_{1j} + \mu_{2j})^2}\right].$$

In addition, using the identities

$$\lambda_j = \frac{\mu_{1j}\mu_{2j}}{\mu_{1j} + \mu_{2j}} \quad \text{and} \quad \mu_{1j} + \mu_{2j} = \frac{2\lambda_j}{1 - C_j}$$

we obtain

$$f_j(t) = \lambda_j^2 t^2 + \lambda_j(C_j - 1)t + \frac{(1 - C_j)^2}{2}\left(1 - \mathrm{e}^{-2\lambda_j t/(1-C_j)}\right). \tag{17}$$

Finally, replacing (17) in (11) we get

$$C_{W_i} = (1 - p_i)^2 C_i + p_i \sum_{j \neq i} p_j C_j + \left(\frac{p_i^2}{2}\right) \sum_{j \neq i} (1 - C_j)^2 \left[1 - \Phi_i\left(\frac{2\lambda_j}{1 - C_j}\right)\right]. \tag{18}$$

We can see that relation (18) generalizes relations (14) and (16). In fact, using $C_j = 1$ (exponential distribution) or $\lambda_j \approx 0$ (asymptotic case) in (18) we recover the previous relations.

Case 4. Mixture of two exponentials

Let $X_j = Y_{1j}$ with probability θ_{1j} and $X_j = Y_{2j}$ with probability $\theta_{2j} = 1 - \theta_{1j}$. In this situation, the Laplace–Stieljes transform of X_j is

$$\Phi_j(s) = \frac{\theta_{1j}\mu_{1j}}{\mu_{1j} + s} + \frac{\theta_{2j}\mu_{1j}}{\mu_{2j} + s}.$$

If, in addition, we assume balanced mean, i.e., $\theta_{1j}/\mu_{1j} = \theta_{2j}/\mu_{2j}$ (see [7] for details), we have

$$\theta_{rj} = \frac{\sqrt{C_j + 1} \pm \sqrt{C_j - 1}}{2\sqrt{C_j + 1}} \quad \text{and} \quad \mu_{rj} = 2\lambda_j\theta_{rj}, \quad r = 1, 2.$$

In this case, we can again compute in closed form $f_j(t)$. We omit here the details of the derivation.

$$f_j(t) = \lambda_j^2 t^2 + \lambda_j (C_j - 1)t + 2\bigl(C_j^2 - 1\bigr)\bigl(1 - \mathrm{e}^{-2\lambda_j t/(c_j+1)}\bigr). \tag{19}$$

From (19) and (11), we conclude:

$$C_{W_i} = (1 - p_i)^2 C_i + p_i \sum_{j\neq i} p_j C_j + 2p_i^2 \sum_{j\neq i} \bigl(C_j^2 - 1\bigr)\left[1 - \Phi_i\left(\frac{2\lambda_j}{C_j + 1}\right)\right]. \tag{20}$$

Acknowledgements

The author thanks Gabriel Bitran and Lawrence Wein for helping him improve this note.

Appendix

Proof of proposition 1. Let us first recall that $W_i = Z_i + 1 = \sum_{j\neq i} Z_{ji} + 1$. Thus,

$$E[W_i] = \sum_{j\neq i} E[Z_{ji}] + 1 \quad \text{and} \quad \mathrm{Var}[W_i] = \mathrm{Var}[Z_i].$$

From (9) $E[Z_{ji}(t)] = \lambda_j t$, therefore $E[Z_{ji}] = \lambda_j E[X_i] = p_j/p_i$, and

$$E[W_i] = \sum_{j\neq i} \frac{p_j}{p_i} + 1 = \frac{1}{p_i}. \tag{A.1}$$

On the other hand, from the definition of $f_j(t)$ we have

$$E\bigl[\bigl(Z_{ji}(t)\bigr)^2\bigr] = f_j(t) + E\bigl[Z_{ji}(t)\bigr] = f_j(t) + \lambda_j t,$$

which implies

$$\begin{aligned} E\bigl[\bigl(Z_i(t)\bigr)^2\bigr] &= E\left[\left(\sum_{j\neq i} Z_{ji}(t)\right)^2\right] \\ &= \sum_{j\neq i}\bigl[f_j(t) + \lambda_j t\bigr] + 2\sum_{j\neq i}\sum_{k<j;\,k\neq i} E\bigl[Z_{ji}(t) Z_{ki}(t)\bigr]. \end{aligned} \tag{A.2}$$

But conditioned on $X_i = t$, the random variables $Z_{ji}(t)$ and $Z_{ki}(t)$ $(k \neq j)$ are independent, therefore (A.2) becomes

$$E\bigl[\bigl(Z_i(t)\bigr)^2\bigr] = \sum_{j\neq i} f_j(t) + \lambda(1 - p_i)t + 2\lambda^2 t^2 \sum_{j\neq i}\sum_{k<j;\,k\neq i} p_j p_k. \tag{A.3}$$

Taking expectation in (A.3) with respect to X_i we get

$$E\bigl[Z_i^2\bigr] = \sum_{j\neq i} E\bigl[f_j(X_i)\bigr] + \frac{1 - p_i}{p_i} + 2\lambda E\bigl[X_i^2\bigr] \sum_{j\neq i}\sum_{k<j;\,k\neq i} p_j p_k.$$

Thus, combining this equation and (A.1), we end up with

$$\begin{aligned}\mathrm{Var}[W_i] &= E\big[Z_i^2\big] - \big(E[Z_i]\big)^2 \\ &= \sum_{j\neq i} E\big[f_j(X_i)\big] + \frac{(1-p_i)(2p_i-1)}{p_i^2} + 2\lambda E\big[X_i^2\big]\sum_{j\neq i}\sum_{k<j;\, k\neq i} p_j p_k. \qquad (A.4)\end{aligned}$$

Combining (A.1) and (A.4), and after some manipulations we have

$$C_{W_i} = (1-p_i)\big[(1-p_i)C_i + p_i\big] - (C_i+1)\sum_{j\neq i} p_j^2 + p_i^2 \sum_{j\neq i} E\big[f_j(X_i)\big], \qquad (A.5)$$

completing the proof of the first part. For the second part, we use the following result:

$$sL\big(f_j(t)\big) = 2\lambda_j\left(\frac{\Phi_j(s)}{s(1-\Phi_j(s))}\right) = 2\lambda_j \sum_{k=1}^{\infty} \frac{\Phi_j^k(s)}{s}.$$

We note that $\Phi_j^k(s)$ represents the Laplace–Stieljes transform for the k-fold convolution of the distribution of X_j. The proof is completed combining this observation and the following identities:

- $sL(f(t)) = L(f'(t)) + f(0)$.
- $L(f(s))/s = L(F(s))$, where $F(t) = \int_0^t f(z)\,\mathrm{d}z$.
- If X is non-negative random variable with cdf $F(x)$, then for any $t \geqslant 0$,

$$\int_0^t F(z)\,\mathrm{d}z = E\big[(t-X)^+\big]. \qquad \square$$

References

[1] J. Abate and W. Whitt, Numerical inversion of laplace transforms of probability distributions, ORSA J. Comput. 7 (1995) 36–43.

[2] G. Bitran and S. Dasu, A review of open queueing network models of manufacturing systems, Queueing Systems 12 (1992) 95–134.

[3] G. Bitran and D. Tirupati, Multiproduct queueing networks with deterministic routing: Decomposition approach and the notion of interference, Managm. Sci. 34(1) (1988) 75–100.

[4] D. Cox, *Renewal Theory* (Methuen, London, 1972).

[5] A. Khintchine, *Mathematical Methods in Queueing Theory* (Griffin, London, 1960).

[6] W. Kraemer and M. Langenbach-Belz, Approximate formulae for the delay in the queueing system GI/G/1, in: *Congressbook, 8th ITC*, Melbourne, 1976, pp. 235.1–235.8.

[7] P. Kuehn, Approximate analysis of general queueing networks by decomposition, IEEE Trans. Commun. 27 (1979) 113–126.

[8] W. Whitt, Approximating a point process by a renewal process: Two basic methods, Oper. Res. 30(1) (1982) 125–147.

[9] W. Whitt, The queueing network analyzer, Bell System Techn. J. 62 (1983) 2779–2843.

[10] W. Whitt, Approximations for departure processes and queues in series, Naval Res. Logist. Quart. 31 (1984) 499–521.

[11] W. Whitt, Towards better multi-class parametric-decomposition approximations for open queueing networks, Ann. Oper. Res. 48 (1994) 221–248.