ANALYSIS OF THE SYMMETRIC SHORTEST QUEUE PROBLEM

I. J. B. F. ADAN
J. WESSELS

Department of Mathematics and Computer Science
Eindhoven University of Technology
P.O. Box 513, 5600 MB - Eindhoven
The Netherlands

W. H. M. ZIJM

Department of Mathematics and Computer Science
Eindhoven University of Technology
P.O. Box 513, 5600 MB - Eindhoven
The Netherlands

and

Nederlandse Philips Bedrijven B.V.
Centre for Quantitative Methods, Building HCM-721
P.O. Box 218, 5600 MD - Eindhoven
The Netherlands

ABSTRACT

In this paper we study a system consisting of two identical servers, each with exponentially distributed service times. Jobs arrive according to a Poisson stream. On arrival a job joins the shortest queue and in case both queues have equal lengths, he joins either queue with probability ½. By using a compensation method, we show that the stationary queue length distribution can be expressed as an infinite linear combination of product forms. Explicit relations are found for these product forms, as well as for the coefficients in the linear combination. These analytic results offer an elegant and efficient numerical algorithm, with effective bounds on the error of each partial sum.

Key Words: difference equation, product form, similar queues in parallel, stationary queue length distribution.

1. Introduction

Consider a system consisting of two identical servers, each with exponentially distributed service times. Jobs arrive according to a Poisson stream. On arrival a job joins the shortest queue and in case both queues have equal length, he joins either queue with probability ½. This problem is known as the symmetric shortest queue problem and has been addressed by many authors. Haight [17] originally introduced the problem. Kingman [21] and Flatto and McKean [11] treated the problem by using a generating function analysis. They show that the generating function for the equilibrium distribution of the lengths of the two queues is a meromorphic function. Then, by decomposition of the generating function into partial fractions, it follows that the equilibrium probabilities can be expressed as an infinite linear combination of product forms. However, the decomposition leads to cumbersome formulae for the equilibrium probabilities. Another analytic approach is found in Cohen and Boxma [7] and Fayolle and Iasnogorodski [9],[20],[10]. They show that the analysis of the symmetric shortest queue problem can be reduced to that of a Riemann-Hilbert boundary value problem. These approaches do not lead to an explicit characterization of the equilibrium probabilities.

The approach presented in this paper is not based on a generating function analysis. Instead the probabilities are found directly from the equilibrium equations. The solution method is initialized by inserting a product form, describing the asymptotic behaviour of the probabilities, and next consists of adding on product forms so as to compensate for the error of its preceding term on one of the boundaries of the state space. The main improvement to the analytic results of Kingman [21] and Flatto and McKean [11] is that our method yields explicit relations for the coefficients in the infinite linear combination of product forms and thereby an explicit characterization of the equilibrium probabilities. Moreover, the compensation idea sheds new light on the existence of this type of solution.

So far, the available analytic results, though mathematically elegant, offered no practical means for evaluating many of the performance characteristics and therefore didn't close the matter in this aspect. For this reason, many numerical studies appeared on the present problem. Most studies, however, deal with the evaluation of *approximating* models. For instance, Gertsbakh [14], Grassmann [15], Rao and Posner [23] and Conolly [8] treated the shortest queue problem by truncating one or more state variables. Using linear programming, Halfin [18] obtained upper and lower bounds for the queue length distribution. Foschini and Salz [12] obtained heavy traffic diffusion approximations for the queue length distribution. Knessl, Matkowsky, Schuss and Tier [22] derived asymptotic expressions of the queue length distribution. These studies are all restricted to systems with two parallel queues. Hooghiemstra, Keane and Van de Ree [19] developed a power series method to calculate the stationary queue length distribution for fairly general multidimensional exponential queueing systems. Their method is not restricted to systems with two queues, but applies equally well to systems with more queues. So far as the shortest queue problem is concerned, Blanc [4][5] reported that the power series method is numerically satisfactory for the shortest queue system with up to 25 parallel queues. The theoretical foundation of this method is, however, still incomplete. Finally, a common disadvantage of the numerical methods mentioned is that in general no error bounds can be given.

As already mentioned, the compensation method yields explicit relations for the product forms, as well as for their coefficients and hence for the equilibrium probabilities. These analytic results are exploited to construct an efficient numerical algorithm, with tight bounds on the error of each partial sum. Also, expressions are obtained for the mean and second moment of the waiting time, which are suitable for numerical evaluation. *These algorithms apply to the exact model.*

The paper is organized as follows. In Section 2 we present the equilibrium equations. In the next section, we develop, step by step, the compensation procedure. Section 4 presents the formal definition of the compensation procedure and the main result, which states that the probabilities can be expressed as an infinite linear combination of product forms. In the following two sections we complete the proof of the main result. In Section 7 we derive an explicit form for the normalizing constant. Section 8 extends the asymptotic expressions for the probabilities, obtained by Kingman [21], Flatto and McKean [11] and Knessl, Matkowsky, Schuss and Tier [22]. Section 9 presents numerical results and the final section is devoted to comments and extensions.

2. Equilibrium Equations

For simplicity of notation the exponential servers have service times with unit mean and the Poisson arrival process has a rate 2ρ with $0 < \rho < 1$. The parallel queue system can be represented by a continuous time Markov process, whose state space consists of the pairs (m, n), $m, n = 0, 1, \ldots$ where m and n are the lengths of the two queues. The transition rates in the upper wedge $n \geq m$ are illustrated in figure 1a, the rates in the lower wedge $n \leq m$ follow by reflection in the diagonal.

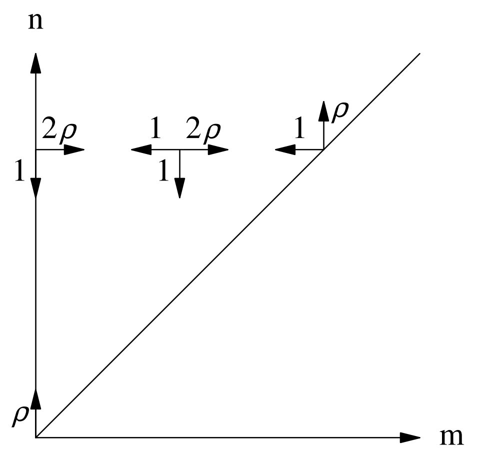

Figure 1a: m-n transition rate diagram

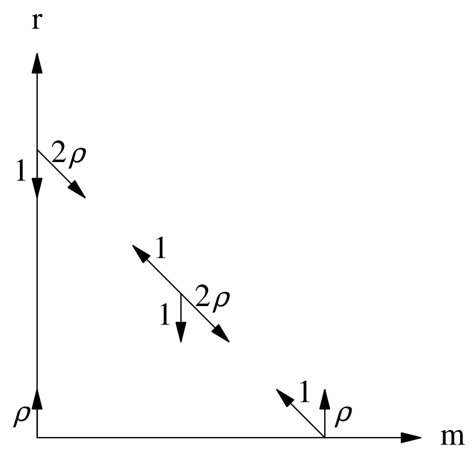

Figure 1b: m-r transition rate diagram

Let $\{p_{m,n}\}$ be the equilibrium distribution of the lengths of the two queues. By symmetry $p_{m,n} = p_{n,m}$, for all values of m and n. Therefore, we can restrict the analysis to the probabilities $p_{m,n}$ in the wedge $n \geq m$. The equilibrium equations state that for all $n > m$:

$$p_{m,n}\, 2(\rho+1) = p_{m-1,n}\, 2\rho + p_{m,n+1} + p_{m+1,n} \qquad \text{if } m > 0,\ n > m+1$$

$$p_{m,m+1}\, 2\,(\rho+1) = p_{m-1,m+1}\, 2\rho + p_{m,m+2} + p_{m+1,m+1} + p_{m,m}\, \rho \qquad \text{if } m > 0,\ n = m+1$$

$$p_{0,n}\, (2\rho+1) = p_{0,n+1} + p_{1,n} \qquad \text{if } n > 1$$

$$p_{0,1}(2\rho+1) = p_{0,2} + p_{1,1} + p_{0,0}\, \rho$$

and, by symmetry, the equations on $n = m$ simplify to

$$p_{m,m}\, (\rho+1) = p_{m-1,m} 2\rho + p_{m,m+1} \qquad \text{if } m > 0$$

$$p_{0,0}\, \rho = p_{0,1}$$

The probabilities $p_{m,m}$ can be eliminated in the equations on the subdiagonal $n = m + 1$ by substituting the equations on the diagonal. Then the analysis can be further restricted to the probabilities $p_{m,n}$ in the upper wedge $n > m$. The equations on the diagonal are used henceforth as *definition* for the probabilities $p_{m,m}$. For the analysis that follows, it is preferable to have the coordinate axes along the boundaries of the upper wedge. Therefore, instead of the coordinates m and n, we will work with the coordinates m and $r = n - m$. Then the upper wedge $n \geq m$ in the m-n plane is transformed into the first quadrant in the m-r plane. In Figure 1b we display the transition rate diagram for the new coordinates. Further, set for all $m \geq 0$ and $r \geq 0$,

$$q_{m,r} = p_{m,m+r}.$$

Restating the equilibrium equations in terms of $q_{m,r}$, we get that for all $m \geq 0$ and $r \geq 1$,

$$q_{m,r}\, 2(\rho+1) = q_{m-1,r+1}\, 2\rho + q_{m,r+1} + q_{m+1,r-1} \qquad \text{if } m > 0,\ r > 1 \quad (1)$$

$$q_{m,1}\, 2(\rho+1) = q_{m-1,2}\, 2\rho + q_{m,2} + \qquad \text{if } m > 0,\ r = 1 \quad (2)$$

$$+ (q_{m,1}\, 2\rho + q_{m+1,1})\, \frac{1}{\rho+1} + (q_{m-1,1}\, 2\rho + q_{m,1})\, \frac{\rho}{\rho+1}$$

$$q_{0,r}\, (2\rho+1) = q_{0,r+1} + q_{1,r-1} \qquad \text{if } r > 1 \quad (3)$$

$$q_{0,1}(2\rho+1) = q_{0,2} + (q_{0,1}\, 2\rho + q_{1,1})\, \frac{1}{\rho+1} + q_{0,1} \quad (4)$$

and for all $m \geq 0$ and $r = 0$,

$$q_{m,0}\, (\rho+1) = q_{m-1,1} 2\rho + q_{m,1} \qquad \text{if } m > 0 \quad (5)$$

$$q_{0,0}\ \rho = q_{0,1} \tag{6}$$

As we already stated, the analysis can be restricted to the set $\{(m, r), m \geq 0, r \geq 1\}$ and the equations on the axis $r = 0$ can be used as definition for the probabilities $q_{m,0}$. We will show that there exist parameters α_i and β_i and coefficients c_i such that for all $m \geq 0$ and $r \geq 1$,

$$q_{m,r} = \sum_{i=0}^{\infty} c_i\ \alpha_i^m\ \beta_i^r .$$

Throughout the analysis we use the trivial, but vital property that the equations, on which the analysis is based, are *linear*, i.e. if two functions satisfy an equation, then any linear combination also satisfies that equation.

3. The Compensation Procedure

The objective in this section is to study the structure of the equilibrium probabilities. Particularly, we investigate whether the probabilities have some kind of separable structure. Obviously, the equations (1)-(4) don't allow a separable solution of the form $q_{m,r} = \alpha^m\ \beta^r$. However, numerical experiments indicate that there exist α and β such that, for some K,

$$q_{m,r} \sim K\ \alpha^m\ \beta^r \quad \text{as } m \to \infty \text{ and } r \geq 1. \tag{7}$$

This is illustrated in Figure 2 for the special case $\rho = 0.5$. In Figure 2a we display the ratio of $q_{m,r}$ in the m direction, which yields, at least for large m, the parameter α. In Figure 2b we display the ratio of $q_{m,r}$ in the r direction, yielding the parameter β. At this stage of the analysis, the best we can do, is calculating *approximations* for the probabilities $q_{m,r}$ by solving a *finite* capacity shortest queue system exactly, i.e. by means of a Markov chain analysis. In the example we computed the equilibrium distribution for a system where each queue has a maximum capacity of 15 jobs, which approximates well the infinite capacity system in case $\rho = 0.5$.

Clearly, as $\rho = 0.5$ we have $q_{m,r} \sim K\ 0.25^m\ 0.1^r$ for some K, which holds even for moderate m. The question is, what are in general the parameters α and β? Intuitively, α stands for the ratio of the probability that there are $m + 2$ and m jobs in the system. So a reasonable choice seems $\alpha = \rho^2$, which is supported by the numerical example. The parameter β follows by observing that the form $\alpha^m\ \beta^r$ has to satisfy equation (1) in the interior of the set $\{(m, r), m \geq 0, r \geq 1\}$. Inserting this form into (1) and dividing both sides by the common term $\alpha^{m-1}\ \beta^{r-1}$ we get a quadratic equation for the unknown β. This results in the following lemma.

r ↑	0	1	2	3	4	5
6	0.19	0.24	0.25	0.25	0.25	0.25
5	0.19	0.24	0.25	0.25	0.25	0.25
4	0.19	0.24	0.25	0.25	0.25	0.25
3	0.19	0.24	0.25	0.25	0.25	0.25
2	0.20	0.24	0.25	0.25	0.25	0.25
1	0.28	0.25	0.25	0.25	0.25	0.25
0						

$m \rightarrow$

Figure 2a: $q_{m+1,r} / q_{m,r}$ *for* $\rho = 0.5$

r ↑	0	1	2	3	4	5
6	0.10	0.10	0.10	0.10	0.10	0.10
5	0.10	0.10	0.10	0.10	0.10	0.10
4	0.10	0.10	0.10	0.10	0.10	0.10
3	0.10	0.10	0.10	0.10	0.10	0.10
2	0.11	0.10	0.10	0.10	0.10	0.10
1	0.15	0.11	0.10	0.10	0.10	0.10
0						

$m \rightarrow$

Figure 2b: $q_{m,r+1} / q_{m,r}$ *for* $\rho = 0.5$

Lemma 1: The form $\alpha^m \ \beta^r$ is a solution of equation (1) if and only if α and β satisfy

$$\alpha \ \beta \ 2(\rho + 1) = \beta^2 \ 2\rho + \alpha \ \beta^2 + \alpha^2 \ . \tag{8}$$

By Lemma 1, we obtain two roots $\beta = \rho$ and $\beta = \rho^2 / (2 + \rho)$ for fixed $\alpha = \rho^2$. The root $\beta = \rho$ yields the asymptotic solution $q_{m,r} \sim K \ \rho^{2m} \ \rho^r$ for some K, which corresponds to the equilibrium distribution of two independent $M|M|1$ queues, each with workload ρ. It is very unlikely that the equilibrium distribution of the shortest queue problem behaves asymptotically like this distribution. Therefore, the only reasonable choice is $\beta = \rho^2 / (2 + \rho)$, which is also supported by the numerical example. Hence, we empirically find that, for some K,

$$q_{m,r} \sim K \ \rho^{2m} \left(\frac{\rho^2}{2 + \rho} \right)^r \quad \text{as } m \rightarrow \infty \text{ and } r \geq 1. \tag{9}$$

Actually, Kingman ([21], Theorem 5) and Flatto and McKean ([11], Section 3) gave a rigorous proof for this asymptotic result.

Let $\alpha_0 = \rho^2$ and $\beta_0 = \rho^2 / (2 + \rho)$. As is illustrated in Figure 2 for the special case $\rho = 0.5$, the asymptotic solution $\alpha_0^m \ \beta_0^r$ perfectly describes the behaviour of the equilibrium probabilities in the interior of the set $\{(m, r), m \geq 0, r \geq 1\}$ as well as at the boundary $r = 1$, but it does not capture the behaviour near the boundary $m = 0$. One easily verifies that $\alpha_0^m \ \beta_0^r$ indeed satisfies equation (2) on the boundary $r = 1$ and that it violates equation (3) on the boundary $m = 0$. Obviously, we can further improve this asymptotic solution by adding a term to correct for the error on the boundary $m = 0$. For large m this correction term should be small compared to the term $\alpha_0^m \ \beta_0^r$ in order to avoid that it spoils the behaviour for large m.

Form the linear combination $\alpha_0^m \beta_0^r + c_0 \alpha^m \beta^r$. We try to choose c_0, α and β such that this linear combination satisfies equation (3) and (1). Inserting it into (3) yields for all $r > 1$

$$(\beta_0^r + c_0 \beta^r)(2\rho + 1) = (\beta_0^{r+1} + c_0 \beta^{r+1}) + (\alpha_0 \beta_0^{r-1} + c_0 \alpha \beta^{r-1}) .$$

Since this must hold for all $r > 1$, we have to set $\beta = \beta_0$. Further we want $\alpha^m \beta_0^r$ to satisfy equation (1). By virtue of Lemma 1, there are two α's such that $\alpha^m \beta_0^r$ satisfies equation (1), namely $\alpha_0 = \rho^2$ and $\alpha_1 = 2\rho^3 / (2 + \rho)^2$. So we have to set $\alpha = \alpha_1$. Then for any c_0, the linear combination $\alpha_0^m \beta_0^r + c_0 \alpha_1^m \beta_0^r$ satisfies equation (1), because equation (1) is linear. Finally, dividing the above equation by the common term β_0^{r-1} gives an equation for the unknown c_0. Hence we can choose the coefficient c_0 such that the linear combination also satisfies equation (3). In general, the result of this procedure can be stated as

Lemma 2: Let x_1 and x_2 be the roots of the quadratic equation (8) for fixed β. Then the linear combination $k_1 x_1^m \beta^r + k_2 x_2^m \beta^r$ satisfies the equations (1) and (3) if k_1 and k_2 satisfy

$$k_2 = -\frac{x_2 - \beta}{x_1 - \beta} k_1. \tag{10}$$

Proof: By virtue of Lemma 1, the forms $x_1^m \beta^r$ and $x_2^m \beta^r$ both satisfy equation (1). Since equation (1) is linear, any linear combination also satisfies (1). Inserting the linear combination $k_1 x_1^m \beta^r + k_2 x_2^m \beta^r$ into (3) and dividing by the common term β^{r-1} yields

$$(k_1 + k_2) \beta (2\rho + 1) = (k_1 + k_2) \beta^2 + k_1 x_1 + k_2 x_2,$$

which can be rewritten as

$$k_2 = -\frac{\beta (2\rho + 1) - \beta^2 - x_1}{\beta (2\rho + 1) - \beta^2 - x_2} k_1. \tag{11}$$

Since x_1 and x_2 are the roots of the quadratic equation (8),

$$x_1 + x_2 = \beta \, 2(\rho + 1) - \beta^2.$$

Substituting that equality into (11) yields (10). $\square$

Applying Lemma 2 with $x_1 = \alpha_0$, $x_2 = \alpha_1$, $\beta = \beta_0$, $k_1 = 1$ and $k_2 = c_0$, yields that

$$c_0 = -\frac{\alpha_1 - \beta_0}{\alpha_0 - \beta_0}.$$

Then $\alpha_0^m\ \beta_0^r + c_0\ \alpha_1^m\ \beta_0^r$ satisfies the equations (1) and (3). For the special case of $\rho = 0.5$, we display in Figure 3 the same ratios as in Figure 2 for this asymptotic solution.

↑ r						
6	0.19	0.24	0.25	0.25	0.25	0.25
5	0.19	0.24	0.25	0.25	0.25	0.25
4	0.19	0.24	0.25	0.25	0.25	0.25
3	0.19	0.24	0.25	0.25	0.25	0.25
2	0.19	0.24	0.25	0.25	0.25	0.25
1	0.19	0.24	0.25	0.25	0.25	0.25
0						
	0	1	2	3	4	5
					$m \rightarrow$	

Figure 3a: $\dfrac{(\alpha_0^{m+1} + c_0\alpha_1^{m+1})\beta_0^r}{(\alpha_0^m + c_0\alpha_1^m)\beta_0^r}$

↑ r						
6	0.10	0.10	0.10	0.10	0.10	0.10
5	0.10	0.10	0.10	0.10	0.10	0.10
4	0.10	0.10	0.10	0.10	0.10	0.10
3	0.10	0.10	0.10	0.10	0.10	0.10
2	0.10	0.10	0.10	0.10	0.10	0.10
1	0.10	0.10	0.10	0.10	0.10	0.10
0						
	0	1	2	3	4	5
					$m \rightarrow$	

Figure 3b: $\dfrac{(\alpha_0^m + c_0\alpha_1^m)\beta_0^{r+1}}{(\alpha_0^m + c_0\alpha_1^m)\beta_0^r}$

Comparing Figures 2 and 3, we see that this refinement also captures the behaviour of the equilibrium probabilities near the boundary $m = 0$. Hence, for some K,

$$q_{m,r} \sim K\ (\alpha_0^m\ \beta_0^r + c_0\ \alpha_1^m\ \beta_0^r) \quad \text{as } m + r \rightarrow \infty \text{ and } r \geq 1. \tag{12}$$

Flatto and McKean ([11], Section 3) proved this statement, which is stronger than (9). We added an extra term to compensate for the error on the boundary $m = 0$. On the other hand, we introduced a new error on the boundary $r = 1$, since the extra term violates equation (2). Since $\alpha_1 < \alpha_0$, the term $\alpha_1^m\ \beta_0^r$ is very small compared to $\alpha_0^m\ \beta_0^r$ even for moderate m. Therefore its disturbing effect near the boundary $r = 1$ is practically negligible. However, we can compensate for this second order error on the boundary $r = 1$ in the same way as we did on the boundary $m = 0$, by adding another correction term.

Form the linear combination $\alpha_0^m\ \beta_0^r + c_0\ \alpha_1^m\ \beta_0^r + d_1\ \alpha^m\ \beta^r$. The term $\alpha_0^m\ \beta_0^r$ already satisfies the equations (2) and (1) and we try to choose d_1, α and β such that the linear combination $c_0\ \alpha_1^m\ \beta_0^r + d_1\ \alpha^m\ \beta^r$ also satisfies (2) and (1). Inserting it into (2) gives for all $m > 0$

$$(c_0\ \alpha_1^m\ \beta_0 + d_1\ \alpha^m\ \beta)\ 2(\rho + 1)$$

$$= (c_0\ \alpha_1^{m-1}\ \beta_0^2 + d_1\ \alpha^{m-1}\ \beta^2)\ 2\rho + c_0\ \alpha_1^m\ \beta_0^2 + d_1\ \alpha^m\ \beta^2$$

$$+ [(c_0\ \alpha_1^m\ \beta_0 + d_1\ \alpha^m\ \beta)\ 2\rho + c_0\ \alpha_1^{m+1}\ \beta_0 + d_1\ \alpha^{m+1}\ \beta]\ \frac{1}{\rho+1}$$

$$+ [(c_0\ \alpha_1^{m-1}\ \beta_0 + d_1\ \alpha^{m-1}\ \beta)\ 2\rho + c_0\ \alpha_1^m\ \beta_0 + d_1\ \alpha^m\ \beta]\ \frac{\rho}{\rho+1}\ .$$

Since this must hold for all $m > 0$, we have to set $\alpha = \alpha_1$. Furthermore we want $\alpha_1^m\ \beta^r$ to satisfy equation (1). By virtue of Lemma 1 there are there are two β's such that $\alpha_1^m\ \beta^r$ satisfies equation (1), namely $\beta_0 = \rho^2 / (2+\rho)$ and $\beta_1 = \rho^3 / ((2+\rho)(2+2\rho+\rho^2))$. So we have to set $\beta = \beta_1$. Then for any d_1, the linear combination $c_0\ \alpha_1^m\ \beta_0^r + d_1\ \alpha_1^m\ \beta_1^r$ satisfies equation (1). Finally, dividing the above equation by the common term α_1^{m-1} yields an equation for the unknown d_1. Hence we can choose d_1 such that the linear combination also satisfies (2). In general, we have

Lemma 3: Let y_1 and y_2 be the roots of the quadratic equation (8) for fixed α. Then the linear combination $k_1\ \alpha^m\ y_1^r + k_2\ \alpha^m\ y_2^r$ satisfies the equations (1) and (2) if k_1 and k_2 satisfy

$$k_2 = -\ \frac{(\alpha+\rho)/y_2 - (\rho+1)}{(\alpha+\rho)/y_1 - (\rho+1)}\ k_1. \tag{13}$$

Proof: By virtue of Lemma 1 both $\alpha^r\ y_1^r$ and $\alpha^r\ y_2^r$ satisfy (1) and by linearity, also any linear combination. Inserting the linear combination $k_1\ \alpha^m\ y_1^r + k_2\ \alpha^m\ y_2^r$ into (2) and dividing both sides by the common term α^{m-1} yields

$$(k_1\ \alpha\ y_1 + k_2\ \alpha\ y_2)\ 2(\rho+1) = (k_1\ y_1^2 + k_2\ y_2^2)\ 2\rho + k_1\ \alpha\ y_1^2 + k_2\ \alpha\ y_2^2$$

$$+ [(k_1\ \alpha\ y_1 + k_2\ \alpha\ y_2)\ 2\rho + k_1\ \alpha^2\ y_1 + k_2\ \alpha^2\ y_2]\ \frac{1}{\rho+1}$$

$$+ [(k_1\ y_1 + k_2\ y_2)\ 2\rho + k_1\ \alpha\ y_1 + k_2\ \alpha\ y_2]\ \frac{\rho}{\rho+1}\ .$$

By inserting equation (8) this reduces to

$$k_1\ \alpha^2 + k_2\ \alpha^2 = [(k_1\ \alpha\ y_1 + k_2\ \alpha\ y_2)\ 2\rho + k_1\ \alpha^2\ y_1 + k_2\ \alpha^2\ y_2]\ \frac{1}{\rho+1}$$

$$+ [(k_1\ y_1 + k_2\ y_2)\ 2\rho + k_1\ \alpha\ y_1 + k_2\ \alpha\ y_2]\ \frac{\rho}{\rho+1}\ .$$

Hence

$$k_2 = -\,\frac{y_1\,(\alpha + 2\rho)(\alpha + \rho)\,/\,(\rho + 1) - \alpha^2}{y_2\,(\alpha + 2\rho)(\alpha + \rho)\,/\,(\rho + 1) - \alpha^2}\;k_1. \tag{14}$$

Since y_1 and y_2 are the roots of the quadratic equation (8),

$$y_1\; y_2\;(\alpha + 2\rho) = \alpha^2\;.$$

Using this relation to rewrite y_1 and y_2 in (14), yields relation (13). $\square$

Applying Lemma 3 with $y_1 = \beta_0,\; y_2 = \beta_1,\; \alpha = \alpha_1,\; k_1 = c_0$ and $k_2 = d_1$, yields that

$$d_1 = -\,\frac{(\alpha_1 + \rho)\,/\,\beta_1 - (\rho + 1)}{(\alpha_1 + \rho)\,/\,\beta_0 - (\rho + 1)}\;c_0.$$

Then the linear combination $\alpha_0^m\;\beta_0^r + c_0\;\alpha_1^m\;\beta_0^r + d_1\;\alpha_1^m\;\beta_1^r$ satisfies both equations (1) and (2). Now we compensated the error on the boundary $r = 1$, but we introduced a new one on the boundary $m = 0$, since the compensating term $\alpha_1^m\;\beta_1^r$ violates equation (3). But it is clear how to continue this compensating procedure: it consists of adding on terms so as to compensate alternatingly for the error on the boundary $m = 0$, according to Lemma 2, and for the error on the boundary $r = 1$, according to Lemma 3. The final solution consists of an infinite series of compensation terms. It is formally defined in the next section.

4. Formal Definition of the Compensation Procedure and the Main Theorem

The final solution is an infinite linear combination of terms of the form $\alpha^m\;\beta^r$. Below we first define the parameters α_i and β_i and next the coefficients of the linear combinations. For the initial values $\alpha_0 = \rho^2$ and $\beta_0 = \rho^2\,/\,(2 + \rho)$, define the sequence

$$\alpha_0 \nearrow \beta_0 \searrow \alpha_1 \nearrow \beta_1 \searrow \alpha_2 \nearrow \beta_2 \searrow \;\ldots$$

such that for all $i = 0, 1, 2, \ldots$, the numbers α_i and α_{i+1} are the roots of the quadratic equation (8) for fixed $\beta = \beta_i$ and β_i and β_{i+1} are the roots of the quadratic equation (8) for fixed $\alpha = \alpha_{i+1}$. Therefore the numbers α_i and α_{i+1} satisfy the relations

$$\alpha_i\;\alpha_{i+1} = 2\rho\;\beta_i^2, \tag{15}$$

$$\alpha_i + \alpha_{i+1} = \beta_i\;2(\rho + 1) - \beta_i^2\;, \tag{16}$$

and β_i and β_{i+1} satisfy

$$\beta_i\ \beta_{i+1} = \alpha_{i+1}^2 / (2\rho + \alpha_{i+1})\ , \tag{17}$$

$$\beta_i + \beta_{i+1} = \alpha_{i+1}\ 2(\rho + 1) / (2\rho + \alpha_{i+1})\ . \tag{18}$$

Using the relations (15) and (17), it follows, by induction, that for all i the numbers α_i and β_i are positive. The relations (15) and (17) provide a simple recursive scheme to produce α_i and β_i, but we can say more, since α_i and β_i can be solved explicitly. Combining (17) and (18) yields for all $i \geq 0$, that

$$\frac{1}{\beta_i} + \frac{1}{\beta_{i+1}} = \frac{2(\rho + 1)}{\alpha_{i+1}}\ . \tag{19}$$

Adding that relation for $i - 1$ and i and next eliminating α_i and α_{i+1} by inserting (15) and (16), yields for all $i \geq 1$,

$$\frac{1}{\beta_{i-1}} + \frac{2}{\beta_i} + \frac{1}{\beta_{i+1}} = \frac{\rho + 1}{\rho} \left(\frac{2(\rho + 1)}{\beta_i} - 1 \right).$$

This is an inhomogeneous second order recursion relation for $\{1 / \beta_i\}$ with initial values $1 / \beta_0 = (2 + \rho) / \rho^2$ and $1 / \beta_1 = (2 + \rho)(2 + 2\rho + \rho^2) / \rho^3$, whose solution is routine. Then the numbers α_i follow from (19). Hence, we obtain (cf. Kingman [21], Lemma 3),

Lemma 4: $\alpha_0 = \rho^2$ and for all $i = 0, 1, 2, \ldots$

$$2(\rho + 1) / \alpha_{i+1} = 2A + B\ (1 + \lambda)\ \lambda^i + C\ (1 + \lambda^{-1})\ \lambda^{-i}\ ,$$

$$1 / \beta_i = A + B\ \lambda^i + C\ \lambda^{-i}\ ,$$

where

$$\lambda = (\rho + 1 - \sqrt{\rho^2 + 1}) / (\rho + 1 + \sqrt{\rho^2 + 1})\ ,$$

$$A = (1 + \rho) / 2(1 + \rho^2)\ ,$$

and B and C follow from the initial values

$$1 / \beta_0 = (2 + \rho) / \rho^2\ ,$$

$$1 / \beta_1 = (2 + \rho)(2 + 2\rho + \rho^2) / \rho^3\ .$$

By Lemma 1, all solutions $\alpha_i^m \ \beta_i^r$ and $\alpha_{i+1}^m \ \beta_i^r$ satisfy equation (1) and by linearity, any linear combination of these solutions also satisfies (1). Now, for all $m \geq 0$ and $r \geq 1$, define the final solution $x_{m,r}$ as

$$x_{m,r} = \sum_{i=0}^{\infty} d_i \ (\alpha_i^m + c_i \ \alpha_{i+1}^m) \ \beta_i^r \tag{20}$$

$$= d_0 \ \alpha_0^m \ \beta_0^r + \sum_{i=0}^{\infty} (d_i \ c_i \ \beta_i^r + d_{i+1} \ \beta_{i+1}^r) \ \alpha_{i+1}^m,$$

where in the first sum we formed pairs with a common factor β_i and in the second one with a common factor α_{i+1}. Put $d_0 = 1$ and successively generate the coefficients c_i and d_{i+1} such that $(\alpha_i^m + c_i \ \alpha_{i+1}^m) \ \beta_i^r$ satisfies equation (3) on the boundary $m = 0$ and $(d_i \ c_i \ \beta_i^r + d_{i+1} \ \beta_{i+1}^r) \ \alpha_i^m$ satisfies equation (2) on the boundary $r = 1$. By Lemma 2 and 3, this yields for all $i = 0, 1, \ldots$

$$c_i = - \frac{\alpha_{i+1} - \beta_i}{\alpha_i - \beta_i} , \tag{21}$$

$$d_{i+1} = - \frac{(\alpha_{i+1} + \rho) / \beta_{i+1} - (\rho + 1)}{(\alpha_{i+1} + \rho) / \beta_i - (\rho + 1)} \ c_i \ d_i. \tag{22}$$

$$= \ \ldots \ = (-1)^{i+1} \prod_{j=0}^{i} \frac{(\alpha_{j+1} + \rho) / \beta_{j+1} - (\rho + 1)}{(\alpha_{j+1} + \rho) / \beta_j - (\rho + 1)} \ c_j \ .$$

The numbers $x_{m,0}$ are defined by the equilibrium equations (5) and (6), yielding

$$x_{m,0} = (x_{m-1,1} \ 2\rho + x_{m,1}) / (\rho + 1) \quad \text{for} \ \ m > 0 \ , \quad x_{0,0} = x_{0,1} / \rho \ . \tag{23}$$

The following theorem establishes our main result: up to a normalizing constant, the solution $\{ x_{m,r} \}$ is the equilibrium distribution $\{ q_{m,r} \}$.

Theorem: For all $m \geq 0$ and $r \geq 0$

$$q_{m,r} = C^{-1} \ x_{m,r} \ ,$$

where the normalizing constant C satisfies

$$C = \frac{\rho \ (2 + \rho)}{2 \ (1 - \rho^2) \ (2 - \rho)} \ .$$

In the following sections we shall prove the theorem. First, we show that the series (20), which define the numbers $x_{m,r}$, converge absolutely. Next we prove that $\{x_{m,r}\}$ is a positive and convergent solution, that is,

$$x_{m,r} > 0 \quad \text{and} \quad C = 2 \sum_{m=0}^{\infty} \sum_{r=1}^{\infty} x_{m,r} + \sum_{m=0}^{\infty} x_{m,0} < \infty.$$

Once the above is established, we can conclude that $\{x_{m,r}\}$ forms a well defined, nonnull and convergent solution, satisfying the equations (1), (2) and (3), and by definition, the equations (5) and (6). The remaining equation in (0, 1) is also satisfied, for summing over all other equations yields the desired equation. By a result of Foster ([13], Theorem 1), this proves that the shortest queue system is ergodic. Since the equilibrium distribution for an ergodic system is unique, $\{x_{m,r}\}$ can be normalized to produce the equilibrium distribution. We finally show that the normalizing constant C satisfies the explicit expression in the theorem.

5. Asymptotics

To prove the convergence of the series (20) we need information about the behaviour of α_i, β_i, c_i and d_i. Instead of exploiting the explicit forms in Lemma 4, we obtain the necessary information, in a relatively easy way, from the behaviour of the ratios α_i / β_i and α_{i+1} / β_i (recall that β_i is positive for all i). First, define for $i = 0, 1, 2, \ldots$,

$$u_i = \alpha_i / \beta_i \ , \quad v_i = \alpha_{i+1} / \beta_i \ .$$

Then from (15) and (19),

$$u_i \, v_i = 2\rho \ , \quad v_i + u_{i+1} = 2(\rho + 1) \ . \tag{24}$$

and eliminating v_i, respectively u_i, leads to the iteration schemes

$$u_{i+1} = 2(\rho + 1) - 2\rho / u_i \ ,$$

$$v_i = 2(\rho + 1) - 2\rho / v_{i+1} \ ,$$

with initial values $u_0 = 2 + \rho$ and $v_0 = 2\rho / (2 + \rho)$. These iteration schemes are illustrated in Figure 4. The *fixed points* of the above iteration schemes are the numbers A_1 and A_2, that is, the roots of $A = 2(\rho + 1) - 2\rho / A$. So

$$A_1 = \rho + 1 - \sqrt{\rho^2 + 1} \ , \quad A_2 = \rho + 1 + \sqrt{\rho^2 + 1} \ .$$

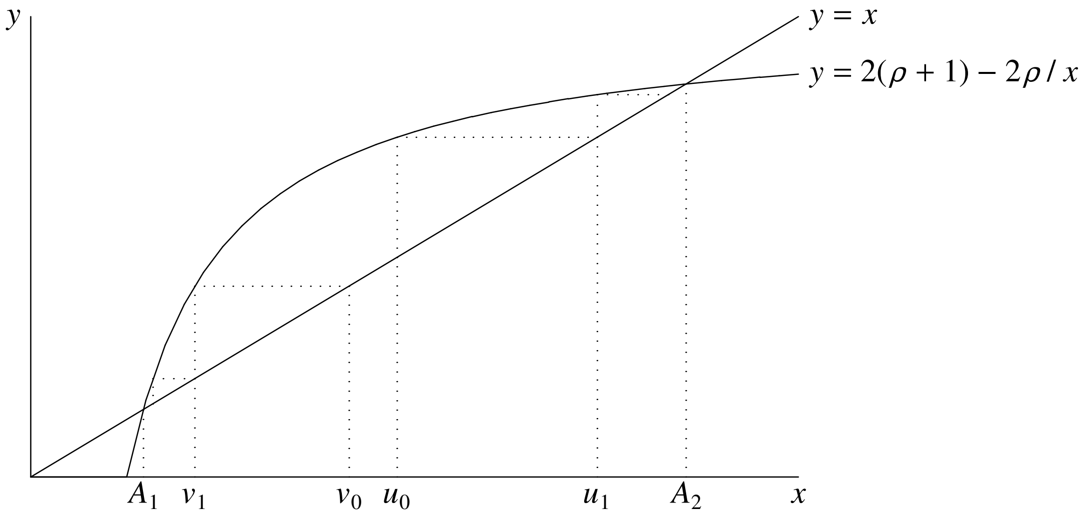

Figure 4: *the iteration schemes for u_i and v_i*

Then, by induction we obtain, as $i \to \infty$,

$$u_i \uparrow A_2 \ , \quad v_i \downarrow A_1 \ . \tag{25}$$

To analyze the behaviour of α_i, β_i, c_i and d_i we first express them in terms of u_i and v_i. Directly from the definition of u_i and v_i it follows that

$$\alpha_{i+1} / \alpha_i = v_i / u_i \ , \quad \beta_{i+1} / \beta_i = v_i / u_{i+1} \ , \tag{26}$$

and dividing the numerator and denominator in (21) by β_i yields

$$c_i = (1 - v_i) / (u_i - 1) \ . \tag{27}$$

To express d_{i+1} / d_i in terms of u_i and v_i, multiply the numerator and denominator in (22) by α_{i+1} and insert the definitions of u_i and v_i, yielding

$$\frac{d_{i+1}}{d_i} = - \frac{(\alpha_{i+1} + \rho)\, u_{i+1} - \alpha_{i+1}\, (\rho + 1)}{(\alpha_{i+1} + \rho)\, v_i - \alpha_{i+1}\, (\rho + 1)}\, c_i \ .$$

Then inserting $\alpha_{i+1} = v_i\, u_{i+1} - 2\rho$, by (17), leads to

$$\frac{d_{i+1}}{d_i} = - \frac{(u_{i+1}^2 - u_{i+1}\, (\rho + 1))\, v_i + \rho\, (2(\rho + 1) - u_{i+1})}{(v_i^2 - v_i\, (\rho + 1))\, u_{i+1} + \rho\, (2(\rho + 1) - v_i)}\, c_i .$$

Finally, inserting $2(\rho + 1) = v_i + u_{i+1}$, by (24), gives the desired expression

$$\frac{d_{i+1}}{d_i} = - \frac{v_i\,(u_{i+1} - \rho)\,(u_{i+1} - 1)}{u_{i+1}\,(v_i - \rho)\,(v_i - 1)}\; c_i \;. \tag{28}$$

By the expressions (26), (27) and (28), the asymptotic behaviour of α_i, β_i, c_i and d_{i+1} / d_i is obtained from the asymptotic behaviour (25) of u_i and v_i. This leads to

Lemma 5: As $i \to \infty$, then

$$\frac{\alpha_{i+1}}{\alpha_i} \text{ and } \frac{\beta_{i+1}}{\beta_i} \downarrow \frac{A_1}{A_2}\;,\quad c_i \to \frac{1 - A_1}{A_2 - 1}\;,\quad \frac{d_{i+1}}{d_i} \to - \frac{A_2}{A_1}\,\frac{1 - A_1}{A_2 - 1}\;.$$

Proof: The limiting behaviour of the ratios α_{i+1} / α_i and β_{i+1} / β_i and the coefficients c_i follows from (25), (26) and (27), and further from (25) and (28), as $i \to \infty$,

$$\frac{d_{i+1}}{d_i} \to - \frac{A_1\,(A_2 - \rho)\,(A_2 - 1)}{A_2\,(A_1 - \rho)\,(A_1 - 1)}\,\frac{1 - A_1}{A_2 - 1}\;.$$

Inserting the identities $A_1\,(A_2 - \rho) = A_2\,\rho(1 - A_1)$ and $A_2\,(A_1 - \rho) = A_1\,\rho(1 - A_2)$ yields the desired limit of d_{i+1} / d_i. □

Further, by (25), it follows that for all i,

$$u_i > u_0 > 1 > \rho > v_0 > v_i > 0.$$

As a consequence, by the expressions (26), (27) and (28),

Lemma 6: $1 > \alpha_0 > \beta_0 > \alpha_1 > \beta_1 > ... > 0$ and $c_i > 0$ and $d_{i+1} / d_i < 0$ for all i.

Thus the terms in expression (20) for $x_{m,r}$ are alternating. The Lemmas 5 and 6 provide the ingredients, needed to prove that the series (20) converges absolutely.

6. The Convergence of the Series of Product Forms

We can now prove that for all $m \geq 0$ and $r \geq 1$ the series (20), which defines $x_{m,r}$, converges absolutely. Consider a fixed $m \geq 0$ and $r \geq 1$. Then from Lemma 5, as $i \to \infty$, both

$$\frac{|\, d_{i+1}\, \alpha_{i+1}^m\, \beta_{i+1}^r \,|}{|\, d_i\, \alpha_i^m\, \beta_i^r \,|} \text{ and } \frac{|\, d_{i+1}\, c_{i+1}\, \alpha_{i+2}^m\, \beta_{i+1}^r \,|}{|\, d_i\, c_i\, \alpha_{i+1}^m\, \beta_i^r \,|} \to \frac{1 - A_1}{A_2 - 1} \left[\frac{A_1}{A_2} \right]^{m+r-1},$$

which is strictly less than unity. Hence, there exist positive constants M and R, with R

strictly less than unity, and both depending on m and r, such that for all i, both terms $|\, d_i\ \alpha_i^m\ \beta_i^r \,|$ and $|\, d_i\ c_i\ \alpha_{i+1}^m\ \beta_i^r \,|$ are bounded by $M\ R^i$. This proves

Lemma 7: For all $m \ge 0$ and $r \ge 1$, both $\sum_{i=0}^{\infty} d_i\ \alpha_i^m\ \beta_i^r$ and $\sum_{i=0}^{\infty} d_i\ c_i\ \alpha_{i+1}^m\ \beta_i^r$ converge absolutely.

By virtue of this lemma, for all $m \ge 0$ and $r \ge 1$ the numbers $x_{m,r}$ are well defined by the series (20), and it is allowed to change the order of summation. As noted at the end of the previous section, the terms in the series (20) are alternating. So it isn't obvious whether the series is positive or negative. The following lemma helps in proving that $\{x_{m,r}\}$ is a positive solution. It states that the terms in (20) are decreasing in modulus, at least with rate $R = 4\,/\,(4 + 2\rho + \rho^2)$.

Lemma 8: Let $R = 4\,/\,(4 + 2\rho + \rho^2) < 1$, then for all $m \ge 0$, $r \ge 1$ and $i \ge 0$,

$$|\, d_{i+1}\ (\alpha_{i+1}^m + c_{i+1}\alpha_{i+2}^m)\ \beta_{i+1}^r \,| \ < \ R\,|\, d_i\ (\alpha_i^m + c_i\alpha_{i+1}^m)\ \beta_i^r \,| \,.$$

Proof: We first prove the lemma for $m = 0$ and $r = 1$. Consider the ratio of both terms. Inserting the expressions (26), (27) and (28), it follows that

$$\begin{aligned}
\frac{|\, d_{i+1}\ (1 + c_{i+1})\ \beta_{i+1} \,|}{|\, d_i\ (1 + c_i)\ \beta_i \,|} &= \frac{v_i^2}{(\rho - v_i)\ (u_i - v_i)}\ \frac{(u_{i+1} - \rho)\ (u_{i+1} - v_{i+1})}{u_{i+1}^2} \\
&< \frac{v_i^2}{(\rho - v_i)\ (u_i - v_i)} \ \le \ \frac{v_0^2}{(\rho - v_0)\ (u_0 - v_0)} \ = \ R \ < \ 1,
\end{aligned}$$

where in the second inequality we used that, by (25), the numbers v_i are positive and decreasing and the numbers $u_i - v_i$ are positive and increasing. This proves the lemma for $m = 0$ and $r = 1$. Now consider an arbitrary $m \ge 0$ and $r \ge 1$. Since the sequences $\{\alpha_i\}$ and $\{\beta_i\}$ are decreasing, by Lemma 6, it follows that for all i,

$$\begin{aligned}
|\, d_{i+1}\ (\alpha_{i+1}^m + c_{i+1}\ \alpha_{i+2}^m)\ \beta_{i+1}^r \,| &< |\, d_{i+1}\ (1 + c_{i+1})\ \beta_{i+1} \,|\ \alpha_{i+1}^m\ \beta_{i+1}^{r-1} \\
&< R\,|\, d_i\ (1 + c_i)\ \beta_i \,|\, \alpha_{i+1}^m\ \beta_{i+1}^{r-1} \ < \ R\,|\, d_i\ (\alpha_i^m + c_i\ \alpha_{i+1}^m)\ \beta_i^r \,| \,.
\end{aligned}$$

□

Corollary: For all $m \ge 0$ and $r \ge 0$, the numbers $x_{m,r}$ are positive.

Proof: The terms in (20) are alternating and, by Lemma 8, strictly decreasing in modulus. Since the first term in (20) is positive, this proves that $x_{m,r}$ is positive for $m \ge 0$ and $r \ge 1$, and, immediate from their definition, also for $m \ge 0$ and $r = 0$. □

We conclude this section by proving that the series

$$C = 2 \sum_{m=0}^{\infty} \sum_{r=1}^{\infty} x_{m,r} + \sum_{m=0}^{\infty} x_{m,0}$$

converges. Inserting the definition of $x_{m,0}$, we obtain that

$$C = 2 \sum_{m=0}^{\infty} \sum_{r=1}^{\infty} x_{m,r} + \frac{1}{\rho + 1} \sum_{m=1}^{\infty} (x_{m-1,1}\, 2\rho + x_{m,1}) + \frac{1}{\rho} x_{0,1} \ , \qquad (29)$$

so that convergence follows once the following lemma is established.

Lemma 9: $\sum_{m=0}^{\infty} \sum_{r=1}^{\infty} x_{m,r} < \infty$.

Proof: By equation (20),

$$\sum_{m=0}^{\infty} \sum_{r=1}^{\infty} x_{m,r} \ = \ \sum_{m=0}^{\infty} \sum_{r=1}^{\infty} \sum_{i=0}^{\infty} d_i \, (\alpha_i^m + c_i \, \alpha_{i+1}^m) \, \beta_i^r ,$$

and we will show that the latter sum converges absolutely. Interchanging summations and using that α_i and β_i are positive and less than unity (cf. Lemma 6), we obtain

$$\sum_{m=0}^{\infty} \sum_{r=1}^{\infty} \sum_{i=0}^{\infty} | \, d_i \, (\alpha_i^m + c_i \, \alpha_{i+1}^m) \, \beta_i^r \, | \ = \ \sum_{i=0}^{\infty} | \, d_i \, | \left(\frac{1}{1 - \alpha_i} + \frac{c_i}{1 - \alpha_{i+1}} \right) \frac{\beta_i}{1 - \beta_i} \, . \qquad (30)$$

By Lemma 5, the ratio of successive terms of the sum at the right hand side of (30) tends to $(1 - A_1) \,/\, (A_2 - 1) < 1$, so that there exist positive constants M and R, with R strictly less than unity, such that for all i,

$$| \, d_i \, | \left(\frac{1}{1 - \alpha_i} + \frac{c_i}{1 - \alpha_{i+1}} \right) \frac{\beta_i}{1 - \beta_i} \ \le \ M \ R^i .$$

Thus the sum (30) converges. □

7. Explicit Form for C

We derive an explicit form for the normalizing constant C, which however, is not essential to the compensation method itself. Substituting the series (20) for $x_{m,r}$ into equation (29) leads to a series of product forms for C, analogously to the one for $x_{m,r}$ (cf. the left hand side of (30)). The method to obtain the explicit form, by means of the generating function, is different from the main arguments in this paper. Therefore we omit details and only sketch the proof. Define the generating function $F(y, z)$ by

$$F(y, z) = \sum_{m=0}^{\infty} \sum_{r=0}^{\infty} q_{m,r}\, y^m\, z^r = C^{-1} \sum_{m=0}^{\infty} \sum_{r=0}^{\infty} x_{m,r}\, y^m\, z^r \ .$$

Substituting (23) to eliminate $x_{m,0}$ and then inserting the series (20), we obtain, by interchanging summations,

$$F(y, z) = C^{-1} \left\{ \sum_{i=0}^{\infty} d_i \left(\frac{1}{1 - \alpha_i\, y} + \frac{c_i}{1 - \alpha_{i+1}\, y} \right) \frac{\beta_i\, z}{1 - \beta_i\, z} \right. \tag{31}$$

$$\left. + \frac{1}{\rho + 1} \sum_{i=0}^{\infty} d_i \left(\frac{(2\rho + \alpha_i)\, y}{1 - \alpha_i\, y} + c_i\, \frac{(2\rho + \alpha_{i+1})\, y}{1 - \alpha_{i+1}\, y} \right) \beta_i + \frac{x_{0,1}}{\rho} \right\} ,$$

valid in $|y| < 1 / \alpha_0$, $|z| < 1 / \beta_0$. It is noted that this partial fraction decomposition of the generating function cannot be obtained, at least in explicit form, from the analysis of Kingman [21] and Flatto and McKean [11]. The equilibrium equations (1)-(6) reduce to the following functional equation for $F(y, z)$,

$$F(y, z)\, g(y, z) = F(y, 0)\, h(y, z) + F(0, z)\, k(y, z) \ ,$$

where

$$g(y, z) = z^2 + y\,(2\rho\, y + 1) - 2(\rho + 1)\, y\, z \ ,$$

$$h(y, z) = y\,(2\rho\, y + 1) - (\rho + 1)\, y\, z - \rho\, y\, z^2 \ ,$$

$$k(y, z) = z\,(z - y) \ .$$

It follows that, if y and z satisfy $|y| < 1 / \alpha_0$, $|z| < 1 / \beta_0$ and $g(y, z) = 0$, then $F(y, 0)$ and $F(0, z)$ are related according to

$$F(y, 0)\, h(y, z) + F(0, z)\, k(y, z) = 0 \ . \tag{32}$$

In the analysis of Kingman [21] and Flatto and McKean [11] this relationship between $F(y, 0)$ and $F(0, z)$ eventually leads to their determination. We use it to establish

$$C = \frac{\rho\,(2 + \rho)}{2\,(1 - \rho^2)\,(2 - \rho)} \ .$$

First, note that $F(0, 1)$ is the fraction of time server 1 (or 2) is idle. Since 2ρ is the offered load, we obtain, by symmetry, $F(0, 1) = 1 - \rho$. Starting with $F(0, 1) = 1 - \rho$, we successively apply relationship (32) to the pairs $(y, z) = (1/2\rho, 1)$ and $(1/2\rho, 1/\rho)$, both satisfying $g(y, z) = 0$, which leads to the determination of $F(0, 1/\rho) = (1-\rho)(2-\rho)$. Next we apply (32) to (y, z), satisfying $g(y, z) = 0$, and let $y \uparrow 1/\rho^2$ $(= 1/\alpha_0)$ and $z \to 1/\rho$. Here note that, by treating y as the parameter, the equation $g(y, z(y)) = 0$, $z(1/\rho^2) = 1/\rho$ is solved for

$$z(y) = (\rho+1)\, y - \sqrt{y\,(\,(\rho^2+1)\, y - 1)}\ .$$

Then inserting $z = z(y)$ into relationship (32) and letting $y \uparrow 1/\rho^2$, we finally obtain the desired expression for C, by using $F(0, 1/\rho) = (1-\rho)(2-\rho)$ and also, as $y \uparrow 1/\rho^2$,

$$h(y, z(y)) = (2+\rho)(\rho-1)(y - 1/\rho^2)/2\rho + o(y - 1/\rho^2)\ ,$$

and by (31),

$$F(y, 0) = (-\,C\ \rho\,(\rho+1)\,(y - 1/\rho^2))^{-1} + O(1)\ .$$

8. Asymptotic Expansion

We now return to the asymptotic equivalence (12), proved in Flatto and McKean [11]. The series (20) extends this result, for it yields a complete asymptotic expansion. First, since the numbers α_j and β_j are decreasing, it follows for all $j \geq 1$, as $m + r \to \infty$ and $r \geq 1$, that

$$d_j\,(\alpha_j^m + c_j\,\alpha_{j+1}^m)\,\beta_j^r = o(d_{j-1}\,(\alpha_{j-1}^m + c_{j-1}\,\alpha_j^m)\,\beta_{j-1}^r)\ .$$

Thus successive terms in (20) are refinements. Since the terms in (20) are alternating and decreasing in modulus, the error of each partial sum is bounded by the final term of the partial sum. Hence, we have for all $j \geq 1$, as $m + r \to \infty$ and $r \geq 1$,

$$q_{m,r}\ =\ C^{-1}\sum_{i=0}^{j-1} d_i\,(\alpha_i^m + c_i\,\alpha_{i+1}^m)\,\beta_i^r + O(d_j\,(\alpha_j^m + c_j\,\alpha_{j+1}^m)\,\beta_j^r)\ . \tag{33}$$

The O-formula for $j = 1$ improves the asymptotic equivalence (12), for as $m + r \to \infty$ and $r \geq 1$,

$$C^{-1} d_0\,(\alpha_0^m + c_0\,\alpha_1^m)\,\beta_0^r + O(d_1\,(\alpha_1^m + c_1\,\alpha_2^m)\,\beta_1^r) =$$

$$C^{-1} d_0 \, (\alpha_0^m + c_0 \, \alpha_1^m) \, \beta_0^r \, (1 + o(1)) \, .$$

Accordingly, the formula for $j = 2$ improves the one for $j = 1$, and so on. The following notation is used in order to represent the whole set (33) by a single formula (see e.g. de Bruijn [6], Section 1.5),

$$\textbf{Lemma 10: } q_{m,r} \approx C^{-1} \sum_{i=0}^{\infty} d_i (\alpha_i^m + c_i \, \alpha_{i+1}) \, \beta_i^r \quad \text{as } m + r \to \infty \text{ and } r \geq 1 \, .$$

9. Numerical Results

The solution (20) is also suitable for numerical evaluation. First, the terms in (20) are easily calculated and they decrease exponentially fast. Secondly, the terms are alternating and decreasing in modulus. This makes that the error in the partial sum can be bounded by the modulus of the final term in the partial sum. In Table 1 we list the numbers $q_{0,1}$, $q_{0,2}$, $q_{1,1}$ and $q_{1,2}$ computed with a relative accuracy of 0.1%. The numbers in parentheses denote the number of terms in (20), needed to attain that accuracy.

ρ	$q_{0,1}$	$q_{0,2}$	$q_{1,1}$	$q_{1,2}$
0.3	0.1591 (14)	0.0100 (3)	0.0156 (3)	0.0007 (2)
0.5	0.1580 (10)	0.0233 (3)	0.0441 (4)	0.0047 (2)
0.7	0.1100 (8)	0.0275 (4)	0.0606 (4)	0.0118 (3)
0.9	0.0380 (6)	0.0140 (4)	0.0350 (4)	0.0104 (3)

Table 1: Computation of $q_{0,1}$, $q_{0,2}$, $q_{1,1}$ and $q_{1,2}$ for increasing values of ρ

Let us investigate the rate of convergence of the terms in the series (20). By Lemma 6, it follows that for all $m \geq 0$ and $r \geq 1$, as $i \to \infty$,

$$\frac{|\, d_{i+1} \, (\alpha_{i+1}^m + c_{i+1} \, \alpha_{i+2}^m) \, \beta_{i+1}^r \,|}{|\, d_i \, (\alpha_i^m + c_i \, \alpha_{i+1}^m) \, \beta_i^r \,|} \to \frac{1 - A_1}{A_2 - 1} \left[\frac{A_1}{A_2} \right]^{m+r-1} .$$

For $0 < \rho < 1$, the factor $(1 - A_1) / (A_2 - 1)$ is decreasing and A_1 / A_2 is increasing, and

$$\lim_{\rho \downarrow 0} \frac{1 - A_1}{A_2 - 1} = 1 \, , \quad \lim_{\rho \uparrow 1} \frac{A_1}{A_2} = \frac{2 - 2^{1/2}}{2 + 2^{1/2}} \, .$$

Hence, if $m > 0$ or $r > 1$, the convergence of the terms in (20) is very fast for all ρ, at least with rate $(2 - 2^{1/2}) / (2 + 2^{1/2}) = 0.1715\ldots$. But if $m = 0$ and $r = 1$, then the rate of convergence is determined only by $(1 - A_1) / (A_2 - 1)$, so, as Table 1 illustrates, convergence is slow for small ρ.

Below we derive expressions for the mean W and second moment $W^{(2)}$ of the waiting time, based on the series for $q_{m,r}$. First, W and $W^{(2)}$ are given by

$$W = 2 \sum_{m=1}^{\infty} \sum_{r=1}^{\infty} m \; q_{m,r} + \sum_{m=1}^{\infty} m \; q_{m,0} \;,$$

$$W^{(2)} = 2 \sum_{m=1}^{\infty} \sum_{r=1}^{\infty} m(m+1) \; q_{m,r} + \sum_{m=1}^{\infty} m(m+1) \; q_{m,0} \;.$$

Substituting (5) to eliminate $q_{m,0}$ and then inserting the series for $q_{m,r}$, we obtain, by interchanging summations (cf. (31)),

$$W = C^{-1} \left\{ 2 \sum_{i=0}^{\infty} d_i \left(\frac{\alpha_i}{(1-\alpha_i)^2} + \frac{c_i \; \alpha_{i+1}}{(1-\alpha_{i+1})^2} \right) \frac{\beta_i}{1-\beta_i} \right.$$

$$\left. + \frac{1}{\rho+1} \sum_{i=0}^{\infty} d_i \left(\frac{2\rho+\alpha_i}{(1-\alpha_i)^2} + \frac{c_i \; (2\rho+\alpha_{i+1})}{(1-\alpha_{i+1})^2} \right) \beta_i \right\},$$

and a similar expression for $W^{(2)}$. The terms in these series are alternating and decreasing (cf. Lemma 8), so the error of each partial sum can be bounded by the modulus of the final term. Accordingly, expressions can be obtained for higher moments of the waiting time or other quantities of interest. In Table 2 we list values of W and $W^{(2)}$, together with the coefficient of variation $cv(W)$ of the waiting time, with a relative accuracy of 0.1%. The numbers of terms, needed to attain that accuracy, are shown in parentheses.

ρ	W	$W^{(2)}$	$cv(W)$
0.3	0.1441 (13)	0.3181 (13)	3.7846
0.5	0.4262 (8)	1.1472 (8)	2.3053
0.7	1.1081 (6)	4.3842 (5)	1.6032
0.9	4.4748 (4)	47.208 (3)	1.1652

Table 2: Computation of the mean W and second moment $W^{(2)}$ of the waiting time, together with the coefficient of variation $cv(W)$, for increasing values of ρ

10. Conclusions and Extensions

We developed a compensation approach to obtain generalized product form expressions for the equilibrium probabilities of the symmetric shortest queue problem. This approach yields explicit relations for the product forms as well as their coefficients and thereby an explicit characterization of the equilibrium probabilities. Based on these explicit relations, qualitative properties of the product forms are derived, which in their turn are exploited to obtain efficient numerical algorithms.

We believe, and this is confirmed by several recent results, that the compensation approach is also useful in other situations. For example, in [3] it is shown that the compensation idea works for the shortest queue problem with non-identical servers. However, in that case the analysis is essentially more complicated, as it requires the construction of solutions on the different regions $m < n$ and $m > n$ which are coupled at the diagonal. In fact, our interest in the present problem arose out of our work in flexible manufacturing systems, which behaved somewhat similar as job-type dependent parallel queueing systems with dynamic routing, see e.g. Schwartz [24], Green [16] and Adan, Wessels and Zijm [1].

Finally we point out that the compensation approach has some flexibility for modifications in the model. For instance, the approach also proceeds if the single servers are replaced by two identical multi-server groups. Then only the compensation on the vertical boundary becomes more complicated. In [2] we showed that the compensation approach can be easily extended to a "simple" asymmetric shortest queue problem, where the symmetric routing probability ½ is replaced by an arbitrary routing probability. In that case the regions $m < n$ and $m > n$ are still a mirror image of each other.

References

1. ADAN, I.J.B.F., WESSELS, J., AND ZIJM, W.H.M., “Queuing analysis in a flexible assembly system with a job-dependent parallel structure” in *Operations Research Proceedings 1988,* pp. 551-558, Springer-Verlag, Berlin (1989).
2. ADAN, I.J.B.F., WESSELS, J., AND ZIJM, W.H.M., “An asymmetric shortest queue problem” in *Operations Research Proceedings 1989,* pp. 434-441, Springer-Verlag, Berlin (1990).
3. ADAN, I.J.B.F., WESSELS, J., AND ZIJM, W.H.M., “Analysis of the asymmetric shortest queue problem,” *Queueing Systems,* 8, pp. 1-58 (1991).
4. BLANC, J.P.C., “On a numerical method for calculating state probabilities for queueing systems with more than one waiting line,” *J. Comput. Appl. Math.,* 20, pp. 119-125 (1987).
5. BLANC, J.P.C., “The power-series algoritm applied to the shortest-queue model,” *Opns. Res.,* 40, pp. 157-167 (1992).
6. BRUIJN, N.G. DE, *Asymptotic methods in analysis,* Dover Publications, Inc., New York (1981).
7. COHEN, J.W. AND BOXMA, O.J., *Boundary value problems in queueing system analysis,* North-Holland, Amsterdam (1983).
8. CONOLLY, B.W., “The autostrada queueing problem,” *J. Appl. Prob.,* 21, pp. 394-403 (1984).
9. FAYOLLE, G., *Méthodes analytiques pour les files d’attente couplées,* Thesis, Univ. de Paris VI, Paris (1979).

10. FAYOLLE, G. AND IASNOGORODSKI, R., "Two coupled processors: the reduction to a Riemann-Hilbert problem," *Z. Wahrsch. Verw. Gebiete,* 47, pp. 325-351 (1979).

11. FLATTO, L. AND MCKEAN, H.P., "Two queues in parallel," *Comm. Pure Appl. Math.,* 30, pp. 255-263 (1977).

12. FOSCHINI, G. J. AND SALZ, J., "A basic dynamic routing problem and diffusion," *IEEE Trans. Commun.,* COM-26, pp. 320-327 (1978).

13. FOSTER, F.G., "On the stochastic matrices associated with certain queuing processes," *Ann. Math. Stat.,* 24, pp. 355-360 (1953).

14. GERTSBAKH, I., "The shorter queue problem: A numerical study using the matrix-geometric solution," *Eur. J. Oper. Res.,* 15, pp. 374-381 (1984).

15. GRASSMANN, W.K., "Transient and steady state results for two parallel queues," *OMEGA Int. J. of Mgmt. Sci.,* 8, pp. 105-112 (1980).

16. GREEN, L., "A queueing system with general-use and limited-use servers," *Opns. Res.,* 33, pp. 168-182 (1985).

17. HAIGHT, F.A., "Two queues in parallel," *Biometrica,* 45, pp. 401-410 (1958).

18. HALFIN, S., "The shortest queue problem," *J. Appl. Prob.,* 22, pp. 865-878 (1985).

19. HOOGHIEMSTRA, G., KEANE, M., AND REE, S. VAN DE, "Power series for stationary distributions of coupled processor models," *SIAM J. Appl. Math.,* 48, pp. 1159-1166 (1988).

20. IASNOGORODSKI, R., *Problèmes-frontières dans les files d'attente,* Thesis, Univ. de Paris VI, Paris (1979).

21. KINGMAN, J.F.C., "Two similar queues in parallel," *Ann. Math. Statist.,* 32, pp. 1314-1323 (1961).

22. KNESSL, C., MATKOWSKY, B.J., SCHUSS, Z., AND TIER, C., "Two parallel queues with dynamic routing," *IEEE Trans. Commun.,* 34, pp. 1170-1175 (1986).

23. RAO, B.M. AND POSNER, M.J.M., "Algorithmic and approximation analysis of the shorter queue model," *Naval Res. Log.,* 34, pp. 381-398 (1987).

24. SCHWARTZ, B.L., "Queuing models with lane selection: a new class of problems," *Opns. Res.,* 22, pp. 331-339 (1974).