A product form solution to a system with multi-type jobs and multi-type servers

Jeremy Visschers · Ivo Adan · Gideon Weiss

Abstract We consider a memoryless single station service system with servers $\mathcal{S} = \{m_1, \ldots, m_K\}$, and with job types $\mathcal{C} = \{a, b, \ldots\}$. Service is skill-based, so that server m_i can serve a subset of job types $\mathcal{C}(m_i)$. Waiting jobs are served on a first-come-first-served basis, while arriving jobs that find several idle servers are assigned to a feasible server randomly. We show that there exist assignment probabilities under which the system has a product-form stationary distribution, and obtain explicit expressions for it. We also derive waiting time distributions in steady state.

Keywords Service system · First-come-first-served policy · Multi-type jobs · Multi-type servers · Partial balance · Product form solution

Mathematics Subject Classification (2000) 60K25 · 90B22

1 Introduction

In this paper we study a service (typically manufacturing) system which serves several types of jobs, labeled $a, b, c, \ldots$, and we denote the set of job types by $\mathcal{C}$. Service

Dedicated to Jaap Wessels (1939–2009).

J. Visschers
Statistics Netherlands, Postbus 4481, 6401 CZ Heerlen, The Netherlands

I. Adan (✉)
Department of Mechanical Engineering, Eindhoven University of Technology, P.O. Box 513,
5600 MB Eindhoven, The Netherlands
e-mail: iadan@tue.nl

G. Weiss
Department of Statistics, The University of Haifa, Mount Carmel 31905, Israel
e-mail: gweiss@stat.haifa.ac.il

is provided by K servers (machines), labeled $m_1, \ldots, m_K$. We denote the set of machines by $\mathcal{M}$. Jobs arrive at the system in independent Poisson streams with rates λ_i, $i \in \mathcal{C}$, and have independent service requirements which are exponentially distributed with rate 1. Each machine is capable of handling a specific subset of job types. Machine m_i can only handle jobs from the set $\mathcal{C}(m_i) \subset \mathcal{C}$. The union of these is $\mathcal{C}$. Machine m_i works at rate μ_{m_i}.

The service discipline in the system is a combination of First-Come-First-Served (FCFS) and random assignment. Arriving jobs which find no feasible available machine wait in a single queue, and are processed in an FCFS order as long as it is possible. This means that as soon as a machine finishes a job it takes the first job in the queue that it can process, possibly skipping several jobs that it cannot process. Jobs which upon arrival find some available feasible machines, are assigned to one of them randomly and go into service immediately. Else they join the end of the queue.

To fully specify the system we need to specify the random assignment: we assume that an arriving job of type i will choose a feasible machine from those which are idle according to a specified probability distribution which depends on i and on the set of idle machines. We call these distributions the *assignment probability distributions*. There is one assignment probability distribution for each type of job and for each subset of idle machines which contains at least one feasible machine for that type of job. We treat these assignment probability distributions as control parameters for the system. Figure 1 shows two examples of such systems, which we will refer to as System A and System B.

A full Markovian description of the system is to list all the jobs in the system in their order of arrivals, including jobs which are being processed, and to imagine that the machines are situated in the queue on the position of the job that they are processing. To illustrate we consider System B in Fig. 1, in which there are three job types a, b, c and three machines with $\mathcal{C}(m_1) = \{a, b\}$, $\mathcal{C}(m_2) = \{a, c\}$, $\mathcal{C}(m_3) = \{a\}$.

In Fig. 2, a possible situation of System B is depicted. The jobs are denoted by circles and the machines by rectangles. Jobs in service have a rectangle drawn around them with the identity of the machine inside. There are 12 jobs in the system, all the machines are busy. Machine m_1 is processing the first job in line, which must therefore be either of types a or b. Following the line, machine m_2 is processing the first job in the line which it can process, which is job 5 in the line, and must therefore be either of types a or c. Machine m_3 is processing the first job in the line (apart from jobs 1 and 5) which it can process, which must be of type a. There are three jobs waiting between machines m_1 and m_2. These cannot be processed by
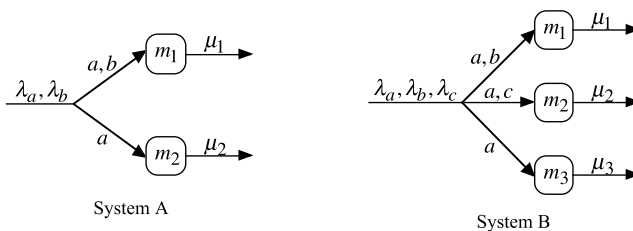

Fig. 1 Some service systems with multi-type jobs and multi-type servers
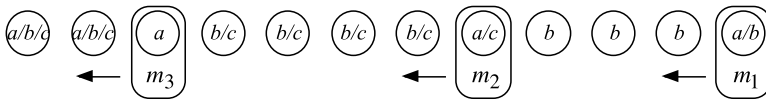
Fig. 2 A possible state for System B

either machine m_2 or by machine m_3, so they must be type-b jobs. There are four jobs waiting between m_2 and m_3. Those cannot be processed by machine m_3, so they must be of types b or c. Finally, there are two jobs at the back of the queue, behind machine m_3, which may be of types a, b, or c.

We will actually aggregate some of the states in this detailed description, to simplify the model while retaining the Markovian property. We will retain the identity and location of the busy machines, but we will not specify the type of job which they are working on, and we will only record the number of jobs between the busy machines, and not specify the string of job types. Thus the situation of Fig. 2 will be denoted as the state $(2, m_3, 4, m_2, 3, m_1)$.

With this reduced description the system is still Markovian. Our main result in this paper is to show that there exist choices of the assignment probability distributions, such that the system has a product-form stationary distribution. This product-form stationary distribution is unique, and we obtain it explicitly.

To motivate and illustrate the results we will first analyze System A of Fig. 1, with two types of jobs $\mathcal{C} = \{a, b\}$ and two servers, where $\mathcal{C}(m_1) = \{a, b\}$ and $\mathcal{C}(m_2) = \{a\}$. Here only jobs of type a have a choice of machines, and that happens only when a job of type a arrives to find both machines available; in other words, when the system is empty. Hence the assignment probability distributions reduce simply to the probability η of assigning an arriving job of type a to machine m_1 when the system is empty. This system has previously been analyzed by Adan, Foley and McDonald [1]. In Sect. 2 we solve the equilibrium equations for this system, and derive the correct assignment probability η and the product form solution. We note that this solution satisfies partial balance equations. We also show that for all other choices of η the system will not have a product form solution. Furthermore, if we choose a more detailed state description, again there will be no product form solution. These results are summarized in Theorem 1.

We then formulate and derive our main result in Sect. 3. We define the states (Sect. 3.1), write down the transition rates (Sect. 3.2), and the equilibrium equations (Sect. 3.3). We then formulate some partial balance equations (Sect. 3.4), and obtain a candidate solution (Sect. 3.5). We derive a necessary *assignment condition* (Sect. 3.6), and show in Sect. 3.7 that if the assignment condition holds then the candidate solution satisfies all the partial balance equations. This leads us to explicit conditions for ergodicity, and an explicit expression for the stationary distribution (Theorem 2). Finally, in Sect. 3.8 we show that it is always possible to choose assignment probability distributions so that the assignment condition holds. Section 3.8 is based on our recent paper [2], which considers a loss system which is similar to our system. It is shown there that the same assignment condition implies that the loss system is time-reversible and insensitive to processing time distributions.

We conjecture that when the traffic intensity approaches 1, the assignment probability distributions become less relevant and the product form solution is then a good approximation for general assignment probability distributions. In particular this leads to a model of FCFS infinite matching discussed in our paper [3].

We conclude in Sect. 4 with the derivation of the waiting time distributions of jobs of various types. It turns out that these waiting times are distributed as mixtures of sums of exponentials (Theorem 3).

Our model in this paper is formulated as a manufacturing model. However, it should find as much use also to describe, for example, skill-based routing of calls to operators in call centers, routing of wireless messages to ad hoc nodes, processing of chips, multiprocessor scheduling and mounting of printed circuit boards; see in particular [4–6, 9].

2 A system with two servers and two job types

In this section we analyze System A of Fig. 1. This system has been analyzed by Adan, Foley and McDonald [1], who found the product form solution for a special value of η, and derived exact asymptotics for the general system. Here we present a full derivation of the product form solution, to illustrate and motivate the general results in Sect. 3, and to reach several additional conclusions, summarized in Theorem 1. Note that the results in Sects. 3 and 4 are independent of the results in this section. For ease of presentation, the notation used in this section will slightly deviate from the one introduced in Sect. 1.

System A has two machines, $\mathbf{m}_1$, $\mathbf{m}_2$, and two types of jobs, a, b, where machine $\mathbf{m}_1$ can serve both types, and machine $\mathbf{m}_2$ can serve only type-a jobs. Arrivals are Poisson, with rates λ_a, λ_b, and service times are exponential with machine-dependent rates μ_i, $i = 1, 2$. Service is on an FCFS basis, and there is one control parameter:

η is the probability that a type-a job that arrives to find an empty system will be assigned to machine $\mathbf{m}_1$.

Clearly, for ergodicity it is necessary that both

$$\frac{\lambda_b}{\mu_1} < 1, \qquad \frac{\lambda_a + \lambda_b}{\mu_1 + \mu_2} < 1 \tag{1}$$

should hold. In this section we will show that the conditions above are also sufficient.

In our Markovian description the system can be in the following states:

- $(n, \mathbf{m}_2, m, \mathbf{m}_1)$, where $m, n \geq 0$, with $m + n + 2$ jobs, machine $\mathbf{m}_1$ is working on the first job followed by m jobs waiting between machine $\mathbf{m}_1$ and machine $\mathbf{m}_2$, all of which must be of type b, followed by machine $\mathbf{m}_2$ working on the $m + 2$nd job in line, followed by additional n jobs that have not yet been identified, and are waiting behind machine $\mathbf{m}_2$.
- $(n, \mathbf{m}_1, 0, \mathbf{m}_2)$, where $n \geq 0$, with $n + 2$ jobs, machine $\mathbf{m}_2$ is working on the first job, machine $\mathbf{m}_1$ is working on the second job, and there are additional n jobs of unidentified type waiting behind the two machines.
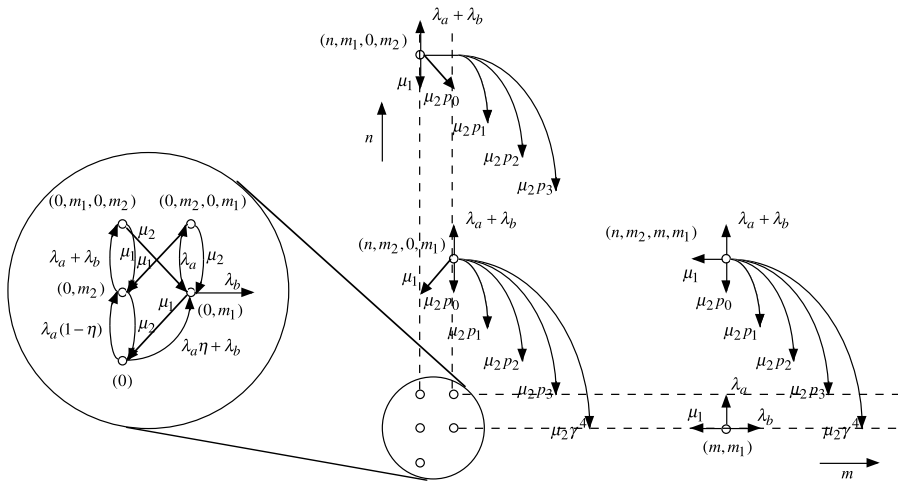
Fig. 3 State transitions for the two machine System A

- $(m, \mathbf{m}_1)$, where $m \geq 0$, machine $\mathbf{m}_1$ is working on the first job, followed by m jobs which are all of type b, and machine $\mathbf{m}_2$ is idle.
- $(0, \mathbf{m}_2)$, machine $\mathbf{m}_2$ is working on a single job in the system, and machine $\mathbf{m}_1$ is idle.
- (0) the empty system state.

The Markov process is a random walk with geometric jumps in the interior of the (m, n) positive quadrant, with modified transitions close to the axes and the origin. Arrivals join the queue or activate an available machine. On completion of service by machine $\mathbf{m}_1$, if the queues are not empty, the earliest job goes into service. On completion of service by machine $\mathbf{m}_2$, the machine moves along the line of n unidentified jobs until it finds a type-a job to process, or if none is available, it becomes idle. The jobs which are passed over in this search are now identified as type b and are queued between the two machines. Each of the unidentified n jobs at the end of the line is of type b with a probability $\gamma = \frac{\lambda_b}{\lambda_a + \lambda_b}$. Thus machine $\mathbf{m}_2$ will identify $j - 1$ jobs of type b and then find a type-a job with probability $p_j = (1 - \gamma)\gamma^{j-1}$, $j = 1, \ldots, n$, or identify all n jobs as type b and become idle with probability γ^n. The transitions and transition rates are illustrated in Fig. 3. Note that the Markov process is irreducible.

The stationary probabilities, denoted by $\pi(x)$ for state (x), need to satisfy the following equilibrium equations, for all the states excluding $(0, \mathbf{m}_1, 0, \mathbf{m}_2)$, $(0, \mathbf{m}_1)$, $(0, \mathbf{m}_2)$, (0). There are four different cases:

$$(\lambda_a + \lambda_b + \mu_1 + \mu_2)\pi(n, \mathbf{m}_2, m, \mathbf{m}_1)$$
$$= (\lambda_a + \lambda_b)\pi(n - 1, \mathbf{m}_2, m, \mathbf{m}_1) + \mu_1\pi(n, \mathbf{m}_2, m + 1, \mathbf{m}_1)$$
$$+ \mu_2(1 - \gamma)\sum_{j=0}^{m}\gamma^j\pi(n + j + 1, \mathbf{m}_2, m - j, \mathbf{m}_1)$$
$$+ \mu_2(1 - \gamma)\gamma^m\pi(n + m + 1, \mathbf{m}_1, 0, \mathbf{m}_2), \quad n > 0, m \geq 0; \tag{2}$$

$$(\lambda_a + \lambda_b + \mu_1 + \mu_2)\pi(0, \mathbf{m}_2, m, \mathbf{m}_1)$$

$$= \lambda_a \pi(m, \mathbf{m}_1) + \mu_1 \pi(0, \mathbf{m}_2, m+1, \mathbf{m}_1)$$

$$+ \mu_2(1-\gamma) \sum_{j=0}^{m} \gamma^j \pi(j+1, \mathbf{m}_2, m-j, \mathbf{m}_1)$$

$$+ \mu_2(1-\gamma)\gamma^m \pi(m+1, \mathbf{m}_1, 0, \mathbf{m}_2), \quad m \geq 0; \tag{3}$$

$$(\lambda_a + \lambda_b + \mu_1)\pi(m, \mathbf{m}_1)$$

$$= \lambda_b \pi(m-1, \mathbf{m}_1) + \mu_1 \pi(m+1, \mathbf{m}_1)$$

$$+ \mu_2 \sum_{j=0}^{m} \gamma^j \pi(j, \mathbf{m}_2, m-j, \mathbf{m}_1) + \mu_2 \gamma^m \pi(m, \mathbf{m}_1, 0, \mathbf{m}_2), \quad m > 0; \tag{4}$$

$$(\lambda_a + \lambda_b + \mu_1 + \mu_2)\pi(n, \mathbf{m}_1, 0, \mathbf{m}_2)$$

$$= (\lambda_a + \lambda_b)\pi(n-1, \mathbf{m}_1, 0, \mathbf{m}_2) + \mu_1 \pi(n+1, \mathbf{m}_1, 0, \mathbf{m}_2)$$

$$+ \mu_1 \pi(n+1, \mathbf{m}_2, 0, \mathbf{m}_1), \quad n > 0. \tag{5}$$

We attempt to derive a product form solution, in other words, we assume the solution is of the form $\pi(n, \mathbf{m}_2, m, \mathbf{m}_1) = x^m y^n$, for all $m, n \geq 0$, with some constants $0 < x, y < 1$. In the following Propositions 1–8 we derive solutions to the equations, ignoring the normalizing constant which will turn them into probabilities.

Proposition 1 *If $\pi(n, \mathbf{m}_2, m, \mathbf{m}_1) = x^m y^n$ for all $m, n \geq 0$, then $\pi(n, \mathbf{m}_1, 0, \mathbf{m}_2) = cy^n$, where $c = \frac{\gamma y}{x - \gamma y}$.*

Proof We substitute the form of the solution $\pi(n, \mathbf{m}_2, m, \mathbf{m}_1) = x^m y^n$ into (2):

$$(\lambda_a + \lambda_b + \mu_1 + \mu_2)x^m y^n = (\lambda_a + \lambda_b)x^m y^{n-1} + \mu_1 x^{m+1} y^n$$

$$+ \mu_2(1-\gamma) \sum_{j=0}^{m} \gamma^j x^{m-j} y^{n+j+1}$$

$$+ \mu_2(1-\gamma)\gamma^m \pi(n+m+1, \mathbf{m}_1, 0, \mathbf{m}_2).$$

Summing the finite geometric term and dividing by $x^m y^{n-1}$ we obtain

$$(\lambda_a + \lambda_b + \mu_1 + \mu_2)y = (\lambda_a + \lambda_b) + \mu_1 xy$$

$$+ \mu_2(1-\gamma)y^2 \frac{1 - (\frac{\gamma y}{x})^{m+1}}{1 - \frac{\gamma y}{x}}$$

$$+ \mu_2(1-\gamma)y^2 \left(\frac{\gamma y}{x}\right)^m \frac{1}{y^{n+m+1}} \pi(n+m+1, \mathbf{m}_1, 0, \mathbf{m}_2).$$

This equation can hold for every $m, n > 0$ only if:

$$\pi(n+m+1, \mathbf{m}_1, 0, \mathbf{m}_2) = \frac{\gamma y}{x - \gamma y} y^{n+m+1}. \qquad \square$$

Proposition 2 *If $\pi(n, \mathbf{m}_2, m, \mathbf{m}_1) = x^m y^n$ for all $m, n \geq 0$, then $\pi(m, \mathbf{m}_1) = dx^m$, where $d = \frac{\lambda_a + \lambda_b}{\lambda_a y}$.*

Proof Substituting the expanded candidate solution (using the result of Proposition 1) into (2) and (3), we get:

$$(\lambda_a + \lambda_b + \mu_1 + \mu_2)x^m y^n = (\lambda_a + \lambda_b)x^m y^{n-1} + \mu_1 x^{m+1} y^n$$
$$+ \mu_2(1 - \gamma)x^{m+1} y^{n+1} \frac{1}{x - \gamma y},$$

$$(\lambda_a + \lambda_b + \mu_1 + \mu_2)x^m = \lambda_a \pi(m, \mathbf{m}_1) + \mu_1 x^{m+1} + \mu_2(1 - \gamma)x^{m+1} y \frac{1}{x - \gamma y}.$$

Dividing the first equation by y^n, and comparing with the second, we get

$$\lambda_a \pi(m, \mathbf{m}_1) = (\lambda_a + \lambda_b)x^m y^{-1}. \qquad \square$$

Proposition 3 *If $\pi(n, \mathbf{m}_2, m, \mathbf{m}_1) = x^m y^n$ for all $m, n \geq 0$, where $0 < x, y < 1$, then $x = \frac{\lambda_b}{\mu_1}$ and $y = \frac{\lambda_a + \lambda_b}{\mu_1 + \mu_2}$.*

Proof Substituting the expanded candidate solution (using the results of Propositions 1 and 2) into (2) and (4), and canceling powers of x, y, we get:

$$(\lambda_a + \lambda_b + \mu_1 + \mu_2) = (\lambda_a + \lambda_b)y^{-1} + \mu_1 x + \mu_2(1 - \gamma)\frac{xy}{x - \gamma y},$$

$$(\lambda_a + \lambda_b + \mu_1)\frac{\lambda_a + \lambda_b}{\lambda_a y} = \lambda_b \frac{\lambda_a + \lambda_b}{\lambda_a y} x^{-1} + \mu_1 \frac{\lambda_a + \lambda_b}{\lambda_a y} x + \mu_2 \frac{x}{x - \gamma y}.$$

Recall that $1 - \gamma = \frac{\lambda_a}{\lambda_a + \lambda_b}$, multiply the second equation by $(1 - \gamma)y$ and subtract from the first to obtain after easy manipulations:

$$y = \frac{(\lambda_a + \lambda_b)x}{\lambda_b + \mu_2 x}.$$

Substituting back into one of the equations we obtain:

$$(\lambda_a + \lambda_b + \mu_1)\frac{\lambda_b + \mu_2 x}{\lambda_a x} = \lambda_b \frac{\lambda_b + \mu_2 x}{\lambda_a x^2} + \mu_1 \frac{\lambda_b + \mu_2 x}{\lambda_a} + \frac{\lambda_b + \mu_2 x}{x}.$$

The resulting cubic equation has roots $x = 1$ and $x = \frac{\lambda_b}{\mu_1}$ (double root). By (1), the latter is the only root in $(0, 1)$. The corresponding value of y is then: $y = \frac{\lambda_a + \lambda_b}{\mu_1 + \mu_2}$. $\square$

So far we have solved (2)–(4) and obtained values for the product form solution. Substituting the values of x, y from Proposition 3 into the expressions obtained in Propositions 1 and 2, we have in summary:

$$x = \frac{\lambda_b}{\mu_1}, \qquad y = \frac{\lambda_a + \lambda_b}{\mu_1 + \mu_2}, \qquad c = \frac{\mu_1}{\mu_2}, \qquad d = \frac{\mu_1 + \mu_2}{\lambda_a}. \tag{6}$$

These form the unique solution for (2)–(4).

Proposition 4 *The product form solution with the values of x, y, c, d of (6) satisfies (5).*

Proof Substituting the values which we obtained into (5), and dividing by y^n, we get:

$$(\lambda_a + \lambda_b + \mu_1 + \mu_2)\frac{\mu_1}{\mu_2} = (\lambda_a + \lambda_b)\frac{\mu_1}{\mu_2}\frac{\mu_1 + \mu_2}{\lambda_a + \lambda_b} + \mu_1\frac{\mu_1}{\mu_2}\frac{\lambda_a + \lambda_b}{\mu_1 + \mu_2} + \mu_1\frac{\lambda_a + \lambda_b}{\mu_1 + \mu_2}$$

which is easily seen to hold.   □

Note that this is a "lucky" solution—we had 4 parameters to determine and 4 sets of equations, but the elimination of the geometric terms already determined one of the parameters, c, and so we had 4 sets of equations for just 3 parameters. Hence, we should not regard this as a method that works for general random walks with diagonal geometric jumps.

It now remains to verify and determine the remaining values around the origin, for which we get the following 3 equilibrium equations:

$$(\lambda_a + \lambda_b + \mu_1 + \mu_2)\pi(0, \mathbf{m}_1, 0, \mathbf{m}_2) = (\lambda_a + \lambda_b)\pi(0, \mathbf{m}_2) + \mu_1\pi(1, \mathbf{m}_1, 0, \mathbf{m}_2)$$
$$+ \mu_1\pi(1, \mathbf{m}_2, 0, \mathbf{m}_1), \tag{7}$$

$$(\lambda_a + \lambda_b + \mu_1)\pi(0, \mathbf{m}_1) = (\lambda_b + \eta\lambda_a)\pi(0) + \mu_1\pi(1, \mathbf{m}_1)$$
$$+ \mu_2\pi(0, \mathbf{m}_1, 0, \mathbf{m}_2)$$
$$+ \mu_2\pi(0, \mathbf{m}_2, 0, \mathbf{m}_1), \tag{8}$$

$$(\lambda_a + \lambda_b + \mu_2)\pi(0, \mathbf{m}_2) = (1 - \eta)\lambda_a\pi(0) + \mu_1\pi(0, \mathbf{m}_1, 0, \mathbf{m}_2)$$
$$+ \mu_1\pi(0, \mathbf{m}_2, 0, \mathbf{m}_1), \tag{9}$$

in which we have two unknowns, $\pi(0, \mathbf{m}_2)$, $\pi(0)$, and three equations, with the additional control parameter η.

Proposition 5 *The unique choice of η and the solution of the remaining stationary probabilities are*:

$$\eta = \frac{\lambda_a}{2\lambda_a + \lambda_b}, \qquad \pi(0, \mathbf{m}_2) = \frac{\mu_1}{\mu_2}\frac{\mu_1 + \mu_2}{\lambda_a + \lambda_b}, \qquad \pi(0) = \frac{2\lambda_a + \lambda_b}{\lambda_a + \lambda_b}\frac{\mu_1}{\lambda_a}\frac{\mu_1 + \mu_2}{\lambda_a + \lambda_b}.$$
$$\tag{10}$$

Proof Substituting the values of the stationary probabilities obtained so far into (7), we get:

$$(\lambda_a + \lambda_b + \mu_1 + \mu_2)\frac{\mu_1}{\mu_2} = (\lambda_a + \lambda_b)\pi(0, \mathbf{m}_2) + \mu_1\frac{\mu_1}{\mu_2}\frac{\lambda_a + \lambda_b}{\mu_1 + \mu_2} + \mu_1\frac{\lambda_a + \lambda_b}{\mu_1 + \mu_2}$$
$$= (\lambda_a + \lambda_b)\pi(0, \mathbf{m}_2) + \frac{\mu_1}{\mu_2}(\lambda_a + \lambda_b),$$

from which we solve $\pi(0, \mathbf{m}_2) = \frac{\mu_1}{\mu_2}\frac{\mu_1 + \mu_2}{\lambda_a + \lambda_b}$.

Substituting all the known values into the remaining equations we get:

$$(\lambda_a + \lambda_b + \mu_1)\frac{\mu_1 + \mu_2}{\lambda_a} = (\lambda_b + \eta\lambda_a)\pi(0) + \mu_1\frac{\mu_1 + \mu_2}{\lambda_a}\frac{\lambda_b}{\mu_1} + \mu_2\frac{\mu_1}{\mu_2} + \mu_2,$$

$$(\lambda_a + \lambda_b + \mu_2)\frac{\mu_1}{\mu_2}\frac{\mu_1 + \mu_2}{\lambda_a + \lambda_b} = (1 - \eta)\lambda_a\pi(0) + \mu_1\frac{\mu_1}{\mu_2} + \mu_1,$$

which after rearrangement and canceling becomes:

$$\mu_1\frac{\mu_1 + \mu_2}{\lambda_a} = (\lambda_b + \eta\lambda_a)\pi(0),$$

$$\mu_1\frac{\mu_1 + \mu_2}{\lambda_a + \lambda_b} = (1 - \eta)\lambda_a\pi(0).$$

From this we obtain

$$\lambda_b + \eta\lambda_a = (\lambda_a + \lambda_b)(1 - \eta),$$

having as unique solution $\eta = \frac{\lambda_a}{2\lambda_a + \lambda_b}$. Substituting back into the equations we finally obtain:

$$\pi(0) = \frac{2\lambda_a + \lambda_b}{\lambda_a + \lambda_b}\frac{\lambda_b}{\lambda_a}\frac{\mu_1}{\lambda_b}\frac{\mu_1 + \mu_2}{\lambda_a + \lambda_b}.$$

□

We make several more observations:

Proposition 6 *Conditions* (1) *are necessary and sufficient for ergodicity.*

Proof Necessity is obvious. To prove sufficiency notice that the obtained product form is a non-null and, due to (1), convergent solution to the equilibrium equations of an irreducible Markov process, which implies by Theorem 1 in Foster [8] that the system is ergodic.   □

Proposition 7 *The stationary probabilities satisfy partial balance, for each state the flux out of the state due to departures equals the flux into the state due to arrivals, and (consequently) the flux out of the state due to arrivals equals the flux into the state due to departures.*

Proof It is immediate that:

$$(\mu_1 + \mu_2)\pi(n, \mathbf{m}_2, m, \mathbf{m}_1) = (\lambda_a + \lambda_b)\pi(n - 1, \mathbf{m}_2, m, \mathbf{m}_1), \quad n > 0, m \geq 0,$$

$$(\mu_1 + \mu_2)\pi(n, \mathbf{m}_1, 0, \mathbf{m}_2) = (\lambda_a + \lambda_b)\pi(n - 1, \mathbf{m}_1, 0, \mathbf{m}_2), \quad n > 0,$$

$$(\mu_1 + \mu_2)\pi(0, \mathbf{m}_2, m, \mathbf{m}_1) = \lambda_a\pi(m, \mathbf{m}_1), \quad m \geq 0,$$

$$\mu_1\pi(m, \mathbf{m}_1) = \lambda_b\pi(m - 1, \mathbf{m}_1), \quad m > 0,$$

$$(\mu_1 + \mu_2)\pi(0, \mathbf{m}_1, 0, \mathbf{m}_2) = (\lambda_a + \lambda_b)\pi(0, \mathbf{m}_2),$$

$$\mu_1 \pi(0, \mathbf{m}_1) = (\lambda_b + \eta \lambda_a) \pi(0),$$

$$\mu_2 \pi(0, \mathbf{m}_2) = (1 - \eta) \lambda_a \pi(0). \qquad \square$$

The following proposition states that use of the aggregated description of the states of System A is crucial to the existence of a product form solution. That is, if we augment the state description, by including the identity of the job being processed by machine $\mathbf{m}_1$ so that the state space includes

$$(0), \qquad (0, \mathbf{m}_2), \qquad (m, \mathbf{m}_{1a}), \qquad (m, \mathbf{m}_{1b}), \qquad (n, \mathbf{m}_{1a}, 0, \mathbf{m}_2),$$

$$(n, \mathbf{m}_{1b}, 0, \mathbf{m}_2), \qquad (n, \mathbf{m}_2, m, \mathbf{m}_{1a}), \qquad (n, \mathbf{m}_2, m, \mathbf{m}_{1b}), \quad m, n \geq 0,$$

where $\mathbf{m}_{1a}, \mathbf{m}_{1b}$ distinguish the type on which machine $\mathbf{m}_1$ is working, then there is no product form solution.

Proposition 8 *If one expands the state space to include the identity of the job being processed by machine $\mathbf{m}_1$ (which is either type a or type b), then there is no solution of the form*

$$\pi(n, \mathbf{m}_2, m, \mathbf{m}_{1a}) = B_1 x_1{}^m y_1{}^n, \qquad \pi(n, \mathbf{m}_2, m, \mathbf{m}_{1b}) = B_2 x_2{}^m y_2{}^n, \quad m, n \geq 0,$$

$$(11)$$

for the system.

Proof Substitution of the product form (11) into the new equilibrium equations for the Markov process with the expanded state space leads to the conclusion that there fail to exist values $0 < x_1, x_2, y_1, y_2 < 1$ and $B_1, B_2 > 0$ such that the product form (11) indeed satisfies all equilibrium equations; for details, see [12]. $\qquad \square$

To get the actual steady-state probabilities we need a normalizing constant B which is found by summing up the obtained solution over all states and taking the reciprocal. We summarize all findings in the following theorem, which includes the value obtained for the normalizing constant:

Theorem 1 *The system is ergodic if and only if $\frac{\lambda_b}{\mu_1} < 1$, $\frac{\lambda_a + \lambda_b}{\mu_1 + \mu_2} < 1$.*
The system has a product form solution if and only if $\eta = \frac{\lambda_a}{2\lambda_a + \lambda_b}$.
The stationary distribution in that case is

$$\pi(n, \mathbf{m}_2, m, \mathbf{m}_1) = B \left(\frac{\lambda_b}{\mu_1}\right)^m \left(\frac{\lambda_a + \lambda_b}{\mu_1 + \mu_2}\right)^n, \quad m, n \geq 0,$$

$$\pi(n, \mathbf{m}_1, 0, \mathbf{m}_2) = B \frac{\mu_1}{\mu_2} \left(\frac{\lambda_a + \lambda_b}{\mu_1 + \mu_2}\right)^n, \quad n \geq 0,$$

$$\pi(m, \mathbf{m}_1) = B \frac{\mu_1 + \mu_2}{\lambda_a} \left(\frac{\lambda_b}{\mu_1}\right)^m, \quad m \geq 0,$$

$$\pi(0, \mathbf{m}_2) = B \frac{\mu_1}{\mu_2} \frac{\mu_1 + \mu_2}{\lambda_a + \lambda_b},$$

$$\pi(0) = B \frac{2\lambda_a + \lambda_b}{\lambda_a + \lambda_b} \frac{\mu_1}{\lambda_a} \frac{\mu_1 + \mu_2}{\lambda_a + \lambda_b},$$

where the normalizing constant is given by

$$B = \frac{\mu_2 \lambda_a (\lambda_a + \lambda_b)^2 (\mu_1 - \lambda_b)(\mu_1 + \mu_2 - \lambda_a - \lambda_b)}{\mu_1(\mu_1 + \mu_2)(\mu_2^2 \lambda_a^2 + \mu_1 \lambda_a (\mu_1 - \lambda_b)(\lambda_a + \lambda_b) + \mu_1 \mu_2 (2\lambda_a + \lambda_b)(\mu_1 + \mu_2 - \lambda_b))}.$$

The product form solution satisfies partial balance: *flux into a state due to arrivals equals flux out of that state due to departures.*

If one expands the state space to include the identity of the job in service by machine $\mathbf{m}_1$, *then there is no product form solution.*

3 The multidimensional model

In this section we consider the general K-machine system. In Sect. 3.1 we describe the state space which we use to describe this system. In Sect. 3.2 we describe the Markovian transition probabilities and rates. In Sect. 3.3 we formulate the equilibrium equations for the model. In Sect. 3.4 we formulate *partial balance equations* and in Sect. 3.5 we use them to obtain a product form candidate solution to the equilibrium equations. In Sect. 3.6 we formulate the *assignment condition* and in Sect. 3.7 we show that under the assignment condition the partial balance equations are satisfied by the candidate solution, i.e. this product form solution is the stationary distribution of the Markov chain. Finally, we use the results of Adan, Hurkens and Weiss [2] to show how it is always possible to calculate *assignment probability distributions* so as to satisfy the assignment condition.

We introduce the following notation:

$M :=$ an arbitrary machine M from the set of machines $\mathcal{M} = \{m_1, \ldots, m_K\}$. The capitalized M points to one of the machines m_i. Note that the names (or labels) of the machines m_i are not capitalized.

$$\lambda_\mathcal{X} := \sum_{c \in \mathcal{X}} \lambda_c, \quad \text{where } \mathcal{X} \subset \mathcal{C}.$$

$$\mu_\mathcal{Y} := \sum_{M \in \mathcal{Y}} \mu_M, \quad \text{where } \mathcal{Y} \subset \mathcal{M}.$$

$\mathcal{C}(\mathcal{Y}) :=$ total set of job types that can be handled by the machines in $\mathcal{Y} \subset \mathcal{M}$, which is equal to $\bigcup_{M \in \mathcal{Y}} \mathcal{C}(M)$.

$\mathcal{U}(\mathcal{Y}) :=$ set of job types unique to the machines in $\mathcal{Y} \subset \mathcal{M}$, thus the set of job types that cannot be handled by machines outside $\mathcal{Y}$. We have $\mathcal{U}(\mathcal{Y}) = \overline{\mathcal{C}(\overline{\mathcal{Y}})}$.

3.1 The state space for the multidimensional system

For models with K machines we use the following state description:

Fig. 4 General system in state $\mathfrak{s} = (n_i, M_i, \ldots, n_2, M_2, n_1, M_1)$
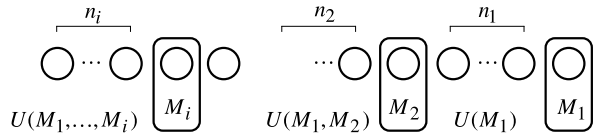
$(n_i, M_i, n_{i-1}, M_{i-1}, \ldots, n_1, M_1)$: *States in which there are i machines busy. These machines are denoted by $M_1, \ldots, M_i$, where $\{M_1, \ldots, M_i\} \subset \mathcal{M}$. The number of jobs between machines M_j and M_{j+1} is denoted by n_j (≥ 0), with $j = 1, \ldots, i - 1$. The number of waiting jobs at the end of the queue, behind machine M_i, is denoted by n_i.*

The state space is denoted by $\mathcal{S}$ and to simplify the notation we use $\mathfrak{s}$ to denote an arbitrary state $(n_i, M_i, \ldots, n_1, M_1) \in \mathcal{S}$. Figure 4 shows a system in state $\mathfrak{s}$. There are a few things that are important to note about this state description:

First, the waiting jobs between machines M_j and M_{j+1} can only be handled by the machines $M_1, \ldots, M_j$ and not by any of the machines $M_{j+1}, \ldots, M_i$ or any of the idle machines. This is due to the first-come-first-served processing order. Thus, waiting jobs between machines M_j and M_{j+1} can only be of type $c \in \mathcal{U}(\{M_1, \ldots, M_j\})$, according to the definition of $\mathcal{U}(\mathcal{Y})$. The n_i waiting jobs in the back of the queue cannot be handled by any of the idle machines and have to be of type $c \in \mathcal{U}(\{M_1, \ldots, M_i\})$.

Second, since each part of the queue between two machines contains jobs from different subsets of job types, it is necessary to keep these sets separated in the state description. It is not possible to aggregate the state description any further without losing the Markov property.

Third, it is possible that the set of job types $\mathcal{U}(\{M_1, \ldots, M_j\})$ is empty for a certain set of machines $\{M_1, \ldots, M_j\}$. In this case there are no jobs which cannot be handled by any of the machines $M_{j+1}, \ldots, M_i$ or the idle machines. Thus there can be no waiting jobs between M_j and M_{j+1}, and therefore n_j can only be equal to zero. Hence, the state space is given by

$$\mathcal{S} = \{0\} \cup \bigcup_{i=1}^{K} \Big\{(n_i, M_i, \ldots, n_1, M_1) \,|\, \{M_1, \ldots, M_i\} \subset \mathcal{M}, n_j \geq 0,$$

$$n_j = 0 \text{ if } \mathcal{U}(\{M_1, \ldots, M_j\}) = \emptyset, j = 1, \ldots, i\Big\},$$

where state 0 denotes the empty state (in which all machines are idle).

Fourth, it is important to note that in this state description we lose job type information about the jobs that are in service, since we only denote the machine that is handling the job and not the job type of the job. This aggregation preserves the Markov property since all types are processed by machine m_i at rate μ_i. In Sect. 2 we already argued that a more detailed state description, where this job type information is included in the state description, does not result in a product form solution for System A. It is reasonable to conjecture that specifying the job types in process will destroy the possibility of a product form solution also in the multidimensional model.

Fifth, note that the Markov process on $\mathcal{S}$ is irreducible. Clearly, it is possible to reach the empty state 0 from any other state $\mathfrak{s} \in \mathcal{S}$. To see that any state $\mathfrak{s} = (n_i, M_i, \ldots, n_1, M_1)$ can be reached from 0, observe that starting from 0, with positive probability, the first K events are arrivals occupying all machines. Then, again with positive probability, the next $K - i$ events are departures from machines $M \notin \{M_1, \ldots, M_i\}$ and we end up in a state where only machines $M_i, \ldots, M_1$ are busy and no jobs are waiting. Then, the next $n_i + \cdots + n_1 + i$ events are arrivals that can exclusively be processed by the busy machines $M_i, \ldots, M_1$. Subsequently, with positive probability, the next i events are departures from each of the machines $M_1, \ldots, M_i$. First M_1 finishes a job and is able to handle the first waiting job. Then M_2 finishes a job, skips that first n_1 waiting jobs and handles the next one, and so on, until we eventually reach state $\mathfrak{s} = (n_i, M_i, \ldots, n_1, M_1)$ when M_i finishes its job, skips the first $n_1 + \cdots + n_{i-1}$ waiting jobs and starts processing the next one.

3.2 The transition behavior

We discuss the possible transitions of the general model below.

From an arbitrary state $\mathfrak{s} = (n_i, M_i, \ldots, n_1, M_1) \in \mathcal{S}$ the following transitions are possible:

(i) *Arrival of jobs*: If a job arrives that cannot be handled by any of the idle machines, it joins the back of the queue. In state $\mathfrak{s}$, arriving jobs that cannot be handled by any of the idle machines must be of type $c \in \mathcal{U}(\{M_1, \ldots, M_i\})$. Thus such a job arrives with rate $\lambda_{\mathcal{U}(\{M_1,\ldots,M_i\})}$. With rate $\lambda_{\mathcal{C}} - \lambda_{\mathcal{U}(\{M_1,\ldots,M_i\})}$ a job arrives that can be handled by one or more idle machines. For this job an assignment probability distribution determines to which machine it is sent. The assignment probability distributions determine the total transition rate from a state $\mathfrak{s}$ to a state $(0, M, \mathfrak{s})$.

To simplify the discussion we introduce the following notation:

$\lambda_M(\{M_1, \ldots, M_i\}) :=$ transition rate from state $\mathfrak{s} = (n_i, \overline{M}_i, \ldots, n_1, \overline{M}_1)$ to state $(0, M, \mathfrak{s})$, for all possible permutations $\overline{M}_i, \ldots, \overline{M}_1$ of the sequence $M_i, \ldots, M_1$ and all machines M that are not busy in state $\mathfrak{s}$; this rate is also referred to as the *activation rate* (of machine M).

Note that the activation rates $\lambda_M(\{M_1, \ldots, M_i\})$ only depend on the set of busy machines $\{M_1, \ldots, M_i\}$ and not on the sequence of machines $M_i, \ldots, M_1$ and thus not on the order of the machines in the queue. Given the assignment probability distributions, the arrival rates can be calculated. Note that the following holds:

$$\sum_{M \in \mathcal{M} \setminus \{M_1,\ldots,M_i\}} \lambda_M(\{M_1, \ldots, M_i\}) = \lambda_{\mathcal{C}} - \lambda_{\mathcal{U}(\{M_1,\ldots,M_i\})}. \tag{12}$$

We discuss the choice of assignment probabilities further in Sects. 3.6, 3.8.

(ii) *Departure of jobs*: If a job is finished on a machine, then that machine scans the queue from right to left until it finds the first job it can handle. There are two possibilities:
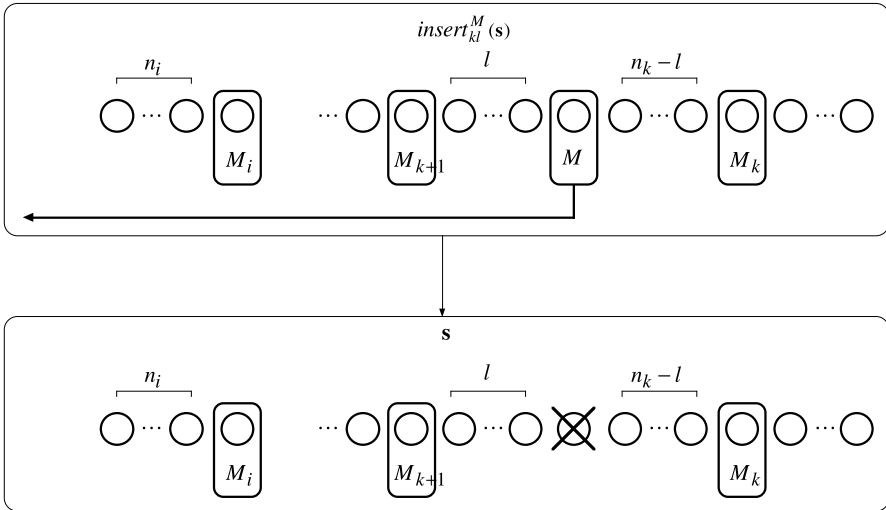
Fig. 5 Transition from state $insert_{kl}^M(\mathfrak{s})$ to state $\mathfrak{s} = (n_i, M_i, \ldots, n_1, M_1)$

(1) *The machine does not find a job it can handle.* This transition is illustrated in Fig. 5. Such a transition is possible to state $\mathfrak{s} = (n_i, M_i, \ldots, n_1, M_1)$ from state $(n_i, M_i, \ldots, l, M, n_k - l, M_k, \ldots, n_1, M_1)$, where machine M is situated in the queue between M_k and M_{k+1}. We denote this state by $insert_{kl}^M(\mathfrak{s})$. In this state there are $n_k - l$ jobs between M_k and M and l jobs between M and M_{k+1}. The jobs between machines M and M_{k+1} can only be of type $c \in \mathcal{U}(\{M_1, \ldots, M_k, M\})$, thus with probability

$$\frac{\lambda_{\mathcal{U}(\{M_1, \ldots, M_k\})}}{\lambda_{\mathcal{U}(\{M_1, \ldots, M_k, M\})}},$$

such a job cannot be handled by machine M. A job between machine M_j and M_{j+1}, $j > k$, cannot be handled by machine M with probability

$$\frac{\lambda_{\mathcal{U}(\{M_1, \ldots, M_j\})}}{\lambda_{\mathcal{U}(\{M_1, \ldots, M_j, M\})}}.$$

Thus if machine M finishes its job then, with probability $p_{kl}^M(\mathfrak{s})$, it will not find a job in the queue that it can handle and will become idle, where $p_{kl}^M(\mathfrak{s})$ is defined as

$$p_{kl}^M(\mathfrak{s}) := \delta_k(M)^l \delta_{k+1}(M)^{n_{k+1}} \cdots \delta_i(M)^{n_i},$$

with

$$\delta_j(M) := \frac{\lambda_{\mathcal{U}(\{M_1, \ldots, M_j\})}}{\lambda_{\mathcal{U}(\{M_1, \ldots, M_j, M\})}}, \quad j = 1, \ldots, i. \tag{13}$$
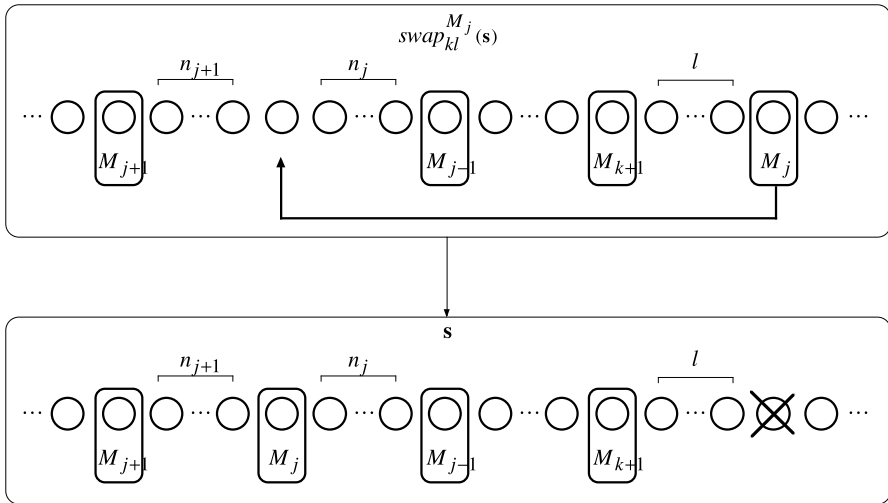
Fig. 6 Transition from state $swap_{kl}^{M_j}(\mathfrak{s})$ to state $\mathfrak{s}$

We set, by convention, $\delta_j(M) = 0$ if $\mathcal{U}(\{M_1, \ldots, M_j\}) \subseteq \mathcal{U}(\{M_1, \ldots, M_j, M\}) = \emptyset$. Thus, with probability $p_{kl}^M(\mathfrak{s})$, a jump is made from state $insert_{kl}^M(\mathfrak{s})$ to state $\mathfrak{s}$, given that machine M finishes its job.

In the special case of $k = 0$, before the transition into $\mathfrak{s}$, machine M is working on the first job in the queue, and in that case $l = 0$, since otherwise all jobs between the first and second busy machines must be of type in $\mathcal{U}(\{M\})$, and machine M will be able to process all of them. Hence, for $k = 0$ the only transition in which machine M becomes idle is from $insert_{00}^M(\mathfrak{s})$ to $\mathfrak{s}$, and, in fact,

$$p_{00}^M(\mathfrak{s}) = p_{1n_1}^M(\mathfrak{s}).$$

(2) *The machine finds a job it can handle.* This transition is illustrated in Fig. 6. In this case one of the busy machines finishes its job and finds somewhere in the queue another job that it can process. The state $\mathfrak{s}$ can be reached by such a transition from state

$$(n_i, M_i, \ldots, n_{j+1}, M_{j+1}, n_j + n_{j-1} + 1, M_{j-1}, \ldots,$$
$$M_{k+1}, l, M_j, n_k - l, M_k, \ldots, n_1, M_1).$$

This state will be denoted by $swap_{kl}^{M_j}(\mathfrak{s})$ with $j - 1 \geq k$. In this state, machine M_j is located between machines M_k and M_{k+1}. Between machines M_{j-1} and M_{j+1} there are $n_{j-1} + n_j + 1(> 0)$ jobs; this is of course only possible if $\mathcal{U}(\{M_1, \ldots, M_j\}) \neq \emptyset$. If machine M_j finishes its job, a transition is made to state $\mathfrak{s}$ if the first job that M_j can handle is the $(n_{j-1} + 1)$-th job (from the right) between M_{j-1} and M_{j+1}. The probability of this event is

$$q_{kl}^{M_j}(\mathfrak{s}) := \delta_k(M_j)^l \delta_{k+1}(M_j)^{n_{k+1}} \cdots \delta_{j-1}(M_j)^{n_{j-1}} \left(1 - \delta_{j-1}(M_j)\right),$$

with $\delta_j(M)$ defined in (13). The system makes a jump from state $swap_{kl}^{M_j}(\mathfrak{s})$ to state $\mathfrak{s}$ with probability $q_{kl}^{M_j}(\mathfrak{s})$, given that M_j finishes its job.

In the special case that $k = j - 1$, machine M_j starts and ends its move between machines M_{j-1} and M_j, and there are initially $n_j + 1 + l$ jobs between M_j and M_{j+1} and $n_{j-1} - l$ jobs between M_{j-1} and M_j. This state is denoted by $swap_{j-1,l}^{M_j}(\mathfrak{s})$ and the probability that a transition is made from this state to state $\mathfrak{s}$ equals

$$q_{j-1,l}^{M_j}(\mathfrak{s}) = \delta_{j-1}(M_j)^l \big(1 - \delta_{j-1}(M_j)\big).$$

Slightly different is the special case that $k = 0$, where there are two possibilities: In the transition from $swap_{0n_1}^{M_1}(\mathfrak{s})$ to $\mathfrak{s}$, machine M_1 was working on the first job, and there were $n_1 + 1$ jobs queued between it and machine M_2, and the machine moved to the next job. Because the next job is in $\mathcal{U}(\{M_1\})$, the probability for this transition is

$$q_{0n_1}^{M_1}(\mathfrak{s}) = 1.$$

Otherwise, if $j > 1$ then $l = 0$ and

$$q_{00}^{M_j}(\mathfrak{s}) = q_{1n_1}^{M_j}(\mathfrak{s}).$$

3.3 Equilibrium equations

We can now formulate the set of equilibrium equations. The equilibrium probability of being in the state $\mathfrak{s}$ is denoted by $\pi(\mathfrak{s})$. The state $\mathfrak{s}$ can be reached by (i) an arrival of a job, (ii) a departure of a job from a machine that finds no new job in the queue, and (iii) a departure of a job from a machine that does find a new job in the queue. The equilibrium equations display these three possibilities. The left-hand side of the equations equals the total probability flux out of state $\mathfrak{s}$. The right-hand side of the equations equals the probability flux into state $\mathfrak{s}$ and consists of three parts, corresponding to respectively (i), (ii) and (iii). In part (ii) we need to sum over all possible states with one more busy machine (machine M) and over all possible positions of this machine in the queue. In part (iii) we need to sum over all machines $M_j \in \{M_1, \ldots, M_i\}$ for which $\mathcal{U}(\{M_1, \ldots, M_j\}) \neq \emptyset$, and over the positions of machine M_j in the queue. For all states $\mathfrak{s} = (n_i, M_i, \ldots, n_1, M_1) \in \mathcal{S} \setminus \{0\}$ the equilibrium equations are given by: for $n_i > 0$,

$$(\lambda_{\mathcal{C}} + \mu_{\{M_1, \ldots, M_i\}})\pi(\mathfrak{s}) = \lambda_{\mathcal{U}(\{M_1, \ldots, M_i\})}\pi(n_i - 1, M_i, \ldots, n_1, M_1) \qquad \text{(i)}$$

$$+ \sum_{M \in \mathcal{M} \setminus \{M_1, \ldots, M_i\}} \mu_M \mathcal{P}_M(\mathfrak{s}) \qquad \text{(ii)}$$

$$+ \sum_{\substack{j=1 \\ \mathcal{U}(\{M_1, \ldots, M_j\}) \neq \emptyset}}^{i} \mu_{M_j} \mathcal{Q}_{M_j}(\mathfrak{s}); \qquad \text{(iii)} \quad \text{(14)}$$

for $n_i = 0$,

$$(\lambda_{\mathcal{C}} + \mu_{\{M_1,\ldots,M_i\}})\pi(\mathfrak{s}) = \lambda_{M_i}(\{M_1,\ldots,M_{i-1}\})\pi(n_{i-1},M_{i-1},\ldots,n_1,M_1) \quad \text{(i)}$$

$$+ \sum_{M \in \mathcal{M}\backslash\{M_1,\ldots,M_i\}} \mu_M \mathcal{P}_M(\mathfrak{s}) \quad \text{(ii)}$$

$$+ \sum_{\substack{j=1 \\ \mathcal{U}(\{M_1,\ldots,M_j\}) \neq \emptyset}}^{i} \mu_{M_j} \mathcal{Q}_{M_j}(\mathfrak{s}); \quad \text{(iii)}$$

$$\quad (15)$$

where

$$\mathcal{P}_M(\mathfrak{s}) = \sum_{k=1}^{i} \sum_{l=0}^{n_k} p_{kl}^M(\mathfrak{s})\pi\big(insert_{kl}^M(\mathfrak{s})\big) + p_{1n_1}^M(\mathfrak{s})\pi(\mathfrak{s},0,M), \quad (16)$$

$$\mathcal{Q}_{M_j}(\mathfrak{s}) = \sum_{k=1}^{j-1} \sum_{l=0}^{n_k} q_{kl}^{M_j}(\mathfrak{s})\pi\big(swap_{kl}^{M_j}(\mathfrak{s})\big) + q_{1n_1}^{M_j}\pi\big(swap_{00}^{M_j}(\mathfrak{s})\big). \quad (17)$$

We may omit the equation for state 0, since the set of equilibrium equations is dependent.

3.4 Partial balance equations

We now decompose the equilibrium equations for the model into partial balance equations. We will show that if these equations are satisfied, the model has a product form solution. It appears that the partial balance equations are satisfied and thus that the model has a product form distribution, if a so-called *assignment condition* is satisfied. The assignment condition is discussed in Sects. 3.6 and 3.8.

Since the job type of a departing job is not known (a departing job from machine M can be of any job type $c \in \mathcal{C}(M)$), it is more appropriate to consider station- or machine-balance, instead of job-type-balance. For state $\mathfrak{s}$ this results in the following equations:

(i) One equation for each idle machine $M \in \mathcal{M}\backslash\{M_1,\ldots,M_i\}$: *the total probability flux out of state $\mathfrak{s}$ due to an arrival of a job that is taken into service by machine M (activation rate $\lambda_M(\{M_1,\ldots,M_i\})$) equals the total probability flux into state $\mathfrak{s}$ due to a departure of a job on machine M after which the machine becomes idle*. This results in the equations:

$$\lambda_M(\{M_1,\ldots,M_i\})\pi(\mathfrak{s}) = \mu_M \mathcal{P}_M(\mathfrak{s}), \quad M \in \mathcal{M}\backslash\{M_1,\ldots,M_i\}. \quad (18)$$

(ii) One equation for the set of busy machines $\{M_1,\ldots,M_i\}$: *the total probability flux out of state $\mathfrak{s}$, due to an arrival of a job that can only be processed on one or more machines in $\{M_1,\ldots,M_i\}$ and not on any of the idle machines in $\mathcal{M}\backslash\{M_1,\ldots,M_i\}$ (arrival rate $\lambda_{\mathcal{U}(\{M_1,\ldots,M_i\})}$), equals the total probability flux*

into state $\mathfrak{s}$, due to the departure of a job on one of the machines, which finds another job in the queue, so that the set of busy machines $\{M_1, \dots, M_i\}$ remains the same. This equation cannot be divided into one equation for every machine $M_j \in \{M_1, \dots, M_i\}$, since upon arrival it is not possible to say which machine M_j will process the arriving job. This results in the equation:

$$\lambda_{\mathcal{U}(\{M_1, \dots, M_i\})} \pi(\mathfrak{s}) = \sum_{\substack{j = 1 \\ \mathcal{U}(\{M_1, \dots, M_j\}) \neq \emptyset}}^{i} \mu_{M_j} \mathcal{Q}_{M_j}(\mathfrak{s}). \tag{19}$$

3.5 Candidate product form solution

With the formulated partial balance equations it is easy to derive the candidate product form solution. This can be done for every $\mathfrak{s}$, by subtracting the sum of (18) for all machines $M \in \mathcal{M} \setminus \{M_1, \dots, M_i\}$ plus (19) from (14) (and by using (12)), and similarly from (15). This yields the following equations:

$$\mu_{\{M_1, \dots, M_i\}} \pi(\mathfrak{s}) = \lambda_{\mathcal{U}(\{M_1, \dots, M_i\})} \pi(n_i - 1, M_i, \dots, n_1, M_1), \quad n_i > 0, \tag{20}$$

$$\mu_{\{M_1, \dots, M_i\}} \pi(\mathfrak{s}) = \lambda_{M_i}\big(\{M_1, \dots, M_{i-1}\}\big) \pi(n_{i-1}, M_{i-1}, \dots, n_1, M_1), \quad n_i = 0. \tag{21}$$

It is clear that these equations yield the candidate product form solution given by

$$\pi(\mathfrak{s}) = \alpha_i^{n_i} \frac{\lambda_{M_i}(\{M_1, \dots, M_{i-1}\})}{\mu_{\{M_1, \dots, M_i\}}} \cdots \alpha_1^{n_1} \frac{\lambda_{M_1}(\emptyset)}{\mu_{\{M_1\}}} \pi(0), \tag{22}$$

where

$$\alpha_j := \frac{\lambda_{\mathcal{U}(\{M_1, \dots, M_j\})}}{\mu_{\{M_1, \dots, M_j\}}}, \quad j = 1, 2, \dots, i. \tag{23}$$

It is also clear that if this product form solution satisfies (18) for all machines $M \in \mathcal{M} \setminus \{M_1, \dots, M_i\}$ and (19), then also (14) and (15) are satisfied. Thus, if the candidate solution (22) satisfies the partial balance (18) and (19), then the model has a product form solution.

3.6 Assignment condition

Because $\lambda_{M_i}(\{M_1, \dots, M_{i-1}\})$ still depends on the control parameters, the partial balance equations (18) and (19) will not be satisfied for every value of the control parameters. We show in the proof of our main result in the following section that these equations are satisfied if the control parameters are chosen such that the following condition is satisfied:

Assignment condition
For $i = 1, \dots, K$, and for every subset $\{M_i, \dots, M_1\} \in \mathcal{M}$ of size i, the following holds:

$$\prod_{j=1}^{i} \lambda_{M_j}\big(\{M_1, \ldots, M_{j-1}\}\big) = \prod_{j=1}^{i} \lambda_{\overline{M}_j}\big(\{\overline{M}_1, \ldots, \overline{M}_{j-1}\}\big)$$

for every permutation $\overline{M}_i, \ldots, \overline{M}_1$ of $M_i, \ldots, M_1$.

This condition implies that the product $\prod_{j=1}^{i} \lambda_{M_j}(\{M_1, \ldots, M_{j-1}\})$ should be independent of the order of the machines $M_i, \ldots, M_1$ in the queue.

We assume for now that the assignment condition is satisfied. We show in Sect. 3.8 that it is always possible to choose the assignment probability distributions in such a way that the assignment condition is satisfied. We also show that while the assignment probability distributions which achieve that may not be unique, the resulting activation rates $\lambda_M(\{M_1, \ldots, M_i\})$ are unique.

We can simplify the notation by introducing

$$\Pi_\lambda\big(\{M_1, \ldots, M_i\}\big) := \prod_{j=1}^{i} \lambda_{M_j}\big(\{M_1, \ldots, M_{j-1}\}\big), \quad \text{for all } \{M_1, \ldots, M_i\} \in \mathcal{M},$$

$$\Pi_\mu(M_i, \ldots, M_1) := \prod_{j=1}^{i} \mu_{\{M_1, \ldots, M_j\}}, \quad \text{for all } (M_i, \ldots, M_1) \in \mathcal{M}^i,$$

where $\mathcal{M}^i$ is defined as the set of all possible sequences $(M_i, \ldots, M_1)$ of all possible subsets $\{M_1, \ldots, M_i\} \subset \mathcal{M}$ of size i. Note that $\Pi_\lambda(\{M_1, \ldots, M_i\})$ is independent of the order of the machines in the sequence $M_i, \ldots, M_1$, since we assumed that the assignment condition is satisfied and that $\Pi_\mu(M_i, \ldots, M_1)$ is *not*. Now that we have formulated the *assignment condition*, we can state our main result.

3.7 Main result

We have formulated the equilibrium equations and derived a candidate product form solution with the use of partial balance. This candidate product form solution is the solution of the global balance equations (14) and (15) if the assignment condition is satisfied. This is shown in the proof of the following theorem.

Theorem 2 *The model described in Sect. 3.1 has a product form solution if the control parameters of the model (the assignment probability distributions) are chosen in such a way that the assignment condition is satisfied. Then, for all states $\mathfrak{s} = (n_i, M_i, \ldots, n_1, M_1) \in \mathcal{S}$, the solution is*

$$\pi(\mathfrak{s}) = \alpha_i^{n_i} \cdots \alpha_1^{n_1} \frac{\Pi_\lambda(\{M_1, \ldots, M_i\})}{\Pi_\mu(M_i, \ldots, M_1)} \pi(0).$$

After normalization, this solution becomes the stationary distribution.

Proof We only have to verify that the partial balance equations are satisfied whenever the assignment condition holds. Therefore we need to verify that the solution (22) satisfies (18) and (19).

To verify (18), we first look at one term of the sum $\mathcal{P}_M(\mathfrak{s})$, which we divide by $\lambda_M(\{M_1, \ldots, M_i\})\pi(\mathfrak{s})$. After substituting the product form solution (22) we obtain the following derivation, which we justify in the following paragraph:

$$
\mu_M p_{kl}^M(\mathfrak{s}) \frac{\pi(insert_{kl}^M(\mathfrak{s}))}{\lambda_M(\{M_1, \ldots, M_i\})\pi(s)}
$$

$$
= \frac{\mu_M}{\lambda_M(\{M_1, \ldots, M_i\})} \delta_k(M)^l \delta_{k+1}(M)^{n_{k+1}} \cdots \delta_i(M)^{n_i} \frac{\pi(0) \frac{\Pi_\lambda(\{M_1, \ldots, M_i, M\})}{\Pi_\mu(M_i, \ldots, M, M_k, \ldots, M_1)}}{\pi(0) \frac{\Pi_\lambda(\{M_1, \ldots, M_i\})}{\Pi_\mu(M_i, \ldots, M_1)}}
$$

$$
\times \frac{\alpha_1^{n_1} \cdots \alpha_k^{n_k - l} (\frac{\lambda_{\mathcal{U}}(\{M_1, \ldots, M_k, M\})}{\mu_{\{M_1, \ldots, M_k, M\}}})^l (\frac{\lambda_{\mathcal{U}}(\{M_1, \ldots, M_{k+1}, M\})}{\mu_{\{M_1, \ldots, M_{k+1}, M\}}})^{n_{k+1}} \cdots (\frac{\lambda_{\mathcal{U}}(\{M_1, \ldots, M_i, M\})}{\mu_{\{M_1, \ldots, M_i, M\}}})^{n_i}}{\alpha_1^{n_1} \cdots \alpha_i^{n_i}}
$$

$$
= \mu_M \left(\frac{\delta_k(M)}{\alpha_k} \frac{\lambda_{\mathcal{U}}(\{M_1, \ldots, M_k, M\})}{\mu_{\{M_1, \ldots, M_k, M\}}} \right)^l \left(\frac{\delta_{k+1}(M)}{\alpha_{k+1}} \frac{\lambda_{\mathcal{U}}(\{M_1, \ldots, M_{k+1}, M\})}{\mu_{\{M_1, \ldots, M_{k+1}, M\}}} \right)^{n_{k+1}} \cdots
$$

$$
\times \left(\frac{\delta_i(M)}{\alpha_i} \frac{\lambda_{\mathcal{U}}(\{M_1, \ldots, M_i, M\})}{\mu_{\{M_1, \ldots, M_i, M\}}} \right)^{n_i} \tag{24}
$$

$$
\times \frac{\mu_{\{M_1, \ldots, M_k, M_{k+1}\}} \cdots \mu_{\{M_1, \ldots, M_i\}}}{\mu_{\{M_1, \ldots, M_k, M\}} \mu_{\{M_1, \ldots, M_k, M_{k+1}, M\}} \cdots \mu_{\{M_1, \ldots, M_i, M\}}}
$$

$$
= \frac{\mu_M}{\mu_{\{M_1, \ldots, M_k, M\}}} \left(\frac{\mu_{\{M_1, \ldots, M_k\}}}{\mu_{\{M_1, \ldots, M_k, M\}}} \right)^l \left(\frac{\mu_{\{M_1, \ldots, M_{k+1}\}}}{\mu_{\{M_1, \ldots, M_{k+1}, M\}}} \right)^{n_{k+1}+1} \cdots
$$

$$
\times \left(\frac{\mu_{\{M_1, \ldots, M_i\}}}{\mu_{\{M_1, \ldots, M_i, M\}}} \right)^{n_i+1}
$$

$$
= (1 - \beta_k)(\beta_k)^l (\beta_{k+1})^{n_{k+1}+1} \cdots (\beta_i)^{n_i+1}.
$$

Here the first equality is just the substitution of (22). The second equality follows by canceling and rearranging terms, expanding $\Pi_\mu(\cdot)$ in the numerator and denominator, and using the assignment condition to see that

$$
\Pi_\lambda(\{M_1, \ldots, M_i, M\}) = \lambda_M(\{M_1, \ldots, M_i\})\Pi_\lambda(\{M_1, \ldots, M_i\}).
$$

The third equality follows from the definitions of δ and α, by further cancelations and rearrangements. For the last equality we define:

$$
\beta_j = \frac{\mu_{\{M_1, \ldots, M_j\}}}{\mu_{\{M_1, \ldots, M_j, M\}}}, \quad j = 1, \ldots, i.
$$

A similar calculation leads to:

$$
\mu_M p_{1,n_1}^M(\mathfrak{s}) \frac{\pi(\mathfrak{s}, 0, M)}{\lambda_M(\{M_1, \ldots, M_i\})\pi(s)} = \prod_{j=1}^{i} (\beta_j)^{n_j+1}. \tag{25}
$$

Equations (24), (25) can be used to simplify $\mu_M \mathcal{P}_M(\mathfrak{s})$. We get:

$$\mu_M \mathcal{P}_M(\mathfrak{s}) = \sum_{k=1}^{i} \sum_{l=0}^{n_k} p_{kl}^M(\mathfrak{s}) \pi\left(insert_{kl}^M(\mathfrak{s})\right) + p_{1n_1}^M(\mathfrak{s}) \pi(\mathfrak{s}, 0, M)$$

$$= \lambda_M\left(\{M_1, \ldots, M_i\}\right) \pi(\mathfrak{s}) \left[\prod_{h=1}^{i} (\beta_j)^{n_j+1} \right.$$

$$\left. + \sum_{k=1}^{i} \sum_{l=0}^{n_k} (1-\beta_k)(\beta_k)^l \prod_{h=k+1}^{i} (\beta_j)^{n_j+1} \right]. \tag{26}$$

We now see that the sum of all the β terms on the right-hand side is 1. We note that the β_k represent probabilities for Bernoulli trials, of which there are altogether $\sum_{k=1}^{i}(n_k + 1)$ trials, starting with $n_i + 1$ trials with success probability of $(1 - \beta_i)$, followed by $n_k + 1$ trials with success probability $(1 - \beta_k)$, for $k = i - 1, \ldots, 2, 1$. The summation of terms $\sum_{k=1}^{i} \sum_{l=0}^{n_k}$ sums up the probabilities that the first success will be on the first, the second, $\ldots$, or the last of the trials, while the first summand is the probability of no success at all. These obviously add up to 1. This verifies (18).

We repeat similarly for (19). Justification of the various steps will follow. Let $j \in \{2, \ldots, i\}$ with $\mathcal{U}(\{M_1, \ldots, M_j\}) \neq \emptyset$. For $j - 1 \geq k \geq 1$ and $0 \leq l \leq n_k$, we get:

$$\mu_{M_j} q_{kl}^{M_j}(\mathfrak{s}) \frac{\pi(swap_{kl}^{M_j}(\mathfrak{s}))}{\pi(s)}$$

$$= \mu_{M_j} \delta_k(M_j)^l \delta_{k+1}(M_j)^{n_{k+1}} \cdots \delta_{j-1}(M_j)^{n_{j-1}} \left(1 - \delta_{j-1}(M_j)\right)$$

$$\times \frac{\pi(0) \frac{\Pi_\lambda(\{M_1, \ldots, M_i\})}{\Pi_\mu(M_i, \ldots, M_{j+1}, M_{j-1}, \ldots, M_j, M_k, \ldots, M_1)}}{\pi(0) \frac{\Pi_\lambda(\{M_1, \ldots, M_i\})}{\Pi_\mu(M_i, \ldots, M_1)}}$$

$$\times \frac{\alpha_1^{n_1} \cdots \alpha_k^{n_k-l} (\frac{\lambda_{\mathcal{U}(\{M_1, \ldots, M_k, M_j\})}}{\mu_{\{M_1, \ldots, M_k, M_j\}}})^l \cdots (\frac{\lambda_{\mathcal{U}(\{M_1, \ldots, M_{j-1}, M_j\})}}{\mu_{\{M_1, \ldots, M_{j-1}, M_j\}}})^{n_{j-1}+1+n_j} \alpha_{j+1}^{n_{j+1}} \cdots \alpha_i^{n_i}}{\alpha_1^{n_1} \cdots \alpha_i^{n_i}}$$

$$= \mu_{M_j} \left(\frac{\delta_k(M_j)}{\alpha_k} \frac{\lambda_{\mathcal{U}(\{M_1, \ldots, M_k, M_j\})}}{\mu_{\{M_1, \ldots, M_k, M_j\}}} \right)^l \cdots$$

$$\times \left(\frac{\delta_{j-1}(M_j)}{\alpha_{j-1}} \frac{\lambda_{\mathcal{U}(\{M_1, \ldots, M_{j-1}, M_j\})}}{\mu_{\{M_1, \ldots, M_{j-1}, M_j\}}} \right)^{n_{j-1}} \left(1 - \delta_{j-1}(M_j)\right)$$

$$\times \frac{\mu_{\{M_1, \ldots, M_k, M_{k+1}\}} \cdots \mu_{\{M_1, \ldots, M_{j-2}, M_{j-1}\}}}{\mu_{\{M_1, \ldots, M_k, M_j\}} \cdots \mu_{\{M_1, \ldots, M_{j-2}, M_j\}}} \frac{\lambda_{\mathcal{U}(\{M_1, \ldots, M_{j-1}, M_j\})}}{\mu_{\{M_1, \ldots, M_{j-1}, M_j\}}}$$

$$= \left(\lambda_{\mathcal{U}(\{M_1, \ldots, M_{j-1}, M_j\})} - \lambda_{\mathcal{U}(\{M_1, \ldots, M_{j-1}\})}\right) \frac{\mu_{M_j}}{\mu_{\{M_1, \ldots, M_k, M_j\}}}$$

$$\times \left(\frac{\mu_{\{M_1, \ldots, M_k\}}}{\mu_{\{M_1, \ldots, M_k, M_j\}}} \right)^l \left(\frac{\mu_{\{M_1, \ldots, M_{k+1}\}}}{\mu_{\{M_1, \ldots, M_{k+1}, M_j\}}} \right)^{n_{k+1}+1} \cdots$$

$$\times \left(\frac{\mu_{\{M_1,\ldots,M_{j-1}\}}}{\mu_{\{M_1,\ldots,M_{j-1},M_j\}}} \right)^{n_{j-1}+1}$$

$$= (\lambda_{\mathcal{U}(\{M_1,\ldots,M_{j-1},M_j\})} - \lambda_{\mathcal{U}(\{M_1,\ldots,M_{j-1}\})})$$

$$\times (1 - \beta_{k,j})(\beta_{k,j})^l (\beta_{k+1,j})^{n_{k+1}+1} (\beta_{j-1,j})^{n_{j-1}+1}.$$

Here in the first equality we note that $\Pi_\lambda(\{M_1,\ldots,M_i\})$ is independent of the swap of the position of M_j by the assignment condition. The second equality is obtained by canceling common terms, and some rearranging. The very last term is there because there is one more job in the swap state. The third equality follows from the definitions of δ and α, and for the last equality we define:

$$\beta_{h,j} = \frac{\mu_{\{M_1,\ldots,M_h\}}}{\mu_{\{M_1,\ldots,M_h,M_j\}}}, \quad 1 \le h \le j-1.$$

A similar calculation leads to

$$\mu_{M_j} q_{1,n_1}^{M_j}(\mathfrak{s}) \frac{\pi(swap_{00}^{M_j}(\mathfrak{s}))}{\pi(s)} = (\lambda_{\mathcal{U}(\{M_1,\ldots,M_{j-1},M_j\})} - \lambda_{\mathcal{U}(\{M_1,\ldots,M_{j-1}\})}) \prod_{h=1}^{j-1} (\beta_{h,j})^{n_h+1}.$$

We now add up the terms for each M_j, $1 < j \le i$, to get:

$$\mu_{M_j} \mathcal{Q}_{M_j}(\mathfrak{s}) = \mu_{M_j} \sum_{k=1}^{j-1} \sum_{l=0}^{n_k} q_{kl}^{M_j}(\mathfrak{s}) \pi\left(swap_{kl}^{M_j}(\mathfrak{s})\right) + \mu_{M_j} q_{1n_1}^{M_j} \pi\left(swap_{00}^{M_j}(\mathfrak{s})\right)$$

$$= \pi(s)(\lambda_{\mathcal{U}(\{M_1,\ldots,M_{j-1},M_j\})} - \lambda_{\mathcal{U}(\{M_1,\ldots,M_{j-1}\})})$$

$$\times \left[\prod_{h=1}^{j-1} \beta_{h,j}^{n_h+1} + \sum_{k=1}^{j-1} \sum_{l=0}^{n_k} (1 - \beta_{k,j})\beta_{k,j}^l \prod_{h=k+1}^{j-1} \beta_{h,j}^{n_h+1} \right]$$

$$= \pi(s)(\lambda_{\mathcal{U}(\{M_1,\ldots,M_{j-1},M_j\})} - \lambda_{\mathcal{U}(\{M_1,\ldots,M_{j-1}\})}),$$

where the argument that the sum of products of β's equals 1 is the same as for the insert transitions.

For $j = 1$ with $\mathcal{U}(\{M_1\}) \ne \emptyset$ we get:

$$\mu_{M_1} \mathcal{Q}_{M_1}(\mathfrak{s}) = \mu_{M_1} q_{1,n_1}^{M_1}(\mathfrak{s}) \pi\left(swap_{0,n_1}^{M_1}(\mathfrak{s})\right)$$

$$= \mu_{M_1} 1 \frac{\lambda_{\mathcal{U}(\{M_1\})}}{\mu_{M_1}} \pi(s)$$

$$= \pi(s)(\lambda_{\mathcal{U}(\{M_1\})} - \lambda_{\mathcal{U}(\emptyset)}),$$

where we use $\lambda_{\mathcal{U}(\emptyset)} = 0$.

Finally, adding up over all $j \in \{1,\ldots,i\}$ with $\mathcal{U}(\{M_1,\ldots,M_j\}) \ne \emptyset$ we get:

$$\sum_{\substack{j=1 \\ \mathcal{U}(\{M_1,\ldots,M_j\}) \ne \emptyset}}^{i} \mu_{M_j} \mathcal{Q}_{M_j}(\mathfrak{s})$$

$$= \pi(s) \sum_{\substack{j=1 \\ \mathcal{U}(\{M_1,\ldots,M_j\}) \neq \emptyset}}^{i} \left(\lambda_{\mathcal{U}(\{M_1,\ldots,M_{j-1},M_j\})} - \lambda_{\mathcal{U}(\{M_1,\ldots,M_{j-1}\})} \right)$$

$$= \pi(s) \sum_{j=1}^{i} \left(\lambda_{\mathcal{U}(\{M_1,\ldots,M_{j-1},M_j\})} - \lambda_{\mathcal{U}(\{M_1,\ldots,M_{j-1}\})} \right)$$

$$= \pi(s) \left(\lambda_{\mathcal{U}(\{M_1,\ldots,M_i\})} - \lambda_{\mathcal{U}(\emptyset)} \right)$$

$$= \pi(s) \lambda_{\mathcal{U}(\{M_1,\ldots,M_i\})},$$

where the second equality is valid since, if $\mathcal{U}(\{M_1,\ldots,M_j\}) = \emptyset$, then $\lambda_{\mathcal{U}(\{M_1,\ldots,M_k\})} = 0$ for $k = 1,\ldots,j$. This verifies (19), and thus completes the proof. □

3.8 Verifying the assignment condition

In order to be able to state that a product form solution exists, we need to verify that indeed it is always possible to find assignment probability distributions which result in activation rates that satisfy the assignment condition. But this is exactly the result obtained in [2] which considers a closely related system: the current system *without waiting room*, so jobs that arrive, and do not find an idle machine which can handle them, are lost. Below we summarize the results from [2].

First of all it is shown that for given arrival rates λ_c, $c \in \mathcal{C}$, and given the job classes served by each machine, $\mathcal{C}(m_j)$, $j = 1,\ldots,K$, the activation rates are uniquely determined by the assignment condition, and can be calculated recursively from the following formulas. For M being the only idle machine,

$$\lambda_M(\mathcal{M} \backslash \{M\}) = \lambda_{\mathcal{C}(M)}, \quad M \in \mathcal{M}.$$

Now we proceed by induction on the number of idle machines, assuming we have determined for all sets $\{M_1,\ldots,M_j\}$ of size j greater than i the activation rates $\lambda_M(\{M_1,\ldots,M_j\})$ for $M \notin \{M_1,\ldots,M_j\}$. Consider then the set of machines $\{M_1,\ldots,M_i\}$ of size i. Clearly,

$$\lambda_{\mathcal{C}(\mathcal{M} \backslash \{M_1,\ldots,M_i\})} = \sum_{M \notin \{M_1,\ldots,M_i\}} \lambda_M(\{M_1,\ldots,M_i\}). \tag{27}$$

Introducing the notation

$$\overline{\prod_\lambda}(\{M_1,\ldots,M_i\}) = \prod_{j=i}^{K-1} \lambda_{M_{j+1}}(\{M_1,\ldots,M_j\}),$$

we conclude from the assignment condition that for any two machines $M_j, M_k \notin \{M_1,\ldots,M_i\}$,

$$\lambda_{M_j}(\{M_1,\ldots,M_i\}) \overline{\prod_\lambda}(\{M_1,\ldots,M_i,M_j\})$$

$$= \lambda_{M_k}\big(\{M_1, \ldots, M_i\}\big) \overline{\prod}_\lambda \big(\{M_1, \ldots, M_i, M_k\}\big). \tag{28}$$

The above equations can be solved for $\lambda_M(\{M_1, \ldots, M_i\})$ which are then determined up to a multiplicative constant, and this constant is then uniquely determined by (27). This shows that activation rates that satisfy the assignment condition exist, and are unique. To calculate them recursively we obtain from (28):

$$\frac{\lambda_{M_j}(\{M_1, \ldots, M_i\})}{\lambda_{M_k}(\{M_1, \ldots, M_i\})} = \frac{\lambda_{M_j}(\{M_1, \ldots, M_i, M_k\})}{\lambda_{M_k}(\{M_1, \ldots, M_i, M_j\})} \tag{29}$$

and substitution of (29) into (27) yields

$$\lambda_{M_j}\big(\{M_1, \ldots, M_i\}\big)$$
$$= \lambda_{\mathcal{C}(\mathcal{M}\setminus\{M_1, \ldots, M_i\})} \Big/ \Big(1 + \sum_{M_k \notin \{M_1, \ldots, M_i, M_j\}} \frac{\lambda_{M_k}(\{M_1, \ldots, M_i, M_j\})}{\lambda_{M_j}(\{M_1, \ldots, M_i, M_k\})} \Big),$$
$$i \le K - 2, \ M_j \notin \{M_1, \ldots, M_i\}.$$

It is then necessary to find assignment probability distributions which result in these activation rates. We define $P(c, M_j \mid \{M_1, \ldots, M_i\})$ to be the probability that an arriving job of type c who finds machines $\{M_1, \ldots, M_i\}$ busy will be assigned to the idle machine M_j. Here we get a distribution for each type of job and each subset of machines $\{M_1, \ldots, M_i\}$ for which there exists at least one machine M_j such that $c \in \mathcal{C}(M_j)$ and $M_j \notin \{M_1, \ldots, M_i\}$.

All that is needed is to find values $P(c, M_j \mid \{M_1, \ldots, M_i\}) \ge 0$ such that

$$\lambda_{M_j}\big(\{M_1, \ldots, M_i\}\big) = \sum_{c \in \mathcal{C}(M_j)} \lambda_c P\big(c, M_j \mid \{M_1, \ldots, M_i\}\big), \tag{30}$$

and it is shown in [2] that there always exist such values (which often are not unique).

It is shown that appropriate values can be obtained by solving a maximal flow problem (cf. [7]) for each subset of busy machines $\{M_1, \ldots, M_i\}$. The maximal flow problem is illustrated in Fig. 7. In this problem there is a source node s, with an arc of capacity λ_c to a node c for every $c \in \mathcal{C}(\mathcal{M}\setminus\{M_1, \ldots, M_i\})$, and a terminal node t with an arc of capacity $\lambda_{M_j}(\{M_1, \ldots, M_i\})$ from a node M_j to node t for every $M_j \notin \{M_1, \ldots, M_i\}$, and an arc of infinite capacity from node c to node M_j if $c \in \mathcal{C}(M_j)$, see Fig. 7. Note that the sum of capacities of arcs from the source and the sum of capacities of arcs to the terminal are both equal to $\lambda_{\mathcal{C}(\mathcal{M}\setminus\{M_1, \ldots, M_i\})}$:

$$\sum_{c \in \mathcal{C}(\mathcal{M}\setminus\{M_1, \ldots, M_i\})} \lambda_c = \sum_{M_j \notin \{M_1, \ldots, M_i\}} \lambda_{M_j}\big(\{M_1, \ldots, M_i\}\big) = \lambda_{\mathcal{C}(\mathcal{M}\setminus\{M_1, \ldots, M_i\})}.$$

Hence, if the maximal flow in this network is $\lambda_{\mathcal{C}(\mathcal{M}\setminus\{M_1, \ldots, M_i\})}$, then all the arcs from the source and all the arcs to the terminal are utilized at full capacity. Let $\nu(c, M_j)$ denote the flow on the arc with infinite capacity from c to M_j, then
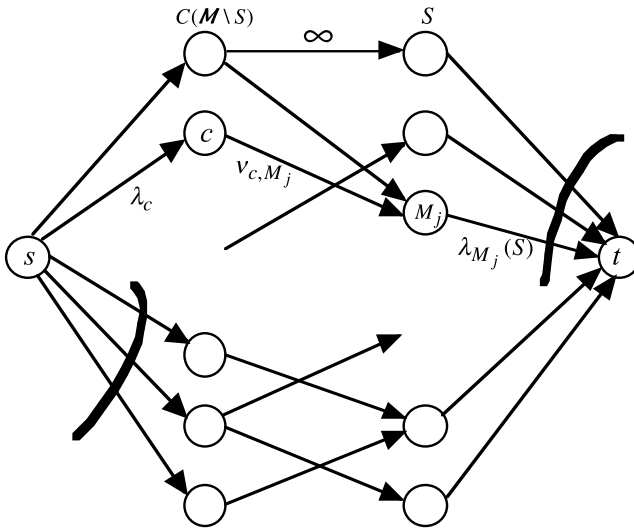
Fig. 7 Max flow problem related to the set of servers $S = \{M_1, \ldots, M_i\}$

$P(c, M_j \mid \{M_1, \ldots, M_i\}) = \nu(c, M_j)/\lambda_c$ solve (30). A sufficient condition for a flow of $\lambda_{C(\mathcal{M}\setminus\{M_1,\ldots,M_i\})}$ is the following monotonicity property of the activation rates:

$$\lambda_{M_j}\big(\{M_1, \ldots, M_i\}\big) \geq \lambda_{M_j}\big(\{M_1, \ldots, M_{i-1}\}\big), \quad i \leq K - 1, \ M_j \notin \{M_1, \ldots, M_i\}.$$

It is proved in [2] that the above monotonicity property always holds.

Below we present an example; see Example 2 in [2] for detailed calculations. It illustrates that the assignment probability distributions need not be unique.

Example There are three job types, numbered 1, 2 and 3, and three servers, also numbered 1, 2, and 3 with $C(1) = \{2, 3\}$, $C(2) = \{1, 3\}$, $C(3) = \{1, 2\}$. Let $\lambda = \lambda_1 + \lambda_2 + \lambda_3$. We get for one or two idle servers and $i \neq j \neq k$:

$$P\big(j, i \mid \{j, k\}\big) = P\big(k, i \mid \{j, k\}\big) = 1, \qquad P\big(i, j \mid \{k\}\big) = P\big(j, i \mid \{k\}\big) = 1,$$

$$P\big(k, i \mid \{k\}\big) = \frac{\lambda_i + \lambda_k}{\lambda_i + \lambda_j + 2\lambda_k}, \qquad P\big(k, j \mid \{k\}\big) = \frac{\lambda_j + \lambda_k}{\lambda_i + \lambda_j + 2\lambda_k}.$$

When all three servers are idle, the assignment probability distributions are not unique. Using the abbreviations $P(i, j) \equiv P(i, j \mid \emptyset)$, the distributions can be parameterized as:

$$\begin{bmatrix} P(i, j) \\ P(j, k) \\ P(k, i) \end{bmatrix} = \begin{bmatrix} 1 - P(i, k) \\ 1 - P(j, i) \\ 1 - P(k, j) \end{bmatrix} = (1 - \theta) \begin{bmatrix} \frac{\max(0, \eta_j - \lambda_k, \lambda_i - \eta_k)}{\lambda_i} \\ \frac{\max(0, \eta_k - \lambda_i, \lambda_j - \eta_i)}{\lambda_j} \\ \frac{\max(0, \eta_i - \lambda_j, \lambda_k - \eta_j)}{\lambda_k} \end{bmatrix}$$

$$+ \theta \begin{bmatrix} \frac{\min(\lambda_i, \eta_j, \lambda_i + \lambda_j - \eta_k)}{\lambda_i} \\ \frac{\min(\lambda_j, \eta_k, \lambda_j + \lambda_k - \eta_i)}{\lambda_j} \\ \frac{\min(\lambda_k, \eta_i, \lambda_i + \lambda_k - \eta_j)}{\lambda_k} \end{bmatrix}$$

where $0 \le \theta \le 1$ and

$$\eta_j = \frac{\lambda(\lambda^2 - \lambda_j^2)}{3\lambda^2 - \lambda_1^2 - \lambda_2^2 - \lambda_3^2}.$$

4 The waiting time distribution

In this section we use the distributional form of Little's law [10] to derive the waiting time distributions of jobs of various types.

The basic formula says that, within a queueing system, a stream of jobs that enter in a Poisson stream of rate λ stay for a while without affecting either future arrivals or the time spent in the system by previous jobs, and leave in the same order as they arrived, satisfy the following relation between steady-state number of jobs in the system N and the time spent in the system W:

$$E(z^N) = E(e^{-\lambda W(1-z)}). \tag{31}$$

Jobs of type c satisfy the above conditions necessary for the distributional form of Little's law. Thus we can use relation (31) to derive explicit expressions for the steady-state waiting time of jobs of type c in the system, from the time that they enter the system and until the time that they enter service. It turns out that this waiting time is distributed as a mixture of sums of exponentials with various parameters.

We introduce the following random variables associated with the system:

N_j is the number of jobs waiting between machines M_j and M_{j+1}, or behind the last machine if machine M_j is the last busy machine, or 0 if there are less then j busy machines, $j = 1, \ldots, K$.

$N_{c,j}$ is the number of jobs of type c among N_j. Recall that all jobs of N_j need to be of types $\mathcal{U}(\{M_1, \ldots, M_j\})$, so $N_{c,j} = 0$ if $c \notin \mathcal{U}(\{M_1, \ldots, M_j\})$.

$N_{c\cdot}$ is the total number of jobs of type c in the queue, obtained by adding up all non-zero $N_{c,j}$ for $j = 1, \ldots, K$. Note that if $c \notin \mathcal{U}(\{M_1, \ldots, M_i\})$, then one of the idle machines can serve jobs of type c, and hence $N_{c\cdot} = 0$.

We now derive $E(z^{N_{c\cdot}})$. To do so we first condition on the set of busy machines. Recall that the steady-state probability of state $\mathfrak{s} = (n_i, M_i, \ldots, n_1, M_1)$ is

$$\pi(\mathfrak{s}) = \alpha_i^{n_i} \cdots \alpha_1^{n_1} \frac{\Pi_\lambda(\{M_1, \ldots, M_i\})}{\Pi_\mu(M_i, \ldots, M_1)} \pi(0).$$

We then have that for the system in steady state,

$$P(N_K = 0, \ldots, N_{i+1} = 0, N_i = n_i, M_i, \ldots, N_1 = n_1, M_1 \mid M_1, \ldots, M_i)$$

$$= \prod_{j=1}^{i} (1 - \alpha_j) \alpha_j^{n_j}.$$

Recall also that

$$\alpha_j = \frac{\lambda_{\mathcal{U}(\{M_1,\dots,M_j\})}}{\mu_{\{M_1,\dots,M_j\}}} \quad j = 1, 2, \dots, i,$$

and that for $c \in \mathcal{U}(\{M_1, \dots, M_j\})$, the conditional distribution of $N_{c,j}$ conditional on N_j is binomial:

$$P(N_{c,j} = m | N_j = n) = \binom{n}{m} \left(\frac{\lambda_c}{\lambda_{\mathcal{U}(\{M_1,\dots,M_j\})}} \right)^m \left(1 - \frac{\lambda_c}{\lambda_{\mathcal{U}(\{M_1,\dots,M_j\})}} \right)^{n-m}.$$

For N a geometric random variable with parameter α and for M conditional on N a binomial random variable with parameters (N, θ), we have:

$$E(z^M) = \sum_{n=0}^{\infty} (1-\alpha)\alpha^n E(z^M | N = n) = \sum_{n=0}^{\infty} (1-\alpha)\alpha^n \sum_{m=0}^{n} \binom{n}{m} (z\theta)^m (1-\theta)^{n-m}$$

$$= \sum_{n=0}^{\infty} (1-\alpha)\big(\alpha(1-\theta+z\theta)\big)^n = \frac{1-\alpha}{1-\alpha(1-\theta+z\theta)} = \frac{1 - \frac{\alpha\theta}{1-\alpha(1-\theta)}}{1 - \frac{\alpha\theta}{1-\alpha(1-\theta)}z},$$

from which we can conclude that the unconditional M has a geometric distribution with parameter $\frac{\alpha\theta}{1-\alpha(1-\theta)}$.

We can now calculate the joint generating function of $N_{c,1}, \dots, N_{c,K}$, conditional on the busy machines $M_1, \dots, M_i$. We note that for $j = 1, \dots, i$, the conditional distribution of N_j is geometric with parameter α_j, and that if $c \in \mathcal{U}(\{M_1, \dots, M_j\})$, then $N_{c,j} | N_j$ is binomial with parameters N_j and $\theta_j = \frac{\lambda_c}{\lambda_{\mathcal{U}(\{M_1,\dots,M_j\})}}$. Hence $N_{c,j}$ is geometric with parameter

$$\eta_{c,j} = \frac{\alpha_j \theta_j}{1 - \alpha_j(1-\theta_j)} = \frac{\lambda_c}{\mu_{\{M_1,\dots,M_j\}} - \lambda_{\mathcal{U}(\{M_1,\dots,M_j\})} + \lambda_c}.$$

Furthermore, from the product form of the stationary distribution it is seen that conditional on the busy machines $M_1, \dots, M_i$, the $N_1, \dots, N_i$ are independent, and hence also $N_{c,1}, \dots, N_{c,i}$ are independent. We therefore obtain for the sequence of busy machines $M_1, \dots, M_i$,

$$E\big(z_1^{N_{c,1}} \cdots z_K^{N_{c,K}} | M_1, \dots, M_i\big) = \prod_{\substack{j=1 \\ c \in \mathcal{U}(\{M_1,\dots,M_j\})}}^{i} \frac{1-\eta_{c,j}}{1 - \eta_{c,j} z_j}.$$

The generating function of $N_{c\cdot} = N_{c,1} + \cdots + N_{c,K}$ conditional on $M_1, \ldots, M_i$ is obtained by substituting $z_j = z$ for all j:

$$E\big(z^{N_{c\cdot}} \mid M_1, \ldots, M_i\big) = \prod_{\substack{j=1 \\ c \in \mathcal{U}(\{M_1, \ldots, M_j\})}}^{i} \frac{1 - \eta_{c,j}}{1 - \eta_{c,j}z}.$$

To un-condition we use the steady-state probability that machines $M_1, \ldots, M_i$ are busy:

$$\pi(\cdot, M_i, \ldots, \cdot, M_1) = \frac{\Pi_\lambda(\{M_1, \ldots, M_i\})}{\Pi_\mu(M_i, \ldots, M_1)} \prod_{j=1}^{i} \frac{1}{1 - \alpha_j} \, \pi(0). \qquad (32)$$

By summing over all possible sequences $(M_i, \ldots, M_1)$, we then have that:

$$E\big(z^{N_{c\cdot}}\big) = \sum_{i=0}^{K} \sum_{(M_i, \ldots, M_1) \in \mathcal{M}^i} \pi(\cdot, M_i, \ldots, \cdot, M_1) \prod_{\substack{j=1 \\ c \in \mathcal{U}(\{M_1, \ldots, M_j\})}}^{i} \frac{1 - \eta_{c,j}}{1 - \eta_{c,j}z}.$$

We now apply the distributional form of Little's law. To obtain the Laplace–Stieltjes transform (LST) of the waiting time we need to use $E(e^{-sW}) = E((\frac{\lambda-s}{\lambda})^N)$, and in particular, if N has a geometric distribution with parameter η, then

$$E\big(e^{-sW}\big) = E\left(\left(\frac{\lambda - s}{\lambda}\right)^N\right) = \frac{1 - \eta}{1 - \eta(\frac{\lambda-s}{\lambda})} = \frac{\frac{1-\eta}{\eta}\lambda}{\frac{1-\eta}{\eta}\lambda + s},$$

which implies that W is an exponential random variable with parameter $\frac{1-\eta}{\eta}\lambda$.

Hence, for LST of the steady-state waiting time W_c of a job of type c, we obtain:

$$E(e^{-sW_c}) = \sum_{i=0}^{K} \sum_{(M_i, \ldots, M_1) \in \mathcal{M}^i} \pi(\cdot, M_i, \ldots, \cdot, M_1) \prod_{\substack{j=1 \\ c \in \mathcal{U}(\{M_1, \ldots, M_j\})}}^{i} \frac{\frac{1-\eta_{c,j}}{\eta_{c,j}}\lambda_c}{\frac{1-\eta_{c,j}}{\eta_{c,j}}\lambda_c + s}.$$

Using the definition of $\eta_{c,j}$ we get the surprising simplification, whenever $c \in \mathcal{U}(\{M_1, \ldots, M_j\})$:

$$\frac{1 - \eta_{c,j}}{\eta_{c,j}}\lambda_c = \frac{1 - \frac{\lambda_c}{\mu_{\{M_1,\ldots,M_j\}} - \lambda_{\mathcal{U}(\{M_1,\ldots,M_j\})} + \lambda_c}}{\frac{\lambda_c}{\mu_{\{M_1,\ldots,M_j\}} - \lambda_{\mathcal{U}(\{M_1,\ldots,M_j\})} + \lambda_c}}\lambda_c = \mu_{\{M_1,\ldots,M_j\}} - \lambda_{\mathcal{U}(\{M_1,\ldots,M_j\})},$$

from which λ_c has miraculously disappeared. Rewriting this we now have:

Theorem 3 *The LST of the steady-state waiting time W_c of a job of type c is equal to*

$$E\big(e^{-sW_c}\big) = \sum_{i=0}^{K} \sum_{(M_i, \ldots, M_1) \in \mathcal{M}^i} \pi(\cdot, M_i, \ldots, \cdot, M_1)$$

$$\times \prod_{\substack{j=1 \\ c \in \mathcal{U}(\{M_1, \ldots, M_j\})}}^{i} \frac{\mu_{\{M_1, \ldots, M_j\}} - \lambda_{\mathcal{U}(\{M_1, \ldots, M_j\})}}{\mu_{\{M_1, \ldots, M_j\}} - \lambda_{\mathcal{U}(\{M_1, \ldots, M_j\})} + s}, \quad (33)$$

where $\pi(\cdot, M_i, \ldots, \cdot, M_1)$ is given by (32).

This has the following interpretation: Consider the system in steady state. Jobs of type c arrive as a Poisson stream, and hence they see the queue in steady state, and find machines $(M_i, \ldots, M_1)$ busy with probability $\pi(\cdot, M_i, \ldots, \cdot, M_1)$. If some of the idle machines can process a job of type c, the arriving job will go into service immediately, and the waiting time will be 0. This is expressed in (33) by noting that in that case $c \notin \mathcal{U}(\{M_1, \ldots, M_i\})$, and so the product is empty (and thus equal to 1). Else, it will have to wait a sum of exponential waiting times.

There will be an exponential waiting time term for machine M_j if $c \in \mathcal{U}(\{M_1, \ldots, M_j\})$. This will include terms for machines $M_i, M_{i-1}, \ldots$ as long as none of them can serve c, and terminate with a term for the first machine M_k which can serve c. To be precise, $c \in \mathcal{C}(M_k)$, while $c \notin \mathcal{C}(M_j)$ for $j = k + 1, \ldots, i$.

The waiting time term for machine M_j is exponential with rate $\mu_{\{M_1, \ldots, M_j\}} - \lambda_{\mathcal{U}(\{M_1, \ldots, M_j\})}$, which is the waiting time in an M/M/1 queue with arrival rate $\lambda_{\mathcal{U}(\{M_1, \ldots, M_j\})}$ and service rate $\mu_{\{M_1, \ldots, M_j\}}$.

Hence, when a job of type c arrives, his waiting time can be interpreted as going through a tandem sequence of M/M/1 queues, until he can be served by the last of them.

References

1. Adan, I., Foley, R., McDonald, D.: Exact asymptotics of the stationary distribution of a Markov chain: a production model. Queueing Syst. **62**, 311–344 (2009)
2. Adan, I.J.B.F., Hurkens, C.A.J., Weiss, G.: A reversible multi-class multi-server loss system. Probab. Eng. Inf. Sci. **24**, 536–548 (2010)
3. Adan, I.J.B.F., Weiss, G.: Exact FCFS matching rates for two infinite multi-type sequences. Oper. Res. (2011, to appear)
4. Adan, I.J.B.F., Wessels, J., Zijm, W.H.M.: Queuing analysis in a flexible assembly system with a job-dependent parallel structure. In: Operations Research Proceedings, 1988, pp. 551–558. Springer, Berlin (1989)
5. Aerts, J., Korst, J., Verhaegh, W.: Load balancing for redundant storage strategies: Multiprocessor scheduling with machine eligibility. J. Sched. **4**, 245–257 (2007)
6. Aksin, Z., Armony, M., Mehrotra, V.: The modern call-center, a multi-disciplinary perspective on operations management research. Prod. Oper. Manag. **16**, 665–688 (2007)

7. Ford, L.R. Jr., Fulkerson, D.R.: Flows in Networks. Princeton University Press, Princeton (1962)
8. Foster, F.G.: On the stochastic matrices associated with certain queuing processes. Ann. Math. Stat. **24**, 355–360 (1953)
9. Garnett, O., Mandelbaum, A.: An introduction to skill-based routing and its operational complexities (2000). http://iew3.technion.ac.il/serveng/Lectures/SBR.pdf
10. Keilson, J., Servi, L.D.: A distributional form of Little's law. Oper. Res. Lett. **7**, 223–227 (1988)
11. Visschers, J.W.C.H.: Random walks with geometric jumps. Ph.D. Thesis, Eindhoven University of Technology (2000)
12. Visschers, J.W.C.H., Adan, I.J.B.F., Weiss, G.: A product form solution to a system with multi-type jobs and multi-type servers. Eurandom report 2011-002 (2011)