

A compensation procedure for multiprogramming queues

I. J. B. F. Adan¹, G. J. van Houtum¹, J. Wessels^{1,2}, and W. H. M. Zijm³

¹ Eindhoven University of Technology, Department of Mathematics and Computing Science, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

² International Institute for Applied Systems Analysis, A-2361 Laxenburg, Austria

³ University of Twente, Department of Mechanical Engineering, P.O. Box 217, 7500 AE Enschede, The Netherlands

Received March 16, 1992 / Accepted in revised form November 24, 1992

Summary. In this paper we study a multiprogramming system consisting of an input-output unit (IO unit) and a central processor (CP). This system can be represented by a continuous time Markov process with states (m, n) , where m and n denote the number of jobs at the CP and the IO unit respectively. The computation of the equilibrium distribution $\{p_{m,n}\}$ of this Markov process is the purpose of the analysis in this paper. The analysis consists of two parts. In the first part, we use a compensation procedure to show that the equilibrium distribution $\{p_{m,n}\}$ in those states (m, n) for which $m + n$ is not too small, can be expressed as an infinite linear combination of product forms. Explicit formulae are given for the product forms and the coefficients of this infinite linear combination. In the second part of the analysis, we pay attention to some numerical aspects of the computation of the equilibrium distribution.

Zusammenfassung. Wir betrachten in dieser Arbeit das Modell eines multiprogrammierten Rechensystems, das aus einer zentralen Recheneinheit (CP) und einer Ein-Ausgabe-Einheit (IO) besteht. Das Modell ist ein Markov-Prozess in kontinuierlicher Zeit mit Zuständen (m, n) , wobei m die Anzahl der Aufträge an der CP, n die Anzahl der Aufträge an der IO angibt. Hauptziel der Arbeit ist die Berechnung der Gleichgewichtswahrscheinlichkeiten $\{p_{m,n}\}$ des Prozesses. Im ersten Teil der Untersuchungen wird eine Kompensationsprozedur verwendet, um zu zeigen, daß $p_{m,n}$ für Zustände mit hinreichend großen $m + n$ als abzählbar unendliche Linearkombination von Produktformausdrücken geschrieben werden kann. Es werden explizite Formeln für diese Produktformausdrücke und die Linearkoeffizienten angegeben. Im zweiten Teil behandeln wir die numerischen Aspekte bei der Berechnung der $\{p_{m,n}\}$.

Key words: multiprogramming queues, Markov process, product forms, compensation procedure

Schlüsselwörter: Multiprogrammierte Wartesysteme, Markov-Prozess, Produktform-Gleichgewicht, Kompensationsmethode

1. Introduction

In this paper we consider a multiprogramming system with two stations, an input-output unit (IO unit) and a central processor (CP), and an infinite source of available new jobs. After being released from the infinite source, a new job makes a number of visits to the IO unit and the CP and subsequently leaves the system. During each visit the job may have to wait before being served, due to the presence of other jobs cycling between the IO unit and the CP. This system can be represented by a continuous time Markov process with states (m, n) , where m and n denote the number of jobs at the CP and the IO unit respectively. Due to the special rule for releasing new jobs, the system is not a Jackson network and therefore the equilibrium distribution $\{p_{m,n}\}$ of this Markov process has a more complex structure than a simple product-form (geometric) distribution. The determination of this distribution is the purpose of this paper and we shall call this problem the *multiprogramming queues problem (MPQP)*.

Hofri [6] studied this problem by exploiting techniques developed by Kingman [7] (see also Flatto and McKean [4]) for the *symmetric shortest queue problem (SSQP)*. For this SSQP, Kingman converts the equilibrium equations into an equation for the bivariate generating function for $\{p_{m,n}\}$, by which this function is given in terms of two univariate generating functions. Subsequently, he shows that these two univariate functions are meromorphic and he gives formulae for the poles and residues. By this result, the equilibrium distribution of the lengths of the two queues *could* be expressed as an infinite linear combination of product forms. However, explicit formulae are given *only for* the equilibrium probabilities at the boundaries of the state space. By using the same kind of analysis for the MPQP, Hofri derived similar results for the

equilibrium distribution $\{p_{m,n}\}$ of the lengths of the queues at the CP and the IO unit. However, for the MPQP these results are restricted to those states (m,n) for which $m+n$ is not too small. The extension of the analysis for the SSQP to the MPQP appears to be far from trivial. For the complete analysis along this line, the reader is referred to a recent paper [3] in which the results of Hofri are amended.

Recently, an elementary method, called the *compensation procedure*, has been developed to compute the equilibrium distribution of the SSQP (cf. [1]). In short this compensation procedure works as follows. First, the method constitutes an initial solution consisting of one product form, which approximately describes the equilibrium distribution far away from the origin of the state space. After that, in all next steps product form terms are added to the solution to correct errors on one of the boundaries of the state space. It can be shown that this procedure yields the solution of the equilibrium equations up to a normalizing constant. The analysis in [1] *directly* leads to the result that the equilibrium distribution of the lengths of the two queues can be expressed as an infinite linear combination of product forms; moreover, explicit and simple formulae are given for *all* equilibrium probabilities. These results easily lead to efficient algorithms for the computation of the equilibrium distribution and other quantities of interest, such as the mean number of jobs at the IO unit.

Noting the extension by Hofri [6] of the analysis of Kingman [7] for the SSQP to the MPQP, it seems natural to investigate whether the analysis in [1] for the SSQP applies to the MPQP as well. Such an extension appears to be possible, but it is indeed far from trivial. Nevertheless, it is much simpler than Hofri's analysis and, as in the SSQP case, it leads to more explicit results.

Besides the SSQP, also the asymmetric shortest queue problem (ASQP) may be solved by using a compensation procedure (see [2]). However, the asymmetric case is much more difficult than the symmetric case. For the ASQP one has to analyse a completely different Markov process (in this case the state space consists of two quadrants instead of one) and the equilibrium distribution appears to be a binary tree of product forms.

Let us end this section with an outline of the contents and main results of this paper. In Sect. 2 we present the model of the multiprogramming system. Next, we apply the compensation procedure to the MPQP in Sect. 3. This procedure generates an infinite linear combination of product forms; explicit and simple formulae are given for the product forms and the coefficients in this linear combination. The main result of this paper is proved in the Sect. 4 and 5 and it states that the equilibrium distribution $\{p_{m,n}\}$ in those states (m,n) for which $m+n$ is not too small, is equal to the infinite linear combination of product forms generated by the compensation procedure. In the Sects. 6 and 7 some numerical aspects of the computation of $\{p_{m,n}\}$ are discussed. For the computation of the equilibrium probabilities in those states (m,n) for which $m+n$ is not too small, bounds are derived for the errors caused by truncating the infinite linear combinations, and after that numerically stable formulae are presented to compute one by one the remaining probabili-

ties. Finally, Sect. 8 is devoted to the computation of a number of relevant quantities, such as the mean value of the number of jobs at the IO unit, while some concluding remarks are made in Sect. 9.

2. Model and problem formulation

In this section we describe the multiprogramming system as studied by Hofri [6] and we present the equilibrium equations for the equilibrium probabilities of the relevant continuous time Markov process.

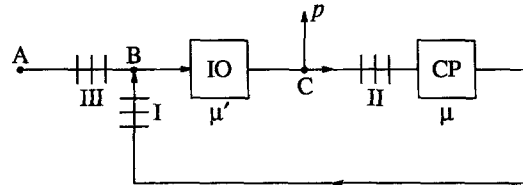


Fig. 1. The multiprogramming system

Consider the multiprogramming system as shown in Fig. 1. This system consists of an input-output unit (IO unit), a central processor (CP) and an infinite source of new jobs (queue III), which are waiting to start their service by the system with a visit to the IO unit. A variable number of "old" jobs cycles between the IO unit and the CP. After being served by the IO unit, a job leaves the system at point C with probability p , $0 < p < 1$, and joins queue II to be served by the CP with probability $1-p$. After a visit to the CP a job is recycled to the IO unit and joins queue I. If the IO unit becomes idle, the IO unit starts servicing a job from queue I. If and only if queue I is empty, the IO unit starts the service of a new job from queue III, where always jobs are available. So, jobs in queue I have nonpreemptive priority with respect to the jobs in queue III.

It is assumed that the IO unit and the CP have a FCFS service discipline, and that service times at the IO unit and the CP are exponentially distributed with parameter μ' , $\mu' > 0$, and μ , $\mu > 0$, respectively. Since the IO unit always has jobs available, it generates a stream of jobs according to a Poisson process with intensity μ' . The stream of jobs leaving the system at point C therefore is a Poisson stream with intensity

$$\eta = p\mu',$$

and the stream of jobs joining queue II is a Poisson stream with intensity

$$\lambda = (1-p)\mu'.$$

As a consequence, the CP process can be modeled as an $M|M|1$ queueing system.

The multiprogramming system may be represented by a continuous time Markov process with states (m,n) , $m, n = 0, 1, \dots$, where m represents the length of queue II,

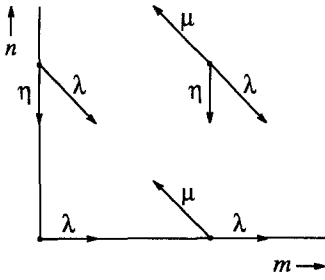


Fig. 2. The transition-rate diagram

including the job being served by the CP, and n represents the length of queue I, excluding the job being served by the IO. This process constitutes a denumerable, irreducible, aperiodic Markov chain. The transition rates of the process are illustrated in Fig. 2.

Since the Markov process is certainly not ergodic if $\lambda \geq \mu$ (in that case the Markov process which describes the length of the queue at the CP is not ergodic), we assume $\lambda < \mu$. In Sect. 5 we will see that this condition is also sufficient to prove the ergodicity of the Markov process.

Let $\{p_{m,n}\}$ be the equilibrium distribution of the Markov process. Then $\{p_{m,n}\}$ satisfies the equilibrium equations

$$\kappa p_{m,n} = \mu p_{m+1,n-1} + \lambda p_{m-1,n+1} + \eta p_{m,n+1} \quad \text{if } m > 0, n > 0, \quad (1)$$

$$(\mu + \lambda)p_{m,0} = \lambda p_{m-1,0} + \lambda p_{m-1,1} + \eta p_{m,1} \quad \text{if } m > 0, n = 0, \quad (2)$$

$$(\lambda + \eta)p_{0,n} = \mu p_{1,n-1} + \eta p_{0,n+1} \quad \text{if } m = 0, n > 0, \quad (3)$$

$$\lambda p_{0,0} = \eta p_{0,1} \quad \text{if } m = 0, n = 0, \quad (4)$$

where

$$\kappa := \mu + \lambda + \eta.$$

The determination of the equilibrium distribution $\{p_{m,n}\}$ is the objective of this paper and we call this problem the *multiprogramming queues problem (MPQP)*. For a more extensive description of the problem we refer to Hofri [6].

Remark 1 (service discipline). For clarity, we have assumed that the IO unit and the CP have service discipline FCFS. However, as long as the equilibrium equations remain the same, assuming another service discipline (for example processor sharing) makes no difference for the results in this paper.

3. The compensation procedure

Comparing the transitions for the MPQP (see Fig. 2) with the ones for the SSQP (see for example [1], Fig. 1b), we see that the structure of the transitions is nearly the same for

both problems. This is probably an explanation for the fact that Hofri [6] was able to analyse the MPQP by using the same technique as Kingman [7] used for the SSQP. To the MPQP we shall now apply the compensation procedure that has been developed in [1] for the SSQP. The compensation procedure itself is described in this section, while in the next two sections it is investigated to what extent this method yields the equilibrium distribution. We will see that the latter question raises difficulties not encountered when answering the same question for the SSQP.

Using the compensation procedure we try to construct a solution of the equilibrium equations with product forms as building blocks. This procedure only succeeds, if the equilibrium distribution can be written as a linear combination of product forms:

$$p_{m,n} = \sum_{i=0}^{\infty} c_i \alpha_i^m \beta_i^n, \quad m \geq 0, n \geq 0.$$

The first step of the compensation procedure consists of choosing an initial product form $\alpha_0^m \beta_0^n$ which asymptotically describes the equilibrium distribution, i.e. far away from the origin of the state space. If we require that $\alpha_0^m \beta_0^n$ describes the equilibrium distribution $\{p_{m,n}\}$ for large m , i.e.

$$p_{m,n} \approx C^{-1} \alpha_0^m \beta_0^n \quad (5)$$

for large m , where C is a constant, then the knowledge about the behavior of the CP yields the choice for α_0 . Since the CP behaves like an $M|M|1$ queueing system with arrival intensity λ and service rate μ , we know the marginal equilibrium distribution $\{p_{m,\cdot}\}$ of the number of jobs at the CP:

$$p_{m,\cdot} = \left(1 - \frac{\lambda}{\mu}\right) \left(\frac{\lambda}{\mu}\right)^m, \quad m \geq 0. \quad (6)$$

On the other hand $p_{m,\cdot}$ equals

$$p_{m,\cdot} = \sum_{n=0}^{\infty} p_{m,n}, \quad m \geq 0. \quad (7)$$

Now, substitution of (5) in this last expression yields

$$p_{m,\cdot} \approx C^{-1} \frac{1}{1 - \beta_0} \alpha_0^m.$$

for large m . Combination of this expression and (6) shows that we have to choose $\alpha_0 = \lambda/\mu$. The choice for the parameter β_0 now follows by the fact that requirement (5) implies that $\alpha_0^m \beta_0^n$ has to satisfy equilibrium equation (1). Substituting $\alpha_0^m \beta_0^n$ in (1) and then dividing both sides of the equation by $\alpha_0^{m-1} \beta_0^{n-1}$ yields a quadratic equation for β_0 . This result is stated more generally in the following lemma.

Lemma 1. *The product form $\alpha^m \beta^n$ is a solution of Eq. (1) if and only if α and β satisfy*

$$\alpha\beta = \mu\alpha^2 + \lambda\beta^2 + \eta\alpha\beta^2 \quad (8)$$

By Lemma 1, we find two possible values for β_0 , namely 1 and $\lambda/(\mu + \eta)$. We set $\beta_0 = \lambda/(\mu + \eta)$, because $\beta_0 = 1$ implies that for large m the equilibrium probability $p_{m,n}$ is independent of n , which is very unlikely. Besides, it does not lead to convergence of the sum of all equilibrium probabilities.

Besides (1) the product form $\alpha_0^m \beta_0^n$ also satisfies Eq. (2) on the boundary $n=0$. Because $\alpha_0^m \beta_0^n$ violates Eq. (3) on the boundary $m=0$, the next step of the compensation procedure consists of adding a correction term $c_1 \alpha^m \beta^n$ to $\alpha_0^m \beta_0^n$, such that $\alpha_0^m \beta_0^n + c_1 \alpha^m \beta^n$ satisfies (3) as well as (1). Since (3) has to be satisfied for all n , it follows immediately by substituting $\alpha_0^m \beta_0^n + c_1 \alpha^m \beta^n$ in (3) that we have to take $\beta = \beta_0$. Next, Lemma 1 yields the possible choices for α in the following way. Because of the linearity of (1), the linear combination $\alpha_0^m \beta_0^n + c_1 \alpha^m \beta_0^n$ is a solution of this equation if $\alpha^m \beta_0^n$ is a solution of this equation. By Lemma 1, the product form $\alpha^m \beta_0^n$ is a solution of (1) if α equals α_0 or α_1 , where α_0 is the already known root of the quadratic equation (8) for fixed $\beta = \beta_0$ and α_1 is the second root, $\alpha_1 = (\lambda/(\mu + \eta))^2$. Because we need a correction term, we take $\alpha = \alpha_1$. Finally, by again substituting $\alpha_0^m \beta_0^n + c_1 \alpha_1^m \beta_0^n$ in (3), we get

$$c_1 = \frac{\beta_0 - \alpha_1}{\alpha_0 - \beta_0}.$$

In a more general formulation this argument gives the following lemma.

Lemma 2.

(i) Let x_1 and x_2 be the roots of the quadratic equation (8) for fixed β , $\beta \neq 0$. Then the linear combination $k_1 x_1^m \beta^n + k_2 x_2^m \beta^n$ satisfies the Eqs. (1) and (3) if k_1 and k_2 satisfy

$$k_2 = -\frac{x_2 - \beta}{x_1 - \beta} k_1. \quad (9)$$

(ii) Let y_1 and y_2 be the roots of the quadratic equation (8) for fixed α , $\alpha \neq 0$. Then the linear combination $k_1 \alpha^m y_1^n + k_2 \alpha^m y_2^n$ satisfies the Eqs. (1) and (2) if k_1 and k_2 satisfy

$$k_2 = \frac{1 - y_2}{1 - y_1} k_1. \quad (10)$$

Proof. We only prove part (i). Part (ii) can be proved along the same lines. By Lemma 1, $x_1^m \beta^n$ and $x_2^m \beta^n$ are solutions of (1), so, by linearity, $k_1 x_1^m \beta^n + k_2 x_2^m \beta^n$ is a solution of (1) for all k_1 and k_2 . Now, by substituting $k_1 x_1^m \beta^n + k_2 x_2^m \beta^n$ in (3), we get

$$k_2 = \frac{(\lambda + \eta)\beta - \eta\beta^2 - \mu x_1}{(\lambda + \eta)\beta - \eta\beta^2 - \mu x_2} k_1.$$

Substituting in this expression the equation for the sum $x_1 + x_2$ of the roots of the quadratic equation (8) for fixed β ,

$$x_1 + x_2 = \frac{\alpha\beta - \eta\beta^2}{\mu},$$

yields (9). \square

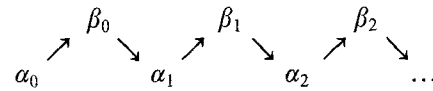
By adding a correction term to our initial solution we have corrected the error on the boundary $m=0$, but at the same time we introduced a new error on the boundary $n=0$. In the same way as above, one can prove that this new error can be corrected by adding a correction term $c_1 d_1 \alpha_1^m \beta_1^n$, where β_1 is defined as the smallest root of the quadratic equation (8) for fixed $\alpha = \alpha_1$, and d_1 is determined by Lemma 2(ii).

It is obvious that the process described above can be continued by correcting the violation of alternately Eq. (3) on the boundary $m=0$ and Eq. (2) on the boundary $n=0$. In this way the compensation procedure yields an infinite linear combination of product forms

$$\begin{aligned} x_{m,n} &= \sum_{i=0}^{\infty} (c_i \alpha_i^m + c_{i+1} \alpha_{i+1}^m) d_i \beta_i^n \\ &= c_0 d_0 \alpha_0^m \beta_0^n + \sum_{i=0}^{\infty} (d_i \beta_i^n + d_{i+1} \beta_{i+1}^n) c_{i+1} \alpha_{i+1}^m, \end{aligned} \quad (11)$$

which, hopefully, provides the equilibrium distribution up to some normalizing constant. First, we define the parameters α_i and β_i and the coefficients c_i and d_i .

The parameters α_i and β_i are defined simultaneously. For the initial values $\alpha_0 = \lambda/\mu$ and $\beta_0 = \lambda/(\mu + \eta)$ we define the sequence



such that for all $i \geq 0$ the parameters 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 $\alpha = \alpha_{i+1}$. From this definition it follows that the parameters α_i and β_i can be computed recursively by the formulae

$$\alpha_i \alpha_{i+1} = \frac{\lambda \beta_i^2}{\mu}, \quad i \geq 0, \quad (12)$$

$$\beta_i \beta_{i+1} = \frac{\mu \alpha_{i+1}^2}{\lambda + \eta \alpha_{i+1}}, \quad i \geq 0, \quad (13)$$

which both are formulae for the product of the roots of the quadratic equation (8). From these formulae it follows by induction that for all i the numbers α_i and β_i are real and positive.

By Lemma 1, all product forms $\alpha_i^m \beta_i^n$ and $\alpha_{i+1}^m \beta_{i+1}^n$ satisfy (1), and thus any linear combination of these product forms also satisfies (1). We now define for $i \geq 0$ the coefficients c_i respectively d_i such that $\{x_{m,n}\}$ also satisfies the boundary conditions. Therefore, for $i=0, 1, \dots$ the coefficients c_{i+1} and d_{i+1} have to be generated such that $(c_i \alpha_i^m + c_{i+1} \alpha_{i+1}^m) \beta_i^n$ satisfies (3) and $(d_i \beta_i^n + d_{i+1} \beta_{i+1}^n) \alpha_{i+1}^m$ satisfies (2). By Lemma 2 this yields

$$c_{i+1} = \frac{\beta_i - \alpha_{i+1}}{\alpha_i - \beta_i} c_i, \quad i \geq 0, \quad (14)$$

$$d_{i+1} = -\frac{1 - \beta_{i+1}}{1 - \beta_i} d_i, \quad i \geq 0. \quad (15)$$

For c_0 and d_0 we may take arbitrary values. We take $c_0 = 1 - \alpha_0$ and $d_0 = 1 - \beta_0$, since then the first term of $\{x_{m,n}\}$ is a normalized two-dimensional product form distribution. Besides, in this case we have the following simple formula for d_i :

$$d_i = (-1)^i (1 - \beta_i), \quad i \geq 0.$$

This completes the definition of the solution $\{x_{m,n}\}$ generated by the compensation procedure. Since we do not know whether the series $x_{m,n}$ converge, $\{x_{m,n}\}$ is called a *formal solution*.

Remark 2 (explicit formulae for α_i and β_i). It is possible to derive explicit formulae for α_i and β_i (cf. Lemma 3 in Kingman [7]). Combining (13) and the formula for the sum $\beta_i + \beta_{i+1}$ of the roots of the quadratic equation (8) for fixed $\alpha = \alpha_{i+1}$ yields

$$\frac{1}{\beta_i} + \frac{1}{\beta_{i+1}} = \frac{\kappa}{\mu \alpha_{i+1}}, \quad i \geq 0.$$

Next, adding this relation for $i-1$ and i , and eliminating α_i and α_{i+1} by using (12) and the formula for the sum $\alpha_i + \alpha_{i+1}$ of the roots of the quadratic equation (8) for fixed $\beta = \beta_i$, yields the following inhomogeneous second order linear recurrence relation for $\{1/\beta_i\}$:

$$\frac{1}{\beta_{i-1}} + \frac{2}{\beta_i} + \frac{1}{\beta_{i+1}} = \frac{\kappa}{\mu} \left(\frac{\kappa}{\lambda \beta_i} - \frac{\eta}{\lambda} \right), \quad i \geq 1.$$

This recurrence relation can be solved by standard means, yielding

$$\frac{1}{\beta_i} = A + B\tau^i + C\frac{1}{\tau^i}, \quad i \geq 0,$$

where

$$\tau = \frac{\kappa - \sqrt{\kappa^2 - 4\lambda\mu}}{\kappa + \sqrt{\kappa^2 - 4\lambda\mu}},$$

$$A = \frac{\eta\kappa}{\kappa^2 - 4\lambda\mu},$$

and B and C follow from the initial values of $\beta_0 = \lambda/(\mu + \eta)$ and $\beta_1 = \lambda^2\mu/((\mu + \eta)((\mu + \eta)^2 + \eta\lambda))$. For α_i we find $\alpha_0 = \lambda/\mu$ and

$$\frac{\kappa}{\mu \alpha_{i+1}} = 2A + B(1 + \tau)\tau^i + C\left(1 + \frac{1}{\tau}\right)\frac{1}{\tau^i}, \quad i \geq 0.$$

4. Absolute convergence of the series of product forms

In this section we prove that there exists a nonnegative integer N such that the series $x_{m,n}$ are absolutely convergent in all states (m, n) , $m \geq 0$, $n \geq 0$, $m + n \geq N$. The series $x_{m,n}$ in the other states are divergent. As a consequence, the constructed solution only satisfies the equilibrium equations in which all equilibrium probabilities $p_{m,n}$ belong to states where $x_{m,n}$ converges. So, $\{x_{m,n}\}$ only satisfies the equilibrium equations in the states (m, n) , $m \geq 0$, $n \geq 0$, $m + n \geq N$, $(m, n) \neq (N, 0)$ (compare Eqs. (1) till (4)).

Theorem 1 (absolute convergence). *There exists an integer N , $N \geq 0$, such that:*

(i) *The series $x_{m,n}$ is absolutely convergent for all $m \geq 0$, $n \geq 0$, $m + n \geq N$.*

$$(ii) \quad \sum_{\substack{m \geq 0, n \geq 0 \\ m+n \geq N}} |x_{m,n}| < \infty. \quad (16)$$

Before we can prove Theorem 1, we need some information about the behavior of the sequences $\{\alpha_i\}$, $\{\beta_i\}$, $\{c_i\}$ and $\{d_i\}$. This information is easily derived by studying the behavior of the sequences $\{u_i\}$ and $\{v_i\}$, where u_i and v_i for all $i \geq 0$ are defined by

$$u_i = \frac{\alpha_i}{\beta_i}, \quad v_i = \frac{\alpha_{i+1}}{\beta_i}. \quad (17)$$

By first considering the sequences $\{u_i\}$ and $\{v_i\}$, and next the sequences $\{\alpha_i\}$, $\{\beta_i\}$, $\{c_i\}$ and $\{d_i\}$, we get the following results.

Lemma 3.

(i) *As $i \rightarrow \infty$, then*

$$u_i \uparrow A_2 = \frac{\kappa + R}{2\mu}, \quad v_i \downarrow A_1 = \frac{\kappa - R}{2\mu}, \quad (18)$$

where

$$R = \sqrt{\kappa^2 - 4\lambda\mu}.$$

$$(ii) \quad 0 < \mu - \lambda + \eta < R < \kappa, \quad (19)$$

$$0 < A_1 < \lambda/\mu < 1 < A_2. \quad (20)$$

(iii) *As $i \rightarrow \infty$, then*

$$\frac{\alpha_{i+1}}{\alpha_i} \text{ and } \frac{\beta_{i+1}}{\beta_i} \downarrow \frac{A_1}{A_2}, \quad \frac{c_{i+1}}{c_i} \rightarrow \frac{1 - A_1}{A_1 - 1}, \quad \frac{d_{i+1}}{d_i} \rightarrow -1. \quad (21)$$

(iv) *The terms of the series (11) for $x_{m,n}$ are alternating on account of:*

$$1 > \alpha_0 > \beta_0 > \alpha_1 > \beta_1 > \dots > 0, \quad c_i > 0 \text{ for all } i \geq 0,$$

$$\frac{d_{i+1}}{d_i} < 0 \text{ for all } i \geq 0. \quad (22)$$

Proof. (i) By substituting

$$u_i v_i = \frac{\lambda}{\mu}$$

in

$$v_i + u_{i+1} = \frac{\kappa}{\mu}$$

(to prove these two equalities, one needs (12), (13) and also the formula for the sum $\beta_i + \beta_{i+1}$ of the roots of the quadratic equation (8) for fixed $\alpha = \alpha_{i+1}$), we get for $\{u_i\}$ and $\{v_i\}$ the iteration schemes

$$u_{i+1} = \frac{1}{\mu} \left(\kappa - \frac{\lambda}{u_i} \right), \quad v_i = \frac{1}{\mu} \left(\kappa - \frac{\lambda}{v_{i+1}} \right),$$

with initial values $u_0 = (\mu + \eta)/\mu$ and $v_0 = \lambda/(\mu + \eta)$. By these iteration schemes, u_i increases to A_2 and v_i decreases to A_1 , where A_1 and A_2 are the fixed points of the iteration schemes, that is, the roots of $A = 1/\mu \cdot (\kappa - \lambda/A)$.

(ii) The inequalities in (19) follow from the definition of R and the fact that $\kappa^2 - 4\lambda\mu$ can be rewritten as $(\mu - \lambda + \eta)^2 + 4\eta\lambda$. The inequalities in (20) follow from (19).

(iii) The first three limits in (21) can be derived from the limits of the sequences $\{u_i\}$ and $\{v_i\}$ after having rewritten α_{i+1}/α_i , β_{i+1}/β_i and c_{i+1}/c_i to:

$$\frac{\alpha_{i+1}}{\alpha_i} = \frac{v_i}{u_i}, \quad \frac{\beta_{i+1}}{\beta_i} = \frac{v_i}{u_{i+1}}, \quad \frac{c_{i+1}}{c_i} = \frac{1 - v_i}{u_i - 1}.$$

For the last limit in (21), one needs (15) and the fact that $\beta_i \rightarrow 0$ as $i \rightarrow \infty$, which follows from the limit behaviour of β_{i+1}/β_i and (20).

(iv) The inequalities for α_i and β_i follow from the fact that $u_i > 1$ and $v_i < 1$ for all $i \geq 0$, while the inequalities for c_i and d_{i+1}/d_i are shown by using the formulae (14) and (15) and the inequalities for α_i and β_i . \square

The knowledge about the behavior of α_i , β_i , c_i and d_i enables us to prove Theorem 1. To show the first part of this theorem, consider a fixed $m \geq 0$ and $n \geq 0$. By Lemma 3(iii), we obtain that

$$\frac{|c_{i+1} d_{i+1} \alpha_{i+1}^m \beta_{i+1}^n|}{|c_i d_i \alpha_i^m \beta_i^n|} \quad \text{and} \quad \frac{|c_{i+2} d_{i+1} \alpha_{i+2}^m \beta_{i+1}^n|}{|c_{i+1} d_i \alpha_{i+1}^m \beta_i^n|} \rightarrow \frac{1 - A_1}{A_2 - 1} \left(\frac{A_1}{A_2} \right)^{m+n}, \quad (23)$$

as $i \rightarrow \infty$. Now define N as follows.

Definition. Let N , $N \geq 0$, be the smallest integer for which

$$\frac{1 - A_1}{A_2 - 1} \left(\frac{A_1}{A_2} \right)^N < 1. \quad (24)$$

Because $A_1/A_2 < 1$ (see (20)), N is well-defined, and the

integer N has been defined such that for $m + n \geq N$ the limit in (23) is smaller than unity. As a consequence, we may conclude that for $m + n \geq N$ the series

$$\sum_{i=0}^{\infty} c_i d_i \alpha_i^m \beta_i^n \quad \text{and} \quad \sum_{i=0}^{\infty} c_{i+1} d_i \alpha_{i+1}^m \beta_i^n$$

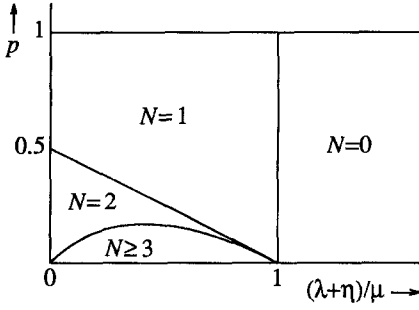
are absolutely convergent. Therefore, also the series $x_{m,n}$ is absolutely convergent. This completes the proof of Theorem 1(i). The set consisting of the states (m, n) , $m \geq 0$, $n \geq 1$, $m + n \geq N$, is called the *convergence region*, while the set consisting of the remaining states is called the *divergency region*. For the states in this divergency region, the limit in (23) is greater than or equal to unity. It is easy to see that $x_{m,n}$ diverges in those states (m, n) in the divergency region for which the limit in (23) is greater than unity. For those states for which the limit in (23) equals unity, this is more difficult to prove. At the end of this section, Remark 3 shows that in general the divergency region is rather small.

The second part of Theorem 1, needed to define a normalizing constant, is proved with the help of Lemma 3(iv) and Theorem 1(i). Due to Lemma 3(iv), the series in (16) is bounded as follows:

$$\begin{aligned} \sum_{\substack{m \geq 0, n \geq 0 \\ m+n \geq N}} |x_{m,n}| &\leq \sum_{\substack{m \geq 0, n \geq 0 \\ m+n \geq N}} \sum_{i=0}^{\infty} (c_i \alpha_i^m + c_{i+1} \alpha_{i+1}^m) |d_i| \beta_i^n \\ &= \sum_{n=0}^{N-1} \sum_{m=N-n}^{\infty} \sum_{i=0}^{\infty} (c_i \alpha_i^m + c_{i+1} \alpha_{i+1}^m) |d_i| \beta_i^n \\ &\quad + \sum_{n=N}^{\infty} \sum_{m=0}^{\infty} \sum_{i=0}^{\infty} (c_i \alpha_i^m + c_{i+1} \alpha_{i+1}^m) |d_i| \beta_i^n \\ &= \sum_{n=0}^{N-1} \sum_{i=0}^{\infty} \left(c_i \frac{\alpha_i^{N-n}}{1 - \alpha_i} + c_{i+1} \frac{\alpha_{i+1}^{N-n}}{1 - \alpha_{i+1}} \right) |d_i| \beta_i^n \\ &\quad + \sum_{i=0}^{\infty} \left(c_i \frac{1}{1 - \alpha_i} + c_{i+1} \frac{1}{1 - \alpha_{i+1}} \right) |d_i| \frac{\beta_i^N}{1 - \beta_i} \\ &\leq \sum_{n=0}^{N-1} \frac{1}{1 - \alpha_0} \sum_{n=0}^{\infty} (c_i \alpha_i^{N-n} + c_{i+1} \alpha_{i+1}^{N-n}) |d_i| \beta_i^n \\ &\quad + \frac{1}{1 - \alpha_0} \frac{1}{1 - \beta_0} \sum_{i=0}^{\infty} (c_i + c_{i+1}) |d_i| \beta_i^N. \end{aligned}$$

Since the series in this last expression converge by Theorem 1(i), we may conclude that the series in (16) converges as well.

The existence of the divergency region pinpoints the error made by Hofri [6] in his analysis of the MPQP. Hofri derived that all equilibrium probabilities could be written as infinite sums of powers, which correspond to our infinite linear combinations of product forms, and he gave explicit formulae for the equilibrium probabilities on the boundaries. However, he overlooked the possibility of divergence of these infinite sums (see [3]). The existence of the divergency region is the essential point which makes the MPQP more complicated than the SSQP.


 Fig. 3. The value of N

Remark 3 (value of N). By considering the definitions of A_1 , A_2 and N (see (18) and (24)), one can see that the value of N completely depends on $(\lambda + \eta)/\mu$ and p . In general the value of N is rather small, which is illustrated in Fig. 3, where the value of N is given as function of $(\lambda + \eta)/\mu$ and p . The boundaries between the regions in this figure are derived by writing $(1 - A_1)/(A_2 - 1) \cdot (A_1/A_2)^i$ for $i = 0, 1, 2$ as a function of $(\lambda + \eta)/\mu$ and p . We see that the divergency region is empty (i.e. $N = 0$), if and only if the IO unit works faster than the CP. By looking at the limiting behavior of $(1 - A_1)/(A_2 - 1)$ and A_1/A_2 , one can prove that $N \rightarrow \infty$ as $0 < (\lambda + \eta)/\mu < 1$ and $p \downarrow 0$.

5. The main theorem

In this section we will show that the formal solution $\{x_{m,n}\}$ restricted to the convergency region is a solution of the equilibrium equations of the Markov process restricted to this convergency region. Therefore, up to a constant $\{x_{m,n}\}$ equals the equilibrium distribution of this restricted Markov process, which in turn equals the equilibrium distribution $\{p_{m,n}\}$ of the original Markov process up to another constant. As a consequence, our main result states that in the convergency region $\{x_{m,n}\}$ equals $\{p_{m,n}\}$ up to a normalizing constant. Due to our suitable choice for the coefficients c_0 and d_0 , this normalizing constant appears to be equal to one. We remark that the restricted Markov process approach has been used earlier for the asymmetric shortest queue problem (ASQP), where a divergency region occurred as well (see [2]).

Theorem 2 (Main Theorem). For all $m \geq 0$, $n \geq 0$, $m + n \geq N$,

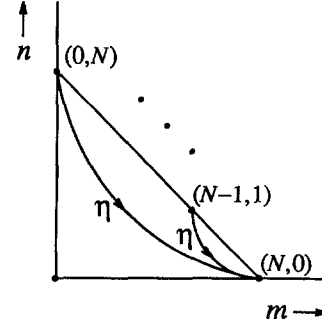
$$p_{m,n} = x_{m,n}. \quad (25)$$

Consider a time interval of infinite length and the visits of the original Markov process to the states during this interval. The process restricted to the convergency region is obtained from the original process by skipping the time intervals during which the original process is in the divergency region. In the remainder of this section the convergency region is denoted by

$$\mathcal{V}^{(N)} = \{(m, n) | m \geq 0, n \geq 0, m + n \geq N\},$$

and the equilibrium distribution of the restricted process will be denoted by $\{p_{m,n}^{(N)}\}$.

For the restricted Markov process all transition rates are equal to the corresponding transition rates for the original unrestricted process, except the transition rates from the states $(m, N - m)$, $0 \leq m \leq N - 1$, to state $(N, 0)$. The latter transition rates correspond with visits to the divergency region in the unrestricted process. Because these visits always end with a transition to state $(N, 0)$ (see Fig. 2), the transition rate of the restricted process from $(m, N - m)$, $0 \leq m \leq N - 1$, to $(N, 0)$ is equal to η (see Fig. 4).


 Fig. 4. The transition rates to state $(N, 0)$ for the process restricted to $\mathcal{V}^{(N)}$

Now, it is easy to verify that all equilibrium equations of the restricted process are equal to the corresponding equilibrium equations of the unrestricted process except for the equation in $(N, 0)$. Because our constructed solution $\{x_{m,n}\}$ satisfies the equilibrium equations in the states $\mathcal{V}^{(N)} \setminus \{(N, 0)\}$ for the unrestricted process, $\{x_{m,n}\}$ also satisfies the same equations for the restricted process. But, for this restricted process, $\{x_{m,n}\}$ then also satisfies the equilibrium equation in the state $(N, 0)$, since inserting $\{x_{m,n}\}$ into the other equations of the restricted process and then summing over these equations and changing summations (which is allowed by virtue of the absolute convergence of $\{x_{m,n}\}$, see (16)) exactly yields the desired equation.

Apart from the fact that $\{x_{m,n}\}$ is a solution of the equilibrium equations of the restricted process, we also need the fact that $\{x_{m,n}\}$ is a nonnull solution (i.e. not identical to the null solution). This latter fact can be proved by showing that for $m \geq N$ the sum of $x_{m,n}$ over $n \geq 0$ is nonnull. For $m \geq N$, we get (use expression (25) for d_{i+1})

$$\begin{aligned} \sum_{n=0}^{\infty} x_{m,n} &= \sum_{n=0}^{\infty} \left(c_0 d_0 \alpha_0^m \beta_0^n + \sum_{i=0}^{\infty} (d_i \beta_i^n + d_{i+1} \beta_{i+1}^n) c_{i+1} \alpha_{i+1}^m \right) \\ &= (1 - \alpha_0) \alpha_0^m \\ &\quad + \sum_{i=0}^{\infty} \left(d_i \frac{1}{1 - \beta_i} + d_{i+1} \frac{1}{1 - \beta_{i+1}} \right) c_{i+1} \alpha_{i+1}^m \\ &= (1 - \alpha_0) \alpha_0^m \\ &= \left(1 - \frac{\lambda}{\mu} \right) \left(\frac{\lambda}{\mu} \right)^m. \end{aligned} \quad (26)$$

Since this expression is greater than null, $\{x_{m,n}\}$ is a nonnull solution. Remark that the form of expression (26) corresponds to the fact that the CP behaves like an $M|M|1$ queueing system.

Because $\{x_{m,n}\}$ is a nonnull solution of the equilibrium equations of the restricted process, and because the sum of all $|x_{m,n}|$ is finite according to Theorem 1(ii), we may conclude by a result of Foster [5] (see Theorem 1 of that paper) that the restricted process is ergodic and the solution $\{x_{m,n}\}$ can be normalized to produce the equilibrium distribution $\{p_{m,n}^{(N)}\}$ of this restricted process:

$$p_{m,n}^{(N)} = \left(\sum_{\substack{m \geq 0, n \geq 0, \\ m+n \geq N}} x_{m,n} \right)^{-1} x_{m,n}, \quad (m,n) \in \mathcal{V}^{(N)}. \quad (27)$$

Since the number of states in the divergency region is finite, the ergodicity of the restricted process implies the ergodicity of the unrestricted process. As a consequence, in the convergency region the equilibrium distribution $\{p_{m,n}\}$ of the unrestricted process is proportional to the equilibrium distribution $\{p_{m,n}^{(N)}\}$ of the restricted process,

$$p_{m,n} = \mathbb{P}(\mathcal{V}^{(N)}) p_{m,n}^{(N)}, \quad (m,n) \in \mathcal{V}^{(N)}, \quad (28)$$

where $\mathbb{P}(\mathcal{V}^{(N)})$ represents the probability that the unrestricted process is in the convergency region. Substitution of (27) in (28) proves that there exists a constant C such that

$$p_{m,n} = C^{-1} x_{m,n}, \quad (m,n) \in \mathcal{V}^{(N)}.$$

Insertion of this relation and equation (26) into formula (7) shows that

$$p_{m,\cdot} = C^{-1} \sum_{n=0}^{\infty} x_{m,n} = C^{-1} \left(1 - \frac{\lambda}{\mu} \right) \left(\frac{\lambda}{\mu} \right)^m$$

for $m \geq N$. Comparing this expression with (6) yields that C has to be equal to $C=1$, which completes the proof of the Main Theorem.

6. Bounds needed for the computation of the $x_{m,n}$'s

By the Main Theorem, we are able to compute the equilibrium distribution $\{p_{m,n}\}$, at least in principle. We now will pay attention to some numerical aspects. The numerical part of the analysis of the MPQP is described in the next two sections. In the present section we will derive bounds for the errors caused by truncating the series $x_{m,n}$ and we will treat the computation of the remaining equilibrium probabilities in the divergency region in the next section.

The computation of the terms of the series $x_{m,n}$ needs no more attention, since the coefficient α_i, β_i, c_i and d_i can be computed recursively by the formulae (12) till (15). What does need more attention is the approximation of the series $x_{m,n}$ by partial sums. For $m \geq 0, n \geq 0, m+n \geq N$ and $k \geq 0$ we define the partial sum $x_{m,n}^k$ by

$$x_{m,n}^k = \sum_{i=0}^k (c_i \alpha_i^m + c_{i+1} \alpha_{i+1}^m) d_i \beta_i^n, \quad (29)$$

and we will derive bounds for the absolute (and relative) error in $x_{m,n}^k$ with respect to $x_{m,n}$. Remark that the terms of $x_{m,n}$ are alternating. So, if they would also be decreasing in absolute value (which is the case for the SSQP), then the error of $x_{m,n}^k$ would be bounded by the k -th term. However, numerical experiments show that in general the terms are not monotonously decreasing. Therefore, a more subtle approach is needed to obtain bounds.

For the derivation of the bound for the absolute error in $x_{m,n}^k$ a preliminary result is formulated in Lemma 4.

Lemma 4. *Let $m \geq 0, n \geq 0, m+n \geq N$ and $k \geq 0$.*

Then for all $i \geq k$,

$$(c_{i+1} \alpha_{i+1}^m + c_{i+2} \alpha_{i+2}^m) |d_{i+1}| \beta_{i+1}^n \leq R(m, n, k) (c_i \alpha_i^m + c_{i+1} \alpha_{i+1}^m) |d_i| \beta_i^n, \quad (30)$$

where $R(m, n, k)$ is defined by

$$R(m, n, k) = \frac{1 - A_1}{(\alpha_k/\beta_k) - 1} \frac{1}{1 - \beta_k} \left(\frac{\alpha_{k+1}}{\alpha_k} \right)^m \left(\frac{\beta_{k+1}}{\beta_k} \right)^n. \quad (31)$$

Proof. By Lemma 3 and the formulae (14) and (15) for c_{i+1} and d_{i+1} , we find for $i \geq k$

$$\begin{aligned} & \frac{c_{i+1} |d_{i+1}| \alpha_{i+1}^m \beta_{i+1}^n}{c_i |d_i| \alpha_i^m \beta_i^n} \\ &= \frac{\beta_i - \alpha_{i+1}}{\alpha_i - \beta_i} \frac{1 - \beta_{i+1}}{1 - \beta_i} \left(\frac{\alpha_{i+1}}{\alpha_i} \right)^m \left(\frac{\beta_{i+1}}{\beta_i} \right)^n \\ &= \frac{1 - v_i}{u_i - 1} \frac{1 - \beta_{i+1}}{1 - \beta_i} \left(\frac{\alpha_{i+1}}{\alpha_i} \right)^m \left(\frac{\beta_{i+1}}{\beta_i} \right)^n \\ &\leq \frac{1 - A_1}{u_i - 1} \frac{1}{1 - \beta_i} \left(\frac{\alpha_{i+1}}{\alpha_i} \right)^m \left(\frac{\beta_{i+1}}{\beta_i} \right)^n \\ &\leq \frac{1 - A_1}{u_k - 1} \frac{1}{1 - \beta_k} \left(\frac{\alpha_{k+1}}{\alpha_k} \right)^m \left(\frac{\beta_{k+1}}{\beta_k} \right)^n \\ &= R(m, n, k), \end{aligned}$$

and in the same way

$$\frac{c_{i+2} |d_{i+1}| \alpha_{i+2}^m \beta_{i+1}^n}{c_{i+1} |d_i| \alpha_{i+1}^m \beta_i^n} \leq R(m, n, k),$$

from which (30) follows immediately. \square

By Lemma 4 the terms of the series

$$x_{m,n} - x_{m,n}^k = \sum_{i=k+1}^{\infty} (c_i \alpha_i^m + c_{i+1} \alpha_{i+1}^m) d_i \beta_i^n$$

are decreasing in absolute value if $R(m, n, k) < 1$. Since the terms of this series are also alternating, in this case the absolute value of $x_{m,n} - x_{m,n}^k$ is bounded by the absolute value of the first term of the above series. This immediately yields Lemma 5.

Lemma 5. *Let $m \geq 0$, $n \geq 0$, $m + n \geq N$ and $k \geq 0$. If $R(m, n, k) < 1$, then*

$$|x_{m,n} - x_{m,n}^k| \leq R(m, n, k)(c_k \alpha_k^m + c_{k+1} \alpha_{k+1}^m) |d_k| \beta_k^n. \quad (32)$$

We remark that the bound in Lemma 5 for $k \rightarrow \infty$ and fixed m and n decreases slowly if $m+n \approx N$, and fastly if $m+n \gg N$, since the function $R(m, n, k)$ satisfies the following properties:

- $R(m, n, k) \downarrow R(m, n)$ as $k \rightarrow \infty$, where $R(m, n)$ is defined by
- $$R(m, n) = \frac{1 - A_1}{A_2 - 1} \left(\frac{A_1}{A_2} \right)^{m+n}.$$
- $R(m, n, k)$ decreases monotonously and exponentially as $m \rightarrow \infty$ for fixed n and k .
 - $R(m, n, k)$ decreases monotonously and exponentially as $n \rightarrow \infty$ for fixed m and k .

We see that the limit $R(m, n)$ is smaller than unity if and only if $m+n \geq N$.

Remark 4 (relative error). In this section we have derived a bound for the absolute error in $x_{m,n}^k$. But from this we can also get a bound for the relative error in $x_{m,n}^k$. Namely, if

$$|x_{m,n} - x_{m,n}^k| \leq \varepsilon_k,$$

and $|x_{m,n}^k| > \varepsilon_k$, then the relative error in $x_{m,n}^k$ is bounded by

$$\frac{|x_{m,n} - x_{m,n}^k|}{|x_{m,n}|} \leq \frac{|x_{m,n} - x_{m,n}^k|}{|x_{m,n}^k| - |x_{m,n} - x_{m,n}^k|} \leq \frac{\varepsilon_k}{|x_{m,n}^k| - \varepsilon_k}.$$

7. Computation of the equilibrium probabilities in the divergency region

In this section we discuss the computation of the equilibrium probabilities in the divergency region. Of course these probabilities can be obtained by solving the system consisting of the equilibrium equations in the divergency region. However, we will derive numerically stable formulae to compute these probabilities one by one.

With level l , $l \geq 0$, we denote the states in which there are $l+1$ jobs in the system, i.e. the states (m, n) , $m \geq 0$, $n \geq 0$, $m+n=l$. We now show that the equilibrium probabilities at level l , $l \geq 0$, can be computed given the equilibrium probabilities at level $l+1$. First, the probability $p_{l,0}$ is obtained by balancing the stream out of the set of states (m, n) , $m \geq 0$, $n \geq 0$, $m+n \leq l$, and the stream into this set of states:

$$\lambda p_{l,0} = \eta \sum_{j=0}^l p_{j,l+1-j}. \quad (33)$$

Next, the remaining equilibrium probabilities at level l are obtained by solving the system consisting of the equilibrium equations in the states $(i, l-i)$, $0 \leq i < l-1$:

$$(\lambda + \eta) p_{0,l} = \mu p_{1,l-1} + \eta p_{0,l+1}, \quad (34)$$

$$\begin{aligned} \alpha p_{i,l-i} &= \mu p_{i+1,l-i-1} + \lambda p_{i-1,l-i+1} + \eta p_{i,l+1-i} \\ 0 < i < l-1. \end{aligned} \quad (35)$$

However, this system of l equations can be reduced to l equations from which the equilibrium probabilities $p_{i,l-i}$, $0 \leq i < l-1$, can be computed one by one. Before we formulate this result in Lemma 6, we define the sequence $\{y_i\}$ by the homogeneous second order linear recurrence relation

$$y_i = \alpha y_{i-1} - \lambda \mu y_{i-2}, \quad i \leq 2, \quad (36)$$

with initial values $y_0 = 1$ and $y_1 = (\lambda + \eta)$. For this sequence an explicit formula is given in Remark 5 below, from which one easily derives that $y_i > 0$ for all $i \geq 0$.

Lemma 6. *For $i = l-1, l-2, \dots, 0$:*

$$y_{i+1} p_{i,l-i} = \mu y_i p_{i+1,l-i-1} + \sum_{j=0}^i \lambda^{i-j} \eta y_j p_{j,l+1-j}. \quad (37)$$

Proof. Relation (37) is proved by induction with respect to i . For $i=0$ relation (37) equals (34). Next, assume that (37) is valid for $i, 0 \leq i < l-1$. Then adding λ times (37) for i and y_{i+1} times (35) for $i+1$ shows that (37) is also valid for $i+1$. \square

In the formulae (33) and (37) all terms at the right hand side are positive. This means that, if the equilibrium probabilities at level $l+1$ have been computed with relative accuracy ε , then the equilibrium probabilities at level l are computed with the same relative accuracy ε , by these formulae. So, the formulae (33) and (37) are numerically stable. As a result, if the equilibrium probabilities at level N have been computed with relative accuracy ε and one computes the equilibrium probabilities at the levels $N-1, N-2, \dots, 1$ with the help of the formulae (33) and (37), then all equilibrium probabilities in the divergency region are computed with the same relative accuracy ε .

Example. The results presented so far in this paper are used in a program that computes the equilibrium probabilities up to a certain level L and with a certain relative accuracy ε . For the equilibrium probabilities that are computed with the help of the series $x_{m,n}$ (i.e. the equilibrium probabilities at the levels N up to L), this program also produces the number of iterations (= number of terms of $x_{m,n}$) needed to compute $x_{m,n}$ with the given relative accuracy ε . For the case $\mu' = 0.75$, $p = 0.2$ (so $\eta = 0.15$ and $\lambda = 0.6$) and $\mu = 1$ (in this case $N = 1$) we have computed with this program the equilibrium probabilities up to level $L = 10$ and with relative accuracy $\varepsilon = 10^{-4}$. The results are presented in Tables 1 and 2. These results have

Table 1. The equilibrium probabilities for the example

↑	10	0.0012																					
n	9	0.0023	0.0008																				
	8	0.0044	0.0015	0.0006																			
	7	0.0083	0.0029	0.0012	0.0006																		
	6	0.0158	0.0056	0.0023	0.0011	0.0006																	
	5	0.0293	0.0106	0.0044	0.0021	0.0011	0.0006																
	4	0.0520	0.0196	0.0082	0.0039	0.0021	0.0012	0.0007															
	3	0.0833	0.0346	0.0151	0.0074	0.0039	0.0022	0.0013	0.0008														
	2	0.1048	0.0546	0.0264	0.0135	0.0074	0.0042	0.0025	0.0015	0.0009													
	1	0.0779	0.0661	0.0405	0.0232	0.0134	0.0079	0.0047	0.0028	0.0017	0.0010												
	0	0.0195	0.0427	0.0446	0.0341	0.0227	0.0143	0.0088	0.0053	0.0032	0.0019	0.0012											
		0	1	2	3	4	5	6	7	8	9	10											
													m →										

Table 2. The numbers of iterations needed for the example

↑	10	2																					
n	9	2	2																				
	8	2	2	2																			
	7	2	2	2	2																		
	6	3	2	2	2	2																	
	5	3	3	2	2	2	2																
	4	4	3	3	2	2	2	2															
	3	5	4	3	3	2	2	2	2														
	2	9	5	4	3	3	2	2	2	2													
	1	33	9	5	4	3	3	2	2	2	2												
	0		34	9	6	4	3	3	2	2	2	2											
													m →										

been computed in less than one second on an IBM-compatible PC\AT with a numerical coprocessor.

Remark 5 (explicit formula for y_i). Since y_i has been defined by a homogeneous second order linear recurrence relation (see (36)), we can derive an explicit formula for y_i by standard means. We find

$$y_i = \frac{\tau_2 - (\lambda + \eta)}{\tau_2 - \tau_1} \tau_1^i + \frac{(\lambda + \eta) - \tau_1}{\tau_2 - \tau_1} \tau_2^i, \quad i \geq 0,$$

where

$$\tau_1 = \frac{\alpha - R}{2}, \quad \tau_2 = \frac{\alpha + R}{2}.$$

By using (19) we can derive that $0 < \tau_1 < (\lambda + \eta) < \tau_2$, from which we obtain $y_i > 0$ for all $i \geq 0$.

Remark 6 (reduction of computing time). As is indicated in Table 2, the computation of the equilibrium probabilities at level N with the help of the series $x_{m,n}$ needs relative large numbers of iterations, while these equilibrium probabilities can be computed with evidently less effort with the help of the formulae (33) and (37) given the equilibrium probabilities at level $N+1$. So, it may be useful to compute all desired equilibrium probabilities at the levels M and higher, $M > N$, with the help of the series

$x_{m,n}$, and the equilibrium probabilities at the levels $M-1$ and lower with the help of the formulae (33) and (37).

8. Computation of relevant quantities

Besides the equilibrium distribution $\{p_{m,n}\}$, we are also interested in quantities such as the number of jobs at the IO unit, which inform us about the performance of the multiprogramming system. A number of relevant quantities is discussed in this section.

Let X_{CP} be the number of jobs at the CP, X_{IO} the number of jobs at the IO unit, including the job being served by the IO unit, and X_{tot} the total number of jobs cycling between the CP and the IO unit, i.e. $X_{tot} = X_{CP} + X_{IO}$. Since the CP is an $M|M|1$ queue, we know the distribution of X_{CP} (see (6)) and formulae for its moments are easily derived. Expressions for the distribution and moments of X_{IO} and X_{tot} are found by expressing each quantity as a function of the equilibrium probabilities $p_{m,n}$ and subsequently using the main result stated by Theorem 2. For example, in this way we find for the first moment of X_{IO} :

$$\begin{aligned} \mathbb{E}X_{IO} &= 1 + \sum_{n=0}^{\infty} n \left(\sum_{m=0}^{\infty} p_{m,n} \right) \\ &= 1 + \sum_{n=0}^{N-1} n \left(\sum_{m=0}^{N-n-1} p_{m,n} + \sum_{m=N-n}^{\infty} x_{m,n} \right) \\ &\quad + \sum_{n=N}^{\infty} n \left(\sum_{m=0}^{\infty} x_{m,n} \right) \\ &= 1 + \sum_{n=0}^{N-1} n \left(\sum_{m=0}^{N-n-1} p_{m,n} \right) \\ &\quad + \sum_{i=0}^{\infty} \left(c_i \frac{\alpha_i^{N-n}}{1 - \alpha_i} + c_{i+1} \frac{\alpha_{i+1}^{N-n}}{1 - \alpha_{i+1}} \right) d_i \beta_i^n \\ &\quad + \sum_{i=0}^{\infty} \left(c_i \frac{1}{1 - \alpha_i} + c_{i+1} \frac{1}{1 - \alpha_{i+1}} \right) \\ &\quad \frac{\beta_i + N(1 - \beta_i)}{1 - \beta_i} d_i \beta_i^N. \end{aligned} \tag{38}$$

Table 3. Some relevant quantities computed with relative accuracy $\varepsilon = 10^{-4}$ for the case with $p = 0.2$, $\mu = 1$ and varying values for μ'

μ'	λ	N	$r(N)$	iter.	$\mathbb{E}X_{CP}$	$\sigma(X_{CP})$	$\mathbb{E}X_{IO}$	$\sigma(X_{IO})$	$\mathbb{E}X_{tot}$	$\sigma(X_{tot})$	S	$\varrho(X_{CP}, X_{IO})$
0.0125	0.01	2	0.04	3	0.01	0.10	1.84	0.46	1.85	0.45	739.3	-0.172
0.0625	0.05	2	0.17	6	0.05	0.24	1.96	0.62	2.01	0.60	161.2	-0.278
0.2500	0.20	2	0.41	12	0.25	0.56	2.28	0.94	2.53	0.91	50.6	-0.350
0.5000	0.40	2	0.39	12	0.67	1.05	2.66	1.35	3.33	1.40	33.3	-0.343
0.7500	0.60	1	0.69	34	1.50	1.94	3.11	1.88	4.61	2.27	30.8	-0.293
1.0000	0.80	1	0.38	14	4.00	4.47	3.77	2.72	7.77	4.75	38.9	-0.197
1.1875	0.95	0	0.68	38	19.00	19.49	4.59	3.87	23.59	19.62	99.3	-0.067
1.2375	0.99	0	0.62	31	99.00	99.50	4.91	4.34	103.91	99.53	419.8	-0.015

Another relevant quantity we can compute is the mean value S of the system response time, i.e. the time between the moment that a job passes point B for the first time and the moment that a job leaves the system at point C. Since the intensity with which jobs leave the system equals η , by Little's formula we have $S = \mathbb{E}X_{tot}/\eta$. The last quantity we like to mention is the coefficient of correlation $\varrho(X_{CP}, X_{IO})$ between the number of jobs at the CP and the number of jobs at the IO unit. The only extra variable that we need for this quantity is $\mathbb{E}X_{CP}X_{IO}$, for which an expression is derived in the same way as for the moments of X_{IO} and X_{tot} .

The expressions for the distribution and moments of X_{IO} and X_{tot} contain series which are similar to the series $x_{m,n}$. For these series, the same bounds hold as the bounds that we found for $x_{m,n}$, see Lemma 5. For example, formula (38) for X_{IO} contains $N+1$ series. The absolute value of the difference between

$$\sum_{i=0}^{\infty} \left(c_i \frac{\alpha_i^{N-n}}{1-\alpha_i} + c_{i+1} \frac{\alpha_{i+1}^{N-n}}{1-\alpha_{i+1}} \right) d_i \beta_i^n, \quad 0 \leq n \leq N-1,$$

and its k -th partial sum is bounded by $R(N-n, k)$ times the absolute value of the k -th term and the absolute value of the difference between

$$\sum_{i=0}^{\infty} \left(c_i \frac{1}{1-\alpha_i} + c_{i+1} \frac{1}{1-\alpha_{i+1}} \right) \frac{\beta_i + N(1-\beta_i)}{1-\beta_i} d_i \beta_i^N$$

and its k -th partial sum is bounded by $R(0, N, k)$ times the absolute value of the k -th term.

For $p=0.2$, $\mu=1$ and varying values for μ' , we have computed the mean value and the standard deviation of X_{CP} , X_{IO} and X_{tot} , the mean system response time S and the coefficient of correlation $\varrho(X_{CP}, X_{IO})$; see the right part of Table 3. In the first three columns the value of μ' and the corresponding values of λ and N are denoted. Since $\mu=1$, λ denotes the workload for the CP. Since this workload is required to be smaller than unity, μ' had to be chosen smaller than 1.25.

To gather the results listed in the table, for each case one first computes all equilibrium probabilities $p_{m,n}$ at level N , i.e. all series $x_{m,n}$ at level N , whereafter the equilibrium probabilities in the divergency region can be computed with the help of the formulae (33) and (37). Subsequently, one computes the quantities listed in Table 3. During these computations, for all series involved

$$r(N) = \frac{1-A_1}{A_2-1} \left(\frac{A_1}{A_2} \right)^N;$$

denotes the rate with which the terms of each of these series decrease in the limit. Therefore, this rate is expected to be a measure for the maximum number of iterations (= terms) needed to compute all series involved with the given accuracy. The rate $r(N)$ and this maximum number of iterations are listed in the forth and fifth column of the table.

As we see, $r(N)$ is a good measure for the maximum number of iterations. For all cases considered, the number of iterations needed for the series $x_{m,n}$ at level N appeared to be larger than for the series in the expressions of the moments of X_{IO} and X_{tot} . Of course the numbers of iterations can be decreased by computing the equilibrium probabilities at level $M > N$ by means of the series $x_{m,n}$, the equilibrium probabilities at all lower levels with the help of the formulae (33) and (37) and the moments of X_{IO} and X_{CP} as before, but with the N replaced by M in their formulae (see also Remark 6). In that case the terms of all series involved decrease with rate $r(M)$ in the limit.

Studying the right part of Table 3, we observe that for increasing μ' the values for the mean and the deviation of X_{CP} , X_{IO} and X_{tot} are also increasing. For $\mu' \uparrow 1.25$ the values for $\mathbb{E}X_{IO}$ and $\sigma(X_{IO})$ are only slowly increasing, which is due to the fact that the IO unit behaves like an $M|M|1$ queue for large μ' . The mean system response time S appears to be a convex function of μ' and $S \rightarrow \infty$ as $\mu' \downarrow 0$ or $\mu' \uparrow 1.25$. Finally, the coefficient of correlation $\varrho(X_{CP}, X_{IO})$ appears to be a concave function of μ' and $\varrho(X_{CP}, X_{IO}) \uparrow 0$ as $\mu' \downarrow 0$ or $\mu' \uparrow 1.25$.

9. Conclusions

We developed a compensation procedure to derive explicit expressions for the equilibrium probabilities of the multiprogramming queues problem (MPQP). In the convergency region (consisting of the states which are not too close to the origin of the state space) the equilibrium distribution of the lengths of the queues at the CP and the IO unit can be expressed as an infinite linear combination of product forms. Explicit formulae have been given for the product forms and the coefficients of this linear combination. The reason that this result is restricted to the convergency region is the fact that this infinite linear

combination can be divergent near the origin. This latter fact was the essential point which made the analysis for the MPQP much more difficult than for the SSQP. Apart from the above result, we derived numerical procedures for the computation of the equilibrium distribution. In particular, we determined bounds for the errors in the partial sums which approximate the infinite linear combination, and we derived numerically stable formulae to compute one by one the equilibrium probabilities in a bounded region near the origin. By these results it has been possible to obtain an efficient numerical algorithm for the computation of the equilibrium distribution. Similar results have been derived for a number of relevant quantities.

Comparing the results of this paper to the results derived by Hofri [6], we can say the following. Hofri also derived that the equilibrium distribution can be expressed as an infinite linear combination of product forms. However, he only gave explicit formulae for the equilibrium probabilities on the boundaries, and he overlooked the fact that the infinite linear combination can be divergent near the origin, by which the result is restricted

to the convergency region (compare [3] for a more complete treatment). Moreover, the analysis in the present paper is based directly on the set of equilibrium equations.

References

1. Adan IJBF, Wessels J, Zijm WHM (1990) Analysis of the symmetric shortest queue problem. *Stochastic Models* 6:691–713
2. Adan IJBF, Wessels J, Zijm WHM (1991) Analysis of the asymmetric shortest queue problem. *Queueing Syst* 8:1–58
3. Adan IJBF, Wessels J, Zijm WHM (1993) Analysing multiprogramming queues by generating functions. *SIAM J Appl Math* (to appear 1993)
4. Flatto L, McKean HP (1977) Two queues in parallel. *Commun. Pure Appl Math* 30:255–263
5. Foster FG (1953) On the stochastic matrices associated with certain queueing process. *Ann Math Statist* 24:355–360
6. Hofri M (1978) A generating-function analysis of multiprogramming queues. *Int J Comp Inform Sci* 7:121–155
7. Kingman JFC (1961) Two similar queues in parallel. *Ann Math Statist* 32:1314–1323