Combined Analysis of Transient Delay Characteristics and Delay Autocorrelation Function in the $\text{Geo}^{\text{X}}/G/1$ Queue

Joris Walraevens, Dieter Fiems and Herwig Bruneel

Department of Telecommunications and Information Processing (EA07)
Ghent University - UGent
Sint-Pietersnieuwstraat 41, B-9000 Gent, Belgium.
Phone: +32-9-2648902, Fax: +32-9-2644295
E-mail: {jw,df,hb}@telin.UGent.be

Abstract

We perform a discrete-time analysis of customer delay in a buffer with batch arrivals. The delay of the k-th customer that enters the FIFO buffer is characterized under the assumption that the numbers of arrivals per slot are independent and identically distributed. By using supplementary variables and generating functions, $z-$transforms of the transient delays are calculated. Numerical inversion of these transforms lead to results for the moments of the delay of the k-th customer. For computational reasons k cannot be too large. Therefore, these numerical inversion results are complemented by explicit analytic expressions for the asymptotics for large k. We further show how the results allow us to characterize jitter-related variables, such as the autocorrelation of the delay in steady state.

Key words: Generating functions, transient analysis, delay, autocorrelation

1 Introduction

The delay (or waiting time) experienced by customers is one of the most important characteristics of a queueing system. There is a vast literature dealing with the analysis of the steady-state delay in a broad range of queueing systems. Fewer research efforts have addressed transient delay measures. The latter is the topic of this paper. More specifically, we are interested in the (expected) delay of the k-th customer that arrives in a discrete-time $Geo^X/G/1$ queue with a FIFO (First-In-First-Out) scheduling discipline.

The study of the delay of the k-th customer is intrinsically different from the study of the transient system content or the transient unfinished work (or virtual delay), as these measures are *time-dependent* (for instance the system content at time t) while the measure that we are interested in is *customer-dependent* (the delay of the k-th customer). In other words, we are not interested in the delay of a customer arriving at time t (that is strongly related to the virtual delay at time t) but of the delay of the k-th customer (irrespective of its arrival epoch).

Some literature on transient delay measures has concentrated on time-dependent analyses. Particularly, the virtual delay at time t has been the subject of [3, 8]. Abate en Whitt [3] analyze the workload process in the $M/G/1$ queue using Laplace-Stieltjes transforms. In [8], transient relationships are established between the system content at time t and the delay of a customer arriving at time t for a broad range of different (non-stationary) continuous-time queueing systems. Studies on customer-dependent delay are the focus of [6, 10, 13, 15, 19, 22, 23]. In the book of Cohen [10], the delay of the k-th arriving customer is analyzed for some continuous-time single-server queueing models like the $M/M/1$ and $G/M/1$ queues. Furthermore, in [22, 23], the transient delay of subsequently arriving customers is analyzed for the (continuous-time) $M/G/1$ queue and $GI/G/1$ queue respectively. The author further argues that his results can help to gain variance reduction in simulations. The delay of the k-th arriving customer in the $G/M/1$ queue is the focus of [19], with alternating as well as with regular service. Baron [6] analyses the delay of the k-th arriving customer in a busy period (provided that at least k customers arrive in the busy period) of a $G/M/c$ queue. Finally, some studies of discrete-time queueing systems can be reported on as well. Hofkens et al [13] calculate performance measures of the delay of the k-th arriving customer in a discrete-time $DMAP/G/1$ queue. These authors analyze the transient system content at departure epochs first and then relate the delay of an arriving customer to the system content. They also apply their results to a play-out buffer for variable bit rate video streams. Janssen and van Leeuwaarden [15] calculate relaxation times (which characterize the speed of convergence to the stationary solution) of waiting times in the discrete-time $D/G/1$ queue.

Although the above mentioned studies on customer-dependent delay measures involve many different queueing systems, they have a common feature, namely, they are all studies on *single arrival* systems. These studies are thus not directly applicable to queueing systems with *batch arrivals* which is the topic of the present contribution. In particular, we study the transient delay in a discrete-time $Geo^X/G/1$ queue. The paper is a (large) extension of our conference paper [20], where we initiated our method of analysis for a simpler model. Due to the batch nature of the arrivals, one has to keep track of the ordinal number of the customers in their arrival batch. This makes the analysis more complicated, as will be apparent in section 3. Our results include moments of the delay of the k-th customer as well as the autocorrelation function of the stationary delay. For small k, the measures corresponding with the k-th customer (expected delay, autocorrelation function, ...) are calculated by numerically inverting the generating function. These numerical results are further complemented by approximate analytic results (asymptotics), which are accurate for large enough k.

The autocorrelation function of the delay is an important function from a practical point of view, as it quantifies the correlation of the delay of a certain customer and the delays of consecutive customers. We extend our delay analysis so that the autocorrelation function of the steady-state

delay can be calculated. Our analysis thus provides transient delay measures and autocorrelation functions of delays in a unified framework. This seems to be a novel feature, as all studies of autocorrelation functions (of delays) seem to be separated from analyses of transient delays, although some authors study them in one and the same paper, see for instance [19]. Calculation of the autocorrelation function of the virtual delay in an $M/G/1$ queue is adressed in [17]. The autocorrelation function of the actual delay in the $G/M/1$ queue is analyzed in [19]. The sum of the correlation coefficients in $G/M/1$ and $G/G/1$ queues is calculated in [11, 18].

The remainder of this paper is organized as follows. The queueing model is described in more detail in the following section and the transient delay analysis is presented in section 3. The final result of the latter section is an expression for a transform of a bivariate probability generating function of the delay of the first and of the k-th customer. Some important transforms, such as the transforms of the transient moments, the transform of the autocorrelation function and the limiting steady-state probability generating function, relate to this joint transform. These corrollary results and some special cases are the subject of section 4. We then discuss the inversion of the transforms in section 5 with the purpose of obtaining transient performance measures. Several applications of the results are discussed in section 6, and some final comments are given in section 7.

2   Queueing model and notations

We consider a discrete-time single-server queueing system with infinite buffer space. We assume the system to be a Late Arrival System with Delayed Access (LAS-DA) [14]. The numbers of arrivals during consecutive slots are independent and identically distributed (i.i.d.) stochastic variables. Denote the number of arrivals in an arbitrarily chosen slot by a. We use the notations $a(n)$ and $A(z)$ to indicate its probability mass function (pmf) and probability generating function (pgf) respectively, i.e., $a(n) \triangleq \mathrm{Prob}\,[a = n]$, $n \geq 0$ and $A(z) \triangleq \mathrm{E}\,[z^a] = \sum_{n=0}^{\infty} a(n)z^n$. Furthermore, the mean and the variance of the number of arrivals in a slot is given by $\lambda = A'(1)$ and $\sigma_A^2 = A''(1) - A'(1)^2 + A'(1)$ respectively. We make no assumptions on the specific arrival instants within a slot. However, it is implicitly assumed that customers arrive in a certain order within their arrival slot. This is necessary to make "the k-th arriving customer" a valid term.

Customers are served on a First Come First Served (FCFS) basis and start service at the slot boundary following their arrival instant at the earliest (delayed access). The service time of customer k is denoted by s_k. The service times are i.i.d. and have a common pgf $S(z)$. The mean and variance of the service times are given by $1/\mu = S'(1)$ and $\sigma_S^2 = S''(1) - S'(1)^2 + S'(1)$ respectively. Finally, the system load ρ equals λ/μ.

For future use, we define

$$b_n(m) \triangleq \mathrm{Prob}\,[a \geq n+m | a \geq n] = \frac{\sum_{j=n+m}^{\infty} a(j)}{\sum_{j=n}^{\infty} a(j)},\ n, m \geq 0, \tag{1}$$

$$c_n(m) \triangleq b_n(m) - b_n(m+1),\ n, m \geq 0,$$

$$B_n(z) \triangleq \sum_{m=0}^{\infty} b_n(m)z^m,\ n \geq 0, \tag{2}$$

$$C_n(z) \triangleq \sum_{m=0}^{\infty} c_n(m)z^m = \frac{(z-1)B_n(z)+1}{z},\ n \geq 0. \tag{3}$$

Function $b_n(m)$ has the property that $b_n(m)b_{n+m}(l) = b_n(m+l)$, $n, m, l \geq 0$. As a corrollary, we find

$$\prod_{j=n}^{n+m} b_j(1) = b_n(m+1),\ n, m \geq 0. \tag{4}$$

Finally, $B_1(z)$ and $C_1(z)$ are written in terms of $A(z)$ as

$$B_1(z) = \frac{A(z) - 1}{(z-1)(1 - A(0))}, \tag{5}$$

$$C_1(z) = \frac{A(z) - A(0)}{z(1 - A(0))}. \tag{6}$$

3 Analysis

Next to the delay of subsequently arriving customers, we first define another variable such that the delay and the supplementary variable capture the state of the queueing system in the Markovian sense. We further keep track of the delay of the first customer in our analysis, which is convenient to calculate the autocorrelation function of the steady-state delay later on, see subsection 4.2.4. We subsequently compose the set of system equations, transform these equations to generating functions and finally solve the model in the z-domain.

3.1 System equations

The delay d_k of the k-th arriving customer is defined as the number of slots between the end of the customer's arrival slot and the end of his departure slot (thus excluding his arrival slot and including his departure slot). Because of the general distribution of the arrival batch sizes, $\{d_k, k \geq 1\}$ does not constitute a Markov chain. Therefore we introduce additional stochastic variables r_k ($k \geq 1$) defined as the ordinal number of the k-th arriving customer in his arrival batch, i.e., $r_k = n$ if the k-th arriving customer is the n-th customer of all customers arriving in his arrival batch. The resulting sequence $\{(r_k, d_k), k \geq 1\}$ forms a Markov chain. We allow the pair (r_1, d_1) to have a general (given) bivariate distribution. Define the following pgfs:

$$D_1(z) \triangleq \mathrm{E}\left[z^{d_1}\right], \tag{7}$$

$$D_1^{(i)}(z) \triangleq \mathrm{E}\left[z^{d_1}\mathbf{1}_{\mathbf{r_1=i}}\right], \ i \geq 1. \tag{8}$$

Here, $\mathbf{1_X}$ denotes the indicator function of the event X.

Let t_{k-1} denote the interarrival time (in number of slots) between customers $k-1$ and k. Basically, two cases can be distinguished. Firstly, customers $k-1$ and k might arrive in the same batch, which results in an interarrival time t_{k-1} equal to zero. Secondly, customers $k-1$ and k might arrive in different batches. Then customer $k-1$ is the last customer in his arrival batch, and customer k the first in the following arrival batch. As a result, the interarrival time t_{k-1} is the interarrival time between two consecutively arriving (non-empty) batches in this second case. The following equations are then easily established ($k > 1$):

1. If $t_{k-1} = 0$:

$$r_k = r_{k-1} + 1, \tag{9}$$

$$d_k = d_{k-1} + s_k. \tag{10}$$

2. If $t_{k-1} > 0$:

$$r_k = 1, \tag{11}$$

$$d_k = [d_{k-1} - t_{k-1}]^+ + s_k, \tag{12}$$

with $[x]^+$ defined as $\max(x, 0)$.

Finally, note that the random variable t_{k-1} depends on r_{k-1}. In particular, $\mathrm{Prob}[t_{k-1} = 0 | r_{k-1} = n] = b_n(1)$, $k > 1, n \geq 1$, with $b_n(m)$ defined in Eq. (1).

3.2 Translation to generating functions

We are interested in the characteristics of the sequence $\{d_k\}_{k=1}^{\infty}$. Furthermore, for the calculation of the autocorrelation function of the steady-state delay later on, the dependence of d_k ($k \geq 1$) on d_1 is of concern as well. Therefore, let $P_k(y,z)$ denote the bivariate pgf of the delay of the first and the delay of the k-th customer, i.e.,

$$P_k(y,z) \triangleq \mathrm{E}\left[y^{d_1} z^{d_k}\right], \ k \geq 1. \tag{13}$$

Ultimately, the goal of this and the following subsection is to find an expression for $P(x,y,z)$, defined as (with some abuse of notation)

$$P(x,y,z) \triangleq \sum_{k=1}^{\infty} P_k(y,z) x^k. \tag{14}$$

We first condition on the value of r_k in Eq. (13):

$$P_k(y,z) = \sum_{i=1}^{\infty} P_k^{(i)}(y,z), \ k \geq 1, \tag{15}$$

with $P_k^{(i)}(y,z) \triangleq \mathrm{E}\left[y^{d_1} z^{d_k} \mathbf{1}_{\mathbf{r_k=i}}\right]$, $k,i \geq 1$. For $k,i > 1$, the following recursive expression is found:

$$P_k^{(i)}(y,z) = S(z) b_{i-1}(1) P_{k-1}^{(i-1)}(y,z). \tag{16}$$

This expression is retrieved by noting that for $k,i > 1$, $r_k = i$ iff $r_{k-1} = i-1$ and $t_{k-1} = 0$, and by further using Eq. (10). By iteratively applying Eq. (16), we find

$$P_k^{(i)}(y,z) = S(z)^{\min(k,i)-1} b_{\max(i-k,0)+1}(\min(i,k)-1) P_{\max(k-i,0)+1}^{(\max(i-k,0)+1)}(y,z), \ k,i \geq 1, \tag{17}$$

where we have also used Eq. (4). Loosely speaking, Eq. (17) relates the (bivariate partial) pgf corresponding with the k-th customer to either the pgf corresponding with the first customer of his batch (when $k \geq i$) or to the pgf corresponding with the delay of the very first customer (when $k \leq i$). Eq. (17) permits us to find a relationship between $P(x,y,z)$, $P^{(1)}(x,y,z)$ and $\left\{P_1^{(i)}(y,z), i \geq 2\right\}$. Here, $P^{(1)}(x,y,z)$ is defined as

$$P^{(1)}(x,y,z) \triangleq \sum_{k=1}^{\infty} P_k^{(1)}(y,z) x^k. \tag{18}$$

This is achieved by substituting Eqs. (15) and (17) in Eq. (14) and by means of some tedious (yet standard) mathematical manipulations. This relationship reads

$$P(x,y,z) = B_1(xS(z)) P^{(1)}(x,y,z) + x \sum_{i=2}^{\infty} B_i(xS(z)) P_1^{(i)}(y,z), \tag{19}$$

with $B_n(z)$ as defined in Eq. (2). In the remainder of this subsection, we first establish a functional equation for $P^{(1)}(x,y,z)$ which is then transformed into a functional equation for $P(x,y,z)$ by use of Eq. (19). The latter functional equation is solved in subsection 3.3.

To establish a functional equation for $P^{(1)}(x,y,z)$, we first use Eqs. (11) and (12). Since $r_k = 1$ iff $t_{k-1} > 0$, we find

$$P_k^{(1)}(y,z) = \mathrm{E}\left[y^{d_1} z^{[d_{k-1}-t_{k-1}]^+ + s_k} \mathbf{1}_{\mathbf{r_k=1}}\right], \ k > 1,$$

by using Eq. (12). Conditioning on the values of r_{k-1}, d_{k-1} and t_{k-1}, we find

$$P_k^{(1)}(y,z) = S(z) \sum_{i=1}^{\infty} \sum_{n=1}^{\infty} \mathrm{E}\left[y^{d_1} \mathbf{1}_{\mathbf{r_{k-1}=i,d_{k-1}=n,r_k=1}}\right] \left(\sum_{m=1}^{n} \mathrm{Prob}\left[t_{k-1} = m | r_k = 1\right] z^{n-m}\right.$$

$$\left. + \sum_{m=n+1}^{\infty} \mathrm{Prob}\left[t_{k-1} = m | r_k = 1\right]\right), \quad k > 1, \tag{20}$$

where we have used that t_{k-1} is independent of the values of r_{k-1} and d_{k-1}, under the condition that $r_k = 1$. In this case, t_{k-1} is the interarrival time between two consecutively arriving (non-empty) batches. Since in each slot a non-empty batch arrives with probability $1 - A(0)$, t_{k-1} is (shifted) geometrically distributed with parameter $A(0)$ when conditioned on $r_k = 1$, i.e., $\mathrm{Prob}\left[t_{k-1} = m | r_k = 1\right] = (1 - A(0))(A(0))^{m-1}$, $k > 1, m \geq 1$. Using this expression and $\mathrm{Prob}\left[r_k = 1 | r_{k-1} = i, d_{k-1} = n\right] = 1 - b_i(1)$, $k > 1, i, n \geq 1$ in Eq. (20), we retrieve

$$P_k^{(1)}(y,z) = S(z) \sum_{i=1}^{\infty} \sum_{n=1}^{\infty} (1 - b_i(1)) \mathrm{E}\left[y^{d_1} \mathbf{1}_{\mathbf{r_{k-1}=i,d_{k-1}=n}}\right] \frac{(z-1)A(0)^n + (1 - A(0))z^n}{z - A(0)}, \quad k > 1.$$

Finally, the sum over n can be calculated which yields the following result:

$$P_k^{(1)}(y,z) = S(z) \sum_{i=1}^{\infty} (1 - b_i(1)) \frac{(z-1)P_{k-1}^{(i)}(y, A(0)) + (1 - A(0))P_{k-1}^{(i)}(y, z)}{z - A(0)}, \quad k > 1. \tag{21}$$

This expression relates the pgf corresponding to the delay of customer k given that he is the first in his arrival batch with the pgf corresponding to the delay of customer $k - 1$ conditioned on his ordinal number in his arrival batch. Substitution of Eq. (21) in Eq. (18) yields

$$P^{(1)}(x,y,z) = P_1^{(1)}(y,z)x + \frac{xS(z)[(z-1)Q(x, y, A(0)) + (1 - A(0))Q(x, y, z)]}{z - A(0)}, \tag{22}$$

with

$$Q(x,y,z) = \sum_{k=1}^{\infty} \sum_{i=1}^{\infty} (1 - b_i(1)) P_k^{(i)}(y, z) x^k. \tag{23}$$

We calculate Q next. Substitution of Eq. (17) in Eq. (23) and some further manipulations lead to:

$$Q(x,y,z) = C_1(xS(z)) P^{(1)}(x, y, z) + x \sum_{i=2}^{\infty} C_i(xS(z)) P_1^{(i)}(y, z), \tag{24}$$

with $C_n(z)$ defined in Eq. (3). By plugging Eq. (24) into Eq. (22), we obtain the desired functional equation for $P^{(1)}(x, y, z)$:

$$P^{(1)}(x,y,z) = \frac{x}{z - A(xS(z))} \left[S(z)C_1(xS(A(0)))(z-1)P^{(1)}(x, y, A(0)) + (z - A(0))P_1^{(1)}(y, z) \right.$$

$$\left. + xS(z) \left((z-1) \sum_{i=2}^{\infty} C_i(xS(A(0))) P_1^{(i)}(y, A(0)) + (1 - A(0)) \sum_{i=2}^{\infty} C_i(xS(z)) P_1^{(i)}(y, z) \right) \right],$$

where we have also used Eq. (6). This expression together with Eq. (19) finally leads to the functional equation for $P(x, y, z)$. The functional equation reads

$$
\begin{aligned}
P(x,y,z) = \frac{1}{z - A(xS(z))} &\left[\frac{B_1(xS(z))S(z)(z-1)(A(xS(A(0))) - A(0))P(x, y, A(0))}{B_1(xS(A(0)))S(A(0))(1 - A(0))} \right. \\
&+ xB_1(xS(z))(1 - A(0))D_1(yz) + \frac{xS(z)B_1(xS(z))(z-1)D_1(yA(0))}{S(A(0))} \\
&+ x(z-1) \sum_{i=1}^{\infty} B_i(xS(z))D_1^{(i)}(yz) - \left. \frac{B_1(xS(z))xS(z)(z-1)\sum_{i=1}^{\infty} B_i(xS(A(0)))D_1^{(i)}(yA(0))}{B_1(xS(A(0)))S(A(0))} \right].
\end{aligned}
$$
(25)

Here we used that $P_1(y, z) = D_1(yz)$ and $P_1^{(i)}(y, z) = D_1^{(i)}(yz)$ (see Eqs. (7) and (8)). The right-hand side (RHS) of Eq. (25) is fully determined except for the unknown boundary term $P(x, y, A(0))$. A method to determine this boundary function is discussed in the next subsection.

3.3 Solution of the functional equation

In order to determine the boundary function $P(x, y, A(0))$, we return to Eqs. (14) and (25). From the definition in Eq. (14), it follows that $P(x, y, z)$ is an analytic function of y and z with $\{(y, z) : |y| < 1 \wedge |z| < 1\}$, for all values of x in $\{x : |x| < 1\}$. It can be shown by Rouché's theorem that the denominator of Eq. (25) has exactly one zero inside the unit disk of the z-plane for all $|x| < 1$, namely, the factor $z - A(xS(z))$ has one zero inside the unit disk of the z-plane for all $|x| < 1$ while the other factors of the denominator have no zeroes inside that region. Let $V(x)$ denote this zero. Then $V(x)$ must be a zero of the second factor of the RHS of Eq. (25) in order to guarantee the analiticity of $P(x, y, z)$ for all y and z in $\{(y, z) : |y| < 1 \wedge |z| < 1\}$. This condition leads to the following expression for $P(x, y, A(0))$:

$$
\begin{aligned}
P(x, y, A(0)) = \frac{x(1 - A(0))}{A(0) - A(xS(A(0)))} &\left[B_1(xS(A(0)))D_1(yA(0)) - \sum_{i=1}^{\infty} B_i(xS(A(0)))D_1^{(i)}(yA(0)) \right. \\
&+ \frac{(1 - A(0))S(A(0))B_1(xS(A(0)))D_1(yV(x))}{S(V(x))(V(x) - 1)} \\
&+ \left. \frac{S(A(0))B_1(xS(A(0)))\sum_{i=1}^{\infty} B_i(xS(V(x)))D_1^{(i)}(yV(x))}{S(V(x))B_1(xS(V(x)))} \right].
\end{aligned}
$$
(26)

Finally, substitution of this expression in Eq. (25) leads to following theorem.

Theorem 1. *The trivariate pgf $P(x, y, z) = \sum_{k=1}^{\infty} E\left[y^{d_1} z^{d_k}\right] x^k$ is given by*

$$
\begin{aligned}
P(x,y,z) = \frac{x}{z - A(xS(z))} &\left[(1 - A(0))B_1(xS(z)) \left(D_1(yz) - \frac{S(z)(z-1)D_1(yV(x))}{S(V(x))(V(x) - 1)} \right) \right. \\
&- \frac{S(z)(z-1)B_1(xS(z))\sum_{i=1}^{\infty} B_i(xS(V(x)))D_1^{(i)}(yV(x))}{S(V(x))B_1(xS(V(x)))} + \left. (z-1) \sum_{i=1}^{\infty} B_i(xS(z))D_1^{(i)}(yz) \right],
\end{aligned}
$$
(27)

with $V(x)$ the solution of the equation

$$
z - A(xS(z)) = 0, \; |x| < 1,
$$
(28)

in the complex unit circle in the z-plane.

The analytic determination of $V(x)$ is, in general, not possible. However, $V(x)$ can be calculated numerically/iteratively for each specific value of x. We will show how this is done in section 5 and that it is sufficient to numerically calculate the desired performance measures.

7

4 Special cases and corollary transforms

From the general result (27), we can derive numerous interesting results. In a first subsection, we will look into some special cases for the distributions of the number of arrivals in a slot, of the service times and/or of the delay and ordinal number of the first customer. In a second subsection, we calculate the transforms of some performance measures, such as the (mean) delay of the k-th customer and the autocorrelation function of the steady-state delay.

4.1 Special cases

We here treat some special cases for which expression (27) simplifies considerably.

First of all, we look at the natural case that $r_1 = 1$ with probability (w.p.) 1.

Corollary 1. *In the case $D_1(z) = D_1^{(1)}(z)$, the trivariate pgf $P(x, y, z)$ is given by*

$$P(x, y, z) = \frac{xB_1(xS(z))\left[(z - A(0))D_1(yz) + \dfrac{S(z)(z - 1)(V(x) - A(0))D_1(yV(x))}{S(V(x))(1 - V(x))}\right]}{z - A(xS(z))}. \quad (29)$$

Secondly, we look at the case of single arrivals. Note that it directly follows that $r_1 = 1$ w.p. 1, so we can start from Eq. (29).

Corollary 2. *In the case $A(z) = A(0) + (1 - A(0))z$, the trivariate pgf $P(x, y, z)$ is given by*

$$P(x, y, z) = \frac{x\left[(z - A(0))D_1(yz) + \dfrac{xS(z)(z - 1)(1 - A(0))D_1(yV(x))}{1 - V(x)}\right]}{z - A(0) - (1 - A(0))xS(z)}, \quad (30)$$

with $V(x)$ the solution of the equation

$$z - A(0) - (1 - A(0))xS(z) = 0, \ |x| < 1, \quad (31)$$

in the complex unit circle in the z-plane.

A last special case we would like to treat is $(r_1, d_1) = (1, 1)$ w.p. 1 and deterministic service times of one slot.

Corollary 3. *In the case $D_1(z) = D_1^{(1)}(z) = z$ and $S(z) = z$, the trivariate pgf $P(x, y, z)$ is given by*

$$P(x, y, z) = \frac{xyz(1 - A(xz))(z - V(x))}{(1 - xz)(1 - V(x))(z - A(xz))}, \quad (32)$$

with $V(x)$ the solution of the equation

$$z - A(xz) = 0, \ |x| < 1, \quad (33)$$

in the complex unit circle in the z-plane. In the special case of geometrically distributed arrival batch sizes, i.e., $A(z) = 1/(1 + \lambda - \lambda z)$, $V(x)$ can be calculated explicitly and is given by

$$V(x) = \frac{1 + \lambda - \left[(1 + \lambda)^2 - 4\lambda x\right]^{1/2}}{2\lambda x}. \quad (34)$$

4.2 Corollary transforms

We here show how to derive some important transforms from the expression(s) we obtained for $P(x, y, z)$. It leads us to expressions of (i) transforms of time-dependent delay characteristics, of (ii) the steady-state delay distribution, and of (iii) the transform of the autocorrelation function of the steady-state delay.

In most practical situations, one is interested in the natural case that the first customer is also the first customer in his batch, i.e., $r_1 = 1$ w.p. 1. In the first part of this subsection, it is assumed this is the case. We calculate the transform of the pgfs of the delay, given by

$$D(x, z) \triangleq \sum_{k=1}^{\infty} \mathrm{E}\left[z^{d_k}\right] x^k, \tag{35}$$

and the transform of the mean delays

$$\bar{D}(x) \triangleq \sum_{k=1}^{\infty} \bar{d}_k x^k. \tag{36}$$

Here, $\bar{d}_k$ denotes the mean delay of the k-th customer. The latter expression is important as it can be numerically or analytically inverted to obtain the (tail of the) sequence of the mean delays $\{\bar{d}_k, k \geq 1\}$ (we will do so in sections 5 and 6). Through the final value theorem of generating functions, we also calculate the pgf of the steady-state delay from Eq. (29). Finally, we will go back to the more general setting for r_1 in order to calculate the transform of the autocorrelation function.

4.2.1 Transform of the pgfs of the delay

By substituting $y = 1$ in Eq. (29), we find the generating function $D(x, z)$ of the sequence $\{\mathrm{E}\left[z^{d_k}\right], k \geq 1\}$.

Corollary 4. *The generating function* $D(x, z) \triangleq \sum_{k=1}^{\infty} E\left[z^{d_k}\right] x^k$ *in case* $r_1 = 1$ *w.p. 1 is given by*

$$D(x, z) = \frac{x B_1(x S(z)) \left[(z - A(0)) D_1(z) + \dfrac{S(z)(z-1)(V(x) - A(0)) D_1(V(x))}{S(V(x))(1 - V(x))} \right]}{z - A(x S(z))}. \tag{37}$$

4.2.2 Transform of the mean delays

We calculate the transform function $\bar{D}(x)$ of the sequence $\{\bar{d}_k, k \geq 1\}$ in this subsection. We have $\bar{d}_k = \partial P_k(y, z)/\partial z|_{y=z=1}$ and therefore

$$\bar{D}(x) = \left.\frac{\partial D(x, z)}{\partial z}\right|_{z=1}. \tag{38}$$

By substituting Eq. (37) in this expression, taking the partial derivative to z and substituting $z = 1$, we retrieve $\bar{D}(x)$ as a function of the pgfs A and S of the arrival batch sizes and service times respectively, of the pgf D_1 and the mean $\bar{d}_1$ of the delay of the first customer, and of the implicitly defined function V.

Corollary 5. *The transform function* $\bar{D}(x) \triangleq \sum_{k=1}^{\infty} \bar{d}_k x^k$ *in case* $r_1 = 1$ *w.p. 1 is given by*

$$\bar{D}(x) = \frac{x \bar{d}_1}{1 - x} + \frac{x^2}{\mu(1-x)^2} + \frac{x(A(x) - A(0))}{(1-x)(1 - A(0))(A(x) - 1)} + \frac{x(V(x) - A(0)) D_1(V(x))}{(1 - A(0))(1 - x) S(V(x))(1 - V(x))}, \tag{39}$$

Calculation of higher derivatives to z of Eq. (37) and substitution of $z = 1$ provide transform functions of higher order moments of the delays as well.

4.2.3 Steady-state delay

In this subsection, we show how to obtain the steady-state delay in a stable system, since it gives some insight in the calculation of steady-state results from transient results and in the behaviour of the function V. The pgf $D_\infty(z)$ of the delay in steady state can be found from $D(x, z)$ by the final value theorem $D_\infty(z) = \lim_{x \to 1^-}(1-x)D(x, z)$. We note that $D_\infty(z)$ not necessary exists (for instance if the delay experiences some periodicity). In this case, $\lim_{x \to 1^-}(1-x)D(x, z)$ equals $< D_k(z) >= \lim_{T \to \infty} \frac{1}{T} \sum_{k=1}^{T} D_k(z)$, see [12]. However, since for the more interesting cases $D_\infty(z)$ does exist, we use this notation in the remainder of this subsection.

The limit $\lim_{x \to 1^-}(1-x)D(x, z)$ is calculated by first substituting Eq. (37) in this expression. Doing this we obtain

$$D_\infty(z) = S(z)(z-1) \lim_{x \to 1^-} \frac{x(1-x)B_1(xS(z))(A(0) - V(x))D_1(V(x))}{S(V(x))(V(x)-1)(z - A(xS(z)))}. \tag{40}$$

It turns out that $\lim_{x \to 1^-} V(x)$ has to be calculated. It can be shown through the implicit function theorem that $V(1)$ is a solution of $z - A(S(z)) = 0$ in the closed unit disk. This latter equation has exactly one solution in the closed unit disk for stable systems (see Rouché's theorem or see [5]). Since $z = 1$ is an obvious solution of this equality it turns out that $V(1) = 1$. Eq. (40) then leads to

$$D_\infty(z) = \frac{S(z)(z-1)(A(S(z)) - 1)}{V'(1)(z - A(S(z)))(S(z) - 1)}. \tag{41}$$

Here, we used Eq. (5) and l'Hôpital's rule. Taking the first derivative of both sides of the functional equation defining $V(x)$, solving to $V'(x)$ and putting $x = 1$ further yields $V'(1) = \lambda/(1-\rho)$. Substituting this expression in Eq. (41), we find the result for the pgf of the steady-state delay in a $Geo^X/G/1$ queue (see e.g. [9]):

$$D_\infty(z) = \frac{1-\rho}{\lambda} \frac{S(z)(z-1)}{z - A(S(z))} \frac{A(S(z)) - 1}{S(z) - 1}. \tag{42}$$

A similar procedure can be used to obtain the mean delay in steady state from Eq. (39), leading to

$$\bar{d}_\infty = \frac{1}{2\mu} + \frac{\sigma_A^2}{2\lambda\mu(1-\rho)} + \frac{\lambda\sigma_S^2}{2(1-\rho)}. \tag{43}$$

Higher moments of the steady-state delay, such as its variance, can also be calculated.

We may finally note that in the case of unstable systems, $z = 1$ is not the only solution inside the unit disk. Therefore, $\lim_{x \to 1} V(x)$ is then not equal to 1 but is equal to the other solution inside the unit disk of $z - A(S(z)) = 0$. In that case, Eq. (40) tends to 0, no longer leading to a proper probability generating function. Note that $V^*(1) = 1$ is then the analytic continuation of a solution $V^*(x)$ ($|x| < 1$) of $z - A(xS(z)) = 0$ *outside* the unit disk (see also [21]).

4.2.4 Transform of the autocorrelation function of the delay

We finally calculate the transform of the autocorrelation function of the steady-state delay. Therefore, denote an arbitrarily chosen customer in steady state by customer J, his delay by $d_{J,1}$ and the delay of the k-th customer arriving after customer J by $d_{J,k+1}$ ($k \geq 1$). We further denote the ordinal number of customer J by $r_{J,1}$. The autocorrelation function is then defined as

$$R_k \triangleq \frac{\mathrm{E}\left[\left(d_{J,1} - \bar{d}_\infty\right)\left(d_{J,k+1} - \bar{d}_\infty\right)\right]}{\sigma_{d_\infty}^2}, \tag{44}$$

with $\bar{d}_\infty$ and $\sigma_{d_\infty}^2$ the mean and variance of the steady-state delay. The transform of the sequence $\{R_k, k \geq 1\}$ is given by $R(x) \triangleq \sum_{k=0}^{\infty} R_k x^k$. In the remainder of this subsection, we intend to find an expression for $R(x)$. Therefore, we consider the analysis in section 3 with the starting stochastic vector (d_1, r_1) equal to $(d_{J,1}, r_{J,1})$. From Eqs. (13), (14) and (44), it is seen that

$$R(x) = \frac{1}{\sigma_{d_\infty}^2} \left(\frac{1}{x} \left. \frac{\partial^2 P(x,y,z)}{\partial y \partial z} \right|_{y=z=1} - \frac{\bar{d}_\infty^2}{1-x} \right). \tag{45}$$

In order to use the results from section 3 (and more specifically Eq. (27)), we need expressions for the pgfs related to the starting stochastic vector $(d_{J,1}, r_{J,1})$, namely, $D_{J,1}(z) \triangleq \mathrm{E}\left[z^{d_{J,1}}\right]$ and $D_{J,1}^{(i)}(z) \triangleq \mathrm{E}\left[z^{d_{J,1}} \mathbf{1}_{\mathbf{r}_{J,1}=\mathbf{i}}\right]$, $i \geq 1$, i.e., for the pgf of the steady-state delay and for the partial pgf of the steady-state delay given that the customer is the i-th customer in his arrival batch. Since customer J is a randomly chosen customer in steady state,

$$D_{J,1}(z) = D_\infty(z). \tag{46}$$

To calculate $D_{J,1}^{(i)}(z)$, we first use Eq. (16) $i-1$ times (with $k = \infty$ and $y = 1$), leading to

$$D_{J,1}^{(i)}(z) = D_\infty^{(i)}(z) = S(z)^{i-1} b_1(i-1) D_\infty^{(1)}(z), \; i > 1, \tag{47}$$

with $D_\infty^{(i)}(z)$ the partial pgf of the steady-state delay given that the corresponding customer is the i-th customer of his arrival batch. We further have $D_\infty(z) = \sum_{i=1}^{\infty} D_\infty^{(i)}(z)$. Substitution of Eq. (47) in this expression and calculation of the sum leads to $D_\infty^{(1)}(z) = D_\infty(z)/B_1(S(z))$. Substitution of the latter expression in Eq. (47) yields

$$D_{J,1}^{(i)}(z) = \frac{S(z)^{i-1} b_1(i-1) D_\infty(z)}{B_1(S(z))}. \tag{48}$$

Final expressions of $D_{J,1}(z)$ and $D_{J,1}^{(i)}(z)$ are found by substitution of Eq. (42) in Eqs. (46) and (48). Finally, we replace $D_1(z)$ and $D_1^{(i)}(z)$ in Eq. (27) with these calculated expressions for $D_{J,1}(z)$ and $D_{J,1}^{(i)}(z)$ respectively, take the necessary derivatives to calculate $\partial^2 P(x,y,z)/\partial y \partial z$, replace y and z by 1 and plug this expression in Eq. (45). After some (extensive) standard mathematical manipulations, this leads to an expression for R.

Corollary 6. *The transform function $R(x)$ of the autocorrelation function of the steady-state delay $\{R_k, k \geq 1\}$ is given by*

$$R(x) = \frac{1}{1-x} - \frac{x}{\sigma_{d_\infty}^2 (1-x)^2} \left[\frac{S''(1)}{2} - \frac{x + \rho(1-x)}{\lambda \mu (1-x)} + \frac{1}{\mu(1-A(x))} + \frac{(1-\rho)xV(x)S'(V(x))}{\lambda(1-x)S(V(x))} \right.$$
$$\left. + \frac{(1-\rho)V(x)(1 - A(S(V(x))))}{\lambda(1-V(x))(V(x) - A(S(V(x))))} \right]. \tag{49}$$

5 Inversion of the transforms

In section 4, transforms for various customer-dependent performance measures are found. In the remainder, we focus on the inversion of these transforms. This task is, in general, not straight-forward. We propose the application of two techniques in this section. Firstly, we show how the transforms can be inverted numerically. Secondly, we apply an analytic technique to approximate the transient sequence. Both techniques are complementary, as we can only obtain results with sufficient precision for the first (say) N values of the sequence (N is usually chosen in the range of 500-2000) with the numerical technique, while the analytic technique generally provides good approximations for the tail of the sequence: it provides asymptotics for large k.

5.1 Numerical inversion

The main problem in the inversion of the transforms is the appearance of the implicitly defined function $V(x)$ in these transform expressions (see e.g. Eqs. (39) and (49)). We here make use of a numerical technique proposed in [2]. In this paper, it is shown that the functional equations that traditionally occur in transforms of transient performance measures can be iteratively calculated for complex arguments. Furthermore, it is proved that these iterations converge to the correct value if a stochastic interpretation can be given to the implicitly defined function. The transform can then be numerically inverted, as e.g. described in [1].

So, we first numerically calculate the function $V(x)$, characterized by $V(x) = A(xS(V(x)))$ by using the procedure described in [2]. We therefore iteratively calculate $V_{i+1}(x)$ as $V_{i+1}(x) = A(xS(V_i(x)))$. If we can prove that $V(x)$ is the (possibly defective) pgf of a random variable, it follows from [2] that $\lim_{i\to\infty} V_i(x) = V(x)$, regardless of the startingpoint $V_0(x)$ in the complex unit disk. $V(x)$ is indeed a pgf, namely it is the pgf of a random variable V defined as

$$V = \sum_{i=1}^{A} \left(1 + \sum_{j=1}^{S_i} V_{i,j} \right),$$

with A the number of arrivals in a randomly chosen slot, say slot I, S_i the service time of the i-th customer arriving in slot I and $V_{i,j}$ i.i.d. random variables with the same distribution as V.

Finally, the transforms have to be inverted numerically. Phrased in terms of the mean transient delay, we want to calculate a sequence of real numbers $\{\bar{d}_k, k \geq 1\}$ for all k upto a certain number using the generating function (39), where x is a complex number. Most mathematical software packages provide procedures for numerical inversion of transform functions, for instance FFTs. However, the accuracy of FFTs suffers from the windowing effect. An alternative procedure with known error bounds is suggested in [1]. Here, the calculated error bounds are only practical (i.e. small enough) for stable systems, i.e., where the mean delay of the k-th customer goes to a finite steady-state value for $k \to \infty$. Summarizing, the numerical inversion techniques yield suitable results for stable systems only. The analytical technique discussed next, on the other hand, can be used for unstable systems as well.

5.2 Analytical inversion

In [21], we showed how singularity analysis of a transform of a transient sequence leads to an approximation for (the tail of) the sequence. Given the expression of a transform $F(x)$, the corresponding sequence $\{f_k, k \geq 1\}$ is approximated according to the following procedure:

Procedure.

(i) Calculate $f_\infty = \lim_{x\to 1}(1-x)F(x)$.

(ii) If $0 < f_\infty < \infty$, $H(x) = F(x) - f_\infty/(1-x)$; otherwise $H(x) = F(x)$.

(iii) Determine the radius of convergence R of $H(x)$,

(iv) determine the singularities α_n of $H(x)$ on its circle of convergence,

(v) determine the behavior of $H(x)$ in the neighbourhood of these α_n, according to the following form $H(x) \sim \left(1 - \frac{x}{\alpha_n} \right)^{-\omega_n} G_n(x)$, where $\omega_n \neq 0, -1, -2, \ldots$ and G_n denotes a nonzero analytic function near α_n, and

(vi) calculate $h_k \sim \sum_n \frac{G_n(\alpha_n)}{\Gamma(\omega_n)} k^{\omega_n-1} \alpha_n^{-k}$.

(vii) If $0 < f_\infty < \infty$, calculate $f_k = h_k + f_\infty$ for all $k \geq 0$; otherwise $f_k = h_k$.

Application of the procedure on Eq. (39) yields an approximation of the mean transient delay. Three cases can be distinguished, namely $\rho < 1$, $\rho = 1$ and $\rho > 1$. When $\rho < 1$, f_∞ equals the mean steady-state delay $\bar{d}_\infty$ (Eq. (43)) and $H(x) = \bar{D}(x) - \bar{d}_\infty/(1-x)$ (step (ii) in the procedure). In the determination of the singularities of the function $H(x)$ in step (iv) of the procedure, the implicitly defined function V plays an important role. Given that $V(x)$ fulfills some mild conditions (see [7]), it can be proved that $V'(x) \to \infty$ for x going to the radius of convergence R_V of V, while $V(R_V)$ is finite. Then, $V(x)$ has a square-root type behavior in the neighbourhood of R_V, i.e., $V(x) \sim V(R_V) - K(1 - x/R_V)^{1/2}$. The dominant singularity (singularities) of $H(x)$ is (are) either this square-root branchpoint R_V of V or singularities of D_1. In the case $\rho > 1$, $\bar{d}_\infty = \infty$ and the radius of convergence equals 1. In the case $\rho = 1$, $R = 1$ as well, except for some pathological cases.

Secondly, by applying the procedure on Eq. (49), we find approximations for the autocorrelation function of the steady-state delay in stable systems. In this case, $f_\infty = 0$ and the dominant singularity of $H(x)$ is the branchpoint R_V of V.

We will give explicit expressions for the approximations of the mean delay and the autocorrelation function in a few examples in the next section. When the sequence converges, we are also interested in the time to convergence (relaxation time), defined as

$$\gamma_f \triangleq \sup\{k : (|f_\infty - f_k| > \epsilon)\} + 1, \tag{50}$$

for some (small enough) ϵ. In words, γ_f is the first element of the sequence $\{f_k, k \geq 1\}$ with the property that all elements of the sequence $\{f_k, k \geq \gamma_f\}$ lie in a neighbourhood ϵ of f_∞. Replacement of f_k with the approximated value provided by the analytic inversion technique yields an approximation of the time to convergence, which is generally accurate since we are interested in small values for ϵ.

6 Some applications

In this section, we show examples of transient delay characteristics for a couple of arrival processes and demonstrate some applications of the analysis. Single-slot service times are assumed throughout this section, i.e., $S(z) = z$.

6.1 Zero initial conditions

We assume throughout this subsection that the buffer is empty at the beginning, resulting in a delay of 1 slot for the first customer (due to the LAS-DA assumption in the model). We invert Eq. (39) with $D_1(z) = S_1(z) = z$ and $\bar{d}_1 = \mu = 1$ to find results for the mean delays for three distinct arrival batch sizes.

In a first example, we assume that batch sizes are geometrically distributed stochastic variables, i.e., $A(z) = 1/(1 + \lambda - \lambda z)$. Then $V(x)$ can be calculated explicitly and is given by Eq. (34). Therefore the transform $\bar{D}(x)$ - Eq. (39) - can be explicitly calculated as well. We have first inverted Eq. (39) numerically and obtain the mean delays of the first N arriving customers, for some different values of the mean batch size λ. In Figure 1a., we have shown these mean delays of the first 512 arriving customers for $\lambda = 0.3$, 0.6 and 0.9 respectively. The respective mean delays in the steady state for these λ's are 1.43, 2.5 and 10 respectively. It can be seen from this figure that the mean transient delays tend to these steady-state values rather quickly for small λ while the convergence is much slower for larger λ. This is even more obvious from Figure 1b., where we show the mean delays of the first 1024 customers for λ equal to 0.92, 0.94 and 0.96 respectively. We have also shown the steady-state values for the three cases. The higher λ the slower the convergence to the steady-state values. The latter conclusion can also be drawn from the analytical result
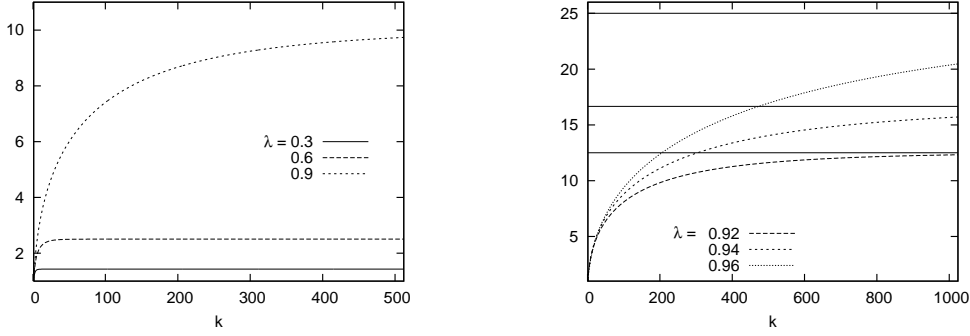
a. mean arrival rate of 0.3, 0.6 and 0.9   b. mean arrival rate of 0.92, 0.94 and 0.96

Figure 1: Mean transient delays for customers arriving in geometrically distributed batch sizes. The system is empty at the beginning.
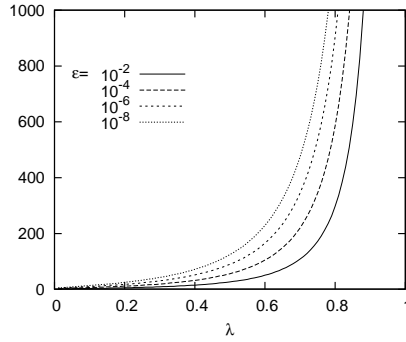

Figure 2: Time to convergence of the mean delay versus the arrival rate λ for geometrically distributed batch sizes and for different values of ϵ. The system is empty at the beginning.

obtained with the analytical approach of subsection 5.2. The expected delay of the k-th customer is approximated by

$$
\bar{d}_k \sim \begin{cases} \dfrac{1}{1-\lambda} - \dfrac{(1+\lambda)^3}{\sqrt{\pi}(1-\lambda)^4 k^{3/2}} \left(\dfrac{(1+\lambda)^2}{4\lambda} \right)^{-k} & \text{if } \lambda < 1 \\[2ex] \dfrac{2k^{1/2}}{\sqrt{\pi}} & \text{if } \lambda = 1 \\[2ex] \dfrac{\lambda - 1}{\lambda} k & \text{if } \lambda > 1 \end{cases} \quad \cdot
$$

This formula thus gives the asymptotics for large k, not only for the stable system, but also in overload. To characterize the convergence rate in stable systems, we depict in Figure 2 the (approximated) time to convergence $\gamma_{\bar{d}}$, as defined in Eq. (50), as a function of the arrival rate. We assume different values for ϵ. Figure 2 confirms that the time to convergence grows rapidly with λ.

In a second example, we assume that the batch sizes have a Poisson distribution, i.e., $A(z) = e^{\lambda(z-1)}$. In this case, we used the iterative procedure, discussed in subsection 5.1, to calculate $V(x)$ in the required points in the complex x-plane. Next, we numerically inverted Eq. (39) and obtain the mean delays of the first N arriving customers, for some values of the mean batch sizes λ. In Figure 3, we have shown the mean delay of the first 512 arriving customers for $\lambda = 0.3$, 0.6 and 0.9 respectively. The respective mean delays in steady state for these λ's are 1.21, 1.75 and 5.5 respectively. We experienced that a reasonable precision can only be obtained with enough
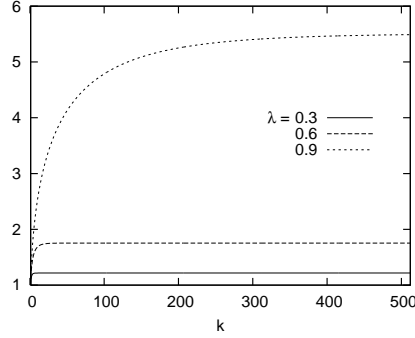
14

Figure 3: Mean transient delays for customers arriving in Poisson distributed batch sizes with means 0.3, 0.6 and 0.9 respectively. The system is empty at the beginning

iterations for the calculation of $V(x)$, especially for high λ (more iterations were necessary to attain reasonable results for $\lambda = 0.9$). The analytic asymptotics in this case are

$$\bar{d}_k \sim \begin{cases} \dfrac{2-\lambda}{2(1-\lambda)} - \dfrac{e^{1-\lambda}}{(1-\lambda e^{1-\lambda})(1-e^{1-\lambda})^2 \, k^{3/2}\sqrt{2\pi}} \left(\dfrac{1}{\lambda e^{1-\lambda}}\right)^{-k} & \text{if } \lambda < 1 \\[2ex] \dfrac{\sqrt{2}k^{1/2}}{\sqrt{\pi}} & \text{if } \lambda = 1 \\[2ex] \dfrac{\lambda-1}{\lambda}k & \text{if } \lambda > 1 \end{cases} \quad .$$

In a last example in this subsection, we assume batch size distributions with finite support. In particular, the batch sizes are binomially distributed with parameters λ/m and m, i.e. $A(z) = (1 - \lambda/m + \lambda z/m)^m$. We chose $\lambda = 0.9$ (fixed) and vary m. Note that the mean size of a batch remains constant whereas the variance changes (higher m leads to higher variance). In table 1, we show the mean delay of the k-th customer for some k and some m, obtained by the numerical inversion procedure and the iterative procedure for the calculation of $V(x)$. The respective steady-state values are also given. We observe that convergence is slower when m is larger, i.e., when the variance of the number of arrivals per slot is higher. This can also be concluded from table 2 that displays the times to convergence $\gamma_{\bar{d}}$ for $\lambda = 0.9$ and several values for the parameters ϵ and m. Here, we used the analytic approximation for the calculation of the tail of the transient delay sequence.

6.2 Transition in traffic patterns

In this subsection, we study the transition of the mean delay when the arrival process changes. More precisely, we assume geometric arrivals and assume the arrival rate is switched from λ_1 to λ_2 ($\lambda_1, \lambda_2 < 1$). We start count at the transition point, i.e., we assume that the first customer arrives in steady state of the queue with arrival rate λ_1 and immediately after the arrival of this customer the arrival rate is switched to λ_2 (and remains λ_2 from then on). In other words, we perform a quasi-stationary analysis by assuming that the periods that traffic patterns are unchanged are (significantly) longer than the time to reach steady state. In Figure 4a., we depict the mean transient delays of the first 512 customer arriving in a system when the arrival rate is switched from 0.3 to 0.9 and from 0.9 to 0.3 respectively. The results are obtained by means of numerical inversion of the transform of the mean delay with the first customer being a customer arriving in steady state in the original system. This figure captures the behavior of the delays of customers when a transition occurs in the system from busy times to less busy times (or the other way around).

15

	$m = 2$	$m = 4$	$m = 8$	$m = 16$
$k = 50$	2.8357	3.5207	3.8284	3.9753
$k = 100$	3.0944	3.9790	4.3874	4.5843
$k = 150$	3.1825	4.1675	4.6316	4.8570
$k = 200$	3.2185	4.2594	4.7572	5.0005
$k = 250$	3.2347	4.3080	4.8272	5.0823
$k = 300$	3.2423	4.3352	4.8684	5.1313
$k = 350$	3.2461	4.3509	4.8934	5.1617
$k = 400$	3.2479	4.3601	4.9089	5.1810
$k = 450$	3.2489	4.3657	4.9187	5.1935
$k = 500$	3.2494	4.3691	4.9250	5.2017
$k = \infty$	3.2500	4.3750	4.9375	5.2188

Table 1: Mean transient delays for customers arriving in binomially distributed batch sizes. The system is empty at the beginning.

	$m = 2$	$m = 4$	$m = 8$	$m = 16$
$\epsilon = 10^{-2}$	326	505	598	645
$\epsilon = 10^{-4}$	675	1014	1187	1274
$\epsilon = 10^{-6}$	1066	1577	1838	1969
$\epsilon = 10^{-8}$	1480	2174	2526	2703

Table 2: Time to convergence of the mean delay for customers arriving in binomially distributed batch sizes. The system is empty at the beginning.

We observe that the mean delay converges more quickly to the steady-state value when the arrival rate is decreased than when it is increased.
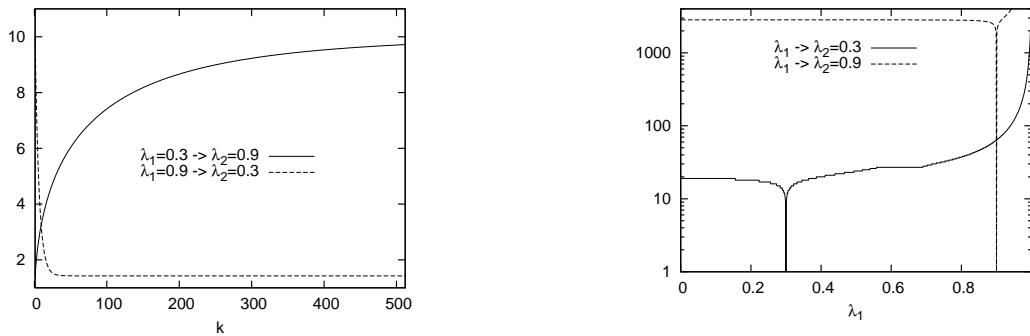
a. mean transient delays when the mean arrival rate is switched from 0.3 to 0.9 or vice-versa

b. time to convergence when the arrival rate is switched from λ_1 to 0.3 or 0.9. $\epsilon = 10^{-4}$

Figure 4: The mean of the batch sizes is switched from λ_1 to λ_2 at the beginning. Customers arrive in geometrically distributed batch sizes.
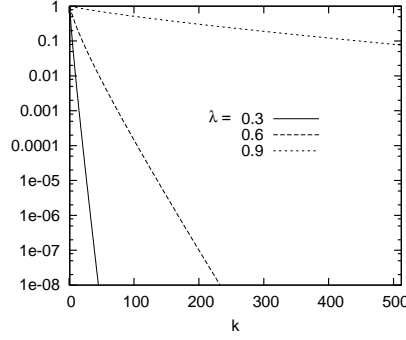
Figure 5: Autocorrelation function of the steady-state delay in a system with geometrically distributed batch sizes with means 0.3, 0.6 and 0.9 respectively

The analytic asymptotics of the mean delay are given by

$$
\bar{d}_k \sim
\begin{cases}
\dfrac{1}{1-\lambda_2} + \dfrac{\lambda_1(2\lambda_1 - 1 - \lambda_2)(1-\lambda_1+\lambda_2)}{\lambda_2(1-\lambda_1)(\lambda_1-\lambda_2)} \left(\dfrac{\lambda_1(1-\lambda_1+\lambda_2)}{\lambda_2}\right)^{-k} & \text{if } \lambda_2 < (2\lambda_1 - 1)^+ \\[3mm]
\dfrac{1}{2(1-\lambda_1)} + \dfrac{\lambda_1^3}{(1-\lambda_1)^2(2\lambda_1-1)k^{1/2}\sqrt{\pi}} \left(\dfrac{\lambda_1^2}{2\lambda_1-1}\right)^{-k} & \text{if } \lambda_2 = (2\lambda_1 - 1)^+ \\[3mm]
\dfrac{1}{1-\lambda_2} + \dfrac{(1+\lambda_2)^5(\lambda_1-\lambda_2)(1-\lambda_1)}{\lambda_2(1-\lambda_2)^4(\lambda_2-2\lambda_1+1)^2 k^{3/2}\sqrt{\pi}} \left(\dfrac{(1+\lambda_2)^2}{4\lambda_2}\right)^{-k} & \text{if } (2\lambda_1 - 1)^+ < \lambda_2 < 1
\end{cases}
.
$$

These formulas show that the way the system adapts to the changed arrival rate can be very distinct, depending on both arrival rates. Apparently, whether λ_2 is smaller or larger than $2\lambda_1 - 1$ plays an important role in the type of the asymptotic. When λ_2 is larger than $2\lambda_1 - 1$, the actual decay is independent of the starting value of the arrival rate (λ_1), while the decay depends on both λ_1 and λ_2 when $\lambda_2 < 2\lambda_1 - 1$. An intuitive reason for this distinction is not clear, but could be related to a similar observation for steady-state distributions in priority queues and tandem queues (see [4,16]). Finally, Figure 4b. depicts the time to convergence $\gamma_{\bar{d}}$ when the arrival rate is switched from λ_1 to $\lambda_2 = 0.3$ or to $\lambda_2 = 0.9$ respectively, as a function of λ_1. When $\lambda_2 = \lambda_1$ (the arrival rate is not changed), $\gamma_{\bar{d}} = 1$. When λ_1 is decreased from the value of λ_2 on, the time to convergence quickly tends to a constant value (around 2830 for $\lambda_2 = 0.9$). When λ_1 is increased, the time to convergence increases to ∞ for $\lambda_1 \to 1$. This figure further confirms the observation of Figure 4a. that the mean delay converges quicker to the steady-state value when $\lambda_2 = 0.3$ than when $\lambda_2 = 0.9$. For instance, the time to convergence for $\lambda_1 = 0.3$ and $\lambda_2 = 0.9$ equals 2830, while it is only 64 for $\lambda_1 = 0.9$ and $\lambda_2 = 0.3$. It also shows that the time to convergence is largely invariant of λ_1, except for high values of this parameter or for $\lambda_1 \approx \lambda_2$.

6.3 Autocorrelation function of the steady-state delay

In this final subsection, we invert the transform $R(x)$ (Eq. (49)) of the autocorrelation function of the steady-state delay. We consider geometric batch sizes with mean λ ($\lambda < 1$). Figure 5 depicts the autocorrelation function for the first 512 lags, after numerical inversion of Eq. (49). The autocorrelation function starts in 1 and tends to 0 for $k \to \infty$. It is seen that it decreases more slowly for higher λ, meaning that the correlation between consecutive delays fades out much more slowly when the system is heavily loaded. This is also concluded from the asymptotics of the autocorrelation function, which are, in this case, given by

$$
R_k \sim \frac{4(1+\lambda)^3}{(1-\lambda)^3 k^{3/2}\sqrt{\pi}} \left(\frac{(1+\lambda)^2}{4\lambda}\right)^{-k} .
$$

17

7 Conclusions

In this paper, we have analyzed the transient delay in a discrete-time $Geo^X/G/1$ queue by the supplementary variable technique and by further making extensive use of generating functions. A functional equation for the transform of the joint probability generating function of the delay of the k-th arriving customer, the ordinal number of this customer within his arrival batch and the delay of the first arriving customer was obtained. By Rouché's theorem, this functional equation is solved in terms of an implicitly defined function. Important transforms can be computed, most notably the transforms of the mean transient delays and of the autocorrelation function of the steady-state delay. Inversion of these transforms is not straightforward because of the implicitly defined function that was introduced in the analysis. Special attention is therefore given to the inversion of these transforms, numerically as well as by obtaining analytic asymptotics by means of singularity analysis.

References

[1] J. Abate and W. Whitt. Numerical inversion of probability generating functions. *Operations Research Letters*, 12(4):245–251, 1992.

[2] J. Abate and W. Whitt. Solving probability transform functional equations for numerical inversion. *Operations Research Letters*, 12(5):275–281, 1992.

[3] J. Abate and W. Whitt. Transient behavior of the M/G/1 workload process. *Operations Research*, 42(4):750–764, 1994.

[4] I. Adan, M. Mandjes, W. Scheinhardt, and E. Tzenova. On a generic class of two-node queueing systems. *Queueing Systems*, 61(1):37–63, 2009.

[5] I. Adan, J. Van Leeuwaarden, and E. Winands. On the application of Rouché's theorem in queueing theory. *Operations Research Letters*, 34(3):355–360, 2006.

[6] O. Baron. On the law of the ith waiting time in a busy period of G/M/c queues. *Probability in the Engineering and Informational Sciences*, 22(1):75–80, 2008.

[7] E. Bender. Asymptotic methods in enumeration. *SIAM Review*, 16(4):485–515, 1974.

[8] D. Bertsimas and G. Mourtizinou. Transient laws of non-stationary queueing systems and their applications. *Queueing Systems*, 25(1-4):115–155, 1997.

[9] H. Bruneel and B. Kim. *Discrete-time models for communication systems including ATM*. Kluwer Academic Publisher, Boston, 1993.

[10] J. Cohen. *The single server queue*. North-Holland, Amsterdam, 1982.

[11] D. Daley. The serial correlation coefficients of waiting times in the stationary GI/M/1 queue. *The Annals of Mathematical Statistics*, 42(5):1724–1737, 1971.

[12] E. Gluskin and J. Walraevens. On two generalizations of the final value theorem: scientific relevance, first applications, and physical foundations. *International Journal of Systems Science*, accepted for publication.

[13] T. Hofkens, K. Spaey, and C. Blondia. Transient analysis of the D-BMAP/G/1 queue with an application to the dimensioning of a playout buffer for VBR video. *Lecture Notes in Computer Science*, 3042:1338–1343, 2004.

[14] J. Hunter. *Mathematical Techniques of Applied Probability, Volume 2, Discrete Time Models: Techniques and Applications.* Academic Press, New York, 1983.

[15] A. Janssen and J. van Leeuwaarden. Relaxation time for the discrete D/G/1 queue. *Queueing Systems*, 50(1):53–80, 2005.

[16] T. Maertens, J. Walraevens, and H. Bruneel. Priority queueing systems: from probability generating functions to tail probabilities. *Queueing Systems*, 55(1):27–39, 2007.

[17] T. Ott. The covariance function of the virtual waiting-time process in an M/G/1 queue. *Advances in Applied Probability*, 9:158–168, 1977.

[18] B. Pagurek and C. Woodside. The sum of serial correlations of waiting and system time in GI/G/1 queues. *Operations Research*, 27(4):755–766, 1979.

[19] M. Vlasiou and B. Zwart. Time-dependent behaviour of an alternating service queue. *Stochastic Models*, 23(2):235–263, 2007.

[20] J. Walraevens, D. Fiems, and H. Bruneel. Analysis of the transient delay in a discrete-time buffer with batch arrivals. In *Proceedings of the 5th International Conference on Information Technology: New Generations (ITNG 2008)*, Las Vegas, April 2008.

[21] J. Walraevens, D. Fiems, and M. Moeneclaey. Using singularity analysis to approximate transient characteristics in queueing systems. *Probability in the Engineering and Informational Sciences*, 23(2):333–355, 2009.

[22] C. Wang. On the transient delays of M/G/1 queues. *Journal of Applied Probability*, 36(3):882–893, 1999.

[23] C. Wang. An identity of the GI/G/1 transient delay and its applications. *Probability in the Engineering and Informational Sciences*, 16(1):47–66, 2002.