Probability in the Engineering and Informational Sciences, **23**, 2009, 333–355. Printed in the U.S.A.
doi:10.1017/S0269964809000199

USING SINGULARITY ANALYSIS TO APPROXIMATE TRANSIENT CHARACTERISTICS IN QUEUEING SYSTEMS

JORIS WALRAEVENS, DIETER FIEMS, AND MARC MOENECLAEY
Department of Telecommunications and Information Processing (IR07)
Ghent University
B-9000 Gent, Belgium
E-mail: {jw,df,mm}@telin.ugent.be

In this article, we develop a simple method to approximate the transient behavior of queueing systems. In particular, it is shown how singularity analysis of a known generating function of a transient sequence of some performance measure leads to an approximation of this sequence. To illustrate our approach, several specific transient sequences are investigated in detail. By means of some numerical examples, we validate our approximations and demonstrate the usefulness of the technique.

1. INTRODUCTION

Queueing models and queueing theory have been used for a number of decades to model and analyze the performance of queueing systems appearing in various applications, most notably in (digital) communication systems. In general, *input* processes are characterized and various *output* variables are analyzed. The input processes comprise the arrival process, the service times process, and the scheduling discipline. The steady-state system content, customer delay, and unfinished work are examples of output variables that are regularly analyzed.

A popular technique for analyzing queueing systems is the *generating function* technique. With this technique, the relation between stochastic variables is translated into a relation between their Laplace–Stieltjes transforms or z-transforms when dealing with continuous variables or discrete variables, respectively. The transform of the

 0269-9648/09 $25.00

stochastic variable of interest is then obtained and interesting performance measures are calculated, either analytically or numerically. These performance measures range from the mean value to the density function or probability mass function. The standard books of Takagi [18] give a nice overview of analyses of some basic queueing models by means of the transform approach.

Although there is a vast literature on transform-based *steady-state* analyses of queueing systems, results on the *transient* behavior of these queueing systems are much scarcer. For particular queueing systems, one might find explicit expressions for the time-dependent probability generating function of the queue content. This is, for example, the case for the $M/M/1$ queueing system [17] and the $M/E-r/1$ queueing system [15]. Inversion of these generating functions is possible but involve Bessel functions. Various authors have obtained expressions of the z-transform of the series of the probability generating functions of the queue content at consecutive epochs in time. Bruneel [8] performed a transient analysis of the discrete-time $M^X/Geo/1$ queue. The z-transform of the probability generating functions of the queue content is obtained in terms of known probability generating functions and the z-transform of the probability that the queue is empty at the consecutive slot boundaries. An ad hoc method is provided to find the latter probabilities. Later, Walraevens, Fiems, and Bruneel [20] extended Bruneel's approach to queueing systems with priorities. Asrin and Kamoun [5] and Kamoun [16] investigated the transient behavior of an ATM buffer with arrival traffic stemming from a fixed number of on-/off-sources with geometric on- and off-times and with geometric off-times and deterministic on-times, respectively. The authors obtained amongst others the z-transform of the probability generating functions of the queue content.

From literature, one observes that transform-based approaches often lead to "time transforms" of the performance measures of interest. For instance, if $f(t)$, $t \geq 0$, denotes the mean system content at time t, then $F^*(s) = \int_0^\infty e^{-st} f(t)\, dt$ is the corresponding Laplace transform. Another example is when f_n, $n \geq 0$, denotes the variance of the delay of the $(n+1)$st customer; then $F(z) = \sum_{n=0}^\infty f_n z^n$ is the corresponding z-transform.

Given the transforms $F^*(s)$ or $F(z)$, a second nontrivial part of the analysis is then finding $f(t), t \geq 0$ or $\{f_n, n \geq 0\}$, respectively. This is especially difficult in the context of transient analyses, since most generating functions in these analyses can only be characterized *implicitly* via functional equations. In this article, we introduce a technique to approximately invert these transforms. In particular, we focus on the inversion of z-transforms. However, a similar technique can be used to invert Laplace–Stieltjes transforms.

Different approaches to invert generating functions can be found in the literature. A first approach consists of the numerical inversion of the transform [1,3,7,10]. A second approach uses contour integration and/or Taylor series expansion to invert the transform [8,20]. Both methods, however, encounter their own problems. The numerical inversion technique involves the calculation of the generating function in a number of (complex-valued) arguments. It thus suffers from the fact that numerous calculations are necessary to accurately obtain the required transient characteristics.

https://doi.org/10.1017/S0269964809000199 Published online by Cambridge University Press

The occurrence of implicitly defined functions makes this even more cumbersome since these functions might have to be calculated iteratively in each argument. Second, since transient sequences do not necessarily approach zero (or do not even converge), the aliasing error introduced by numerical inversion techniques can be large. Abate and Whitt [2] showed, for instance, a technique in which the error depends on the maximum of $|f_n|$ for all n. This technique is effective when *probability generating functions* are inverted, since the f_n are all probabilities in that case and are thus bounded by 1, but it leads to incorrect results when used to invert a generating function of an unbounded sequence. The contour integrals approach, on the other hand, is quite ad hoc, which makes it difficult to apply to different inversion problems. Furthermore, this technique usually leads to a recursive procedure for the calculation of the transient sequence, which can get quite cumbersome.

Therefore, in this article, we look for an *approximate easy-to-use* technique to calculate transient characteristics from their generating functions. The typical inversion problem is described as follows: If $F(z)$, defined as the z-transform of a sequence $\{f_n, n \geq 0\}$, that is,

$$F(z) = \sum_{n=0}^{\infty} f_n z^n, \tag{1}$$

is a given function (either explicitly or as a function of implicitly defined functions) that is analytic at least in the open unit disk, calculate the sequence $\{f_n, n \geq 0\}$. We modify the aim in this article slightly to the calculation of as much information as possible about the behavior of the sequence $\{f_n, n \geq 0\}$ using singularity analysis. In transient analyses, f_n might be a real number between 0 and 1, a real positive number, or even a complex-valued function. Examples are respectively the probability that a discrete-time system is empty at the beginning of slot $n+1$, the mean packet delay of the $(n+1)$st customer arriving in a queueing system, and the probability generating function (p.g.f.) of the system content at the beginning of slot $n+1$ in a discrete-time system. We concentrate on real nonnegative numbers in this article, but the technique can potentially be extended to negative and complex-valued numbers or functions. So we assume the f_n to be real nonnegative numbers in the remainder.

We describe the developed technique in Section 2. The technique is based on the *dominant-singularity approximation* and is widely used in the case of *probability* generating functions, but it is largely unknown as a technique to invert generating functions of transient characteristics. We first look at some convergence/divergence properties of the sequence at hand before explaining the main procedure. We will also pay special attention on how to handle the implicitly defined functions generally appearing in the transforms.

We then apply the technique to some particular queueing systems in Section 3. We first investigate the discrete-time $M^X/Geo/1$ queue and approximate the transient probability that the system is empty at the beginning of slots, the transient mean system content, and the transient mean packet delay. We then look at a queue with a correlated arrival process, namely an arrival process originating from on–off sources. Finally, we study the transient behavior of the low-priority system content in a two-class priority queue. In all of these applications, we demonstrate that the approximate technique yields reasonable results in most cases. In some scenarios, the approximation is too crude, but, at the very least, the approximate results show how the sequence reaches its limiting value in case of converging sequences, such as the mean system content of a stable system, or how the sequence diverges (e.g., the mean system content of a nonstable system).

2. ANALYSIS

We first discuss some convergence/divergence properties of the sequence before using singularity analysis to approximate the complete sequence.

2.1. Convergence Properties of the Sequence

In this subsection, we show how convergence properties of a sequence $\{f_n, n \geq 0\}$ can be deduced from its generating function $F(z)$. We therefore use the generalization of the *final value theorem* [14], formulated as follows.

THEOREM 1 (Generalized Final Value Theorem): *If* $L = \lim_{N\to\infty} \sum_{n=0}^{N} f_n/N$ *exists, then* $\lim_{z\to 1}(1-z)F(z) = L$.

Here, L equals the *average* of all f_n. Obviously, for sequences that converge, $L = \lim_{n\to\infty} f_n$. However, Theorem 1 also includes periodic or almost-periodic functions (which do *not* converge) with a finite average.

2.2. Approximation of the Sequence

The approximate calculation of the sequence $\{f_n, n \geq 0\}$ from its generating function $F(z)$ is based on singularity analysis of generating functions. This is widely used to calculate the probability mass function from probability generating functions [9] and even more frequently in combinatorics [12]. However, it does not seem to be used yet in case of the analysis of the transient behavior of queues. As mentioned in [13], the basic principle of singularity analysis is "the existence of a correspondence between *the asymptotic expansion of a function near its dominant singularities* and *the asymptotic expansion of the function's coefficients*". We especially make use of the following theorem [6]:

THEOREM 2 (Darboux's theorem): *Suppose* $H(z) = \sum_{n=0}^{\infty} h_n z^n$ *with positive real coefficients* h_n *is analytic near zero and has only algebraic singularities* α_k *on its circle of convergence* $|z| = R$. *In other words, in a neighborhood of* α_k *we have*

$$H(z) \sim \left(1 - \frac{z}{\alpha_k}\right)^{-\omega_k} G_k(z), \tag{2}$$

https://doi.org/10.1017/S0269964809000199 Published online by Cambridge University Press

where $\omega_k \neq 0, -1, -2, \ldots$ *and* $G_k(z)$ *denotes a nonzero analytic function near* α_k. *Let* $\omega = \max_k Re(\omega_k)$ *denote the maximum of the real parts of the* ω_k. *Then we have*

$$h_n = \sum_j \frac{G_j(\alpha_j)}{\Gamma(\omega_j)} n^{\omega_j - 1} \alpha_j^{-n} + o(n^{\omega-1} R^{-n}), \quad \textbf{(3)}$$

with the sum taken over all j with $\mathrm{Re}(\omega_j) = \omega$.

Here, $H(z) \sim G(z)$ means that $H(z)/G(z) \to 1$ as z goes to the chosen complex number. Further, $h_n = o(g_n)$ means that $h_n/g_n \to 0$ as $n \to \infty$. Finally, $\Gamma(\omega)$ denotes the Gamma function of ω (with $\Gamma(n) = (n-1)!$ for n discrete).

Applying Darboux's theorem on the generating function $F(z)$ of the transient sequence $\{f_n, n \geq 0\}$, we conclude that once the behavior of $F(z)$ is characterized in its dominant singularities, the first term of (3) yields an approximation of f_n for sufficiently high n. In order to avoid a too obvious approximation of the sequence $\{f_n, n \geq 0\}$, we will use Darboux's theorem on a newly defined function $H(z)$—related to $F(z)$—rather than on $F(z)$ itself in some cases. This is discussed next, before formally stating the procedure.

We denote $\lim_{z\to 1}(1-z)F(z)$ by L, L being the average of the sequence $\{f_n, n \geq 0\}$ as in Theorem 1. Assume for the moment that L is positive and finite. Then $F(z)$ has a singularity in $z = 1$, since $\lim_{z\to 1}(1-z)F(z) \neq 0$. If we further have that

$$F(z) \sim \frac{G(z)}{1-z}$$

in the neighborhood of 1, with $G(z)$ a nonzero analytic function near 1 (with $G(1) = L$), then $z = 1$ is a pole with multiplicity 1 of $F(z)$. If this is, furthermore, the only pole on the circle of unity (which is the circle of convergence in this case), Darboux's theorem results in

$$f_n \approx L$$

for all n. Although this is obviously an approximation of the f_n (it is the average of the numbers in the sequence), we would like some more information on the behavior of the sequence. Therefore, we avoid this pole in 1 by performing Darboux's theorem on

$$H(z) = F(z) - \frac{L}{1-z} \quad \textbf{(4)}$$

rather than on $F(z)$. Note that $H(z)$ is the generating function of the sequence $\{f_n - L, n \geq 0\}$. Note further that $H(z)$ can still have a singularity in $z = 1$, if $z = 1$ was not a simple pole of $F(z)$.

The general procedure to approximate the sequence $\{f_n, n \geq 0\}$ from its generating function $F(z)$ is described as follows:

Approximation procedure

(i) Calculate $L = \lim_{z\to 1}(1-z)F(z)$.

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

https://doi.org/10.1017/S0269964809000199 Published online by Cambridge University Press

(iii) Determine the radius of convergence R of $H(z)$.
(iv) Determine the singularities α_k of $H(z)$ on its circle of convergence.
(v) Determine the behavior of $H(z)$ in the neighborhood of these α_k (see (2)).
(vi) Calculate h_n as in the first term of the right-hand side of (3).
(vii) If $0 < L < \infty$, calculate $f_n = h_n + L$ for all $n \geq 0$; otherwise $f_n = h_n$.

Remark: If $L = \infty$, the average of all numbers in the sequence $\{f_n, n \geq 0\}$ equals ∞. In this case, $H(z) = F(z)$, this function has a singularity in $z = 1$ and R equals 1. If the singularity in 1 is a pole, it has a multiplicity of at least 2.

2.3. Implicitly Defined Functions

Determining the dominant singularities of $H(z)$ and the behavior of $H(z)$ in the neighborhood of these singularities is a remaining difficulty, especially because of the occurrence of implicitly defined functions in the expressions of generating functions of transient characteristics. In general, the expression of $F(z)$ contains a function $Y(z)$ defined as

$$Y(z) = g(Y(z), z), \tag{5}$$

with $g(x, z)$ a known function. We thus want to find the dominant singularity of $Y(z)$ and the behavior of $Y(z)$ in the neighborhood of this singularity. This has been studied in numerous papers in the field of combinatorics [6,11]. Under the condition that $g(x, z)$ fulfills some mild requirements (see [6] for details),

$$Y'(z) \to \infty$$

for z going to the dominant singularity z_b, whereas $Y(z_b)$ is finite. Note that the conditions for this to be true are usually met except for some pathological cases. Then $Y(z)$ has a square-root type behavior in the neighbourhood of z_b; that is,

$$Y(z) \sim Y(z_b) - K_Y \left(1 - \frac{z}{z_b}\right)^{1/2}.$$

By calculating $Y'(z)$ from (5) it can be seen that z_B is a solution of

$$\frac{\partial g}{\partial x}(Y(z_b), z_b) = 1. \tag{6}$$

We further remark that the pair $(z_b, Y(z_b))$ can be calculated (numerically) from the set of (5) and (6).

3. SOME APPLICATIONS

In this section, we discuss some applications. First, we work out the analysis of some transient sequences in the discrete-time $M^X/Geo/1$ queue in great detail in Section 3.1. Some numerical examples are also shown for these sequences. We then look at the transient mean system content in some more general discrete-time queueing systems, namely in a queue with an arrival process governed by on–off sources in Section 3.2 and the transient mean low-priority content in a two-class priority queue in Section 3.3. By means of some figures, the approximations are compared with exact values and some conclusions are drawn about the accuracy of the approximations.

3.1. Discrete-Time $M^X/Geo/1$ Queue

3.1.1. The probability of an empty buffer. In [8], the transient system content at the beginning of slots is analyzed for a discrete-time $M^X/Geo/1$ queue. The number of arrivals per slot are independent and identically distributed (i.i.d.) and the service times are geometrically distributed with mean $1/\sigma$. It is shown that the generating function of the sequence $\{V_j(0), j \geq 0\}$—with $V_j(0)$ the probability that the system is empty at the beginning of slot $j + 1$—plays a key role in the transient analysis of the system content. Therefore, we hereby first analyze the transient probabilities that the system is empty at the beginning of slots. Denoting the generating function of the sequence $\{V_j(0), j \geq 0\}$ by $V(z)$, that is,

$$V(z) \triangleq \sum_{j=0}^{\infty} V_j(0) z^j,$$

the following expression is found in [8] for this generating function:

$$V(z) = \frac{\sigma + (1 - \sigma)Y(z)}{\sigma[1 - Y(z)]} U_0(Y(z)), \tag{7}$$

with $Y(z)$ the unique solution inside the unit disk of the x-plane for all $|z| < 1$ of

$$x - z[\sigma + (1 - \sigma)x]E(x) = 0. \tag{8}$$

Here, $E(z)$ is the p.g.f. of the number of arrivals during a random slot and $U_0(z)$ is the p.g.f. of the system content at the beginning of the first slot. We denote ρ as the load of this system (i.e., $\rho = E'(1)/\sigma$).

To apply the analysis of the previous section, we first calculate $\lim_{z\to 1}(1 - z)V(z)$, which equals the average of the sequence $\{V_j(0), j \geq 0\}$. Since the $V_j(0)$ are probabilities, this average will be a number between 0 and 1. By substituting expression (7) in this limit, we get

$$\lim_{z\to 1}(1 - z)V(z) = \lim_{z\to 1}\left[\frac{[\sigma + (1 - \sigma)Y(z)]U_0(Y(z))}{\sigma} \frac{1 - z}{1 - Y(z)}\right]. \tag{9}$$

Since $Y(z)$ is analytic and $|Y(z)| < 1$ inside the unit circle, $|\lim_{z\to 1} Y(z)| \leq 1$. Thus, the right-hand side of (9) can only be different from zero if $\lim_{z\to 1} Y(z) = 1$. Therefore,

https://doi.org/10.1017/S0269964809000199 Published online by Cambridge University Press

we turn to (8). $Y(z)$ is the unique solution of (8) inside the unit disk for all $|z| < 1$. We first note that $(x, z) = (1, 1)$ is *always* a solution of (8). However, this does not necessarily mean that $Y(1) = 1$, since $(x, z) = (1, 1)$ could also be the limit of a solution $Y^*(z)$ of (8) *outside* the unit disk of the x-plane for $|z| < 1$. We thus look for all solutions inside and on the unit disk of (8) for $z = 1$, since one of these solutions equals $Y(1)$. $Y(1)$ is a solution of

$$x - \frac{\sigma E(x)}{1 - (1 - \sigma)E(x)} = 0, \qquad \textbf{(10)}$$

which is obtained by substituting z by 1 in (8) and by multiplying both sides with $1/[1 - (1 - \sigma)E(x)]$. We note that the multiplying factor has no zero inside or on the unit disk. Further, note that $\sigma E(x)/[1 - (1 - \sigma)E(x)]$ is the p.g.f. of the number of arrivals during the service time of a random customer.

It turns out—not unexpectedly—that three cases can be distinguished, namely $\rho < 1$, $\rho = 1$, and $\rho > 1$. We treat the three cases separately in the remainder.

Case 1: $\rho < 1$. In this case, it can be proved my means of Rouché's theorem (or by the generalized version proved in [4]) that (10) has exactly one solution inside and on the unit disk. Since $x = 1$ is a solution, this is the unique solution inside and on the unit disk of (10) in this case. Therefore, $Y(1) = 1$. So, in order to calculate $\lim_{z\to 1}(1 - z)V(z)$, we use de l'Hôpital's rule in (9), yielding

$$\lim_{z\to 1}(1 - z)V(z) = \frac{1}{\sigma} \lim_{z\to 1} \frac{1}{Y'(z)}.$$

The first derivative of $Y(z)$ can be found by substituting x by $Y(z)$ in (8) and taking the first derivative of both sides of this equation. The limit for z to 1 then yields

$$\lim_{z\to 1} Y'(z) = \frac{1}{\sigma(1 - \rho)} \qquad \textbf{(11)}$$

and, thus,

$$\lim_{z\to 1}(1 - z)V(z) = 1 - \rho. \qquad \textbf{(12)}$$

This was expected since this is indeed the steady-state probability of an empty buffer in a stable system. We then have to use Darboux's theorem on (see the approximation procedure)

$$V(z) - \frac{1 - \rho}{1 - z} = \frac{\sigma + (1 - \sigma)Y(z)}{\sigma[1 - Y(z)]} U_0(Y(z)) - \frac{1 - \rho}{1 - z}.$$

The dominant singularity of this function is either the square-root branch point of $Y(z)$ or a singularity of $U_0(Y(z))$. In a later paragraph, we discuss this for some specific arrival processes and a given initial system content distribution.

https://doi.org/10.1017/S0269964809000199 Published online by Cambridge University Press

Case 2: $\rho = 1$. In the special subcase that $\sigma = 1$ and $E(x) = x$, the only solution of (8) inside the unit disk of the x-plane for all $|z| < 1$ equals $Y(x) = 0$. Thus, $Y(1) = 0$ in this case and $\lim_{z\to 1}(1 - z)V(z) = 0$. In all other subcases of the case $\rho = 1$, it can be proved (see the Appendix) that $Y(1) = 1$ if $\rho = 1$. This is thus the same as in the case $\rho < 1$. The reasoning of this latter case thus applies to the case $\rho = 1$. Equations (11) and (12) lead to

$$\lim_{z\to 1} Y'(z) = \infty$$

and

$$\lim_{z\to 1}(1 - z)V(z) = 0,$$

respectively. $z_b = 1$ is a square-root branch point of $Y(z)$ since $Y(1) = 1$ and $Y'(1) \to \infty$. Clearly, $z = 1$ is the dominant singularity in this case.

We conclude that $\lim_{z\to 1}(1 - z)V(z) = 0$ when $\rho = 1$.

Case 3: $\rho > 1$. Since $A(z) = \sigma E(z)/[1 - (1 - \sigma)E(z)]$ is a monotonously increasing function in $[0, 1]$ and since $A(0) \geq 0$, $A(1) = 1$, and $A'(1) > 1$, (10) has a real solution in the segment $[0, 1]$, denoted by r. r is the limit of one of the solutions of (8) for $z \to 1$. Since a real number in $[0, 1]$ cannot be the limit of a function *outside* the unit disk and since $Y(z)$ is the only solution of (8) *inside* the unit disk, $r = Y(1)$. Thus, $Y(1) < 1$ and $\lim_{z\to 1}(1 - z)V(z) = 0$. The dominant singularity of $V(z)$ is again the square-root branch point of $Y(z)$ or a singularity of $U_0(Y(z))$.

Examples: We now calculate the transient probabilities of an empty buffer for some specific input distributions and input parameters. We assume that the service times equal 1 slot ($\sigma = 1$) and that the system is empty at the beginning ($U_0(z) = 1$). We discuss the results for two different distributions of the arrival batch sizes.

In a first example, we assume the number of per-slot arrivals to be geometrically distributed with mean ρ; that is,

$$E(z) = \frac{1}{1 + \rho - \rho z}.$$

In this case, an explicit expression can be found for $Y(z)$; $V(z)$ is given by

$$V(z) = \frac{1 - \rho + ((1 + \rho)^2 - 4\rho z)^{1/2}}{2(1 - z)}. \tag{13}$$

As a result, two singularities of $V(z)$ may be dominant—depending on the value of ρ—namely $z_r = 1$ and/or the square-root branch point z_b of $Y(z)$ given by

$$z_b = \frac{(1 + \rho)^2}{4\rho}. \tag{14}$$

https://doi.org/10.1017/S0269964809000199 Published online by Cambridge University Press

In accordance with the approximation procedure, we have to invert the following functions:

$$V(z) - \frac{1-\rho}{1-z} = -\frac{1-\rho}{2(1-z)} + \left(1 - \frac{z}{z_b}\right)^{1/2} G(z) \qquad (\rho < 1),$$

$$V(z) = (1-z)^{-1/2} \qquad (\rho = 1),$$

$$V(z) = -\frac{\rho - 1}{2(1-z)} + \left(1 - \frac{z}{z_b}\right)^{1/2} G(z) \qquad (\rho > 1).$$

Here,

$$G(z) = \frac{1+\rho}{2(1-z)}.$$

Further applying the approximation procedure, we find the probability that the system is empty at the beginning of slot $j+1$:

$$V_j(0) \approx \begin{cases} 1 - \rho + \dfrac{\rho(1+\rho)}{(1-\rho)^2 j^{3/2}\sqrt{\pi}} \left(\dfrac{(1+\rho)^2}{4\rho}\right)^{-j} & \text{if } \rho < 1 \\ \dfrac{1}{j^{1/2}\sqrt{\pi}} & \text{if } \rho = 1 \\ \dfrac{\rho(1+\rho)}{(\rho-1)^2 j^{3/2}\sqrt{\pi}} \left(\dfrac{(1+\rho)^2}{4\rho}\right)^{-j} & \text{if } \rho > 1. \end{cases}$$

We illustrate the approximate analysis by means of some figures. In Figure 1, the transient probability $V_j(0)$ of having an empty system is plotted versus the discrete-time parameter j for $\rho = 0.2, 0.4, 0.6, 0.8$, and 1. We also show the exact results, which are calculated by using the iterative procedure discussed in [8]. These exact results are represented by dots on the figures. We see from Figure 1 that the approximation goes to the correct steady-state value. For low loads, it seems that the approximation is already good for rather low j. For higher loads (<1), the approximation is less accurate. However, for $\rho = 1$, the approximation is excellent. Figure 2 shows a logarithmic plot of the transient probability $V_j(0)$ of having an empty system versus the discrete-time parameter j for some overload scenarios, namely for $\rho = 1$, 2, 3, 4, and 5. We again also depict the exact values, found via recursion. For high loads, the approximation is excellent.

In the next example, the number of per-slot arrivals are assumed to be Poisson distributed with mean ρ; that is,

$$E(z) = e^{\rho(z-1)}.$$

In this case, $V(z)$ is given by

$$V(z) = \frac{1}{1 - Y(z)},$$

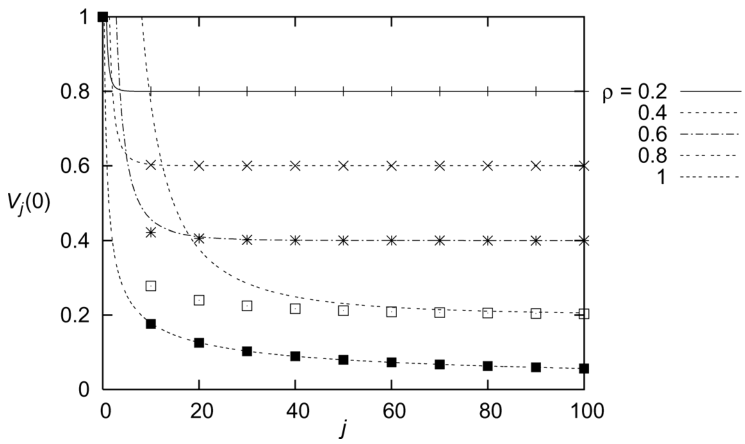

FIGURE 1. Transient probabilities of having an empty system for underload scenarios.

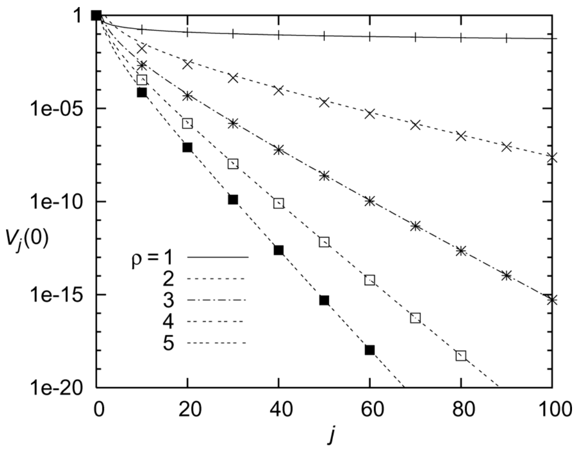

FIGURE 2. Transient probabilities of having an empty system for overload scenarios.

with $Y(z)$ implicitly defined by

$$Y(z) = ze^{\rho(Y(z)-1)} \tag{15}$$

such that $|Y(z)| < 1$ for $|z| < 1$. Again, two singularities might be dominant—depending on the value of ρ—namely $z_r = 1$ and the square-root branch point z_b of $Y(z)$ given by

$$z_b = \frac{1}{\rho e^{1-\rho}}. \tag{16}$$

https://doi.org/10.1017/S0269964809000199 Published online by Cambridge University Press

Expression (16) is found by taking the derivative of both sides of (15), substituting z by $Y(z)/e^{\rho(Y(z)-1)}$ and noting that $\lim_{z\to z_b} Y'(z) = \infty$. This yields $Y(z_b) = 1/\rho$, which, in turn, finally yields (16). So what remains to be found is the behavior of $V(z)$ in the neighborhood of z_b. First $Y(z)$ can be written as

$$Y(z) \sim Y(z_b) - K_Y(1 - z/z_b)^{1/2} \qquad \textbf{(17)}$$

in the neighborhood of z_b. K_Y is found as the square root of

$$K_Y^2 = z_b \lim_{z\to z_b} \frac{(Y(z_b) - Y(z))^2}{z_b - z},$$

which leads to

$$K_Y = \frac{\sqrt{2}}{\rho},$$

by using de l'Hôpital's rule, by writing $Y'(z)$ as a function of z and $Y(z)$, and by taking the limit for z going to z_b. Thus, in the neighborhood of z_b, $Y(z)$ and $V(z)$ are, respectively, given by

$$Y(z) \sim \frac{1}{\rho} - \frac{\sqrt{2}}{\rho}(1 - z/z_b)^{1/2}$$

and

$$V(z) \sim \frac{1 - 1/\rho - \sqrt{2}(1 - z/z_b)^{1/2}/\rho}{(1 - 1/\rho)^2 - 2(1 - z/z_b)/\rho^2}.$$

The following expressions for the probability that the system is empty at the beginning of slot j are then found by applying the approximation procedure:

$$V_j(0) \approx \begin{cases} 1 - \rho + \dfrac{\rho}{\sqrt{2}(1-\rho)^2 j^{3/2}\sqrt{\pi}}\left(\dfrac{1}{\rho e^{1-\rho}}\right)^{-j} & \text{if } \rho < 1 \\ \dfrac{1}{\sqrt{2} j^{1/2}\sqrt{\pi}} & \text{if } \rho = 1 \\ \dfrac{\rho}{\sqrt{2}(\rho-1)^2 j^{3/2}\sqrt{\pi}}\left(\dfrac{e^{\rho-1}}{\rho}\right)^{-j} & \text{if } \rho > 1. \end{cases}$$

Similar figures can be plotted as in the case of geometrically distributed batch sizes and the same conclusions can be drawn.

3.1.2. The mean system content. Next, we look at the mean transient system content in the $M^X/Geo/1$ queue. Again, we start from a result obtained in [8]:

https://doi.org/10.1017/S0269964809000199 Published online by Cambridge University Press

the generating function of the sequence $\{\bar{u}_j,\ j \geq 0\}$—with $\bar{u}_j$ the expected system content at the beginning of slot $j+1$—is denoted by $\bar{U}(z)$ and given by

$$\bar{U}(z) = \frac{\bar{u}_0}{1-z} + \frac{z[\sigma + (1-\sigma)Y(z)]}{(1-z)(1-Y(z))} U_0(Y(z)) - \frac{(\sigma - E'(1))z}{(1-z)^2}.$$

As in Section 3.1.1, the service times are geometrically distributed with mean $1/\sigma$, $E(z)$ is the p.g.f. of the number of arrivals during a slot, and $U_0(z)$ denotes the p.g.f. of the system content at the beginning of the first slot. Further, $Y(z)$ is again the unique solution inside the unit disk of the x-plane for all $|z| < 1$ of (8). The same three cases as in the previous example can be distinguished. We briefly summarize some properties for the three cases. For more details, we refer to Section 3.1.1.

Case 1: $\rho < 1$. In this case, $Y(1)$ equals 1. We have

$$\begin{aligned}&\lim_{z\to 1}(1-z)\bar{U}(z)\\ &\quad = \bar{u}_0 + \lim_{z\to 1} \frac{z[(\sigma + (1-\sigma)Y(z))U_0(Y(z))(1-z) + (\sigma - E'(1))(1-Y(z))]}{(1-Y(z))(1-z)}.\end{aligned}$$

By using de l'Hôpital's rule and the implicit definition of $Y(z)$, this expression is transformed to

$$\lim_{z\to 1}(1-z)\bar{U}(z) = \frac{\rho(1-\lambda)}{1-\rho} + \frac{E''(1)}{2\sigma(1-\rho)}.$$

This is the mean steady-state system content at the beginning of a random slot in a stable system, as expected. Following the approximation procedure, we use Darboux's theorem on

$$\bar{U}(z) - \left(\frac{\rho(1-\lambda)}{1-\rho} + \frac{E''(1)}{2\sigma(1-\rho)}\right)(1-z)^{-1}.$$

The dominant singularity is again either the square-root branch point of $Y(z)$ or a singularity of $U_0(Y(z))$.

Case 2: $\rho = 1$. In the special subcase that $\sigma = 1$ and $E(x) = x$, we have $Y(1) = 0$. We then find

$$\lim_{z\to 1}(1-z)\bar{U}(z) = \bar{u}_0 + U_0(0).$$

We could thus use the approximation procedure to find an approximation of the probabilities of the mean system content. However, in this pathological subcase, $\bar{U}(z)$ can be easily inverted exactly. Except for this special subcase, $Y(1)$ equals 1 when $\rho = 1$. The reasoning of case 1 thus applies, leading to

$$\lim_{z\to 1}(1-z)\bar{U}(z) = \infty.$$

In this case, $R = 1$ is the radius of convergence.

https://doi.org/10.1017/S0269964809000199 Published online by Cambridge University Press

Case 3: $\rho > 1$. In this case, $Y(1) < 1$, and as a result, we have

$$\lim_{z \to 1} (1 - z)\bar{U}(z) = \infty.$$

Again, the radius of convergence R equals 1.

Example: In this example, we assume that the service times equal 1 slot ($\sigma = 1$) and we assume that the system is empty at the beginning ($U_0(z) = 1$ and $\bar{u}_0 = 0$). We discuss the results for geometrically distributed arriving batch sizes with mean ρ; that is,

$$E(z) = \frac{1}{1 + \rho - \rho z}.$$

In this case, $\bar{U}(z)$ is given by

$$\bar{U}(z) = \frac{z[-(1-\rho) + ((1+\rho)^2 - 4\rho z)^{1/2}]}{2(1-z)^2}. \tag{18}$$

The same two singularities as for the probability of an empty system might be dominant, namely $z_r = 1$ and the square-root branch point z_b of $Y(z)$ as given in (14). The following expressions for the mean system content at the beginning of slot j are found:

$$\bar{u}_j \approx \begin{cases} \dfrac{\rho}{1-\rho} - \dfrac{\rho(1+\rho)^3}{(1-\rho)^4 j^{3/2}\sqrt{\pi}} \left(\dfrac{(1+\rho)^2}{4\rho}\right)^{-j} & \text{if } \rho < 1 \\ \dfrac{2j^{1/2}}{\sqrt{\pi}} & \text{if } \rho = 1 \\ (\rho - 1)j & \text{if } \rho > 1. \end{cases}$$

We illustrate the approximate analysis by means of some figures. In Figure 3, the transient mean system content $\bar{u}_j$ is plotted versus the discrete-time parameter j for $\rho = 0.2, 0.4, 0.6, 0.8$, and 1. We have also shown the exact results (dots on the figure), which are again calculated by using the iterative procedure explained in [8]. Again for low loads, it seems that the approximation is already good for rather low j. For higher loads (<1), the approximation is less accurate. However, for $\rho = 1$, the approximation is excellent. Figure 4 shows a logarithmic plot of the transient mean system content $\bar{u}_j$ versus the discrete-time parameter j for some overload scenarios, namely for $\rho = 1, 2$, 3, 4, and 5. We have again shown the exact values, found via recursion. For high loads, the approximation seems to be rather good for all j and improves for increasing load.

3.1.3. The mean packet delay. We have chosen this example to demonstrate the approximation when the sequence is an almost-periodic function. The sequence under consideration is $\{\bar{d}_j,\ j \geq 0\}$, with $\bar{d}_j$ the mean transient customer delay of the $(j+1)$st arriving customer in a discrete-time FIFO $M^X/D/1$ queue with single-slot

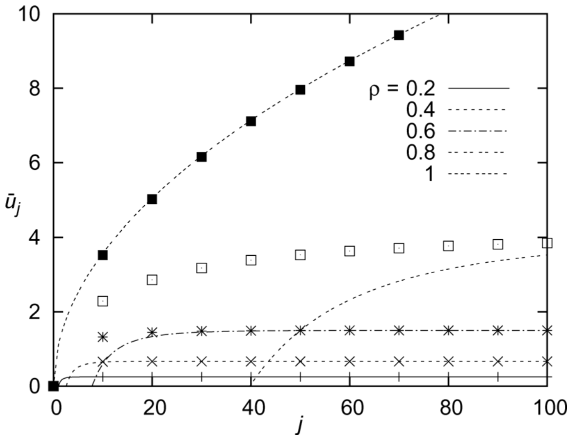

FIGURE 3. Mean transient system content for underload scenarios.

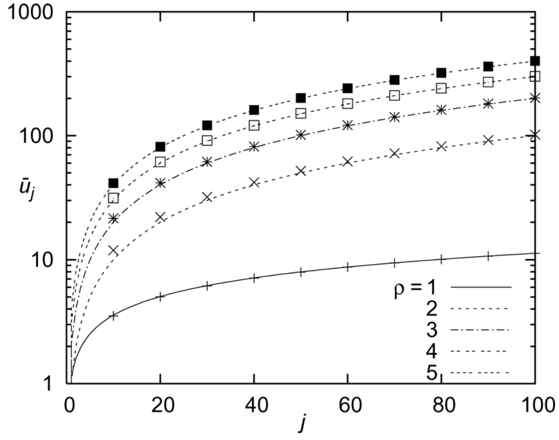

FIGURE 4. Mean transient system content for overload scenarios.

service times. The p.g.f. of the arriving batch sizes is given by

$$E(z) = 1 - \frac{\rho}{2} + \frac{\rho}{2} z^2. \tag{19}$$

The customers thus arrive in pairs. The generating function of the sequence $\{\bar{d}_j, j \geq 0\}$ is calculated in [19] (for general $E(z)$) and is given by

$$\bar{D}(z) = \frac{\bar{d}_0}{1-z} + \frac{z}{(1-z)^2} + \frac{E(z) - E(0)}{(1-E(0))(1-z)(E(z)-1)} + \frac{(Y(z) - E(0))D_0(Y(z))}{(1-E(0))(1-z)Y(z)(1-Y(z))}, \tag{20}$$

with $D_0(z)$ the p.g.f. of the customer delay of the first arriving customer and $Y(z)$ the unique solution inside the unit disk of the x-plane for all $|z| < 1$ of

$$x - E(xz) = 0.$$

Again, the three possible cases $\rho < 1$, $\rho = 1$, and $\rho > 1$ can be distinguished as in the previous examples. We only focus on the stable case here, (i.e., $\rho < 1$), since the purpose of this example is to show that oscillating behavior is "detected" using singularity analysis. For $\rho < 1$, we have

$$\lim_{z\to 1}(1 - z)\bar{D}(z) = 1 + \frac{E''(1)}{2\rho(1 - \rho)}.$$

This expression is valid for general batch sizes; it is indeed the mean delay of a randomly arriving customer in the steady state of a stable $M^X/D/1$ queue with a FIFO scheduling discipline.

For the batch size distribution as specified in (19), $Y(z)$ is given by

$$Y(z) = \frac{1 - [1 - \rho(2 - \rho)z^2]^{1/2}}{\rho z^2}. \qquad \textbf{(21)}$$

Assuming that the first customer does not have to wait (i.e., its delay equals 1 slot and thus $D_0(z) = z$ and $\bar{d}_0 = 1$), (20) has four possible dominant singularities, namely $z_r = 1$,

$$z_b = \sqrt{\frac{1}{\rho(2 - \rho)}},$$

$-z_r$, and $-z_b$. We use the approximation procedure to obtain an approximation of the sequence $\{\bar{d}_j, j \geq 0\}$. The dominant singularity of

$$\bar{D}(z) - \left(1 + \frac{E''(1)}{2\rho(1 - \rho)}\right)(1 - z)^{-1} \qquad \textbf{(22)}$$

is -1 and we obtain

$$\bar{d}_j \approx 1 + \frac{E''(1)}{2\rho(1 - \rho)} - \frac{(-1)^j}{2}.$$

Thus, using the procedure on this example, the *oscillating* behavior rather than the transient behavior is exposed. Indeed, the previous formula gives the steady-state mean delay of a customer arriving first in his batch (j even) or arriving second (j odd).

Note that if one would want to examine the transient behavior in this case, one could do something similar to what was done in the approximation procedure: subtract $1/[2(1 + z)]$ of the expression in (22) to avoid the singularity in -1. In this way, the singularities in -1 and 1 are both avoided and the dominant singularities are z_b and $-z_b$. Thus, using Darboux's theorem on this function approximates the mean transient customer delay.

https://doi.org/10.1017/S0269964809000199 Published online by Cambridge University Press

3.2. A Queue with On–Off Sources

In this example, we analyze the transient probabilities that a discrete-time queue fed by N on–off sources is empty at the beginning of slots. We assume that the sources send no packets when they are in the off state and send messages of a fixed number m of packets at the rate of one packet per slot when they are in the on state. A source that is in the off state during a certain slot switches to the on state at the end of that slot with probability q. It then stays in the on state for at least m slots to send a message and then either goes back to the off state with probability $1 - q$ or stays in the on state with probability q to generate another message.

This queueing system is analyzed by Kamoun [16]. Among other characteristics, an expression for $V_j(0)$, the probability that the system is empty at the beginning of slot $j + 1, j \geq 0$, and their generating function $V(z)$, is given in the case that $m = 2$. Here, we will approximate the $V_j(0)$ by inverting $V(z)$ using the approximate technique from this article and compare them with the exact results given in [16]. $V(z)$ is given by

$$V(z) = \frac{1}{1 - Y(z)}, \tag{23}$$

with $Y(z)$ the unique root inside the unit disk of the x-plane of the equation

$$x = z\lambda(x)^N$$

for $|z| \leq 1$. Here, $\lambda(z)$ denotes the unique root for x of the characteristic equation

$$x^m - (1 - q)x^{m-1} - qz^m$$

that equals 1 in $z = 1$.

We apply the approximation procedure to expression (23) in the case that $m = 2$. Again, a distinction can be made based on the value of the load ρ, which is in this queueing system given by (see [16])

$$\rho = \frac{Nqm}{1 + (m - 1)q}.$$

The important difference with the previous analysis is that in this case $Y(z)$ has two branch points on its circle of convergence: one on the positive axis and one on the negative axis. The further calculations are similar to those in Section 3.1. Therefore, we omit them and show a numerical example instead. Figure 5 depicts the probability of having an empty buffer at the beginning of the $(j + 1)$st slot for a queue with $N = 4$ on–off sources and a load ρ equal to 0.2, 1, and 2, respectively. We have shown the consequent exact results (obtained from [16]) with marks. From Figure 5, we can once again see that the obtained approximations are good. Particularly striking in this example is that the plots are not converging monotonously to the steady-state value but that some oscillating behavior with a period of 2 slots is observed. This effect is beautifully predicted by our approximate analysis and thus matched by the

https://doi.org/10.1017/S0269964809000199 Published online by Cambridge University Press

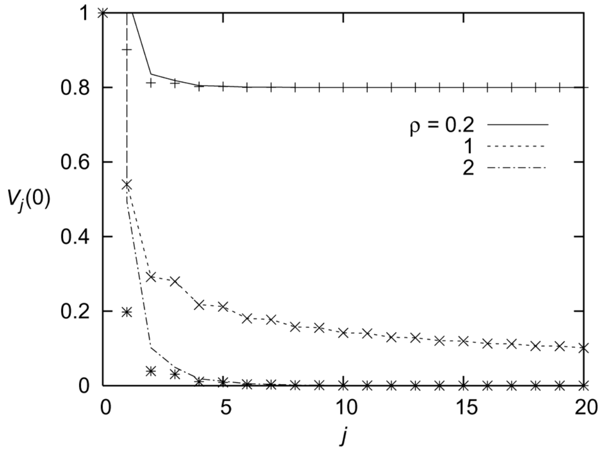

FIGURE 5. Transient probabilities of having an empty system in a queue with on–off sources.

curves for our approximations. Indeed, instead of only one dominant singularity on the positive real axis, the generating function $V(z)$ has a second dominant singularity on the negative real axis, which accounts for the oscillating behavior (see also the discussion in Section 3.1.3). We can thus conclude that this oscillating behavior is directly related to the number and location of the dominant singularities on the circle of convergence. Our procedure quantifies this oscillating behavior as well as the converging or diverging course of the sequences.

3.3. Low-Priority System Content in a Two-Class Priority Queue

As a final application, we discuss the low-priority system content in a discrete-time priority queue. The numbers of per-slot packet arrivals are i.i.d. and the numbers of high-priority and low-priority packet arrivals in a slot have a general two-dimensional distribution. The service times are equal to 1 slot. The transient behavior of this system is analyzed in [20]. In this subsection, we apply our procedure to the transform function $\bar{U}_2(z)$ of $\{\bar{u}_2^{(j)}, j \geq 0\}$, the mean low-priority system content at the beginning of slots. This transform function is given by (see [20])

$$\bar{U}_2(z) = \frac{\bar{u}_2^{(0)}}{1-z} + \frac{zU_T^{(0)}(Y_T(z))}{(1-z)(1-Y_T(z))} - \frac{zU_1^{(0)}(Y_1(z))}{(1-z)(1-Y_1(z))} + \frac{\rho_2 z}{(1-z)^2}, \quad \textbf{(24)}$$

with $U_T^{(0)}(z)$ and $U_1^{(0)}(z)$ the p.g.f.'s of the total and high-priority system content at the beginning of the first slot and $Y_1(z)$ and $Y_T(z)$ the unique solutions for x inside the unit disk of $x = zA_1(x)$ and $x = zA_T(x)$, respectively, for $|z| < 1$; $A_1(z)$ and $A_T(z)$ are the p.g.f.'s of the numbers of per-slot high-priority arrivals and the total number of arrivals in a slot, respectively.

https://doi.org/10.1017/S0269964809000199 Published online by Cambridge University Press

We again use the approximation procedure on the expression of the generating function (24) of the time-dependent sequence $\{\bar{u}_2^{(j)}, j \geq 0\}$ to obtain approximations of this sequence. We discuss an example in the remainder.

Example: We assume the system to be empty at the beginning; thus, $U_1^{(0)}(z) = 1$, $U_T^{(0)}(z) = 1$ and $\bar{u}_2^{(0)} = 0$. The high-priority and low-priority arriving batch sizes are Poisson distributed with mean ρ_1 and ρ_2, respectively. ρ_T is defined as the total load and is given by $\rho_1 + \rho_2$.

In this case, $\bar{U}_2(z)$ is given by

$$\bar{U}_2(z) = \frac{z}{(1-z)(1-Y_T(z))} - \frac{z}{(1-z)(1-Y_1(z))} + \frac{\rho_2 z}{(1-z)^2}.$$

Three singularities can play a role, namely 1, the square-root branch point $z_{1,b}$ of $Y_1(z)$, and the square-root branch point $z_{T,b}$ of $Y_T(z)$. It can easily be proved that these branch points are given by

$$z_{1,b} = \frac{1}{\rho_1 e^{1-\rho_1}}$$

and

$$z_{T,b} = \frac{1}{\rho_T e^{1-\rho_T}}.$$

Which of the singularities is dominant depends on the load of both classes. One can, for instance, show that $z_{1,b}$ is never dominant, except when it equals 1 (i.e., when $\rho_1 = 1$). The calculations are again rather similar to those in Section 3.1. The following expressions for the mean low-priority system content at the beginning of slot j are found:

$$\bar{u}_2^{(j)} \approx \begin{cases} \bar{u}_2^{(\infty)} + \dfrac{\rho_T (1/\rho_T e^{1-\rho_T})^{-j}}{\sqrt{2\pi}(\rho_T e^{1-\rho_T} - 1)(1-\rho_T)^2 j^{3/2}} & \text{if } \rho_T < 1 \\ \dfrac{\sqrt{2} j^{1/2}}{\sqrt{\pi}} & \text{if } \rho_T = 1 \\ (\rho_T - 1) j & \text{if } \rho_1 < 1 < \rho_T \\ \rho_2 j & \text{if } \rho_1 \geq 1, \end{cases}$$

with $\bar{u}_2^{(\infty)}$ the mean low-priority system content of a stable system in the steady state and given by [21]

$$\bar{u}_2^{(\infty)} = \rho_2 + \frac{\rho_T^2}{2(1-\rho_T)} - \frac{\rho_1^2}{2(1-\rho_1)}.$$

We illustrate the approximate results by means of two figures. In Figure 6, the mean transient low-priority system content $\bar{u}_2^{(j)}$ is depicted versus the discrete-time parameter j for $\rho_1 = 0.2$ and $\rho_T = 0.4, 0.6, 0.8$, and 1. We have also shown the exact results

https://doi.org/10.1017/S0269964809000199 Published online by Cambridge University Press

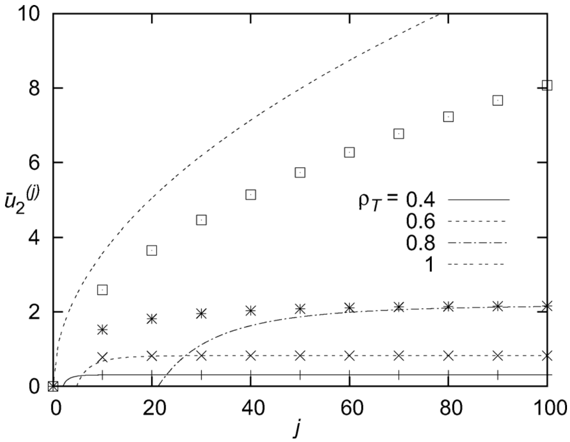

FIGURE 6. Mean transient low-priority system content for underload scenarios.

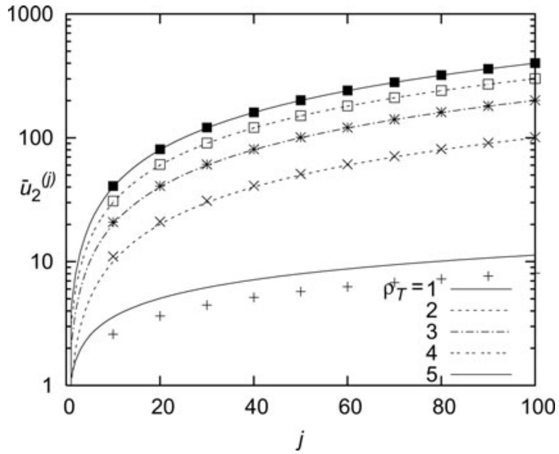

FIGURE 7. Mean transient low-priority system content for overload scenarios.

(dots on the figure), which are calculated by using the iterative procedure explained in [20]. For low loads, it seems that the approximation is already good for rather low j. For higher loads (<1), the approximation is less accurate. In this case, we observe that the curve for the case $\rho_T = 1$ is not satisfactory. (We note that a small adjustment of the method also yields accurate results in this case, but this is outside the scope of the current article.) Figure 7 shows a logarithmic plot of the transient mean system content $\bar{u}_2^{(j)}$ versus the discrete-time parameter j for some overload scenarios, namely for $\rho_1 = 0.2$ and $\rho_T = 1, 2, 3, 4,$ and 5. We have again shown the exact values, found via recursion. For high loads, the approximation is once again good for all j.

4. CONCLUSIONS

In this article, we have developed a general technique to approximate transient sequences from their generating function. The technique is based on singularity analysis: By studying the behavior of the generating function in its dominant singularities, we obtain an asymptotically exact approximation of the sequence. The main advantages of the approach are that the technique is generally applicable, easy to use, and yields *analytic* results. We have also shown that quite some characteristics of an unknown transient sequence can be found by studying the dominant singularities of its generating function, most prominently, converging or diverging behavior and possible oscillations.

We have applied the technique to analyze the transient behavior of the discrete-time $M^X/Geo/1$ queue, of a queue fed by on–off sources, and of a priority queue. It was demonstrated that the technique yields good results in most cases. At the very least, the results show the asymptotic behavior for the time index going to infinity. In some cases, however (especially for loads around 1), the approximation is too crude for the lower slot indexes. Further research is necessary to investigate whether the analysis can be adapted to yield better approximations in those cases as well. This could be an ad hoc method for a specific sequence or—preferably—a general applicable extension of the approach of this article. In [13], for example, such singularity analysis extensions are explained, but it remains to be seen if they work in the context of transient performance analysis of queues. We must note though that more accurate results are only possible through a more complex analysis.

Acknowledgments

The first two authors are postdoctoral fellows with the Fund for Scientific Research, Flanders (F.W.O.-Vlaanderen), Belgium. The authors wish to thank the referee for insightful comments, which improved the presentation of the article and the validity of the technique.

References

1. Abate, J., Choudhury, G. & Whitt, W. (1995). Calculating the $M/G/1$ busy-period density and LIFO waiting-time distribution by direct numerical inversion. *Operations Research Letters* 18(3): 113–119.
2. Abate, J. & Whitt, W. (1992). Numerical inversion of probability generating functions. *Operations Research Letters* 12(4): 245–251.
3. Abate, J. & Whitt, W. (1992). Solving probability transform functional equations for numerical inversion. *Operations Research Letters* 12(5): 275–281.
4. Adan, I., Van Leeuwaarden, J. & Winands, E. (2006). On the application of Rouché's theorem in queueing theory. *Operations Research Letters* 34(3): 355–360.
5. Asrin, M. & Kamoun, F. (1998). A transient discrete-time analysis of the ATM multiplexer. *Performance Evaluation* 32(3): 153–183.
6. Bender, E. (1974). Asymptotic methods in enumeration. *SIAM Review* 16(4): 485–515.
7. Blanc, J. (2002). On the numerical inversion of busy-period related transforms. *Operations Research Letters* 30(1): 33–42.
8. Bruneel, H. (1991). Exact derivation of transient behavior for buffers with random output interruptions. *Computers Networks and ISDN Systems* 22: 277–285.

9. Bruneel, H., Steyaert, B., Desmet, E. & Petit, G. (1994). Analytic derivation of tail probabilities for queue lengths and waiting times in ATM multiserver queues. *European Journal of Operational Research* 76(3): 563–572.
10. Choudhury, G., Lucantoni, D. & Whitt, W. (1994). Multidimensional transform inversion with applications to the transient $M/G/1$ queue. *Annals of Applied Probability* 4(3): 719–740.
11. Drmota, M. (1997). Systems of functional equations. *Random Structures & Algorithms* 10(1–2): 103–124.
12. Flajolet, P. & Odlyzko, A. (1990). Singularity analysis of generating functions. *SIAM Journal on Discrete Mathematics* 3(2): 216–240.
13. Flajolet, P. & Sedgewick, R. (2009). *Analytic Combinatorics*. Cambridge: Cambridge University Press.
14. Gluskin, E. (2003). Let us teach this generalization of the final-value theorem. *European Journal of Physics* 24(6): 591–597.
15. Griffiths, J., Leonenko, G. & Williams, J. (2006). The transient solution to $M/E-k/1$ queue. *Operations Research Letters* 34(3): 349–354.
16. Kamoun, F. (2006). Performance analysis of a discrete-time queuing system with a correlated train arrival process. *Performance Evaluation* 63(4–5): 315–340.
17. Parthasarathy, P. (1987). A transient solution to an $M/M/1$ queue: a simple approach. *Advances in Applied Probability* 19(4): 997–998.
18. Takagi, H. (1991–1993). *Queueing analysis: A foundation of performance evaluation*. Amsterdam: North-Holland, Vols I–III.
19. Walraevens, J., Fiems, D. & Bruneel, H. (2008). 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.
20. Walraevens, J., Fiems, D. & Bruneel, H. (2008). Time-dependent performance analysis of a discrete-time priority queue. *Performance Evaluation* 65(9): 641–652.
21. Walraevens, J., Steyaert, B. & Bruneel, H. (2003). Performance analysis of a single-server ATM queue with a priority scheduling. *Computers and Operations Research* 30(12): 1807–1829.

APPENDIX

In this Appendix, we prove that $Y(1) = 1$ when $\rho = 1$. We do this by proving that (10) has only one solution inside and on the unit circle when $\rho = 1$, namely $x = 1$. We in fact prove it in a more general setting: We prove that

$$x - A(x) = 0 \tag{A.1}$$

has no solution inside $\bar{D}\backslash\{1\}$, the closed complex unit disk minus the point 1. Here, $A(x)$ is a p.g.f. with $\rho = A'(1) = 1$. (Note that $A(x) = x$ is excluded here since this special case was already treated in the article.) The wanted result then follows by substituting $A(x)$ by

$$\frac{\sigma E(x)}{1 - (1 - \sigma)E(x)}.$$

We denote the stochastic variable corresponding with $A(x)$ by a; that is,

$$A(x) = \sum_{n=0}^{\infty} \text{Prob}[a = n]x^n.$$

We further introduce $A_c(x)$ defined as

$$A_c(x) = \sum_{n=0}^{\infty} \text{Prob}[a > n]x^n.$$

The following relation between $A_c(x)$ and $A(x)$ is then easily established:

$$A_c(x) = \frac{A(x) - 1}{x - 1}.$$

Note that $A_c(x)$ is a p.g.f. in the special case that $\rho = 1$. We will use this property later. Introducing $A_c(x)$, (A.1) can be transformed into

$$(x - 1)(1 - A_c(x)) = 0.$$

$x - 1$ has no zero in $\bar{D}\backslash\{1\}$, so the solutions of (A.1) in $\bar{D}\backslash\{1\}$ equal the solutions of

$$1 - A_c(x) = 0. \tag{A.2}$$

Since $|A_c(x)| < 1$ for $|x| < 1$—$A_c(x)$ is a p.g.f. when $\rho = 1$—Rouché's theorem yields that (A.2) has no solution inside an arbitrary contour in the unit disk. As a result, (A.2) has no solution inside the unit circle. It can further be proved that (A.2) has no solutions *on* the unit circle either except for $x = 1$. Indeed, $A_c(x)$ on the complex unit disk can be written as

$$A_c(e^{2\pi t}) = \text{Prob}[a > 0] + \text{Prob}[a > 1]e^{2\pi t} + \sum_{n=2}^{\infty} \text{Prob}[a > n]e^{2\pi nt}. \tag{A.3}$$

For this expression to equal 1 for a $t \in [0, 1]$, Prob$[a > 1]$ has to be zero. This leads to Prob$[a = 1] = 1$ since $\rho = 1$, which results in the excluded special case $A(x) = x$. So for all other cases, $x = 1$ is the only solution inside and on the unit circle of (A.2) and, as a consequence, $Y(1) = 1$.