Analysis of a 2-state discrete-time queue with stochastic state-period lengths and state-dependent server availability and arrivals

Freek Verdonck[a,*], Herwig Bruneel[a], Sabine Wittevrongel[a]

[a] *Ghent University (UGent), Department of Telecommunications and Information Processing (TELIN), Sint-Pietersnieuwstraat 41, 9000 Gent, Belgium*

Abstract

In this work we study a discrete-time multiserver queueing system with an infinite storage capacity and deterministic service times equal to 1 slot. Specific to the model under study is that the system is assumed to be in one of two different states (state-1 or state-2) and that both the distribution of the number of available servers and the arrival process depend on the system state. State changes can only occur at slot boundaries and mark the beginning and end of state-1-periods and state-2-periods. The lengths of these state-1-periods and state-2-periods, expressed as a number of slots, are assumed to be two independent sets of independent and identically distributed stochastic variables. The number of available servers during a slot is a stochastic variable with a distribution that is completely determined by the system state during that slot. Likewise, the distribution of the number of arrivals during a slot only depends on the system state during that slot. The only restrictions we put on the distributions of the state-1-periods, state-2-periods and number of available servers is that they have rational probability generating functions (pgfs), and that during each slot at least one server is available. For the considered queueing system we present a method to determine the pgf of the steady-state system content at various observation instants. Several numerical examples demonstrate the possibilities of this model.

Keywords: Queueing theory, discrete-time, multiserver, correlation

1. Introduction

Queueing theory has been the subject of mathematical analysis for a number of decades. Its applications are wide and vary from production lines over service industries to the Internet. The research domain of queueing systems is subdivided based on a large number of criteria, some of which are the choice of time line (continuous or discrete), the used solution technique (simulation, analytical or semi-analytical solution techniques, numerical procedures, etc.) and the presence of specific restrictions or assumptions in the model (with respect to the number of servers, the arrival process, the distribution of service times, etc.)

In many applications of queueing theory the number of available servers is not constant over time. This can be due to *internal* reasons (e.g. a server that takes a vacation when it finds its queue idle) or due to *external* reasons (e.g. a machine that breaks down in a production facility). In this paper, we focus on discrete-time multiserver queueing systems with a server availability that varies independently of the system content, so due to external reasons. A way to model such varying server availability is the introduction of server interruptions. The analysis of discrete-time queueing systems with various types of random server interruptions has received considerable attention so far; we refer to [1] for a recent survey. Most of the

*Corresponding author.

Email addresses: `freek.verdonck@ugent.be` (Freek Verdonck), `herwig.bruneel@ugent.be` (Herwig Bruneel), `sabine.wittevrongel@ugent.be` (Sabine Wittevrongel)

https://doi.org/10.1016/j.peva.2019.102026

available research with respect to server interruptions, however, concerns the single-server case, see e.g. [2, 3, 4, 5, 6] for some related results.

For the multiserver case (with $m > 1$ servers), on the contrary, only a limited number of studies on queues with server interruptions are available in the literature. In the simplest studies [7, 8, 9] the server availability is assumed to be independent from slot to slot. Specifically, [7] considers the case where interruptions of the m servers occur simultaneously, i.e., either no servers or all m servers are available during a slot, and independently from slot to slot, where a single parameter σ indicates the probability that the servers are available. In [8, 9], the number of available servers during a slot can take all values from 0 to m, and is independent and identically distributed (i.i.d.) from slot to slot. Some extensions to a time-correlated server-interruption process are reported in [10, 11, 12]. In particular, the number of available servers is assumed to be a first-order Markov chain in [10]. In the model of [11], the time horizon is divided into on-periods (of geometrically distributed lengths) and off-periods (of arbitrarily distributed lengths). Here, during off-periods no servers are available, while the number of available servers during a slot of an on-period takes a random value and changes independently from slot to slot. In the recent analysis [12], a two-server queueing system is considered, where one server is permanently available and the other server is only intermittently available according to a correlated interruption process with alternating geometric up-periods and arbitrarily distributed down-periods.

All the previous studies [7, 8, 9, 10, 11, 12] on multiserver queues with varying server availability have in common that the numbers of arrivals are considered to be i.i.d. from slot to slot. The simultaneous presence of time-correlation in both the arrival process and the server availability has to the best of our knowledge only been considered in the single-server case, see e.g. [4, 5, 6]. Even for discrete-time queues without server interruptions, the effect of time-correlated arrivals has mainly been studied for single-server queues, see e.g. [13, 14, 15, 16, 17].

The main aim of the current paper exactly is to analyze a discrete-time multiserver queueing model that includes a more general time-correlated description of both the server-availability process and the arrival process. In our paper, we will assume the queueing system to alternate between two different system states, each with arbitrarily distributed sojourn times. In addition, each state is characterized by its own arbitrary distribution for the number of available servers during a slot. Moreover, in the considered model, the arrival process to the queueing system may also depend on the system state. The analysis in this paper is an extension of previous work in the sense that we allow both states to have a stochastic number of servers available, we do not limit neither of the state periods to have geometrically distributed lengths and we introduce correlation in the arrival process by allowing the number of arrivals during a slot to depend on the system state. The objective of our analysis is to find the distribution of the system content at several observation points.

The motivation for this model are the many applications with both (correlated) variation in the number of available servers and correlated (e.g. bursty) arrivals. Examples are models of the airport check-in process [18] and models of production inventory systems [19]. Another practical case is the increasing use of peer-to-peer logistics in cities, e.g. for food deliveries [20, 21]. Courier services use smartphone applications to assign self-employed drivers to pick up and deliver food orders. Here both the number of available servers (bike couriers) and the number of customers can be stochastically dependent on the time of the day, weather situations, etc.

The outline of the paper is as follows. In the next section we describe the queueing model under study in detail and in Section 3 we perform the mathematical queueing analysis. Specifically, in Section 3, we develop a method to determine the pgf of the steady-state system content at the beginning of an arbitrary state-1-period, state-2-period, state-1-slot and state-2-slot and at the beginning of an arbitrary slot. The results are obtained after finding a finite number of roots within the complex unit disk of a non-polynomial function and solving a set of linear equations. In Section 4 we give an expression for some important performance measures of the system. We discuss some numerical examples in Section 5, before concluding the paper in Section 6.

2. Queueing model under study

In discrete-time queueing models the time axis is divided into time slots of equal length. For the model considered in this paper we take the slot length equal to the deterministic service time of customers. Let us have a closer look at the events that can occur around slot k. Suppose that at the beginning of slot k, g_{k-1} customers are present in the queueing system and s_k servers are available. The minimum of g_{k-1} and s_k customers enter service at the beginning of slot k. During the slot, a total of c_k new customers arrive; these cannot enter service before the beginning of the next slot, even if there are servers idle at the moment of arrival. We assume infinite storage capacity, so an arriving customer will always join the system. At the end of the slot, the customers that were taken into service leave the system. Linking the system content from slot to slot in this case leads to the following equation:

$$g_k = (g_{k-1} - s_k)^+ + c_k \,, \tag{1}$$

with the operator $(...)^+$ defined as $\max(..., 0)$. Here the s_k and c_k are given, or at least their distributions are given, while we aim to solve the equation for the distribution of g_k in steady state. When the series $\{s_k\}$ and $\{c_k\}$ are both sets of i.i.d. random variables, the analysis is rather straightforward, but then the model does not allow any correlation in the arrivals or in the server availability. In this paper we introduce a way to allow more correlation while keeping the corresponding queueing analysis mathematically tractable.

For this purpose, we assume a queueing system with two states; we refer to the states as state-1 and state-2. Each state has its own distribution for the number of servers available and the number of arrivals per slot. Within a given state, these variables are i.i.d. Slots are either state-1-slots or state-2-slots and state transitions are only allowed at slot boundaries. These transitions mark the beginnings and ends of state-1-periods and state-2-periods. Each period has its own distribution for its period lengths. If we denote by r_{ik} the length of the kth state-i-period, then the series $\{r_{1k}\}$ and $\{r_{2k}\}$ are two (different) sets of i.i.d. random variables.

In the remainder of this section we define the necessary distributions and parameters. We note here that throughout this paper we will always use i to indicate a system state and we will further not explicitly repeat that it can only take values 1 and 2. Furthermore, we will use the notation $\bar{\imath}$ to refer to the *other* state:

$$\bar{\imath} \triangleq 3 - i \,. \tag{2}$$

For the period lengths we have

$$r_i(n) \triangleq \text{Prob}[\text{state-}i\text{-period has } n \text{ slots}] \,; \tag{3}$$

$$R_i(z) \triangleq \sum_{n=1}^{\infty} r_i(n) z^n \,; \tag{4}$$

$$\bar{r}_i \triangleq \sum_{n=1}^{\infty} n r_i(n) = R_i'(1) \,. \tag{5}$$

The probability that an arbitrary slot is a state-1-slot corresponds to the fraction of time during which the system is in state-1 and is given by

$$\sigma \triangleq \frac{\bar{r}_1}{\bar{r}_1 + \bar{r}_2} \,. \tag{6}$$

The distributions of the number of servers available during a slot are

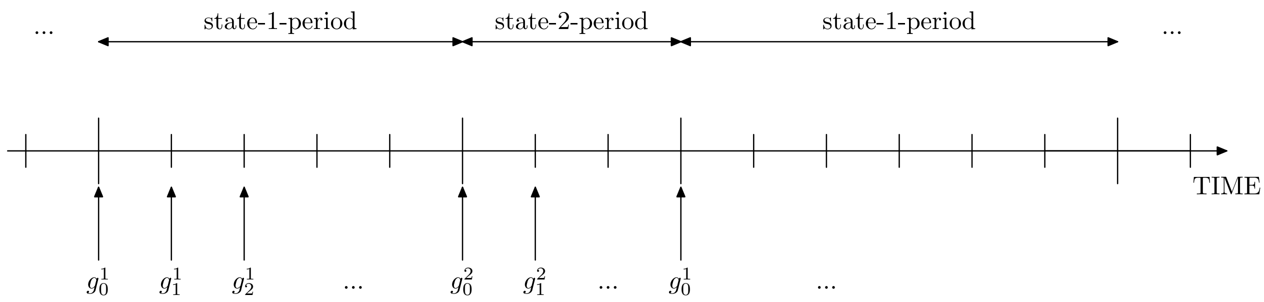

Figure 1: Illustration of the random variables g_k^i.

$$s_i(n) \triangleq \text{Prob}[n \text{ servers available during state-}i\text{-slot}] \; ; \tag{7}$$

$$S_i(z) \triangleq \sum_{n=1}^{\infty} s_i(n) z^n \; ; \tag{8}$$

$$\overline{s}_i \triangleq \sum_{n=1}^{\infty} n s_i(n) = S_i'(1) \; . \tag{9}$$

The arrival process is described by

$$c_i(n) \triangleq \text{Prob}[n \text{ customers arrive during state-}i\text{-slot}] \; ; \tag{10}$$

$$C_i(z) \triangleq \sum_{n=0}^{\infty} c_i(n) z^n \, ; \tag{11}$$

$$\lambda_i \triangleq \sum_{n=1}^{\infty} n c_i(n) = C_i'(1) \; . \tag{12}$$

The average arrival intensity λ is then given by

$$\lambda \triangleq \sigma \lambda_1 + (1-\sigma)\lambda_2 \, . \tag{13}$$

For the system to be stable the average arrival intensity needs to be strictly smaller than the average number of customers that can be served [22]. The stability condition is therefore expressed as

$$\lambda < \sigma \overline{s}_1 + (1-\sigma)\overline{s}_2 \, . \tag{14}$$

It will prove useful to introduce the following notation:

$$Y_i(z) \triangleq C_i(z) \, S_i\left(\frac{1}{z}\right) \; . \tag{15}$$

3. Mathematical analysis

To analyze the behavior of the considered 2-state queueing system, we introduce system equations similar to (1). Let us first define the stochastic variables g_k^i, $k \geq 0$, as the system content at the beginning of the

$(k+1)$st slot of a state-i-period, with corresponding probability generating function (pgf) $G_k^i(z)$. Figure 1 illustrates the definition of these random variables. For the system equations we get

$$g_k^i = (g_{k-1}^i - s_k^i)^+ + c_k^i \,, \tag{16}$$

with s_k^i the number of servers available in the kth slot of a state-i-period and with c_k^i the number of arrivals in such a slot.

The next step is the transformation of (16) into the z-domain. By the law of total expectation the pgf $G_k^i(z)$ then follows as

$$\begin{aligned}
G_k^i(z) &= C_i(z)\,\mathrm{E}_{s_k^i}\left[\mathrm{E}_{g_{k-1}^i}\left[z^{(g_{k-1}^i - s_k^i)^+}\middle| s_k^i\right]\right] \\
&= C_i(z)\sum_{j=1}^{\infty}\left[\sum_{l=j}^{\infty}\mathrm{Prob}[g_{k-1}^i = l]s_i(j)z^{l-j} + \sum_{l=0}^{j-1}\mathrm{Prob}[g_{k-1}^i = l]s_i(j)\right] \\
&= C_i(z)\sum_{j=1}^{\infty}\left[\sum_{l=0}^{\infty}\mathrm{Prob}[g_{k-1}^i = l]z^l s_i(j)z^{-j} + \sum_{l=0}^{j-1}\mathrm{Prob}[g_{k-1}^i = l]s_i(j)(1-z^{l-j})\right] \\
&= Y_i(z)\,G_{k-1}^i(z) + C_i(z)\sum_{l=0}^{\infty}\sum_{j=1}^{\infty}\mathrm{Prob}[g_{k-1}^i = l]s_i(l+j)\left(1-z^{-j}\right)\,,
\end{aligned} \tag{17}$$

where to remove the $(.)^+$ operator we first have distinguished between the cases $g_{k-1}^i \geq s_k^i$ and $g_{k-1}^i < s_k^i$; in the last step we have introduced the appropriate pgfs in the first double summation, while in the second we have changed the order of the summations over j and l and then used the substitution $j \to j+l$. Recursive application of (17) leads to

$$G_k^i(z) = [Y_i(z)]^k\,G_0^i(z) + \frac{1}{S_i\left(\frac{1}{z}\right)}\sum_{m=1}^{k}[Y_i(z)]^m\sum_{l=0}^{\infty}\sum_{j=1}^{\infty}\mathrm{Prob}\left[g_{k-m}^i = l\right]s_i(l+j)(1-z^{-j})\,. \tag{18}$$

Based on (18), we derive in the next subsections the pgfs of the system content at the beginning of a state-i-period, at the beginning of an arbitrary state-i-slot and at the beginning of an arbitrary slot.

3.1. System content at the beginning of a period

In order to obtain a relationship between $G_0^i(z)$ and $G_0^{\bar{\imath}}(z)$, we express that the system content at the beginning of a state-i-period equals the system content at the end of a state-$\bar{\imath}$-period. We then get the following expression for the function $G_0^{\bar{\imath}}(z)$:

$$G_0^{\bar{\imath}}(z) = \sum_{k=1}^{\infty} r_i(k)G_k^i(z)\,. \tag{19}$$

Let us further work out (19) using (18). This yields

$$
\begin{aligned}
\sum_{k=1}^{\infty} r_i(k) G_k^i(z) =& \sum_{k=1}^{\infty} r_i(k) \left[Y_i(z)\right]^k G_0^i(z) \\
&+ \frac{1}{S_i\left(\frac{1}{z}\right)} \sum_{k=1}^{\infty} r_i(k) \sum_{m=1}^{k} \left[Y_i(z)\right]^m \sum_{l=0}^{\infty} \sum_{j=1}^{\infty} \mathrm{Prob}\left[g_{k-m}^i = l\right] s_i(l+j)(1-z^{-j}) \\
=& R_i(Y_i(z))\, G_0^i(z) \\
&+ \frac{1}{S_i\left(\frac{1}{z}\right)} \sum_{m=1}^{\infty} \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \sum_{j=1}^{\infty} \mathrm{Prob}\left[g_k^i = l\right] s_i(l+j) r_i(k+m) \left[Y_i(z)\right]^m (1-z^{-j}) \\
=& R_i(Y_i(z))\, G_0^i(z) + \frac{1}{S_i\left(\frac{1}{z}\right)} \left[Q_i(Y_i(z)\,,1) - Q_i\left(Y_i(z)\,, \frac{1}{z}\right)\right] . \qquad (20)
\end{aligned}
$$

Note that in the second step we have changed the order of the summations over k and m and used the substitution $k \to k+m$. The unknown functions $Q_i(x,z)$ are defined as

$$
Q_i(x,z) \triangleq \sum_{m=1}^{\infty} \sum_{j=1}^{\infty} q_i(m,j) x^m z^j \,; \qquad (21)
$$

$$
q_i(m,j) \triangleq \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \mathrm{Prob}\left[g_k^i = l\right] s_i(l+j) r_i(k+m)\,. \qquad (22)
$$

From (19) and (20) we then get a set of 2 linear equations for the pgfs $G_0^1(z)$ and $G_0^2(z)$:

$$
G_0^i(z) = R_{\bar{i}}(Y_{\bar{i}}(z))\, G_0^{\bar{i}}(z) + \frac{1}{S_{\bar{i}}\left(\frac{1}{z}\right)} \left[Q_{\bar{i}}(Y_{\bar{i}}(z)\,,1) - Q_{\bar{i}}\left(Y_{\bar{i}}(z)\,, \frac{1}{z}\right)\right] , \quad i = 1,2\,. \qquad (23)
$$

Solving this set of equations we find the following explicit expression for the function $G_0^i(z)$:

$$
G_0^i(z) = \frac{S_{\bar{i}}\left(\frac{1}{z}\right) R_{\bar{i}}(Y_{\bar{i}}(z)) \left[Q_i(Y_i(z)\,,1) - Q_i\left(Y_i(z)\,,\frac{1}{z}\right)\right] + S_i\left(\frac{1}{z}\right) \left[Q_{\bar{i}}(Y_{\bar{i}}(z)\,,1) - Q_{\bar{i}}\left(Y_{\bar{i}}(z)\,,\frac{1}{z}\right)\right]}{S_i\left(\frac{1}{z}\right) S_{\bar{i}}\left(\frac{1}{z}\right) \left[1 - R_i(Y_i(z))\, R_{\bar{i}}(Y_{\bar{i}}(z))\right]} . \qquad (24)
$$

3.2. System content at the beginning of an arbitrary slot of a period

Similar to the definition of g_k^i as the system content at the beginning of the $(k+1)$st slot of a state-i-period, we introduce the variable g^i as the system content at the beginning of an arbitrary slot of a state-i-period, with corresponding pgf $G^i(z)$. The ordinate of this arbitrary slot within its period will be noted as K_i. We can then write

$$
G^i(z) = \sum_{k=1}^{\infty} \mathrm{Prob}[K_i = k]\, G_{k-1}^i(z)\ . \qquad (25)
$$

It is known that the random variable K_i has the following probability mass function (pmf) and pgf, see e.g. [23]:

$$
\mathrm{Prob}[K_i = k] = \frac{\sum_{n=k}^{\infty} r_i(n)}{\overline{r}_i}\,; \qquad (26)
$$

$$
\sum_{k=1}^{\infty} \mathrm{Prob}[K_i = k]\, z^k = \frac{z\left[R_i(z) - 1\right]}{(z-1)\overline{r}_i}\,. \qquad (27)
$$

Substituting (18) into (25), then in a similar way as before changing the order of the summations over k and m, now using the substitution $k \to k+m+1$, and finally introducing (26) and (27), we can derive

$$\begin{aligned}
G^i(z) =& \frac{1}{\overline{r}_i} G_0^i(z) + \sum_{k=2}^{\infty} \mathrm{Prob}[K_i = k] \Bigg\{ Y_i(z)^{k-1}\, G_0^i(z) \\
&+ \frac{1}{S_i\left(\frac{1}{z}\right)} \sum_{m=1}^{k-1} Y_i(z)^m \sum_{l=0}^{\infty}\sum_{j=1}^{\infty} \mathrm{Prob}\left[g_{k-1-m}^i = l\right] s_i(l+j)(1-z^{-j}) \Bigg\} \\
=& \sum_{k=1}^{\infty} \mathrm{Prob}[K_i = k]\, Y_i(z)^{k-1}\, G_0^i(z) \\
&+ \frac{1}{S_i\left(\frac{1}{z}\right)} \sum_{m=1}^{\infty}\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\sum_{j=1}^{\infty} \mathrm{Prob}[K_i = k+m+1]\, \mathrm{Prob}\left[g_k^i = l\right] s_i(l+j) Y_i(z)^m\, (1-z^{-j}) \\
=& G_0^i(z)\, \frac{R_i(Y_i(z))-1}{[Y_i(z)-1]\,\overline{r}_i} + \frac{1}{S_i\left(\frac{1}{z}\right)} \left[T_i(Y_i(z)\,,1) - T_i\left(Y_i(z)\,, \frac{1}{z}\right)\right] ,
\end{aligned} \tag{28}$$

where the unknown functions $T_i(x,z)$ are defined as

$$\begin{aligned}
T_i(x,z) &\triangleq \sum_{m=1}^{\infty}\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\sum_{j=1}^{\infty} \frac{\sum_{n=k+m+1}^{\infty} r_i(n)}{\overline{r}_i} \mathrm{Prob}\left[g_k^i = l\right] s_i(l+j) x^m z^j \\
&= \frac{1}{\overline{r}_i} \sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\sum_{j=1}^{\infty}\sum_{n=2}^{\infty}\sum_{m=1}^{n-1} \mathrm{Prob}\left[g_k^i = l\right] s_i(l+j) r_i(n+k) x^m z^j \\
&= \frac{1}{\overline{r}_i} \sum_{n=2}^{\infty}\sum_{j=1}^{\infty}\sum_{k=0}^{\infty}\sum_{l=0}^{\infty} \mathrm{Prob}\left[g_k^i = l\right] s_i(l+j) r_i(n+k) \frac{(x^n - x)}{x-1} z^j \\
&= \frac{Q_i(x,z) - xQ_i(1,z)}{\overline{r}_i\,(x-1)} ,
\end{aligned} \tag{29}$$

in view of (21) and (22). Combination of (28) and (29) then leads to

$$G^i(z) = G_0^i(z)\, \frac{R_i(Y_i(z))-1}{\overline{r}_i\,[Y_i(z)-1]} + \frac{Q_i(Y_i(z)\,,1) - Y_i(z)\, Q_i(1,1) - Q_i\left(Y_i(z)\,, \frac{1}{z}\right) + Y_i(z)\, Q_i\left(1, \frac{1}{z}\right)}{\overline{r}_i S_i\left(\frac{1}{z}\right)\, [Y_i(z)-1]} . \tag{30}$$

Finally, the pgf $G(z)$ of the system content at the beginning of an arbitrary slot can be expressed as

$$G(z) = \sigma G^1(z) + (1-\sigma) G^2(z) , \tag{31}$$

where the probability σ of having a state-1-slot is given by (6).

3.3. Finding the remaining unknowns

Until now, our results still contain an infinite number of unknowns, present in the functions $Q_i(x,z)$. In this subsection we will impose a restriction on some of the system characteristics. We will prove that this restriction reduces the unknowns to a finite number and we will provide a method to determine them.

In particular, in the remainder of our analysis we limit the pgfs $R_i(z)$ and $S_i(z)$ to be rational functions of their arguments. Let us thus write

$$R_i(z) = \frac{A_r^i(z)}{B_r^i(z)}\,, \tag{32}$$

with $A_r^i(z)$ and $B_r^i(z)$ mutually prime polynomials given by

$$A_r^i(z) \triangleq \sum_{l=1}^{m_\alpha^i} \alpha_l^i z^l \quad , \quad A_r^i(1) = 1\,; \tag{33}$$

$$B_r^i(z) \triangleq \sum_{j=0}^{m_\beta^i} \beta_j^i z^j \quad , \quad B_r^i(1) = 1\,, \tag{34}$$

and

$$S_i(z) = \frac{A_s^i(z)}{B_s^i(z)}\,, \tag{35}$$

with $A_s^i(z)$ and $B_s^i(z)$ mutually prime polynomials given by

$$A_s^i(z) \triangleq \sum_{l=1}^{m_\gamma^i} \gamma_l^i z^l \quad , \quad A_s^i(1) = 1\,; \tag{36}$$

$$B_s^i(z) \triangleq \sum_{j=0}^{m_\delta^i} \delta_j^i z^j \quad , \quad B_s^i(1) = 1\,. \tag{37}$$

We also introduce the following notations:

$$m_r^i \triangleq \max\bigl(m_\alpha^i, m_\beta^i\bigr)\ ; \tag{38}$$

$$m_s^i \triangleq \max\bigl(m_\gamma^i, m_\delta^i\bigr)\ . \tag{39}$$

Next, we recall the equations (21) and (22) for the definition of the unknown functions $Q_i(x,z)$. We can combine and rewrite these into

$$Q_i(x,z) = \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} \mathrm{Prob}\bigl[g_k^i = l\bigr]\, S_l^i(z)\, R_k^i(x)\ , \tag{40}$$

where we have introduced the following functions:

$$S_l^i(z) \triangleq \sum_{j=1}^{\infty} s_i(l+j) z^j \quad , \; l \geq 0\,; \tag{41}$$

$$R_k^i(x) \triangleq \sum_{m=1}^{\infty} r_i(k+m) x^m \quad , \; k \geq 0\,. \tag{42}$$

We will now prove the functions $S_l^i(z)$ and $R_k^i(x)$ have the following properties:

(a) they are rational functions of their arguments;

(b) their denominator equals $B_s^i(z)$ and $B_r^i(x)$ respectively;

(c) their numerator is of maximum degree m_s^i and m_r^i respectively and

(d) they are divisible by their argument, i.e. $S_l^i(0) = R_k^i(0) = 0$.

Proof: Property (d) clearly holds by definition. To prove (a) to (c) we use mathematical induction. The base step follows from the observation that

$$\begin{aligned} S_0^i(z) &= S_i(z) \ ; \\ R_0^i(z) &= R_i(z) \ . \end{aligned}$$

Now for the induction step, assume that the propositions are valid for $S_l^i(z)$, then we get for $S_{l+1}^i(z)$:

$$\begin{aligned} S_{l+1}^i(z) &= \sum_{j=1}^{\infty} s_i(l+1+j) z^j \\ &= \frac{\sum_{j=1}^{\infty} s_i(l+j) z^j}{z} - s_i(l+1) \\ &= \frac{S_l^i(z)}{z} - s_i(l+1) \, , \end{aligned}$$

which proves (a) to (c). Analogously, when the propositions are valid for $R_k^i(x)$, then we get for $R_{k+1}^i(x)$:

$$\begin{aligned} R_{k+1}^i(x) &= \sum_{m=1}^{\infty} r_i(k+1+m) x^m \\ &= \frac{\sum_{m=1}^{\infty} r_i(k+m) x^m}{x} - r_i(k+1) \\ &= \frac{R_k^i(x)}{x} - r_i(k+1) \, , \end{aligned}$$

which proves (a) to (c). This concludes the proof. $\square$

As $Q_i(x,z)$ is a linear combination of all $S_l^i(z)\, R_k^i(x)$ (with $k, l \geq 0$), we can state based on the above properties that $Q_i(x,z)$ is a rational function in x and z with a numerator of degree m_r^i in x and of degree m_s^i in z and with denominator equal to $B_r^i(x)\, B_s^i(z)$:

$$Q_i(x,z) = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \mathrm{Prob}\big[g_k^i = l\big]\, S_l^i(z)\, R_k^i(x) = \frac{A_q^i(x,z)}{B_r^i(x)\, B_s^i(z)} \, , \tag{43}$$

where $A_q^i(x,z)$ is of the following form:

$$A_q^i(x,z) \triangleq \sum_{m=1}^{m_r^i} \sum_{j=1}^{m_s^i} \epsilon_{mj}^i x^m z^j \ . \tag{44}$$

This reduces the infinite number of unknowns given by all $\mathrm{Prob}\big[g_k^i = l\big]$ to the finite number $m_r^i m_s^i$, so that the total number of unknowns M is given by

$$M \triangleq m_r^1 m_s^1 + m_r^2 m_s^2 \,. \tag{45}$$

In order to determine these M unknowns, we will rely on the properties of pgfs, namely that they are normalized and that they are analytical within the complex unit disk. Let us consider (24) for $i = 1$ while taking into account (32), (35) and (43). After moving all the poles to the denominator, we get the following expressions for the numerator $N_1(z)$ and denominator $D_1(z)$ of $G_0^1(z)$:

$$\begin{aligned} N_1(z) =& A_r^2(Y_2(z))\, B_s^2\left(\frac{1}{z}\right)^{m_r^2} z^{m_r^2 m_s^2} A_s^2\left(\frac{1}{z}\right) z^{m_s^2} \\ & \left[A_q^1(Y_1(z)\,,1)\, B_s^1\left(\frac{1}{z}\right) - A_q^1\left(Y_1(z)\,,\frac{1}{z}\right)\right] B_s^1\left(\frac{1}{z}\right)^{m_r^1} z^{m_r^1 m_s^1} z^{m_s^1} \\ & + B_r^1(Y_1(z))\, B_s^1\left(\frac{1}{z}\right)^{m_r^1} z^{m_r^1 m_s^1} A_s^1\left(\frac{1}{z}\right) z^{m_s^1} \\ & \left[A_q^2(Y_2(z)\,,1)\, B_s^2\left(\frac{1}{z}\right) - A_q^2\left(Y_2(z)\,,\frac{1}{z}\right)\right] B_s^2\left(\frac{1}{z}\right)^{m_r^2} z^{m_r^2 m_s^2} z^{m_s^2}\,; \end{aligned} \tag{46}$$

$$\begin{aligned} D_1(z) =& A_s^1\left(\frac{1}{z}\right) z^{m_s^1} A_s^2\left(\frac{1}{z}\right) z^{m_s^2} \\ & \Bigg\{ B_r^1(Y_1(z))\, B_s^1\left(\frac{1}{z}\right)^{m_r^1} z^{m_r^1 m_s^1} B_r^2(Y_2(z))\, B_s^2\left(\frac{1}{z}\right)^{m_r^2} z^{m_r^2 m_s^2} \\ & - A_r^1(Y_1(z))\, B_s^1\left(\frac{1}{z}\right)^{m_r^1} z^{m_r^1 m_s^1} A_r^2(Y_2(z))\, B_s^2\left(\frac{1}{z}\right)^{m_r^2} z^{m_r^2 m_s^2} \Bigg\}\,. \end{aligned} \tag{47}$$

Let us look more closely at the first parts of (47):

- The location of the zeros of $A_s^1\left(\frac{1}{z}\right) z^{m_s^1} A_s^2\left(\frac{1}{z}\right) z^{m_s^2}$ is unspecified. However, we can conclude that the numerator of $G_0^1(z)$ is divisible by this part, by observing that the function $A_q^i(x,z)$ is divisible by both of its arguments, such that in view of (15) $A_q^i(Y_i(z)\,,z)$ is divisible by $A_s^i\left(\frac{1}{z}\right)$.
- $B_r^i(x)$ has m_β^i zeros x_j^i outside the complex unit disk. By application of Rouché's theorem (see e.g. [24]) we get that $B_s^i\left(\frac{1}{z}\right) z^{m_s^i} x_j^i - C_i(z)\, A_s^i\left(\frac{1}{z}\right) z^{m_s^i} = 0$ has m_s^i solutions within the complex unit disk. Thus $B_r^i(Y_i(z))$ has $m_s^i m_\beta^i$ zeros within the complex unit disk. Multiplying this expression with $B_s^i\left(\frac{1}{z}\right)^{m_r^i} z^{m_r^i m_s^i}$ removes its poles and adds another $(m_r^i - m_\beta^i) m_s^i$ zeros which are all inside the complex unit disk. We can thus conclude that $B_r^i(Y_i(z))\, B_s^i\left(\frac{1}{z}\right)^{m_r^i} z^{m_r^i m_s^i}$ as a whole has $m_s^i m_r^i$ zeros within the complex unit disk.

The first part of (47) thus has $M = m_s^1 m_r^1 + m_s^2 m_r^2$ zeros within the complex unit disk. By application of Rouché's theorem to $D_1(z)$ we can conclude that the denominator of $G_0^1(z)$ as a whole also has M zeros within the complex unit disk. For these zeros, the numerator must then vanish due to the property that pgfs are analytical within the complex unit disk. As it can be easily observed that $z = 1$ is already a zero of both numerator and denominator, independently of the unknowns, we obtain $M - 1$ linearly independent equations to determine the remaining unknowns. The required extra equation can be found by expressing the normalization condition for pgfs:

$$\lim_{z\to 1} G_0^1(z) = \frac{\frac{\partial}{\partial z} Q_1(x,z)\big|_{(x,z)=(1,1)} + \frac{\partial}{\partial z} Q_2(x,z)\big|_{(x,z)=(1,1)}}{\overline{r}_1(\overline{s}_1 - \lambda_1) + \overline{r}_2(\overline{s}_2 - \lambda_2)} = 1\,. \tag{48}$$

4. Important performance measures

The determination of the exact form of the functions $Q_i(x,z)$ is only an intermediate step in the calculation of some important performance characteristics of the queueing system. In this section we are particularly interested in the mean value of the system content at several observation epochs:

- mean system content at the beginning of a state-i-period: $\mathrm{E}\big[g_0^i\big] = \frac{d}{dz} G_0^i(z)\big|_{z=1}$;
- mean system content at the beginning of an arbitrary state-i-slot: $\mathrm{E}\big[g^i\big] = \frac{d}{dz} G^i(z)\big|_{z=1}$;
- mean system content at the beginning of an arbitrary slot: $\mathrm{E}[g] = \frac{d}{dz} G(z)\big|_{z=1}$.

Let us first introduce the following notations:

$$
\begin{aligned}
Q_z^i(1,1) &\triangleq \frac{\partial}{\partial z} Q_i(x,z)\bigg|_{(x,z)=(1,1)} \\
&= \frac{\partial}{\partial z} A_q^i(x,z)\bigg|_{(x,z)=(1,1)} - \frac{d}{dz} B_s^i(z)\bigg|_{z=1} A_q^i(1,1) \\
&= \sum_{m=1}^{m_r^i} \sum_{j=1}^{m_s^i} \Big[j - {B_s^i}'(1) \Big]\, \epsilon_{mj}^i \; ;
\end{aligned}
\tag{49}
$$

$$
\begin{aligned}
Q_{xz}^i(1,1) &\triangleq \frac{\partial^2}{\partial x \partial z} Q_i(x,z)\bigg|_{(x,z)=(1,1)} \\
&= \frac{\partial^2}{\partial x \partial z} A_q^i(x,z)\bigg|_{(x,z)=(1,1)} - \frac{\partial}{\partial x} A_q^i(x,z)\bigg|_{(x,z)=(1,1)} \frac{d}{dz} B_s^i(z)\bigg|_{z=1} - Q_z^i(1,1)\, \frac{d}{dx} B_r^i(x)\bigg|_{x=1} \\
&= \sum_{m=1}^{m_r^i} \sum_{j=1}^{m_s^i} m \Big[j - {B_s^i}'(1) \Big]\, \epsilon_{mj}^i - {B_r^i}'(1)\, Q_z^i(1,1) \; ;
\end{aligned}
\tag{50}
$$

$$
\begin{aligned}
Q_{zz}^i(1,1) &\triangleq \frac{\partial^2}{\partial z^2} Q_i(x,z)\bigg|_{(x,z)=(1,1)} \\
&= \frac{\partial^2}{\partial z^2} A_q^i(x,z)\bigg|_{(x,z)=(1,1)} - 2\, \frac{d}{dz} B_s^i(z)\bigg|_{z=1} Q_z^i(1,1) - \frac{d^2}{dz^2} B_s^i(z)\bigg|_{z=1} A_q^i(1,1) \\
&= \sum_{m=1}^{m_{ri}} \sum_{j=1}^{m_s^i} \Big[j(j-1) - {B_s^i}''(1) \Big]\, \epsilon_{mj}^i - 2{B_s^i}'(1)\, Q_z^i(1,1) \; .
\end{aligned}
\tag{51}
$$

Here the prime and double prime are used as concise notations for the first and second derivative of the functions of one variable. Furthermore we introduce ψ and χ as

$$
\begin{aligned}
\psi \triangleq \sum_{i=1}^{2} \Big\{ & 2(\lambda_i - \overline{s}_i) Q_{xz}^i(1,1) + R_i''(1)\,(\overline{s}_i - \lambda_i)^2 - Q_{zz}^i(1,1) - 2(1+\overline{s}_i) Q_z^{\overline{i}}(1,1) \\
& - \overline{r}_i \left[4\lambda_i \overline{s}_i + 2\lambda_i \overline{s}_{\overline{i}} - C_i''(1) - 2\overline{s}_i - S_i''(1) - 2\overline{s}_i \overline{s}_{\overline{i}} - 2\overline{s}_i^2 \right] \Big\} \; ;
\end{aligned}
\tag{52}
$$

$$
\chi \triangleq \prod_{i=1}^{2} \overline{r}_i (\overline{s}_i - \lambda_i) \, . \tag{53}
$$

Calculation of the first derivative of $G_0^i(z)$ from (23) and evaluation at $z=1$ leads to the mean system content at the start of a state-i-period. The result is given by

$$\mathrm{E}\big[g_0^i\big] = \frac{2\left[\lambda_{\bar{\imath}}\overline{r}_{\bar{\imath}} - \overline{r}_{\bar{\imath}}\overline{s}_{\bar{\imath}}\right] Q_z^i(1,1) + \psi + \chi}{2\overline{r}_1(\overline{s}_1 - \lambda_1) + 2\overline{r}_2(\overline{s}_2 - \lambda_2)} \,. \tag{54}$$

Note that as the average arrival intensity goes to the average number of available servers per slot, i.e. as the system becomes unstable, the denominator factor $2\overline{r}_1(\overline{s}_1 - \lambda_1) + 2\overline{r}_2(\overline{s}_2 - \lambda_2)$ in (54) goes to zero. Numerical results confirm that the mean system content indeed goes to infinity in this case.

The mean system content at the beginning of an arbitrary state-i-slot is given by

$$\mathrm{E}\big[g^i\big] = \mathrm{E}\big[g_0^i\big] - \frac{Q_z^i(1,1)}{\overline{r}_i} + \frac{Q_{xz}^i(1,1)}{\overline{r}_i} + \frac{R''(1)\,(\lambda_i - \overline{s}_i)}{2\overline{r}_i} \,. \tag{55}$$

Finally, for the mean system content at the beginning of an arbitrary slot we get

$$\mathrm{E}[g] = \sigma \mathrm{E}\big[g^1\big] + (1-\sigma)\mathrm{E}\big[g^2\big] \,. \tag{56}$$

Note that we can make the following specific choices for the input distributions: $C_1(z) = C_2(z)$, $S_1(z) = z^2$, $S_2(z) = z$ and a geometrical distribution for the state-1-period lengths. We then have a queueing system that can also be analyzed by the method described in [12]. We verified that the current analysis and the method of [12] lead to the same results for this specific special case.

5. Numerical examples

In this section we will look at some examples of queueing systems that can be analyzed with the developed method. The section is also used to validate the method by simulation. On all graphs in this section, the lines represent the theoretical results, while the marks represent values obtained by simulation.

Let us first introduce an interesting probability distribution: the mixture of two geometrics. This distribution will prove to be a convenient choice for the period lengths and for the number of available servers. For a random variable f, the corresponding pmf is as follows:

$$\mathrm{Prob}[f=n] \triangleq \omega(1-\alpha_1)\alpha_1^{n-1} + (1-\omega)(1-\alpha_2)\alpha_2^{n-1} \,, \quad n \geq 1 \,, \tag{57}$$

where all 3 parameters have a value between 0 and 1: $0 < \omega,\ \alpha_1,\ \alpha_2 < 1$. The average $\overline{f}$ for this distribution is given by

$$\overline{f} = \frac{\omega}{1-\alpha_1} + \frac{1-\omega}{1-\alpha_2} \,. \tag{58}$$

As a measure of the variance of a distribution we will use the squared coefficient of variation C_f^2:

$$C_f^2 \triangleq \frac{\mathrm{var}[f]}{\overline{f}^2} \,. \tag{59}$$

In order to reduce the number of parameters, we still add the following (arbitrary) constraint:

$$\frac{\omega}{1-\alpha_1} = \frac{1-\omega}{1-\alpha_2} \,. \tag{60}$$

Then we can fully describe the distribution of f based on two meaningful parameters: $\overline{f}$ and C_f^2. The advantage of this distribution is that it gives a large freedom of choice for average value ($1 < \overline{f} < \infty$) and variance ($1 - \frac{1}{\overline{f}} \leq C_f^2 < \infty$), while remaining fairly simple (only 2 parameters) and its pgf $F(z)$ is given by the ratio of two second-degree polynomials:

$$F(z) = \frac{(1+\overline{f}C_f^2)z - \overline{f}C_f^2 z^2}{(1+\overline{f}+\overline{f}C_f^2)\frac{\overline{f}}{2} - (1+\overline{f}+\overline{f}C_f^2)(\overline{f}-1)z + \frac{2-(1+\overline{f}+\overline{f}C_f^2)(2-\overline{f})}{2}z^2} . \tag{61}$$

Note that if $1 - \frac{1}{\overline{f}} = C_f^2$, expression (61) reduces to $\frac{z}{\overline{f}-(\overline{f}-1)z}$, which corresponds to a geometric distribution.

In a first set of examples we study the impact of the mean values and variances of the numbers of servers available and the period lengths and in particular the effect of the server unbalance between the 2 states on the system performance. A second set of examples focuses on the influence of traffic burstiness on the system behavior.

5.1. Unbalanced server availability

One way to use the model is to look at situations where the arrival process to the system is always of the same nature, but the server availability is not. Throughout this subsection, we choose the lengths of both periods to have the same distribution (which leads to $\sigma = 0.5$) according to a mixture of 2 geometrics as described above.

$$R_1(z) = R_2(z) . \tag{62}$$

For the distribution of the numbers of servers available we also use a mixture of 2 geometrics. We choose an overall average of 10 available servers:

$$\sigma\overline{s}_1 + (1-\sigma)\overline{s}_2 = 10 \iff \overline{s}_1 + \overline{s}_2 = 20 . \tag{63}$$

We introduce the server unbalance Δs as

$$\Delta s \triangleq |\overline{s}_2 - \overline{s}_1| . \tag{64}$$

We consider arrivals either according to the well-known Poisson process or according to a so-called "disaster" process. In the latter case, there is a high probability of no arrivals and a low probability of a very high number of arrivals. Their respective pgfs are given by

$$C_{\text{Poisson}}(z) = e^{\lambda(z-1)} ; \tag{65}$$

$$C_{\text{disaster}}(z) = (1-\alpha) + \alpha z^N . \tag{66}$$

Note that for the disaster-process, N is an integer number and the arrival intensity is given by $\lambda = \alpha N$. We use for both states the same type of arrivals with the same parameters.

$$C_1(z) = C_2(z) . \tag{67}$$

In a first example we look at the influence of the average period lengths and the average numbers of available servers on the system performance. Specifically, in Figure 2 we plot the mean system content at the beginning of an arbitrary slot versus the server unbalance Δs for the case of Poisson arrivals. We look at 4 cases of correlated server availability (thick lines), all with $C_{ri}^2 = 1$ and different mean period lengths. The

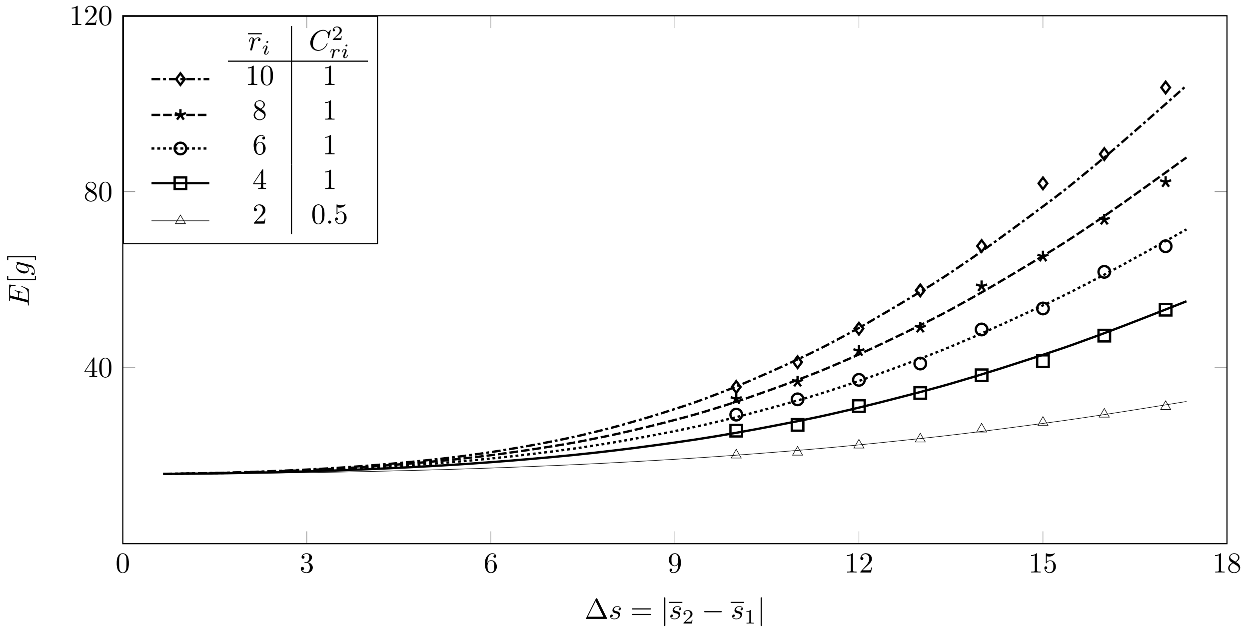

Figure 2: Mean system content in function of server unbalance, Poisson arrivals with $\lambda = 7$, $C^2_{si} = 1$ and for different mean period lengths.

average arrival intensity considered in Figure 2 is $\lambda = 7$, and $C^2_{si} = 1$. We see that the mean system content increases monotonously with increasing Δs. The effect is stronger with longer average period lengths (i.e. with higher correlation on the number of servers available from slot to slot). We also consider the case of $C^2_{ri} = 0.5$ and $\overline{r}_i = 2$, which corresponds to a geometric distribution with parameter 0.5 for the period lengths. In this case the correlation on the number of available servers is removed; the resulting mean queue content is plotted with thin line. We observe that correlation has a considerable effect on the system performance.

In Figure 3 we repeat the experiment of Figure 2, but now we consider arrivals according to a disaster process. We choose $N = 70$ and $\alpha = 0.1$ which corresponds to an average arrival intensity $\lambda = 7$. The variance of the number of arrivals from slot to slot is now much larger, which results in larger mean system contents. The effects of server unbalance in relation to the average period lengths remains, the system content increases with increasing Δs and the effect is stronger with longer average period lengths. The effect is not negligible even if the mean system content is already increased due to the higher variance of the arrival process.

When the server unbalance is small, and thus both periods have an average number of available servers close to 10, the system is able to efficiently clear all the incoming customers independently of its state, and there is little influence of the average period lengths on the mean system content. For a larger server unbalance, we get a situation where the system states can be identified as a state of work accumulation and a state of work removal: the average number of available servers is smaller, respectively larger than the average number of incoming customers per slot. When this occurs, the average period lengths have a higher impact on the mean system content. This is because the longer the periods are, the longer the system accumulates work before the queue can be emptied during a state of work removal. The larger the server unbalance is, the stronger the effect of longer periods, as the work will accumulate at a higher rate during periods of work accumulation.

In the next 2 examples we investigate the effect of the variance of the period lengths and the numbers of available servers, we do this for Poisson arrivals. In Figure 4 we consider a scenario with average period lengths of 10 slots and a small server unbalance of $\Delta s = 4$. We plot the mean system content versus the arrival intensity λ. The solid line depicts the base situation with all squared coefficients of variation equal to 1. The dashed line shows the effect of an increased variance of the period lengths and the dotted line shows

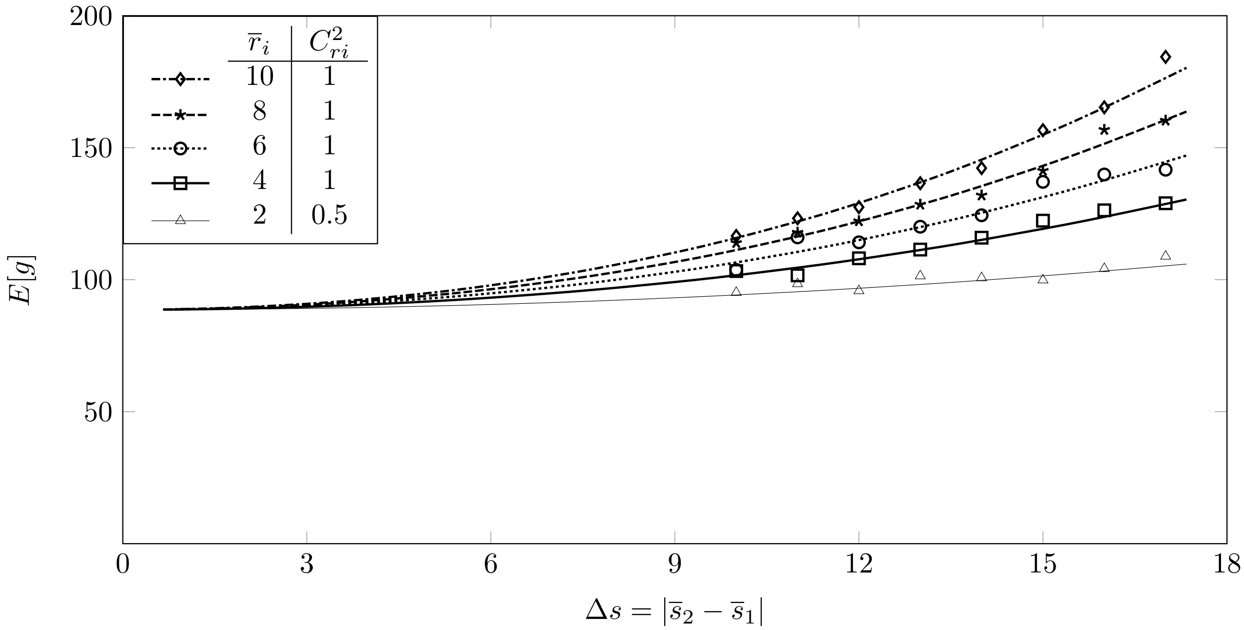

Figure 3: Mean system content in function of server unbalance, disaster arrivals with $\alpha = 0.1$, $N = 70$, $\lambda = \alpha N = 7$, $C_{si}^2 = 1$ and for different mean period lengths.

the effect of an increased variance of the numbers of available servers. We can see that in this scenario the variance of the numbers of available servers has a large impact on the mean system content; this impact is already noticeable for lower arrival intensities. This observation is explained by the fact that with a large variance on the numbers of available servers, it occurs more frequently that, regardless of the system state, in consecutive slots only a few servers are available and thus the system content increases strongly.

We also see from Figure 4 that the variance of the period lengths only has an impact when the average arrival intensity to the system goes to 8, which is the average number of servers available during state-1-slots. This is in line with the observations from Figure 2 and Figure 3; the length of the periods plays a role when one of the system states can be identified as work accumulating. A large variance of the period lengths then means that occasionally a very long period of work accumulation occurs, which significantly increases the mean system content. This delayed effect appears to be less prominent when the variance of the number of available servers in state-1-slots is lightly increased, as can be seen from the dash-dotted line. In this case, it occurs more frequently that work accumulates during state-1-periods, even for lower arrival intensities.

From Figure 4 we can conclude that in the case of a small server unbalance, a higher variance of the numbers of available servers is much worse for the system performance in terms of mean system content compared to a higher variance of the period lengths.

In Figure 5 we consider a situation with average period lengths of 10 slots and a large server unbalance of $\Delta s = 12$, and we again plot the mean system content versus the arrival intensity λ. The solid line depicts the base situation with all squared coefficients of variation equal to 1. This base situation yields a larger mean system content as compared to the base situation of Figure 4, which is due to the larger server unbalance. The dotted line shows the mean system content when the variance of the numbers of available servers is increased, and the dashed line when the variance of the period lengths is increased. It can be seen that the relative effects of the two types of variance have reversed. The dashed line (with increased C_{ri}^2) shows the highest mean system content. Because of the large server unbalance, the system is in a state of either work accumulation or work removal. When the variance of the period lengths is increased, it occurs more frequently that the system resides in a very long period of work accumulation, resulting in large mean system contents.

With the final example of this subsection we illustrate that the effect of variance is continuous and strongly

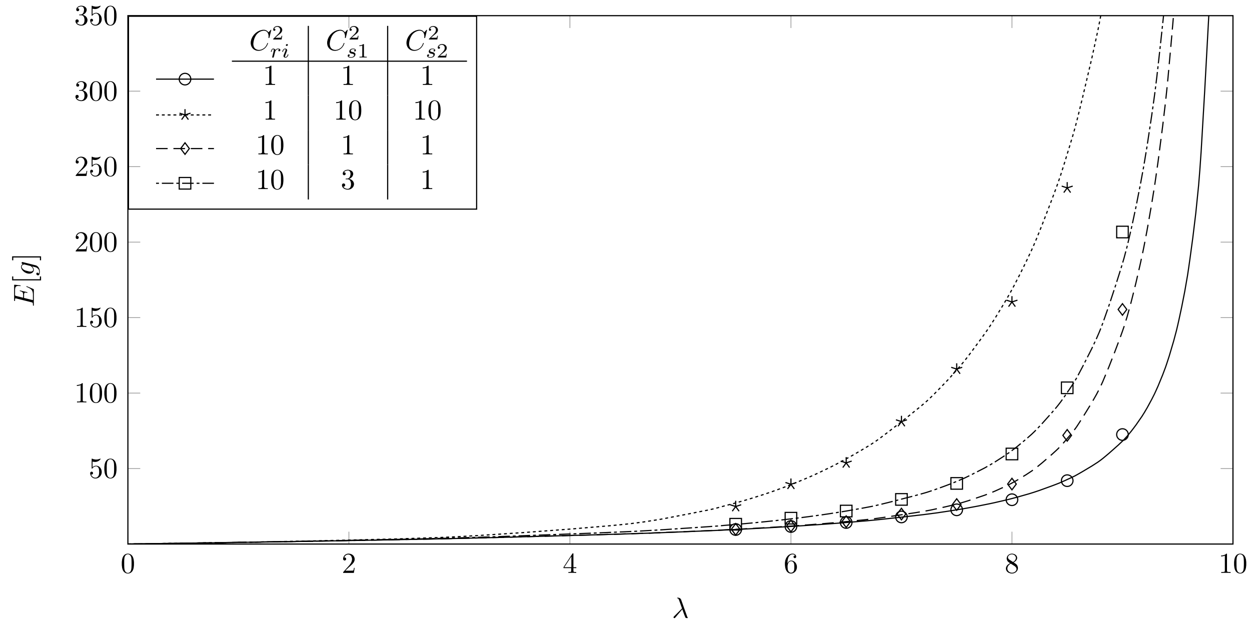

Figure 4: Mean system content in function of average arrival intensity λ, for Poisson arrivals, $\overline{r}_1 = \overline{r}_2 = 10$, $\overline{s}_1 = 8$, $\overline{s}_2 = 12$ and different values of the squared coefficients of variation.

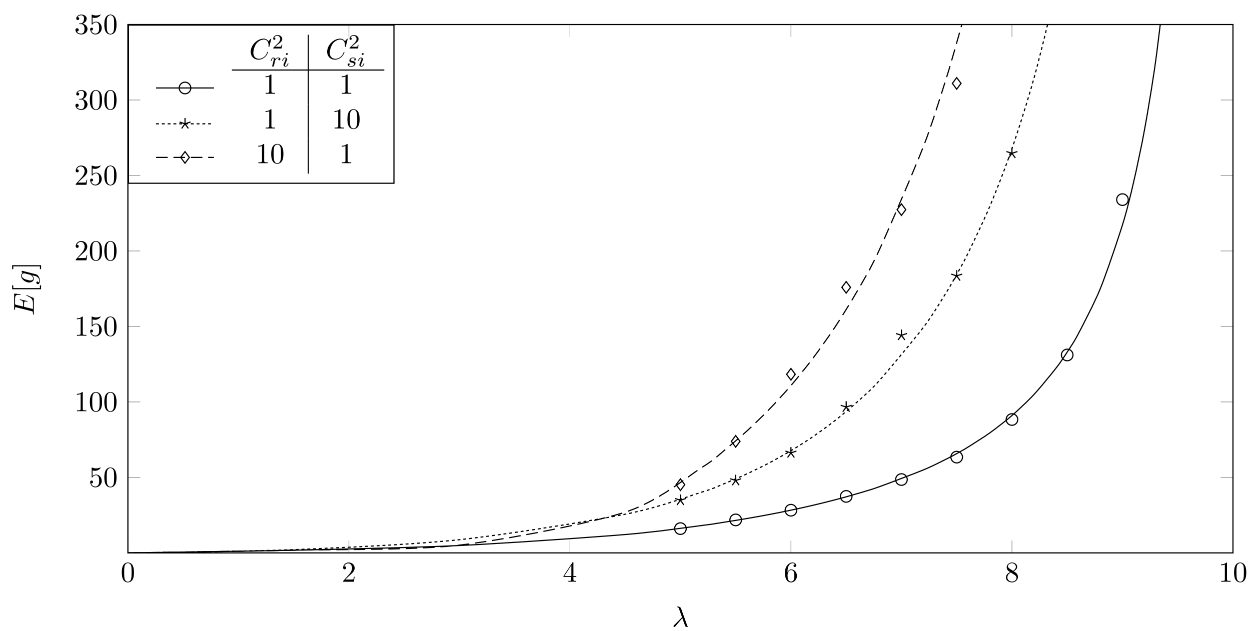

Figure 5: Mean system content in function of average arrival intensity λ, for Poisson arrivals, $\overline{r}_1 = \overline{r}_2 = 10$, $\overline{s}_1 = 4$, $\overline{s}_2 = 16$ and different values of the squared coefficients of variation.

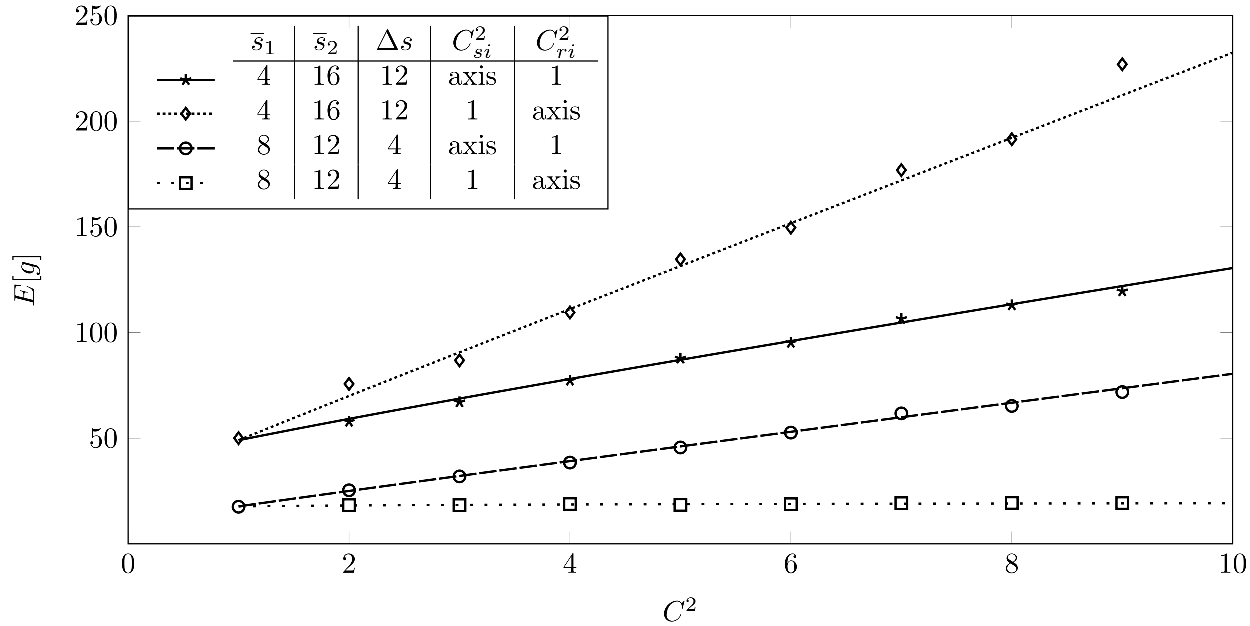

Figure 6: Mean system content in function of squared coefficient of variation, for Poisson arrivals with $\lambda = 7$, $\bar{r}_1 = \bar{r}_2 = 10$ and with different Δs.

depending on the exact configuration of a queueing system. We consider Poisson arrivals with $\lambda = 7$ and mean period lengths of 10 slots. In Figure 6 we plot the mean system content in function of the squared coefficient of variation of either the period lengths or the numbers of available servers, while keeping the other squared coefficient of variation fixed at 1. The top set of 2 lines is for a large Δs, while the bottom set is for a small Δs. For both sets, the dotted line shows the influence of the variance of the period lengths, while the solid and dashed line show the influence of the variance of the numbers of servers.

Variance is obviously bad for the performance of a queueing system, but Figure 4, Figure 5 and Figure 6 demonstrate that not all systems are equally sensitive to the same kind of variance. In general, the variance of the period lengths will be more meaningful for a queueing system when the server unbalance is larger.

5.2. Bursty arrivals

Another way to use the model of this paper is to look at a bursty arrival process. Many applications deal with some form of bursty traffic where short periods of high arrival intensity are alternated with long periods of low arrival intensity. In this subsection we will assume that 2 servers are constantly available:

$$S_1(z) = S_2(z) = z^2 \,. \tag{68}$$

We choose state-1 to be the high-traffic state with $\lambda_1 = 10\lambda_2$. To limit the number of parameters in this example we choose the period lengths to be geometrically distributed:

$$R_1(z) = \frac{(1-\alpha_1)z}{1-\alpha_1 z} \,; \tag{69}$$

$$R_2(z) = \frac{(1-\alpha_2)z}{1-\alpha_2 z} \,. \tag{70}$$

We assume Poisson arrivals for the state-2-slots:

$$C_2(z) = e^{\lambda_2(z-1)} . \tag{71}$$

In state-1-slots we consider arrivals according to a Poisson process or according to a disaster process, as introduced in the previous subsection.

$$C_{1,\text{Poisson}}(z) = e^{\lambda_1(z-1)} ; \tag{72}$$

$$C_{1,\text{disaster}}(z) = (1-\alpha) + \alpha z^N , \tag{73}$$

with N an integer and the arrival intensity given by $\lambda_1 = \alpha N$, where we take $\alpha = 0.1$.

In Figure 7 we plot the mean system content versus the average arrival intensity for these two different arrival processes during high-traffic (state-1). The thick lines in this figure correspond to $\alpha_1 = 0.8$ and $\alpha_2 = 0.96$, which leads to $\overline{r}_1 = 5$, $\overline{r}_2 = 25$ and $\sigma = \frac{1}{6}$. We observe that the mean system content is considerably higher in the disaster case, which is logical as the variability of the arrival process is a lot higher in this case. For the same mean number of arrivals per slot, it then occurs more frequently that a high number of customers enter the system during the same slot, and it takes a long time for the 2 servers to recover from this.

For the thin lines in Figure 7, the parameters α_1 and α_2 are set as $\alpha_1 = \frac{1}{6}$ and $\alpha_2 = 1 - \alpha_1 = \frac{5}{6}$. This choice changes the values of $\overline{r}_1$ and $\overline{r}_2$ without changing their ratio, i.e. without changing the value of σ. The probability for any slot to be a state-1-slot then equals α_1 and the probability for any slot to be a state-2-slot becomes $1 - \alpha_1$, regardless of the state of the previous slot, so the correlation in the arrival process is now removed. From the position of the curves in Figure 7 we again observe that the effect of correlation is clearly not negligible.

In a final experiment we investigate whether the variability of the arrival process is the main cause of the higher mean system content for the disaster case as compared to the Poisson case in Figure 7, or also other factors such as the shape of the distribution function play a role. We therefore again set $\alpha_1 = 0.8$ and $\alpha_2 = 0.96$, and consider a third type of arrival process in state-1-slots, namely a mixture of two Poisson processes, with the following pgf:

$$C_{1,\text{mix Poisson}}(z) = \omega e^{\Lambda_1(z-1)} + (1-\omega) e^{\Lambda_2(z-1)} . \tag{74}$$

We look at various values of $0 < \omega < 1$ and choose the remaining parameters $0 < \Lambda_1, \Lambda_2$ of the arrival process in such a way that the first two moments of $C_{1,\text{disaster}}(z)$ and $C_{1,\text{mix Poisson}}(z)$ are equal, i.e. the average and squared coefficient of variation are equal. The resulting mean system content for this third arrival process turns out to be indistinguishable of the disaster case (and is therefore not plotted in Figure 7). This suggests that the exact shape of the arrival process is not significant for the mean system content as soon as its first two moments are given.

6. Conclusion

In this paper we studied a discrete-time multiserver queueing model with the possibility to add correlation to the arrival process and/or the server availability. This is achieved by the introduction of two different system states. Each state is characterized by its own distribution for the number of arrivals per slot and its own distribution for the number of available servers during a slot. State changes can only occur at slot boundaries and mark the beginnings and ends of state-periods. The lengths of these periods are assumed to be two independent sets of i.i.d. stochastic variables and within a given state-period there is no correlation between the numbers of arrivals from slot to slot nor between the numbers of available servers from slot to slot. Because we do not limit the distributions of the period lengths to be geometrical, our model goes beyond the first-order Markov approach that is regularly used in literature where state changes are only dependent on the state of the last slot, and not on the time the system is already in a certain state. The model is suited

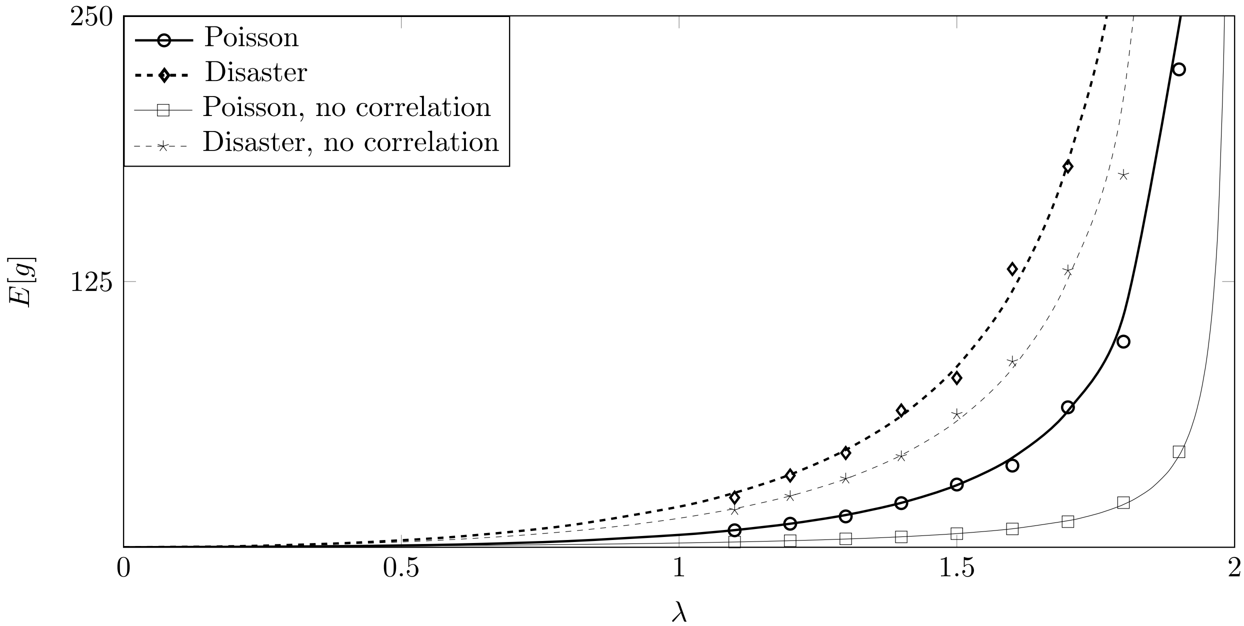

Figure 7: Mean system content in function of average arrival intensity λ, for two different arrival distributions during high-traffic periods with and without correlation.

for situations where for example long periods of light traffic are alternated with short periods of heavy traffic, while also the number of available servers can be different during light and heavy traffic.

To keep the mathematical analysis tractable, we imposed small restrictions on the distributions of the numbers of available servers and period lengths, namely they need to have rational pgfs. We obtained the distributions of the number of customers present in the queueing system at the beginning of a period, at the beginning of an arbitrary slot within a certain period and at the beginning of an arbitrary slot. For this, it is necessary to compute the zeros of a non-polynomial function within the complex unit disk, and solve a set of linear equations. For many common choices of the distributions this does not require large computational effort.

We illustrated the analysis with several numerical examples to give further insight in the practical use of the model. The examples showed that correlation in the arrival process and in the availability of servers can have a significant effect on the mean system content. The model also allows us to investigate the effects of different types of variance on the system performance. Further research could include models with more than 2 system states (where the transition from state to state is either a cyclical process or based on a Markovian transition process).

References

[1] A. Krishnamoorthy, P.K. Pramod, and S.R. Chakravarthy. Queues with interruptions: a survey. *TOP*, 22:290–320, 2014.

[2] H. Bruneel. A general treatment of discrete-time buffers with one randomly interrupted output line. *European Journal of Operations Research*, 27(1):67–81, 1986.

[3] T. Takine and B. Sengupta. A single server queue with service interruptions. *Queueing Systems*, 26(3-4):285–300, 1997.

[4] F. Kamoun. Performance evaluation of a queueing system with correlated packet-trains and server interruption. *Telecommunication Systems*, 41:267–277, August 2009.

[5] B. Feyaerts, S. De Vuyst, H. Bruneel, and S. Wittevrongel. Performance analysis of buffers with train arrivals and correlated output interruptions. *Journal of Industrial and Management Optimization*, 11(3):829–848, July 2015.

[6] A. Senthil Vadivu, R. Arumuganathan, and M. Senthil Kumar. Analysis of discrete-time queues with correlated arrivals, negative customers and server interruption. *Rairo-operations Research*, 50(1):67–81, 2016.

[7] N. D. Georganas. Buffer behavior with poisson arrivals and bulk geometric service. *IEEE Transactions on Communications*, 24:938–940, 1976.

[8] H. Bruneel. A general model for the behavior of infinite buffers with periodic service opportunities. *European Journal of Operations Research*, 16(1):98–106, 1984.

[9] K. Laevens and H. Bruneel. Delay analysis for discrete-time queueing systems with multiple randomly interrupted servers. *European Journal of Operations Research*, 85:161–177, 1995.

[10] H. Bruneel. A mathematical model for discrete-time buffer systems with correlated output process. *European Journal of Operational Research*, 18:98–110, 1984.

[11] H. Bruneel. A discrete-time queueing system with a stochastic number of servers subjected to random interruptions. *Opsearch*, 22:215–231, 1985.

[12] H. Bruneel and S. Wittevrongel. Analysis of a discrete-time single-server queue with an occasional extra server. *Performance Evaluation*, 116:119–142, 2017.

[13] S. Li. Performance of a nonblocking space-division packet switch with correlated input traffic. *IEEE Transactions on Communications*, 40(1):97–108, January 1992.

[14] J. N. Daigle, Y. Lee, and M. N. Magalhaes. Discrete time queues with phase dependent arrivals. *IEEE Transactions on Communications*, 42(234):606–614, Feb 1994.

[15] K. Elsayed and H. Perros. The superposition of discrete-time markov renewal processes with an application to statistical multiplexing of bursty traffic sources. *Appl Math Comput*, 115:43–62, 2000.

[16] J. Wang, Y. Huang, and T. V. Do. A single-server discrete-time queue with correlated positive and negative customer arrivals. *Applied Mathematical Modelling*, 37(9):6212–6224, May 2013.

[17] H. Bruneel, D. Fiems, J. Walraevens, and S. Wittevrongel. Queueing models for the analysis of communication systems. *TOP*, 22(2):421–448, 2014.

[18] Y. Zhou, R. Anderson, H. Vakilzadian, D. Moeller, and A. Deutschmann. Developing a dynamic queueing model for the airport check-in process. *2018 IEEE International Conference on Electro/Information Technology (EIT)*, pages 871–876, 2018.

[19] D. Yue and Y. Qin. A production inventory system with service time and production vacations. *Journal of Systems Science and Systems Engineering*, 28:168–180, 2019.

[20] L. Dablanc, E. Morganti, N. Arvidsson, J. Woxenius, M. Browne, and N. Saidi. The rise of on-demand 'instant deliveries' in European cities. *Supply Chain Forum: An International Journal*, 18(4):203–217, 2017.

[21] M. Schmid-Drüner. The situation of workers in the collaborative economy. *In-depth-analysis*, 2016. European Parliament, PE 587.316.

[22] L. Kleinrock. *Queueing Systems, Volume I: Theory.* Wiley, 1975.

[23] H. Bruneel and B. G. Kim. *Discrete-Time Models for Communication Systems Including ATM.* Kluwer Academic Publishers Group, 1993.

[24] M. Gonzalez. *Classical Complex Analysis.* CRC Press, 1991.