# Repair systems with exchangeable items and the longest queue mechanism

Queueing Syst (2013) 73:295–316
DOI 10.1007/s11134-012-9319-5

# Repair systems with exchangeable items and the longest queue mechanism

**R. Ravid · O.J. Boxma · D. Perry**

Received: 25 August 2011 / Revised: 1 May 2012 / Published online: 27 June 2012


**Abstract** We consider a repair facility consisting of one repairman and two arrival streams of failed items, from bases 1 and 2. The arrival processes are independent Poisson processes, and the repair times are independent and identically exponentially distributed. The item types are exchangeable, and a failed item from base 1 could just as well be returned to base 2, and vice versa. The rule according to which backorders are satisfied by repaired items is the *longest queue* rule: At the completion of a service (repair), the repaired item is delivered to the base that has the largest number of failed items.

We point out a direct relation between our model and the classical longer queue model. We obtain simple expressions for several probabilities of interest, and show how all two-dimensional queue length probabilities may be obtained. Finally, we derive the sojourn time distributions.

**Keywords** Repair system · Longest queue · Queue lengths · Sojourn time

**Mathematics Subject Classification** 60K25 · 90B22

The research of O.J. Boxma was partly conducted during a visit to the University of Haifa.

R. Ravid · D. Perry
Department of Statistics, University of Haifa, Haifa 31909, Israel

R. Ravid
College of Engineering, ORT Braude, Karmiel 20101, Israel

O.J. Boxma
EURANDOM, Eindhoven, The Netherlands

O.J. Boxma (✉)
Department of Mathematics and Computer Science, Technische Universiteit Eindhoven, Eindhoven, The Netherlands
e-mail: boxma@win.tue.nl

## 1 Introduction

In this paper, we consider a repair facility consisting of one repairman and two arrival streams of failed items, from bases 1 and 2. The arrival processes are independent Poisson processes with rates $\lambda_1$ and $\lambda_2$. The repair times are independent and identically distributed, with $\exp(\mu)$ distribution regardless of the type of failed item. The item types are exchangeable, and a failed item from base 1 could just as well be returned to base 2, and vice versa. The rule according to which backorders are satisfied by repaired items is the *longest queue* rule: At the completion of a service (repair), the repaired item is delivered to the base that has the largest number of failed items. In case of a tie, the item will be delivered to base 1 or base 2 with probability $\frac{1}{2}$.

We are interested in key performance measures of this repair facility, such as the (joint) queue length distribution of failed items of both types, and their sojourn time distribution (the time between arrival and departure of a failed item). In the literature, several studies have appeared about the so-called longest queue system. That is a queueing system with one server and (typically) two queues with customers of two different types; the server choosing a customer from the longest queue upon service completion. Cohen [2] has studied the case of two customer types with Poisson arrival streams with rates $\lambda_1$ and $\lambda_2$, having service time distributions $B_1(\cdot)$ and $B_2(\cdot)$. If the server has completed a service, then the next customer to be served is the one at the head of the longest queue if the queue lengths are not equal; if both queues have equal length, then the next customer in service is of type $i$ with some probability $\alpha_i$. He determines the generating function of the joint steady-state queue length distribution right after service completions, by solving a boundary value problem of Riemann–Hilbert type.

Zheng and Zipkin [9] consider the completely symmetric exponential case ($\lambda_1 = \lambda_2$; $\alpha_1 = \alpha_2 = \frac{1}{2}$; $B_1(x) = B_2(x) = 1 - \mathrm{e}^{-\mu x}$). They calculate the steady-state distribution of the difference between the two queue lengths, and they provide a recursive scheme for the calculation of the joint queue length distribution and the marginal distributions. They also briefly consider the case that $\lambda_1 \neq \lambda_2$. Flatto [5] also considers the symmetric exponential case. He allows preemption, and derives an expression for the probability generating function of the joint queue length distribution. He uses this expression to derive asymptotic results.

Van Houtum et al. [7] also focus on the completely symmetric exponential model. They consider two variants: a longest queue system with threshold rejection of customers and one with threshold addition of customers. They show that these systems can be analyzed in detail using matrix-geometric methods, and that this provides lower and upper bounds for the longest queue system.

The repair facility with exchangeable items, that is the subject of our paper, has already been studied by Daryanani and Miller [4]. Using taboo sets and taboo probabilities, they derive various relations between the steady-state queue length probabilities; however, they do not solve those equations.

*Remark 1* While the above described classical longer queue model is closely related to the model studied in [4] and the present paper, there are significant differences. To demonstrate these, let us consider the classical system and our repair system, receiving exactly the same input and having exactly the same service times. Suppose both systems start empty, and then a type-1 item arrives. The server starts serving. During the service, there are no type-1 arrivals and three type-2 arrivals. In the classical system, the type-1 customer leaves and the server starts serving the first of the three type-2 customers. The state of the classical system now is (0, 3): one type-2 customer is just entering service and the other two type-2 customers are waiting.

In our repair system, the items are exchangeable. The repaired item is assigned to the first type-2 customer, even though it was brought as a type-1 item. Hence, the state of the system right after the repair is $(1, 2)$: there is still one waiting type-1 customer and there are two waiting type-2 customers.

*Motivation* The longer queue model is related to the join-the-shortest-queue model: both models feature a mechanism that tends to equate the queue lengths. The longer queue model is a very natural one, but it has received much less attention than the join-the-shortest-queue model. We believe our paper yields valuable new insight into the longer queue model and a variant of it.

*Contributions* Our main contributions are: (i) We point out a direct relation between our model and the classical longer queue model of Cohen [2]; (ii) we obtain simple expressions for several probabilities of interest, and we show how all two-dimensional probabilities may be obtained; (iii) we derive the sojourn time distributions—this performance measure was not studied in the papers mentioned above; and (iv) we present some methodological ideas which might be more broadly applicable; one example is the use of the "difference busy period."

*Organization of the paper* In Sect. 2, we first give the balance equations for the joint steady-state queue length distribution. We then study its generating function (GF), deriving various special results like the distribution of the difference of the two queue lengths and the probability that there are $n_1$ customers of one type and none of the other type. Then we point out a direct relation between Cohen's model and our model, which in principle allows us to use his results for obtaining the GF of the joint queue length distribution. However, we are interested in providing explicit results for the probabilities $P(i,i)$ of having $i$ customers of either type, $i = 1, 2, \dots,$ which will also give us the marginal distributions explicitly. These probabilities are studied in Sect. 3; in that section, we also give an iterative method for obtaining all queue length probabilities. Finally, Sect. 4 is devoted to the determination of the sojourn time distribution of a customer of either type.

## 2 Queue lengths—a generating function approach

Let $N_i(t)$ denote the number of failed items of type $i = 1, 2$ at time $t$. Clearly, $\{(N_1(t), N_2(t)), t \geq 0\}$ is a Markov process. We restrict ourselves to the case that the total load $\rho := \frac{\lambda_1+\lambda_2}{\mu} < 1$. Then the limiting distribution $P(n_1, n_2) = P(N_1 = n_1, N_2 = n_2) := \lim_{t\to\infty} P(N_1(t) = n_1, N_2(t) = n_2 | N_1(0) = k_1, N_2(0) = k_2)$ exists and is independent of the initial state; indeed, notice that the total number of customers has the same distribution as the number of customers in an $M/M/1$ queue with arrival rate $\lambda_1+\lambda_2$ and service rate $\mu$. The $P(n_1,n_2)$ satisfy the following balance equations: for $n_1,n_2\geq 1$, with $I(\cdot)$ denoting an indicator function,

$$\begin{aligned}(\lambda_1+\lambda_2+\mu)P(n_1,n_2)\\ &=\lambda_1 P(n_1-1,n_2)+\lambda_2 P(n_1,n_2-1)+\mu P(n_1+1,n_2)I(n_1\geq n_2)\\ &\quad+\mu P(n_1,n_2+1)I(n_2\geq n_1)+\frac{\mu}{2}P(n_1,n_2+1)I(n_1=n_2+1)\\ &\quad+\frac{\mu}{2}P(n_1+1,n_2)I(n_2=n_1+1).\end{aligned}\tag{1}$$

For $n_1=0$ and/or $n_2=0$, the same equations hold with minor adaptations. Multiplying these equations with $z_1^{n_1}z_2^{n_2}$ and summing, one gets:

$$\begin{aligned}&\big(\lambda_1(1-z_1)+\lambda_2(1-z_2)+\mu\big)E\big[z_1^{N_1}z_2^{N_2}\big]\\ &\quad=P(0,0)\left(\frac{\mu}{2}\left(1-\frac{1}{z_1}\right)+\frac{\mu}{2}\left(1-\frac{1}{z_2}\right)\right)\\ &\qquad+\frac{\mu}{z_1}E\big[z_1^{N_1}z_2^{N_2}I(N_1>N_2)\big]+\frac{\mu}{z_2}E\big[z_1^{N_1}z_2^{N_2}I(N_2>N_1)\big]\\ &\qquad+\left(\frac{\mu}{2z_1}+\frac{\mu}{2z_2}\right)E\big[z_1^{N_1}z_2^{N_2}I(N_1=N_2)\big].\end{aligned}\tag{2}$$

We delay the solution of this equation, first showing how one can derive various special probabilities from this equation.

### 2.1 $P(N_1+N_2=n)$

Taking $z_1=z_2=z$ in (2), it easily follows that

$$\left((\lambda_1+\lambda_2)(1-z)+\mu\left(1-\frac{1}{z}\right)\right)E\big[z^{N_1+N_2}\big]=P(0,0)\left(\mu\left(1-\frac{1}{z}\right)\right),\tag{3}$$

so

$$E\big[z^{N_1+N_2}\big]=\frac{P(0,0)}{1-\rho z}.\tag{4}$$

This implies (by taking $z=1$) that $P(0,0)=1-\rho$, and that the total number of customers is geom($\rho$) distributed. This confirms that the total number of customers behaves as if the system is an $M/M/1$ queue with arrival rate $\lambda_1+\lambda_2$ and service rate $\mu$.

### 2.2 The probabilities $P(n_1,0)$ and $P(0,n_2)$

Taking $z_1=z$ and $z_2=0$ in (2), we obtain by carefully considering the $1/z$ terms:

$$\begin{aligned}&\left(\lambda_1(1-z)+\lambda_2+\mu\left(1-\frac{1}{z}\right)\right)E\big[z^{N_1}I(N_1>0,N_2=0)\big]\\ &\quad=\mu P(0,1)-\big(\lambda_1(1-z)+\lambda_2\big)P(0,0)+\frac{\mu}{2}zP(1,1).\end{aligned}\tag{5}$$

Hence,

$$E\big[z^{N_1} I(N_1 > 0, N_2 = 0)\big] = -z \frac{\mu P(0,1) - (\lambda_1(1-z) + \lambda_2)P(0,0) + \frac{\mu}{2} z P(1,1)}{\lambda_1 z^2 - (\lambda_1 + \lambda_2 + \mu)z + \mu}. \tag{6}$$

Consider the denominator of the right-hand side of (6). Partial fraction yields

$$E\big[z^{N_1} I(N_1 > 0, N_2 = 0)\big] = \frac{C_1 z}{1 - z/z_+} + \frac{C_2 z}{1 - z/z_-}, \tag{7}$$

where $C_1$ and $C_2$ remain to be determined. From $M/M/1$ theory (cf. Cohen [1], Chap. II.4), it follows that the zero $z_-$ with "minus the square root" of the denominator of (6) is the Laplace–Stieltjes transform (LST) with argument $\lambda_2$ of the length of the busy period $P_1$ in an $M/M/1$ queue with arrival rate $\lambda_1$ and service rate $\mu$: $z_- = E[\mathrm{e}^{-\lambda_2 P_1}]$. It may also be interpreted as the probability of zero arrivals from base 2 during a busy period of type 1. Since this zero $z_-$ has absolute value less than one, and the left-hand side of (6) is analytic inside the unit circle, $C_2$ must be zero. The product of $z_+$ and $z_-$ is $\mu/\lambda_1 > 1$, so $z_+$ has absolute value larger than one. Hence, we conclude that the probabilities $P(N_1 = n_1, N_2 = 0)$, for $n_1 > 0$, are geometric with parameter $\frac{1}{z_+} = \frac{\lambda_1 z_-}{\mu}$:

$$P(j,0) = C_1 \left(\frac{1}{z_+}\right)^{j-1}, \quad j = 1, 2, \ldots. \tag{8}$$

Similarly, $P(0,j) = \tilde{C}_1(\frac{1}{\tilde{z}_+})^{j-1}$, $j = 1, 2, \ldots$, where $\tilde{z}_+$ is obtained from $z_+$ by interchanging $\lambda_1$ and $\lambda_2$. It remains to determine the constants $C_1$ and $\tilde{C}_1$. We shall return to this in Sect. 3.2, where an expression for all $P(i+1, i)$ and $P(i, i+1)$ is derived (cf. (27)). An alternative approach to determining these two constants is to use the two equations $P(1,0) + P(0,1) = \rho(1-\rho)$ and $P(2,0) + P(1,1) + P(0,2) = \rho^2(1-\rho)$ (cf. Sect. 2.1), in combination with an expression for $P(1,1)$ which will be obtained in Sect. 3.1. One thus gets two equations for $P(1,0) = C_1$ and $P(0,1) = \tilde{C}_1$.

### 2.3 $P(N_1 - N_2 = n | N_1 > N_2)$ and $P(N_2 - N_1 = n | N_2 > N_1)$

Taking $z_1 = z = 1/z_2$ in (2), we get a relation between the generating functions $E[z^{N_1 - N_2} I(N_1 > N_2)]$ and $E[z^{N_1 - N_2} I(N_2 > N_1)]$:

$$\begin{aligned} &(\lambda_2 + \mu - \lambda_1 z) E\big[z^{N_1 - N_2} I(N_1 > N_2)\big] \\ &\quad + \left(\lambda_2 + \frac{\mu}{2}(1-z) - \lambda_1 z\right) P(N_1 = N_2) - \frac{\mu}{2}(1-z) P(0,0) \\ &= -\big(\lambda_2 - (\lambda_1 + \mu) z\big) E\big[z^{N_1 - N_2} I(N_2 > N_1)\big]. \end{aligned} \tag{9}$$

Now observe that the terms in the left-hand side are analytic in $z$ for $|z| < 1$, whereas the term in the right-hand side is analytic in $z$ for $|z| > 1$. Application of Liouville's theorem, using the fact that the right-hand side has a finite limit for $|z| \to \infty$, yields that both sides are equal to a constant, say $C$, respectively, for $|z| < 1$ and for $|z| > 1$. In particular, taking $y = 1/z$,

$$E\big[y^{N_2-N_1} I(N_2 > N_1)\big] = \frac{Cy}{\lambda_1 + \mu - \lambda_2 y}, \tag{10}$$

implying that $N_2 - N_1$, when positive, is geom($\frac{\lambda_2}{\lambda_1+\mu}$) distributed. By symmetry (or by studying the left-hand side of (9), which equals $C$ for $|z| < 1$, and by considering the $z^0$ terms in the left-hand side) one may conclude that $N_1 - N_2$, when positive, is geom($\frac{\lambda_1}{\lambda_2+\mu}$) distributed. In the symmetric case $\lambda_1 = \lambda_2$, this was already observed in [9]. Below we would like to interpret this result. Consider the system from the moment $N_2$ reaches the value $N_1 + k$ for some positive $k$, until the level $N_1 + k - 1$ is reached again for the first time. In between, the server will invariably be giving repaired items back to base 2, and never to base 1. The value $k$, when positive, plays no role in this. Hence, $N_2 - N_1$, when positive, is memoryless: $P(N_2 - N_1 \geq k + l | N_2 - N_1 \geq k) = P(N_2 - N_1 \geq l)$. In fact, all the time that $N_2 - N_1 \geq k$, the system behaves like an $M/M/1$ queue with arrival rate $\lambda_2$ and service rate $\lambda_1 + \mu$: the difference $N_2 - N_1$ decreases with rate $\lambda_1 + \mu$. The events when both queues are of equal length will be particularly important; in the next section, we shall derive $P(N_1 = N_2 = i)$, $i = 0, 1, \ldots$.

So far, we have not yet tackled the general problem of finding $E[z_1^{N_1} z_2^{N_2}]$; we only showed that various relevant performance measures have a geometric distribution. The natural approach to the general solution of (2) seems to be to translate the problem into a boundary value problem, like a Riemann–Hilbert problem (cf. [3]). That was also the approach chosen by Cohen [2] in his analysis of the two-dimensional queue length process right after departures in the case of Poisson arrivals and *generally* distributed service times; see also Flatto [5] for the case of exponential service times.

When we compared (2) with formula (1.7) of Cohen [2], we came to the conclusion that his formula for exp($\mu$) service times reduces to our formula (2). This is surprising in view of Remark 1, where it is explained that the exchangeability feature of our repair system leads to different queue length behavior in both models. Below we shall show that, despite that different behavior, the steady-state joint queue length distributions in both models are *the same*.

Cohen [2] studies $(x_n^{(1)}, x_n^{(2)})$, where these are the numbers of customers of types 1 and 2 right after a service completion. He states in his formula (1.6): If $x_n^{(1)} > x_n^{(2)}$, then

$$x_{n+1}^{(1)} = x_n^{(1)} - 1 + \nu_{n+1}^{(1)}, \qquad x_{n+1}^{(2)} = x_n^{(2)} + \nu_{n+1}^{(2)}, \tag{11}$$

where the $\nu_{n+1}^{(i)}$ are numbers of type-$i$ arrivals during the $(n+1)$th service (actually, he distinguishes between arrivals during a service of type 1 and type 2, but we assume all service times are exp($\mu$)). He has similar equations for the other cases. In particular, if $x_n^{(1)} = x_n^{(2)} = 0$, then $x_{n+1}^{(i)} = \nu_{n+1}^{(i)}$, $i = 1, 2$.

In our repair system, we study $(N_1, N_2)$, where $N_i$ is the steady-state number of type-$i$ requests. Let us, however, instead look at subsequent departure epochs (i.e., repair completion epochs), with one significant difference: Remove the idle periods of the system. Moreover, ignore the customer who is the first to arrive after such an idle period (both his arrival and his departure). Call the numbers of requests waiting just *before* the $n$th departure epoch: $(N_n^{(1)}, N_n^{(2)})$. Since we now view the system at independent $\exp(\mu)$ intervals, PASTA (Poisson Arrivals See Time Averages) applies [8]. PASTA states that the distribution of $(N_1, N_2)$ equals the distribution of numbers of requests just before those $\exp(\mu)$ departure intervals. We claim that those satisfy exactly the same recursion as Cohen's $x_n^{(i)}$. Indeed, one may easily verify that, *exactly* as for the $x_n^{(i)}$ in (11), one has: If $N_n^{(1)} > N_n^{(2)}$ then

$$N_{n+1}^{(1)} = N_n^{(1)} - 1 + \nu_{n+1}^{(1)}, \qquad N_{n+1}^{(2)} = N_n^{(2)} + \nu_{n+1}^{(2)}, \tag{12}$$

and other similar equations hold; in particular, if $N_n^{(1)} = N_n^{(2)} = 0$, then $N_{n+1}^{(i)} = \nu_{n+1}^{(i)}$, $i = 1, 2$. The tricky case is when we have $(1, 0)$ or $(0, 1)$ just *before* a service completion. Now an idle period will start. It will be ended with an arrival. As said before, we ignore the idle period *and* the arrival that ends it. Just before the end of the next $\exp(\mu)$, we shall have $(\nu_{n+1}^{(1)}, \nu_{n+1}^{(2)})$ plus that one ignored customer. However, as he is ignored, we have exactly the same recursion relations for $(N_n^{(1)}, N_n^{(2)})$ as Cohen [2] gets for $(x_n^{(1)}, x_n^{(2)})$.

So, although Cohen [2] and we study different quantities (cf. Remark 1), the above reasoning shows that in the case of $\exp(\mu)$ service times, his $(x_n^{(1)}, x_n^{(2)})$ and our $(N_1(t), N_2(t))$ have the same limiting distribution. This is confirmed by the fact that Cohen's formula (1.7) for the generating function, when taking $\exp(\mu)$ service times, agrees with our formula (2).

For general service times, this reasoning fails because then successive service times do not generate a Poisson process, and PASTA cannot be applied.

## 3 Queue lengths—a probabilistic approach

In this section, we shall first determine the probabilities $P(i, i)$, and then present a procedure to obtain all $P(i, j)$.

### 3.1 Determination of $P(i, i)$

We use an argument from Markov renewal theory to derive an expression for (the generating function of) $P(i, i)$.

*Step 1: relate $P(i, i)$ to the steady-state probabilities $\pi_i$ of an underlying Markov chain.*

The successive busy cycles, where a busy cycle $BC$ is the sum of an idle period and consecutive busy period of the server, constitute renewal cycles. Let $\theta_i$ denote the mean number of visits to state $(i, i)$ during a cycle. Since the mean visit time to state $(i, i)$ equals $\frac{1}{\lambda_1+\lambda_2+\mu}$ for $i \geq 1$, we have

$$P(i, i) = \frac{\frac{\theta_i}{\lambda_1+\lambda_2+\mu}}{E[BC]}, \tag{13}$$

where $E[BC] = \frac{1}{\lambda_1+\lambda_2} + \frac{1}{\mu-\lambda_1-\lambda_2} = \frac{1}{\lambda_1+\lambda_2}\frac{1}{1-\rho}$.

Now consider a discrete-time Markov chain $K := \{K_n,\ n = 0, 1, \dots\}$, with state space $\{(0,0), (1,1), \dots, (i,i), \dots\}$, and with transition probabilities $P_{(j,j),(k,k)}$ which will be determined in step 2. This is a Markov chain where we only consider the states where both queue lengths are equal. Its limiting distribution is $\pi_i := \lim_{n\to\infty} P(K_n = i)$. Clearly, $\pi_i$ is proportional to $\theta_i$ for $i \geq 1$. More specifically, since the state $(0,0)$ is visited exactly once per busy cycle, we have

$$\theta_i = \frac{\pi_i}{\pi_0}, \quad i \geq 1. \tag{14}$$

*Step 2: determination of the $\pi_i$.*
The steady-state solution of the Markov chain $K$ satisfies the normalizing condition $\sum_{i=0}^{\infty} \pi_i = 1$ and the balance equations

$$\pi_i = \sum_{k=0}^{\infty} \pi_k P_{(k,k),(i,i)}, \quad i = 0, 1, \dots. \tag{15}$$

Let us now determine the transition probabilities $P_{(k,k),(i,i)}$. Suppose that the original queue length process is in state $(k,k)$. There are three possible events: an arrival from base 1, an arrival from base 2, and a service completion.

(i) An arrival from base 1 occurs first. As indicated in the previous section, one may argue that this arrival starts a busy period, call it $B_1$, with arrival rate $\lambda_1$ and service rate $\lambda_2 + \mu$. Indeed, all the time until equality of the two queue lengths occurs again for the first time (at some level $(i,i)$), repaired items will be handed back to base 1 and not to base 2, since queue 1 is the longer queue. Notice that the stability condition $\lambda_1 + \lambda_2 < \mu$ implies that $\lambda_1 < \lambda_2 + \mu$.
(ii) Similarly, if an arrival from base 2 occurs first when the queue length process is in state $(k,k)$, then it takes a busy period $B_2$ with arrival rate $\lambda_2$ and service rate $\lambda_1 + \mu$ until the system is back at some state $(i,i)$ with equal queue lengths.
(iii) Finally, if a service completion occurs first, then with probability $\frac{1}{2}$ the queue length process moves to $(k, k-1)$, respectively, to $(k-1, k)$, and again a busy period $B_1$ respectively $B_2$ occurs. We shall sometimes speak of a 'difference busy period'.

We now determine the probability that the underlying Markov chain jumps from $(k,k)$ to $(i,i)$, in each of these three possible events.

Let us define $L^{(m)}$ as the number of arrivals from base $3-m$ during $B_m$, $m = 1, 2$. Furthermore, define $K^{(m)}$ as the number of services in the busy period $B_m$, $m = 1, 2$; it is the *sum* of the number of arrivals $L^{(m)}$ and the number of item departures during $B_m$. Since $K^{(m)}$ is the number of customers served in a busy period of an $M/M/1$ queue with arrival rate $\lambda_m$ and service rate $\lambda_{3-m} + \mu$, it follows from $M/M/1$ theory (cf. Chap. II.4 of Cohen [1]) that, for $m = 1, 2$,

$$E\left[z^{K^{(m)}}\right] = \frac{\lambda_1 + \lambda_2 + \mu - \sqrt{\{(\lambda_1 + \lambda_2 + \mu)^2 - 4\lambda_m(\lambda_{3-m} + \mu)z\}}}{2\lambda_m}. \tag{16}$$

$E[z^{L^{(m)}}]$ now follows, since given $K^{(m)} = j$, we have that $L^{(m)} = \text{bin}(j, \frac{\lambda_{3-m}}{\lambda_{3-m}+\mu})$:

$$\begin{aligned} E\bigl[z^{L^{(m)}}\bigr] &= \sum_{j=1}^{\infty} P\bigl(K^{(m)} = j\bigr)\left(\frac{\mu + \lambda_{3-m}z}{\mu + \lambda_{3-m}}\right)^j \\ &= \frac{\lambda_1 + \lambda_2 + \mu - \sqrt{\{(\lambda_1 + \lambda_2 + \mu)^2 - 4\lambda_m(\lambda_{3-m}z + \mu)\}}}{2\lambda_m}. \end{aligned} \tag{17}$$

The key observation in determining the transition probabilities $P_{(k,k),(i,i)}$ in case (i) (so via a difference busy period $B_1$) is the following. During the difference busy period there is no equality of queue lengths. If the busy period starts from $(k+1,k)$ and $L^{(1)} = i-k$, then there have been $i-k$ arrivals from base 2 during $B_1$, and none of these is served in $B_1$. Hence, the next level at which there are equal queue lengths is level $i$ (in state $(i,i)$). Using similar reasonings in cases (ii) and (iii) finally results in the following transition probabilities of the Markov chain $K$:

$$P_{(k,k),(k-1,k-1)} = \frac{\mu}{2(\lambda_1 + \lambda_2 + \mu)}\bigl[P\bigl(L^{(1)} = 0\bigr) + P\bigl(L^{(2)} = 0\bigr)\bigr],$$

$$\begin{aligned} P_{(k,k),(i,i)} &= \frac{\mu}{2(\lambda_1 + \lambda_2 + \mu)}\bigl[P\bigl(L^{(1)} = i-k+1\bigr) + P\bigl(L^{(2)} = i-k+1\bigr)\bigr] \\ &\quad + \frac{\lambda_1}{\lambda_1 + \lambda_2 + \mu}P\bigl(L^{(1)} = i-k\bigr) \\ &\quad + \frac{\lambda_2}{\lambda_1 + \lambda_2 + \mu}P\bigl(L^{(2)} = i-k\bigr), \quad i \geq k, \end{aligned} \tag{18}$$

$$P_{(k,k),(i,i)} = 0, \quad i \leq k-2,$$

$$P_{(0,0),(i,i)} = \frac{\lambda_1}{\lambda_1 + \lambda_2}P\bigl(L^{(1)} = i\bigr) + \frac{\lambda_2}{\lambda_1 + \lambda_2}P\bigl(L^{(2)} = i\bigr).$$

Introducing $a_i := P_{(0,0),(i,i)}$, and $b_l := P_{(k,k),(k+l-1,k+l-1)}$, $l \geq 0$, $k \geq 1$, we notice that we have an $M/G/1$-type Markov chain that satisfies (cf. (15)) the following balance equations:

$$\pi_i = \pi_0 a_i + \sum_{k=1}^{i+1} \pi_k b_{i-k+1}, \quad i = 0, 1, \ldots. \tag{19}$$

Introducing the GF $A(z) := \sum_{i=0}^{\infty} a_i z^i$, $B(z) := \sum_{i=0}^{\infty} b_i z^i$, and $\Pi(z) := \sum_{i=0}^{\infty} \pi_i z^i$, it is easily seen that

$$\Pi(z) = \pi_0 \frac{zA(z) - B(z)}{z - B(z)}. \tag{20}$$

Here,

$$A(z) = \frac{\lambda_1}{\lambda_1 + \lambda_2}E\bigl[z^{L^{(1)}}\bigr] + \frac{\lambda_2}{\lambda_1 + \lambda_2}E\bigl[z^{L^{(2)}}\bigr],$$

$$B(z) = \frac{(\lambda_1 z + \frac{\mu}{2})E[z^{L^{(1)}}] + (\lambda_2 z + \frac{\mu}{2})E[z^{L^{(2)}}]}{\lambda_1 + \lambda_2 + \mu}.$$

$\pi_0$ follows by applying l'Hopital's formula to (20):

$$\pi_0 = \frac{B'(1) - 1}{B'(1) - 1 - A'(1)}, \tag{21}$$

where

$$A'(1) = \frac{\lambda_1}{\lambda_1 + \lambda_2} E\big[L^{(1)}\big] + \frac{\lambda_2}{\lambda_1 + \lambda_2} E\big[L^{(2)}\big],$$

$$B'(1) = \frac{\lambda_1 + \lambda_2 + (\lambda_1 + \frac{\mu}{2})E[L^{(1)}] + (\lambda_2 + \frac{\mu}{2})E[L^{(2)}]}{\lambda_1 + \lambda_2 + \mu}.$$

Finally, let us determine the GF of $P(i,i)$ using (13) and (14):

$$\begin{aligned} \sum_{i=0}^{\infty} z^i P(i,i) &= 1 - \rho + \frac{\rho(1-\rho)}{1+\rho} \sum_{i=1}^{\infty} z^i \frac{\pi_i}{\pi_0} \\ &= 1 - \rho + \frac{\rho(1-\rho)}{1+\rho} \left[ \frac{\Pi(z)}{\pi_0} - 1 \right] \\ &= 1 - \rho + \frac{\rho(1-\rho)}{1+\rho} \frac{z(1 - A(z))}{B(z) - z}. \end{aligned} \tag{22}$$

In particular, it follows that

$$P(N_1 = N_2) = \sum_{i=0}^{\infty} P(i,i) = 1 - \rho + \frac{\rho(1-\rho)}{1+\rho} \frac{A'(1)}{1 - B'(1)}. \tag{23}$$

*Remark 2* Zheng and Zipkin [9], studying the fully symmetric case with $\lambda_1 = \lambda_2$, give a relation between the probabilities $P(i,i)$ and the *marginal queue lengths* $P(N_1 = i)$ in that symmetric case:

$$P(N_1 = i) = \frac{\rho}{2} P(N_1 = i - 1) + \frac{1}{2} P(i,i) + \frac{1}{\rho} P(i+1, i+1), \quad i = 1, 2, \ldots, \tag{24}$$

and $P(N_1 = 0) = (1-\rho)(1 + \sqrt{1+\rho^2})/(1 - \rho + \sqrt{1+\rho^2})$. It is trivial to use (24) to establish a relation between $\sum_{i=0}^{\infty} z^i P(i,i)$ and $\sum_{i=0}^{\infty} z^i P(N_1 = i)$; the latter GF now follows from (22).

### 3.2 Determination of $P(i+1, i)$ and $P(i, i+1)$

In Sects. 2.1, 2.2, and 2.3, we successively derived simple geometric expressions for $P(N_1 + N_2 = n)$, $P(n_1, 0)$ and $P(0, n_2)$, and for $P(N_1 - N_2 = n | N_1 > N_2)$ and $P(N_2 - N_1 = n | N_2 > N_1)$. In Sect. 3.1, we obtained a slightly more complicated expression for the (GF of) $P(i,i)$. In the present subsection, we shall determine $P(i+1, i)$ and $P(i, i+1)$; that is an important building block for obtaining all $P(i, j)$. Once more, there is a crucial role for the idea of having a "difference busy period" that starts at the moment the queue length vector leaves the state $(i, i)$, and that lasts until equality is reached again. In the next subsection, we shall subsequently show how all $P(i, j)$ can be determined once we know the above mentioned probabilities.

Consider the discrete-time Markov chain denoting the state at the times in which the difference between the two queue lengths becomes 0, 1 or $-1$. Its state space $S^*$ consists of the states $(i, i)$, $(i+1, i)$ and $(i, i+1)$, $i = 0, 1, \dots$. It has the following one-step transition probabilities:

$$P_{(0,0),(1,0)} = \frac{\lambda_1}{\lambda_1 + \lambda_2}, \quad P_{(0,0),(0,1)} = \frac{\lambda_2}{\lambda_1 + \lambda_2},$$

and for $i = 1, 2, \dots$,

$$P_{(i,i),(i+1,i)} = \frac{\lambda_1}{\lambda_1 + \lambda_2 + \mu}, \quad P_{(i,i),(i,i+1)} = \frac{\lambda_2}{\lambda_1 + \lambda_2 + \mu},$$

$$P_{(i,i),(i-1,i)} = P_{(i,i),(i,i-1)} = \frac{\mu}{2(\lambda_1 + \lambda_2 + \mu)},$$

$$P_{(i+1,i),(i,i)} = P_{(i,i+1),(i,i)} = \frac{\mu}{\lambda_1 + \lambda_2 + \mu},$$

$$P_{(i,i+1),(i+1,i+1)} = \frac{\lambda_1}{\lambda_1 + \lambda_2 + \mu}, \quad P_{(i+1,i),(i+1,i+1)} = \frac{\lambda_2}{\lambda_1 + \lambda_2 + \mu},$$

$$P_{(i+1,i),(i+l+1,i+l)} = \frac{\lambda_1 P(L^{(1)} = l)}{\lambda_1 + \lambda_2 + \mu}, \quad P_{(i,i+1),(i+l,i+l+1)} = \frac{\lambda_2 P(L^{(2)} = l)}{\lambda_1 + \lambda_2 + \mu}.$$

The last line perhaps requires an explanation. If a type-1 arrival occurs in state $(i+1, i)$, then the difference becomes 2. It now takes a "difference busy period" until the difference returns to 1. With probability $P(L^{(1)} = l)$, $l$ customers of type 2 arrive during this busy period, so in the discrete-time Markov chain under consideration we then move from state $(i+1, i)$ to state $(i+l+1, i+l)$.

Let us denote the limiting probabilities of the Markov chain by $\pi(i+1, i)$, $\pi(i, i+1)$ and $\pi(i, i)$, $i = 0, 1, \dots$. Since $1/\pi(0, 0)$ is the expected number of visits in states of $S^*$ during the server busy period, we get that $\frac{\pi(i,i)}{\pi(0,0)}$, $\frac{\pi(i+1,i)}{\pi(0,0)}$ and $\frac{\pi(i,i+1)}{\pi(0,0)}$, $i = 1, 2, \dots$ are, respectively, the expected numbers of visits to states $(i, i)$, $(i+1, i)$, and $(i, i+1)$ during the server busy period. Notice that, hence, $\pi(i, i)/\pi(0, 0)$ is identical to the $\pi_i/\pi_0$ of the previous subsection; they both represent the mean number of visits to state $(i, i)$ during a server busy period. Let us concentrate on $P(i+1, i)$; $P(i, i+1)$ then follows by symmetry (interchanging $\lambda_1$ and $\lambda_2$). We find for $i = 0, 1, \dots$:

$$\pi(i+1, i) = \sum_{k=0}^{i} \pi(k+1, k) \frac{\lambda_1 P(L^{(1)} = i-k)}{\lambda_1 + \lambda_2 + \mu} + \pi(i, i) \frac{\lambda_1}{\lambda_1 + \lambda_2 + \mu} + \pi(i+1, i+1) \frac{\mu}{2(\lambda_1 + \lambda_2 + \mu)}. \tag{25}$$

Dividing both sides by $\pi(0,0)$ and introducing $F_+(z) := \sum_{i=0}^{\infty} \frac{\pi(i+1,i)}{\pi(0,0)} z^i$ and $\Theta(z) := \sum_{i=0}^{\infty} \theta_i z^i$ where, as before, $\theta_i = \pi_i/\pi_0 = \pi(i,i)/\pi(0,0)$, we find after a straightforward calculation:

$$F_+(z) = \sum_{i=0}^{\infty} \frac{\pi(i+1,i)}{\pi(0,0)} z^i = \frac{\frac{\lambda_1}{\lambda_1+\lambda_2+\mu}\Theta(z) + \frac{\mu}{2(\lambda_1+\lambda_2+\mu)}\frac{\Theta(z)-1}{z}}{1 - \frac{\lambda_1}{\lambda_1+\lambda_2+\mu} E[z^{L^{(1)}}]}. \tag{26}$$

The GF $F_+(z)$ of the $\frac{\pi(i+1,i)}{\pi(0,0)}$ now follows from the known expressions for $E[z^{L^{(1)}}]$ (cf. (17)) and for $\Theta(z)$ (via $\Pi(z)$, cf. (20)).

We finally obtain $P(i+1,i)$. Remembering that $\frac{\pi(i+1,i)}{\pi(0,0)}$ is the expected number of visits to state $(i+1,i)$ during the server busy period, and using Markov renewal theory, it follows that, like in (13),

$$P(i+1,i) = \frac{\frac{1}{\lambda_1+\lambda_2+\mu}}{E[BC]} \frac{\pi(i+1,i)}{\pi(0,0)} = \frac{\rho(1-\rho)}{1+\rho} \frac{\pi(i+1,i)}{\pi(0,0)}, \quad i = 0, 1, \dots. \tag{27}$$

Using (26), which gives the GF of $P(i+1,i)$, this determines all $P(i+1,i)$ for $i = 0, 1, \dots$, and by symmetry the $P(i,i+1)$ also follow. Notice that $P(j,0)$ and $P(0,j)$ were given in Sect. 2.2 *up to a multiplicative constant*; that constant now follows since $P(1,0)$ also is given by (27).

### 3.3 Determination of $P(i,j)$

We now describe a recipe to find the remaining $P(i,j)$, $j \geq i+2$, and $i \geq j+2$. By symmetry, we can concentrate on the former case. First, we indicate how to obtain all $P(1,j)$, $j \geq 3$; $P(1,2)$ follows from the results of the previous subsection. From (1), we have

$$(\lambda_1 + \lambda_2 + \mu) P(1,j) = \lambda_1 P(0,j) + \lambda_2 P(1,j-1) + \mu P(1,j+1). \tag{28}$$

For $j = 2$, this immediately determines $P(1,3)$ as we know all other terms: $P(0,2)$ from Sect. 2.2, $P(1,1)$ from Sect. 3.1 and $P(1,2)$ from Sect. 3.2. For $j \geq 3$, one can now use (28) to get $P(1,j+1)$ once we have $P(1,j-1)$ and $P(1,j)$. We refrain from giving details, but we would like to observe the following. Equation (28) is an inhomogeneous second-order difference equation. The general solution of the *homogeneous* second-order difference equation reads $K_1 a_+^j + K_2 a_-^j$, $j \geq 2$, where $1/a_+$ and $1/a_-$ are the two zeros of the equation $\lambda_2 x^2 - (\lambda_1 + \lambda_2 + \mu)x + \mu = 0$. We already encountered this expression, with $\lambda_1$ and $\lambda_2$ interchanged, in the denominator of (6). There we concluded that one of the two roots has absolute value smaller than one. Accordingly, we may conclude that the general solution of the homogeneous second-order differential equation reads $K_1 a_+^j$, $j \geq 2$. Turning to the inhomogeneous equation, it should be observed that $P(0,j) = \tilde{C}_1 a_+^j$, with exactly the same parameter $a_+$ as for the homogeneous equation. This follows, by symmetry, from (8). The theory of inhomogeneous difference equations now implies that the general solution of (28) is given by $(L_0 + L_1 j) a_+^{j-1}$. This would also readily follow via an approach with generating functions, expressing $\sum P(1,j) z^j$ into $\sum P(0,j) z^j$.

Next, we indicate how to obtain all $P(i, j)$, $i \geq 2$, $j \geq i + 2$. From (1), we have

$$(\lambda_1 + \lambda_2 + \mu)P(i, j) = \lambda_1 P(i - 1, j) + \lambda_2 P(i, j - 1) + \mu P(i, j + 1). \tag{29}$$

After having obtained $P(i, j - 1)$ and $P(i, j)$, and the (lower level) $P(i - 1, j)$, the $P(i, j + 1)$ follow. Again observe that (29) is an inhomogeneous second-order difference equation of exactly the same form as (28). By induction, one may show that its solution has the form:

$$P(i, j) = \sum_{m=0}^{i} h_m j^m a_+^{j-i}, \quad j = i + 1, i + 2, \ldots. \tag{30}$$

In fact, in the completely symmetric case this was proven in [9], and an algorithm was provided to determine the $h_m$.

*Remark 3* An interesting feature of the present model is that, for general service times, it has only been solved via the boundary value method (cf. Cohen [2]). In almost all two-dimensional queueing problems which have been solved via the boundary value method, taking exponential service times does not simplify the problem to such an extent that one no longer needs to rely on that method. In the present problem, though, there seems to be so much structure that, in the exponential case, the $P(i, j)$ have the nice form indicated in (30).

### 3.4 Numerical example

We have implemented the formulas and algorithms for calculating $P(i, j)$, as outlined above, in MATLAB. We have first computed the $P(i, j)$ in a completely symmetric case ($\lambda_1 = \lambda_2$) that was already studied in [9], and we verified that we obtained the same numbers as in their Table III. Next, we took an asymmetric case: $\lambda_1 = 2$, $\lambda_2 = 1$ and $\mu = 4$; so the load $\rho = 0.75$. The results are presented in Table 1. First observe that $P(0, 0) = 1 - \rho = 0.25$. Next, one may observe that $\sum_{j=0}^{i} P(j, i - j) = (1 - \rho)\rho^i$ (Sect. 2.1) and that the $P(i, 0)$ and $P(0, j)$ decrease geometrically fast (Sect. 2.2). We have computed the $P(i, i)$ using (22). Notice that $P(1, 1)$ was used in calculating the constants $C_1 = P(1, 0)$ and $\tilde{C}_1 = P(0, 1)$ as outlined at the end of Sect. 2.2. Next, we determined $P(i + 1, i)$ and $P(i, i + 1)$ using (25) and (27). Finally, we have computed $P(i, j)$ for $|i - j| \geq 2$ using (28)–(30). Notice that $P(i, j) > P(j, i)$ for $i > j$; this makes sense as $\lambda_1 > \lambda_2$.

*Remark 4* In the case of equal arrival rates, the computation of the steady-state probabilities simplifies a bit. In particular, one now has $P(i, j) = P(j, i)$. $P(j, 0) = P(0, j)$ follow as before from (8). The $P(i, i)$ follow from (22). $P(i, i + 1) = P(i + 1, i)$ follow using the relations $(2\lambda + \mu)P(i, i) = 2\lambda P(i - 1, i) + 2\mu P(i + 1, i)$. Note that $P(1, 0) = P(0, 1)$ were already determined. Finally, $P(i, j) = P(j, i)$ follow for $|i - j| \geq 2$ as in Sect. 3.3.

**Table 1** Queue length probabilities $P(i, j)$ for the case $\lambda_1 = 2$, $\lambda_2 = 1$, $\mu = 4$

| $ij$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| 0 | 0.250000 | 0.066432 | 0.010425 | 0.001636 | 0.000257 | 0.000040 | 0.000006 | 0.000001 |
| 1 | 0.121068 | 0.086662 | 0.030848 | 0.005411 | 0.000938 | 0.000161 | 0.000028 | 0.000005 |
| 2 | 0.043537 | 0.057328 | 0.043391 | 0.015993 | 0.002840 | 0.000502 | 0.000088 | 0.000015 |
| 3 | 0.015657 | 0.024413 | 0.030185 | 0.023189 | 0.008623 | 0.001531 | 0.000271 | 0.000048 |
| 4 | 0.005630 | 0.010145 | 0.013431 | 0.016414 | 0.012686 | 0.004734 | 0.000838 | 0.000149 |
| 5 | 0.002025 | 0.004139 | 0.005875 | 0.007430 | 0.009052 | 0.007018 | 0.002624 | 0.000464 |
| 6 | 0.000728 | 0.001665 | 0.002532 | 0.003326 | 0.004131 | 0.005028 | 0.003906 | 0.001462 |
| 7 | 0.000262 | 0.000662 | 0.001076 | 0.001473 | 0.001871 | 0.002305 | 0.002805 | 0.002182 |

## 4 Sojourn times

The main purpose of this section is to express the LST of the sojourn time distribution of a customer into the joint queue length distribution that was derived in the previous sections. We focus on a customer who has brought a failed item of type 1; by interchanging indices 1 and 2 (in particular, the arrival rates), we then also obtain the sojourn time LST for items of type 2.

### 4.1 The LST of the sojourn time distribution

Let $X_{k,j} :=$ sojourn time of a type-1 customer who increases the number of customers waiting in line 1 from $j-1$ to $j$, and whose arrival increases the difference between numbers of waiting customers in lines 1 and 2 from $k-1$ to $k$, $k \geq 1$, $j \geq 1$. We similarly define $X_{k,j}$ for $k \leq 0$; in that case, the difference between lines 1 and 2 again decreases from $k-1$ to $k$. Define $\Psi_{k,j}(\alpha) := E[\mathrm{e}^{-\alpha X_{k,j}}]$.

*Case I: $k \geq 0$*

Let us first concentrate on the case $k \geq 1$, $j \geq 1$. Notice that $X_{k,j}$ lasts until the moment that $j$ items have been returned to base 1. Conditioning on the amount of time until the first event occurs, we can write for $k \geq 1$, $j \geq 1$:

$$\Psi_{k,j}(\alpha) = \frac{\mu}{\mu + \lambda_1 + \lambda_2 + \alpha}\Psi_{k-1,j-1}(\alpha) + \frac{\lambda_1}{\mu + \lambda_1 + \lambda_2 + \alpha}\Psi_{k+1,j}(\alpha) + \frac{\lambda_2}{\mu + \lambda_1 + \lambda_2 + \alpha}\Psi_{k-1,j}(\alpha). \tag{31}$$

Here, we define $\Psi_{k,0} = 1$, $k \geq 1$. Similarly, we obtain for $k = 0$, $j \geq 1$:

$$\Psi_{0,j}(\alpha) = \frac{\mu}{2(\mu + \lambda_1 + \lambda_2 + \alpha)}\Psi_{-1,j-1}(\alpha) + \frac{\mu}{2(\mu + \lambda_1 + \lambda_2 + \alpha)}\Psi_{1,j}(\alpha) + \frac{\lambda_1}{\mu + \lambda_1 + \lambda_2 + \alpha}\Psi_{1,j}(\alpha) + \frac{\lambda_2}{\mu + \lambda_1 + \lambda_2 + \alpha}\Psi_{-1,j}(\alpha). \tag{32}$$

We shall solve this set of recurrence relations using generating functions. Let

$$G(z_1, z_2; \alpha) := \sum_{k=0}^{\infty}\sum_{j=0}^{\infty} z_1^k z_2^j \Psi_{k,j}(\alpha), \tag{33}$$

$$G_0(z_2; \alpha) := \sum_{j=0}^{\infty} z_2^j \Psi_{0,j}(\alpha), \tag{34}$$

$$G_1(z_2; \alpha) := \sum_{j=0}^{\infty} z_2^j \Psi_{1,j}(\alpha). \tag{35}$$

Multiplying both sides of (31) by $z_1^k z_2^j$ and summing over $k \geq 1$, $j \geq 1$ yields:

$$\begin{aligned} G(z_1, z_2; \alpha) &- \frac{z_1}{1-z_1} - G_0(z_2; \alpha) \\ &= \frac{\mu z_1 z_2}{\mu + \lambda_1 + \lambda_2 + \alpha} G(z_1, z_2; \alpha) \\ &\quad + \frac{\lambda_1}{z_1(\mu + \lambda_1 + \lambda_2 + \alpha)}\left[G(z_1, z_2; \alpha) - \frac{z_1^2}{1-z_1} - G_0(z_2; \alpha) - z_1 G_1(z_2; \alpha)\right] \\ &\quad + \frac{\lambda_2 z_1}{\mu + \lambda_1 + \lambda_2 + \alpha}\left[G(z_1, z_2; \alpha) - \frac{1}{1-z_1}\right]. \end{aligned} \tag{36}$$

We need to determine $G_0(z_2; \alpha)$ and $G_1(z_2; \alpha)$. One relation between these two functions is obtained via (32). First, we rewrite that equation by observing that

$$\Psi_{-1,j}(\alpha) = \Gamma(\alpha)\Psi_{0,j}(\alpha), \tag{37}$$

where $\Gamma(\alpha) = E[\mathrm{e}^{-\alpha B_2}]$, the LST of the difference busy period corresponding to an $M/M/1$ queue with arrival rate $\lambda_2$ and service rate $\lambda_1 + \mu$. The idea behind (37) is the following. If the tagged type-1 customer arrives to find $j-1$ customers in $Q_1$ and $j+1$ customers in $Q_2$, leading to a state $(j, j+1)$, then it first takes a difference busy period $B_2$ until the two queue lengths are again equal. No type-1 items are returned during that busy period. At the end of $B_2$, the state has become $(j+m, j+m)$ for some $m \geq 0$, with the tagged customer still in position $j$ of $Q_1$. For the sojourn time of the tagged customer, it makes no difference whether the system is in state $(j+m, j+m)$ or in state $(j, j)$.

*Remark 5* To see that it indeed makes no difference for the sojourn time of the tagged customer whether the system is in state $(j+m, j+m)$ or in state $(j, j)$, suppose that a type-1 customer (just for convenience we refer to this customer as the *red customer*), arrives to find the system at state $(j-1, j)$. That means that he finds $j-1$ type-1 customers in front of him in line and also $j$ type-2 customers; however, it is not yet determined how many of them are *served* before him. The reason for that is that in principle, the sojourn time of the *red customer* depends on future arrivals of both types of customers. After the admittance of the *red customer*, the state of the system becomes $(j, j)$, and the LST of the sojourn time of the *red customer* is $\Psi_{0,j}(\alpha)$. However, the latter dependence on future arrivals has a special regenerative property. For example, suppose that $m$ type-1 customers and $m$ type-2 customers were admitted to the system after the arrival of the *red customer* and before the service completion of the item being served. Then, by the memoryless property of the service, the residual sojourn time of the *red customer* (the time it takes from the arrival of the above $2m$th customer until the red customer leaves the system), is stochastically equal to the sojourn time of the red customer. Thus, the LST of the above residual sojourn time is also $\Psi_{0,j}(\alpha)$.

Thus, rewriting (32) yields for $j = 1$:

$$\begin{aligned}\Psi_{0,1}(\alpha) &= \frac{\mu}{2(\mu+\lambda_1+\lambda_2+\alpha)} + \frac{\mu+2\lambda_1}{2(\mu+\lambda_1+\lambda_2+\alpha)}\Psi_{1,1}(\alpha) \\ &\quad + \frac{\lambda_2}{\mu+\lambda_1+\lambda_2+\alpha}\Gamma(\alpha)\Psi_{0,1}(\alpha) \end{aligned} \tag{38}$$

and for $j \geq 2$:

$$\begin{aligned}\Psi_{0,j}(\alpha) &= \frac{\mu}{2(\mu+\lambda_1+\lambda_2+\alpha)}\Gamma(\alpha)\Psi_{0,j-1}(\alpha) \\ &\quad + \frac{\mu+2\lambda_1}{2(\mu+\lambda_1+\lambda_2+\alpha)}\Psi_{1,j}(\alpha) + \frac{\lambda_2}{\mu+\lambda_1+\lambda_2+\alpha}\Gamma(\alpha)\Psi_{0,j}(\alpha), \end{aligned} \tag{39}$$

or equivalently, for $j \geq 2$:

$$\begin{aligned}\left[1 - \frac{\lambda_2\Gamma(\alpha)}{\mu+\lambda_1+\lambda_2+\alpha}\right]\Psi_{0,j}(\alpha) &= \frac{\mu\Gamma(\alpha)}{2(\mu+\lambda_1+\lambda_2+\alpha)}\Psi_{0,j-1}(\alpha) \\ &\quad + \frac{\mu+2\lambda_1}{2(\mu+\lambda_1+\lambda_2+\alpha)}\Psi_{1,j}(\alpha). \end{aligned} \tag{40}$$

Multiplying the terms of the last equation by $z_2^j$ and summing over $j \geq 1$ gives:

$$\begin{aligned}&\left[1 - \frac{\lambda_2\Gamma(\alpha)}{\mu+\lambda_1+\lambda_2+\alpha}\right]\left[G_0(z_2;\alpha) - 1\right] \\ &\quad = \frac{\mu\Gamma(\alpha)z_2}{2(\mu+\lambda_1+\lambda_2+\alpha)}\left[G_0(z_2;\alpha) - 1\right] + \frac{\mu}{2(\mu+\lambda_1+\lambda_2+\alpha)} \\ &\qquad + \frac{\mu+2\lambda_1}{2(\mu+\lambda_1+\lambda_2+\alpha)}\left[G_1(z_2;\alpha) - 1\right]. \end{aligned} \tag{41}$$

Subsequently, we derive a second relation between $G_0(z_2;\alpha)$ and $G_1(z_2;\alpha)$. We first rewrite (36), by grouping all $G(z_1, z_2;\alpha)$ terms, and multiplying all terms by $\mu + \lambda_1 + \lambda_2 + \alpha$:

$$\left[\mu(1 - z_1z_2) + \lambda_1\left(1 - \frac{1}{z_1}\right) + \lambda_2(1 - z_1) + \alpha\right]G(z_1, z_2;\alpha)$$

$$= \left(\mu + \lambda_1\left(1 - \frac{1}{z_1}\right) + \lambda_2 + \alpha\right) G_0(z_2;\alpha) - \lambda_1 G_1(z_2;\alpha) + \frac{z_1}{1-z_1}(\mu+\alpha). \tag{42}$$

Consider the factor in front of $G(z_1, z_2; \alpha)$ in the left-hand side of (42). Multiplying this factor by $z_1$ and equating the result to zero yields the following quadratic equation in $z_1$:

$$(\lambda_2 + \mu z_2)z_1^2 - (\mu + \lambda_1 + \lambda_2 + \alpha)z_1 + \lambda_1 = 0. \tag{43}$$

We shall show that this equation has one root $z_1^+(z_2)$ inside the unit circle, and one root $z_1^-(z_2)$ outside the unit circle. Notice that $1/z_1^+(z_2)$ and $1/z_1^-(z_2)$ are the two roots of the equation

$$\lambda_1 y^2 - (\mu + \lambda_1 + \lambda_2 + \alpha)y + (\lambda_2 + \mu z_2) = 0. \tag{44}$$

Following the reasoning that led to (16) and (17), it can be shown that the minus root of (44) is given by $E[z_2^{K^{(1)}-L^{(1)}} \mathrm{e}^{-\alpha B_1}]$, where $K^{(1)} - L^{(1)}$ equals the number of item departures during the difference busy period $B_1$. This interpretation immediately shows that this $y$ lies inside the unit circle. The sum of the two $y$-roots equals $\frac{\mu+\lambda_1+\lambda_2+\alpha}{\lambda_1}$, which in absolute value is at least 2 since $\lambda_1 + \lambda_2 < \mu$. Hence, one $y$-root is inside the unit circle and the other one lies outside, implying that the same holds for the roots of (43).

Since $G(z_1, z_2; \alpha)$ is analytic in $z_1$ for $|z_1| < 1$, the right-hand side of (42) must be zero for $z_1 = z_1^+$. This yields a second relation between $G_0(z_2; \alpha)$ and $G_1(z_2; \alpha)$:

$$\bigl[\mu z_1^+(z_2) + \lambda_1\bigl(z_1^+(z_2) - 1\bigr) + (\lambda_2 + \alpha)z_1^+(z_2)\bigr] G_0(z_2;\alpha) - \lambda_1 z_1^+(z_2) G_1(z_2;\alpha) + \frac{(z_1^+(z_2))^2}{1 - z_1^+(z_2)}(\mu + \alpha) = 0. \tag{45}$$

We are thus able to determine those functions, and finally $G(z_1, z_2; \alpha)$ follows from (42).

In this way, we find the double GF of the sojourn time LST for a customer of type 1 who arrives to find $j - 1$ customers waiting at $Q_1$, and whose arrival increases the difference between numbers of waiting customers in lines 1 and 2 from $k - 1$ to $k$, for both $k \geq 1$ and $k = 0$ (see $G_0(z_2; \alpha)$).

*Case II: $k < 0$*

In handling the case $k < 0$, we use the same argument as in (37): if a customer arrives at base 1 and finds there $j - 1$ waiting customers while the difference with the queue length in $Q_2$ decreases to $k < 0$, then

$$\Psi_{k,j}(\alpha) = \Gamma(\alpha)^{-k} \Psi_{0,j}(\alpha). \tag{46}$$

Indeed, $\Gamma(\alpha)^k = (E[\mathrm{e}^{-\alpha B_2}])^k$ is the LST of the time it takes to reduce the difference between the two queue lengths to zero. Of course, during that process the queue lengths move to some state $(j+m, j+m)$, but for the tagged customer in position $j$ of $Q_1$, the value of $m \geq 0$ is irrelevant.

We thus obtain all conditional sojourn time LST's. Multiplying by the probabilities $P(j-1, j-k)$ as seen by an arriving customer (use PASTA to conclude that these arrival probabilities are exactly the probabilities which were calculated in Sects. 2 and 3) and summing yields the unconditional sojourn time LST.

### 4.2 Further results for the sojourn time distribution

One might use the results of the previous subsection to obtain mean conditional sojourn times $E(X_{k,j})$. In the present subsection, we present an alternative approach for obtaining those means.

Let $T_j$ be the number of arrivals into base 1 minus the number of arrivals into base 2 that occur between the $(j-1)$th and the $j$th departure. Clearly, $\{T_j; j \geq 1\}$ is a sequence of i.i.d. random variables; let $T$ be the generic random variable of the sequence.

**Lemma 1**

$$P(T=n) = \mu\left(\frac{\lambda_1}{\lambda_2}\right)^{\frac{n}{2}}\left[2\sqrt{\lambda_1\lambda_2\left(a^2-1\right)}\right]^{-1}\left\{a-\sqrt{a^2-1}\right\}^{|n|}, \quad n = 0, \pm 1, \pm 2, \ldots$$

*where*

$$a = \frac{\lambda_1+\lambda_2+\mu}{2\sqrt{\lambda_1\lambda_2}} > 1.$$

*Proof* Let $V_i$ be the number of arrivals into base $i$ $(i = 1, 2)$ between two successive departures and let $X$ be the generic service time of an item. By conditioning on $X$, $V_i$ is a Poisson random variable with mean $\lambda_i/\mu$. We have for $n = 0, \pm 1, \pm 2, \ldots$

$$\begin{aligned}
P(T=n) &= \int_0^\infty P(V_1 - V_2 = n \mid X = x) f_X(x)\, dx \\
&= \sum_{k=0}^{\infty} \int_0^\infty e^{-\lambda_1 x}\frac{(\lambda_1 x)^{k+n}}{(k+n)!} e^{-\lambda_2 x}\frac{(\lambda_2 x)^k}{k!}\mu e^{-\mu x}\, dx \\
&= \mu\left(\frac{\lambda_1}{\lambda_2}\right)^{\frac{n}{2}}(2\sqrt{\lambda_1\lambda_2})^{-1}\int_0^\infty \exp\left\{-y(\lambda_1+\lambda_2+\mu)/2\sqrt{\lambda_1\lambda_2}\right\} \\
&\quad \times \sum_{k=0}^{\infty}\frac{(y/2)^{n+2k}}{k!(k+n)!}\, dy \\
&= \mu\left(\frac{\lambda_1}{\lambda_2}\right)^{\frac{n}{2}}(2\sqrt{\lambda_1\lambda_2})^{-1}\int_0^\infty \exp\left\{-y(\lambda_1+\lambda_2+\mu)/2\sqrt{\lambda_1\lambda_2}\right\} \cdot I_n(y)\, dy
\end{aligned} \tag{47}$$

where $I_n(y) = \sum_{k=0}^{\infty} \frac{(y/2)^{n+2k}}{k!(k+n)!}$ is the Bessel function (of order $n$) of a purely imaginary argument. Using [6], we get

$$\int_0^{\infty} e^{-ay} I_n(y)\,dy = \left(\sqrt{a^2-1}\right)^{-1}\left(a-\sqrt{a^2-1}\right)^n, \quad n = 0, 1, 2, \dots\ a > 0 \tag{48}$$

where

$$a = \frac{\lambda_1 + \lambda_2 + \mu}{2\sqrt{\lambda_1 \lambda_2}}.$$

□

Note that

$$P(T = n) = \begin{cases} cb_1^n, & n \ge 0, \\ cb_2^{-n}, & n < 0, \end{cases} \tag{49}$$

where

$$c = (\mu/2)\left[\lambda_1\lambda_2\left(a^2-1\right)\right]^{-1/2},$$

$$b_1 = (\lambda_1/\lambda_2)^{1/2}\left(a - \left(a^2-1\right)^{1/2}\right),$$

and

$$b_2 = (\lambda_2/\lambda_1)^{1/2}\left(a - \left(a^2-1\right)^{1/2}\right).$$

By using Lemma 1, we derive some relevant relations for $\Psi_{k,j}$, $k \ge 0$, $j \ge 1$. The relations are computed via a recursive algorithm that will be implemented to compute $E_{k,j} := E(X_{k,j})$.

$$E\left(e^{-\alpha X_{0,1}} \mid T = n\right) = \begin{cases} \frac{\mu}{\mu+\alpha}, & n \ge 1, \\ \frac{1}{2}\frac{\mu}{\mu+\alpha}[1 + \Psi_{1,1}(\alpha)], & n = 0, \\ \frac{\mu}{\mu+\alpha}[\Gamma(\alpha)]^{-(n+1)}\Psi_{0,1}(\alpha), & n \le -1, \end{cases} \tag{50}$$

so that by Lemma 1

$$\begin{aligned} \Psi_{0,1}(\alpha) = {} & \frac{\mu c b_1}{(\mu+\alpha)(1-b_1)} + \frac{1}{2}\frac{\mu c}{\mu+\alpha}\left[1 + \Psi_{1,1}(\alpha)\right] \\ & + \frac{\mu c b_2}{(\mu+\alpha)(1 - b_2\Gamma(\alpha))}\Psi_{0,1}(\alpha). \end{aligned} \tag{51}$$

For $k \ge 1$,

$$E\left(e^{-\alpha X_{k,1}} \mid T = n\right) = \begin{cases} \frac{\mu}{\mu+\alpha}, & n \ge 1-k, \\ \frac{1}{2}\frac{\mu}{\mu+\alpha}[1 + \Psi_{1,1}(\alpha)], & n = -k, \\ \frac{\mu}{\mu+\alpha}[\Gamma(\alpha)]^{-(n+k+1)}\Psi_{0,1}(\alpha), & n \le -k-1, \end{cases}$$

and

$$\Psi_{k,1}(\alpha) = \frac{\mu}{\mu+\alpha}\left\{\frac{c}{1-b_1} + \frac{cb_2(1-b_2^{k-1})}{1-b_2} + \frac{cb_2^k}{2}\big[1+\Psi_{1,1}(\alpha)\big] + \Psi_{0,1}(\alpha)\frac{cb_2^{k+1}}{(1-b_2\Gamma(\alpha))}\right\}. \tag{52}$$

Substituting $\alpha = 0$ in the first derivative of (51) and (52), we obtain

$$E_{0,1} = \frac{1}{\mu} + \frac{c}{2}E_{1,1} + \frac{cb_2}{1-b_2}E_{0,1} + \frac{cb_2^2}{(1-b_2)^2(\lambda_1+\mu-\lambda_2)} \tag{53}$$

and for $k \geq 1$

$$\begin{aligned} E_{k,1} &= \frac{1}{\mu} + \frac{cb_2^k}{2}E_{1,1} + \frac{cb_2^{k+1}}{1-b_2}E_{0,1} + \frac{cb_2^{k+2}}{(1-b_2)^2(\lambda_1+\mu-\lambda_2)} \\ &= \frac{1}{\mu} + \left[\frac{c}{2}E_{1,1} + \frac{cb_2}{1-b_2}E_{0,1} + \frac{cb_2^2}{(1-b_2)^2(\lambda_1+\mu-\lambda_2)}\right]b_2^k. \end{aligned} \tag{54}$$

Note that in the first phase we compute $E_{0,1}$ and $E_{1,1}$. Then, in the second phase, the $E_{k,1}$ for $k \geq 2$ are obtained recursively.

From (53) and (54), we get

$$E_{k,1} = b_2^k E_{0,1} + \frac{1-b_2^k}{\mu}, \tag{55}$$

where obviously,

$$\lim_{k\longrightarrow\infty} E_{k,1} = \frac{1}{\mu}.$$

Similarly, we create recursive equations for $\Psi_{0,j}(\alpha)$, $j \geq 1$:

$$E\big(e^{-\alpha X_{0,j}} \mid T = n\big) = \begin{cases} \frac{\mu}{\mu+\alpha}\Psi_{n-1,j-1}(\alpha), & n \geq 1, \\ \frac{\mu}{\mu+\alpha}\frac{1}{2}[\Gamma(\alpha)\Psi_{0,j-1}(\alpha) + \Psi_{1,j}(\alpha)], & n = 0, \\ \frac{\mu}{\mu+\alpha}[\Gamma(\alpha)]^{-(n+1)}\Psi_{0,j}(\alpha), & n \leq -1. \end{cases} \tag{56}$$

By Lemma 1, we get (after substituting $\alpha = 0$ in the first derivative)

$$\begin{aligned} E_{0,j} = {} & \frac{1}{\mu} + \sum_{n=1}^{\infty} E_{n-1,j-1}\cdot cb_1^n + \frac{c}{2}E_{0,j-1} + \frac{c}{2(\lambda_1+\mu-\lambda_2)} \\ & + \frac{c}{2}E_{1,j} + \frac{cb_2}{1-b_2}E_{0,j} + \frac{cb_2^2}{(1-b_2)^2(\lambda_1+\mu-\lambda_2)}. \end{aligned} \tag{57}$$

Substituting $\alpha = 0$ in the first derivative of (40), we get for $j \geq 2$:

$$E_{0,j}(\mu+\lambda_1) = \frac{3\mu+2\lambda_1}{2(\mu+\lambda_1-\lambda_2)} + \frac{\mu}{2}E_{0,j-1} + \frac{\mu+2\lambda_1}{2}E_{1,j}. \tag{58}$$

**Table 2** Conditional expected sojourn time $E_{k,j}$ for the case $\lambda_1 = 2$, $\lambda_2 = 1$, $\mu = 4$

| $kj$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| 0 | 0.379555 | 0.811411 | 1.224560 | 1.631373 | 2.035292 | 2.437686 | 2.839205 |
| 1 | 0.269332 | 0.627366 | 1.031134 | 1.434780 | 1.837252 | 2.238883 | 2.639965 |
| 2 | 0.252885 | 0.531077 | 0.873868 | 1.256923 | 1.649990 | 2.046662 | 2.444902 |
| 3 | 0.250430 | 0.506422 | 0.788239 | 1.119573 | 1.486777 | 1.869373 | 2.259407 |
| 4 | 0.250064 | 0.501170 | 0.759749 | 1.042309 | 1.364878 | 1.719494 | 2.092150 |
| 5 | 0.250009 | 0.500022 | 0.752156 | 1.011815 | 1.294536 | 1.609989 | 1.954075 |
| 6 | 0.250001 | 0.500000 | 0.750332 | 1.000703 | 1.263822 | 1.545459 | 1.854200 |
| 7 | 0.250000 | 0.500000 | 0.750039 | 1.000000 | 1.252476 | 1.515039 | 1.792684 |

To compute $E_{k,2}$ for $k \geq 0$, substitute (55) in (57) and use (58).

By substituting $\alpha = 0$ in the first derivative of (31), we get

$$E_{k,j} = \frac{1}{\mu + \lambda_1 + \lambda_2} + \frac{\mu}{\mu + \lambda_1 + \lambda_2} E_{k-1,j-1} + \frac{\lambda_1}{\mu + \lambda_1 + \lambda_2} E_{k+1,j} + \frac{\lambda_2}{\mu + \lambda_1 + \lambda_2} E_{k-1,j}. \tag{59}$$

By (59), $E_{0,2}$, $E_{1,2}$, and $E_{k,1}$, $k \geq 0$, are used for the computation of $E_{k,2}$; the recursion is complete.

### 4.3 Numerical example

We have implemented the formulas and algorithms for calculating $E_{k,j}$, as outlined above, in MATLAB. We have taken $\lambda_1 = 2$, $\lambda_2 = 1$ and $\mu = 4$. The results are displayed in Table 2.

One can see that for a constant $j$, as $k$ increases, the expected sojourn time $E_{k,j}$, tends to $j/4$ which is the expectation of the $Erlang(j, 4)$ random variable. This is not surprising, since as $k$ gets larger, the probability that repaired items will be released only to base 1 during the customer's sojourn time, is getting closer to 1.



## References

1. Cohen, J.W.: The Single Server Queue. North-Holland, Amsterdam (1982)
2. Cohen, J.W.: A two-queue, one-server model with priority for the longer queue. Queueing Syst. **2**, 261–283 (1987)
3. Cohen, J.W., Boxma, O.J.: Boundary Value Problems in Queueing System Analysis. North-Holland, Amsterdam (1983)
4. Daryanani, S., Miller, D.R.: Calculation of steady-state probabilities for repair facilities with multiple sources and dynamic return priorities. Oper. Res. **40**, S248–S256 (1992)
5. Flatto, L.: The longer queue model. Probab. Eng. Inf. Sci. **3**, 537–559 (1989)

6. http://eqworld.ipmnet.ru (2012)
7. van Houtum, G.J., Adan, I., van der Wal, J.: The symmetric longest queue system. Stoch. Models **13**, 105–120 (1997)
8. Wolff, R.W.: Poisson arrivals see time averages. Oper. Res. **30**, 223–231 (1982)
9. Zheng, Y.-S., Zipkin, P.: A queueing model to analyze the value of centralized inventory information. Oper. Res. **38**, 296–307 (1990)