

EURANDOM PREPRINT SERIES  
2010-027

**A reversible loss system  
with multi-type customers  
and multi-type servers**

I. Adan, C. Hurkens, G. Weiss  
ISSN 1389-2355

# A reversible loss system with multi-type customers and multi-type servers

Ivo Adan\*    Cor Hurkens†    Gideon Weiss‡

June 7, 2010

## Abstract

We consider a memoryless loss system with servers  $\mathcal{S} = \{1, \dots, J\}$ , and with customer types  $\mathcal{C} = \{1, \dots, I\}$ . Servers are multi-type, so that server  $j$  can serve a subset of customer types  $C(j)$ . We show that the probabilities of assigning arriving customers to idle servers can be chosen in such a way that the Markov process describing the system is reversible, with a simple product form stationary distribution. Furthermore, the system is insensitive, these properties are preserved for general service time distributions.

*Keywords:* Service system; loss system; multi type customers; multi type servers; product form solution; reversible Markov chain, insensitivity.

## 1 Model

We consider a loss system with servers  $\mathcal{S} = \{1, \dots, J\}$ , and with customer types  $\mathcal{C} = \{1, \dots, I\}$ . Arrivals are Poisson. Customers of type  $i$  arrive at rate  $\lambda_i$ . The service requirements of all customers are i.i.d. exponentially distributed with rate 1. Servers are multi-type, so that server  $j$  can serve a subset of customer types  $C(j)$ . Server  $j$  works at rate  $\mu_j$ .

The system is a loss system: Customers that arrive, and do not find an idle server which can serve them, are lost. We define the state of the system at time  $t$  as  $X(t) = S$ , where  $S \subseteq \mathcal{S}$  is the set of idle servers which are available to receive customers at time  $t$ .

To complete the description of the system we need to specify how arriving customers are assigned to servers: An arriving customer of type  $i$  which arrives when the system is in state  $S$  will choose server  $j \in S$  (where  $i \in C(j)$ ) with probability  $P(i, j|S)$ . With this assignment  $X(t)$  is a continuous time finite state Markov chain (CTMC).

Loss systems with multi-type servers and multi-type customers are motivated by applications such as, e.g., call centers with skill based routing [6, 10], redundant data storage for video on demand [5] or bed capacity planning of hospital wards [7, 12].

---

\*Department of Mathematics and Computer Science, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, the Netherlands, and Department of Quantitative Economics, University of Amsterdam, P.O.Box 19268, 1000 GG Amsterdam, the Netherlands; email iadan@win.tue.nl Research supported in part by the Netherlands Organization for Scientific Research (NWO).

†Department of Mathematics and Computer Science, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, the Netherlands; email wscor@win.tue.nl

‡Department of Statistics, The University of Haifa, Mount Carmel 31905, Israel; email gweiss@stat.haifa.ac.il Research supported in part by Israel Science Foundation Grants 454/05 and 711/09, hospitality of the Newton Institute of Mathematical Sciences is gratefully acknowledged.

Our result in this paper is to show that one can choose  $P(i, j|S)$  in such a way that the Markov process  $X(t)$  is reversible, and as a result, one can then write down the stationary distribution of the process explicitly. This stationary distribution is unique, even though  $P(i, j|S)$  which lead to it may not be unique. It is furthermore also true that for this reversible case, the system is insensitive — the stationary distribution remains the same, and the process remains reversible, even when the processing times at each server have arbitrary distributions.

**Example 1:**

Let  $\mathcal{S} = \{1, 2\}$ ,  $\mathcal{C} = \{1, 2\}$  and  $C(1) = \{1, 2\}$ ,  $C(2) = \{1\}$ . If a type 1 customer arrives in an empty system, then this customer is sent to server 1 or to server 2 with corresponding probabilities  $P(1, 1|\{1, 2\})$ ,  $P(1, 2|\{1, 2\})$ . If we choose

$$P(1, 1|\{1, 2\}) = \frac{\lambda_1}{2\lambda_1 + \lambda_2}, \quad P(1, 2|\{1, 2\}) = \frac{\lambda_1 + \lambda_2}{2\lambda_1 + \lambda_2}, \quad (1)$$

then the stationary distribution is (as we shall show):

$$\pi(\{1\}) = \pi(\emptyset) \frac{\mu_1}{\lambda_1 + \lambda_2}, \quad \pi(\{2\}) = \pi(\emptyset) \frac{\mu_2}{\lambda_1}, \quad \pi(\{1, 2\}) = \pi(\emptyset) \frac{\mu_1 \mu_2 (2\lambda_1 + \lambda_2)}{\lambda_1 (\lambda_1 + \lambda_2)^2}, \quad (2)$$

where  $\pi(\emptyset)$  normalizes the sum to 1.

**Notation:**

To facilitate reading we will use index  $i$  for customer types, and indexes  $j, k$  for servers, and we will use  $S$  for subsets of servers,  $C$  for subsets of customer types. We will denote by  $C(S) = \bigcup_{j \in S} C(j)$  the set of customer types which can be served by at least one server in  $S$ . We will also denote by  $S(i)$  the set of servers that can serve customers of type  $i$ .

## 2 Reversibility and product form

We now show that the assumption of reversibility uniquely determines the transition rates of the CTMC, and induces a simple product form stationary distribution. The process  $X(t)$  is reversible if and only if the CTMC  $X(t)$  satisfies the detailed balance equations (see Theorem 1.2 in [11]).

We denote by  $\eta_j(S)$  the rate at which server  $j \in S$  becomes busy, when the system is in state  $S$ . Detailed balance equations for the stationary probabilities  $\pi(S)$  hold if:

$$\pi(S) \eta_j(S) = \pi(S \setminus \{j\}) \mu_j, \quad \text{for all subsets } S \text{ and } j \in S \quad (3)$$

If detailed balance (3) holds, we get for  $S = \{j_1, \dots, j_m\}$ :

$$\pi(S) = \pi(\emptyset) \frac{\mu_{j_1}}{\eta_{j_1}(\{j_1\})} \frac{\mu_{j_2}}{\eta_{j_2}(\{j_1, j_2\})} \frac{\mu_{j_3}}{\eta_{j_3}(\{j_1, j_2, j_3\})} \cdots \frac{\mu_{j_m}}{\eta_{j_m}(S)}. \quad (4)$$

This of course only makes sense if it is independent of the order in which we put the servers in  $S$ , so it has to hold equally for all permutations of  $j_1, \dots, j_m$ . In particular, for every  $S$  and  $j, k \in S$  we obtain the recursion:

$$\frac{\eta_j(S)}{\eta_k(S)} = \frac{\eta_j(S \setminus \{k\})}{\eta_k(S \setminus \{j\})}. \quad (5)$$

When the system is in state  $S$  we denote by  $\eta(S)$  the rate at which one of the idle servers will become busy. We get two expressions for  $\eta(S)$ : it is the sum of the  $\eta_j(S)$ , and it is the sum of the arrival rates of all the customer types which can be served by the servers in  $S$ :

$$\eta(S) = \sum_{j \in S} \eta_j(S) = \sum_{i \in C(S)} \lambda_i \quad (6)$$

**Proposition 1** *The equations (5), (6) uniquely determine the values of  $\eta_j(S)$  for all  $S$  and  $j \in S$ .*

*Proof.* For singletons  $S = \{j\}$ ,

$$\eta_j(\{j\}) = \sum_{i \in C(j)} \lambda_i$$

We proceed by induction, assuming we have determined the unique values for all states  $S$  of size  $m-1$ . Consider then the state  $S = \{j_1, \dots, j_m\}$ , and a server  $k \in S$ . From (5) and the induction hypothesis we obtain:

$$\frac{\eta(S)}{\eta_k(S)} = \frac{\eta_{j_1}(S) + \dots + \eta_{j_m}(S)}{\eta_k(S)} = 1 + \sum_{j \in S \setminus \{k\}} \frac{\eta_j(S \setminus \{k\})}{\eta_k(S \setminus \{j\})},$$

where  $\eta(S)$  is also known, from (6). Hence:

$$\eta_k(S) = \eta(S) / \left( 1 + \sum_{j \in S \setminus \{k\}} \frac{\eta_j(S \setminus \{k\})}{\eta_k(S \setminus \{j\})} \right). \quad (7)$$

□

**Example 1, continued:**

We calculate the  $\eta_j(S)$ , from the values of  $\lambda_1, \lambda_2$ :

$$\eta(\{1\}) = \eta_1(\{1\}) = \lambda_1 + \lambda_2, \quad \eta(\{2\}) = \eta_2(\{2\}) = \lambda_1, \quad \eta(\{1, 2\}) = \lambda_1 + \lambda_2,$$

and using the recursion step:

$$\begin{aligned} \eta_1(\{1, 2\}) &= \eta(\{1, 2\}) / \left( 1 + \frac{\eta_2(\{2\})}{\eta_1(\{1\})} \right) = (\lambda_1 + \lambda_2) / \left( 1 + \frac{\lambda_1}{\lambda_1 + \lambda_2} \right) = \frac{(\lambda_1 + \lambda_2)^2}{2\lambda_1 + \lambda_2}, \\ \eta_2(\{1, 2\}) &= \eta(\{1, 2\}) / \left( 1 + \frac{\eta_1(\{1\})}{\eta_2(\{2\})} \right) = (\lambda_1 + \lambda_2) / \left( 1 + \frac{\lambda_1 + \lambda_2}{\lambda_1} \right) = \frac{\lambda_1(\lambda_1 + \lambda_2)}{2\lambda_1 + \lambda_2}. \end{aligned}$$

The stationary probabilities (2) follow now from (4).

### 3 Assigning customers to servers

In this section we show that it is possible to choose the assigning probabilities  $P(i, j|S)$  so that the resulting  $X(t)$  will be reversible, with transition rates and stationary distribution as determined in Section 2.

Having calculated the values  $\eta_j(S)$  we now look for the assignment probabilities  $P(i, j|S)$  so that

$$\eta_j(S) = \sum_{i \in C(j)} \lambda_i P(i, j|S). \quad (8)$$

**Proposition 2** *There exist assignment probabilities  $P(i, j|S)$ , for all  $S$ ,  $j \in S$ , and  $i \in C(j)$ , which satisfy (8).*

We prove Proposition 2 in four steps. The first one is the translation to a maximal flow problem [9].

**Proposition 3** *To satisfy (8) for  $S$  we need to solve a maximal flow problem.*

*Proof.* Summing over all the servers in  $S$ ,

$$\eta(S) = \sum_{j \in S} \eta_j(S) = \sum_{j \in S} \sum_{i \in C(j)} \lambda_i P(i, j|S) = \sum_{i \in C(S)} \lambda_i.$$

We formulate a maximal flow problem with nodes  $a, b$  and nodes  $j \in S$ ,  $i \in C(S)$ , where there is an arc with infinite capacity from  $i$  to  $j$  if  $i \in C(j)$ , and there are arcs from  $a$  to  $i$  with capacity  $\lambda_i$  and arcs from  $j$  to  $b$  with capacity  $\eta_j(S)$  (see Fig 1).

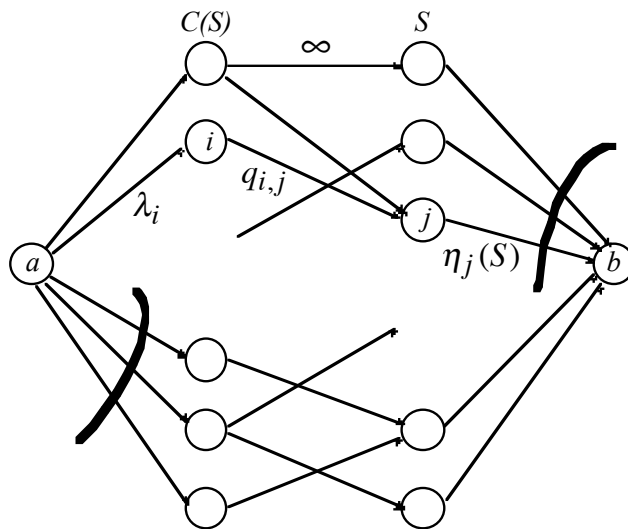


Figure 1: A maximal flow problem for finding  $P(i, j|S)$

If the maximal flow in this network is  $\eta(S)$ , and  $q_{i,j}$  is the flow on the arc from  $i$  to  $j$ , then  $P(i, j|S) = q_{i,j}/\lambda_i$  solve (8).  $\square$

**Example 1, concluded:**

We calculate the assignment probabilities by solving:

$$\eta_2(\{1, 2\}) = \lambda_1 P(1, 2|\{1, 2\}), \quad \eta_1(\{1, 2\}) = \lambda_2 + \lambda_1 P(1, 1|\{1, 2\}),$$

to obtain the values (1).

**Proposition 4** *A necessary and sufficient condition for the existence of a flow of  $\eta(S)$  in the network is: for every  $R \subseteq S$*

$$\sum_{i \in C(R)} \lambda_i \geq \sum_{j \in R} \eta_j(S). \tag{9}$$

*Proof.* See the proof of Proposition 4 in [8].  $\square$

**Proposition 5** *A sufficient condition for (9) is that  $\eta$  satisfy the following monotonicity condition: for all  $j \in R \subseteq S$*

$$\eta_j(R) \geq \eta_j(S). \quad (10)$$

*Proof.* Note that (9) actually says:

$$\sum_{j \in R} \eta_j(R) = \eta(R) = \sum_{i \in C(R)} \lambda_i \geq \sum_{j \in R} \eta_j(S).$$

which is clearly implied by (10).  $\square$

**Proposition 6** *The monotonicity condition (10) always holds.*

*Proof.* The proof is by induction on the size of  $S$ , the case  $R = S$  (in particular  $R = S = \{j\}$ ) is trivial. It is enough to verify the condition for  $R$  and  $S$  differing by only one element, say  $S = R \cup \{q\}$ . Suppose  $S$  has two or more elements and monotonicity has been established for smaller sets.

Then, for  $k \in R$  ( $k \neq q$ ), by (7),

$$\begin{aligned} \eta_k(R) &= \eta(R) / \left( 1 + \sum_{j \in R \setminus \{k\}} \frac{\eta_j(R \setminus \{k\})}{\eta_k(R \setminus \{j\})} \right), \\ \eta_k(S) &= \eta(S) / \left( 1 + \sum_{j \in R \setminus \{k\}} \frac{\eta_j(S \setminus \{k\})}{\eta_k(S \setminus \{j\})} + \frac{\eta_q(S \setminus \{k\})}{\eta_k(S \setminus \{q\})} \right). \end{aligned}$$

The monotonicity property  $\eta_k(R) \geq \eta_k(S)$  can be rewritten as

$$\eta(S) \left[ 1 + \sum_{j \in R \setminus \{k\}} \frac{\eta_j(R \setminus \{k\})}{\eta_k(R \setminus \{j\})} \right] \leq \eta(R) \left[ 1 + \sum_{j \in R \setminus \{k\}} \frac{\eta_j(S \setminus \{k\})}{\eta_k(S \setminus \{j\})} + \frac{\eta_q(S \setminus \{k\})}{\eta_k(S \setminus \{q\})} \right].$$

We can rearrange the left hand side:

$$\eta(S) \left[ 1 + \sum_{j \in R \setminus \{k\}} \frac{\eta_j(R \setminus \{k\})}{\eta_k(R \setminus \{j\})} \right] = \eta(R) \left[ 1 + \sum_{j \in R \setminus \{k\}} \frac{\eta_j(R \setminus \{k\})}{\eta_k(R \setminus \{j\})} \right] + (\eta(S) - \eta(R)) \frac{\eta(R)}{\eta_k(R)}.$$

Hence we need to verify that

$$\left[ 1 + \sum_{j \in R \setminus \{k\}} \frac{\eta_j(R \setminus \{k\})}{\eta_k(R \setminus \{j\})} \right] + \frac{(\eta(S) - \eta(R))}{\eta_k(R)} \leq \left[ 1 + \sum_{j \in R \setminus \{k\}} \frac{\eta_j(S \setminus \{k\})}{\eta_k(S \setminus \{j\})} + \frac{\eta_q(S \setminus \{k\})}{\eta_k(S \setminus \{q\})} \right]. \quad (11)$$

Note that for  $j \in R \setminus \{k\}$ :

$$\begin{aligned} \frac{\eta_j(R \setminus \{k\})}{\eta_k(R \setminus \{j\})} &= \frac{\eta_j(R \setminus \{k\}) - \eta_j(S \setminus \{k\})}{\eta_k(R \setminus \{j\})} + \frac{\eta_j(S \setminus \{k\})}{\eta_k(R \setminus \{j\})} \\ &\leq \frac{\eta_j(R \setminus \{k\}) - \eta_j(S \setminus \{k\})}{\eta_k(R)} + \frac{\eta_j(S \setminus \{k\})}{\eta_k(S \setminus \{j\})}, \end{aligned}$$

where the inequality follows by applying the induction hypothesis three times, yielding  $\eta_j(R \setminus \{k\}) - \eta_j(S \setminus \{k\}) \geq 0$ ,  $\eta_k(R \setminus \{j\}) \geq \eta_k(R)$ , and  $\eta_k(R \setminus \{j\}) \geq \eta_k(S \setminus \{j\})$ . Taking the sum over  $j \in R \setminus \{k\}$  yields

$$\begin{aligned} \sum_{j \in R \setminus \{k\}} \frac{\eta_j(R \setminus \{k\})}{\eta_k(R \setminus \{j\})} &\leq \sum_{j \in R \setminus \{k\}} \frac{\eta_j(R \setminus \{k\}) - \eta_j(S \setminus \{k\})}{\eta_k(R)} + \sum_{j \in R \setminus \{k\}} \frac{\eta_j(S \setminus \{k\})}{\eta_k(S \setminus \{j\})} \\ &= \frac{\eta(R \setminus \{k\}) - \eta(S \setminus \{k\}) + \eta_q(S \setminus \{k\})}{\eta_k(R)} + \sum_{j \in R \setminus \{k\}} \frac{\eta_j(S \setminus \{k\})}{\eta_k(S \setminus \{j\})}. \end{aligned} \quad (12)$$

Finally we note that

$$\eta(S) - \eta(R) \leq \eta(S \setminus \{k\}) - \eta(R \setminus \{k\}), \quad (13)$$

since the left hand side is the sum of arrival rates over  $C(q) \cap (C \setminus C(R))$ , while the right hand side is the sum of arrival rates of the larger or equal set  $C(q) \cap (C \setminus C(R \setminus \{k\}))$ .

Combining (12) and (13) proves (11).  $\square$

### Example 2:

There are three customer types and three servers, with  $C(1) = \{2, 3\}$ ,  $C(2) = \{1, 3\}$ ,  $C(3) = \{1, 2\}$ . Let  $\lambda = \lambda_1 + \lambda_2 + \lambda_3$ . For  $i \neq j \neq k$  (note the symmetry) we have:

$$\begin{aligned} \eta_i(\{i\}) &= \lambda_j + \lambda_k, & \eta_i(\{i, j\}) &= \frac{\lambda(\lambda_j + \lambda_k)}{\lambda_i + \lambda_j + 2\lambda_k}, \\ \eta_i(\{i, j, k\}) &= \frac{\lambda(\lambda^2 - \lambda_i^2)}{3\lambda^2 - \lambda_1^2 - \lambda_2^2 - \lambda_3^2}, \end{aligned}$$

and hence:

$$\begin{aligned} \pi(\{i\}) &= \pi(\emptyset) \frac{\mu_i}{\lambda_j + \lambda_k}, & \pi(\{i, j\}) &= \pi(\emptyset) \frac{\mu_i \mu_j (\lambda_i + \lambda_j + 2\lambda_k)}{\lambda(\lambda_i + \lambda_k)(\lambda_j + \lambda_k)}, \\ \pi(\{i, j, k\}) &= \pi(\emptyset) \frac{\mu_i \mu_j \mu_k (3\lambda^2 - \lambda_1^2 - \lambda_2^2 - \lambda_3^2)}{\lambda^2 (\lambda_i + \lambda_j)(\lambda_i + \lambda_k)(\lambda_j + \lambda_k)}. \end{aligned}$$

We now look for the assignment probabilities. We get immediately for  $S$  of one or two servers:

$$\begin{aligned} P(j, i|\{i\}) &= P(k, i|\{i\}) = 1, & P(i, j|\{i, j\}) &= P(j, i|\{i, j\}) = 1, \\ P(k, i|\{i, j\}) &= \frac{\lambda_i + \lambda_k}{\lambda_i + \lambda_j + 2\lambda_k}, & P(k, j|\{i, j\}) &= \frac{\lambda_j + \lambda_k}{\lambda_i + \lambda_j + 2\lambda_k}. \end{aligned}$$

When all three servers are idle, the equations to be solved are the three equations of the form:

$$\lambda_j P(j, i|\{1, 2, 3\}) + \lambda_k P(k, i|\{1, 2, 3\}) = \frac{\lambda(\lambda^2 - \lambda_i^2)}{3\lambda^2 - \lambda_1^2 - \lambda_2^2 - \lambda_3^2} \quad (14)$$

As we have seen, these equations do have positive solutions, but here there are three unknowns and only two equations (the three equations are dependent), so the solution is not unique. Using the abbreviations  $P(i, j) \equiv P(i, j|\{1, 2, 3\})$  and  $\eta_j \equiv \eta_j(\{1, 2, 3\})$ , the solutions to (14) 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  $0 \leq \theta \leq 1$ .

Example 2 illustrates two important points. First, the assignment probabilities need not be unique. Second, one can ask: Is it true that  $P(i, j|S_1) = P(i, j|S_2)$  if  $S(i) \cap S_1 = S(i) \cap S_2$ ? In other words, given the set of idle servers which can serve  $i$ , the  $P(i, j|\cdot)$  do not depend on additional available servers which cannot serve  $i$ . This is false, as Example 2 shows: If we take  $P(i, k|\{1, 2, 3\}) = P(i, k|\{j, k\})$ ,  $P(j, k|\{1, 2, 3\}) = P(j, k|\{i, k\})$  and  $P(k, i|\{1, 2, 3\}) = P(k, i|\{i, j\})$ , this choice will not satisfy the equations (14). So, if we want to have a product form solution, the routing rates have to change every time the state changes, even if the routing options for some of the customer types do not change. This shows how fragile the phenomena of product form is.

## 4 Insensitivity

In this section we show that, like the Erlang loss system, our reversible multi-type system is insensitive, in that the stationary distribution depends on the service time distributions only through their means. Furthermore, we also show that for arbitrary service time distributions with finite means this system remains reversible, and at stationarity all busy servers have attained service and remaining service which are distributed according to the equilibrium distribution of the processing time. We show insensitivity by the supplementary variable method, and our proof closely follows the 1957 proof of Sevastyanov [13], for the Erlang loss system.

We now assume that service times of server  $j$  are i.i.d. with distribution  $F_j$  with  $F_j(0) = 0$ , and finite mean  $1/\mu_j$ . We supplement the description of the state of the system at time  $t$  by specifying the attained service times of the busy servers. We let  $Z(t)$  be the supplemented process, with  $Z_j(t) = z_j = *$  if server  $j$  is idle at time  $t$ , and  $Z_j(t) = z_j = x_j \geq 0$  if server  $j$  is serving a customer, and the attained service time of that customer is  $x_j$ . For state  $z$  we let  $S(z)$  be the subset of idle servers, i.e. the set of coordinates  $j$  with  $z_j = *$ . We will denote by  $P_t(z)$  the distribution of  $Z(t)$ ,

$$P_t(z) = \mathbb{P}(Z_j(t) = *, j \in S(z), Z_j(t) \leq x_j, j \notin S(z))$$

and by  $p_t(z)$  its density (which is shown in the proof to exist). We will denote by  $P(z), p(z)$  the stationary distribution and density.

**Proposition 7** *The process  $Z(t)$  is ergodic with stationary probability density given by:*

$$p(z) = \pi(S(z)) \prod_{j \notin S(z)} \mu_j (1 - F_j(x_j)) \quad (15)$$

with  $\pi(S)$  given in (4).

*Proof.* Let  $P_t(z)$  be the distribution of  $Z(t)$ , with initial distribution  $P_0$ . It follows exactly as in Theorem 2 of [13] that for arbitrary  $P_0$  and for any state  $z$ ,  $P_t$  has a density at the coordinates  $x_j, j \notin S(z)$ , if  $t > \max\{x_j : j \notin S(z)\}$ .

The process  $Z(t)$  is a Markov process with transitions for large  $t$  and small  $\Delta$  given by:

$$p_{t+\Delta} \left\{ z : S(z) = S, z_j = x_j, j \notin S \right\} = p_t \left\{ z : S(z) = S, z_j = x_j - \Delta, j \notin S \right\} (1 - \eta(S)\Delta) \prod_{j \notin S} \frac{1 - F_j(x_j)}{1 - F_j(x_j - \Delta)}$$



$$+ \sum_{k \in S(z)} \int_0^\infty p_t \left\{ z : S(z) = S \setminus k, s_k = y, s_j = x_j - \Delta, j \notin S \right\} \\ \prod_{j \notin S} \frac{1 - F_j(x_j)}{1 - F_j(x_j - \Delta)} \frac{F_k(y + \Delta) - F_k(y)}{1 - F_k(y)} dy + o(\Delta)$$

and for any  $k \notin S$ ,

$$p_{t+\Delta} \left\{ z : S(z) = S, z_j = x_j, j \notin S \cup k, x_k = 0 \right\} \Delta = \\ p_t \left\{ z : S(z) = S \cup k, z_j = x_j - \Delta, j \notin S \right\} \prod_{j \notin S \cup k} \frac{1 - F_j(x_j)}{1 - F_j(x_j - \Delta)} \eta_k(S \cup k) \Delta + o(\Delta).$$

Define now

$$p_t^*(z) = p_t(z) / \prod_{j \notin S(z)} (1 - F_j(x_j)),$$

to obtain:

$$p_{t+\Delta}^* \left\{ z : S(z) = S, z_j = x_j, j \notin S \right\} = p_t^* \left\{ z : S(z) = S, z_j = x_j - \Delta, j \notin S \right\} (1 - \eta(S) \Delta) \\ + \sum_{k \in S(z)} \int_0^\infty p_t^* \left\{ z : S(z) = S \setminus k, s_k = y, s_j = x_j - \Delta, j \notin S \right\} (F_k(y + \Delta) - F_k(y)) dy + o(\Delta)$$

and for any  $k \notin S$ ,

$$p_{t+\Delta}^* \left\{ z : S(z) = S, z_j = x_j, j \notin S \cup k, x_k = 0 \right\} \Delta = \\ p_t^* \left\{ z : S(z) = S \cup k, z_j = x_j - \Delta, j \notin S \cup k \right\} \eta_k(S \cup k) \Delta + o(\Delta).$$

From these equations (and assuming that  $p_t^*(z)$  is differentiable) we get a set of integro-differential equations:

$$\frac{\partial p_t^*(z)}{\partial t} + \sum_{j \notin S(z)} \frac{\partial p_t^*(z)}{\partial x_j} = -\eta(S) p_t^*(z) + \sum_{k \in S(z)} \int_0^\infty p_t^*(z : z_k = y) dF_k(y)$$

with boundary conditions:

$$p_t^*(z : z_k = 0) = p_t^*(z : s_k = *) \eta_k(S \cup k), \quad k \notin S.$$

In stationarity the derivatives with respect to  $t$  cancel, so that we have:

$$\sum_{j \notin S(z)} \frac{\partial p^*(z)}{\partial x_j} = -\eta(S) p^*(z) + \sum_{k \in S(z)} \int_0^\infty p^*(z : z_k = y) dF_k(y)$$

with boundary conditions:

$$p^*(z : z_k = 0) = p^*(z : s_k = *) \eta_k(S \cup k), \quad k \notin S.$$

We now put in the trial solution

$$p^*(z) = \pi(S) \prod_{j \notin S(z)} \mu_j.$$

Note that  $x_j$  do not appear in this trial solution. We obtain in the second equation, for any  $S$  and  $k \notin S$ :

$$\pi(S)\mu_k \prod_{j \notin S \cup k} \mu_j = \pi(S \cup k)\eta_k(S \cup k) \prod_{j \notin S \cup k} \mu_j,$$

which is exactly the detailed balance equation (3) satisfied by  $\pi$  for each  $S$  and  $k \notin S$ , and in the first equation we get:

$$\pi(S) \prod_{j \notin S} \mu_j \eta(S) = \sum_{k \in S} \pi(S \setminus k) \mu_k \prod_{j \notin S} \mu_j,$$

which is also, according to (3), satisfied by  $\pi$  for any  $S$ .

This confirms that (15) is a stationary density for the Markov process  $Z(t)$ . It can now be shown exactly as in [13] that  $Z(t)$  is ergodic with a unique stationary density.  $\square$

We now consider a different way to supplement our process  $X(t)$ . We specify at time  $t$  the set of idle machines, supplemented by the remaining processing time on the busy machines (rather than the attained service time). We let  $Y(t)$  be the supplemented process, with  $Y_j(t) = y_j = *$  if server  $j$  is idle, and  $Y_j(t) = y_j = x_j \geq 0$  if server  $j$  is serving a customer at time  $t$  with remaining service time  $x_j$ .

**Proposition 8** *The process  $Y(t)$  is ergodic with the same stationary probability density as  $Z(t)$ . Furthermore, if we consider the stationary versions of  $Z(t)$  and  $Y(t)$  then  $Z(t)$  is equal in distribution (for the whole process) to the reversed process  $Y(-t)$ .*

*Proof.* Both  $Z(\cdot)$  and  $Y(\cdot)$  are Markov processes. They both move on the same state space. Denote by  $p_Y(z, t, z')$ ,  $p_Z(z, t, z')$  the transition kernels of the two processes, from state  $z$  to  $z'$  in time  $t$ . Let  $P$  be the stationary distribution of  $Z(\cdot)$ , with density  $p$ , as given in Proposition 7. Assume that  $Y(t - \Delta)$  and  $Z(t)$  are both distributed like  $P$ . We will show that for small  $\Delta$ , the joint probability densities of  $(Y(t), Y(t - \Delta))$  and  $(Z(t), Z(t + \Delta))$  differ only by a term of order  $o(\Delta)$ . This will show that  $P$  is also the stationary probability distribution of  $Y(\cdot)$ , and that the forward stationary Markov process  $Z(t)$  has the same transition kernel as the reversed Markov process  $Y(-t)$ , and thus prove the theorem.

Consider a fixed general state  $z$  with a set  $S$  of idle servers, and values  $x_j \geq 0, j \notin S$ ; we will for convenience denote it by  $z = (S, x_j, j \notin S)$ . We will calculate the joint probability density of  $(Z(t), Z(t + \Delta)) = (z, z')$ , and of  $(Y(t), Y(t - \Delta)) = (z, z')$ , or equivalently  $(Y(t - \Delta), Y(t)) = (z', z)$ , for all  $z'$  and small  $\Delta$ . We wish to show that the order 1 and order  $\Delta$  terms of both densities are the same. Excluding events of probability  $o(\Delta)$  the states  $z'$  that we need to be consider are:

$$z' = (S, x_j + \Delta, j \notin S), \quad z' = (S \cup k, x_j + \Delta, j \notin S \cup k), \quad k \notin S, \quad z' = (S \setminus k, x_j + \Delta, j \notin S, x_k = 0), \quad k \in S.$$

We now perform the six probability calculations to prove the required equalities. For  $z' = (S, x_j + \Delta, j \notin S)$  we get:

$$\begin{aligned} p(z)p_Z(z, t, z') &= \pi(S) \prod_{j \notin S} \mu_j (1 - F_j(x_j)) \cdot (1 - \eta(S)\Delta) \prod_{j \notin S} \frac{1 - F_j(x_j + \Delta)}{1 - F_j(x_j)} + o(\Delta), \\ p(z')p_Y(z', t, z) &= \pi(S) \prod_{j \notin S} \mu_j (1 - F_j(x_j + \Delta)) \cdot (1 - \eta(S)\Delta) + o(\Delta), \end{aligned}$$

which are obviously equal up to order  $\Delta$  terms.

For  $k \notin S$  and  $z' = (S \cup k, x_j + \Delta, j \notin S \cup k)$  we get:

$$p(z)p_Z(z, t, z') = \pi(S) \prod_{j \notin S} \mu_j (1 - F_j(x_j)) \cdot (1 - \eta(S)\Delta) \prod_{j \notin S \cup k} \frac{1 - F_j(x_j + \Delta)}{1 - F_j(x_j)} \frac{F_k(x_k + \Delta) - F_k(x_k)}{1 - F_k(x_k)} + o(\Delta),$$

$$p(z')p_Y(z', t, z) = \pi(S \cup k) \prod_{j \notin S \cup k} \mu_j (1 - F_j(x_j + \Delta)) \cdot \eta_k(S) (F_k(x_k + \Delta) - F_k(x_k)) + o(\Delta).$$

After discarding the order  $\Delta^2$  term in the first expression and canceling common terms, we get that these expressions are equal (up to order  $\Delta$  terms) by the detailed balance of  $\pi$ , for  $k \notin S$ :

$$\pi(S)\mu_k = \pi(S \cup k)\eta_k(S).$$

Finally, for  $k \in S$  and  $z' = (S \setminus k, x_j + \Delta, j \notin S, x_k = 0)$  we get:

$$p(z)p_Z(z, t, z') = \pi(S) \prod_{j \notin S} \mu_j (1 - F_j(x_j)) \cdot \prod_{j \notin S} \frac{1 - F_j(x_j + \Delta)}{1 - F_j(x_j)} \eta_k(S)\Delta + o(\Delta),$$

$$p(z')p_Y(z', t, z) = \pi(S \setminus k) \prod_{j \notin S} \mu_j (1 - F_j(x_j + \Delta)) \mu_k (1 - F_k(0))\Delta \cdot (1 - \eta(S \setminus k)\Delta) + o(\Delta),$$

and again, after discarding the order  $\Delta^2$  term in the second expression and canceling common terms and using that  $F_k(0) = 0$ , we get that these are equal (up to order  $\Delta$  terms) by the detailed balance of  $\pi$ , for  $k \in S$ :

$$\pi(S \setminus k)\mu_k = \pi(S)\eta_k(S).$$

This completes the proof.  $\square$

## 5 Discussion

In this paper we considered a loss system. It is interesting to also investigate the same system with no losses: this is a single queueing station, with multi-type customers queueing in  $I$  different queues, and  $J$  servers, which are heterogeneous, server  $j$  serving the queues of customers of types  $C(j)$ , at rate  $\mu_j$ .

In that case one needs to specify the service policy. A very common service policy is FCFS: whenever a server becomes available he will serve the longest waiting customer which is compatible with him, or else he will idle. One needs also to specify assignment rules for customers which arrive and find suitable servers which are idle. It again turns out that the assignment probabilities can be chosen so that this FCFS system will satisfy partial balance, and have a product form stationary distribution. This topic was explored in [2, 3, 17, 4], and finally resolved in [15, 16]. Remarkably, it turns out that the assignment probabilities are exactly those derived here for the reversible loss system.

This model is also related to the overloaded system with abandonments discussed in [14], and to the model of FCFS matching of infinite sequences of customers and servers proposed in [8]. Recently [1] showed that the model of FCFS matching of infinite sequences has a product form solution which is similar to that of [15, 16], for which no assignment condition is necessary.

### Acknowledgement

The authors would like to thank Scott Provan for many stimulating discussions on this problem. The authors would like to thank Frank Kelly and Peter Whittle for discussions of insensitivity.

## References

- [1] Adan, I.J.B.F., Weiss, G. (2010) Exact FCFS matching rates for two infinite multi-type sequences. Submitted for publication.
- [2] Adan, I.J.B.F., Wessels, J., Zijm, W.H.M. (1989) Queuing analysis in a flexible assembly system with a job-dependent parallel structure. *Operations Research Proceedings 1988* Springer-Verlag, Berlin, pp. 551-558.
- [3] Adan, I.J.B.F., Wessels, J., Zijm, W.H.M. (1991) Flexible assembly and shortest queue problems. *Modern production concepts, theory and applications*, G. Fandel, G. Zaepfel (eds.), Springer-Verlag, Berlin, pp. 644-659.
- [4] Adan, I. , Foley, R. and McDonald, D. (2009) Exact Asymptotics of the Stationary Distribution of a Markov Chain: a Production Model. *Queueing Systems* 62: 311-344.
- [5] Aerts, J., Korst, J., Verhaegh, W., (2007) Load balancing for redundant storage strategies: Multiprocessor scheduling with machine eligibility. *Journal of Scheduling* 4: 245-257.
- [6] Aksin, Z., Armony, M., Mehrotra, V., (2007) The modern call-center, a multi-disciplinary perspective on operations management research. *Production and Operations Management* 16 (6), forthcoming.
- [7] de Bruin, A.M., Bekker, R., van Zanten, L., Koole, G.M. (2008) Dimensioning hospital wards using the Erlang loss model. *Annals of Operations Research*, to appear.
- [8] Caldentey, R., Kaplan, E.H., Weiss, G., (2009) FCFS infinite bipartite matching of servers and customers. *Advances in Applied Probability* 41:695-730.
- [9] Ford, L. R., Jr., Fulkerson, D. R. (1962) *Flows in Networks*, Princeton University Press, Princeton.
- [10] Garnett, O. and Mandelbaum, A. (2000) An Introduction to Skill-Based Routing and its Operational Complexities. <http://iew3.technion.ac.il/serveng/Lectures/SBR.pdf>
- [11] Kelly, F.P. (1979) *Reversibility and Stochastic Networks*, Wiley.
- [12] McManus, M.L., Long, M.C., Copper, A., Litvak, E. (2004) Queuing theory accurately models the need for critical care resources. *Anesthesiology* 100: 1271-1276.
- [13] Sevastyanov, B.A. (1957) An ergodic theorem for Markov processes and its application to telephone systems with refusals *Theory of Probability and its Applications* 2:104-112.
- [14] Talreja, R. and Whitt, W. (2007) Fluid Models for Overloaded Multi-class many-service queueing systems with FCFS routing. *Management Science* 54:1513-1527.
- [15] Visschers, J.W.C.H.. (2000) *Random walks with geometric jumps*. Ph.D. Thesis, Eindhoven University of Technology.
- [16] Visschers, J.W.C.H., Adan, I.J.B.F., Weiss, G.(2009). A multi-class multi-server system with customer-class service restrictions. In preparation, 2009.
- [17] Zijm, W.H.M., Laarhoven, P.J.M. (1993) Production Preparation and Numerical Control in PCB assembly *International Journal of Flexible Manufacturing Systems* 5(3):187-207.