Analyzing $E_k|E_r|c$ Queues

I.J.B.F. Adan*
W.A. van de Waarsenburg*
J. Wessels†*

Abstract

In this paper we study a system consisting of c parallel identical servers and a common queue. The service times are Erlang-r distributed and the interarrival times are Erlang-k distributed. The service discipline is first-come first-served. The waiting process may be characterized by $(n_{-1}, n_0, n_1, \ldots, n_c)$ where n_{-1} represents the number of remaining arrival stages, n_0 the number of waiting jobs and n_i, $i = 1, \ldots, c$, the number of remaining service stages for server i. Bertsimas has proved that the equilibrium probability for saturated states (i.e. states with all servers busy) can be written as a linear combination of geometric terms with n_0 as exponent. In the present paper it is shown that the coefficients also have a geometric form with respect to $n_{-1}, n_1, \ldots, n_c$. It is also shown how the factors may be found efficiently. The present paper uses a direct approach for solving the equilibrium equations rather than a generating function approach as Bertsimas does. The direct approach is based on separation of variables and has been inspired by previous work of two of the authors on the shortest queue problem in particular and the two-dimensional random walk more generally. The characterization of the equilibrium probabilities leads to exact expressions for performance measures such as the moments of the queue length and the waiting time, which are useful for numerical computations. Numerical results are presented.

1 Introduction

The $E_k|E_r|c$ queueing system is a typical example of an elegantly modeled and seemingly simple system which, nevertheless, has never been analyzed satisfactorily. Apparently, the system behavior is more complex than the simple formulation suggests.

In the literature many solution approaches have been proposed. Shapiro [30] studies the $M|E_2|c$ system and proves that the probability of n jobs in the system can be expressed as a linear combination of powers. Mayhugh and McCormick [22] and Heffer [13] treat the $M|E_r|c$ system and determine the generating function of the stationary state probabilities. Their results seem to be computationally useful for small values of r and c only; Heffer presents some numerical results for r and c ranging from 1 to 3. Poyntz and Jackson [25] analyze the $E_k|E_r|2$ and $E_k|E_r|3$ system making use of generating functions. By generalizing the approach of Mayhugh and McCormick [22], Yu [33] solves the $E_k|E_r|c$ queue with heterogeneous servers. Numerical results based on this procedure for the special case of homogeneous servers have been obtained by Hillier and Lo [14].

Yu [34] uses the imbedded Markov chain approach to analyze the $GI|E_r|c$ queue. By defining an appropriate state space partitioning and exploiting the cyclic structure of the transition probability matrix of the imbedded chain, he shows that the equilibrium probabilities for saturated states (i.e. states with all servers busy) are geometric in one of the state variables. Utilizing

*Eindhoven University of Technology, Department of Mathematics and Computer Science, p.o. Box 513, 5600 MB - Eindhoven, The Netherlands.
†International Institute for Applied Systems Analysis, A-2361 Laxenburg, Austria.

this structure of the equilibrium probabilities, Avis [5] develops a procedure for the computation of the waiting time distribution and reports numerical results for several $GI|E_2|2$ queueing systems.

The $C_k|C_r|c$ queue, which is more general than the $E_k|E_r|c$ queue, is treated by Bertsimas [7], who shows, by using a generating function technique, that the equilibrium probabilities for saturated states are geometric in the number of waiting jobs. This characterization of the equilibrium probabilities is exploited by Bertsimas [8] to prove that the waiting time distribution for the $C_k|C_r|c$ system under the first-come first-served (FCFS) discipline can be expressed as a mixture of exponential distributions. Bertsimas includes some numerical results for the $C_2|C_2|c$ system.

Ishikawa [18] tries to tackle the $GI|E_r|c$ queue by using the elapsed time since the last arrival as supplementary variable, but mainly restricts his attention to the $GI|E_3|3$ queue. By using the remaining service times as supplementary variables, Hokstad [15] and Knessl, Matkowsky, Schuss and Tier [20] study the $M|G|2$ system and find for a couple of examples, in particular for the $M|E_2|2$ queue, an explicit form for the generating function of the queue length. Hokstad [16] uses a refined version of the technique in [15] to analyze the $M|K_2|c$ queue and includes a few numerical examples for the case $c = 3$.

It is well-known that the steady state probability distributions of the embedded Markov chain for the $GI|Ph|c$ queue are modified matrix-geometric, see e.g. section 4.5. in Neuts' book [24], where it is pointed out that this result is of limited value from a computational point of view, because it requires the determination of the high dimensional rate matrix R, which is the solution of a non-linear matrix equation. By exploiting the structural properties of the model, Ramaswami and Lucantoni [26] extend the numerical feasibility of the method to a wide set of problems by developing efficient schemes to compute R. They also provide an easily computable form for the waiting time distribution in terms of R, and report numerical results for several $GI|H_2|15$ models. The $Ph|Ph|c$ queueing system allows certain simplifications. This system can be modeled as quasi-birth-and-death process, for which R can be computed efficiently. Latouche and Ramaswami [21] provide an efficient iterative algorithm for obtaining R. Daigle and Lucantoni [10] propose to determine R by spectral decomposition.

An alternative approach, related to [10], that may be used to find the equilibrium distribution of $E_k|E_r|c$ queues is the spectral expansion method of Mitrani and Mitra [23]. It is based on reducing the equilibrium equations to a vector difference equation with constant coefficients, the solution of which can be expressed in terms of eigenvalues and eigenvectors of the associated characteristic polynomial (see also section 1.6. in [24]).

The structure of the equilibrium equations of $Ph|Ph|c$ queues enables application of special iterative methods to solve the equations numerically. Takahashi and Takami [31] propose an iterative aggregation-disaggregation algorithm exploiting a convergence property of probabilities conditioned by the number of jobs in the system. Seelen [28] presents an algorithm which may be viewed as an extremely effective modification of the method in [31]. The algorithm of Seelen is capable of solving models with a large number of servers and, in fact, it has been used to produce the tablebook of Seelen, Tijms and Van Hoorn [29]. For an excellent survey of computational and approximation methods for $GI|GI|c$ queues, the reader is referred to Tijms [32].

In the present paper we show, by using a direct solution approach (without generating functions), that the equilibrium probabilities for saturated states can be expressed as a linear combination of terms that are geometric in each of the state variables. This characterization of the equilibrium probabilities directly leads, by a result of Bertsimas [8], to an exact expression for the waiting time distribution under FCFS in the form of a mixture of exponential distributions.

The solution approach is based on separation of variables, well known from the theory of partial differential equations (see e.g. Garabedian [12]). The inspiration for using this approach came from its relative success in solving the equilibrium equations for the shortest queue problem (cf. [1], [2]) and its generalization to the two-dimensional random walk on the integer grid in the first quadrant (cf. [1], [3]). Generating function approaches integrate the equations for boundary states with the equations for states in the inner region to construct the usual functional equation for the generating function. The essential idea of the present approach is to avoid the integration of boundary and inner conditions. This is achieved by first constructing a sufficiently rich basis of separable solutions of the equations in the interior of the state space and then to use this basis to construct a linear combination that also satisfies the conditions at the boundaries of the state space. For the present model a finite linear combination does the job, whereas in [1], [2] and [3] infinitely many terms are needed. In the latter case a more complicated procedure is needed to select the appropriate basis solutions from the even larger class of available solutions. In the present case, there is no selection problem, since we need all available basis solutions.

To some extent, a similar approach is followed by Bertsimas [7] for the $C_k|C_r|c$ problem. In the present paper it is shown that Bertsimas' representation can be refined considerably for the $E_k|E_r|c$ problem. The refinement is obtained by a complete direct analysis of the equilibrium equations, contrary to the approach of Bertsimas, who also separates the equilibrium equations for boundary and inner states, but still treats the latter by generating functions. Extensions of the present approach to more complicated distributions are possible, however, in such cases the results are correspondingly more complex. Therefore, it seems sensible to introduce our approach for $E_k|E_r|c$ problems separately.

Our results lead to an efficient algorithm for the computation of system performance measures such as the delay probability and the moments of the queue size and the (conditional) waiting time. Also, the state probabilities and the waiting time distribution can be exactly evaluated. The computational complexity of the algorithm is proportional to $k\binom{c+r-1}{r}^3 + \binom{c+r-1}{r-1}^3$. It appears that the basis solutions depend on the zeros of polynomials. In general, it may be difficult to compute such zeros accurately. In the present model, however, it is shown that the zeros can be found efficiently and accurately via contraction schemes. Compared to the numerical methods of Takahashi and Takami [31] and Seelen [28], the present approach offers qualitative insight by providing explicit expressions, which apart from their theoretical interest, are also numerically useful and lead to numerical procedures, which are efficient and stable.

This paper is organized as follows. In section 2 the model is introduced and the relevant equilibrium equations are formulated. In section 3 the basis solutions are constructed satisfying the equations in the states with all servers busy. Section 4 presents the main theorem, stating that the equilibrium probabilities can be expressed as a linear combination of products of powers. In section 5 the symmetry of the model is exploited and in section 6 explicit expressions for performance measures are presented. Numerical issues are discussed in section 7 and numerical results are reported in section 8. The final section is devoted to comments and conclusions. Two technical lemmas are proved in the appendices A and B.

2 Model and equations

Consider a system with c parallel identical servers and a common queue. The service times are Erlang-r distributed with mean r/μ and the interarrival times are Erlang-k distributed with mean k/λ. An Erlang-l distributed time with mean l/ν is interpreted as to be composed of l stages, each with a negative-exponentially distributed length (parameter value ν). The service discipline is FCFS. An arriving job, who finds any free servers, chooses each free server with equal probability. This queueing system can be modeled as a continuous-time Markov process with a state space consisting of the vectors $\underline{n} = (n_{-1}, n_0, n_1, \ldots, n_c)$, where n_{-1} is the number of remaining arrival stages, n_0 is the number of waiting jobs and n_i is the number of remaining service stages for server i, $i = 1, \ldots, c$. So n_{-1} ranges from 1 to k, n_0 from 0 to ∞ and n_i, $i = 1, \ldots, c$, from 0 to r, where $n_i = 0$ is only possible if $n_0 = 0$. Note that completion of an arrival stage leads to an arrival if $n_{-1} = 1$ and completion of a service stage at server i leads to a departure if $n_i = 1$, $i = 1, \ldots, c$. Below we introduce some notations and formulate the equilibrium equations that will be relevant to the analysis.

Let $\underline{e}_i = (0,\ldots,0,1,0,\ldots,0)$ have $c+2$ components, with the one at the same place as n_i in $\underline{n}$ and let $\delta_i(l)$ be 1 if $l=i$ and 0 otherwise. Further, we introduce $a(c,r) = \binom{c+r-1}{r-1}$.

By equating the rate out of and the rate into state $\underline{n}$ we obtain

$$\begin{aligned} p(\underline{n})(\lambda + c\mu) \quad &= \quad p(\underline{n}+\underline{e}_{-1})\lambda(1-\delta_k(n_{-1})) + p(\underline{n}-(k-1)\underline{e}_{-1}-\underline{e}_0)\lambda\delta_k(n_{-1}) \\ &\quad + \sum_{i=1}^{c} p(\underline{n}+\underline{e}_i)\mu(1-\delta_r(n_i)) + \sum_{i=1}^{c} p(\underline{n}+\underline{e}_0-(r-1)\underline{e}_i)\mu\delta_r(n_i)\,, \end{aligned} \tag{1}$$

which is valid for all states $\underline{n}$ with $n_{-1} = 1,\ldots,k$, $n_0 = 0,1,\ldots$ and $n_i = 1,\ldots,r$ for $i=1,\ldots,c$, except for the ones with $n_{-1} = k$ and $n_0 = 0$ (in that case namely, the second term at the right-hand side of (1) has to be modified). The equations (1) form the inner conditions, the equations in states with $n_{-1} = k$ and $n_0 = 0$ (the ones excluded from (1)) and states with $n_i = 0$ for some $i \in \{1,\ldots,c\}$ form the boundary conditions. The formulation of the boundary conditions is more complicated and would require extra notations. However, the precise form of these conditions is not relevant to the analysis, and therefore it is omitted.

Bertsimas directly exploits that servers are identical by using, instead of $(n_1,\ldots,n_c)$, the aggregated description $\underline{m} = (m_1,\ldots,m_r)$ of the service process, where m_i denotes the number of servers with i remaining service stages. He proves in [7] that, under the assumption that the factors w_j are distinct (this is proved for the special case $r=2$), the probabilities $p(n_{-1},n_0,\underline{m})$ may be expressed as

$$\begin{aligned} p(n_{-1},n_0,\underline{m}) \quad &= \quad \sum_{j=1}^{a(c,r)} B_j v_j^{n_{-1}} w_j^{n_0} f_j(\underline{m}) \\ &\quad n_{-1} = 1,\ldots,k, \quad n_0 = 0,1,\ldots, \quad m_i = 1,\ldots,c,\ \sum_{i=1}^{r} m_i = c\,, \end{aligned} \tag{2}$$

where for each j the factor v_j is characterized by a system of nonlinear equations involving the Laplace transforms of the interarrival and service time distributions, the factor w_j satisfies $w_j = 1/v_j^k$, the coefficients $f_j(\underline{m})$ have to be solved from a system of $a(c,r)$ linear equations involving w_j and, finally, the coefficients B_j have to be solved from the boundary conditions.

Based on the characterization (2) for the equilibrium probabilities for saturated states, Bertsimas shows in [8] that the complementary distribution of the waiting time W is given by

$$P(W > t) = k \sum_{j=1}^{a(c,r)} B_j G_j(\mathbf{1}) \frac{1}{1-w_j} v_j e^{\lambda(1-v_j)t}\,, \qquad t > 0\,, \tag{3}$$

where $\mathbf{1}$ is the row vector of ones and $G_j(\underline{z})$ is the generating function of $f_j(\underline{m})$ defined by

$$G_j(\underline{z}) = \sum_{\substack{m_i=1,\ldots,c \\ m_1+\ldots+m_r=c}} f_j(\underline{m}) z_1^{m_1}\ldots z_r^{m_r}\,, \quad \underline{z} = (z_1,\ldots,z_r)\,. \tag{4}$$

In section 3 we will develop a refinement of representation (2) by employing the detailed rather than the aggregated state concept. In the refined representation also the coefficients f_j are replaced by geometric forms, i.e.,

$$\begin{aligned} p(\underline{n}) \quad &= \quad \sum_{j=1}^{r^c} C_j \beta_{-1,j}^{n_{-1}} \beta_{0,j}^{n_0} \beta_{1,j}^{n_1} \cdots \beta_{c,j}^{n_c} \\ &\quad n_{-1} = 1,\ldots,k, \quad n_0 = 0,1,\ldots, \quad n_i = 1,\ldots,r,\ i = 1,\ldots,c\,. \end{aligned} \tag{5}$$

It appears that, contrary to Bertsimas' approach, the assumption of different factors $\beta_{0,j}$ is not needed to obtain this representation. Since we do not use the assumption of identical servers explicitly, the results of sections 3 and 4 can easily be extended to the case of non-identical servers. In section 5 the characterization above will be translated to a form in the aggregated state concept yielding explicit expressions for (the generating functions of) the coefficients f_j.

Result (5) may be related to the modified matrix-geometric representation of the equilibrium distribution. The $E_k|E_r|c$ queue can be modeled as a quasi-birth-and-death process, the equilibrium distribution of which has the form (see e.g. chapter 3 in Neuts [24])

$$p_n = p_1 R^{n-1}\,, \qquad n \geq 1\,,$$

where p_n is the vector of equilibrium probabilities $p(\underline{n})$ with $n_0 = n$. When the rate matrix R is diagonalizable, it is directly seen that the factors $\beta_{0,j}$ in the form (5) are eigenvalues of R (the other eigenvalues being zero) and the components $x(n_{-1}, n_1, \ldots, n_c)$ of the associated eigenvectors x are $\beta_{-1,j}^{n_{-1}} \beta_{1,j}^{n_1} \ldots \beta_{c,j}^{n_c}$. Similarly, result (2) may be linked to the matrix-geometric representation. See also section 1.6. in Neuts [24] for a discussion on results of the form (2) and (5) and their relationship with matrix-geometric results.

3 Analysis of the inner conditions

In this section we will construct a basis of solutions satisfying the inner conditions (1). The boundary conditions are considered in the next section. To construct the solution basis we use the method of separation of variables. The idea of this method is to seek solutions of the form

$$p(\underline{n}) = F_{-1}(n_{-1}) F_0(n_0) F_1(n_1) \ldots F_c(n_c)\,. \tag{6}$$

Insertion of this form in (1) and then dividing by $F_{-1}(n_{-1}) \ldots F_c(n_c)$ yields

$$\begin{aligned} \lambda + c\mu &= \frac{F_{-1}(n_{-1}+1)}{F_{-1}(n_{-1})} \lambda (1 - \delta_k(n_{-1})) + \frac{F_{-1}(1)}{F_{-1}(k)} \frac{F_0(n_0 - 1)}{F_0(n_0)} \lambda \delta_k(n_{-1}) \\ &\quad + \sum_{i=1}^{c} \frac{F_i(n_i+1)}{F_i(n_i)} \mu (1 - \delta_r(n_i)) + \sum_{i=1}^{c} \frac{F_0(n_0+1)}{F_0(n_0)} \frac{F_i(1)}{F_i(r)} \mu \delta_r(n_i)\,. \end{aligned} \tag{7}$$

For $n_{-1} = 1, \ldots, k-1$ and $n_i = 1, \ldots, r-1$, $i = 1, \ldots, c$ the equations above reduce to

$$\lambda + c\mu = \frac{F_{-1}(n_{-1}+1)}{F_{-1}(n_{-1})} \lambda + \sum_{i=1}^{c} \frac{F_i(n_i+1)}{F_i(n_i)} \mu\,.$$

Each of the terms at the right-hand side is a function of one variable only, independent of the other variables. Hence, the sum of these terms is constant only if each term is constant, so

$$\frac{F_{-1}(n_{-1}+1)}{F_{-1}(n_{-1})} = \beta_{-1}\,, \quad n_{-1} = 1, \ldots, k-1\,,$$

$$\frac{F_i(n_i+1)}{F_i(n_i)} = \beta_i\,, \quad n_i = 1, \ldots, r-1\,,\; i = 1, \ldots, c\,,$$

for some constants $\beta_{-1}, \beta_1, \ldots, \beta_c$, usually referred to as separation constants. These relations directly imply that (up to a factor)

$$F_{-1}(n_{-1}) = \beta_{-1}^{n_{-1}}\,, \quad F_i(n_i) = \beta_i^{n_i}\,,\; i = 1, \ldots, c\,. \tag{8}$$

Finally, to find the form of $F_0(n_0)$, we set $n_{-1} = k$ and $n_i = r$ for $i = 1, \ldots, c$ in (7). The general solution of the resulting difference equation for $F_0(n_0)$ is a linear combination of functions of the form

$$F_0(n_0) = \beta_0^{n_0}\,. \tag{9}$$

Hence, we may conclude that the solutions (6) are of the form

$$p(\underline{n}) = \beta_{-1}^{n_{-1}} \beta_0^{n_0} \beta_1^{n_1} \ldots \beta_c^{n_c}\,,$$

where, by substitution of (8) and (9) in (7), it is seen that the constants $\beta_{-1}, \beta_0, \beta_1, \ldots, \beta_c$ satisfy

$$\lambda + c\mu = \beta_{-1}\lambda(1-\delta_k(n_{-1})) + \frac{\beta_{-1}}{\beta_{-1}^k\beta_0}\lambda\delta_k(n_{-1}) + \sum_{i=1}^c \beta_i\mu(1-\delta_r(n_i)) + \sum_{i=1}^c \frac{\beta_0\beta_i}{\beta_i^r}\mu\delta_r(n_i)\,.$$

Since the right-hand side of this equation must be constant, we have to require that the coefficients of the functions $\delta_k(n_{-1})$ and $\delta_r(n_i)$ vanish, which amounts to

$$\beta_{-1}^{-k} = \beta_0 = \beta_1^r = \ldots = \beta_c^r\,.$$

So, by introducing parameters $x_{-1}, x_1, \ldots, x_c$ satisfying $x_{-1}^k = x_1^r = \ldots = x_c^r = 1$, we may write

$$\beta_{-1} = \frac{x_{-1}}{\alpha^r} \quad \beta_0 = \alpha^{kr}, \quad \beta_i = x_i\alpha^k, \quad i = 1, \ldots, c, \tag{10}$$

for some α. Insertion of (10) in the equation for $\beta_{-1}, \ldots, \beta_c$ and multiplying both sides of the resulting equation by α^r leads to the following equation for α,

$$\alpha^r(\lambda + c\mu) = x_{-1}\lambda + \alpha^{r+k}\sum_{i=1}^c x_i\mu\,. \tag{11}$$

These findings are summarized in the following lemma.

Lemma 1: *The product $p(\underline{n}) = \beta_{-1}^{n_{-1}}\beta_0^{n_0}\ldots\beta_c^{n_c}$ satisfies (1) iff $\beta_{-1}, \beta_0, \ldots, \beta_c$ satisfy*

$$\beta_{-1} = x_{-1}/\alpha^r\,, \quad \beta_0 = \alpha^{kr}\,, \quad \beta_i = x_i\alpha^k\,, \quad i = 1, \ldots, c,$$

where $x_{-1}^k = x_i^r = 1$ for $i = 1, \ldots, c$ and α is a root of equation (11).

This lemma characterizes the set of products $\beta_{-1}^{n_{-1}}\beta_0^{n_0}\ldots\beta_c^{n_c}$ satisfying (1). Clearly, only products that can be normalized, i.e., the sum over all states $\underline{n}$ converges absolutely, are useful. This implies that $|\beta_0| < 1$, or equivalently, $|\alpha| < 1$. The next lemma, which is proved in appendix A, states how many roots of equation (11) lay inside the unit circle. The condition in that lemma states that the offered workload per unit time must be less than the service capacity.

Lemma 2: *Provided $\lambda r/k\mu < c$, equation (11) has exactly r simple roots inside the unit circle for each set of parameters x_i satisfying $x_{-1}^k = x_i^r = 1$ for $i = 1, \ldots, c$.*

We henceforth assume that the utilization condition in lemma 2 is satisfied.

Assumption: $\lambda r/k\mu < c$.

According to the lemmas 1–2, there are r products $\beta_{-1}^{n_{-1}}\beta_0^{n_0}\ldots\beta_c^{n_c}$ with $|\beta_0| < 1$ satisfying (1) for each feasible choice of $x_{-1}, x_1, \ldots, x_c$. So we find rkr^c products. However, below we show that some of them are identical. We may write

$$r = \rho d\,, \quad k = \kappa d\,,$$

where $d = \gcd(r, k)$ and $\gcd(\rho, \kappa) = 1$. For each α satisfying (11) it follows that $u\alpha$ with $u^d = 1$ also satisfies (11) and both roots lead to the same products, since the factors β_i depend only upon α^d. Therefore we set $\gamma = \alpha^d$. Then equation (11) reduces to

$$\gamma^\rho(\lambda + c\mu) = x_{-1}\lambda + \gamma^{\rho+\kappa}\sum_{i=1}^c x_i\mu\,. \tag{12}$$

and lemma 1 can be restated as follows.

Lemma 3: *The product $p(\underline{n}) = \beta_{-1}^{n_{-1}}\beta_0^{n_0}\ldots\beta_c^{n_c}$ satisfies (1) iff $\beta_{-1}, \beta_0, \ldots, \beta_c$ satisfy*

$$\beta_{-1} = x_{-1}/\gamma^\rho\,, \quad \beta_0 = \gamma^{\kappa r}\,, \quad \beta_i = x_i\gamma^\kappa\,, \quad i = 1, \ldots, c,$$

where $x_{-1}^k = x_i^r = 1$ for $i = 1, \ldots, c$ and γ is a root of equation (12).

For two roots γ_1 and γ_2 of (12) inside the unit circle, it is easily verified that $\gamma_1^\kappa \neq \gamma_2^\kappa$ (see also lemma 6), and so γ_1 and γ_2 lead to different values for β_i. Hence, from the equivalent of lemma 2 for equation (12) (stating that this equation has ρ, instead of r, roots inside the unit circle) it follows that for each feasible choice of the parameters x_i there are ρ products $\beta_{-1}^{n-1}\beta_0^{n_0}\ldots\beta_c^{n_c}$ with $|\beta_0| < 1$ satisfying (1). So, by letting run the parameters x_i through all feasible values, we find $\rho k r^c$ products. However, there still are duplicates among these products, since the two sets of parameters $x_{-1}/u^{\rho/\kappa}, ux_1, \ldots, ux_c$ with $u^r = 1$ and $vx_{-1}, x_1, \ldots, x_c$ with $v^\kappa = 1$ lead to exactly the same products as the original set $x_{-1}, x_1, \ldots, x_c$. This follows by observing that if γ is a root of (12) for the original parameter set, then $\gamma/u^{1/\kappa}$ and $w\gamma$ with w satisfying $w^\kappa = 1$ and $w^\rho = v$, are roots of (12) for the first, respectively second parameter set, and the three roots lead to the same product. Hence, there are $r\kappa$ copies of each product. To avoid these copies we arbitrarily decide to restrict the feasible values for x_1 to $x_1 = 1$ and the feasible values for x_{-1} to the first d roots of the equation $x_{-1}^k = 1$, i.e., $x_{-1} = e^{\frac{2\pi i}{k}n}$ for some $n = 0, 1, \ldots, d-1$. Now, by letting run the parameters x_i through this restricted set of feasible values, we find $\rho d r^{c-1} = r^c$ products and it is readily verified that there are no duplicates among them. We label these products $\beta_{-1,j}^{n-1}\beta_{0,j}^{n_0}\ldots\beta_{c,j}^{n_c}$, $j = 1, 2, \ldots, r^c$. In this way we have characterized the set of product forms satisfying the equilibrium equations (1). In the next section, we will prove our main result stating that $p(\underline{n})$ can be expressed as a linear combination of the products $\beta_{-1,j}^{n-1}\beta_{0,j}^{n_0}\ldots\beta_{c,j}^{n_c}$, $j = 1, 2, \ldots, r^c$ for states $\underline{n}$ with all servers busy.

4 Satisfying the boundary conditions

Theorem 1: *There exist unique coefficients C_j such that*

$$\begin{aligned} p(\underline{n}) \quad = \quad & \sum_{j=1}^{r^c} C_j \beta_{-1,j}^{n-1}\beta_{0,j}^{n_0}\beta_{1,j}^{n_1}\ldots\beta_{c,j}^{n_c} \\ & n_{-1} = 1, \ldots, k, \quad n_0 = 0, 1, \ldots, \quad n_i = 1, \ldots, c,\ i = 1, \ldots, c. \end{aligned} \tag{13}$$

Proof: For each choice of the coefficients C_j the sequence $p(\underline{n})$ given by (13) satisfies the equations (1). The remaining equations are the equations for states $\underline{n}$ with $n_{-1} = k$, $n_0 = 0$ and $n_i > 0$, $i = 1, \ldots, c$, and for states with $n_i = 0$ for some $i \in \{1, \ldots, c\}$. These equations form a linear, homogeneous system for the unknowns C_j and the unknown quantities $p(\underline{n})$ in states with $n_i = 0$ for some $i \in \{1, \ldots, c\}$. The number of equations is equal to the number of unknowns. Hence, by first omitting the equation in $\underline{0}$, say, the reduced system has a nonnull solution. The equation in $\underline{0}$ is automatically satisfied, since inserting the solution $p(\underline{n})$ in the equations in states $\underline{n} \neq \underline{0}$ and then summing over these equations and changing summations exactly yields the desired equation. Changing summations is allowed, since the sum of $p(\underline{n})$ over all states converges absolutely. This follows from the fact that $|\beta_{0,j}| < 1$ for all j. Hence, $p(\underline{n})$ is an absolutely convergent solution of all equilibrium equations. It remains to show that $p(\underline{n})$ is a nonnull solution. This readily follows if at least one of the quantities $p(\underline{n})$ in states with $n_i = 0$ for some $i \in \{1, \ldots, c\}$ is nonnull. If all these quantities are null, then at least one of the coefficients C_j must be nonnull. In this case the property that the products $\beta_{-1,j}^{n-1}\beta_{0,j}^{n_0}\ldots\beta_{c,j}^{n_c}$, $j = 1, \ldots, r^c$ are linearly independent on the set of saturated states implies that $p(\underline{n})$ is a nonnull solution. The independence is formulated in the lemma below and will be proved in appendix B. Taking the independence for granted at the moment we can finish the proof of the theorem as follows. From a result of Foster (theorem 1 in [11], or see section I.3.4. in Cohen [9]), we can conclude that the Markov process is ergodic and normalization of the $p(\underline{n})$ produces the equilibrium probabilities. The uniqueness of the coefficients C_j follows from that of the equilibrium distribution and the independence property of the basis solutions. $\square$

Lemma 4: *The products $\beta_{-1,j}^{n_{-1}}\beta_{0,j}^{n_0}\dots\beta_{c,j}^{n_c}$, $j=1,\dots,r^c$ are linearly independent on the set of saturated states, i.e.,*

$$\sum_{j=1}^{r^c} C_j\beta_{-1,j}^{n_{-1}}\beta_{0,j}^{n_0}\beta_{1,j}^{n_1}\dots\beta_{c,j}^{n_c}=0$$

$$n_{-1}=1,\dots,k,\quad n_0=0,1,\dots,\quad n_i=1,\dots,r,\ i=1,\dots,c \tag{14}$$

iff $C_j\equiv 0$.

5 Exploiting the symmetry

Since we have not used the fact that all servers are identical, the results in the previous sections are still valid in case server i works with rate μ_i, where the rates μ_i are not necessarily identical. In this section, however, we exploit the fact that all servers are identical and we show that the number of coefficients C_j to be determined, can be reduced from r^c to $a(c,r)$. The same result can be obtained directly by chosing the same aggregated state concept as in [7]. This will also be worked out.

By symmetry,

$$p(n_{-1},n_0,n_1,\dots,n_c)=p(n_{-1},n_0,n_{\pi(1)},\dots,n_{\pi(c)})$$

for each permutation π of $\{1,\dots,c\}$. Hence, by theorem 1,

$$\sum_{j=1}^{r^c} C_j\beta_{-1,j}^{n_{-1}}\beta_{0,j}^{n_0}\beta_{1,j}^{n_1}\dots\beta_{c,j}^{n_c}=\sum_{j=1}^{r^c} C_j\beta_{-1,j}^{n_{-1}}\beta_{0,j}^{n_0}\beta_{\pi(1),j}^{n_1}\dots\beta_{\pi(c),j}^{n_c}$$

for saturated states. Since, by lemma 4, the products in the sums above are linearly independent on the set of saturated states, it follows that the coefficients of products that are equal up to a permutation of the last c factors are identical. So the number of coefficients to be determined can be reduced from r^c to b say, i.e., the number of products that cannot be obtained from one another by permuting the last c factors. To find b, we first let the parameters x_i in equation (12) run through all feasible values, yielding $\rho k r^c$ products satisfying (1) (including, of course, exactly κr copies of each product). Since equation (12) is invariant under permutations of $x_1,\dots,x_c$, it follows for each $\beta_{-1}^{n_{-1}}\beta_0^{n_0}\dots\beta_c^{n_c}$ satisfying (1) that all products obtained by permuting $\beta_1,\dots,\beta_c$ also satisfy (1). So the set of $\rho k r^c$ products can be split up into $\rho k a(c,r)$ subsets of products that are equal up to a permutation of the last c factors. Since each product has exactly κr copies, each subset also has exactly $\kappa r(=\rho k)$ copies. Hence, since b is equal to the number of different subsets, we find $b=a(c,r)$. The number of boundary equations required to determine the $a(c,r)$ unknown coefficients can be reduced accordingly. Namely, it suffices to consider the equations in boundary states $\underline{n}$ with $n_1\le\dots\le n_c$.

The same reduction can be obtained directly by using aggregated states $(n_{-1},n_0,m_1,\dots,m_r)$, where the variable m_i denotes the number of servers with i remaining service stages. Translation of the basis solution $p(\underline{n})=\beta_{-1,j}^{n_{-1}}\beta_{0,j}^{n_0}\dots\beta_{c,j}^{n_c}$ to the aggregated state description yields $p(n_{-1},n_0,\underline{m})=\beta_{-1,j}^{n_{-1}}\beta_{0,j}^{n_0}f_j(\underline{m})$ where $f_j(\underline{m})$ is the sum of $\beta_{1,j}^{n_1}\dots\beta_{c,j}^{n_c}$ over all states $(n_1,\dots,n_c)$ satisfying $\#\{n_i|n_i=l\}=m_l$ for $l=1,\dots,r$. So $f_j(\underline{m})$ is the coefficient of $z_1^{m_1}\dots z_r^{m_r}$ in the generating function $G_j(\underline{z})$ defined by

$$G_j(\underline{z})=\prod_{i=1}^{c}\left(\beta_{i,j}z_1+\beta_{i,j}^2z_2+\dots+\beta_{i,j}^r z_r\right),\quad \underline{z}=(z_1,\dots,z_r).$$

Clearly, basis solutions that are equal up to a permutation of the last c factors are translated to exactly the same basis solution in the new state description. Hence, by renumbering the basis solutions $\beta_{-1,j}^{n_{-1}}\beta_{0,j}^{n_0}\dots\beta_{c,j}^{n_c}$, $j=1,\dots,r^c$ such that the first $a(c,r)$ solutions cannot be obtained from one another by permuting the last c factors, this basis is translated to the reduced set of solutions $\beta_{-1,j}^{n_{-1}}\beta_{0,j}^{n_0}f_j(\underline{m})$, $j = 1,\ldots,a(c,r)$. From lemma 4, it is readily seen that the new solutions are linearly independent on the set of saturated states, so the same arguments as in the proof of theorem 1 may be used to prove its equivalent in the aggregated state concept.

Theorem 2: *There exist unique coefficients B_j such that*

$$p(n_{-1},n_0,\underline{m}) = \sum_{j=1}^{a(c,r)} B_j\beta_{-1,j}^{n_{-1}}\beta_{0,j}^{n_0}f_j(\underline{m})$$

$$n_{-1}=1,\ldots,k, \quad n_0=0,1,\ldots, \quad m_i=1,\ldots,c, \ \sum_{i=1}^{r} m_i = c\,.$$

6 Global performance measures

In this section we present expressions for the waiting time distribution and the moments of the waiting time and the queue length. From theorem 2 and Bertsimas' result in [8] we may conclude that the complementary distribution of the waiting time W is given by (see also (3))

$$P(W>t) = k\sum_{j=1}^{a(c,r)} B_jG_j(\mathbf{1})\frac{1}{1-\beta_{0,j}}\beta_{-1,j}e^{\lambda(1-\beta_{-1,j})t}\,, \qquad t>0\,,$$

where $\mathbf{1}$ is the row vector of ones and

$$G_j(\mathbf{1}) = \prod_{i=1}^{c}\beta_{i,j}\frac{1-\beta_{i,j}^r}{1-\beta_{i,j}}\,.$$

Hence, we get the following expressions for the delay probability $\Pi_W(=P(W>0))$ and the moments of the waiting time W,

$$\Pi_W = k\sum_{j=1}^{a(c,r)} B_jG_j(\mathbf{1})\frac{1}{1-\beta_{0,j}}\beta_{-1,j}\,,$$

$$E(W^l) = k\sum_{j=1}^{a(c,r)} B_jG_j(\mathbf{1})\frac{1}{1-\beta_{0,j}}\beta_{-1,j}\frac{l!}{(\lambda(\beta_{-1,j}-1))^l}\,, \qquad l=1,2,\ldots\,.$$

From the characterization of the equilibrium probabilities in theorem 2 it is also possible to derive expressions for measures of system performance such as the factorial moments of the queue size L_q, yielding (cf. section 2 in [8])

$$E(L_q(L_q-1)\ldots(L_q-l+1)) = \sum_{j=1}^{a(c,r)} B_jG_j(\mathbf{1})\frac{l!\beta_{0,j}^{l-1}}{(1-\beta_{0,j})^l}\frac{\beta_{-1,j}}{\beta_{-1,j}-1}\,, \qquad l=1,2,\ldots\,.$$

Using the matrix-geometric approach, Ramaswami and Lucantoni [26] develop an infinite sum expression for the waiting time distribution involving the rate matrix R. Compared with this approach, the present approach offers finite sum expressions, moreover, the calculation of the basis solutions (see the next section) requires definitely less effort than that of the rate matrix R.

For the particular case of the $E_k|E_r|1$ queue, Ikeda [19] gives a double series expression for the mean waiting time. The present approach, however, offers a finite sum expression and in the special case $c=1$ it is even possible to give explicit expressions for the coefficients B_j by exploiting the structure of the boundary equations.

7 Numerical aspects

The analytical results of the previous sections can be used for numerical calculations. By doing so, the following steps may be distinguished:

1. Determine the basis solutions $p(n_0, n_{-1}, \underline{m}) = \beta_{-1,j}^{n_{-1}} \beta_{0,j}^{n_0} f_j(\underline{m})$ for all $j = 1, \ldots, a(c, r)$;

2. Solve the coefficients B_j from the boundary conditions and the normalization equation.

In the first step we have to find the roots γ with $|\gamma| < 1$ of the polynomial equations (12) where the parameters x_i run through the set of feasible values. In general, the determination of roots of polynomial equations is a difficult numerical problem. Luckily, in the present case, equation (12) can be solved very efficiently. It is readily seen that equation (12) can be rewritten as

$$\gamma = \phi F(\gamma, x_{-1}, \sum_{i=1}^{c} x_i), \tag{15}$$

where ϕ satisfies $\phi^\rho = 1$ and $F(\gamma, x, y) = \sqrt[\rho]{x\lambda} / \sqrt[\rho]{\lambda + c\mu - \gamma^\kappa y \mu}$. The following lemma states that for feasible values of the parameters x_i the function F in equation (15) is a contraction in γ. The proof of the lemma is straightforward and therefore it is omitted (cf. section 8 in [4]).

Lemma 5: *For all x and y with $|x| \leq 1$ and $|y| \leq c$, the function $F(\gamma, x, y)$ is a contraction on $|\gamma| \leq \Gamma$ where Γ is the positive root, less than one, of the equation $\gamma = F(\gamma, 1, c)$.*

Let us denote the roots of $\phi^\rho = 1$ by $\phi_1, \ldots, \phi_\rho$. Then, for fixed feasible values of the parameters x_i, it follows from lemma 5 that for each $j = 1, \ldots, \rho$ equation (15) with $\phi = \phi_j$ has exactly one root $\tilde{\gamma_j}$ satisfying $|\tilde{\gamma_j}| \leq \Gamma$ and this root can be found by using the iteration scheme

$$\gamma_{n+1} = \phi_j F(\gamma_n, x_{-1}, \sum_{i=1}^{c} x_i), \quad n = 0, 1, \ldots$$

starting with $\gamma_0 = 0$. The sequence $\gamma_0, \gamma_1, \ldots$ converges exponentially fast to $\tilde{\gamma_j}$. Remark that, based on lemma 5, the equivalent of lemma 2 for equation (12) can be proved directly, i.e., without using Rouché's theorem as done in appendix A.

In the second step, the coefficients B_j and the probabilities $p(n_0, n_{-1}, \underline{m})$ for unsaturated states have to be solved from a system of $ka(c-1, r+1) + a(c, r)$ linear equations. The computational effort to solve this system can be reduced essentially by eliminating the unsaturated probabilities. By partitioning the set of unsaturated states into k subsets of size $a(c-1, r+1)$, where the l-th subset consists of the states $(n_{-1}, n_0, \underline{m})$ with $n_{-1} = l$, and exploiting the cyclic transition structure between these subsets, the unsaturated probabilities can be eliminated efficiently. The computational complexity of the elimination is proportional to $ka(c-1, r+1)^3$ and that of the solution of the reduced set of $a(c, r)$ equations for the coefficients B_j is proportional to $a(c, r)^3$.

The algorithm has been tested for a large number of $E_k|E_r|c$ systems, and the results appear to be in full agreement with the ones in the tablebook of Seelen, Tijms and Van Hoorn [29]. Numerical results are reported in the next section.

8 Numerical results

As an illustration of the stability and accuracy of the present algorithm, we present some examples. In table 1 we list the delay probability Π_W, the mean and coefficient of variation $c_{W|W>0}$ of the conditional waiting time $W|W > 0$, the mean and coefficient of variation c_{L_q} of the queue length L_q and the waiting time probability $P(W > \frac{1}{2})$ for a range of values for the system parameters. In each example we have set $\mu = r$, so that the mean service time is equal to one. The parameter SU denotes the server utilization defined by $SU = \lambda r / ck\mu$. The computation times (in minutes) on a SUN-4 workstation are listed in the column below CT.

c	r	k	SU	Π_W	$E(W\|W>0)$	$c_{W\|W>0}$	EL_q	c_{L_q}	$P(W>\frac{1}{2})$	CT
2	3	2	0.8	0.61491	1.09437	0.94893	1.07670	1.6168	0.40476	0.0
			0.95	0.89659	4.21476	0.98588	7.17994	1.1275	0.80616	0.0
2	3	10	0.8	0.46335	0.61435	0.94432	0.45545	1.9112	0.21231	0.0
			0.95	0.84424	2.23188	0.98149	3.58005	1.1680	0.68535	0.0
3	5	10	0.8	0.34808	0.31169	0.89621	0.26038	2.3149	0.06897	0.1
			0.95	0.79441	1.05591	0.95803	2.39065	1.2160	0.51051	0.1
5	3	2	0.8	0.44030	0.45525	0.92982	0.80177	1.9837	0.15019	0.0
			0.95	0.83578	1.70211	0.97765	6.75730	1.1901	0.63481	0.0
5	3	10	0.8	0.29337	0.26382	0.92073	0.30958	2.4705	0.04226	0.0
			0.95	0.76922	0.90953	0.96672	3.32321	1.2476	0.45410	0.0
5	5	2	0.8	0.42648	0.39232	0.90302	0.66927	2.0271	0.12074	0.5
			0.95	0.82976	1.43886	0.96844	5.67103	1.1945	0.60144	0.5
5	5	10	0.8	0.25246	0.20017	0.88527	0.20213	2.7422	0.01712	1.4
			0.95	0.74450	0.64622	0.94387	2.28527	1.2676	0.35315	1.4
8	4	1	0.8	0.44496	0.42027	0.91674	1.19682	1.9929	0.13737	1.6
			0.98	0.93323	3.93328	0.98941	28.7778	1.0748	0.82997	1.5
8	4	3	0.8	0.27905	0.21784	0.90036	0.38904	2.5677	0.02459	2.7
			0.98	0.89980	1.85523	0.98253	13.0877	1.1002	0.69733	2.6
10	3	4	0.8	0.21900	0.17652	0.91478	0.30925	2.9201	0.01060	0.6
			0.98	0.88440	1.48638	0.98314	12.8827	1.1153	0.64016	0.6
10	3	10	0.8	0.16474	0.14080	0.91375	0.18557	3.3811	0.00335	1.1
			0.98	0.86617	1.11307	0.97857	9.44826	1.1279	0.56104	1.1

Table 1: *Performance measures are displayed for a range of values for c, r and k. In all examples the mean service time is normalised to one (so $\mu = r$). For each combination two values of λ are used, the first corresponding to a server utilisation $SU = 0.8$, the second corresponding to a server utilisation of at least* 0.95. *The computation times (in minutes) on a SUN-4 workstation are listed in the column below CT.*

In table 2 it is demonstrated that the exact solution behaves numerically well also for more extreme values of the system parameters k, r, c and the server utilization SU.

c	r	k	SU	Π_W	$E(W\|W>0)$	$c_{W\|W>0}$	EL_q	c_{L_q}	$P(W>\frac{1}{2})$	CT
3	10	10	0.95	0.76752	0.72444	0.93165	1.58465	1.2436	0.40071	3.3
			0.995	0.97479	6.72273	0.99178	19.5615	1.0200	0.91209	3.3
5	5	2	0.99	0.96450	7.03794	0.99329	33.6010	1.0367	0.90418	0.6
			0.995	0.98216	14.0378	0.99662	68.5921	1.0182	0.95094	0.5
25	2	2	0.99	0.92540	2.01359	0.99350	46.1184	1.0762	0.72579	0.9
			0.995	0.96222	4.01355	0.99669	96.0655	1.0377	0.85209	1.0
25	2	10	0.99	0.90333	1.21516	0.99027	27.1676	1.0938	0.60262	2.2
			0.995	0.95080	2.41511	0.99499	57.1203	1.0463	0.77628	2.3

Table 2: *Performance measures for more extreme values of the system parameters.*

9 Comments and conclusions

It has been shown, by using a separation of variables technique, that the equilibrium distribution for the $E_k|E_r|c$ queue may be characterized as a finite linear combination of products of powers. In these products each component of the state is represented by one power-factor. These power-factors have been characterized completely. The weights of the linear combination are determined by a set of linear equations. Although the analysis has been worked out for the case of identical servers, the main result (theorem 1) can easily be extended to the case of nonidentical servers. In that case the linear combination contains r^c terms. If there are some groups of identical servers, then the required number of terms decreases considerably. In case all servers are equal, the number decreases to $\binom{c+r-1}{r-1}$. The expressions for the equilibrium probabilities lead to similar expressions for measures of system performance such as the moments of the queue length and the waiting time. These results are exploited to develop an efficient and stable numerical algorithm, which still works well for relatively large systems with high traffic intensities.

With respect to possible extensions, there are several directions of interest. One interesting direction is constituted by the case of separate queues for the c servers with some sort of allocation of the incoming jobs to the respective servers. For such systems it has been proved that the so-called shortest-delay routing is optimal under some conditions (see Hordijk, Koole [17]). Shortest-delay routing allocates an incoming job to the server with the lowest number of service stages in its queue. This situation has been treated to some extent by Adan and Wessels in [4] for the case $c = 2$, $k = 1$. In this situation one gets a linear combination of countably many terms which can be computed according to a tree-like recurrent scheme. For $c > 2$, it is unlikely that the approach will work, since it apparently does not work for the shortest queue problem with 3 servers, Poisson arrivals, and exponentially distributed service times. However, it might work for $c = 2$, $k > 1$.

Another interesting direction of extension consists of replacing E_k and/or E_r by more general distribution classes. One may think here of Coxians (cf. Bertsimas [7]) or of finite mixtures of Erlang distributions with the same scale parameters, but also of general phase-type distributions (cf. Neuts [24]). The class of finite mixtures of Erlang distributions with the same scale parameters is dense in the set of all distributions (cf. Schassberger [27]) and leads to the same state description as the one in the present paper, however, with slightly more complicated transition behavior. A simplification would be to accept only mixtures of two Erlang distributions with $l-1$ and l phases or with 1 and l phases, since (as Tijms [32] points out) this class seems to be sufficiently rich for most practical purposes. Particularly for the two last-mentioned situations, it seems most likely that similar results may be derived as in the present paper.

Appendix A: Proof of lemma 2

Equation (11) can be rewritten in the form $f(z) + g(z) = 0$, where $f(z) = z^{r+k} \sum_{i=1}^{c} x_i \mu$ and $g(z) = -z^r(\lambda + c\mu) + x_{-1}\lambda$. It follows for all z with $z \neq 0$ that

$$|f(z)| - |g(z)| \leq |z|^{r+k} c\mu - |z|^r(\lambda + c\mu) + \lambda\,.$$

It is readily verified that the right-hand side $h(x) = x^{r+k}c\mu - x^r(\lambda + c\mu) + \lambda$ is convex for $x \geq 0$, $h(0) = \lambda > 0$, $h(1) = 0$ and $h'(1) = kc\mu - r\lambda > 0$. Hence, $h(x) = 0$ has two positive roots, namely $x = 1$ and one in $(0, 1)$, say $x = B$. Applying Rouché's Theorem to a circle $|z| = C$ with $B < C < 1$ proves that equation (11) has r roots inside the unit circle. It remains to show that all zeros are simple. The derivative of $f(z) + g(z)$ vanishes for $\tilde{z}$ satisfying

$$\tilde{z}^k = \frac{r(\lambda + c\mu)}{(r+k)\sum_{i=1}^{c} x_i \mu}\,.$$

So

$$|\tilde{z}|^k \geq \frac{r(\lambda + c\mu)}{(r+k)c\mu}\,.$$

Since

$$|f(\tilde{z}) + g(\tilde{z})| = |\tilde{z}^r \frac{k(\lambda + c\mu)}{r+k} - x_{-1}\lambda| \geq |\tilde{z}|^r \frac{k(\lambda + c\mu)}{r+k} - \lambda \geq \frac{kc\mu}{r}\left(\frac{r(\lambda + c\mu)}{(r+k)c\mu}\right)^{\frac{r+k}{k}} - \lambda > 0\,,$$

we can conclude that all zeros of $f(z) + g(z)$ are simple, which completes the proof. □

Appendix B: Proof of lemma 4

The '$\Leftarrow$' is trivial; to prove '$\Rightarrow$' we first rewrite equation (14) by substituting $\beta_{0,j} = \beta_{1,j}^r$ and $\beta_{i,j} = x_{i,j}\beta_{1,j}$, $i = 2,\ldots,c$, where $x_{i,j}$ are solutions of $x_{i,j}^r = 1$, into equation (14) and introducing the variable m denoting the total number of uncompleted service stages in the system, i.e., $m = n_0 r + n_1 + \ldots + n_c$, yielding

$$\sum_{j=1}^{r^c} C_j \beta_{-1,j}^{n_{-1}} \beta_{1,j}^{m} x_{2,j}^{n_2} \ldots x_{c,j}^{n_c} = 0$$

$$n_{-1} = 1,\ldots,k, \quad m = cr, cr+1,\ldots, \quad n_i = 1,\ldots,r,\ i = 2,\ldots,c. \tag{16}$$

For each feasible choice of $n_{-1}, m, n_3, \ldots, n_c$ we obtain from equation (16) that

$$\sum_{n=0}^{r-1} \left(\sum_{x_{2,j}=e^{\frac{2\pi i}{r}n}} C_j \beta_{-1,j}^{n_{-1}} \beta_{1,j}^{m} x_{3,j}^{n_3} \ldots x_{c,j}^{n_c} \right) \left(e^{\frac{2\pi i}{r}n}\right)^{n_2} = 0\,, \qquad n_2 = 1,\ldots,r\,.$$

By well known properties of the *VanderMonde matrix* (see e.g. Bellman [6], p. 186) this implies that for all $n = 0,\ldots,r-1$

$$\sum_{x_{2,j}=e^{\frac{2\pi i}{r}n}} C_j \beta_{-1,j}^{n_{-1}} \beta_{1,j}^{m} x_{3,j}^{n_3} \ldots x_{c,j}^{n_c} = 0$$

$$n_{-1} = 1,\ldots,k, \quad m = cr, cr+1,\ldots, \quad n_i = 1,\ldots,r,\ i = 3,\ldots,c.$$

By repeatedly applying this procedure, equation (16) eventually decomposes into r^{c-1} sets of equations, i.e., for each choice of x_j satisfying $x_j^r = 1$, $j = 2,\ldots,c$ we obtain that

$$\sum_{x_{2,j}=x_2,\ldots,x_{c,j}=x_c} C_j \beta_{-1,j}^{n_{-1}} \beta_{1,j}^{m} = 0\,, \qquad n_{-1} = 1,\ldots,k, \quad m = cr, cr+1,\ldots\,. \tag{17}$$

The sum in equation (17) runs over exactly r products, the factors $\beta_{-1,j}$ and $\beta_{1,j}$ of which are given by $\beta_{-1,j} = x_{-1}/\gamma^\rho$ and $\beta_{1,j} = \gamma^\kappa$ respectively, where $x_{-1} = e^{\frac{2\pi i}{k}n}$ for some $n = 0,1,\ldots,d-1$ and γ is a root inside the unit circle of equation (12). From the next lemma it follows that the products in equation (17) have different factors $\beta_{1,j}$ implying that $C_j \equiv 0$, which completes the proof of lemma 4. $\square$

Lemma 6: *Let γ_1 and γ_2 be roots inside the unit circle of equation (12) with $x_{-1} = e^{\frac{2\pi i}{k}n_1}$ and $x_{-1} = e^{\frac{2\pi i}{k}n_2}$ respectively, where $0 \le n_1 < n_2 < d$. Then $\gamma_1^\kappa \ne \gamma_2^\kappa$. The same result holds if $n_1 = n_2$ provided that $\gamma_1 \ne \gamma_2$.*

Proof: Let us suppose the contrary, i.e., $\gamma_1^\kappa = \gamma_2^\kappa$. The roots γ_j satisfy

$$\gamma_j^\rho(\lambda + c\mu) = y_j\lambda + \gamma_j^{\rho+\kappa}\sum_{i=1}^{c} x_i\mu\,, \tag{18}$$

where $y_j = e^{\frac{2\pi i}{k}n_j}$, $j = 1,2$. By multiplying equation (18) for $j = 1$ with y_2 and for $j = 2$ with y_1 and then subtracting both equations and inserting $\gamma_2^\kappa = \gamma_1^\kappa$ we get

$$(y_2\gamma_1^\rho - y_1\gamma_2^\rho)(\lambda + c\mu) = \gamma_1^\kappa(y_2\gamma_1^\rho - y_1\gamma_2^\rho)\sum_{i=1}^{c} x_i\mu. \tag{19}$$

If we can prove that $y_2\gamma_1^\rho - y_1\gamma_2^\rho \ne 0$, then division of equation (19) by this term results in

$$\lambda + c\mu = \gamma_1^\kappa \sum_{i=1}^{c} x_i\mu\,,$$

which contradicts that $|\gamma_1| < 1$ and thus completes the proof of lemma 6. Hence, it remains to prove that $y_2\gamma_1^\rho \neq y_1\gamma_2^\rho$. Let us, again, suppose the contrary, i.e.,

$$\gamma_2^\rho = \gamma_1^\rho y_2/y_1 = \gamma_1^\rho e^{\frac{2\pi i}{k}(n_2-n_1)}. \tag{20}$$

Since $\gamma_2 \neq \gamma_1$ and $\gamma_2^\kappa = \gamma_1^\kappa$, it follows that $\gamma_2 = e^{\frac{2\pi i}{\kappa}n}\gamma_1$ for some $n = 1, 2, \ldots, \kappa - 1$. Substitution of this relation in equality (20) yields that $e^{\frac{2\pi i}{\kappa}n\rho} = e^{\frac{2\pi i}{k}(n_2-n_1)}$, implying (recall that $k = \kappa d$)

$$n\rho = (n_2 - n_1)/d + m\kappa \tag{21}$$

for some integer m. Since $n_2 - n_1 < d$ it follows that equality (21) leads to a contradiction if $n_2 > n_1$. Otherwise, if $n_2 = n_1$, then, since $\gcd(\rho, \kappa) = 1$, equality (21) implies that n can be divided by κ contradicting that $0 < n < \kappa$. □

References

[1] I.J.B.F. Adan, *A compensation approach for queueing problems,* CWI Tract No. 104, CWI, Amsterdam, 1994.

[2] I.J.B.F. Adan, J. Wessels, W.H.M. Zijm, "Analysis of the symmetric shortest queue problem", *Stochastic Models,* vol. 6, pp. 691–713, 1990.

[3] I.J.B.F. Adan, J. Wessels, W.H.M. Zijm, "A compensation approach for two-dimensional Markov processes," *Adv. Appl. Prob.*, vol. 25, pp. 783–817, 1993.

[4] I.J.B.F. Adan, J. Wessels, "Analysing shortest expected delay routing for Erlang servers," Memorandum COSOR 93-38, Department of Mathematics and Computing Science, Eindhoven University of Technology, 1993 (submitted for publication).

[5] D.M. Avis, "Computing waiting times in $GI|E_k|c$ queueing systems", *TIMS Studies in the Management Sciences,* vol. 7, pp. 215–232, 1977.

[6] R. Bellman, *Introduction to matrix analysis,* McGraw-Hill, London, 1960.

[7] D. Bertsimas, "An analytic approach to a general class of $G|G|s$ queueing systems", *Opns. Res.,* vol. 38, pp. 139–155, 1990.

[8] D. Bertsimas, "An exact FCFS waiting time analysis for a class of $G|G|s$ queueing systems", *QUESTA,* vol. 3, pp. 305–320, 1988.

[9] J.W. Cohen, *The single server queue,* North Holland, Amsterdam, 1982.

[10] J.N. Daigle, D.M. Lucantoni, "Queueing systems having phase-dependent arrival and service rates", in *Numerical solution of Markov chains,* ed. W.J. Stewart, pp. 161–202, Marcel Dekker, New York, 1991.

[11] F.G. Foster, "On the stochastic matrices associated with certain queueing processes", *Ann. Math. Stat.,* vol. 24, pp. 355–360, 1953.

[12] P.R. Garabedian, *Partial differential equations,* John Wiley & Sons, Chichester, 1967.

[13] J.C. Heffer, "Steady-state solution of the $M|E_k|c$ $(\infty, FIFO)$ queueing system", INFOR *J. Canadian O. R. S.,* vol. 17, pp. 16-30, 1969.

[14] F.S. Hillier, F.D. Lo, "Tables for multi-server queueing systems involving Erlang distributions", TR-14, Operations Research Department, Stanford University, 1971.

[15] P. HOKSTAD "On the steady-state solution of the $M|G|2$ queue", *Adv. Appl. Prob.*, vol. 11, pp. 240–255, 1979.

[16] P. HOKSTAD "The steady-state solution of the $M|K_2|m$ queue", *Adv. Appl. Prob.*, vol. 12, pp. 799–823, 1980.

[17] A. HORDIJK, G. KOOLE, "On the optimality of the generalized shortest queue policy", *Probab. Engineer. Inform. Sci.*, vol. 4, pp. 477–487, 1990.

[18] A. ISHIKAWA, "On the equilibrium solution for the queueing system; $GI|E_k|m$", *TRU Mathematics,* vol. 15, pp. 47–66, 1979.

[19] Z. IKEDA, "Mean waiting time of a Gamma/Gamma/1 queue", *Opns. Res. Let.,* vol. 10, pp. 177–181, 1991.

[20] C. KNESSL, B.J. MATKOWSKY, Z. SCHUSS, C. TIER, "An integral equation approach to the $M|G|2$ queue", *Opns. Res.,* vol. 38, pp. 506–518, 1990.

[21] G. LATOUCHE, V. RAMASWAMI, "A logarithmic reduction algorithm for quasi-birth-death processes", *J. Appl. Prob.,* vol. 30, pp. 650–674, 1993.

[22] J.O. MAYHUGH, R.E. MCCORMICK, "Steady state solution of the queue $M|E_k|r$", *Mgmt. Sci.,* vol. 14, pp. 692–712, 1968.

[23] I. MITRANI, D. MITRA, "A spectral expansion method for random walks on semi-infinite strips", in *Iterative methods in Linear Algebra,* eds. R. Beauwens, P. de Groen, North-Holland, Amsterdam, pp. 141–149, 1992.

[24] M.F. NEUTS, *Matrix-geometric solutions in stochastic models,* Johns Hopkins University Press, Baltimore 1981.

[25] C.D. POYNTZ, R.R.P. JACKSON, "The steady-state solution for the queueing process $E_k|E_m|r$", *O. R. Quart.,* vol. 24, pp. 615–625, 1973.

[26] V. RAMASWAMI, D.M. LUCANTONI, "Algorithms for the multi-server queue with phase type service", *Stochastic Models,* vol. 1, pp. 393–417, 1985.

[27] R.S. SCHASSBERGER, *Warteschlangen,* Springer-Verlag, Berlin 1973.

[28] L.P. SEELEN, "An algorithm for $Ph|Ph|c$ queues", *EJOR,* vol. 23, pp. 118–127, 1986.

[29] L.P. SEELEN, H.C. TIJMS, M.H. VAN HOORN, *Tables for multi-server queues,* North Holland, Amsterdam, 1986.

[30] S. SHAPIRO, "The *M*-server queue with Poisson input and Gamma-distributed service of order two", *Opns. Res.,* vol. 14, pp. 685–694, 1966.

[31] Y. TAKAHASHI, Y. TAKAMI "A numerical method for the steady-state probabilities of a $GI|G|C$ queueing system in a general class", *J. Oper. Res. Soc. Japan,* vol. 19, pp. 147–157, 1976.

[32] H.C. TIJMS, *Stochastic modelling and analysis: a computational approach,* John Wiley & Sons, Chichester, 1986.

[33] O.S. YU, "On the steady-state solution of an $E_m|E_k|r$ queue with heterogeneous servers", TR-38, Operations Research Department, Stanford University, 1971.

[34] O.S. YU, "Steady-state solution of a heterogeneous-server queue with Erlang service times", *TIMS Studies in the Management Sciences,* vol. 7, pp. 199–213, 1977.