Journal of Computational and Applied Mathematics 20 (1987) 119–125
North-Holland

On a numerical method for calculating state probabilities for queueing systems with more than one waiting line

J.P.C. BLANC *
Department of Mathematics and Informatics, Delft University of Technology, 2600 AJ Delft, The Netherlands

Received 26 May 1986

Abstract: Keane, Hooghiemstra and Van de Ree have proposed a new numerical method for calculating state probabilities for queueing systems with more than one waiting line in parallel. The method is based on power series expansions of state probabilities as functions of the traffic intensity of a system. The coefficients of these power series can be recursively calculated. The coefficients of the power series expansions of moments of queue length distributions can be derived from those of the state probabilities in a straightforward manner. The above method is discussed for a rather general class of exponential queueing systems. The asymptotic behaviour of moments in heavy traffic is used to obtain extrapolations of the coefficients of their power series expansions at the origin. The calculation of moments is strongly improved by means of these extrapolations.

Keywords: Queues in parallel, moments, power series expansions, traffic intensity.

1. Introduction

Queueing systems with more than one waiting line are in general hard to analyse if the joint queue length distribution does not possess a product form. During the last decade a rather general analytical method has been developed for the study of systems with two waiting lines. This method is based on application of the theory of Riemann–Hilbert type boundary value problems to solve functional equations for bivariate generating functions of joint queue length distributions, see a [5] and [2]. At present it does not seem likely that this method can be generalised as useful tool for the analysis of systems with more than two waiting lines. Moreover, in many instances non-trivial algorithms (e.g. the numerical solution of integral equations) are required to obtain numerical data from the formulas which result by application of this method.

Numerical methods for calculating state probabilities without the use of generating functions have been proposed by several authors. They include Conolly [4], who uses truncation of the state space (i.e. of the waiting rooms) and inversion of the matrix of transition rates, and Gertsbakh [6], who uses truncation of the state space and numerical evaluation of the matrix-ge-

* Present address: Catholic University of Brabant, Faculty of Economics, P.O. Box 90153, 5000 LE Tilburg, The Netherlands.

ometric solution of the truncated system. Keane et al. [7] propose another method for calculating state probabilities directly from the equilibrium equations. They consider power series expansions of the state probabilities at the origin as functions of the traffic intensity. For a wide class of queueing systems the coefficients of these power series expansions can be calculated recursively. Keane et al. [7] found that for several models a (bilinear) conformal transformation has to be used to ensure convergence of the power series over the whole range of traffic intensities for which a system is stable. The coefficients of the power series expansions of the moments of the queue length distributions follow directly from those of the state probabilities. The number of coefficients required to reach a certain accuracy can be strongly reduced by extrapolation of these coefficients on the basis of the singular behaviour of the moments near the end of the region of stability of the queueing system.

The idea of the method of Keane et al. [7] will be outlined in Section 2 for a class of exponential queueing systems; Section 3 is devoted to the procedure for calculating moments of queue length distributions. In Section 4 the procedure will be illustrated by means of some examples, and experiences with the method will be discussed.

2. Calculating state probabilities

Consider the class of stable exponential queueing systems with s ($s > 1$) waiting lines which can be described in the following way. Let $\bar{n} = (n_1, \dots, n_s)$ be a vector with non-negative entries; let $\rho a_j(\bar{n})$ be the arrival rate to queue j and $d_j(\bar{n})$ be the departure rate from queue j, $j = 1, \dots, s$, when the system is in state $\bar{n}$. Let N_j be the number of customers in queue j in the stationary situation, $j = 1, \dots, s$, and define the stationary state probabilities

$$p(\rho;\ \bar{n}) = \Pr\{N_j = n_j,\ j = 1, \dots, s\}. \tag{2.1}$$

Let $I(E)$ stand for the indicator function of the event E, and $\bar{e}_j = (e_{j1}, \dots, e_{js})$ for the vector with $e_{jj} = 1$ and $e_{ji} = 0$ if $i \neq j$, $j = 1, \dots, s$. The equations for the state probabilities (2.1) read:

$$\left[\rho \sum_{j=1}^{s} a_j(\bar{n}) + \sum_{j=1}^{s} d_j(\bar{n}) I(n_j > 0)\right] p(\rho;\ \bar{n})$$
$$= \sum_{j=1}^{s} d_j(\bar{n} + \bar{e}_j) p(\rho;\ \bar{n} + \bar{e}_j) + \rho \sum_{j=1}^{s} a_j(\bar{n} - \bar{e}_j) p(\rho;\ \bar{n} - \bar{e}_j) I(n_j > 0). \tag{2.2}$$

The following assumptions are made:

(A1) the system is stable for $0 \leqslant \rho < 1$;

(A2) the state probabilities $p(\rho;\ \bar{n})$ possess analytic continuations as functions of ρ into a domain which includes the disk $|\rho - \frac{1}{2}| \leqslant \frac{1}{2}$;

(A3) not all servers are idle when customers are present in the system, i.e.

$$\sum_{j=1}^{s} d_j(\bar{n}) I(n_j > 0) > 0 \quad \text{for all } \bar{n},\ \bar{n} \neq \bar{0} = (0, \dots, 0).$$

Introducing the notations, for $m = 0, 1, 2, \dots$, $|\bar{n}| = \sum_{j=1}^{s} n_j$,

$$D(\rho;\ m) = \sum_{|\bar{n}|=m} p(\rho;\ \bar{n}) \sum_{j=1}^{s} d_j(\bar{n}) I(n_j > 0),$$

$$A(\rho;\ m) = \sum_{|\bar{n}|=m} p(\rho;\ \bar{n}) \sum_{j=1}^{s} a_j(\bar{n}), \tag{2.3}$$

we obtain by summation of the equations in (2.2) over all states $\bar{n}$ with $|\bar{n}| = m$, for $m = 1, 2, \dots$,

$$\rho A(\rho;\ m) + D(\rho;\ m) = D(\rho;\ m+1) + \rho A(\rho;\ m-1), \quad \rho A(\rho;\ 0) = D(\rho;\ 1).$$

This implies, by induction, that

$$\rho A(\rho;\ m) = D(\rho;\ m+1), \quad m = 0, 1, 2, \dots . \tag{2.4}$$

From (2.3) and assumption (A3) it follows, for $m = 1, 2, \dots$, that if $D(\rho;\ m) = O(\rho^k)$, $\rho \downarrow 0$, for some k, then also $p(\rho;\ \bar{n}) = O(\rho^k)$, $\rho \downarrow 0$, for all $\bar{n}$ with $|\bar{n}| = m$, and hence $A(\rho;\ m) = O(\rho^k)$, $\rho \downarrow 0$. Therefore, relations (2.4) imply that $D(\rho;\ m) = O(\rho^m)$, $\rho \downarrow 0$, $m = 1, 2, \dots$ (note that $p(\rho;\ \bar{0}) = O(1)$, $\rho \downarrow 0$). Hence, if state probabilities $p(\rho;\ \bar{n})$ satisfy (2.2) and if assumption (A3) holds, then

$$p(\rho;\ \bar{n}) = O(\rho^{|\bar{n}|}), \quad \rho \downarrow 0. \tag{2.5}$$

Motivated by this property Keane et al. [7] propose the following procedure. Write

$$p(\rho;\ \bar{n}) = \rho^{|\bar{n}|} \sum_{k=0}^{\infty} \rho^k u(k;\ \bar{n}), \tag{2.6}$$

and substitute these power series expansions into (2.2). Equating coefficients of corresponding powers of ρ leads to: for $k = 0, 1, \dots$,

$$\begin{aligned} &\sum_{j=1}^{s} d_j(\bar{n}) I(n_j > 0) u(k;\ \bar{n}) \\ &\quad = -\sum_{j=1}^{s} a_j(\bar{n}) u(k-1;\ \bar{n}) I(k > 0) + \sum_{j=1}^{s} d_j(\bar{n} + \bar{e}_j) u(k-1;\ \bar{n} + \bar{e}_j) I(k > 0) \\ &\qquad + \sum_{j=1}^{s} a_j(\bar{n} - \bar{e}_j) I(n_j > 0) u(k;\ \bar{n} - \bar{e}_j). \end{aligned} \tag{2.7}$$

To obtain a relation for $u(k;\ \bar{0})$, $k = 0, 1, \dots$, the law of total probability is used: substituting (2.6) and equating corresponding powers of ρ gives

$$u(0;\ \bar{0}) = 1; \qquad u(k;\ \bar{0}) = -\sum_{0 < |\bar{n}| \leq k} \cdots \sum u(k - |\bar{n}|;\ \bar{n}), \quad k = 1, 2, \dots . \tag{2.8}$$

It is readily verified that all coefficients $u(k;\ \bar{n})$ can be recursively calculated from (2.8) and (2.7). The state probabilities $p(\rho;\ \bar{n})$ can then be approximated with any degree of accuracy for ρ smaller than the radius of convergence of their power series (2.6). When these radii of convergence are smaller than one, Keane et al. [7] propose to apply a bilinear conformal transformation of the real interval [0,1] onto itself:

$$\theta = \frac{1+G}{1+G\rho}\rho, \quad \rho = \frac{\theta}{1+G-G\theta}, \qquad G \geq 0. \tag{2.9}$$

Studying the state probabilities as functions of θ it can be shown in a similar way as above that, cf. (2.6),

$$\tilde{p}(\theta;\ \bar{n}) = p\left(\frac{\theta}{1+G-G\theta},\ \bar{n}\right) = \theta^{|\bar{n}|} \sum_{k=0}^{\infty} \theta^k b(k;\ \bar{n}), \tag{2.10}$$

where, cf. (2.8), (2.9),

$$b(0;\ \bar{0}) = 1; \qquad b(k;\ \bar{0}) = -\sum_{0<|\bar{n}|\leq k}\cdots\sum b(k-|\bar{n}|;\ \bar{n}), \quad k = 1, 2, \ldots, \tag{2.11}$$

and for $\bar{n} \neq \bar{0}$,

$$\begin{aligned}
&(1+G)\sum_{j=1}^{s} d_j(\bar{n}) I(n_j>0) b(k;\ \bar{n}) \\
&\quad = \sum_{j=1}^{s} a_j(\bar{n}-\bar{e}_j) I(n_j>0) b(k;\ \bar{n}-\bar{e}_j) \\
&\quad\quad + \sum_{j=1}^{s} d_j(\bar{n}+\bar{e}_j)\left[(1+G) b(k-1;\ \bar{n}+\bar{e}_j) I(k>0) - G b(k-2;\ \bar{n}+\bar{e}_j) I(k>1)\right] \\
&\quad\quad - \left[\sum_{j=1}^{s} a_j(\bar{n}) - G \sum_{j=1}^{s} d_j(\bar{n}) I(n_j>0)\right] b(k-1;\ \bar{n}) I(k>0), \quad k=0, 1, \ldots .
\end{aligned} \tag{2.12}$$

By assumption (A2) the constant G can be chosen in such a way that the radii of convergence of the power series (2.10) are at least one.

Remark. Note that the coefficient of $u(k;\ \bar{n})$ in (2.7) and that of $b(k;\ \bar{n})$ in (2.12) does not vanish when $\bar{n} \neq \bar{0}$ by assumption (A3).

3. Calculating moments

The coefficients of the power series expansions of the moments of queue length distributions can be directly calculated from those of the state probabilities. For $i,\ j = 1, \ldots, s$, let, cf. (2.9), (2.10),

$$E\{N_j\} = \sum_{k=1}^{\infty} f_j(k)\theta^k, \qquad E\{N_i N_j\} = \sum_{k=1}^{\infty} h_{ij}(k)\theta^k. \tag{3.1}$$

It is easily verified that for $i,\ j = 1, \ldots, s,\ k = 1, 2, \ldots,$

$$f_j(k) = \sum_{0\leq|\bar{n}|\leq k}\cdots\sum n_j b(k-|\bar{n}|;\ \bar{n}), \qquad h_{ij}(k) = \sum_{0\leq|\bar{n}|\leq k}\cdots\sum n_i n_j b(k-|\bar{n}|;\ \bar{n}). \tag{3.2}$$

As can be expected the power series of the moments converge very slowly for θ close to one (heavy traffic, assumption (A1)), because the nth moments will have a pole of (at most) order n at $\theta = 1$. The latter implies for instance that $f_j(k)$ tends to some constant and $h_{ij}(k)$ tends to some linearly increasing sequence as $k \to \infty$ ($i, j = 1, \ldots, s$). Therefore, we propose to use

$$\sum_{k=1}^{M} f_j(k)\theta^k + f_j(M)\frac{\theta^{M+1}}{1-\theta},$$

$$\sum_{k=1}^{M} h_{ij}(k)\theta^k + \left[h_{ij}(M) + \frac{h_{ij}(M) - h_{ij}(M-1)}{1-\theta}\right]\frac{\theta^{M+1}}{1-\theta}, \tag{3.3}$$

as approximations for $E\{N_j\}$ and $E\{N_iN_j\}$ respectively, when the coefficients of the power series expansions have been calculated up to the Mth power of θ. These extrapolations of the power series of moments strongly improve the rates of convergence to the exact values of the moments when θ is close to one. E.g. in a typical case with $\rho = 0.9$ the relative error between the approximate and the exact value of the average total number of customers in a system is 0.03 for $M = 60$ without the use of the extrapolation, and 0.001 for $M = 24$ with the aid of the extrapolation.

4. Examples

In this section two queueing models will be discussed for which the equations to be satisfied by the state probabilities fall within the general framework of (2.2). The interarrival times and the service times have negative exponential distributions in both models.

Example 1. The shortest queue model: customers choose with equal probabilities one of the shortest of s queues upon arrival; customers in queue j are served by server j with service rate r_j, $j = 1, \ldots, s$; $\Sigma_{j=1}^{s} r_j = 1$. For all states $\bar{n}$ in this model, for $j = 1, \ldots, s$,

$$d_j(\bar{n}) = r_j, \qquad a_j(\bar{n}) = 0 \quad \text{if } \exists i\ n_i < n_j,$$
$$= 1/\nu, \ \text{if } \forall i\ n_i \geq n_j, \quad \nu = \#\{i;\ n_i = n_j\}. \tag{4.1}$$

The special case $s = 2$, $r_1 = r_2 = \frac{1}{2}$, has been studied by several authors; Conolly [4] and Gertsbakh [6] provided numerical data. In [1] moments of the waiting time distribution have been calculated for more general cases ($s = 2, 3$), based on the present method. In Table 1 values of the average total number of customers in the system and the correlation between the number of customers in two queues are given for the symmetrical models with two and three servers for several values of the traffic intensity ρ.

Table 1
The shortest queue model

ρ	$s = 2$, $r_1 = r_2 = \frac{1}{2}$		$s = 3$, $r_1 = r_2 = r_3 = \frac{1}{3}$	
	$E\{N_1 + N_2\}$	Corr(N_1, N_2)	$E\{N_1 + N_2 + N_3\}$	Corr(N_1, N_2)
0.10	0.2035	0.09982	0.3010	0.09098
0.30	0.6864	0.3032	0.9579	0.1984
0.50	1.426	0.5257	1.867	0.3925
0.70	2.951	0.7659	3.589	0.6560
0.90	9.855	0.9647	10.75	0.9404
0.98	49.97	0.9984	51.00	0.9973

Table 2
The pre-emptive longer queue model ($s=2$, $\mu_1=4\mu_2$)

ρ	$\lambda_1=4\lambda_2$		$\lambda_1=\lambda_2$		$4\lambda_1=\lambda_2$	
	$E\{N_1\}$	$E\{N_2\}$	$E\{N_1\}$	$E\{N_2\}$	$E\{N_1\}$	$E\{N_2\}$
0.10	0.05588	0.05547	0.02283	0.08874	0.006823	0.1045
0.30	0.2211	0.2126	0.1061	0.3378	0.03656	0.4005
0.50	0.5264	0.4934	0.3283	0.7679	0.1455	0.9195
0.70	1.230	1.151	1.037	1.724	0.6286	2.073
0.90	4.621	4.470	5.343	6.314	4.743	7.417
0.98	24.65	24.46	32.44	33.54	34.67	38.17

Example 2. The pre-emptive longest queue model: s independent customer streams form s queues, one server provides service to a customer which is selected with equal probabilities from one of the longest queues; at each arrival instant the server reselects the customer to be served according to this rule. In this model we have for each state $\bar{n}$, for $j=1,\ldots,s$,

$$a_j(\bar{n})=\lambda_j, \qquad d_j(\bar{n})=0, \qquad \text{if } \exists i\ n_i>n_j,$$
$$=\mu_j/\nu, \ \text{if } \forall i\ n_i \leqslant n_j, \quad \nu=\#\{i;\ n_i=n_j\}. \tag{4.2}$$

The arrival rates are normalized by the relation

$$\sum_{j=1}^{s}\lambda_j/\mu_j=1, \tag{4.3}$$

in order to satisfy the assumption (A1). In the Tables 2 and 3 values of mean queue lengths and correlation between queue lengths are displayed for the case of two queues, a ratio between service rates of 4 : 1, and ratios between arrival rates of 4 : 1, 1 : 1, 1 : 4, respectively. Models with s queues and one server, without a pre-emptive service discipline, such as the longer queue model studied by Cohen [3], where a customer is selected from the longer of two queues and then served without interruption, and the alternating service model studied by Cohen and Boxma [2], are not contained in the framework of equations (2.2). But by extending the state space with a variable indicating the queue from which a customer is being served these models are readily adapted to the present procedure for calculating state probabilities and moments.

In general, the choice of G in (2.9), i.e. the choice of the conformal mapping, is difficult, because the state probabilities $p(\rho;\ \bar{n})$, as functions of ρ, may possess different singularities

Table 3
The pre-emptive longer queue model ($s=2$, $\mu_1=4\mu_2$)

ρ	$\lambda_1=4\lambda_2$		$\lambda_1=\lambda_2$		$4\lambda_1=\lambda_2$	
	Corr(N_1, N_2)	$E\{\|N_1-N_2\|\}$	Corr(N_1, N_2)	$E\{\|N_1-N_2\|\}$	Corr(N_1, N_2)	$E\{\|N_1-N_2\|\}$
0.10	0.06604	0.09978	0.06451	0.1035	0.04175	0.1084
0.30	0.2629	0.2964	0.3107	0.3310	0.2252	0.3904
0.50	0.5062	0.4901	0.6055	0.5838	0.4765	0.8089
0.70	0.7612	0.6858	0.8402	0.8603	0.7131	1.482
0.90	0.9638	0.8909	0.9793	1.159	0.9307	2.708
0.98	0.9984	0.9777	0.9991	1.284	0.9957	3.525

inside the unit disk for different states $\bar{n}$. G should be at least so large that all singularities of all state probabilities and moments (which are unknown in general) are mapped outside the unit disk (when all values of the traffic intensity ρ, $0 < \rho < 1$, are considered). However, choosing G much larger often leads to slower convergence of the power series, because other singularities (e.g. on the interval $(1, \infty)$) are mapped closer to the unit disk as G increases. In some cases the power series expansions of moments did converge for a specific value of G, while those of several state probabilities did not. A value which often gives good results is $G = 1$.

It seems that limitations on application of the method are caused more by the amount of memory space required to store the coefficients of the power series than by the amount of computation time. When power series are needed up to the Mth power of θ coefficients $b(k; \bar{n})$ have to be calculated for all k and $\bar{n}$ with $k + |\bar{n}| \leqslant M$. To avoid reservation of superfluous memory space points $(k, \bar{n})$ could be mapped onto the integers, e.g. in the following way

$$(k, n_1, \ldots, n_s) \to \sum_{j=0}^{s} \binom{k + j + \sum_{i=1}^{j} n_i}{j+1}.$$

Then $\binom{M+s+1}{M}$ memory positions are required instead of $(M+1)^{s+1}$.

Remark. Assumption (A1) requires that the stability condition of a system is known in order to normalize the arrival rates $a_j(\bar{n})$ for all states $\bar{n}$ and $j = 1, \ldots, s$. If the stability condition is not known, it can be obtained from the coefficients of the power series expansion of the average number of customers in the system which will tend to a geometrical sequence. In a similar way inspection of the coefficients of the power series expansions of the state probabilities will give information about the radii of convergence of these series. This information can be used to choose an appropriate conformal mapping.

References

[1] J.P.C. Blanc, A note on waiting times in systems with queues in parallel, *J. Appl. Probab.*, to appear.

[2] J.W. Cohen, O.J. Boxma, *Boundary Value Problems in Queueing System Analysis* (North-Holland, Amsterdam, 1983).

[3] J.W. Cohen, A two queue, one server model with priority for the longer queue, University of Utrecht, Dept. of Math., preprint 379, 1985.

[4] B.W. Conolly, The autostrada queueing problem, *J. Appl. Probab.* **21** (1984) 394–403.

[5] G. Fayolle and R. Iasnogorodski, Two coupled processors: the reduction to a Riemann–Hilbert problem, *Z. Wahrscheinlichkeitsth.* *47* (1979) 325–351.

[6] I. Gertsbakh, The shorter queue problem: a numerical study using the matrix-geometric solution, *Eur. J. Oper. Res.* *15* (1984), 374–381.

[7] M.S. Keane, G. Hooghiemstra and S. van de Ree, Private communication.