J. Appl. Prob. **39,** 590–603 (2002)
Printed in Israel

THE SOJOURN TIME DISTRIBUTION IN AN INFINITE SERVER RESEQUENCING QUEUE WITH DEPENDENT INTERARRIVAL AND SERVICE TIMES

TIJS HUISMAN * AND
RICHARD J. BOUCHERIE,** *University of Amsterdam*

Abstract

We consider an infinite server resequencing queue, where arrivals are generated by jumps of a semi-Markov process and service times depend on the jumps of this process. The stationary distribution of the sojourn time, conditioned on the state of the semi-Markov process, is obtained both for the case of hyperexponential service times and for the case of a Markovian arrival process. For the general model, an accurate approximation is derived based on a discretisation of interarrival and service times.

Keywords: Resequencing queue; sojourn time distribution; semi-Markovian arrival process

AMS 2000 Subject Classification: Primary 60K25
Secondary 90B22

1. Introduction

Simultaneous processing of units that have to leave the system in the order of arrival is a feature of various systems with practical relevance. In communication networks (such as the Internet) messages are transferred as independent packets that must be received in the order of transmission, and in single lane traffic systems (such as single lane highways and unidirectional railway tracks), where overtaking is not possible, vehicles move simultaneously and at different speeds, but have to maintain the order of arrival; the constraint of order preservation also naturally arises in a parallel programming environment.

Under the assumption that the number of units that can be processed simultaneously is unlimited, such a system can be modeled as an infinite server resequencing queue, i.e. an infinite server queue where a customer can leave the system only if all earlier customers have left the system. Assuming independent and identically distributed (i.i.d.) interarrival and service times, the infinite server resequencing queue has been extensively studied; see [7] for the M/M/∞ resequencing queue, [4], [6], [8] for the M/G/∞ queue, and [2] for the G/M/∞ resequencing queue.

In modern applications, however, the assumption of independent interarrival and service times is not justified. In communication networks, dependence between consecutive interarrival times arises due to bursty arrivals, and dependence between consecutive service times occurs when the network is heavily used by other sources. In a traffic jam, vehicles pile up causing dependence between speeds of consecutive vehicles, which may be further increased by weather

Received 14 February 2000; revision received 10 June 2002.
* Current address: Railned BV, Department of Innovation, PO Box 2101, 3500 GC Utrecht, The Netherlands.
Email address: t.huisman@railned.nl
** Current address: Stochastic Operations Research Group, Faculty of Mathematical Sciences, University of Twente, PO Box 217, 7500 AE Enschede, The Netherlands.

conditions. Moreover, traffic lights before a stretch of road may lead to bursty arrivals. On railway sections, scheduling of train movements leads to dependence between interarrival times, and interarrival and service times.

This last consideration motivated the study [5] of the sojourn time in an infinite server resequencing queue with dependence between interarrival times and service times. This dependence is modeled by a special case of a Markovian arrival process (MAP), with service times dependent on the state of the MAP. The present note generalises this model to arrivals generated by jumps of a *semi*-Markov process, and service times dependent on the *jumps* of this process. Section 2 presents this model and preliminary results on the stationary distribution of the sojourn time, conditioned on the state of the arrival process. The sojourn time distribution is obtained explicitly for the case of hyperexponentially distributed service times in Section 3, and for the case of arrivals generated by an ordinary MAP in Section 4. Section 5 presents an approximation for the sojourn time distribution in models with general interarrival and service times by a discretisation of time.

2. The model and preliminary results

Consider a service facility with an infinite number of servers. Customers arriving at this facility are labeled $1, 2, \ldots$ in order of their arrival. Let $\mathcal{A}_n$ denote the interarrival time between customer n and customer $n+1$, and let $\mathcal{B}_n$ denote the service time of customer n.

Assume that the arrival of customer n is generated by the nth jump of a semi-Markov process $(\mathcal{X}(\tau),\ \tau > 0)$ with finite state space $\mathcal{S}_\mathcal{X} = \{1, 2, \ldots, J\}$, transition probabilities p_{ij} from state $i \in \mathcal{S}_\mathcal{X}$ to state $j \in \mathcal{S}_\mathcal{X}$, and sojourn time distribution $A_i(t)$ in state $i \in \mathcal{S}_\mathcal{X}$. Let $(\mathcal{X}_n,\ n = 1, 2, \ldots)$ be the embedded Markov chain immediately after jump epochs, i.e. $\mathcal{X}_n$ is the state of $(\mathcal{X}(\tau),\ \tau > 0)$ immediately after the nth jump. Then we have

$$\mathrm{P}(\mathcal{A}_n \le t \mid \mathcal{X}_n = i) = A_i(t), \qquad i \in \mathcal{S}_\mathcal{X},\ t > 0. \tag{2.1}$$

A jump from state i to state j generates a customer with service time distribution $B_{ij}(t)$, drawn independent of $A_i(t)$. The nth customer is generated by the jump of $(\mathcal{X}(\tau),\ \tau > 0)$ from state $\mathcal{X}_{n-1}$ to $\mathcal{X}_n$, so that

$$\mathrm{P}(\mathcal{B}_n \le t \mid \mathcal{X}_{n-1} = i, \mathcal{X}_n = j) = B_{ij}(t), \qquad i, j \in \mathcal{S}_\mathcal{X},\ t > 0. \tag{2.2}$$

Note that the interarrival time $\mathcal{A}_n$ between customers n and $n+1$ depends on the state $\mathcal{X}_n$ after the nth jump, whereas the service time $\mathcal{B}_n$ of customer n depends on the states $\mathcal{X}_{n-1}$ and $\mathcal{X}_n$. As a consequence, the service time $\mathcal{B}_n$ of customer n and the interarrival time $\mathcal{A}_n$ to the next customer $n+1$ are dependent only through the state of $(\mathcal{X}(\tau),\ \tau > 0)$. This dependence is natural in railroad systems for example, where the departure time of the next train will depend on the speed of the train that has just departed, see [5].

Upon the arrival of a customer, its service is started immediately, since the number of servers is unlimited. After completion of its service, however, customer n has to wait until customer $n-1$ has left the system. Thus, the time that customer n leaves the system is the later of the time that his service is completed and the time that customer $n-1$ leaves the system. The sojourn time $\mathcal{S}_n$ of customer n can thus be expressed recursively as

$$\begin{aligned} \mathcal{S}_1 &= \mathcal{B}_1, \\ \mathcal{S}_n &= \max(\mathcal{B}_n, \mathcal{S}_{n-1} - \mathcal{A}_{n-1}), \qquad n = 2, 3, \ldots. \end{aligned} \tag{2.3}$$

This recursion was obtained in [1], where it is also shown that $\mathcal{S}_n$ converges in distribution to a proper random variable when n tends to infinity if the sequences $\mathcal{A}_n$ and $\mathcal{B}_n$ are stationary and $\mathrm{E}(\mathcal{A}_n) > 0$. In our model, however, we see from (2.1) and (2.2) that the sequences $\mathcal{A}_n$ and $\mathcal{B}_n$ are generally not stationary: $\mathcal{A}_n$ and $\mathcal{B}_n$ depend on $\mathcal{X}_n$, and thus $\mathcal{A}_n$ and $\mathcal{B}_n$ are stationary sequences if and only if $(\mathcal{X}(\tau),\ \tau > 0)$ is stationary.

In the next lemma, we show that, independent of the convergence of the sequence S_n, the subsequences $\mathcal{S}_{n_1(j)}, \mathcal{S}_{n_2(j)}, \dots$ consisting of all $\mathcal{S}_n$ with $\mathcal{X}_n = j$ converge for all $j \in \mathcal{S}_\mathcal{X}$ to a proper random variable if there is at least one state $i \in \mathcal{S}_\mathcal{X}$ for which the interarrival time is not always zero.

Lemma 2.1. *If the service times $\mathcal{B}_1, \mathcal{B}_2, \dots$ are proper random variables, if the Markov chain $(\mathcal{X}_n,\ n = 1, 2, \dots)$ is irreducible, and if there exists an $i \in \{1, \dots, J\}$ such that $\int_0^\infty t\,\mathrm{d}A_i(t) > 0$, then for all $j = 1, \dots, J$, $\mathcal{S}_{n_k(j)}$ converges in distribution to a proper random variable when $k \to \infty$.*

Proof. We will first show that the subsequences can be expressed in a form similar to (2.3), with stationary interarrival and service times. Defining $\mathcal{S}_0 = 0$, $\mathcal{A}_0 = 0$, and $n_0(j) = 0$, we have, by repeated use of the recursion (2.3),

$$\mathcal{S}_{n_k(j)} = \max(\tilde{\mathcal{B}}_{n_k(j)}, \mathcal{S}_{n_{k-1}(j)} - \tilde{\mathcal{A}}_{n_{k-1}(j)}), \qquad k = 1, 2, \dots,$$

with

$$\tilde{\mathcal{B}}_{n_k(j)} = \max_{\{m=0,\dots,n_k(j)-n_{k-1}(j)-1\}} \biggl(\mathcal{B}_{n_k(j)-m} - \sum_{\ell=1}^{m} \mathcal{A}_{n_k(j)-\ell}\biggr),$$

and

$$\tilde{\mathcal{A}}_{n_{k-1}(j)} = \sum_{\ell=1}^{n_k(j)-n_{k-1}(j)} \mathcal{A}_{n_k(j)-\ell}.$$

Under the assumption that $(\mathcal{X}_n,\ n = 1, 2, \dots)$ is irreducible, the recurrence time of state j is almost surely finite. Hence, $\tilde{\mathcal{B}}_{n_k}(j)$ is the maximum of an almost surely finite number of random variables, which implies that $\tilde{\mathcal{B}}_{n_k(j)}$ itself is a proper random variable. Moreover, by the semi-Markov property of $\mathcal{X}(\tau)$, the sequence $\tilde{\mathcal{B}}_{n_1(j)}, \tilde{\mathcal{B}}_{n_2(j)}, \dots$ is a stationary sequence. Similarly, $\tilde{\mathcal{A}}_{n_k(j)}$ is a proper random variable, since it is the sum of an almost surely finite number of random variables, and the sequence $\tilde{\mathcal{A}}_{n_1(j)}, \tilde{\mathcal{A}}_{n_2(j)}, \dots$ is a stationary sequence. Moreover, $\mathrm{E}(\mathcal{A}_{n_k(j)}) > 0$, since by the irreducibility of $(\mathcal{X}_n,\ n = 1, 2, \dots)$ the state i is an element of the sequence $n_{k-1}(j), n_{k-1}(j) + 1, \dots, n_k(j) - 1$ with positive probability. Thus, $\mathcal{S}_{n_k(j)}$ can be expressed as the sojourn time in an infinite server resequencing queue with stationary service times and stationary interarrival times with positive expectation. By Lemma 3 of [1], $\mathcal{S}_{n_k(j)}$ converges in distribution to a proper random variable as $k \to \infty$.

Remark 2.1. Note that Lemma 2.1 allows the sequence $(\mathcal{A}_n, \mathcal{B}_n)$ to be periodic, which prohibits a direct use of the results of [1] via, for example, coupling. The main part of the proof above is a construction of a sequence of random variables that satisfies the assumptions of [1].

In the sequel, assume that the conditions of Lemma 2.1 are satisfied. Then

$$S_j(t) = \lim_{k\to\infty} \mathrm{P}(\mathcal{S}_{n_k(j)} \le t), \qquad j = 1, 2, \dots, J,\ t > 0,$$

the limiting distribution of $\mathcal{S}_{n_k(j)}$, is well defined. The following lemma shows that these distributions can be obtained by solving a system of integral equations. The formulation of this lemma uses the time-reversed transition probability from state $j \in \mathcal{S}_{\mathcal{X}}$ to state $i \in \mathcal{S}_{\mathcal{X}}$ of the Markov chain $(\mathcal{X}_n,\ n = 1, 2, \dots)$ given by

$$\tilde{p}_{ji} = \frac{\pi_i p_{ij}}{\pi_j},$$

with π_j the unique solution (up to normalisation) of $\pi_j = \sum_{i=1}^{J} \pi_i p_{ij},\ j = 1, \dots, J$.

Lemma 2.2. *The limiting distributions $S_j(t),\ j \in \mathcal{S}_{\mathcal{X}}$, are the unique solution of the following system of integral equations:*

$$S_j(t) = \sum_{i=1}^{J} \tilde{p}_{ji} B_{ij}(t) \int_{x=0}^{\infty} S_i(t+x)\,\mathrm{d}A_i(x), \qquad t > 0, \tag{2.4}$$

$$\lim_{t\to\infty} S_j(t) = 1. \tag{2.5}$$

Proof. By the recursion (2.3), we have

$$\mathrm{P}(\mathcal{S}_{n_k(j)} \le t) = \mathrm{P}(\mathcal{B}_{n_k(j)} \le t,\ \mathcal{S}_{n_k(j)-1} - \mathcal{A}_{n_k(j)-1} \le t).$$

Conditioning on all states $\mathcal{X}_{n_k(j)-1}$ from which a jump to j is possible gives

$$\mathrm{P}(\mathcal{S}_{n_k(j)} \le t) = \sum_{\{i\in\{1,\dots,J\}:p_{ij}>0\}} \mathrm{P}(\mathcal{B}_{n_k(j)} \le t,\ \mathcal{S}_{n_k(j)-1} - \mathcal{A}_{n_k(j)-1} \le t \mid \mathcal{X}_{n_k(j)-1} = i) \times \mathrm{P}(\mathcal{X}_{n_k(j)-1} = i).$$

As $\mathcal{X}_{n_k(j)} = j$ by the definition of $n_k(j)$, $\mathrm{P}(\mathcal{X}_{n_k(j)-1} = i)$ is the probability that state j is preceded by state i, i.e. the time-reversed transition probability from state j to state i of the Markov chain $(\mathcal{X}_n,\ n = 1, 2, \dots)$. Thus,

$$\mathrm{P}(\mathcal{S}_{n_k(j)} \le t) = \sum_{\{i\in\{1,\dots,J\}:p_{ij}>0\}} \mathrm{P}(\mathcal{B}_{n_k(j)} \le t,\ \mathcal{S}_{n_k(j)-1} - \mathcal{A}_{n_k(j)-1} \le t \mid \mathcal{X}_{n_k(j)-1} = i)\tilde{p}_{ji}.$$

From (2.2), the conditional distribution of $\mathcal{B}_{n_k(j)}$ on the states $\mathcal{X}_{n_k(j)-1} = i$ and $\mathcal{X}_{n_k(j)} = j$ is given by $B_{ij}(t)$, independent of the other random variables. Similarly, according to (2.1), the distribution of $\mathcal{A}_{n_k(j)-1}$ conditioned on $\mathcal{X}_{n_k(j)-1} = i$ is given by $A_i(t)$, also independent of the other random variables. Hence, conditioning on $\mathcal{A}_{n_k(j)-1}$,

$$\mathrm{P}(\mathcal{S}_{n_k(j)} \le t) = \sum_{\{i\in\{1,\dots,J\}:p_{ij}>0\}} B_{ij}(t) \int_{x=0}^{\infty} \mathrm{P}(\mathcal{S}_{n_k(j)-1} \le t + x \mid \mathcal{X}_{n_k(j)-1} = i)\,\mathrm{d}A_i(x)\tilde{p}_{ji}.$$

Taking the limit for $k \to \infty$, noting that interchanging the limit and the sum or the integral is allowed by Lebesgue's dominated convergence theorem, and observing that

$$\lim_{k\to\infty} \mathrm{P}(\mathcal{S}_{n_k(j)-1} \le t \mid \mathcal{X}_{n_k(j)-1} = i) = \lim_{k\to\infty} \mathrm{P}(\mathcal{S}_{n_k(i)} \le t) = S_i(t),$$

we get

$$S_j(t) = \sum_{i=1}^{J} \tilde{p}_{ji} B_{ij}(t) \int_{x=0}^{\infty} S_i(t+x)\,\mathrm{d}A_i(x).$$

The boundary condition is an immediate consequence of the fact that $S_j(t)$ is a probability distribution.

It remains to prove the uniqueness of the solution. Let $S'_j(t)$ and $S''_j(t)$ be solutions of the system (2.4) and (2.5), and let $D_j(t) = S'_j(t) - S''_j(t)$. Then $D_j(t)$ solves (2.4), and $\lim_{t\to\infty} D_j(t) = 0$. Suppose that $\max_{j\in\mathcal{S}_X} |D_j(t')| = \varepsilon > 0$ for some $t' > 0$. Since $\lim_{t\to\infty} D_j(t) = 0$ for all $j \in \mathcal{S}_X$, there is an $M' > 0$ such that $\max_{j\in\mathcal{S}_X} |D_j(t)| < \varepsilon$ for $t > M'$. Let M be the minimal M' for which this holds, i.e. $\max_{j\in\mathcal{S}_X} |D_j(M)| \geq \varepsilon$ and $\max_{j\in\mathcal{S}_X} |D_j(t)| < \varepsilon$ for all $t > M$. As $D_j(t)$ solves (2.4), we have, for all $j \in \mathcal{S}_X$,

$$\begin{aligned}
|D_j(M)| &\leq \sum_{i=1}^{J} \tilde{p}_{ji} B_{ij}(M) \int_{x=0}^{\infty} |D_i(M+x)| \,\mathrm{d}A_i(x) \\
&\leq \sum_{i=1}^{J} \tilde{p}_{ji} \int_{x=0}^{\infty} |D_i(M+x)| \,\mathrm{d}A_i(x) \\
&= \sum_{i=1}^{J} \tilde{p}_{ji} \biggl(|D_i(M)| A_i(0) + \int_{x=0+}^{\infty} |D_i(M+x)| \,\mathrm{d}A_i(x) \biggr). \qquad (2.6)
\end{aligned}$$

Hence, if $k \in \arg\max_{j\in\mathcal{S}_X} |D_j(M)|$ and $A_i(0) = 1$ for all $i \in \mathcal{S}_X$ with $\tilde{p}_{ki} > 0$, then

$$|D_k(M)| \leq \sum_{i=1}^{J} \tilde{p}_{ki} |D_i(M)| \leq \sum_{i=1}^{J} \tilde{p}_{ki} |D_k(M)| = |D_k(M)|,$$

implying that $i \in \arg\max_{j\in\mathcal{S}_X} |D_j(M)|$ for all $i \in \mathcal{S}_X$ with $\tilde{p}_{ki} > 0$. Combining this observation with the assumptions of Lemma 2.1, we conclude that there is an $\ell \in \arg\max_{j\in\mathcal{S}_X} |D_j(M)|$, such that there is an $i \in \mathcal{S}_X$ for which $A_i(0) < 1$ and $p_{\ell i} > 0$. For this ℓ we have, using (2.6) and the definition of M,

$$\begin{aligned}
|D_\ell(M)| &\leq \sum_{i=1}^{J} \tilde{p}_{\ell i} \biggl(|D_i(M)| A_i(0) + \int_{x=0+}^{\infty} |D_i(M+x)| \,\mathrm{d}A_i(x) \biggr) \\
&\leq \sum_{i=1}^{J} \tilde{p}_{\ell i} \biggl(|D_\ell(M)| A_i(0) + \int_{x=0+}^{\infty} |D_\ell(M+x)| \,\mathrm{d}A_i(x) \biggr) \\
&< \sum_{i=1}^{J} \tilde{p}_{\ell i} |D_\ell(M)| \biggl(A_i(0) + \int_{x=0+}^{\infty} \mathrm{d}A_i(x) \biggr) = |D_\ell(M)|.
\end{aligned}$$

This contradicts the assumption that $\max_{j\in\mathcal{S}_X} |D_j(t')| > 0$. Hence, $D_j(t) = 0$ for all $j \in \mathcal{S}_X$ and all t, and the solution of the system (2.4) and (2.5) is unique.

The integral equation (2.4) can be written as

$$S_j(t) = \sum_{i=1}^{J} \tilde{p}_{ji} \int_{x=t}^{\infty} B_{ij}(t) S_i(x) \,\mathrm{d}A_i(t-x),$$

and thus is of Volterra type (with the special property that the fixed integration bound is infinite). Various results and methods for solving this type of integral equation are known (cf. [9]), but generally require finite integration bounds or additional assumptions on the form of the kernel that are not satisfied in our case. In applied probability, a common method of solving integral equations is based on Laplace–Stieltjes transforms. For the case that $J = 1$, interarrival times are exponential, and service times are hyperexponential, this method was used in [1] to solve (2.4) for the Laplace–Stieltjes transform of the sojourn time.

In the remainder of the current paper, we will provide some alternative methods for solving the system (2.4) and (2.5). These methods yield direct expressions for the sojourn time distribution under various assumptions on the distributions $A_i(t)$ and $B_{ij}(t)$.

Remark 2.2. An immediate generalisation of the model in this section is provided by assuming that not only the service time distributions, but also the interarrival time distributions, depend on the state of $(\mathcal{X}(\tau),\ \tau > 0)$ before, as well as after the jump, i.e. $A_{ij}(t) := \mathrm{P}(\mathcal{A}_n \le t \mid \mathcal{X}_n = i, \mathcal{X}_{n+1} = j)$. With the exception of Section 4 below, the results in this paper are readily generalised to this more complicated form of dependence.

3. Service times with hyperexponential distribution

In this section, we consider the case that the service time distributions $B_{ij}(t)$ are hyperexponential, i.e. $B_{ij}(t)$ is a mixture of exponential distributions. This class of distributions is of interest for communication systems (see, for example, [2] for a study of the GI/M/∞ queue in an application to virtual circuits). For transportation systems, hyperexponential service times are generally unrealistic.

We assume that the intensities of the exponential distributions in the mixture are rational. Then time can be scaled such that $B_{ij}(t)$ can be written as

$$B_{ij}(t) = \sum_{k=0}^{K} b_{ij}(k)\mathrm{e}^{-kt}, \qquad i, j = 1, \dots, N,\ t > 0, \tag{3.1}$$

with $b_{ij}(0) = 1$, $-1 \le b_{ij}(k) \le 1$ for $k > 0$, and $\sum_{k=1}^{K} b_{ij}(k) = -1$. To obtain a solution of the system (2.4) and (2.5), observe that distributions of the form (3.1) have two interesting, and useful, properties. First, the distribution of the maximum of two independent random variables with such distributions, say $\sum_{k=0}^{K} b'(k)\mathrm{e}^{-kt}$ and $\sum_{k=0}^{K} b''(k)\mathrm{e}^{-kt}$, is given by

$$\sum_{k=0}^{K} b'(k)\mathrm{e}^{-kt} \sum_{\ell=0}^{K} b''(\ell)\mathrm{e}^{-\ell t} = \sum_{k=0}^{\infty}\sum_{\ell=0}^{k} b'(k-\ell)b''(\ell)\mathrm{e}^{-kt},$$

with $b'(k) = 0$ and $b''(k) = 0$ for $k > K$. Thus, the maximum is distributed similarly to (3.1), but with $b_{ij}(k) \in \mathbb{R}$. The same observation holds for subtracting a nonnegative random variable, say with distribution $A(x)$, from a random variable with distribution (3.1): the resulting distribution is given by

$$\int_{x=0}^{\infty} \sum_{k=0}^{K} b_{ij}(k)\mathrm{e}^{-k(t+x)}\,\mathrm{d}A(x) = \sum_{k=0}^{\infty} b_{ij}(k)\mathrm{e}^{-kt}\alpha(k),$$

with $\alpha(k) = \int_{x=0}^{\infty} \mathrm{e}^{-kx}\,\mathrm{d}A(x)$, the Laplace–Stieltjes transform of $A(x)$.

As the recursive evaluation (2.3) of the sojourn time requires these two operations only, we may expect the sojourn time to have a distribution similar to (3.1). The following theorem shows that this intuition is correct. To state the theorem, we need some additional notation. Define the $J \times J$ matrices $\tilde{\boldsymbol{A}}(k)$ and $\tilde{\boldsymbol{B}}(k)$ as

$$\tilde{\boldsymbol{A}}(k) = \operatorname{diag}(\alpha_1(k), \dots, \alpha_J(k)), \qquad k = 0, 1, 2, \dots,$$
$$\tilde{\boldsymbol{B}}(k) = \begin{cases} (\tilde{p}_{ji} b_{ij}(k))_{i,j=1,\dots,J} & \text{for } k = 0, \dots, K, \\ 0 & \text{for } k > K, \end{cases}$$

with $\alpha_i(s) = \int_{t=0}^{\infty} \mathrm{e}^{-st}\, \mathrm{d}A_i(t)$, the Laplace–Stieltjes transform of $A_i(t)$. Let the J-dimensional row vectors $\boldsymbol{c}(k)$ be recursively defined by

$$\boldsymbol{c}(0) = \boldsymbol{e},$$
$$\boldsymbol{c}(k) = \sum_{\ell=0}^{k-1} \boldsymbol{c}(\ell)\tilde{\boldsymbol{A}}(\ell)\tilde{\boldsymbol{B}}(k-\ell)(\boldsymbol{I} - \tilde{\boldsymbol{A}}(k)\tilde{\boldsymbol{B}}(0))^{-1}, \qquad k = 1, 2, \dots, \tag{3.2}$$

with $\boldsymbol{e}$ the J-dimensional row vector of 1s, and $\boldsymbol{I}$ the $J \times J$ identity matrix.

Theorem 3.1. *Assume that the service time distributions $B_{ij}(t)$ are given by (3.1), and assume that $\int_{t=0}^{\infty} t\, \mathrm{d}A_i(t) > 0$ for all $i = 1, \dots, J$. Then the series $\sum_{k=0}^{\infty} \boldsymbol{c}(k)$ is absolutely convergent and the sojourn time distribution $\boldsymbol{S}(t) = (S_1(t), \dots, S_J(t))$ is given by*

$$\boldsymbol{S}(t) = \sum_{k=0}^{\infty} \boldsymbol{c}(k)\mathrm{e}^{-kt}, \qquad t > 0. \tag{3.3}$$

Proof. Let $\|\cdot\|$ be the matrix norm induced by the l_∞-norm on row vectors:

$$\|A\| = \max_{\{j=1,\dots,J\}} \sum_{i=1}^{J} |a_{ij}|.$$

Then $\|\tilde{\boldsymbol{A}}(k)\tilde{\boldsymbol{B}}(0)\| < 1$, since the conditions on $A_i(t)$ imply that $\alpha_i(k) < 1$ for $k > 0$ and the column sums of $\tilde{\boldsymbol{B}}(0)$ are 1. Hence, $I - \tilde{\boldsymbol{A}}(k)\tilde{\boldsymbol{B}}(0)$ is invertible and $\boldsymbol{c}(k)$ is well defined.

To prove the absolute convergence of the series $\sum_{k=0}^{\infty} \boldsymbol{c}(k)$, we first rewrite (3.2) as

$$\boldsymbol{c}(k) = \sum_{\ell=0}^{k} \boldsymbol{c}(\ell)\tilde{\boldsymbol{A}}(\ell)\tilde{\boldsymbol{B}}(k-\ell).$$

Since $\tilde{\boldsymbol{B}}(k-\ell) = 0$ for $k - \ell > K$, we have for $k \geq K$

$$\boldsymbol{c}(k) = \sum_{\ell=k-K}^{k} \boldsymbol{c}(\ell)\tilde{\boldsymbol{A}}(\ell)\tilde{\boldsymbol{B}}(k-\ell).$$

Since $|b_{ij}(k)| < 1$ and $\sum_{i=1}^{J} \tilde{p}_{ji} = 1$, it follows that $\|\tilde{\boldsymbol{B}}(k)\| \leq 1$, and, hence, by the triangle inequality and the Schwartz inequality,

$$\|\boldsymbol{c}(k)\| \leq \sum_{\ell=k-K}^{k} \|\boldsymbol{c}(\ell)\|\|\tilde{\boldsymbol{A}}(\ell)\|.$$

By the assumptions on $A_i(t)$, $\alpha_i(s)$ is strictly decreasing in s for all $i = 1, \dots, J$, implying that there exists a finite $k' > K$ such that $\|\tilde{\boldsymbol{A}}(k-K)\| < 1/(K+2)$ for $k \geq k'$. Hence, we have, for all $M > k'$,

$$\begin{aligned}\sum_{k=k'}^{M} \|\boldsymbol{c}(k)\| &\leq \sum_{k=k'}^{M} \sum_{\ell=k-K}^{k} \|\boldsymbol{c}(\ell)\| \|\tilde{\boldsymbol{A}}(\ell)\| \\ &\leq \sum_{k=k'}^{M} \sum_{\ell=k-K}^{k} \|\boldsymbol{c}(\ell)\| \frac{1}{K+2} \\ &= \sum_{\ell=k'-K}^{k'-1} \|\boldsymbol{c}(\ell)\| \sum_{k=k'}^{\ell+K} \frac{1}{K+2} + \sum_{\ell=k'}^{M} \|\boldsymbol{c}(\ell)\| \sum_{k=\ell}^{\min(\ell+K,M)} \frac{1}{K+2} \\ &\leq \frac{K+1}{K+2} \sum_{\ell=k'-K}^{M} \|\boldsymbol{c}(\ell)\|.\end{aligned}$$

Thus, for all $M > k'$,

$$\sum_{k=k'}^{M} \|\boldsymbol{c}(k)\| \leq (K+1) \sum_{\ell=k'-K}^{k'-1} \|\boldsymbol{c}(\ell)\| < \infty,$$

which implies that $\sum_{k=k'}^{\infty} \|\boldsymbol{c}(k)\|$ is finite.

We are now ready to prove that the proposed form for $S_j(t)$ satisfies the system (2.4) and (2.5). The boundary condition is almost immediate: as $\sum_{k=0}^{\infty} \boldsymbol{c}(k)\mathrm{e}^{-kt}$ is absolutely convergent, interchanging limit and summation is allowed, and therefore

$$\lim_{t\to\infty} \boldsymbol{S}(t) = \boldsymbol{c}(0) + \sum_{k=1}^{\infty} \lim_{t\to\infty} \boldsymbol{c}(k)\mathrm{e}^{-kt} = \boldsymbol{c}(0) = \boldsymbol{e}.$$

Substituting (3.1) and (3.3) into (2.4) yields

$$\sum_{k=0}^{\infty} c_j(k)\mathrm{e}^{-kt} = \sum_{i=1}^{J} \tilde{p}_{ji} \sum_{\ell=0}^{\infty} b_{ij}(\ell)\mathrm{e}^{-\ell t} \int_{x=0}^{\infty} \sum_{m=0}^{\infty} c_i(m)\mathrm{e}^{-m(t+x)}\,\mathrm{d}A_i(x).$$

Using the absolute convergence and the definition of $\alpha_i(s)$,

$$\begin{aligned}\sum_{k=0}^{\infty} c_j(k)\mathrm{e}^{-kt} &= \sum_{i=1}^{J} \tilde{p}_{ji} \sum_{\ell=0}^{\infty} b_{ij}(\ell)\mathrm{e}^{-\ell t} \sum_{m=0}^{\infty} c_i(m)\mathrm{e}^{-mt}\alpha_i(m) \\ &= \sum_{i=1}^{J} \tilde{p}_{ji} \sum_{\ell=0}^{\infty} b_{ij}(\ell) \sum_{k=\ell}^{\infty} \mathrm{e}^{-kt} c_i(k-\ell)\alpha_i(k-\ell) \\ &= \sum_{k=0}^{\infty} \mathrm{e}^{-kt} \sum_{i=1}^{J} \tilde{p}_{ji} \sum_{\ell=0}^{k} b_{ij}(k-\ell) c_i(\ell)\alpha_i(\ell).\end{aligned}$$

By the definition of $\tilde{\boldsymbol{A}}(k)$ and $\tilde{\boldsymbol{B}}(k)$, the above equation can be written in matrix notation as

$$\sum_{k=0}^{\infty} \boldsymbol{c}(k)\mathrm{e}^{-kt} = \sum_{k=0}^{\infty} \mathrm{e}^{-kt} \sum_{\ell=0}^{k} \boldsymbol{c}(\ell)\tilde{\boldsymbol{A}}(\ell)\tilde{\boldsymbol{B}}(k-\ell). \tag{3.4}$$

By the recursive definition of $\boldsymbol{c}(k)$, $\boldsymbol{c}(k) = \sum_{\ell=0}^{k} \boldsymbol{c}(\ell)\tilde{\boldsymbol{A}}(\ell)\tilde{\boldsymbol{B}}(k-\ell)$, and therefore (3.4) is satisfied. Hence, the proposed form (3.3) satisfies the system (2.4) and (2.5), and, by Lemma 2.2, the proof is complete.

For the case of exponential service times, identically distributed in each state of $(\mathcal{X}(\tau), \tau > 0)$, the formulae for $\boldsymbol{c}(k)$ simplify considerably. Assuming that time is scaled such that the mean service time is 1, we then have $\tilde{\boldsymbol{B}}(1) = -\tilde{\boldsymbol{B}}(0)$ and $\tilde{\boldsymbol{B}}(k) = 0$ for $k > 1$. Substitution in the definition (3.2) of $\boldsymbol{c}(k)$ yields

$$\boldsymbol{c}(k) = -\boldsymbol{c}(k-1)\tilde{\boldsymbol{A}}(k-1)\tilde{\boldsymbol{B}}(0)(I - \tilde{\boldsymbol{A}}(k)\tilde{\boldsymbol{B}}(0))^{-1},$$

and thus $\boldsymbol{S}(t)$ can be written in closed form as

$$\boldsymbol{S}(t) = \boldsymbol{e}\sum_{k=0}^{\infty}(-1)^k \prod_{\ell=1}^{k} \tilde{\boldsymbol{A}}(\ell-1)\tilde{\boldsymbol{B}}(0)(I - \tilde{\boldsymbol{A}}(\ell)\tilde{\boldsymbol{B}}(0))^{-1}\mathrm{e}^{-kt}, \qquad t > 0. \tag{3.5}$$

When the state space of the arrival process consists of one state only, (3.5) gives the sojourn time distribution in the GI/M/∞ resequencing queue:

$$S(t) = \sum_{k=0}^{\infty}(-1)^k \prod_{\ell=1}^{k} \frac{\alpha(\ell-1)}{1-\alpha(\ell)}\mathrm{e}^{-kt}, \qquad t > 0, \tag{3.6}$$

where $S(t) = S_1(t)$ and $\alpha(s) = \alpha_1(s)$.

Remark 3.1. In Section 4 of [2], the sojourn time in the GI/M/∞ queue is obtained as

$$S(t) = \sum_{n=0}^{\infty}\sum_{k=n}^{\infty}(-1)^{k-n}\binom{k}{n}\prod_{\ell=1}^{k}\frac{\alpha(\ell)}{1-\alpha(\ell)}(1-\mathrm{e}^{-t})^{n+1}.$$

It can readily be shown that this expression is equivalent to (3.6). Theorem 3.1 thus generalises the results of [2] to hyperexponential service times and dependence between interarrival and service times modeled by a semi-Markov process.

The sojourn time in the M/M/∞ resequencing queue with arrival rate λ follows from (3.6) by substitution of $\alpha(\ell) = \lambda/(\lambda+\ell)$. For $k > 1$,

$$\prod_{\ell=1}^{k}\frac{\lambda}{\lambda+\ell-1}\left(1-\frac{\lambda}{\lambda+\ell}\right)^{-1} = \frac{\lambda^k}{k!} + \frac{\lambda^{k-1}}{(k-1)!},$$

and hence

$$S(t) = 1 + \sum_{k=1}^{\infty}(-1)^k\left(\frac{\lambda^k}{k!} + \frac{\lambda^{k-1}}{(k-1)!}\right)\mathrm{e}^{-kt} = (1-\mathrm{e}^{-t})\mathrm{e}^{-\lambda\mathrm{e}^{-t}}, \qquad t > 0. \tag{3.7}$$

This result can also be explained by probabilistic arguments. If interarrival and service times are i.i.d. sequences, and independent of each other, the service time is independent of the remaining sojourn time of the preceding customer. By (2.3), the sojourn time is then given by the product of the service time distribution $1 - \mathrm{e}^{-t}$ and the distribution of the remaining sojourn time R of the preceding customer. The remaining sojourn time R is equal to the maximum of the remaining service times of all customers that are still in service. By the memoryless property of the exponential distribution, these remaining service times are exponential with mean equal to 1. Observing that the distribution of the number K of customers that have not completed their service equals the distribution of the number of customers in an ordinary M/M/∞ queue, i.e. K is Poisson distributed with parameter λ, by conditioning on K, we get

$$\sum_{k=0}^{\infty} \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!} (1 - \mathrm{e}^{-t})^k = \mathrm{e}^{-\lambda \mathrm{e}^{-t}}.$$

4. Arrivals generated by an ordinary Markov process

We will now consider the case that $A_i(t)$ is an exponential distribution with mean $1/\alpha_i$:

$$A_i(t) = 1 - \mathrm{e}^{-\alpha_i t}. \tag{4.1}$$

The process $(\mathcal{X}(\tau),\ \tau > 0)$ is then a Markov process, and our model reduces to a MAP/G/∞ resequencing queue with service times dependent on the jumps of the MAP. To derive a solution of (2.4) and (2.5) for this model, we consider the distribution

$$R_j(t) = \int_{x=0}^{\infty} S_j(t + x)\, \mathrm{d}A_j(x), \qquad j = 1, 2, \ldots, J,\ t > 0, \tag{4.2}$$

the limiting distribution of the remaining service time immediately before a jump, conditioned on the event that the state of $(\mathcal{X}(\tau),\ \tau > 0)$ is j. By (2.4), we then have

$$S_j(t) = \sum_{i=1}^{J} \tilde{p}_{ji} B_{ij}(t) R_i(t), \qquad j = 1, 2, \ldots, J. \tag{4.3}$$

Combining (4.2) and (4.3) shows that the limiting distributions $R_j(t),\ j = 1, 2, \ldots, J$, of the remaining service time immediately before a jump, conditioned on the event that the state of $(\mathcal{X}(\tau),\ \tau > 0)$ is j, is the solution of

$$R_j(t) = \sum_{i=1}^{J} \tilde{p}_{ji} \int_{x=0}^{\infty} B_{ij}(t + x) R_i(t + x)\, \mathrm{d}A_j(x), \qquad t > 0, \tag{4.4}$$

$$\lim_{t\to\infty} R_j(t) = 1. \tag{4.5}$$

Again, the boundary condition is a consequence of $R_j(t)$ being a distribution function.

The following theorem shows that the system (2.4) of integral equations can be translated into a set of linear differential equations of the form $(\mathrm{d}/\mathrm{d}t)\boldsymbol{Y}(t) = -\boldsymbol{Y}(t)\boldsymbol{Q}(t),\ t \in (0, \infty)$. The fundamental set of solutions of such a system is formally given by the columns of the matrix $\prod_t^{\infty} \mathrm{e}^{\boldsymbol{Q}(u)\,\mathrm{d}u}$, called the *product integral* of $\boldsymbol{Q}(u)$ over $[t, \infty)$. For a piecewise constant matrix $\boldsymbol{Q}(t)$ such that there exist $t_0, \dots, t_n$ with $0 = t_0 < t_1 < \cdots < t_n < t_{n+1} = \infty$ and matrices $\boldsymbol{Q}_i$ such that $\boldsymbol{Q}(t) = \boldsymbol{Q}_i$ for $t \in [t_i, t_{i+1}),\ i = 1, 2, \dots, n$, say, the product integral is

$$\prod_t^{\infty} \mathrm{e}^{\boldsymbol{Q}(u)\,\mathrm{d}u} = \mathrm{e}^{\boldsymbol{Q}_k(t_k - t)} \mathrm{e}^{\boldsymbol{Q}_{k+1}(t_{k+1}-t_k)} \mathrm{e}^{\boldsymbol{Q}_{k+2}(t_{k+2}-t_{k+1})} \cdots \mathrm{e}^{\boldsymbol{Q}_{n-1}(t_n - t_{n-1})} \mathrm{e}^{-\boldsymbol{Q}_n t_n}, \qquad t \in [t_{k-1}, t_k). \tag{4.6}$$

Since the matrices $\boldsymbol{Q}_j$ and $\boldsymbol{Q}_{j+1}$ generally do not commute, the order of the terms in the product (4.6) cannot be changed. For general $\boldsymbol{Q}(t)$, the product integral is defined by analogy with (4.6) as the limiting result for $n \to \infty$ and $t_k - t_{k-1} \to 0$. For a more detailed treatment of product integration, the reader is referred to [3].

For the formulation of the theorem, we define the $J \times J$ matrices $\tilde{\boldsymbol{A}}$ and $\tilde{\boldsymbol{B}}(t)$ as

$$\tilde{\boldsymbol{A}} = \operatorname{diag}(\alpha_1, \dots, \alpha_J),$$
$$\tilde{\boldsymbol{B}}(t) = (\tilde{p}_{ji} B_{ij}(t))_{i,j=1,\dots,J}.$$

Theorem 4.1. *Assume that $A_i(t)$ is given by (4.1). Then $\boldsymbol{R}(t) = (R_1(t), \dots, R_J(t))$ is given by*

$$\boldsymbol{R}(t) = \boldsymbol{e}\biggl(\prod_t^{\infty} \mathrm{e}^{(\tilde{\boldsymbol{B}}(u) - \boldsymbol{I})\tilde{\boldsymbol{A}}\,\mathrm{d}u}\biggr), \qquad t > 0, \tag{4.7}$$

and the sojourn time distribution $\boldsymbol{S}(t) = (S_1(t), \dots, S_J(t))$ is given by

$$\boldsymbol{S}(t) = \boldsymbol{R}(t)\tilde{\boldsymbol{B}}(t), \qquad t > 0. \tag{4.8}$$

Proof. Differentiation of (4.4) with respect to t yields

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t} R_j(t) &= \sum_{i=1}^{J} \tilde{p}_{ji} \int_{x=0}^{\infty} \frac{\mathrm{d}}{\mathrm{d}t}(B_{ij}(t+x)R_i(t+x))\alpha_j \mathrm{e}^{-\alpha_j x}\,\mathrm{d}x \\ &= \sum_{i=1}^{J} \tilde{p}_{ji} \int_{x=0}^{\infty} \frac{\mathrm{d}}{\mathrm{d}x}(B_{ij}(t+x)R_i(t+x))\alpha_j \mathrm{e}^{-\alpha_j x}\,\mathrm{d}x.\end{aligned}$$

By partial integration,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t} R_j(t) &= \biggl[\sum_{i=1}^{J} \tilde{p}_{ji} B_{ij}(t+x)R_i(t+x)\alpha_j \mathrm{e}^{-\alpha_j x}\biggr]_{x=0}^{\infty} \\ &\quad + \sum_{i=1}^{J} \tilde{p}_{ji} \int_{x=0}^{\infty} B_{ij}(t+x)R_i(t+x)\alpha_j^2 \mathrm{e}^{-\alpha_j x}\,\mathrm{d}x \\ &= -\alpha_j \sum_{i=1}^{J} \tilde{p}_{ji} B_{ij}(t) R_i(t) + \alpha_j R_j(t).\end{aligned}$$

Rewriting this system in matrix notation using the definitions of $\tilde{\boldsymbol{A}}$ and $\tilde{\boldsymbol{B}}(t)$, we arrive at the following system:

$$\frac{\mathrm{d}}{\mathrm{d}t} \boldsymbol{R}(t) = -\boldsymbol{R}(t)(\tilde{\boldsymbol{B}}(t) - \boldsymbol{I})\tilde{\boldsymbol{A}}.$$

Noting that (4.7) is the unique solution of this system with boundary condition $\lim_{t\to\infty} \boldsymbol{R}(t) = \boldsymbol{e}$ proves the first statement of the theorem. The equality (4.8) is an immediate consequence of (4.3).

Remark 4.1. Theorem 4.1 generalises the results of [5] to general Markovian arrival processes and to service times that may depend on the jumps of $\mathcal{X}(\tau)$.

For the computation of the product integral, several numerical methods are available. In some special cases, however, the product integral can be obtained explicitly.

If the state space of the arrival process consists of one state only, the matrices $\tilde{\boldsymbol{A}}$ and $\tilde{\boldsymbol{B}}(t)$ reduce to scalars, and the product integral $R(t)$ is given explicitly by

$$R(t) = \exp\left(-\alpha \int_t^{\infty} (1 - B(u))\,\mathrm{d}u\right),$$

where $R(t) = R_1(t)$ and $B(t) = B_{11}(t)$. By (4.8), we then have

$$S(t) = B(t) \exp\left(-\alpha \int_t^{\infty} (1 - B(u))\,\mathrm{d}u\right), \qquad t > 0, \tag{4.9}$$

corresponding to the results obtained in [8, Section 3.1.1], for the M/G/∞ resequencing queue. Similar to the sojourn time distribution (3.7) in the M/M/∞ resequencing queue, (4.9) can be explained by probabilistic arguments. Following the derivation of (3.7), observing that the number of customers in the M/G/∞ queue depends on the service time distribution only through its mean, and noting that the distribution of the remaining service time is given by

$$\frac{1}{\mathrm{E}(B)} \int_{u=0}^{t} (1 - B(u))\,\mathrm{d}u,$$

with $E(B) = \int_{u=0}^{\infty} (1 - B(u))\,\mathrm{d}u$, the mean of the service time, we get

$$S(t) = \sum_{k=0}^{\infty} \mathrm{e}^{-\alpha \mathrm{E}(B)} \frac{(\alpha \mathrm{E}(B))^k}{k!} \left(\frac{1}{\mathrm{E}(B)} \int_{u=0}^{t} (1 - B(u))\,\mathrm{d}u\right)^k,$$

resulting in (4.9).

Another case for which $\boldsymbol{R}(t)$ can be obtained explicitly is the case of discrete service time distributions. The matrix $\tilde{\boldsymbol{B}}(t)$ is then piecewise constant, since its coefficients depend on t through the service time distributions $B_{ij}(t)$ only. Denoting the intervals on which $\tilde{\boldsymbol{B}}(t)$ is constant by $[t_k, t_{k+1})$, $k = 1, 2, \ldots,$ the product integral of $(\tilde{\boldsymbol{B}}(u) - \boldsymbol{I})\tilde{\boldsymbol{A}}$ is then given in analogy with (4.6) by

$$\boldsymbol{R}(t) = \boldsymbol{e} \prod_{\ell=k}^{\infty} \mathrm{e}^{(\tilde{\boldsymbol{B}}(t_\ell) - \boldsymbol{I})\tilde{\boldsymbol{A}}(t_{\ell+1} - t_\ell)} \mathrm{e}^{(\tilde{\boldsymbol{B}}(t_{k-1}) - \boldsymbol{I})\tilde{\boldsymbol{A}}(t_k - t)} \quad \text{for } t_{k-1} \le t < t_k,\ k = 1, 2, \ldots.$$

5. Discretisation

Observing that the right-hand side of the integral equation (2.4) contains values $S_i(u)$ for $u \ge t$ only, we might expect that a discretisation of the distributions $A_i(t)$ allows a recursive computation of $S_j(t)$. To get a starting point for this recursion, observe that when all service times are bounded by some value T, the sojourn time will never exceed T. The following theorem makes these considerations precise.

Theorem 5.1. *Let $A_i(t)$ and $B_{ij}(t)$ be discrete distribution functions on $t = 0, 1, 2, \dots$, and let $a_i(t) = A_i(t) - A_i(t-1)$ be the probability mass function of $A_i(t)$. If there exists a finite T such that $B_{ij}(t) = 1$ for $t \geq T$, then*

$$\boldsymbol{S}(t) = \begin{cases} \boldsymbol{e}, & t \geq T, \\ \left(\boldsymbol{e}(\boldsymbol{I} - \tilde{\boldsymbol{A}}(0)) + \sum_{x=1}^{T-t-1} (\boldsymbol{S}(t+x) - \boldsymbol{e})\tilde{\boldsymbol{A}}(x)\right)\tilde{\boldsymbol{B}}(t)(\boldsymbol{I} - \tilde{\boldsymbol{A}}(0)\tilde{\boldsymbol{B}}(t))^{-1}, & t \leq T, \end{cases} \tag{5.1}$$

with $\tilde{\boldsymbol{A}}(x) = \operatorname{diag}(a_1(x), \dots, a_J(x))$ and $\tilde{\boldsymbol{B}}(t)$ the $J \times J$ matrix with entries $\tilde{p}_{ji} B_{ij}(t)$.

Proof. First note that (5.1) is well defined: as $a_j(x)$ is a probability, $a_j(x) \leq 1$ for all $j = 1, \dots, J$. Furthermore, by the conditions of Lemma 2.1, there exists an $i \in \mathcal{S}_{\mathcal{X}}$ with $a_i(0) < 1$. Therefore $\|\tilde{\boldsymbol{A}}(0)\tilde{\boldsymbol{B}}(t)\| < 1$, and $\boldsymbol{I} - \tilde{\boldsymbol{A}}(0)\tilde{\boldsymbol{B}}(t)$ is invertible.

We will now prove that (5.1) solves the system (2.4) and (2.5). Obviously, the boundary condition (2.5) of this system is satisfied. Under the assumption of discrete interarrival times, the integral equation (2.4) is given by

$$S_j(t) = \sum_{i=1}^{J} \tilde{p}_{ji} B_{ij}(t) \sum_{x=0}^{\infty} S_i(t+x) a_i(x).$$

For $t \geq T$, it is easily verified that $\boldsymbol{S}(t) = \boldsymbol{e}$ solves this equation. Using $S_j(t) = 1$ for $t \geq T$, we get, for $t < T$,

$$\begin{aligned} S_j(t) &= \sum_{i=1}^{J} \tilde{p}_{ji} B_{ij}(t) \left(\sum_{x=0}^{T-t-1} S_i(t+x) a_i(x) + \sum_{x=T-t}^{\infty} a_i(x) \right) \\ &= \sum_{i=1}^{J} \tilde{p}_{ji} B_{ij}(t) \left(1 + \sum_{x=0}^{T-t-1} (S_i(t+x) - 1) a_i(x) \right). \end{aligned}$$

Rewriting this equation in matrix notation and inserting the definitions of $\tilde{\boldsymbol{A}}$ and $\tilde{\boldsymbol{B}}(t)$ gives

$$\boldsymbol{S}(t) = \left(\boldsymbol{e} + \sum_{x=0}^{T-t-1} (\boldsymbol{S}(t+x) - \boldsymbol{e}) \tilde{\boldsymbol{A}}(x) \right) \tilde{\boldsymbol{B}}(t),$$

which is equivalent to

$$\boldsymbol{S}(t)(\boldsymbol{I} - \tilde{\boldsymbol{A}}(0)\tilde{\boldsymbol{B}}(t)) = \left(\boldsymbol{e}(I - \tilde{\boldsymbol{A}}(0)) + \sum_{x=1}^{T-t-1} (\boldsymbol{S}(t+x) - \boldsymbol{e}) \tilde{\boldsymbol{A}}(x) \right) \tilde{\boldsymbol{B}}(t).$$

As this equation coincides with (5.1), $S_j(t)$ solves the system (2.4) and (2.5), and application of Lemma 2.2 completes the proof.

Computing the inverse of $\boldsymbol{I} - \tilde{\boldsymbol{A}}(0)\tilde{\boldsymbol{B}}(t)$ will be the most time consuming in practical application of (5.1). Note, however, that computing this inverse is often not necessary: if customers do not arrive in batches, interarrival times are always greater than zero, implying that $\tilde{\boldsymbol{A}}(0) = 0$. The recursion (5.1) then reduces to

$$\boldsymbol{S}(t) = \left(\boldsymbol{e} + \sum_{x=1}^{T-t-1} (\boldsymbol{S}(t+x) - \boldsymbol{e}) \tilde{\boldsymbol{A}}(x) \right) \tilde{\boldsymbol{B}}(t). \tag{5.2}$$

A further simplification is obtained if arrivals are generated by a discrete-time Markov process. Then $a_i(1) = 1$ and $a_i(x) = 0$ for $x \neq 1$, and (5.2) simply reads

$$\boldsymbol{S}(t) = \boldsymbol{S}(t+1)\tilde{\boldsymbol{B}}(t).$$

Consequently, the sojourn time distribution is given explicitly by

$$\boldsymbol{S}(T-t) = \boldsymbol{e}\prod_{u=1}^{t} \tilde{\boldsymbol{B}}(T-u), \qquad 0 \leq t < T.$$

Acknowledgements

The authors would like to thank the anonymous referees for their useful comments that helped to improve the paper.

References

[1] Baccelli, F., Gelenbe, E. and Plateau, B. (1984). An end-to-end approach to the resequencing problem. *J. Assoc. Comput. Mach.* **31,** 474–485.

[2] Chowdhury, S. K. (1991). Distribution of the total delay of packets in virtual circuits. In *Proc. IEEE Infocom '91* (Bal Harbour, FL, 7–11 April 1991), IEEE, New York, pp. 911–918.

[3] Dollard, J. D. and Friedman, C. N. (1979). *Product Integration with Applications to Differential Equations* (Encyclopedia Math. Appl. **10**). Addison–Wesley, Reading, MA.

[4] Harrus, G. and Plateau, B. (1982). Queueing analysis of a reordering queue. *IEEE Trans. Software Eng.* **8,** 113–123.

[5] Huisman, T. and Boucherie, R. J. (2001). Running times on railway sections with heterogeneous train traffic. *Transportation Res.* **35,** 271–292.

[6] Kamoun, F., Djerad, M. B. and Lann, G. L. (1982). Queueing analysis of the reordering issue in a distributed database concurrency control mechanism: a general case. In *Proc. 3rd Internat. Conf. Distributed Computing Systems* (Miami, FL, 18–22 October 1982), IEEE Computer Society Press, Los Alamitos, CA, pp. 447–452.

[7] Kleinrock, L., Kamoun, F. and Muntz, R. (1981). Queueing analysis of the reordering queue in a distributed database concurrency control mechanism. In *Proc. 2nd Internat. Conf. Distributed Computing Systems*, IEEE Computer Society Press, Los Alamitos, CA, pp. 13–23.

[8] Ren, J. F., Takahashi, Y. and Hasegawa, T. (1996). Analysis of impact of network delay on multiversion conservative timestamp algorithms in DDBS. *Performance Evaluation* **26,** 21–50.

[9] Zabreyko, P. P. *et al.* (1975). *Integral Equations—A Reference Text*. Noordhoff, Leyden.