Queueing Syst (2017) 86:277–299
DOI 10.1007/s11134-017-9528-z

Queue-length balance equations in multiclass multiserver queues and their generalizations

Marko A. A. Boon[1] · **Onno J. Boxma**[1] · **Offer Kella**[2] · **Masakiyo Miyazawa**[3]

Received: 14 October 2016 / Revised: 25 April 2017 / Published online: 22 May 2017

Abstract A classical result for the steady-state queue-length distribution of single-class queueing systems is the following: The distribution of the queue length just before an arrival epoch equals the distribution of the queue length just after a departure epoch. The constraint for this result to be valid is that arrivals, and also service completions, with probability one occur individually, i.e., not in batches. We show that it is easy to write down somewhat similar balance equations for *multidimensional* queue-length processes for a quite general network of multiclass multiserver queues. We formally derive those balance equations under a general framework. They are called distributional relationships and are obtained for any external arrival process and state-dependent routing as long as certain stationarity conditions are satisfied and external arrivals and service completions do not simultaneously occur. We demonstrate

O. J. Boxma: Partly funded by the NWO Gravity Project NETWORKS, Grant Number 024.002.003. O. Kella: Supported in part by Grant 1462/13 from the Israel Science Foundation and the Vigevani Chair in Statistics. M. Miyazawa: Supported in part by JSPS KAKENHI Grant Number 16H027860001.

✉ Marko A. A. Boon
m.a.a.boon@tue.nl

Onno J. Boxma
o.j.boxma@tue.nl

Offer Kella
offer.kella@gmail.com

Masakiyo Miyazawa
miyazawa@rs.tus.ac.jp

1 Department of Mathematics and Computer Science, Eindhoven University of Technology, Eindhoven, The Netherlands

2 Department of Statistics, The Hebrew University of Jerusalem, 9190501 Jerusalem, Israel

3 Department of Information Sciences, Tokyo University of Science, Noda, Chiba 278 8510, Japan

the use of these balance equations, in combination with PASTA, by (1) providing very simple derivations of some known results for polling systems and (2) obtaining new results for some queueing systems with priorities. We also extend the distributional relationships for a nonstationary framework.

Keywords Queue length · Steady-state distribution · Balance equations · Distributional relationship · Palm distribution · Nonstationary framework

Mathematics Subject Classification 60K25 · 90B22

1 Introduction

A classical result for the steady-state queue-length distribution of single-class queueing systems is the following: The distribution of the queue length just before an arrival epoch equals the distribution of the queue length just after a departure epoch. The constraint for this result to be valid is that, with probability one, arrivals, and also service completions, occur individually, i.e., not in batches. The result then follows by a simple level-crossing argument: In steady state the event that a customer arrives to find j customers present occurs just as often as the event that a customer leaves j customers behind, for all $j = 0, 1, \ldots$. See [7], pp. 154–156, for a formal statement and proof (due to P.J. Burke, unpublished) of this result.

At first sight this level-crossing argument breaks down in higher dimensions, for example, in the case of multiple customer classes. Indeed, with $\boldsymbol{x} \geq \boldsymbol{0}$ and $\boldsymbol{e}_k$ being a unit vector with 1 in the kth coordinate and zero elsewhere, an m-dimensional process can leave state $\boldsymbol{x}$ because of an arrival of a customer of type i, and enter that state from state $\boldsymbol{x} + \boldsymbol{e}_k$ because of a departure of a customer of *another* type k. However, we shall argue that it is easy to write down a more global balance equation for multidimensional queue length processes for a large class of queues and queueing networks—also when service times are not exponentially distributed, and even when arrivals may occur in batches. We shall explore that fact to obtain a simple relation between the steady-state joint queue-length distribution at arrival epochs (which under various circumstances is equal to the time average distribution) and at service completion epochs. Once one has a relation between the probability generating function (PGF) at arbitrary epochs and at service completion epochs, one can find the former when one has the latter. The latter results are indeed known in an $M/G/1$ setting, where it is natural to look at departure epochs. This will yield both new results (for multiclass queueing models with fixed priorities and for the longer-queue model), as well as new and simple derivations of known results, for example, polling models.

The research for the present paper was initially motivated by the desire to provide an intuitive explanation of a result in [4] regarding the steady-state joint queue-length distribution in a large class of polling models. That distribution turned out to have a remarkably simple relation with a weighted sum of the joint queue length distributions at departure epochs of customers from each of the queues. In Sect. 2 we provide such an explanation. Although balance equations are intuitively appealing, their mathematical verification may require a large amount of work. This motivates us to derive

distributional relationships for queue lengths in a unified way using a general tool. The so-called rate conservation law is such a tool as demonstrated in [14] (see also [1,15]). This method is applicable to a general model, but requires Palm distributions, which may not be easy to understand. In Sect. 3 of this paper we take another approach, based on a time evolution of a sample path. This approach is parallel to the rate conservation law, but does not require Palm distributions, which are replaced by sample averages. We apply it to a general model and derive a distributional relationship among different embedded epochs. In Sect. 4 a nonstationary version of the distributional relationship is derived with some error term, which vanishes as time goes to infinity. Our main result, viz. Theorem 1, as well as the nonstationary results, is novel to the best of our knowledge.

Literature review Hébuterne [11] provides a generalization of the above-mentioned classical result of Burke in two directions: He allows (i) batch arrivals, with batches of random size, and (ii) batch services, with batches of fixed size. He also points out that emptying the queue up to N customers is beyond the scope of the analysis, because then the batch sizes are not independent of the system state. Fakinos [9] manages to treat a quite general group-arrival group-departure queue. He treats the batch size problem by assuming that customers within a departing group are randomly ordered, and that they leave the system according to their order. Papaconstantinou and Bertsimas [16] generalize Burke's result to the multiserver $E_k/G/s$ queue. Kim [13] combines the features of batch arrivals, batch services and multiple servers, also allowing multiple customer classes. He does not explicitly address the issue of customers in a departing group being randomly ordered. Hébuterne and Rosenberg [12] focus on the $G/G/1$ queue with batch services and finite capacity. Takine has obtained several relations between queue lengths at random instants and at departure instants; see in particular the very general Theorem 1 in [19], for a single server queue with multiple Markovian arrival streams—an extension of Markovian arrival processes to (possibly correlated) multiple arrival streams.

Organization of the paper Section 2 provides a short proof of a result in [4] by using a multidimensional queue-length balance argument. Section 3 derives the distributional relationship for an open queueing network under a very general setting in Theorem 1. Extensions to the nonstationary case are discussed in Sect. 4. Some applications are presented in Sect. 5. Section 6 contains concluding remarks.

2 A balance equation for a class of polling models

In this section we provide a simple relation between the steady-state joint queue-length distribution at arbitrary epochs and at departure epochs for polling models. This relation, which is derived by introducing a multidimensional queue-length balance argument, is used to provide a short, but somewhat intuitive, derivation of Theorem 1 of [4]. In the next section we shall extend that balance equation in a very general setting and give a rigorous derivation. Let us first describe the polling model studied in [4].

Consider a system of $m \geq 1$ infinite-buffer queues $Q_1, \ldots, Q_m$ and a single server S. Queues are indexed by $J = \{1, 2, \ldots, m\}$. The service times of customers in Q_i

are i.i.d. (independent, identically distributed) positive random variables generically denoted by B_i, with means $b_i := \mathbb{E}B_i$. Denote the Laplace–Stieltjes transform (LST) of B_i by $\tilde{B}_i(\cdot)$. The server moves among the queues in a cyclic order. When S moves from Q_i to Q_{i+1}, it incurs a switchover period. The durations of successive switchover times are i.i.d. nonnegative random variables, which we generically denote by S_i. Denote the LST of S_i by $\tilde{S}_i(\cdot)$ and assume that $s_i := \mathbb{E}S_i < \infty$; let $s := \sum_{i=1}^{m} s_i$. Customers arrive at Q_i according to a Poisson process with rate λ_i; let $\lambda := \sum_{i=1}^{m} \lambda_i$. We do not assume anything about the service disciplines at Q_i. Define $\rho_i := \lambda_i b_i$ as the traffic intensity at Q_i; let $\rho := \sum_{i=1}^{m} \rho_i$. We assume that $\rho < 1$, which is a necessary condition for the system to be stable. In what follows we shall write z for an m-dimensional vector in $\mathbb{R}^m$, $z = (z_1, \ldots, z_m)$, and we assume that $|z_i| \leq 1$ for every $i \in J$. We implicitly use the convention that any index summation is modulo m, for example, $Q_{m+1} \equiv Q_1$.

Assume that all the usual independence assumptions hold between the service times, the switchover times and the interarrival times. We assume that the ergodicity conditions are fulfilled and we restrict ourselves to results for the stationary situation.

Now introduce the PGF of various joint queue-length distributions: $V_i^b(z)$ and $V_i^c(z)$ denote the PGFs of the joint queue-length distribution at visit beginnings and visit completions at Q_i, while $S_i^b(z)$ and $S_i^c(z)$ denote the PGFs of the joint queue-length distribution at service beginnings and service completions at Q_i; $L(z)$ denotes the PGF of the joint queue-length distribution at an arbitrary time in steady state. Theorem 1 of [4] states that, with mean cycle time $\mathbb{E}C = \frac{s}{1-\rho}$:

$$L(z) = \frac{1}{\mathbb{E}C} \sum_{i=1}^{m} \left(\frac{V_i^b(z) - V_i^c(z)}{\Sigma(z)} \frac{z_i \left(1 - \tilde{B}_i(\Sigma(z))\right)}{z_i - \tilde{B}_i(\Sigma(z))} + \frac{V_i^c(z) - V_{i+1}^b(z)}{\Sigma(z)} \right), \quad (1)$$

with $\Sigma(z) := \sum_{j=1}^{m} \lambda_j (1 - z_j)$.

Its proof in [4] is based on the following relations:

(i) a balance relation for polling systems, which is due to Eisenberg [8] and which was generalized in [3]:

$$\gamma_i V_i^b(z) + S_i^c(z) = S_i^b(z) + \gamma_i V_i^c(z), \quad i \in J. \quad (2)$$

Here $\gamma_i := 1/\lambda_i \mathbb{E}C$ represents the reciprocal of the mean number of customers served at Q_i per visit, i.e., the long-term ratio of visit beginnings to service beginnings.

(ii) an obvious relation between queue lengths at the beginning and end of a service time:

$$S_i^c(z) = S_i^b(z) \frac{\tilde{B}_i(\Sigma(z))}{z_i}, \quad i \in J. \quad (3)$$

(iii) an obvious relation between queue lengths at the beginning and end of a switchover time:

$$V_{i+1}^b(z) = V_i^c(z) \tilde{S}_i(\Sigma(z)), \quad i \in J. \quad (4)$$

(iv) a stochastic mean value theorem, expressing $L(z)$ as an average over the PGFs of the joint queue-length distribution at an arbitrary moment during a visit to Q_i ($X_i(z)$) and during a switchover period between Q_i and Q_{i+1} ($Y_i(z)$):

$$L(z) = \frac{1}{\mathbb{E}C} \sum_{i=1}^{m} \left(\frac{b_i}{\gamma_i} X_i(z) + s_i Y_i(z) \right), \tag{5}$$

where, for $i \in J$,

$$X_i(z) = S_i^b(z) \tilde{B}_i^{\text{past}}(\Sigma(z)), \tag{6}$$

$$Y_i(z) = V_i^c(z) \tilde{S}_i^{\text{past}}(\Sigma(z)), \tag{7}$$

where $\tilde{B}_i^{\text{past}}(\cdot)$ and $\tilde{S}_i^{\text{past}}(\cdot)$ are the LST's of the past (elapsed) parts of B_i and S_i, respectively, that is, they are defined as

$$\tilde{B}_i^{\text{past}}(\Sigma(z)) = \frac{1 - \tilde{B}_i(\Sigma(z))}{b_i \Sigma(z)}, \qquad \tilde{S}_i^{\text{past}}(\Sigma(z)) = \frac{1 - \tilde{S}_i(\Sigma(z))}{s_i \Sigma(z)}.$$

Starting from (5), substituting (6) and (7), and using (2) and (3) to eliminate all $S_i^c(z)$ and $S_i^b(z)$, yields (1).

Remark 1 In [4], zero switchover times are also allowed; the same result (1) is shown to hold.

In Theorem 1 of [4] it was subsequently observed that one may simplify (1) as follows, by using (2) and (3):

$$L(z) = \frac{\sum_{i=1}^{m} \lambda_i (1 - z_i) S_i^c(z)}{\sum_{i=1}^{m} \lambda_i (1 - z_i)}. \tag{8}$$

This formula is remarkably simple; please notice that it does not involve the service time distributions, and that the service disciplines at the various queues also do not play a role, which suggests that (1) is based on very general principles. This is the formula for which we would like to provide a short proof. In combination with (2)—(4) it also gives a short proof of (1). In other words, one can obtain an expression for the PGF of the joint steady-state queue-length distribution in a large class of polling systems by just using the elementary balance equations (2) and (9), combined with the obvious relations (3) and (4).

Short proof of (8).

First rewrite (8) as

$$\sum_{i=1}^{m} \lambda_i (1 - z_i) L(z) = \sum_{i=1}^{m} \lambda_i (1 - z_i) S_i^c(z). \tag{9}$$

Secondly observe that, because of the Poisson arrival processes, $L(\boldsymbol{z})$ is also the PGF of the joint queue-length distribution just before an arrival at Q_i, $i \in J$ by PASTA (Poisson Arrival See Time Averages, for example, see [1,15]).

Thirdly invert the transform expressions on both sides of (9), yielding, for $\boldsymbol{x} \geq \boldsymbol{0}$ and $\boldsymbol{e}_i$ being the unit vector with 1 in the ith coordinate and zero elsewhere,

$$\sum_{i=1}^{m} \lambda_i \pi_i^e(\boldsymbol{x}) - \sum_{i=1}^{m} \lambda_i \pi_i^e(\boldsymbol{x} - \boldsymbol{e}_i) = \sum_{i=1}^{m} \lambda_i \pi_i^d(\boldsymbol{x}) - \sum_{i=1}^{m} \lambda_i \pi_i^d(\boldsymbol{x} - \boldsymbol{e}_i), \tag{10}$$

where $\pi_i^d(\cdot)$ indicates that we consider the joint queue-length distribution immediately *after* a departure from Q_i, and $\pi_i^e(\cdot)$ denotes that we view the system just *before* an external arrival at Q_i. Fourthly we reshuffle the terms:

$$\sum_{i=1}^{m} \lambda_i \pi_i^e(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i \pi_i^d(\boldsymbol{x} - \boldsymbol{e}_i) = \sum_{i=1}^{m} \lambda_i \pi_i^d(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i \pi_i^e(\boldsymbol{x} - \boldsymbol{e}_i). \tag{11}$$

Finally observe that the left-hand side of (11) represents the rate out of state $\boldsymbol{x}$, and the right-hand side represents the rate into that state. Indeed the first term on the left-hand side corresponds to arrivals which find $\boldsymbol{x}$ customers in the system. The second term on the left-hand side is slightly less obvious. It corresponds to departures that take place in state $\boldsymbol{x}$. Notice that the rate at which customers depart from Q_i equals λ_i (although the departure process will not be a Poisson process), and that $\pi_i^d(\boldsymbol{x} - \boldsymbol{e}_i)$ is the fraction of departures from Q_i which take the system out of state $\boldsymbol{x}$. Similarly interpret the terms on the right-hand side. We conclude that (8) amounts to a simple flow balance formula.

Remark 2 A similar flow balance argument was used in [5] to derive a queue-length expression in an $M/G/1$ FCFS queue with multiple customer classes.

Remark 3 Observe that (8) immediately gives the formula for the marginal distributions. Indeed, for a vector $\boldsymbol{z}_{m,i} = (1, \ldots, 1, z_i, 1, \ldots, 1)$, $L(\boldsymbol{z}_{m,i}) = S_i^c(\boldsymbol{z}_{m,i})$. From the well-known "step" (level-crossing) argument it follows that $S_i^c(\boldsymbol{z}_{m,i})$ is also the PGF of the queue-length distribution in Q_i at an *arrival* epoch at Q_i. By PASTA it is also the PGF of the steady-state distribution of Q_i.

Next take $\boldsymbol{z}_T = (z, \ldots, z)$. (8) now states that the PGF of the distribution of the total queue length (in terms of z) equals $\sum_{i=1}^{m} \lambda_i S_i^c(\boldsymbol{z}_T) / \sum_{j=1}^{m} \lambda_j$. This formula may be interpreted as follows. By PASTA, $L(\boldsymbol{z}_T)$ is also the PGF of the distribution of the total queue length at an arrival epoch. By a level-crossing argument it follows that this equals the PGF of the distribution of the total queue length just after a departure epoch. The result now follows from the observation that a fraction $\lambda_i / \sum_{j=1}^{m} \lambda_j$ of the departure epochs refers to a departure from Q_i.

Remark 4 Relation (8) may be viewed as an m-dimensional version of the above-mentioned one-dimensional "step" (level-crossing) relation that holds for queues with single arrivals and single departures.

3 Formal derivations under a general framework

In this section we aim to derive distributional relationships at arrival and departure instants for various queues and their network models in a unified way, under general settings. Roughly speaking these settings allow simultaneous external arrivals, simultaneous departures and routing at different stations; however, we do not allow an external arrival to coincide with a departure. We use their time evolutions in sample paths for deriving the relationships rather than using flow balance.

We describe a queueing network system under a fairly general framework. We consider an open queueing network system with m queues, where queues uniquely belong to service facilities, which are called stations. Queues in the same station may be distinguished by customer classes. Each station may have multiple servers, which may change in time. External arrivals at queues are general as long as they satisfy certain stationarity conditions. Customers completing service may be routed among queues depending on the state of the whole system. Thus, this model is quite general and very flexible.

To describe this model we introduce a stochastic process. Queues are still indexed by $J = \{1, 2, \ldots, m\}$. Let

$$\boldsymbol{X}(t) = (X_1(t), \ldots, X_m(t)),$$

where $X_i(t)$ represents the length of queue i at time t, which includes customers in service. Here each queue belongs to a single station. There is a mapping from queues to stations, which will be given when needed.

In addition to $\boldsymbol{X}(t)$, the following counting processes count the number of specified events until time $t \geq 0$, for $i \in J$:

- $N_i^e(t)$—external arrivals at queue i,
- $N_i^d(t)$—departures from queue i,
- $N_i^r(t)$—internal arrivals at queue i (transition from some queue).

With $\boldsymbol{N}^u(t) = (N_1^u(t), \ldots, N_m^u(t))$ for $u = e, d, r$, we consider the process

$$\boldsymbol{Z}(t) \equiv (\boldsymbol{X}(t), \boldsymbol{N}^e(t), \boldsymbol{N}^d(t), \boldsymbol{N}^r(t)) .$$

All processes are assumed right-continuous with left limits. Let $\Delta\boldsymbol{X}(t) = \boldsymbol{X}(t) - \boldsymbol{X}(t-)$. $\Delta\boldsymbol{N}^u(t)$ is similarly defined and is in $\mathbb{Z}_+^m$ for $u = e, d, r$, where $\mathbb{Z}_+$ is the set of nonnegative integers.

For $u = e, d, r$, denote

$$|\boldsymbol{N}^u|(t) = \sum_{i \in J} N_i^u(t)$$

and assume that

(i) $\boldsymbol{X}(0), \boldsymbol{N}^e(t), \boldsymbol{N}^d(t), \boldsymbol{N}^r(t)$ are all finite (in $\mathbb{Z}_+^m$) for each $t \geq 0$.
(ii) $\Delta|\boldsymbol{N}^e|(t)\Delta|\boldsymbol{N}^d|(t) = 0$ for each $t \geq 0$. That is, external arrivals and service completions cannot occur simultaneously.

We also need to define the intermediate state

$$\boldsymbol{X}^d(t) = \boldsymbol{X}(t) - \Delta \boldsymbol{N}^r(t) \in \mathbb{Z}_+^m \,. \tag{12}$$

This differs from $\boldsymbol{X}(t)$ only at departure epochs, and it describes the state "after" a departure and "before" an internal arrival at a different queue.

Clearly the following dynamics hold:

$$\boldsymbol{X}(t) = \boldsymbol{X}(0) + \boldsymbol{N}^e(t) - \boldsymbol{N}^d(t) + \boldsymbol{N}^r(t) \in \mathbb{Z}_+^m \,. \tag{13}$$

Because of (i) $\boldsymbol{X}(t)$ and $\boldsymbol{X}^d(t)$ are also finite. It may be natural to assume that $|\boldsymbol{N}^r|(t) \le |\boldsymbol{N}^d|(t)$ for $t \ge 0$, but we do not require it in this section.

Thus, $\boldsymbol{X}(t)$ is the state of the system at time t of an input–output system driven by the counting processes $\boldsymbol{N}^e$, $\boldsymbol{N}^d$, $\boldsymbol{N}^r$. The dynamics of (12) and (13) indicates that we adopt the *departure first* framework. We have used queueing terminologies, but our results are valid as long as the above mathematical assumptions and (13) are satisfied.

In general $|\boldsymbol{N}^e|(t)$, $|\boldsymbol{N}^d|(t)$ and $N_i^e(t)$ and $N_i^d(t)$ may have jumps greater than one, which is not convenient to describe the time evolution of $\boldsymbol{Z}(t)$. Thus, for $u = e, d$, we introduce

$$|\tilde{\boldsymbol{N}}^u|(t) = \sum_{0<s\le t} 1(\Delta|\boldsymbol{N}^u|(s) \ge 1), \qquad \tilde{N}_i^u(t) = \sum_{0<s\le t} 1(\Delta N_i^u(s) \ge 1),$$

then $\Delta|\tilde{\boldsymbol{N}}^u|(t) \le 1$ and $\Delta\tilde{N}_i^u(t) \le 1$, that is, $|\tilde{\boldsymbol{N}}^u|$ and $\tilde{N}_i^u$ are simple point processes for $u = e, d$. Set $t_0^e = t_0^d = t_{i,0}^e = t_{i,0}^d = 0$ for $i \in J$, and, for $n \ge 1$ and $i \in J$, let $t_n^e, t_n^d, t_{i,n}^e, t_{i,n}^d$ be the n^{th} jump epoch of $|\tilde{\boldsymbol{N}}^e|$, $|\tilde{\boldsymbol{N}}^d|$, $\tilde{N}_i^e$, $\tilde{N}_i^d$, respectively (of course, if the corresponding process is not terminating and such an epoch exists).

Another basic assumption on the counting processes is

(iii) There exist finite and positive numbers λ^u, $u = e, d$ such that

$$\lambda^u = \lim_{t\to\infty} \frac{1}{t} |\tilde{\boldsymbol{N}}^u|(t) \,, \tag{14}$$

a.s. (almost surely) w.r.t. the underlying probability measure $\mathbb{P}$.

We further assume the following ergodic type conditions:

(iv) There exist probability distributions π^e and π^d such that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{\ell=1}^{n} 1(\boldsymbol{X}(t_\ell^e-) = \boldsymbol{x}, \Delta\boldsymbol{N}^e(t_\ell^e) = \boldsymbol{y}) = \pi^e(\boldsymbol{x}, \boldsymbol{y}), \quad \text{a.s.}, \quad \boldsymbol{x}, \boldsymbol{y} \in \mathbb{Z}_+^m, \tag{15}$$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{\ell=1}^{n} 1(\boldsymbol{X}^d(t_\ell^d) = \boldsymbol{x}, \Delta\boldsymbol{N}^d(t_\ell^d) = \boldsymbol{y}, \Delta\boldsymbol{N}^r(t_\ell^d) = \boldsymbol{z}) = \pi^d(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}), \quad \text{a.s.},$$
$$\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathbb{Z}_+^m. \tag{16}$$

From the definitions in (iv) π^e and π^d are considered as the embedded stationary distributions just before arrival epochs and just after departure epochs but before internal arrivals, respectively. They correspond to Palm distributions concerning their counting processes in the time stationary framework (for example, see [1]).

Since the process $\boldsymbol{X}(t)$ is vector-valued, it is not so convenient for manipulations. So we introduce a test function $f : \mathbb{Z}_+^m \to \mathbb{R}$. In the setting (i)–(iv) we will derive distributional relationships among characteristics at different embedded instants using the test function f. For this we need the following lemma.

Lemma 1 *If* (15) *holds, then, for any bounded function* $g : \mathbb{Z}_+^{2m} \to \mathbb{R}$, *we have*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{\ell=1}^{n} g(\boldsymbol{X}(t_\ell^e-), \Delta \boldsymbol{N}^e(t_\ell^e)) = \sum_{\boldsymbol{x},\boldsymbol{y}\in\mathbb{Z}_+^m} g(\boldsymbol{x}, \boldsymbol{y})\pi^e(\boldsymbol{x}, \boldsymbol{y}), \quad \text{a.s.} \tag{17}$$

Similarly, if (16) *holds, then, for any bounded function* $h : \mathbb{Z}_+^{3m} \to \mathbb{R}$, *we have*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{\ell=1}^{n} h(\boldsymbol{X}^d(t_\ell^d), \Delta \boldsymbol{N}^d(t_\ell^d), \Delta \boldsymbol{N}^r(t_\ell^d)) = \sum_{\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}\in\mathbb{Z}_+^m} h(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})\pi^d(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}), \quad \text{a.s.} \tag{18}$$

This lemma may look obvious, but its proof is not immediate because we need to verify the exchange of limits. We prove it in the appendix.

We are now ready to prove distributional relationships. First we denote the expectations under π^e and π^d by $\mathbb{E}^e$ and $\mathbb{E}^d$, respectively. That is,

$$\mathbb{E}^e g(\boldsymbol{X}, \boldsymbol{Y}) = \sum_{\boldsymbol{x},\boldsymbol{y}\in\mathbb{Z}_+^m} g(\boldsymbol{x}, \boldsymbol{y})\pi^e(\boldsymbol{x}, \boldsymbol{y}), \tag{19}$$

$$\mathbb{E}^d h(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}) = \sum_{\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}\in\mathbb{Z}_+^m} h(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})\pi^d(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}). \tag{20}$$

Note that $\boldsymbol{Y}$ in $\mathbb{E}^e$ represents sizes of externally arriving batches, while $\boldsymbol{Y}$ in $\mathbb{E}^d$ represents sizes of departing batches.

Theorem 1 *In the setting (i)–(iv), for any bounded function* $f : \mathbb{Z}_+^m \to \mathbb{R}$, *we have*

$$\lambda^e \mathbb{E}^e\big[f(\boldsymbol{X}+\boldsymbol{Y}) - f(\boldsymbol{X})\big] + \lambda^d \mathbb{E}^d\big[f(\boldsymbol{X}+\boldsymbol{Z}) - f(\boldsymbol{X})\big] = \lambda^d \mathbb{E}^d\big[f(\boldsymbol{X}+\boldsymbol{Y}) - f(\boldsymbol{X})\big]. \tag{21}$$

Proof Since $f(\boldsymbol{X}(t))$ changes in time only at the counting instants t_n^e or t_n^d, we have (with Δ being defined as earlier in this section)

$$f(\boldsymbol{X}(t)) - f(\boldsymbol{X}(0)) = \sum_{\ell=1}^{|\tilde{\boldsymbol{N}}^e|(t)} \Delta f(\boldsymbol{X}(t_\ell^e)) + \sum_{\ell=1}^{|\tilde{\boldsymbol{N}}^d|(t)} \Delta f(\boldsymbol{X}(t_\ell^d)). \tag{22}$$

Recalling (12), we have $\boldsymbol{X}(t_\ell^d) = \boldsymbol{X}^d(t_\ell^d) + \Delta \boldsymbol{N}^r(t_\ell^d)$, and this and (13) yield

$$\boldsymbol{X}(t_\ell^d -) = \boldsymbol{X}(t_\ell^d) + \Delta \boldsymbol{N}^d(t_\ell^d) - \Delta \boldsymbol{N}^r(t_\ell^d) = \boldsymbol{X}^d(t_\ell^d) + \Delta \boldsymbol{N}^d(t_\ell^d).$$

From these $\boldsymbol{X}(t_\ell^d)$ and $\boldsymbol{X}(t_\ell^d -)$ we have

$$\begin{aligned}\sum_{\ell=1}^{|\tilde{\boldsymbol{N}}^d|(t)} \Delta f(\boldsymbol{X}(t_\ell^d)) &= \sum_{\ell=1}^{|\tilde{\boldsymbol{N}}^d|(t)} (f(\boldsymbol{X}(t_\ell^d)) - f(\boldsymbol{X}^d(t_\ell^d))) + \sum_{\ell=1}^{|\tilde{\boldsymbol{N}}^d|(t)} (f(\boldsymbol{X}^d(t_\ell^d)) \\ &\quad - f(\boldsymbol{X}(t_\ell^d -))) \\ &= \sum_{\ell=1}^{|\tilde{\boldsymbol{N}}^d|(t)} (f(\boldsymbol{X}^d(t_\ell^d) + \Delta \boldsymbol{N}^r(t_\ell^d)) - f(\boldsymbol{X}^d(t_\ell^d))) \\ &\quad + \sum_{\ell=1}^{|\tilde{\boldsymbol{N}}^d|(t)} (f(\boldsymbol{X}^d(t_\ell^d)) - f(\boldsymbol{X}^d(t_\ell^d) + \Delta \boldsymbol{N}^d(t_\ell^d))). \end{aligned} \tag{23}$$

It follows from (22) and (23) that

$$\begin{aligned}&\sum_{\ell=1}^{|\tilde{\boldsymbol{N}}^e|(t)} (f(\boldsymbol{X}(t_\ell^e -) + \Delta \boldsymbol{N}^e(t_\ell^e)) - f(\boldsymbol{X}(t_\ell^e -))) + \sum_{\ell=1}^{|\tilde{\boldsymbol{N}}^d|(t)} (f(\boldsymbol{X}^d(t_\ell^d) + \Delta \boldsymbol{N}^r(t_\ell^d)) \\ &\quad - f(\boldsymbol{X}^d(t_\ell^d))) \\ &= \sum_{\ell=1}^{|\tilde{\boldsymbol{N}}^d|(t)} (f(\boldsymbol{X}^d(t_\ell^d) + \Delta \boldsymbol{N}^d(t_\ell^d)) - f(\boldsymbol{X}^d(t_\ell^d))) + f(\boldsymbol{X}(t)) - f(\boldsymbol{X}(0)). \end{aligned} \tag{24}$$

Dividing both sides of this equation by t and letting $t \to \infty$ yields (21) by (14)–(16) and Lemma 1 because f is bounded. □

The assumptions of Theorem 1 exclude arrivals and departures occurring simultaneously, but allow them to occur separately as multiple simultaneous external arrivals or multiple simultaneous departures and routing. The model as well as the distributional relationship may be too general for queueing networks. To make them more specific we make the following assumption:

(v) There exist finite and nonnegative numbers λ_A^d for nonempty $A \subset J$, that is, $A \in 2^J \setminus \{\emptyset\}$, such that

$$\lambda_A^d = \lim_{t\to\infty} \frac{1}{t} \tilde{N}_A^d(t), \qquad \text{a.s.}, \tag{25}$$

where, with the notation $S_A \equiv \{\boldsymbol{x} \in \mathbb{Z}_+^m; x_i > 0, i \in A, x_j = 0, j \in J\backslash A\}$,

$$\tilde{N}_A^d(t) = \sum_{0<s\le t} 1(\Delta \boldsymbol{N}^d(s) \in S_A). \tag{26}$$

Note that $\tilde{N}_A^d$ counts instants when departures occur simultaneously from queues $i \in A$, but there is no departure from queue $j \in J\backslash A$, while $\Delta\tilde{N}_A^d(t)\Delta\tilde{N}_B^d(t) = 0$ if $A \neq B$. Thus, the setting (i)–(v) still allow batch arrivals and batch departures and simultaneous transfer of customers in a departing batch.

We will use the following notation. For each $A \in 2^J\backslash\{\emptyset\}$, let, for $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathbb{Z}_+^m$,

$$\pi_A^d(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) = \begin{cases} \frac{\lambda_A^d}{\lambda_A^d}\pi^d(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}), & \lambda_A^d > 0, \\ 0, & \lambda_A^d = 0. \end{cases}$$

Since $\tilde{N}_A^d$ exclusively counts the increasing epochs of $|\tilde{\boldsymbol{N}}^d|$ for different As, we have

$$|\tilde{\boldsymbol{N}}^d|(t) = \sum_{A\in 2^J\backslash\{\emptyset\}} \tilde{N}_A^d(t), \tag{27}$$

which implies that $\lambda^d = \sum_{A\in 2^J\backslash\{\emptyset\}} \lambda_A^d$, and for $A \in \{B \in 2^J | \lambda_B^d > 0\}$, π_A^d is a probability distribution on $\mathbb{Z}_+^{3m}$, which can be restricted to $\mathbb{Z}_+^m \times S_A \times \mathbb{Z}_+^m$.

Let $t_{A,n}^d$ be the n^{th} jump epoch of $\tilde{N}_A^d$. Just as Lemma 1 does, the following lemma plays a key role; it is proved in the appendix.

Lemma 2 *In the setting (i)–(v) there exist probability distributions π_A^d such that, for any bounded function $h : \mathbb{Z}_+^{3m} \to \mathbb{R}$, with $A \in 2^J\backslash\{\emptyset\}$,*

$$\begin{aligned} &\lim_{n\to\infty} \frac{1}{n} \sum_{\ell=1}^{n} h(\boldsymbol{X}^d(t_{A,\ell}^d), \Delta\boldsymbol{N}^d(t_{A,\ell}^d), \Delta\boldsymbol{N}^r(t_{A,\ell}^d)) \\ &\quad = \sum_{\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}\in\mathbb{Z}_+^m} h(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})\pi_A^d(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}), \text{ a.s.} \end{aligned} \tag{28}$$

By (27), Theorem 1 and Lemma 2 yield the following corollary. As with $\mathbb{E}^e$ and $\mathbb{E}^d$, $\mathbb{E}_A^d$ stands for the expectation under π_A^d.

Corollary 1 *In the setting (i)–(v), for any bounded function $f : \mathbb{Z}_+^m \to \mathbb{R}$,*

$$\begin{aligned} &\lambda^e\mathbb{E}^e[f(\boldsymbol{X}+\boldsymbol{Y}) - f(\boldsymbol{X})] + \sum_{A\in 2^J\backslash\{\emptyset\}} \lambda_A^d\mathbb{E}_A^d[f(\boldsymbol{X}+\boldsymbol{Z}) - f(\boldsymbol{X})] \\ &\quad = \sum_{A\in 2^J\backslash\{\emptyset\}} \lambda_A^d\mathbb{E}_A^d[f(\boldsymbol{X}+\boldsymbol{Y}) - f(\boldsymbol{X})]. \end{aligned} \tag{29}$$

Remark 5 If $\Delta\tilde{N}_i^d(t_{j,n}^d) = 0$ for all $i \neq j$, then $\lambda_A^d > 0$ only if A is a singleton. In this case the summations over A in (29) can be reduced to those over $i \in J$, replacing A by i.

Until now our distributional relationship may still be too general because no assumption is made on how the counting processes are generated from $\boldsymbol{X}(t)$ and other information. To describe this a filtration is convenient. Let $\mathcal{F}_t$ be the σ-field generated by all events up to time t, and let $\mathcal{F}_{t-} = \sigma(\cup_{u<t}\mathcal{F}_u)$, that is, $\mathcal{F}_{t-}$ is a σ-field generated by all events before time t. For a stopping time τ, let $\mathcal{F}_{\tau-} = \sigma(\mathcal{F}_0, \{A\cap\{t < \tau\} \in \mathcal{F}_t\})$, where $\sigma(\mathcal{A})$ is the σ-field generated by a family of events $\mathcal{A}$. Using the filtration the following assumptions are typically used in the setting (i)–(v):

(a1) t_n^e, $t_{i,n}^d$ are stopping times with respect to $\{\mathcal{F}_t; t \geq 0\}$. This can always be realized by choosing a sufficiently large $\mathcal{F}_t$.
(a2) $\Delta\boldsymbol{N}^e(t_n^e)$ is independent of $\mathcal{F}_{t_n^e-}$. That is, the sizes of batch arrivals are independent of the state of the system just before their arrival epochs.
(a3) $\Delta|\boldsymbol{N}^d|(t_n^d) = 1$. That is, departures singly occur from one queue at a time.
(a4) $\Delta N_j^r(t_{i,n}^d) \leq 1$ for $j \in J$, and $\Delta\boldsymbol{N}^r(t_{i,n}^d)$ is in the σ-field generated by $\mathcal{F}_{t_{i,n}^d-}$ and $\Delta\boldsymbol{N}^d(t_{i,n}^d)$.

By (a3), $\tilde{N}_A^d(t) \equiv 0$ if A is not a singleton. Thus, we write $\tilde{N}_A^d(t)$ as $\tilde{N}_i^d(t)$ for $A = \{i\}$. Similarly π_A^d is written as π_i^d for $A = \{i\}$. In the setting (i)–(v) and under the assumptions (a1)–(a4), $\Delta N_j^r(t_{i,\ell}^d) \leq 1$, and therefore Lemma 2 yields

$$\lim_{n\to\infty}\frac{1}{n}\sum_{\ell=1}^{n} 1\left(\boldsymbol{X}^d(t_{i,\ell}^d) = \boldsymbol{x}, \Delta N_i^d(t_{i,\ell}^d) = 1, \Delta N_j^r(t_{i,\ell}^d) = 1\right) = \pi_i^d(\boldsymbol{x}, \boldsymbol{e}_i, \boldsymbol{e}_j),$$

which is denoted by $\pi_{ij}^d(\boldsymbol{x})$. We here recall that $\boldsymbol{e}_i \in \mathbb{Z}_+^m$ is the unit vector whose i-th entry is one and the other entries are zero. Thus, applying Corollary 1 for $f(\boldsymbol{x}) = \boldsymbol{z}^{\boldsymbol{x}}$, where we recall that $\boldsymbol{z}^{\boldsymbol{x}} = \prod_{i\in J} z_i^{x_i}$, we have the following relationship.

Corollary 2 *In the setting (i)–(v) and under assumptions (a1)–(a4), for* $\boldsymbol{z} = (z_1, \ldots, z_m)$ *satisfying* $|z_i| \leq 1$ *for* $i \in J$,

$$\lambda^e\left(1 - \mathbb{E}^e[\boldsymbol{z}^{\boldsymbol{Y}}]\right)\varphi^e(\boldsymbol{z}) + \sum_{j\in J}(1 - z_j)\sum_{i\in J}\lambda_i^d\varphi_{ij}^d(\boldsymbol{z}) = \sum_{i\in J}\lambda_i^d(1 - z_i)\varphi_i^d(\boldsymbol{z}), \quad (30)$$

where

$$\varphi^e(\boldsymbol{z}) = \mathbb{E}^e[\boldsymbol{z}^{\boldsymbol{X}}], \qquad \varphi_i^d(\boldsymbol{z}) = \mathbb{E}_i^d[\boldsymbol{z}^{\boldsymbol{X}}], \qquad \varphi_{i,j}^d(\boldsymbol{z}) = \sum_{\boldsymbol{x}\in\mathbb{Z}_+^m}\boldsymbol{z}^{\boldsymbol{x}}\pi_{ij}^d(\boldsymbol{x}), \qquad i, j \in J.$$

Remark 6 Under the assumptions of this corollary the routing of departing customers may depend on all queue lengths in the network.

Corollary 2 is specialized to Corollary 3 if external arrivals to queues occur one at a time. Namely,

(vi) No simultaneous arrivals occur, and there exist finite numbers (some, but not all, possibly zero) λ_k^e for $k \in J$ such that

$$\lambda_k^e = \lim_{t\to\infty} \frac{1}{t} \tilde{N}_k^e(t), \quad \text{a.s.}, \qquad k \in J. \tag{31}$$

Corollary 3 *Under the assumptions of Corollary 2, assume that (vi) also holds. Define* π_k^e *as*

$$\pi_k^e(\boldsymbol{x}, y_k) = \begin{cases} \frac{\lambda_k^e}{\lambda_k^e} \pi^e(\boldsymbol{x}, y_k), & \lambda_k^e > 0, \\ 0, & \lambda_k^e = 0; \end{cases}$$

then, for $k \in J_e \equiv \{i \in J \,|\, \lambda_i^e > 0\}$, π_k^e *is a probability distribution on* $\mathbb{Z}_+^{m+1}$, *and (30) becomes*

$$\sum_{k\in J} \lambda_k^e \big(1 - \mathbb{E}^e\big[z_k^{Y_k}\big]\big)\varphi_k^e(\boldsymbol{z}) + \sum_{j\in J}(1 - z_j) \sum_{i\in J} \lambda_i^d \varphi_{ij}^d(\boldsymbol{z}) = \sum_{i\in J} \lambda_i^d (1 - z_i)\varphi_i^d(\boldsymbol{z}), \tag{32}$$

where φ_k^e *is the generating function of* $\boldsymbol{X}$ *under the conditional distribution* π_k^e.

Corollary 3 immediately implies the following corollary.

Corollary 4 *Under the assumptions of Corollary 3, if the event* $\{\Delta N_j^r(t_{i,\ell}^d) = 1\}$ *is independent of* $\mathcal{F}_{t_{i,\ell}^d -}$, *then there exist* $p_{ij} \geq 0$ *such that* $\pi_i^d(\boldsymbol{x}, 1, \boldsymbol{e}_j) = \pi_i^d(\boldsymbol{x}, 1)\, p_{ij}$, *and (32) becomes*

$$\sum_{k\in J_e} \lambda_k^e \big(1 - \mathbb{E}^e\big[z_k^{Y_k}\big]\big)\varphi_k^e(\boldsymbol{z}) + \sum_{i\in J} \lambda_i^d \varphi_i^d(\boldsymbol{z}) \sum_{j\in J} p_{ij}(1 - z_j) = \sum_{i\in J} \lambda_i^d (1 - z_i)\varphi_i^d(\boldsymbol{z}). \tag{33}$$

Remark 7 In Sect. 5 we shall present several applications of the above theorem and corollaries. In particular the polling result (8) of Sect. 2 is there shown to be a special case of Corollary 1.

Notice that the setup of this section includes the finite buffer case. This is done by having no arrivals to a queue during times in which it is saturated. This type of dependence is allowed by our setup. Some results for the single server queue with finite capacity are contained in [12].

4 Distributional relationship up to a given time

The purpose of this section is to derive a nonstationary version of Theorem 1, a distributional relationship *up to a given time*. We adopt the setting (i)–(iv) of Sect. 3 and consider the process $\boldsymbol{Z}(t)$ introduced in the beginning of that section. We first define

the expected relative frequencies for bounded test functions g, h from $\mathbb{Z}_+^{2m}, \mathbb{Z}_+^{3m}$ to $\mathbb{R}$ up to time t as

$$R_t^e g = \frac{1}{|\tilde{\boldsymbol{N}}^e|(t)} \sum_{n=1}^{|\tilde{\boldsymbol{N}}^e|(t)} g(\boldsymbol{X}(t_n^e-), \Delta \boldsymbol{N}^e(t_n^e)) 1(|\tilde{\boldsymbol{N}}^e|(t) > 0),$$

$$R_t^d h = \frac{1}{|\tilde{\boldsymbol{N}}^d|(t)} \sum_{n=1}^{|\tilde{\boldsymbol{N}}^e|(t)} h(\boldsymbol{X}^d(t_n^d), \Delta \boldsymbol{N}^d(t_n^d), \Delta \boldsymbol{N}^r(t_n^d)) 1(|\tilde{\boldsymbol{N}}^d|(t) > 0).$$

For each bounded function $f : \mathbb{Z}_+^m \to \mathbb{R}$ we define the following test functions:

$$g^e(\boldsymbol{x}, \boldsymbol{y}) = f(\boldsymbol{x}), \qquad g_+^e(\boldsymbol{x}, \boldsymbol{y}) = f(\boldsymbol{x} + \boldsymbol{y}),$$
$$h^d(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) = f(\boldsymbol{x}), \qquad h_-^d(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) = f(\boldsymbol{x} + \boldsymbol{y}), \qquad h_+^d(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) = f(\boldsymbol{x} + \boldsymbol{z}).$$

Let

$$\lambda^e(t) = \frac{1}{t}|\tilde{\boldsymbol{N}}^e|(t), \qquad \lambda^d(t) = \frac{1}{t}|\tilde{\boldsymbol{N}}^d|(t).$$

Then (24) yields the following lemma.

Lemma 3 *In the setting (i)–(iv), for any bounded function $f : \mathbb{Z}_+^m \to \mathbb{R}$, we have, for any $t > 0$,*

$$\begin{aligned} &\lambda^e(t)\big(R_t^e g_+^e - R_t^e g^e\big) + \lambda^d(t)\big(R_t^d h_+^d - R_t^d h^d\big) \\ &\quad - \lambda^d(t)\big(R_t^d h_-^d - R_t^d h^d\big) = \frac{1}{t}\big(f(\boldsymbol{X}(t)) - f(\boldsymbol{X}(0))\big). \end{aligned} \tag{34}$$

We may interpret Lemma 3 as a transient version of Theorem 1. It is notable that (34) holds without any stability condition, and its right-hand side vanishes as $t \to \infty$ at most in linear order of t^{-1} because f is bounded. If there exists a unique probability measure such that $(\boldsymbol{X}(t), \Delta \boldsymbol{N}^e(t), \Delta \boldsymbol{N}^d(t), \Delta \boldsymbol{N}^r(t))$ is stationary, then $R_t^e g$, $R_t^d h$ converge to the corresponding expectations under the Palm distributions involving $|\tilde{\boldsymbol{N}}^e|$, $|\tilde{\boldsymbol{N}}^d|$, respectively. Thus, we have

$$\lim_{t\to\infty} R_t^e g^e = \mathbb{E}^e f(\boldsymbol{X}), \qquad \lim_{t\to\infty} R_t^e g_+^e = \mathbb{E}^e f(\boldsymbol{X} + \boldsymbol{Y}),$$
$$\lim_{t\to\infty} R_t^d h^d(\boldsymbol{x}) = \mathbb{E}^d f(\boldsymbol{X}), \qquad \lim_{t\to\infty} R_t^d h_-^d = \mathbb{E}^e f(\boldsymbol{X} + \boldsymbol{Y}),$$
$$\lim_{t\to\infty} R_t^d h_+^d = \mathbb{E}^d f(\boldsymbol{X} + \boldsymbol{Z}),$$

and we recover (21) from (34). Corollary 1, (30) and (32) are similarly obtained. We omit the routine details.

5 Some special cases and applications

In this section we consider several applications of the theorem and corollaries of Sect. 3. We first note that, if nonzero N_k^e for $k \in J$ are independent compound Poisson processes, then by PASTA the embedded stationary distributions π^e and π_k^e are identical with the time stationary distributions.

Case 1: *An m-class queue with batch arrivals*

We consider an m-class single-node service facility, with $m \geq 1$. We allow multiple servers. Customers arrive according to a Poisson process, possibly in batches. Customers of class i require service at the service facility according to service time distribution $B_i(\cdot)$, $i \in J$. These distributions are assumed to be continuous, but not otherwise specified. No customers are lost; there is an infinite waiting room. After completion of their service customers immediately leave. We assume that the steady-state joint queue-length distribution (numbers of customers of all classes in the system) exists. Its PGF is denoted by $L(\boldsymbol{z})$. We also again (as in Sect. 2) denote the PGF of the steady-state joint queue-length distributions immediately after departure epochs of a class i customer by $S_i^c(\boldsymbol{z})$, $i \in J$. We do not specify according to which service discipline the customers are served; polling with FCFS within each class is just one of many options.

Theorem 2 *Consider the above-described m-class single-node service facility. Assume that customers arrive according to a batch Poisson process with rate* λ *and that customers are served individually, in some nonspecified order. Let an arbitrary batch arrival have size* $\boldsymbol{G} = (G_1, \ldots, G_m)$ *with PGF* $\mathbb{E}[\boldsymbol{z}^{\boldsymbol{G}}] = \mathbb{E}[z_1^{G_1} \ldots z_m^{G_m}]$. *Then the following relation holds between the PGF* $L(\boldsymbol{z})$ *and the PGFs* $S_i^c(\boldsymbol{z})$, $i \in J$:

$$(1 - \mathbb{E}[\boldsymbol{z}^{\boldsymbol{G}}])L(\boldsymbol{z}) = \sum_{i=1}^{m}(1 - z_i)\mathbb{E}G_i S_i^c(\boldsymbol{z}). \tag{35}$$

Proof After using PASTA, Theorem 2 is a special case of (30) of Corollary 2 in which there is no routing. □

Remark 8 Special cases of the above theorem are obtained by assuming that batches always contain only customers of one type. For the special case that batches have just one customer of class i with probability $\frac{\lambda_i}{\lambda}$, $i \in J$, (35) reduces to (8) that was obtained for the polling system that provided the initial motivation for the present study (but (8) obviously holds for a much more general class of service disciplines).

Case 2: *Generalization of Theorem 2 to the case of batch services*

The following theorem generalizes the main result in [11], but is a special case of Theorem 2 of [19] which allows a more general arrival process (but in that theorem batch service is not considered).

Theorem 3 *Consider the m-class single-node service facility of Theorem 2, with the additional assumption that customers of class i are always served in batches of fixed size* K_i, $i \in J$; *the start of a service of class i customers is delayed until* K_i *customers*

are present. Then the following relation holds between the PGF $L(\boldsymbol{z})$ and the PGFs $S_i^c(\boldsymbol{z})$, $i \in J$:

$$(1 - \mathbb{E}[\boldsymbol{z}^{\boldsymbol{G}}])L(\boldsymbol{z}) = \sum_{i=1}^{m} \frac{1 - z_i^{K_i}}{K_i} \mathbb{E}G_i S_i^c(\boldsymbol{z}). \tag{36}$$

Proof In view of Remark 5, and after using PASTA, Theorem 3 is a special case of (29) of Corollary 1 in which there is no routing, the external arrival batch $\boldsymbol{Y}(t_n^e)$ is independent of $\mathcal{F}_{t_n^e-}$ and the departing batch size Y_i from queue i is some constant K_i. In this case it is easy to see that $\lambda_i^d \mathbb{E}[Y_i] = \lambda^e / K_i$, and we obtain (36) from (29). □

Case 3: *Nonpreemptive priority queues*

In this example we consider a nonpreemptive priority queue with P customer classes. We first verify the equality between the PGFs as given by Theorem 2 for $P = 2$ and subsequently point out how one may use the theorem to obtain the steady-state joint queue-length distribution in that example for a P-class queue.

Consider the $M/G/1$ queue with P classes of customers, with nonpreemptive priority in descending order $1, 2, \ldots, P$ (so class 1 has the highest priority). Let λ_i denote the arrival rate of customers of class i, $i = 1, 2$. Takagi ([18], Formula (2.87) on p. 311) presents the PGF $\Pi(z_1, z_2, \ldots, z_P)$ of the steady-state joint queue-length distribution immediately after an arbitrary customer departure epoch. For $P = 2$ he also obtains the PGF $P(z_1, z_2, \ldots, z_P)$ of the steady-state joint queue-length distribution at an arbitrary epoch ([18], Formula (5.82b) on p. 397). We have verified that, indeed, for $P = 2$ classes one has (cf. Theorem 2 with single arrivals),

$$(\lambda_1(1 - z_1) + \lambda_2(1 - z_2))P(z_1, z_2) = \lambda_1(1 - z_1)S_1^c(z_1, z_2) + \lambda_2(1 - z_2)S_2^c(z_1, z_2).$$

The starting point for this verification was the following obvious set of relations, with $\beta_i(z_1, z_2)$ the PGF of the numbers of arrivals at both queues during one service of a class i customer, $i = 1, 2$:

$$\begin{aligned} \Pi_1(z_1, z_2) &:= \frac{\lambda_1}{\lambda} S_1^c(z_1, z_2) = \frac{\Pi(z_1, z_2) - \Pi(0, z_2)}{z_1} \beta_1(z_1, z_2) \\ &\quad + \Pi(0, 0) \frac{\lambda_1}{\lambda} \beta_1(z_1, z_2), \end{aligned} \tag{37}$$

$$\begin{aligned} \Pi_2(z_1, z_2) &:= \frac{\lambda_2}{\lambda} S_2^c(z_1, z_2) = \frac{\Pi(0, z_2) - \Pi(0, 0)}{z_2} \beta_2(z_1, z_2) \\ &\quad + \Pi(0, 0) \frac{\lambda_2}{\lambda} \beta_2(z_1, z_2). \end{aligned} \tag{38}$$

Here $\Pi_i(z_1, z_2)$ is the PGF of the steady-state joint queue-length distribution immediately after the departure of a class i customer, with the indicator function 1(departing customer is of class i), $i = 1, 2$, and $\Pi(z_1, z_2)$ is as defined above. The factors $\frac{\lambda_i}{\lambda}$ on the left-hand side of (37) and (38) are needed because the $S_i^c(z_1, z_2)$ are conditional PGFs, the condition being that the departing customer is of class i.

This example clearly demonstrates the value of our general balance equations. Besides providing a much shorter proof for Takagi's Formula (5.82b) they also allow us to extend his result to the case of $P(> 2)$ customer classes, by using the expressions for $\frac{\lambda_i}{\lambda} S_i^c(z_1, z_2, \ldots, z_P)$ that follow from Takagi's Formula (2.87) for $\Pi(z_1, z_2, \ldots, z_P)$.

Case 4: *Priority for the longer queue*

Consider a model of one server and two queues. Each queue has its own Poisson arrival process and service time distribution. After a service completion the server proceeds with a customer from the longest queue, if the queue lengths are unequal; if the queue lengths are equal, the server chooses a customer from queue Q_i with probability α_i, $i = 1, 2$. Cohen [6] has derived the PGF $\Pi(z_1, z_2) = \mathbb{E}[z_1^{X_1} z_2^{X_2}]$ of the steady-state joint queue-length distribution immediately after an arbitrary customer departure epoch, by solving a Riemann-type boundary value problem. In the process he also obtained the following PGFs that naturally arise in this *Priority for the longer queue* model: $\mathbb{E}[z_1^{X_1} z_2^{X_2} 1_{\{X_1 > X_2\}}]$, $\mathbb{E}[z_1^{X_1} z_2^{X_2} 1_{\{X_1 < X_2\}}]$ and $\mathbb{E}[z_1^{X_1} z_2^{X_2} 1_{\{X_1 = X_2 > 0\}}]$. Below we first show how one can obtain the PGFs $\Pi_i(z_1, z_2)$ of the steady-state joint queue-length distribution immediately after the departure of a customer from Q_i, $i = 1, 2$ (we stick as much as possible to the notation of Case 3). By considering the joint queue-length distribution at two consecutive departure epochs, and with $\beta_i(z_1, z_2)$ denoting the PGF of the numbers of arrivals at both queues during one service of a customer from Q_i, we can write

$$\begin{aligned}
\Pi_1(z_1, z_2) &= \mathbb{E}\left[z_1^{X_1} z_2^{X_2} 1_{\{X_1 > X_2\}}\right] \frac{\beta_1(z_1, z_2)}{z_1} \\
&\quad + \alpha_1 \mathbb{E}\left[z_1^{X_1} z_2^{X_2} 1_{\{X_1 = X_2 > 0\}}\right] \frac{\beta_1(z_1, z_2)}{z_1} \\
&\quad + \mathbb{P}(X_1 = X_2 = 0) \frac{\lambda_1}{\lambda_1 + \lambda_2} \beta_1(z_1, z_2), \qquad (39) \\
\Pi_2(z_1, z_2) &= \mathbb{E}\left[z_1^{X_1} z_2^{X_2} 1_{\{X_1 < X_2\}}\right] \frac{\beta_2(z_1, z_2)}{z_2} \\
&\quad + \alpha_2 \mathbb{E}\left[z_1^{X_1} z_2^{X_2} 1_{\{X_1 = X_2 > 0\}}\right] \frac{\beta_2(z_1, z_2)}{z_2} \\
&\quad + \mathbb{P}(X_1 = X_2 = 0) \frac{\lambda_2}{\lambda_1 + \lambda_2} \beta_2(z_1, z_2). \qquad (40)
\end{aligned}$$

The queue-length PGFs on the two right-hand sides are derived by Cohen [6], and thus we obtain $\Pi_i(z_1, z_2)$, $i = 1, 2$. This immediately leads to $S_i^c(z_1, z_2)$, $i = 1, 2$, as in Case 3. Subsequently Theorem 2 gives the PGF of the steady-state joint queue-length distribution at an arbitrary epoch. It should be noticed that it is not at all easy to obtain this PGF in another way, for this non-Markovian model; the *priority for the longer queue* model is a difficult queueing model. In the case of exponential service time distributions, with equal arrival and service rates at the two queues and $\alpha_1 = \alpha_2$, Zheng and Zipkin [20] present a recursive method to obtain this PGF, while Flatto [10] for this case (but allowing preemption) obtains the queue-length PGF by solving a boundary value problem.

Case 5: *A simple network*

Consider a network of m service facilities, with independent external Poisson arrival processes, and with continuous service time distributions. We have Markovian routing, a customer moving from Q_i to Q_k with probability p_{ik} and leaving the system after its service completion in Q_i with probability p_{i0}, $i, k \in J$. Define Λ_i as the total flow through Q_i per time unit, $i \in J$; these Λ_i are the unique solution of the set of equations

$$\Lambda_i = \lambda_i + \sum_{k=1}^{m} \Lambda_k p_{ki}, \quad i \in J. \tag{41}$$

Let A_i indicate that the system is viewed just before an arrival at Q_i, D_i that the system is viewed just after a departure from Q_i and I_{ik} that the system is viewed just after a departure from Q_i and just before the arrival of the departing customer at Q_k. Letting $\boldsymbol{j} = (j_1, j_2, \ldots, j_m)$, one can write down the following balance equations for the queue length vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_m)$:

$$\begin{aligned}
&\sum_{i=1}^{m} \lambda_i \mathbb{P}(\boldsymbol{X} = \boldsymbol{j} | A_i) + \sum_{i=1}^{m} \Lambda_i p_{i0} \mathbb{P}(\boldsymbol{X} = \boldsymbol{j} - \boldsymbol{e}_i | D_i) \\
&\quad + \sum_{i=1}^{m} \sum_{k=1}^{m} \Lambda_i p_{ik} \mathbb{P}(\boldsymbol{X} = \boldsymbol{j} - \boldsymbol{e}_i | I_{ik}) \\
&= \sum_{i=1}^{m} \Lambda_i p_{i0} \mathbb{P}(\boldsymbol{X} = \boldsymbol{j} | D_i) + \sum_{i=1}^{m} \lambda_i \mathbb{P}(\boldsymbol{X} = \boldsymbol{j} - \boldsymbol{e}_i | A_i) \\
&\quad + \sum_{i=1}^{m} \sum_{k=1}^{m} \Lambda_i p_{ik} \mathbb{P}(\boldsymbol{X} = \boldsymbol{j} - \boldsymbol{e}_k | I_{ik}).
\end{aligned} \tag{42}$$

The (PGF of the) probabilities, given that we observe just after a real departure from Q_i or that we observe just after a departure from Q_i that will in an instant result in an arrival at Q_k, are obviously the same. If one takes PGFs, one quickly sees that a special case of (33) is obtained.

We now use (42) to provide an alternative proof for the joint queue-length distribution in a queueing network with a single roving server as studied in [2,17]. Again consider a network of m queues with Markovian customer routing, as described above. In this particular example we assume that a *single* server visits the queues in a fixed, cyclic order, requiring a switchover time S_i to move from Q_i to Q_{i+1}. We do not make any assumptions regarding the service disciplines at each queue. This model, which can be regarded as a polling model with customer routing, has been studied by Sidi, Levy and Fuhrmann [17], who refer to this model as a queueing network with a roving server. Sidi et al. obtain the joint queue-length distribution at arbitrary moments, as well as the joint queue-length distribution at departure epochs. The waiting-time distributions are obtained in a different paper [2]. For us it is slightly more convenient to refer to this latter paper in the analysis as described, because the authors in [2] use the same definition of $V_i^c(z)$, the PGF of the joint queue-length at departure epochs,

just after a departure from Q_i and just *before* the arrival of the departing customer at the next queue.

Take the formulas (3.2)–(3.6) of [2]. From (3.2), which is the counterpart of our (2), one can express (in the notation of the present paper) the differences of PGFs at visit beginning and visit completion epochs into those at service beginning and service completion epochs:

$$\frac{V_i^b(z) - V_i^c(z)}{\Lambda_i \mathbb{E}C} = S_i^b(z) - S_i^c(z)P_i(z), \quad i = 1, 2, \ldots, m. \tag{43}$$

Here $P_i(z) := p_{i0} + \sum_{k=1}^m p_{ik} z_k$, and $\mathbb{E}C = s/(1-\rho)$ with $\rho := \sum_{i=1}^m \Lambda_i b_i$. Next, use our relation (3) to express $S_i^b(z)$ in terms of $S_i^c(z)$. Subsequently express $L(z)$, in (3.4) of [2], which is the counterpart of (1) above, in terms of differences $V_i^b(z) - V_i^c(z)$, as was also done in [4]. This gives

$$\sum_{i=1}^m \lambda_i (1 - z_i) L(z) = \sum_{i=1}^m \Lambda_i (P_i(z) - z_i) S_i^c(z). \tag{44}$$

This is indeed in agreement with (42): The left-hand side of (44) gives the first and the fifth term on (42). The last term on the right-hand side gives the second plus the third term in (42), once we realize that $p_{i0} + \sum_{k=1}^m p_{ik} = 1$, and that the conditional probabilities both refer to a service completion in Q_i, no matter whether the condition is D_i or I_{ik}. The first term on the right-hand side gives the fourth plus fifth term in (42). One could argue that some results in [2] and [17] could have been derived faster by starting from (44).

6 Concluding remarks

This paper derives a distributional relationship, at different embedded epochs, for analyzing queues and their networks. As shown in Sect. 3, it has different forms according to the abstraction level of the model. This may both lead to new results and easier derivations of some known results. In Sect. 5 this is demonstrated for a few examples.

The relationship in Sect. 4 has a different nature than the rest of this paper because it does not require any stationarity of the processes of interest. Namely, it suggests that such an asymptotic relationship may enable us to obtain queueing characteristics with some error bounds, not assuming any stationarity condition. This is completely different from the standard analysis in queueing theory. Thus, it would be interesting to see whether it can yield useful results for the performance evaluation of queueing models. We leave this for future studies.

Acknowledgements We are grateful to a referee for providing useful references and insightful remarks. Funding was provided by Israel Science Foundation (Grant No. 1462/13), Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO Gravitation Program NETWORKS, Grant No. 024.002.003), and Japan Society for the Promotion of Science (Grant No. 16H027860001).

Appendix

In the appendices we omit "a.s." because countably many events, each of which occurs w.p. 1, simultaneously occur w.p. 1.

Proof of Lemma 1

Since the proofs of (17) and (18) are similar, we only prove (17). Since π^e is a probability distribution, we can choose a sufficiently large a for each $\epsilon > 0$ such that

$$\sum_{\max(|\boldsymbol{x}|,|\boldsymbol{y}|)\geq a} \pi^e(\boldsymbol{x},\boldsymbol{y}) < \epsilon.$$

Let $\mathcal{S}_a = \{(\boldsymbol{x},\boldsymbol{y}) \in \mathbb{Z}_+^{2m}; \max(|\boldsymbol{x}|,|\boldsymbol{y}|) < a\}$, then $\mathcal{S}_a$ is a finite set. Hence summing both sides of (15) for $(\boldsymbol{x},\boldsymbol{y}) \in \mathcal{S}_a$ yields

$$\lim_{n\to\infty} \frac{1}{n}\sum_{\ell=1}^{n} \sum_{(\boldsymbol{x},\boldsymbol{y})\in\mathcal{S}_a} 1\left(\boldsymbol{X}(t_\ell^e-) = \boldsymbol{x}, \Delta\boldsymbol{N}^e(t_\ell^e) = \boldsymbol{y}\right) = \sum_{(\boldsymbol{x},\boldsymbol{y})\in\mathcal{S}_a} \pi^e(\boldsymbol{x},\boldsymbol{y}),$$

and therefore,

$$\begin{aligned}
&\lim_{n\to\infty} \frac{1}{n}\sum_{\ell=1}^{n} \sum_{(\boldsymbol{x},\boldsymbol{y})\notin\mathcal{S}_a} 1(\boldsymbol{X}(t_\ell^e-) = \boldsymbol{x}, \Delta\boldsymbol{N}^e(t_\ell^e) = \boldsymbol{y}) \\
&= 1 - \lim_{n\to\infty} \frac{1}{n}\sum_{\ell=1}^{n} \sum_{(\boldsymbol{x},\boldsymbol{y})\in\mathcal{S}_a} 1(\boldsymbol{X}(t_\ell^e-) = \boldsymbol{x}, \Delta\boldsymbol{N}^e(t_\ell^e) = \boldsymbol{y}) \\
&= 1 - \sum_{(\boldsymbol{x},\boldsymbol{y})\in\mathcal{S}_a} \pi^e(\boldsymbol{x},\boldsymbol{y}) = \sum_{\max(|\boldsymbol{x}|,|\boldsymbol{y}|)\geq a} \pi^e(\boldsymbol{x},\boldsymbol{y}) < \epsilon. \qquad (45)
\end{aligned}$$

Multiplying both sides of (15) by $g(\boldsymbol{x},\boldsymbol{y})$ and summing them for $(\boldsymbol{x},\boldsymbol{y}) \in \mathcal{S}_a$ yields

$$\lim_{n\to\infty} \frac{1}{n}\sum_{\ell=1}^{n} \sum_{(\boldsymbol{x},\boldsymbol{y})\in\mathcal{S}_a} g(\boldsymbol{x},\boldsymbol{y}) 1(\boldsymbol{X}(t_\ell^e-) = \boldsymbol{x}, \Delta\boldsymbol{N}^e(t_\ell^e) = \boldsymbol{y}) = \sum_{(\boldsymbol{x},\boldsymbol{y})\in\mathcal{S}_a} g(\boldsymbol{x},\boldsymbol{y})\pi^e(\boldsymbol{x},\boldsymbol{y}).$$

Let $\|g\| = \sup_{\boldsymbol{x},\boldsymbol{y}} g(\boldsymbol{x},\boldsymbol{y})$, which is finite by assumption. Since (45) implies that

$$\limsup_{n\to\infty}\frac{1}{n}\sum_{\ell=1}^{n}\sum_{(\boldsymbol{x},\boldsymbol{y})\notin\mathcal{S}_a} g(\boldsymbol{x},\boldsymbol{y})1(\boldsymbol{X}(t_\ell^e-)=\boldsymbol{x},\Delta\boldsymbol{N}^e(t_\ell^e)=\boldsymbol{y})$$
$$\le \|g\|\limsup_{n\to\infty}\frac{1}{n}\sum_{\ell=1}^{n}\sum_{(\boldsymbol{x},\boldsymbol{y})\notin\mathcal{S}_a} 1(\boldsymbol{X}(t_\ell^e-)=\boldsymbol{x},\Delta\boldsymbol{N}^e(t_\ell^e)=\boldsymbol{y}) < \|g\|\epsilon,$$
$$\sum_{(\boldsymbol{x},\boldsymbol{y})\notin\mathcal{S}_a} g(\boldsymbol{x},\boldsymbol{y})\pi^e(\boldsymbol{x},\boldsymbol{y}) < \|g\|\epsilon,$$

we have

$$\limsup_{n\to\infty}\Big|\frac{1}{n}\sum_{\ell=1}^{n}\sum_{\boldsymbol{x},\boldsymbol{y}} g(\boldsymbol{x},\boldsymbol{y})1(\boldsymbol{X}(t_\ell^e-)=\boldsymbol{x},\Delta\boldsymbol{N}^e(t_\ell^e)=\boldsymbol{y}) - \sum_{\boldsymbol{x},\boldsymbol{y}} g(\boldsymbol{x},\boldsymbol{y})\pi^e(\boldsymbol{x},\boldsymbol{y})\Big| < 2\|g\|\epsilon.$$

Letting $\epsilon\downarrow 0$, we arrive at (17).

Proof of Lemma 2

In view of Lemma 1, to prove (28) it suffices to prove that, for $A\in 2^J\setminus\{\emptyset\}$,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{\ell=1}^{n}1(\boldsymbol{X}(t_{A,\ell}^d-)=\boldsymbol{x},\Delta\boldsymbol{N}^d(t_{A,\ell}^d)=\boldsymbol{y},\Delta\boldsymbol{N}^r(t_{A,\ell}^d)=\boldsymbol{z})=\pi_A^d(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}). \tag{46}$$

It follows from (v) that, for each $i\in J,\ell\ge 1,\boldsymbol{y}\in S_A,\boldsymbol{z}\in\mathbb{Z}_+^m$, there is a unique $k\ge 1$ such that $\ell\le k$ and

$$1(\boldsymbol{X}(t_{A,\ell}^d-)=\boldsymbol{x},\Delta\boldsymbol{N}^d(t_{A,\ell}^d)=\boldsymbol{y},\Delta\boldsymbol{N}^d(t_{A,\ell}^d)=\boldsymbol{z})$$
$$=1(\boldsymbol{X}(t_k^d-)=\boldsymbol{x},\Delta\boldsymbol{N}^d(t_k^d)=\boldsymbol{y},\Delta\boldsymbol{N}^d(t_k^d)=\boldsymbol{z}),$$

and (14) and (25) imply

$$\lim_{t\to\infty}\frac{\tilde{N}_A^d(t)}{|\tilde{\boldsymbol{N}}^d|(t)}=\lim_{t\to\infty}\frac{\frac{1}{t}\tilde{N}_A^d(t)}{\frac{1}{t}|\tilde{\boldsymbol{N}}^d|(t)}=\frac{\lambda_A^d}{\lambda^d}.$$

Hence, for $\boldsymbol{y} \in S_A$,

$$
\begin{aligned}
&\pi^d(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} 1\left(\boldsymbol{X}(t_k^d-) = \boldsymbol{x}, \Delta\boldsymbol{N}^d(t_k^d) = \boldsymbol{y}, \Delta\boldsymbol{N}^r(t_k^d) = \boldsymbol{z}\right) \\
&= \lim_{t\to\infty} \frac{1}{|\tilde{\boldsymbol{N}}^d|(t)} \sum_{k=1}^{|\tilde{\boldsymbol{N}}^d|(t)} 1\left(\boldsymbol{X}(t_k^d-) = \boldsymbol{x}, \Delta\boldsymbol{N}^d(t_k^d) = \boldsymbol{y}, \Delta\boldsymbol{N}^r(t_k^d) = \boldsymbol{z}\right) \\
&= \lim_{t\to\infty} \frac{\tilde{N}_A^d(t)}{|\tilde{\boldsymbol{N}}^d|(t)} \frac{1}{\tilde{N}_A^d(t)} \sum_{\ell=1}^{\tilde{N}_A^d(t)} 1\left(\boldsymbol{X}(t_{A,\ell}^d-) = \boldsymbol{x}, \Delta\boldsymbol{N}^d(t_{A,\ell}^d) = \boldsymbol{y}, \Delta\boldsymbol{N}^r(t_{A,\ell}^d) = \boldsymbol{z}\right) \\
&= \frac{\lambda_A^d}{\lambda^d} \lim_{n\to\infty} \frac{1}{n} \sum_{\ell=1}^{n} 1\left(\boldsymbol{X}(t_{A,\ell}^d-) = \boldsymbol{x}, \Delta\boldsymbol{N}^d(t_{A,\ell}^d) = \boldsymbol{y}, \Delta\boldsymbol{N}^r(t_{A,\ell}^d) = \boldsymbol{z}\right).
\end{aligned}
$$

This proves (46) by the definition π_A^d, and therefore (28) holds. The fact that π_A^d is a probability distribution is immediate from (28) with $h(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \equiv 1$.

References

1. Baccelli, F., Brémaud, P.: Elements of Queueing Theory: Palm Martingale Calculus and Stochastic Recurrences, vol. 26 of Applications of Mathematics, 2nd edn. Springer, Berlin (2003)
2. Boon, M.A.A., van der Mei, R.D., Winands, E.M.M.: Waiting times in queueing networks with a single shared server. Queueing Syst. **74**, 403–429 (2013)
3. Borst, S.C., Boxma, O.J.: Polling models with and without switchover times. Oper. Res. **45**, 536–543 (1997)
4. Boxma, O.J., Kella, O., Kosinski, K.M.: Queue lengths and workloads in polling systems. Oper. Res. Lett. **39**, 401–405 (2011)
5. Boxma, O.J., Takine, T.: The $M/G/1$ FIFO queue with several customer classes. Queueing Syst. **45**, 185–189 (2003)
6. Cohen, J.W.: A two-queue, one-server model with priority for the longer queue. Queueing Syst. **2**, 261–283 (1987)
7. Cooper, R.B.: Introduction to Queueing Theory. Macmillan, New York (1972)
8. Eisenberg, M.: Queues with periodic service and changeover time. Oper. Res. **20**, 440–451 (1972)
9. Fakinos, D.: The relation between limiting queue size distributions at arrival and departure epochs in a bulk queue. Stoch. Proces. Appl. **37**, 327–329 (1991)
10. Flatto, L.: The longer queue model. Probab. Eng. Inf. Sci. **3**, 537–559 (1989)
11. Hébuterne, G.: Relation between states observed by arriving and departing customers in bulk systems. Stoch. Proces. Appl. **27**, 279–289 (1988)
12. Hébuterne, G., Rosenberg, C.: Arrival and departure state distributions in the general bulk-service queue. Nav. Res. Logist. **46**, 107–118 (1999)
13. Kim, K.: A relation between queue-length distributions during server vacations in queues with batch arrivals, batch services, or multiclass arrivals: an extension of Burke's theorem. Indian J. Sci. Technol. **8**, 1–5 (2015)
14. Miyazawa, M.: Palm calculus, reallocatable GSMP and insensitivity structure, chap. 4 of Queueing networks: A fundamental approach. International Series in Operations Research and Management Science, pp. 141–215. Springer (2010)
15. Miyazawa, M.: Rate conservation laws: a survey. Queueing Syst. **15**, 1–58 (1994)

16. Papaconstantinou, X., Bertsimas, D.: Relations between the pre-arrival and the post-departure state probabilities and the FCFS waiting time distribution in the $E_k/G/s$ queue. Nav. Res. Logist. **37**, 135–149 (1990)
17. Sidi, M., Levy, H., Fuhrmann, S.W.: A queueing network with a single cyclically roving server. Queueing Syst. **11**, 121–144 (1992)
18. Takagi, H.: Queueing Analysis. A Foundation of Performance Evaluation. Volume 1: Vacation and Priority Systems. North-Holland Publications, Amsterdam (1991)
19. Takine, T.: Distributional form of Little's law for FIFO queues with multiple Markovian arrival streams and its applications to queues with vacations. Queueing Syst. **37**, 31–63 (2001)
20. Zheng, Y.-S., Zipkin, P.: A queueing model to analyze the value of centralized inventory information. Oper. Res. **38**, 296–307 (1990)