EURANDOM PREPRINT SERIES

2016-004

March 14, 2016

An ASIP model with general gate opening intervals

O. Boxma, O. Kella, U. Yechiali
ISSN 1389-2355

An ASIP model with general gate opening intervals

Onno Boxma* , Offer Kella† , and Uri Yechiali‡

March 14, 2016

Abstract

We consider an Asymmetric Inclusion Process (ASIP), which can also be viewed as a model of n queues in series. Each queue has a gate behind it, which can be seen as a server. When a gate opens, all customers in the corresponding queue instantaneously move to the next queue and form a cluster with the customers there. When the nth gate opens, all customers in the nth site leave the system. For the case that the gate openings are determined by a Markov renewal process, and for a quite general arrival process of customers at the various queues during intervals between successive gate openings, we obtain the following results: (i) steady-state distribution of the total number of customers in the first k queues, $k = 1, \ldots, n$; (ii) steady-state joint queue length distributions for the two-queue case. In addition to the case that the numbers of arrivals in successive gate opening intervals are independent, we also obtain explicit results for a two-queue model with renewal arrivals.
Keywords: Asymmetric inclusion process; tandem network; synchronized service; queue length distribution.

1 Introduction

The ASIP (Asymmetric Inclusion Process), introduced and analyzed in [5]-[9] is a one-dimensional lattice of n sites (queues), where particles (e.g. customers) arrive randomly into the first site (Q_1), stay there ('served') for a random time, continue moving simultaneously and uni-directionally from site to site while staying a random time in each site, until finally exiting the last site (Q_n) and leaving the system. The ASIP defines the missing link between the celebrated Tandem Jackson Network (TJN) and the Asymmetric Exclusion Process (ASEP) [1, 2, 3] which plays the role of a paradigm in non-equilibrium statistical mechanics. Imagine that each site has a gate behind it that opens every exponentially distributed random time, allowing particles in the site to move forward to the next site. Denoting by C_{capacity} the capacity of a site (i.e. maximal number of particles that can reside in the site), and by C_{gate} the capacity of the site's gate (i.e., the maximal number of particles that can move forward

*EURANDOM and Department of Mathematics and Computer Science, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands (o.j.boxma@tue.nl); research done in the framework of the IAP BESTCOM project, funded by the Belgian government and by the Gravity program NETWORKS of the Dutch government

†Department of Statistics, The Hebrew University of Jerusalem, Mount Scopus, Jerusalem 91905, Israel (offer.kella@huji.ac.il); supported in part by grant 1462/13 from the Israel Science Foundation and the Vigevani Chair in Statistics

‡Department of Statistics and Operations Research, School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel (uriy@post.tau.ac.il)

when the gate opens), then, for the TJN, $C_{\text{capacity}} = \infty$ and $C_{\text{gate}} = 1$, while for the ASEP, $C_{\text{capacity}} = C_{\text{gate}} = 1$. When $C_{\text{capacity}} = C_{\text{gate}} = \infty$, one obtains the ASIP model where, when a gate of a site opens, then *all* particles (customers) present there move simultaneously to the next site, joining the particles already there and forming a cluster of particles that continues moving as one unit. In the present work we generalize the ASIP model by assuming that gate openings are determined by a Markov renewal process such that if, at some time, gate i opens, then with probability p_{ij} the next gate to open is gate j, and the time until that gate opens is a random variable O_{ij}. We derive the Probability Generating Function (PGF) of the total occupancy (i.e., total number of customers) of sites 1 to k ($k = 1, 2, \ldots, n$), while further studying the case when $p_{ij} = q_j$. We obtain the joint queue length distribution for the two-queue case, and analyze the system assuming Binomial movement of particles. That is, when gate i (say) opens, each particle from the X_i particles present in site i (Q_i) will move forward (independently of the other particles) to site $i + 1$ (Q_{i+1}) with probability a_i, such that the total number of particles moving from Q_i to Q_{i+1} is Binomially distributed with parameters X_i and a_i.
The ASIP model, first studied in [6], presents a one-dimensional lattice of n queues with Poissonian flow only into the first site. Each gate opens, distinctly and independently of the others, every Exponential time with rate μ_k for site k, implying that O_{ij} is the same for all (i, j) combinations, and Exponentially distributed with rate $\sum \mu_k$, while $p_{ij} = \mu_j / \sum \mu_k$, $i, j = 1, 2, \ldots, n$. The multi-dimensional PGF of the occupancy vector $(X_1, \ldots, X_n)$ was studied, and it was shown that this PGF does not exhibit the famous product-form solution characterizing Jackson Networks. Accordingly, an iterative solution procedure was developed. However, the PGF of the total-load-up-to-site k, $k = 1, 2, \ldots, n$, was shown to have a product-form solution of Geometric variables. For various objective functions it was shown that the optimal intensities of the gate openings should be equal to each other. Considering large size ASIP, it was observed in [8], via simulations, that $P(X_k > 0) \sim k^{-1/2}$, $E[X_k | X_k > 0] \sim k^{1/2}$; and that (standard deviation of X_k)$/E[X_k] \sim k^{1/4}$. Those observations were later proved analytically in [7], where limit laws when $n \to \infty$ were derived. Various measures were investigated: (i) a particle's (customer's) traversal time, T, in a homogeneous ASIP, is distributed as $T \sim nm + n^{1/2} m Z$, where m = mean time between successive gate openings, m^2 is its variance, and Z is the Gaussian $(0, 1)$ random variable. (ii) The Laplace-Stieltjes Transform (LST) and mean of the Busy Period (the time from the first arrival of a customer at an empty system until the first moment thereafter that the network becomes empty again). (iii) The LST and mean of the Draining time (the time from an arbitrary moment when the system is in steady state and the inflow is stopped, until the first moment thereafter that the system becomes empty). Occupation probabilities were considered in [9]. Closed-form results were obtained for the probabilities that the total occupation of 'lattice intervals' of m sites, sites k to $k + m - 1$, is equal to l, $l = 0, 1, 2, \ldots$. In particular, when $l = 0$, the problem becomes a discrete boundary value problem and the probabilities are derived with the aid of Catalan numbers.

The main contribution of this paper is that it considerably extends the exact analysis of ASIP tandem models: We allow the gate openings to be determined by a Markov renewal process, instead of assuming that each gate opens after exponentially distributed intervals, and we extend the Poisson arrival assumption by allowing a quite general arrival process of customers at the various queues during intervals between successive gate openings. Under these assumptions, we determine the steady-state distribution of the total number of customers in the first k queues, $k = 1, \ldots, n$. We obtain some additional results for the 2-queue case, and solve three optimization problems, thus obtaining insight into the design of ASIPs.

The paper is organized as follows. Section 2 contains the model description. Section 3 is devoted to the analysis of the steady-state joint distribution of the numbers of customers in the various queues just after gate openings. The section is ended by a brief discussion on optimization of the system performance. A few more detailed two-queue results are presented in Section 4. We conclude with some suggestions for further research in Section 5.

2 Model Description

Consider the following model of n queues $Q_1, \ldots, Q_n$ in series. Each queue has one gate behind it, which may be viewed as a server. Gates are almost all the time closed. When gate i (the gate behind Q_i) opens, all customers present in Q_i are instantaneously transferred to Q_{i+1}, $i = 1, 2, \ldots, n-1$; when gate n opens, all customers present in Q_n instantaneously leave the system. After the transfer, the gate immediately closes again. Gate openings are determined by a Markov renewal process. If, at some time t, gate i opens, then with probability p_{ij} the next gate to open is gate j; and the time until that gate opens is a random variable O_{ij} with distribution $O_{ij}(\cdot)$ and LST $\omega_{ij}(\cdot)$. We assume that the Markov chain governing the successive gate openings is irreducible and we denote its steady-state distribution by π_i, $i = 1, \ldots, n$.

During an O_{ij} period, customers may arrive at all queues. We assume that the vectors of arrival numbers in successive gate opening intervals are independent, but may depend on the indices i and j. The generating function of the numbers of arrivals into $Q_1, \ldots, Q_n$ during an O_{ij} period is given by $A_{ij}(z_1, \ldots, z_n)$. In addition, we denote the generating function of the cumulative number of arrivals into $Q_1, \ldots, Q_k$ during an O_{ij} period by $A_{ijk}(z) := A_{ij}(z, \ldots, z, 1, \ldots, 1)$, where the last z occurs at position k. Notice that one example is provided by a batch Poisson arrival process, possibly with dependence between batch sizes at different queues, and with arrival rates which may depend on the type of gate opening interval.

3 Analysis

We are interested in the steady-state joint distribution of the numbers of customers $(X_1, \ldots, X_n)$ just after a gate opening. To argue the existence of such a distribution, let $\xi_k = (\xi_{k1}, \ldots, \xi_{kn})$ be the state of the network right after the kth gate opening and let q_k be the gate that opened. Then, because the external arrival process is independent of the process of the gate openings, (ξ_k, q_k) is a Markov chain. To argue that it is positive recurrent (on an appropriate state space), let us define an auxiliary process as follows. Let $\eta_{ki} = 1_{\{\xi_{ki} \geq 1\}}$. Then (η_k, q_k) is also a Markov chain and ξ_k is the zero vector if and only if η_k is. We note that for every station j, the state $(0, j)$ is accessible from every other state. This is because if we block external arrivals, the time until the network becomes empty is finite (actually has a finite expected time). When this happens, we are in some state $(0, \ell)$. Since q_k is irreducible, then, if we once again block all arrivals the state $(0, j)$ is accessible from $(0, \ell)$ (actually, without arrivals, it will also be reached after finite expected time). With positive probability the time until the first arrival is greater than the (independent) time to reach $(0, j)$ without arrivals and thus $(0, j)$ is accessible from any other state. Thus, on the states (y, j) which are accessible from $(0, j)$ (which include $(0, \ell)$ for all $1 \leq \ell \leq n$ and all states that are accessible from $(0, \ell)$ for any such ℓ), we have that (η_k, q_k) is an irreducible Markov chain and since the state space is finite (contained in or equal to $\{0,1\}^n \times \{1,\dots,n\}$) it follows that it is positive recurrent. Therefore, for any j, the time between visits to state $(0,j)$ has a finite mean. This implies that the time between visits of (ξ_k, q_k) to $(0,j)$ also has a finite mean and thus the (ξ_k, q_k) is also positive recurrent on an appropriate state space (all the states which are accessible from $(0,j)$ for some, hence all, j). We note that this idea can be used to argue stability for the continuous time process, which although it is not semi-Markov due to the arrival process, is nevertheless regenerative with finite mean regeneration epochs, provided that O_{ij} have finite means. Since we do not need it here, we omit the details.

In the present section we shall in particular focus on $X_{(k)} := X_1 + \cdots + X_k$, viz., the total number of customers in the first k queues right after a gate opening. Introducing M, the index of the gate that has just opened, we consider

$$G_{ki}(z) := \mathbb{E}[z^{X_{(k)}} I(M=i)], \quad k,i = 1,\dots,n, \tag{1}$$

where $I(\cdot)$ denotes an indicator function. The fact that customers can only move to downstream queues (i.e., with higher index), will allow us to express all $G_{ki}(z)$ for a fixed k into functions $G_{k-1,j}(z)$, and finally into the functions $G_{1j}(z)$, which can be determined explicitly.

We begin by giving the equations for $G_{1j}(z)$, $j = 1,\dots,n$. Obviously

$$G_{11}(z) = \mathbb{P}(M=1) = \pi_1; \tag{2}$$

indeed, after gate 1 has opened, Q_1 instantaneously has become empty. Now consider two successive gate openings in steady state, the latter one being an opening of gate j, and sum over all possible gates i opened at the previous gate opening, to obtain:

$$G_{1j}(z) = \sum_{i=1}^{n} G_{1i}(z) p_{ij} A_{ij1}(z), \quad j \neq 1. \tag{3}$$

Here we have used that $A_{ij1}(z)$ is the generating function of the number of arrivals at Q_1 in the gate opening interval.

Notice that we can rewrite (3) as

$$G_{1j}(z) = \sum_{i=2}^{n} G_{1i}(z) p_{ij} A_{ij1}(z) + G_{11}(z) p_{1j} A_{1j1}(z), \quad j \neq 1. \tag{4}$$

Introducing the $(n-1)$-dimensional vector

$$\bar{G}_1(z) := (G_{12}(z), \dots, , G_{1n}(z)),$$

the $(n-1)$-dimensional vector

$$R_1(z) := (p_{12} A_{121}(z), \dots, p_{1n} A_{1n1}(z)),$$

and the matrix $P_1(z)$ with as (i,j) element $p_{ij} A_{ij1}(z)$, we can write (4) as

$$\bar{G}_1(z) = \bar{G}_1(z) P_1(z) + G_{11}(z) R_1(z), \tag{5}$$

and hence, with I the matrix with ones on the diagonal and zeroes outside the diagonal, we have $\bar{G}_1(z)(I - P_1(z)) = G_{11}(z) R_1(z)$, yielding

$$\bar{G}_1(z) = G_{11}(z) R_1(z) (I - P_1(z))^{-1}. \tag{6}$$

All the terms in the righthand side of (6) are known; in particular, $G_{11}(z) = \pi_1$ is given in (2). Hence we have determined $G_{11}(z), G_{12}(z), \ldots, G_{1n}(z)$.

Now let us show how the terms $G_{kj}(z)$, $j = 1, \ldots, n$, are for $2 \leq k \leq n$ expressed into $G_{k-1,i}(z)$, $i = 1, \ldots, n$. Considering two successive gate openings in steady state, the last one being of gate j, and summing over all possible gates i for the first gate opening, we have for $k = 2, \ldots, n$, $j \neq k$:

$$G_{kj}(z) = \sum_{i=1}^{n} G_{ki}(z) p_{ij} A_{ijk}(z), \tag{7}$$

whereas

$$G_{kk}(z) = \sum_{i=1}^{n} G_{k-1,i}(z) p_{ik} A_{ik,k-1}(z). \tag{8}$$

The explanation for the deviating terms ($G_{k-1,i}(z)$ instead of $G_{ki}(z)$ and $A_{ik,k-1}(z)$ instead of $A_{ikk}(z)$) is that Q_k has become empty right after an opening of gate k; so the total number present in $Q_1, \ldots, Q_k$ equals the total number present in $Q_1, \ldots, Q_{k-1}$ after the previous gate opening, plus the number of new arrivals in the first $k-1$ queues.

We can rewrite (7) as follows:

$$G_{kj}(z) = \sum_{i \neq k} G_{ki}(z) p_{ij} A_{ijk}(z) + G_{kk}(z) p_{kj} A_{kjk}(z). \tag{9}$$

Introducing the $(n-1)$-dimensional vector

$$\bar{G}_k(z) := (G_{k1}(z), \ldots, G_{k,k-1}(z), G_{k,k+1}(z), \ldots, G_{kn}(z)),$$

the $(n-1)$-dimensional vector

$$R_k(z) := (p_{k1} A_{k1k}(z), \ldots, p_{k,k-1} A_{k,k-1,k}(z), p_{k,k+1} A_{k,k+1,k}(z), \ldots, p_{kn} A_{knk}(z)),$$

and the matrix $P_k(z)$ with as (i,j) element $p_{ij} A_{ijk}(z)$, we can write (9) as

$$\bar{G}_k(z) = \bar{G}_k(z) P_k(z) + G_{kk}(z) R_k(z), \tag{10}$$

yielding

$$\bar{G}_k(z) = G_{kk}(z) R_k(z) (I - P_k(z))^{-1}. \tag{11}$$

Introducing the column vector
$C_{k-1}^T(z) := (p_{1k} A_{1k,k-1}(z), \ldots, p_{k-2,k} A_{k-2,k,k-1}(z), p_{kk} A_{kk,k-1}(z), \ldots, p_{nk} A_{nk,k-1}(z))^T$,
we can rewrite (8) as

$$G_{kk}(z) = \bar{G}_{k-1}(z) C_{k-1}^T(z) + G_{k-1,k-1}(z) p_{k-1,k} A_{k-1,k,k-1}(z). \tag{12}$$

We have thus expressed $\bar{G}_k(z)$ into $G_{kk}(z)$ via (11), and $G_{kk}(z)$ into $\bar{G}_{k-1}(z)$ and $G_{k-1,k-1}(z)$ via (12). Iterating, defining an empty product to be one and defining $\bar{G}_0(z) C_0^T(z)$ to equal π_1 for notational elegance, we obtain:

$$G_{kk}(z) = \sum_{i=0}^{k-1} \bar{G}_i(z) C_i^T(z) \prod_{j=i+1}^{k-1} p_{j,j+1} A_{j,j+1,j}(z). \tag{13}$$

By carefully studying the structure of the above recursions, and introducing

$$H_i(z) := R_i(z)(I - P_i(z))^{-1} C_i^T(z), \quad i = 1, \dots, n,$$

the following is seen to hold:

$$G_{kk}(z) = \pi_1 \Sigma \prod_{i=1}^{k-1} F_i(z), \quad k = 1, \dots, n, \tag{14}$$

where Σ denotes a sum over the 2^{k-1} terms that arise when each $F_i(z)$, $i = 1, \dots, k-1$, is either $H_i(z)$ or $p_{i,i+1} A_{i,i+1,i}(z)$. For example, for $k = 3$ we get:

$$G_{33}(z) = \pi_1 [H_1(z)H_2(z) + H_1(z)p_{23}A_{232}(z) + p_{12}A_{121}(z)H_2(z) + p_{12}A_{121}(z)p_{23}A_{232}(z).$$

With this explicit Expression (14) for the $G_{kk}(z)$, and Expression (11) for $\bar{G}_k(z)$, we have a recipe to determine all $G_{kj}(z)$ explicitly, for $k, j = 1, \dots, n$.

Example. Let us consider the special case in which $p_{ij} \equiv q_j$, $\forall\ i, j$, and $A_{ijk}(z) =: \hat{A}_{jk}(z)$, $\forall\ i, j, k$. Viz., the Markov renewal process that determines the gate openings and the intervals in between has a simple structure: Each time the next gate opening is of gate j with probability q_j, and the interval length until the next opening also only depends on j. In this case we can obtain a simple expression for $\mathbb{E}[z^{X_{(k)}}] = \sum_{j=1}^{n} G_{kj}(z)$. We have:

$$G_{11}(z) = \pi_1 = q_1, \tag{15}$$

and from (3):

$$G_{1j}(z) = q_j \hat{A}_{j1}(z) \sum_{i=1}^{n} G_{1i}(z), \quad j = 2, \dots, n. \tag{16}$$

Hence

$$\mathbb{E}[z^{X_{(1)}}] = \sum_{j=1}^{n} G_{1j}(z) = q_1 + \sum_{j=2}^{n} q_j \hat{A}_{j1}(z) \mathbb{E}[z^{X_{(1)}}], \tag{17}$$

yielding

$$\mathbb{E}[z^{X_{(1)}}] = \frac{q_1}{1 - \sum_{j=2}^{n} q_j \hat{A}_{j1}(z)}. \tag{18}$$

Furthermore, from (7) and (8),

$$G_{kj}(z) = q_j \hat{A}_{jk}(z) \sum_{i=1}^{n} G_{ki}(z), \tag{19}$$

$$G_{kk}(z) = q_k \hat{A}_{k,k-1}(z) \sum_{i=1}^{n} G_{k-1,i}(z), \tag{20}$$

leading to the following recursive expression of $\mathbb{E}[z^{X_{(k)}}]$ into $\mathbb{E}[z^{X_{(k-1)}}]$:

$$\mathbb{E}[z^{X_{(k)}}] = \frac{q_k \hat{A}_{k,k-1}(z)}{1 - \sum_{j \neq k} q_j \hat{A}_{jk}(z)} \mathbb{E}[z^{X_{(k-1)}}]. \tag{21}$$

Via iteration we obtain:

$$\mathbb{E}[z^{X_{(k)}}] = \prod_{i=1}^{k} \frac{q_i \hat{A}_{i,i-1}(z)}{1 - \sum_{j \neq i} q_j \hat{A}_{jk}(z)}, \tag{22}$$

where $\hat{A}_{10}(z) := 1$.
Notice that (22) represents a decomposition property: The generating function is a product of k terms, all of which are generating functions of random variables, and this implies that $X_{(k)}$ can be represented as the sum of k independent random variables, cf. [6, 7]. In the special case that arrivals only occur at Q_1, and that the generating function of the number of arrivals in all gate intervals is the same, to be denoted by $\hat{A}(z)$, we have:

$$\mathbb{E}[z^{X_{(k)}}] = \hat{A}^{k-1}(z) \prod_{i=1}^{k} \frac{q_i}{1 - \hat{A}(z)(1-q_i)}. \tag{23}$$

When we consider for this case the steady-state number of customers $N_{(n)}$ just *before* a gate opening, we get a slightly more elegant expression. Observing that $\mathbb{E}[x^{N_{(n)}}] = \mathbb{E}[z^{X_{(n)}}]\hat{A}(z)$, we can write:

$$\mathbb{E}[z^{N_{(n)}}] = \prod_{i=1}^{n} \frac{q_i \hat{A}(z)}{1 - \hat{A}(z)(1-q_i)}. \tag{24}$$

This shows that $N_{(n)}$ is distributed like the sum of n independent geometric sums of numbers of arrivals during one gate interval. In particular,

$$\mathbb{E}N_{(n)} = \mathbb{E}A \sum_{i=1}^{n} \frac{1}{q_i}, \tag{25}$$

$$\mathrm{Var}N_{(n)} = (\mathbb{E}A)^2 \sum_{i=1}^{n} (\frac{1}{q_i^2} - \frac{1}{q_i}) + \mathrm{Var}A \sum_{i=1}^{n} \frac{1}{q_i}. \tag{26}$$

The special choice $\hat{A}(z) = z$ (one arrival in each gate interval) yields

$$\mathbb{E}[z^{X_{(k)}}] = z^{k-1} \prod_{i=1}^{k} \frac{q_i}{1-(1-q_i)z}, \tag{27}$$

and hence $X_{(k)} = k - 1 + \sum_{i=1}^{k} B_i$, where $B_i \sim$ geom$(1-q_i)$ for $i = 1, \ldots, k$, and $\mathbb{E}X_{(k)} = k - 1 + \sum_{i=1}^{k} \frac{1-q_i}{q_i} = \sum_{i=1}^{k} \frac{1}{q_i} - 1$.
The special choice $\hat{A}(z) = \frac{\mu}{\mu+\lambda(1-z)}$ (a Poisson distributed number of arrivals in an exp(μ) distributed interval, giving rise to a geometrically distributed number of arrivals in a gate interval) yields

$$\mathbb{E}[z^{X_{(k)}}] = (\frac{\mu}{\mu+\lambda(1-z)})^{k-1} \prod_{i=1}^{k} \frac{q_i(\mu+\lambda(1-z))}{q_i\mu+\lambda(1-z)}, \tag{28}$$

and hence $X_{(k)} = F_{k-1} + \sum_{i=1}^{k} C_i$, where F_{k-1} is negative binomially distributed with parameters $k-1$ and $\frac{\lambda}{\mu+\lambda}$ and where C_i equals zero with probability q_i and is geom$(\frac{q_i\mu}{q_i\mu+\lambda})$ distributed with probability $1-q_i$, $i = 1, \ldots, k$. Hence $\mathbb{E}X_{(k)} = (k-1)\frac{\lambda}{\mu} + \sum_{i=1}^{k}(1-q_i)\frac{\lambda}{q_i\mu}$.
More generally, it follows from (23) that $\mathbb{E}X_{(k)} = [k - 1 + \sum_{i=1}^{k} \frac{1-q_i}{q_i}]\mathbb{E}A$, A denoting the number of arrivals during one gate opening.

3.1 Optimization under constraints

In this subsection we consider three optimization problems, which are very similar to optimization problems studied in [6] for the special case of exponential gate openings. Our goal is to design an efficient ASIP system. We restrict ourselves to the case, leading to (24), in which arrivals only occur in Q_1 while the generating function of the number of arrivals in all gate intervals is the same. For this case we pose the question which choice of $(q_1, \ldots, q_n)$, with $\sum_{i=1}^n q_i = 1$, (i) minimizes the mean number of customers $N_{(n)}$ just before a gate opening, (ii) minimizes the variance of $N_{(n)}$, and (iii) maximizes the probability of zero load (an empty system).

Optimization problem (i): minimization of the mean number of customers
It follows from (25) that the minimization of the mean number of customers amounts to minimizing $\sum_{i=1}^n \frac{1}{q_i}$, sub $\sum_{i=1}^n q_i = 1$. This optimization problem is a special case of the class of resource allocation problems with a separable convex objective function – i.e., the objective function can be separated into n terms, the ith one being a function of q_i only, that is convex in q_i, $i = 1, \ldots, n$. We wish to minimize this separable convex function under a linear constraint. This class of problems is extensively studied in [4]. In particular, if f is convex then

$$f(1/n) = f\left(\frac{1}{n}\sum_{i=1}^n q_i\right) \leq \frac{1}{n}\sum_{i=1}^n f(q_i)$$

and so, $\sum_{i=1}^n f(1/n) \leq \sum_{i=1}^n f(q_i)$ for any $q_i \geq 0$ such that $\sum_{i=1}^n q_i = 1$, thus the optimal solution of our minimization problem is $q_1 = \cdots = q_n = \frac{1}{n}$.

It should be noted that the mean number of customers just before a gate opening is readily expressed in the steady-state mean number of customers at an arbitrary epoch; just subtract the mean number of arrivals in a residual gate opening interval. The latter is linearly related to the mean time in system via Little's law. Hence the above optimization problem also sheds light on the minimization of time in system.

Optimization problem (ii): minimization of the variance of the number of customers
It follows from (26) that the minimization of the variance of the number of customers amounts to minimizing $\sum_{i=1}^n [(\frac{1}{q_i^2} - \frac{1}{q_i})(\mathbb{E}A)^2 + \frac{1}{q_i}\mathrm{Var}A]$. The same reasoning as for (i) applies; we again are faced with a separable convex objective function, and again the optimal solution is $q_1 = \cdots = q_n = \frac{1}{n}$.

Optimization problem (iii): maximization of the probability of an empty system
It follows from (24) that the maximization of the probability of an empty system amounts to maximizing $\prod_{i=1}^n \frac{q_i \hat{A}(0)}{1-(1-q_i)\hat{A}(0)}$, and hence to minimizing $\sum_{i=1}^n \ln[\frac{1-\hat{A}(0)}{q_i} + \hat{A}(0)]$. The same reasoning as for (i) and (ii) applies once again; we have a separable convex objective function, and again the optimal solution is $q_1 = \cdots = q_n = \frac{1}{n}$.

4 Some two-queue results

In this section we study the two-queue case in some more detail. In that case one can sometimes determine the *joint* queue length distribution at gate opening intervals. In Subsection 4.1 we determine the joint queue length distribution at gate openings for a specific choice of the p_{ij} and the same arrival distributions for gate 1 intervals and gate 2 intervals. In Subsection 4.2 we determine the joint queue length distribution for the case in which, when the gate of Q_1 opens, only a *binomially distributed* number of the customers in Q_1 moves to Q_2. These two-queue studies not only lead to more detailed results; they also sometimes give an indication of the limitations of our approach. For example, if one would not only at Q_1, but also at Q_2, allow a binomially distributed number of customers to leave when its gate opens, then a functional equation in the two-dimensional queue length probability generating function results, which seems very difficult to analyze exactly.

4.1 Joint queue length distribution

Let us consider the problem of determining the generating function of the steady-state joint queue length distribution right after gate openings, $G(z_1, z_2) = \mathbb{E}[z_1^{X_1} z_2^{X_2}]$. Take $n = 2$; take only arrivals at Q_1, with generating function $A(z)$ of the number of arrivals per gate opening, regardless whether it is an opening of gate 1 or of gate 2; and take fixed gate opening probabilities $p_{ij} \equiv q_j$. Realizing that, with $X_i^{(r)}$ the number of customers in Q_i right after the rth gate opening, and with A_{r+1} the number of arrivals in the interval between the rth and $(r+1)$st gate openings,

$$X_1^{(r+1)} = 0, \quad X_2^{(r+1)} = X_1^{(r)} + A_{r+1} + X_2^{(r)},$$

if the $(r+1)$st gate opening is of gate 1, and

$$X_1^{(r+1)} = X_1^{(r)} + A_{r+1}, \quad X_2^{(r+1)} = 0,$$

if the $(r+1)$st gate opening is of gate 2, we obtain in steady state:

$$G(z_1, z_2) = q_1 A(z_2) G(z_2, z_2) + q_2 A(z_1) G(z_1, 1). \tag{29}$$

Actually we already know $G(z_1, 1)$, which equals $\mathbb{E}[z_1^{X(1)}]$; but it also follows from (29) by putting $z_2 = 1$. We also already know $G(z_1, z_1)$, which equals $\mathbb{E}[z_1^{X(2)}]$; but it also follows from (29) by putting $z_2 = z_1$. We find:

$$G(z_1, 1) = \frac{q_1}{1 - q_2 A(z_1)}, \tag{30}$$

$$G(z_1, z_1) = \frac{q_1 q_2 A(z_1)}{1 - q_2 A(z_1)} \frac{1}{1 - q_1 A(z_1)}, \tag{31}$$

$$G(z_1, z_2) = \frac{q_1 q_2 A(z_2)}{1 - q_2 A(z_2)} \frac{q_1 A(z_2)}{1 - q_1 A(z_2)} + \frac{q_1 q_2 A(z_1)}{1 - q_2 A(z_1)}. \tag{32}$$

Remark. One could extend the above analysis to the case of arrivals at both queues, and different PGFs for different gate openings. However, this comes at the expense of messier expressions, and we have decided not to include this case in the paper.

One could also in principle analyse the steady-state queue length PGF at an arbitrary epoch. One then would have to average over different gate opening intervals. However, the arrival process must then first be specified in more detail; do arrivals all take place at the beginning of a gate opening interval, or at the end, or maybe according to a Poisson process?

4.2 Binomial movements

Consider the case of $n = 2$ queues in series, with the special feature that, when the gate of Q_1 opens, each customer present in Q_1 (independently from the other customers) moves with probability $a_1 > 0$ to Q_2, and stays with probability $1 - a_1$ in Q_1. We restrict ourselves to the case of a Poisson arrival process, with rate λ, at Q_1, and no external arrivals at Q_2; moreover, we assume that gate openings at Q_i occur after i.i.d., exponentially, distributed intervals with mean $1/\mu_i$, $i = 1, 2$. Denoting by $X_i(t)$ the number of customers in Q_i at time t, $i = 1, 2$, and by $X_1^{bin}(t)$ the number of customers who do move from Q_1 to Q_2 at a gate opening of Q_1 that takes place at time t, we can write (suppressing initial conditions; we shall anyway soon turn to the steady-state situation):

$$\begin{aligned}\mathbb{E}[z_1^{X_1(t+h)} z_2^{X_2(t+h)}] &= (1 - (\lambda + \mu_1 + \mu_2)h)\mathbb{E}[z_1^{X_1(t)} z_2^{X_2(t)}] \\ &+ \lambda h z_1 \mathbb{E}[z_1^{X_1(t)} z_2^{X_2(t)}] \\ &+ \mu_1 h \mathbb{E}[z_1^{X_1(t) - X_1^{bin}(t)} z_2^{X_2(t) + X_1^{bin}(t)}] \\ &+ \mu_2 h \mathbb{E}[z_1^{X_1(t)}] + o(h), \quad h \downarrow 0,\end{aligned} \tag{33}$$

leading to

$$\begin{aligned}\frac{d}{dt}\mathbb{E}[z_1^{X_1(t)} z_2^{X_2(t)}] &= -(\lambda + \mu_1 + \mu_2)\mathbb{E}[z_1^{X_1(t)} z_2^{X_2(t)}] \\ &+ \lambda z_1 \mathbb{E}[z_1^{X_1(t)} z_2^{X_2(t)}] \\ &+ \mu_1 \mathbb{E}[((1 - a_1)z_1 + a_1 z_2)^{X_1(t)} z_2^{X_2(t)}] \\ &+ \mu_2 \mathbb{E}[z_1^{X_1(t)}].\end{aligned} \tag{34}$$

Denoting the probability generating function of the joint distribution of the steady-state queue length vector (X_1, X_2) by $H(z_1, z_2)$, we have

$$[\mu_1 + \mu_2 + \lambda(1 - z_1)]H(z_1, z_2) = \mu_1 H((1 - a_1)z_1 + a_1 z_2, z_2) + \mu_2 H(z_1, 1). \tag{35}$$

We shall first obtain $H(z_1, 1)$. Substituting $z_2 = 1$ into (35) yields:

$$[\mu_1 + \mu_2 + \lambda(1 - z_1)]H(z_1, 1) = \mu_1 H((1 - a_1)z_1 + a_1, 1) + \mu_2 H(z_1, 1), \tag{36}$$

and hence

$$H(z_1, 1) = \frac{\mu_1}{\mu_1 + \lambda(1 - z_1)} H((1 - a_1)z_1 + a_1, 1). \tag{37}$$

Iteration of this relation gives

$$H(z_1, 1) = \prod_{j=0}^{\infty} \frac{\mu_1}{\mu_1 + \lambda(1 - z_1)(1 - a_1)^j}. \tag{38}$$

This infinite product is said to converge iff $\sum_{j=0}^{\infty}(1 - \frac{\mu_1}{\mu_1 + \lambda(1-z_1)(1-a_1)^j}) < \infty$, and hence the infinite product indeed converges if $0 < a_1 < 1$. If $a_1 = 1$ one obtains $H(z_1, 1) = \frac{\mu_1}{\mu_1 + \lambda(1-z_1)}$. This is not a surprising result; it is the generating function of the number of Poisson(λ) arrivals during an exp(μ) interval. According to PASTA it also equals the generating function of the steady-state queue length distribution of Q_1. Observing that $\frac{\mu_1}{\mu_1 + \lambda(1-z_1)(1-a_1)^j}$ is the probability generating function of a geometrically distributed random variable with success parameter $\frac{\lambda(1-a_1)^j}{\mu_1+\lambda(1-a_1)^j}$, one can write

$$X_1 \stackrel{d}{=} \sum_{j=0}^{\infty} H_j, \tag{39}$$

where all H_j are independent, H_j being geometrically distributed with success parameter $\frac{\lambda(1-a_1)^j}{\mu_1+\lambda(1-a_1)^j}$.

Having determined $H(z_1, 1)$, we now turn to the determination of $H(z_1, z_2)$. It follows from (35) that

$$H(z_1, z_2) = Y_1(z_1)H((1-a_1)z_1 + a_1 z_2, z_2) + Y_0(z_1), \tag{40}$$

where

$$Y_1(z_1) := \frac{\mu_1}{\mu_1 + \mu_2 + \lambda(1-z_1)}, \quad Y_0(z_1) := \frac{\mu_2}{\mu_1 + \mu_2 + \lambda(1-z_1)} H(z_1, 1).$$

Iteration of (40) gives:

$$H(z_1, z_2) = \sum_{j=0}^{\infty} Y_0(f_j(z_1, z_2)) \prod_{i=0}^{j-1} Y_1(f_i(z_1, z_2)), \tag{41}$$

an empty product being equal to one and $f_i(z_1, z_2) := (1-a_1)^i z_1 + [1-(1-a_1)^i] z_2$, $i = 0, 1, \ldots$. Using d'Alembert's ratio test one can show that this infinite sum converges. In fact, the sum converges geometrically fast. Indeed, since $a_1 > 0$, one has $f_j(z_1, z_2) \to z_2$, and the ratio of two successive terms in the sum $H(z_1, z_2)$, which is given by $\frac{Y_0(f_{j+1}(z_1,z_2))}{Y_0(f_j(z_1,z_2))} Y_1(f_j(z_1, z_2))$, is for large j bounded by $\mu_1/(\mu_1 + \mu_2)$.

Above we have restricted ourselves to the case of a Poisson arrival process, with rate λ, at Q_1, and no external arrivals at Q_2; moreover, we assumed that gate openings at Q_i occur after i.i.d. exponentially distributed intervals with mean $1/\mu_i$, $i = 1, 2$. Let us now turn to the more general case of Section 2, in which gate openings are determined by a Markov renewal process, and where a gate opening of Q_i is with probability p_{ij} followed by a gate opening of Q_j, while $A_{ij}(z_1, z_2)$ is the generating function of the numbers of arrivals in Q_1 and Q_2 during the period in between those two successive gate openings. Considering the steady-state joint distribution of numbers of customers (X_1, X_2) immediately after gate openings, and letting (cf. (2))

$$G_i(z_1, z_2) := \mathbb{E}[z_1^{X_1} z_2^{X_2} I(M = i)], \quad i = 1, 2, \tag{42}$$

it is easily seen by observing the system at two successive gate openings that

$$\begin{aligned} G_1(z_1, z_2) &= p_{11} A_{11}((1-a_1)z_1 + a_1 z_2, z_2) G_1((1-a_1)z_1 + a_1 z_2, z_2) \\ &+ p_{21} A_{21}((1-a_1)z_1 + a_1 z_2, z_2) G_2((1-a_1)z_1 + a_1 z_2, z_2), \end{aligned} \tag{43}$$

$$G_2(z_1, z_2) = p_{12} A_{12}(z_1, 1) G_1(z_1, 1) + p_{22} A_{22}(z_1, 1) G_2(z_1, 1). \tag{44}$$

It is immediately obvious from (44) that $G_2(z_1, z_2)$ does not depend on z_2, as we could have expected because Q_2 becomes empty after a gate opening at Q_2. Hence it follows from (44) that

$$G_2(z_1, z_2) = G_2(z_1, 1) = \frac{p_{12} A_{12}(z_1, 1)}{1 - p_{22} A_{22}(z_1, 1)} G_1(z_1, 1). \tag{45}$$

Plugging $z_2 = 1$ in (43) and using (45) gives:

$$
\begin{aligned}
G_1(z_1,1) &= p_{11}A_{11}((1-a_1)z_1+a_1,1)G_1((1-a_1)z_1+a_1,1) \\
&+ p_{21}A_{21}((1-a_1)z_1+a_1,1)\frac{p_{12}A_{12}((1-a_1)z_1+a_1,1)}{1-p_{22}A_{22}((1-a_1)z_1+a_1,1)}G_1((1-a_1)z_1+a_1,1), \quad (46)
\end{aligned}
$$

which can be written as

$$G_1(z_1,1) = L(z_1)G_1((1-a_1)z_1+a_1,1), \tag{47}$$

with an obvious choice of the function $L(\cdot)$. Iteration readily yields that

$$G_1(z_1,1) = \prod_{j=0}^{\infty} L(d^{(j)}(z_1)), \tag{48}$$

where

$$d^{(j)}(z_1) := (1-a_1)^j z_1 + 1 - (1-a_1)^j, \quad j = 0,1,\dots\ . \tag{49}$$

The infinite product converges iff the corresponding infinite sum $\sum_{j=0}^{\infty}[1-L(d^{(j)}(z_1))]$ converges. The latter sum convergences geometrically fast. This can be seen by making the following two observations. Observation (i): $L(z_1)$ has the meaning of a probability generating function. Indeed, distinguish between the possibility that a gate opening of Q_1 is followed by another gate opening of Q_1 (probability p_{11}) and the possibility that it is followed by a gate opening of Q_2, followed by a geometric(p_{22}) number of gate openings of Q_2, and finally again a gate opening of Q_1. Observation (ii): $1 - d^{(j)}(z_1) = (1-a_1)^j(1-z_1)$ converges geometrically fast to 0.

Having determined $G_1(z_1,1)$ and hence, using (45), $G_2(z_1,z_2) = G_2(z_1,1)$, we substitute the result in (43), obtaining:

$$G_1(z_1,z_2) = K_1(z_1,z_2)G_1((1-a_1)z_1+a_1z_2,z_2) + K_0(z_1,z_2), \tag{50}$$

where

$$K_1(z_1,z_2) := p_{11}A_{11}((1-a_1)z_1+a_1z_2,z_2), \tag{51}$$

$$K_0(z_1,z_2) := p_{21}A_{21}((1-a_1)z_1+a_1z_2,z_2)\frac{p_{12}A_{12}((1-a_1)z_1+a_1z_2,1)}{1-p_{22}A_{22}((1-a_1)z_1+a_1z_2,1)}G_1((1-a_1)z_1+a_1z_2,1). \tag{52}$$

Iteration of (50) gives:

$$G_1(z_1,z_2) = \sum_{j=0}^{\infty} K_0(f_j(z_1,z_2),z_2)\prod_{i=0}^{j-1} K_1(f_i(z_1,z_2),z_2). \tag{53}$$

Again dÁlembert's ratio test readily shows the convergence of the infinite sum, by using that $|K_1(z_1,z_2)| < p_{11}$. Finally notice that $G_1(z_2,z_2)$, which is the generating function of the total number of customers $X_{(2)} = X_1 + X_2$ in the two queues just after gate openings of Q_1, follows by substituting $z_1 = z_2$ in (53). Since $f_j(z_2,z_2) \equiv z_2$, that formula degenerates into

$$G_1(z_2,z_2) = \frac{K_0(z_2,z_2)}{1-K_1(z_2,z_2)}. \tag{54}$$

After some calculations, this expression is seen to agree with the expression for $G_{21}(z_2)$ that can be derived from (7). This agreement may at first sight seem strange, as we have binomial movements in the present subsection. However, notice that we compare $G_1(z_2,z_2)$ and $G_{21}(z_2)$, both giving the *total* number of customers in both queues. It then does not matter whether some of them are still in Q_1 after a gate opening of Q_1.

4.3 An ASIP model with a renewal arrival process at Q_1

In this subsection we consider the case in which arrivals only take place at Q_1, and follow a renewal process: successive interarrival times are i.i.d., with distribution $A(\cdot)$ and Laplace-Stieltjes transform $\alpha(\cdot)$. We restrict ourselves to $n = 2$ queues. We furthermore restrict ourselves to the case in which openings of the gate of Q_i occur at i.i.d. $\exp(\mu_i)$ distributed intervals, independent of each other and independent of the arrival intervals.

Let $(Y_{n,1}, Y_{n,2})$ denote the vector of numbers of customers in (Q_1, Q_2) just before the nth arrival at Q_1, $n = 1, 2, \dots$. Let A_n denote the arrival interval between customers $n-1$ and n. We need to distinguish between the following five cases.
(i) No gate opening in A_n. This event has probability $\alpha(\mu_1+\mu_2)$; and $(Y_{n,1}, Y_{n,2}) = (Y_{n-1,1} + 1, Y_{n-1,2})$.
(ii) No openings of gate 1 and at least one opening of gate 2 in A_n. This event has probability $\alpha(\mu_1) - \alpha(\mu_1 + \mu_2)$; and $(Y_{n,1}, Y_{n,2}) = (Y_{n-1,1} + 1, 0)$.
(iii) No openings of gate 2 and at least one opening of gate 1 in A_n. This event has probability $\alpha(\mu_2) - \alpha(\mu_1 + \mu_2)$; and $(Y_{n,1}, Y_{n,2}) = (0, Y_{n-1,1} + 1 + Y_{n-1,2})$.
(iv) Both gates open at least once in A_n; the first opening of gate 1 occurs after the last opening of gate 2. This event has probability $\alpha(\mu_1 + \mu_2) - \frac{\mu_2}{\mu_2-\mu_1}\alpha(\mu_2) - \frac{\mu_1}{\mu_1-\mu_2}\alpha(\mu_1)$; and $(Y_{n,1}, Y_{n,2}) = (0, Y_{n-1,1} + 1)$.
(v) Both gates open at least once in A_n; but the first opening of gate 1 occurs before the last opening of gate 2. This event has probability $1 - \frac{\mu_2}{\mu_2-\mu_1}\alpha(\mu_1) - \frac{\mu_1}{\mu_1-\mu_2}\alpha(\mu_2)$, as can, e.g., be seen by writing the probability of this event as the probability that the sum of an $\exp(\mu_1)$ plus an $\exp(\mu_2)$ random variable is less than A_n. We now have $(Y_{n,1}, Y_{n,2}) = (0, 0)$; notice that this is the only way to get into the state $(0, 0)$.
Restricting ourselves to steady-state queue lengths just before arrivals, to be denoted by (Y_1, Y_2), and introducing their generating function $L(z_1, z_2) := \mathbb{E}[z_1^{Y_1} z_2^{Y_2}]$, we obtain:

$$
\begin{aligned}
L(z_1, z_2) &= \alpha(\mu_1 + \mu_2)z_1 L(z_1, z_2) \\
&+ (\alpha(\mu_1) - \alpha(\mu_1 + \mu_2))z_1 L(z_1, 1) \\
&+ (\alpha(\mu_2) - \alpha(\mu_1 + \mu_2))z_2 L(z_2, z_2) \\
&+ [\alpha(\mu_1 + \mu_2) - \frac{\mu_2}{\mu_2 - \mu_1}\alpha(\mu_2) - \frac{\mu_1}{\mu_1 - \mu_2}\alpha(\mu_1)]z_2 L(z_2, 1) \\
&+ 1 - \frac{\mu_2}{\mu_2 - \mu_1}\alpha(\mu_1) - \frac{\mu_1}{\mu_1 - \mu_2}\alpha(\mu_2). \qquad (55)
\end{aligned}
$$

Taking all $L(z_1, z_2)$ terms together, and introducing

$$\zeta := 1 - \frac{\mu_2}{\mu_2 - \mu_1}\alpha(\mu_1) - \frac{\mu_1}{\mu_1 - \mu_2}\alpha(\mu_2),$$

(which actually is $L(0, 0) = P(Y_1 = 0, Y_2 = 0)$, see above) and

$$\omega := \alpha(\mu_1 + \mu_2) - \frac{\mu_2}{\mu_2 - \mu_1}\alpha(\mu_2) - \frac{\mu_1}{\mu_1 - \mu_2}\alpha(\mu_1) = \alpha(\mu_1 + \mu_2) - \alpha(\mu_1) - \alpha(\mu_2) + 1 - \zeta,$$

we obtain

$$
\begin{aligned}
(1 - \alpha(\mu_1 + \mu_2)z_1)L(z_1, z_2) &= (\alpha(\mu_1) - \alpha(\mu_1 + \mu_2))z_1 L(z_1, 1) \\
&+ (\alpha(\mu_2) - \alpha(\mu_1 + \mu_2))z_2 L(z_2, z_2) \\
&+ \omega z_2 L(z_2, 1) + \zeta. \qquad (56)
\end{aligned}
$$

Substitution of $z_2 = 1$ in (56), and using the fact that $\alpha(\mu_2) - \alpha(\mu_1+\mu_2) + \zeta + \omega = 1 - \alpha(\mu_1)$ yields

$$(1-\alpha(\mu_1+\mu_2)z_1)L(z_1,1) = (\alpha(\mu_1)-\alpha(\mu_1+\mu_2))z_1L(z_1,1) + 1 - \alpha(\mu_1), \tag{57}$$

and hence

$$L(z_1,1) = E[z_1^{Y_1}] = \frac{1-\alpha(\mu_1)}{1-\alpha(\mu_1)z_1}. \tag{58}$$

The marginal distribution of Y_1 hence is geometric. The explanation is that Y_1 increases by 1 for a geometrically distributed number of arrival intervals (with parameter $\alpha(\mu_1)$, which is the probability that gate 1 does not close during an arrival interval), and then falls back to zero.

Substituting $z_1 = z_2$ in (56) allows us to express $L(z_2, z_2)$ into $L(z_2, 1)$:

$$(1-\alpha(\mu_2)z_2)L(z_2,z_2) = (\alpha(\mu_1)-\alpha(\mu_1+\mu_2))z_2L(z_2,1) + \zeta + \omega z_2 L(z_2,1), \tag{59}$$

yielding the following expression for the generating function of the total number of customers in the system just before an arrival at Q_1:

$$L(z_2,z_2) = \frac{(\alpha(\mu_1)-\alpha(\mu_1+\mu_2)+\omega)\frac{(1-\alpha(\mu_1))z_2}{1-\alpha(\mu_1)z_2} + \zeta}{1-\alpha(\mu_2)z_2}. \tag{60}$$

Finally, Equations (56), (58) and (60) give the generating function of (Y_1, Y_2):

$$\begin{aligned} L(z_1,z_2) &= \frac{1}{1-\alpha(\mu_1+\mu_2)z_1} \\ &\times [(\alpha(\mu_1)-\alpha(\mu_1+\mu_2))\frac{(1-\alpha(\mu_1))z_1}{1-\alpha(\mu_1)z_1} \\ &+ (\alpha(\mu_2)-\alpha(\mu_1+\mu_2))\frac{z_2}{1-\alpha(\mu_2)z_2}(\zeta + \frac{(1-\alpha(\mu_1))z_2}{1-\alpha(\mu_1)z_2}(\alpha(\mu_1)-\alpha(\mu_1+\mu_2)+\omega)) \\ &+ \omega\frac{(1-\alpha(\mu_1))z_2}{1-\alpha(\mu_1)z_2} + \zeta]. \end{aligned} \tag{61}$$

Substituting $z_2 = 0$ in (61) gives

$$L(z_1,0) = \frac{(\alpha(\mu_1)-\alpha(\mu_1+\mu_2))\frac{(1-\alpha(\mu_1))z_1}{1-\alpha(\mu_1)z_1} + \zeta}{1-\alpha(\mu_1+\mu_2)z_1}. \tag{62}$$

In a similar way we get $L(0, z_2)$ and $L(1, z_2) = E[z_2^{Y_2}]$. In particular,

$$\begin{aligned} L(1,z_2) &= \frac{1}{1-\alpha(\mu_1+\mu_2)} \\ &\times [(\alpha(\mu_1)-\alpha(\mu_1+\mu_2)) \\ &+ (\alpha(\mu_2)-\alpha(\mu_1+\mu_2))\frac{z_2}{1-\alpha(\mu_2)z_2}(\zeta + \frac{(1-\alpha(\mu_1))z_2}{1-\alpha(\mu_1)z_2}(\alpha(\mu_1)-\alpha(\mu_1+\mu_2)+\omega)) \\ &+ \omega\frac{(1-\alpha(\mu_1))z_2}{1-\alpha(\mu_1)z_2} + \zeta]. \end{aligned} \tag{63}$$

Substituting $z_2 = 0$ in (61) gives

$$L(z_1, 0) = \frac{(\alpha(\mu_1) - \alpha(\mu_1 + \mu_2))\frac{(1-\alpha(\mu_1))z_1}{1-\alpha(\mu_1)z_1} + \zeta}{1 - \alpha(\mu_1 + \mu_2)z_1}. \tag{64}$$

It is seen that the marginal distribution of Y_2 has an atom in 0 and furthermore is a weighted sum of (i) a geometric($\alpha(\mu_1)$) distribution, (ii) a geometric($\alpha(\mu_2)$) distribution, and (iii) a convolution of two such geometric distributions.

Finally we determine the generating function of the joint distribution of the steady-state numbers of customers (S_1, S_2) in Q_1 and Q_2 *at an arbitrary epoch*. It is easily seen that this distribution is obtained by considering the queue lengths at a time A^r after the last customer arrival, where this forward recurrence interarrival time or residual interarrival time has LST $\alpha^r(s) = \frac{1-\alpha(s)}{sE[A]}$. We can follow the reasoning leading to (55), simply replacing each $\alpha(\cdot)$ term by $\alpha^r(\cdot)$. Hence

$$\begin{aligned} E[z_1^{S_1} z_2^{S_2}] &= \alpha^r(\mu_1 + \mu_2)z_1 L(z_1, z_2) \\ &+ (\alpha^r(\mu_1) - \alpha^r(\mu_1 + \mu_2))z_1 L(z_1, 1) \\ &+ (\alpha^r(\mu_2) - \alpha^r(\mu_1 + \mu_2))z_2 L(z_2, z_2) \\ &+ \tilde{\omega} z_2 L(z_2, 1) + \tilde{\zeta}, \end{aligned} \tag{65}$$

where $\tilde{\omega}$ and $\tilde{\zeta}$ are obtained from ω and ζ by replacing $\alpha(\cdot)$ by $\alpha^r(\cdot)$ everywhere.

5 Suggestions for further research

The following extensions might be of interest:

1. A batch can move one *or two* queues ahead at a gate opening. The approach taken in Section 3 to obtain expressions for the $G_{ki}(z)$ (cf. (1)) breaks down when batches could move more than one queue ahead after a gate opening. For example, if a batch from Q_{k-1} could move ahead to Q_{k+1}, then the expression for $G_{k,k-1}(z)$ in (7) becomes much more complicated. $G_{k,k-1}(z)$ is the generating function of the number of customers $X_{(k)}$ in the first k queues, right after gate $k-1$ has opened (and it contains the indicator function of the latter event). Q_{k-1} has become empty after that gate opening, *and* its customers no longer are part of the first k queues. So in the righthand side of (7), we need a term that contains the generating function of the number in the first k queues right after the previous gate opening, but *without* the customers who after that previous gate opening still resided in Q_{k-1}. Since our approach is based on considering the total number of customers in the first k queues, it becomes cumbersome to consider the sum $X_1 + \cdots + X_{k-2} + X_k$.

2. At each gate opening, *multiple* gates can open. If, with probability r_i, gates i *and* $i+1$ open, $i = 1, 2, \ldots$, then this amounts to a batch moving *two* queues ahead. So this variant is related to the previous one.

3. Non-tandem configurations. E.g., there are three queues, Q_1 feeding into Q_2 and Q_3 – with fixed probabilities, or via a fixed alternating pattern.

Finally, there are various asymptotic questions of interest. For example, one could let $n \to \infty$, and study, e.g., the fraction of empty stations. We refer to Chapter 6 of [5] and to [7, 8, 9] for an interesting collection of limit laws for three limiting regimes (for the case of only arrivals at Q_1, and exponential gate openings): (i) The heavy-traffic regime, in which the arrival rate at Q_1 goes to infinity; (ii) the large-system regime in which $n \to \infty$; (iii) the balanced-system regime, in which $n \to \infty$, the gate opening intervals tend to zero, and the product of n and the mean gate opening interval tends to a positive limit.

Acknowledgment. Onno Boxma gratefully acknowledges interesting discussions with Corné Suijkerbuijk.

References

[1] R. A. Blythe and M. R. Evans (2007). Nonequilibrium steady states of matrix product form: A solver's guide, J. Phys. A: Math. Theor. 40, R333-R441.

[2] B. Derrida (1998). An exactly soluble non-equilibrium system: The asymmetric simple exclusion process, Physics Reports 301, 65-83.

[3] O. Golinelli and K. Mallick (2006). The asymmetric simple exclusion process: an integrable model for non-equilibrium statistical mechanics, J. Phys. A: Math. Gen. 39, 12679-12705.

[4] T.I. Ibaraki and N. Katoh (1988). *Resource Allocation Problems.* MIT Press, Cambridge.

[5] S. Reuveni (2014). *Tandem Stochastic Systems: The Asymmetric Simple Inclusion Process.* PhD Thesis, Tel-Aviv University.

[6] S. Reuveni, I. Eliazar and U. Yechiali (2011). Asymmetric inclusion process, Physical Review E 84, 041101, 1-16.

[7] S. Reuveni, I. Eliazar and U. Yechiali (2012). Limit laws for the asymmetric inclusion process, Physical Review E 86, 061133, 1-17.

[8] S. Reuveni, I. Eliazar and U. Yechiali (2012). Asymmetric inclusion process as a showcase of complexity, Physical Review Letters 109, 020603, 1-4.

[9] S. Reuveni, O. Hirschberg, I. Eliazar and U. Yechiali (2014). Occupation probabilities and fluctuations in the asymmetric inclusion process, Physical Review E 89, 042109, 1-23.