Open access • Journal Article • DOI:10.1007/S11134-006-9003-8

Priority queueing systems: from probability generating functions to tail probabilities — Source link

Tom Maertens, Joris Walraevens, Herwig Bruneel

Institutions: Ghent University

Published on: 01 Jan 2007 - Queueing Systems (Springer US)

Topics: Layered queueing network, Law of total probability, Mean value analysis, Kendall's notation and Bulk queue

Priority queueing systems: from probability generating functions to tail probabilities

Tom Maertens[*], Joris Walraevens, Herwig Bruneel
Ghent University – UGent
Department of Telecommunications and Information Processing
SMACS Research Group
Sint-Pietersnieuwstraat 41, B-9000 Gent, Belgium
Phone: +32-9-2648901
Fax: +32-9-2644295
E-mail: {tmaerten,jw,hb}@telin.UGent.be

Abstract

Obtaining (tail) probabilities from a transform function is an important topic in queueing theory. To obtain these probabilities in discrete-time queueing systems, we have to invert probability generating functions, since most important distributions in discrete-time queueing systems can be determined in the form of probability generating functions. In this paper, we calculate the tail probabilities of two particular random variables in discrete-time priority queueing systems, by means of the dominant singularity approximation. We show that obtaining these tail probabilities can be a complex task, and that the obtained tail probabilities are not necessarily exponential (as in most 'traditional' queueing systems). Further, we show the impact and significance of the various system parameters on the type of tail behavior. Finally, we compare our approximation results with simulations.

keywords priority queueing systems, tail probabilities, dominant pole, non-exponential behavior

1 Introduction

Many probability distributions of interest in queueing models can be determined in the form of transforms: the Laplace-Stieltjes transforms of *continuous* density functions or the z-transforms of *discrete* probability mass functions. The benefit of using transforms in analyses with stochastic variables has been frequently demonstrated in the past. Transforms are furthermore very useful to extract numerical results, e.g., to calculate moments. However, a seeming disadvantage of working with transforms is that it is not always easy to explicitly calculate the corresponding cumulative distribution functions (cdf's), probability density functions (pdf), or probability mass functions (pmf's).

Often, we are only interested in the tail of the probability distribution. Tail probabilities typically represent the 'exceptional' situations in a queueing system (or more generally, a communication network), of which we want to estimate the frequency of. E.g. the probability that the delay is larger than a given

[*]Corresponding author

value N or the packet loss are examples of interesting performance measures for which the calculation of the (asymptotic behavior of) tail probabilities is usually sufficient. Obtaining tail probabilities of a stochastic variable from its transform, which is basically an *inversion* problem, is thus an important topic in queueing theory. A theoretical solution method is to *analytically* invert the transform, yielding an *explicit*, closed-form expression for the underlying probability distribution. However, this is only possible if the transform expressions are simple enough. In sophisticated queueing models, this method is practically infeasible. Therefore, one has to look for *approximate* solutions.

From our point of view, existing approximate inversion techniques can be roughly divided into two categories: *numerical* inversion methods and *analytical* inversion methods. Abate and Whitt [4] provide an extensive study of various numerical methods for transform inversion. In general, theoretical solutions for the inversion problem can usually be expressed via integrals (see e.g. [9] for a short review): a line integral in the continuous case or a contour integral in the discrete case. These basic inversion integrals can then be calculated by performing a numerical integration. The Fourier-series method numerically integrates a standard inversion integral by means of the trapezoidal rule (see e.g. [4]). In [7] (Laplace transforms) and [5] (probability generating functions), the authors propose Poisson summation formulas which identify the discretization errors associated with this trapezoidal rule. Algorithms based on the Fourier-series method are further a.o. presented in [19] and [21]. Most of these algorithms require the evaluation of the involved transforms at many complex numbers. However, if the transform is only characterised implicitly via a functional equation (e.g., the busy-period distribution in a GI/GI/1 system), it may be quite involved to obtain these values. Abate and Whitt [6] discuss the solution of functional equations for complex arguments, and provide conditions for iterative methods to converge. Variants of these methods can a.o. be found in [9] and [13]. Note finally that the Fourier-series method is closely related to the Laguerre method (see e.g. [2, 3, 16]), i.e., the desired function is in both methods represented as an expansion in terms of *orthogonal functions*, where the coefficients are expressed in terms of the transform.

A second class of approximate inversion techniques exists of analytical methods, which all more or less follow a similar procedure. After determining the (asymptotic) tail behavior, one calculates the corresponding parameters. Finally, approximate expressions for the tail probabilities are derived. The asymptotic tail behavior of a probability distribution can be obtained analytically by calculating the value and type of the *rightmost* singularity of the Laplace transform in the continuous case (see e.g. [8]), or by determining the value and type of the singularity with the smallest modulus of the pgf in the discrete case (see e.g. [10]). Choudhury and Lucantoni [11] showed further that high-order moments of the stochastic variable can be used to estimate the asymptotic parameters of the cdf. Abate et al [1] provide theoretical support for this moment-based algorithm, and present new refined estimators which converge much faster than the estimators proposed in [11]. The techniques in [1] were also used in [22] for computing the asymptotic parameters numerically. When the transforms are not available explicitly, as in models of busy-periods or polling systems, moment-based algorithms prove useful (see e.g. [12]). Finally, when the transforms are only available in matrix-form, one can use an analytical method based on the dominant eigenvalue of the transform-matrix and on the Chernoff large deviations approximation (see e.g. [15]).

In summary, in the analytical approach, an approximate expression is found in terms of a limited number of parameters (such as the dominant singularity of the pgf), while the numerical approach only yield 'a set of numbers'. The analytical method has the advantage that the *behavior* of the pmf is found. On the other hand, numerical inversion techniques are usually more accurate for low arguments of the cdf's or pmf's. The analytical approach thus complements the numerical one.

In this paper, we use the *dominant* singularity method for deriving approximate expressions for tail probabilities of a discrete-time variable from its pgf. E.g. Bruneel et al [10] have shown that for high n, the pmf $x(n)$ of a discrete variable X is dominated by the contribution of the singularity of the corresponding pgf, with the smallest absolute value. This dominant singularity is necessarily positive real and larger than 1. In traditional single-class queueing systems with a FIFO scheduling discipline, the pgf's of the system quantities generally have one type of dominant singularity, usually a simple pole (i.e., a zero of the denominator of the pgf with multiplicity 1) (see e.g. [10]). This leads to the well-known geometric (or exponential) behavior $x(n) \approx K s_*^{-n-1}$, with s_* the dominant pole of the corresponding pgf.

In e.g. priority queueing systems however, several (types of) singularities may play a role (see e.g. [8, 17, 26]). In [17] and [26], the authors analyse "basic" *discrete-time* priority queueing systems: two-class queues with single-slot service times and with a HOL (Head-Of-the-Line) priority scheduling discipline. E.g. in [26], it is shown that two (types of) singularities on the positive real axis of the pgf of the low-priority packet delay in such a priority system play a role: a simple pole and a branch point. This branch point is a result of an implicitly defined function appearing in the pgf. Both singularities can dominate and it depends on the system parameters (arrival rates in that case) which one is dominant. A consequence of the appearance of two types of singularities is that two forms of tail behavior can distinguished be, namely exponential behavior when the simple pole dominates and non-exponential tail behavior when the branch point dominates. Tail behavior in *continuous-time* priority systems has been examined, via analytical methods in e.g. [1, 8, 11, 22], or via numerical methods in e.g. [23]. We finally note that the papers mentioned in this paragraph all assume infinite queue sizes. Different results are however obtained by scaling the number of arrival sources along with the capacity of the system and the queue size (see e.g. [20]).

In this paper, we calculate the tail probabilities of two particular random variables in more complex discrete-time priority queueing systems, whereby more than two (types of) singularities may exist and each of them may (co-)dominate, depending on the values of the various system parameters. We derive expressions for the tail probabilities using the dominant singularity approximation, and we compare our approximations with simulations to validate the used method. The contribution of this paper thus mainly concerns the solution technique that is used, and the extension of this solution technique to rather complicated queueing systems. Hence, also the generated results are a major contribution of the current paper. Specifically, we show that the tail behavior in priority queueing systems can be quite complicated and diverse, highly depending on the values of the system parameters (e.g. arrival and service rates).

The outline of the paper is as follows. In the following section, we describe the tail behavior and derive expressions for the tail probabilities of the system contents of a complex HOL (Head-Of-Line) priority queue. In section 3, we sketch the tail behavior of the delay of a low-priority packet in a HOL-PJ (HOL with Priority Jumps) queue. Some conclusions are drawn in section 4.

2 The total system contents of a HOL priority queue

2.1 Preliminaries

We first concentrate on the total, steady-state system contents of a single-server, two-class priority queueing system with the number of per-slot class-1 and class-2 arrivals characterised by the joint pgf $A(z_1, z_2)$ and the marginal pgf's $A_T(z) = A(z,z)$, $A_1(z) = A(z,1)$ and $A_2(z) = A(1,z)$ (with $\lambda_j = A_j'(1)$ the arrival rate of class j). The waiting room is of infinite capacity and the service times are generally distributed (with $S_j(z)$ the pgf of the service times of class j packets and $\mu_j = S_j'(1)$ the mean service time of a class j packet). Arriving packets are scheduled according to a HOL non-preemptive priority scheduling discipline, where class-1 packets are assumed to have priority over class-2 packets. The pgf of the total system contents has been derived in [25], and is given by

$$U_T(z) = \frac{\left\{\begin{array}{l} (1-\rho_T)(z-1)\Big[S_1(A_T(z))(A_T(z)-1)(z-S_2(A(Y(z),z))) \\ +z\,(A(Y(z),z)-1)(S_2(A_T(z))-S_1(A_T(z)))\Big] \end{array}\right\}}{(z-S_1(A_T(z)))(z-S_2(A(Y(z),z)))(A_T(z)-1)}, \tag{1}$$

where $\rho_T \triangleq \rho_1 + \rho_2 = \lambda_1\mu_1 + \lambda_2\mu_2$ denotes the total load, and where $Y(z)$ is the only solution of $x - S_1(A(x,z)) = 0$ with $|x| < 1$ and $|z| < 1$. The pgf $Y(z)$ is thus implicitly defined as $S_1(A(Y(z),z))$. Note that the total load has to be smaller than 1 (i.e., $\rho_T < 1$) to ensure having a stable system and proper steady-state system distributions. We furthermore assume in the remainder that the pgf's $A_T(z)$, $A_j(z)$ and $S_j(z)\,(j = 1,2)$ and their derivatives go to infinity for z equal to their radii of convergence, or for $z \to \infty$. This includes all 'usual' arrival and service processes, except e.g. processes with a long tail. For the numerical examples and figures in this section, we use a two-dimensional binomial arrival process, with joint pgf $A(z_1, z_2) = (1 - \lambda_1(1-z_1)/N - \lambda_2(1-z_2)/N)^N$ (with $N = 16$), and deterministic or

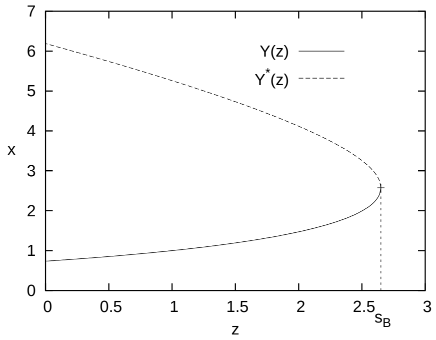


Figure 1: Solutions of $x - S_1(A(x,z)) = 0$

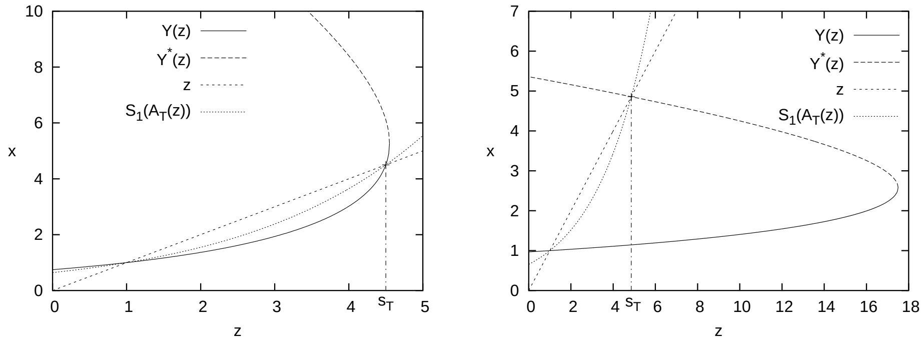
geometric service times, with pgf's $S_j(z) = z^{\mu_j}$ or $S_j(z) = \dfrac{z}{z - \mu_j(z-1)}$ respectively. It should be noted that the figures serve only as an illustration to show that the values of the system parameters play a role in the existence - and thus also in the dominance - of the possible singularities. The derivations in the paper are thus valid for other arrival and service processes as well.

It is obvious that the calculation of the tail probabilities of the total system contents is not straightforward, since it is not a priori clear from expression (1) which singularity of $U_T(z)$ is dominant. Several singularities may play a role, namely the dominant positive real (> 1) zeros of $z - S_1(A_T(z))$ and $z - S_2(A(Y(z), z))$, denoted by s_T and s_L respectively, and the radii of convergence of the pgf's in expression (1). Furthermore, the tail behavior of the total system contents is also influenced by the dominant singularity of the function $Y(z)$, denoted by s_B in the remainder. We first take a closer look at this function on the positive real axis.

2.2 Singularity s_B

First we note that $Y(z)$ is convex on the positive real axis, since $Y(z)$ is a pgf of a stochastic variable (see [24] for a proof). As z increases along the positive real axis, a branch point s_B will be encountered in $Y(z)$ where $Y'(z) \to \infty$ (see e.g. [17] for a similar case). For values of z beyond that point, $Y(z)$ is no longer properly defined (or, in other words, $x - S_1(A(x,z)) = 0$ has no solution for $z > s_B$). However, a second real and positive solution $Y^*(z)$ of the functional equation $x - S_1(A(x,z)) = 0$ for positive real z, exists and decreases as z increases (see Figure 1). E.g. for $z = 1$, it is easily seen that $x - S_1(A_1(x))$ has 2 distinct zeros (i.e., $Y(1) = 1$ and $Y^*(1) > 1$), since $S_1(A_1(x))$ is convex and $\left.\dfrac{dS_1(A_1(x))}{dx}\right|_{x=1} = \rho_1 < 1$. Both solutions coincide for $z = s_B$. s_B is the solution of

$$\begin{cases} Y(s_B) = S_1(A(Y(s_B), s_B)) \\ Y'(s_B) \to \infty \end{cases} \Rightarrow \begin{cases} Y(s_B) - S_1(A(Y(s_B), s_B)) = 0 \\ S_1'(A(Y(s_B), s_B))A^{(1)}(Y(s_B), s_B) = 1 \end{cases}. \tag{2}$$



a. $s_T = Y(s_T)$ b. $s_T = Y^*(s_T)$

Figure 2: The singularity s_T

with $A^{(1)}(z_1, z_2) \triangleq \dfrac{\partial A(z_1, z_2)}{\partial z_1}$. Remark that $Y'(s_B)$ is infinite, but that $Y(s_B)$ remains finite. Applying the results of [14], one can show that in the neighbourhood of s_B, $Y(z)$ is approximately given by

$$Y(z) \approx Y(s_B) - K_Y(s_B - z)^{1/2}, \tag{3}$$

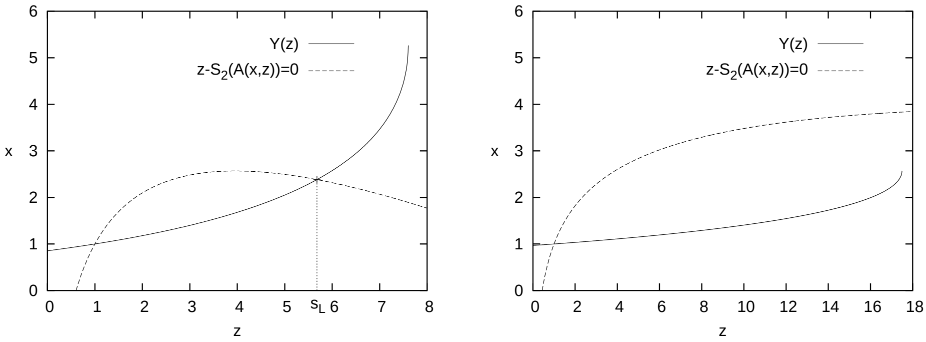
with

$$K_Y = \sqrt{\frac{2A^{(2)}(Y(s_B), s_B)}{S_1''(A(Y(s_B), s_B))(A^{(1)}(Y(s_B), s_B))^3 + A^{(11)}(Y(s_B), s_B))}}, \tag{4}$$

where $A^{(2)}(z_1, z_2) \triangleq \dfrac{\partial A(z_1, z_2)}{\partial z_2}$ and $A^{(11)}(z_1, z_2) \triangleq \dfrac{\partial^2 A(z_1, z_2)}{\partial z_1^2}$. Expression (4) is found by taking the limit $z \to s_B$ in expression (3), and by using the definition of $Y(z)$. Since $Y(z)$ appears in the expression of $U_T(z)$, s_B is also a singularity of $U_T(z)$, and thus plays a role in the tail behavior of the total system contents.

2.3 Singularity s_T

A second potential singularity s_T of $U_T(z)$ on the positive real axis (> 1) is given by the zero of $z - S_1(A_T(z))$. We will show that s_T is however not in all cases a singularity of $U_T(z)$. Since s_T is a zero of $z - S_1(A(z, z))$, it is easily seen that $(x, z) = (s_T, s_T)$ is a solution of $x - S_1(A(x, z)) = 0$. As a consequence, s_T has to be smaller than s_B, since this equation has no solution for $z > s_B$ (see previous subsection). Furthermore, we have shown in the previous paragraph that the equation $x - S_1(A(x, z)) = 0$ has 2 positive real solutions - namely $(Y(z), z)$ and $(Y^*(z), z)$ - for $z < s_B$ positive real. Therefore, $s_T = Y(s_T)$ or $s_T = Y^*(s_T)$, depending on the pgf $S_1(A_T(z))$. Both cases are illustrated in Figure 2, where the functions $Y(z)$, $Y^*(z)$, z and $S_1(A_T(z))$ are shown. Note that $s_T = Y(s_T)$ if $s_T < Y(s_B)$ and $s_T = Y^*(s_T)$ if $s_T > Y(s_B)$, since $Y(z)$ ($Y^*(z)$) is maximal (minimal) in s_B. It is now easily verified that in the case that $s_T = Y(s_T)$, s_T is also a zero of the numerator of $U_T(z)$ (see expression (1)) and that it



a. s_L *exists* b. s_L *does not exist*

Figure 3: The singularity s_L

is thus not a singularity of $U_T(z)$. Note that in this case, it is possible that $s_T \leq 1$ (see subsection 2.5). If $s_T = Y^*(s_T)$ on the other hand, s_T is not a zero of the numerator and is thus indeed a singularity of $U_T(z)$.

So, summarizing, three cases are possible for s_T: $s_T = Y(s_T) < Y(s_B)$, $s_T = Y(s_B)$ (in which case the branch point s_B and s_T coincide, see further) and $s_T = Y^*(s_T) > Y(s_B)$. In the second and third case, s_T is a singularity, with multiplicity 1. Indeed, due to the convexity of $S_1(A_T(z))$, $z - S_1(A_T(z))$ has at most 2 positive real zeros. $z = 1$ is one of them and $z = s_T$ is the other. Since $s_T > 1$ when s_T is a pole, both zeros have multiplicity 1.

2.4 Singularity s_L

A third potential singularity s_L of $U_T(z)$ is given by the (dominant) solution of $z - S_2(A(Y(z), z)) = 0$, on the positive real axis (> 1). Since $S_2(A(Y(z), z))$ is a pgf, and thus a convex function for z positive real, $z - S_2(A(Y(z), z)) = 0$ has at most 2 positive real solutions, i.e., $z = 1$ and $z = s_L$, each with multiplicity 1. Note further that since $Y(z)$ appears in this equation, and since $Y(z)$ does not exist for $z > s_B$, this possible solution s_L has to be smaller than s_B. To characterise this solution s_L, we first transform the equation $z - S_2(A(Y(z), z)) = 0$ into the following system of equations:

$$\begin{cases} z - S_2(A(x, z)) = 0 \\ x = Y(z) \end{cases} . \tag{5}$$

If this system has a solution for $z > 1$ on the real axis, it is $(x, z) = (Y(s_L), s_L)$ (see Figure 3a.). However, this system does not always have a solution, in which case s_L does not exist. This can be seen in Figure 3b., where the function $x = Y(z)$ does not cut the function $z - S_2(A(x, z))$ for $x, z > 1$. In summary, three possible cases can occur for the potential singularity s_L: s_L exists and $s_L < s_B$, s_L exists and $s_L = s_B$ or s_L does not exist.

2.5 Radii of convergence

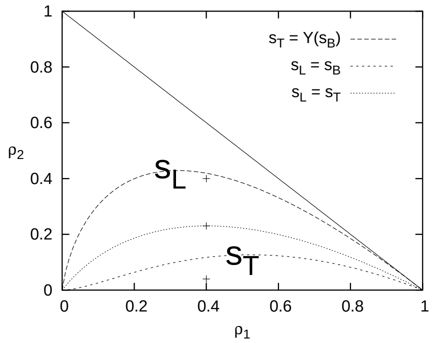
Until now, we have focused on the potential poles of the denominator and on the branch point of the implicitly defined function $Y(z)$. Thereby, we have (implicitly) assumed that all pgf's appearing in expression (1) are analytic in the region of these singularities, i.e., that the radii of convergence of these pgf's do not play a role as possible dominant singularities of $U_T(z)$. In this subsection, we will prove that this is indeed the case for all pgf's appearing in (1), except for one, namely $S_2(A_T(z))$. First, we will show that the radii of convergence of $S_1(A_T(z))$, $A_T(z)$, $S_2(A(Y(z), z))$, $A(Y(z), z)$ are necessarily larger than at least one of the singularities s_B, s_L or s_T.

We first focus on the radius of convergence of $S_1(A_T(z))$. We distinguish two cases: s_T is a singularity or s_T is not a singularity (see subsection 2.3). In the first case, s_T is within the region of convergence of $S_1(A_T(z))$, since $s_T > 1$ is a solution of $z - S_1(A_T(z)) = 0$. In the second case, one can prove that $|z| < |Y(z)|$ for $|z|$ larger than the largest zero of $z - S_1(A_T(z))$ (i.e., 1 or s_T). For those z, $|Y(z)| > |S_1(A_T(z))|$ and as a result, $Y(z)$ will reach its branch point s_B before $S_1(A_T(z))$ diverges. Concluding, the radius of convergence of $S_1(A_T(z))$ is in both cases preceded by another singularity of $U_T(z)$.

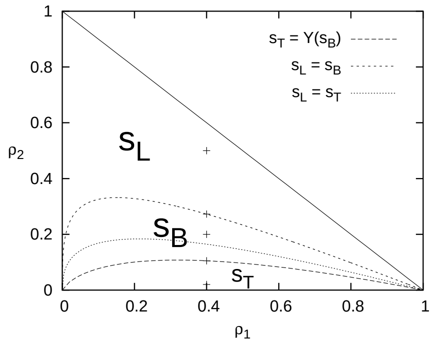
Further, it is easily verified that the radius of convergence of $S_2(A(Y(z), z))$ is not a new potential singularity. Again two cases are distinguished: s_L exists or s_L does not exist (see subsection 2.4). In the first case, $S_2(A(Y(z), z)) < z$ for $1 < z < s_L$ since $\left.\frac{dS_2(A(Y(z), z))}{dz}\right|_{z=1} < 1$, $S_2(A(Y(s_L), s_L)) = s_L$, and $S_2(A(Y(z), z))$ is a convex function. As a result, the radius of convergence of $S_2(A(Y(z), z))$ is larger than s_L. In the second case, $S_2(A(Y(z), z)) < z$ for $1 < z < s_B$ since $\left.\frac{dS_2(A(Y(z), z))}{dz}\right|_{z=1} < 1$ and $Y(z)$ reaches its branch point before $S_2(A(Y(z), z))$ reaches z. As a result, s_B is the radius of convergence of $S_2(A(Y(z), z))$, and we have already treated this singularity in subsection 2.2.

Thirdly, since the service times have means bigger than or equal to 1, $S_1(A_T(z)) \geq A_T(z)$ and $S_2(A(Y(z), z)) \geq A(Y(z), z)$, for z larger than 1 and positive real. As a result, the radii of convergence of $A_T(z)$ and $A(Y(z), z)$ are larger (or equal) than the radii of convergence of $S_1(A_T(z))$ and $S_2(A(Y(z), z))$ respectively, and thus do not play a role (a fortiori) in the tail behavior of $U_T(z)$ (or do not yield new potential singularities).

Finally, as already mentioned, the radius of convergence of $S_2(A_T(z))$, denoted by s_Q, can be the dominant singularity of $U_T(z)$. The reason why this singularity can be dominant is that $S_2(z)$ does not influence the singularities s_B and s_T, as can be seen from subsections 2.2 and 2.3 respectively. Hence, $S_2(z)$ can be such that $S_2(A_T(z)) \to \infty$ before z reaches s_B or s_T. Furthermore, $S_2(z)$ can be such that $S_2(A_T(z)) > S_2(A(Y(z), z))$ for $z > 1$, and even so that for z increasing, $S_2(A_T(z))$ reaches its radius of convergence s_Q before $U_T(z)$ reaches s_L. As a result, s_Q can be smaller than s_B, s_T and s_L, and thus be the dominant singularity of $U_T(z)$. We will give an example of such $S_2(z)$ in the following subsection.



a. *when* $\mu_1 = 2$, $\mu_2 = 4$



b. *when* $\mu_1 = 4$, $\mu_2 = 2$

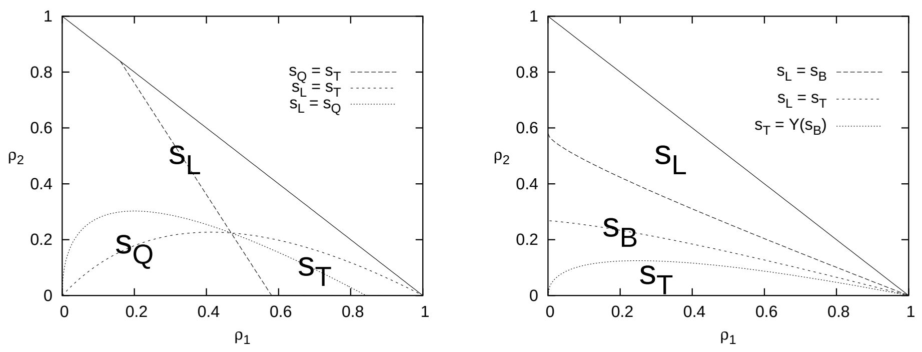
Figure 4: Tail behavior of $U_T(z)$ as a function of ρ_1 and ρ_2, for deterministic service times

2.6 Bringing everything together

Summarizing, the tail behavior of the total system contents is characterised by s_B, s_T, s_L or s_Q, depending on which singularity is dominant. In case of a two-dimensional binomial arrival process and deterministic service times (with pgf $S_j(z) = z^{\mu_j}$), the curves in Figures 4a. and 4b. show for which combination of class-1 and class-2 loads, $s_T = Y(s_B)$ (i.e., s_T and s_B coincide), $s_L = s_B$ and $s_L = s_T$, when $\mu_1 = 2$, $\mu_2 = 4$ and $\mu_1 = 4$, $\mu_2 = 2$ respectively. Note that s_Q does not play a role in the tail behavior of $U_T(z)$ in this case, since the radius of convergence of $S_2(A_T(z)) = (A_T(z))^{\mu_2}$ is infinite. The curve $s_T = Y(s_B)$ represents the following: above the curve $s_T = Y(s_T) < Y(s_B)$, while below the curve $s_T = Y^*(s_T) > Y(s_B)$. Or, in other words, below this curve, s_T is a singularity, while above the curve, s_T is not a singularity (see subsection 2.3). The curve $s_L = s_B$ can be interpreted in a similar way: below this curve, s_L does not exist, while above the curve, it does. Finally, below the curve $s_L = s_T$, $s_T < s_L$, while above the curve, $s_T > s_L$. In the area above the linear line (defined by $\rho_1 + \rho_2 = 1$), the total load is larger than 1, and as a result, the system becomes unstable.

It can easily be seen that these curves split the (ρ_1, ρ_2)-space in several regions. In each region, one particular singularity is dominant, depending on which singularity (of those who exist in that region) has the smallest value. On the curves, singularities coincide. By observing the values of the several existing singularities in a certain region, one can easily determine which singularity is dominant in that particalur region. For instance, when $\mu_1 = 2$ and $\mu_2 = 4$, only two singularities play a role in the tail behavior: s_T and s_L (see Figure 4a.). s_B is thus never dominant in this case. When $\mu_1 = 4$ and $\mu_2 = 2$, the singularity s_B can also be dominant (see Figure 4b.). So, the values of all system parameters play a role in the dominance of the possible singularities and thus in the tail behavior of the total system contents in this HOL priority queue.

Obviously the distributions of the system variables also play a distinctive role. In Figure 5, we show for which combination of class-1 and class-2 loads the important singularities of $U_T(z)$ coincide, for the same



a. *when* $\mu_1 = 2$, $\mu_2 = 4$ b. *when* $\mu_1 = 4$, $\mu_2 = 2$

Figure 5: Tail behavior of $U_T(z)$ as a function of ρ_1 and ρ_2, for geometric service times

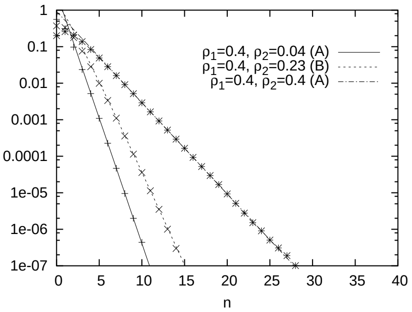
two-dimensional binomial arrival process but for geometric service times with pgf's $S_j(z) = \dfrac{z}{z - \mu_j(z-1)}$ $(j = 1, 2)$, with $\mu_1 = 2$, $\mu_2 = 4$ (Figure 5a.) and $\mu_1 = 4$, $\mu_2 = 2$ (Figure 5b.). We observe that the (ρ_1, ρ_2)-space again is split in several regions. s_T, s_L, or s_Q determine the tail behavior of $U_T(z)$ when $\mu_1 = 2$, $\mu_2 = 4$ (see Figure 5a.), while $U_T(z)$ is dominated by s_T, s_L, or s_B when $\mu_1 = 4$, $\mu_2 = 2$ (see Figure 5b.). When the service times are geometrically distributed, the radius of convergence of $S_2(A_T(z))$ may thus determine the tail behavior of the total system contents.

2.7 The behavior of $U_T(z)$ in its dominant singularity

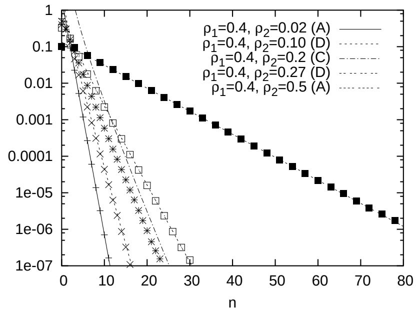
The type of the (co-)dominant singularity has a large impact on the tail behavior (see e.g. [8, 17]). In this subsection, we use s_* as a general notation for the dominant singularity of $U_T(z)$. According to the previous subsection, three possible cases are established for s_*: $s_* \neq s_B$, $s_* = s_B$ and s_B is single-dominant, or $s_* = s_B$ but other singularities are co-dominant. In the remainder, we formulate a procedure to approximate $U_T(z)$ in the neighbourhood of s_*, for each case. We refer to Appendix B for applications of this procedure.

In the first case, the branch point s_B is *not dominant* and a 'regular' pole is the dominant singularity. We replace each factor of $U_T(z)$ by its n^{th}-order Taylor-series approximation in s_*, with n the multiplicity of s_* as zero of that factor. This leads to $U_T(z) \approx \dfrac{K_T^{(*)}}{(s_* - z)^m}$ for $z \to s_*$, with $K_T^{(*)}$ a constant and with m the multiplicity of the dominant singularity.

In the second case, in which the branch point s_B is the only dominant singularity, we first substitute expression (3) of $Y(z)$ in (1). Secondly, the obtained expression is rationalised, i.e., all roots are removed from the denominator. We furthermore replace each factor of the denominator by the 0^{th}-order Taylor-series approximation in s_B (since s_B is not a zero of the denominator). We then get an expression for $U_T(z)$ of the form $U_T(s_B) - K_T^{(*)}(s_B - z)^{1/2} - K_T^{(**)}(s_B - z)$ in the neighourhood of s_B. Since the last term *tends to zero faster* than the second term, we can omit the last term, yielding $U_T(z) \approx U_T(s_B) - K_T^{(*)}(s_B - z)^{1/2}$



a. $\mu_1 = 2$, $\mu_2 = 4$



b. $\mu_1 = 4$, $\mu_2 = 2$

Figure 6: Tail probabilities of the total system contents for some (ρ_1, ρ_2)-combinations

for $z \to s_B$.

Finally, in the third case, the branch point s_B is dominant together with other singularities. First, we replace each factor of (1) in which $Y(z)$ does not appear by its n^{th}-order Taylor-series approximation in s_B. Secondly, we substitute $Y(z)$ by its approximate expression (3). Finally, we separately look in the numerator and the denominator for the term that tends to zero the slowest. This eventually leads to $U_T(z) \approx \frac{K_T^{(*)}}{(s_B - z)^{m/2}}$, with m an integer.

2.8 Obtaining expressions for the tail probabilities

In this subsection, we will focus on the special case of deterministic service times. Other distributions of service times can be treated in a similar way. For each possible combination of dominant singularities appearing in Figure 4, we can approximate $U_T(z)$ in the neighbourhood of its dominant singularity by using the procedure in the previous subsection. We encounter 4 different tail behaviors for $U_T(z)$ in this case, depending on which type of singularity dominates: a simple pole (behavior A), a pole with multiplicity 2 (behavior B), a branch point (behavior C) or a simple pole coexisting with a branch point (behavior D). Using Darboux's theorem (see Appendix A), we finally find the tail probabilities for these 4 different cases:

$$u_T(n) \quad \triangleq \quad \text{Prob}[u_T = n] \approx \begin{cases} K_T^* s_*^{-n-1} & \text{behavior A} \\ K_T^*(n+1) s_*^{-n-2} & \text{behavior B} \\ \dfrac{K_T^* n^{-3/2} s_*^{-n}}{2\sqrt{\pi/s_*}} & \text{behavior C} \\ \dfrac{K_T^* n^{-1/2} s_*^{-n}}{\sqrt{\pi s_*}} & \text{behavior D} \end{cases}, \tag{6}$$

with s_* a general notation for the dominant singularity and $K_T^{(*)}$ easily obtained according to the procedure described in the previous subsection. Behavior A constitutes a typical *geometric* (exponential) behavior - as encountered in many other queueing studies - while the others are *non-geometric*.

Figures 6a. and 6b. show the tail probabilities of the total system contents for the (ρ_1, ρ_2)-combinations indicated by the marks in Figures 4a. and 4b. respectively, when $\mu_1 = 2$ and $\mu_2 = 4$ and $\mu_1 = 4$ and $\mu_2 = 2$. Note that the capital letter next to a (ρ_1, ρ_2)-combination in the legend of Figures 6a. and 6b., indicates the type of behavior (A, B, C or D) to which that particular combination belongs. Further, we have compared our approximations with simulation results (marks in Figures 6a. and 6b.). The figures show that the accuracy of the exponential (behavior A) asympotic approximations is excellent, while the non-exponential approximations are not as accurate as the exponential ones. The minor accuracy of the non-exponential approximations is attributed to the slower rates of convergence in the corresponding expressions (see e.g. [8]).

3 The delay of a low-priority packet in a HOL-PJ queue

3.1 Preliminaries

As a second example of a priority queueing system quantity with a complex tail behavior, we consider the delay of a low-priority packet in a single-server, two-class HOL-PJ (HOL priority with priority jumps) queue. The waiting room is considered infinite and the arrival process is the same as in the first example, i.e., a joint pgf $A(z_1, z_2)$ for the numbers of arrivals of both classes, and marginal pgf's $A_T(z)\,(= A(z,z))$, $A_1(z)\,(= A(z,1))$ and $A_2(z)\,(= A(1,z))$ for the total number, the number of class-1 and the number of class-2 arrivals respectively (with λ_j the arrival rate of class j, and $\lambda_T = \lambda_1 + \lambda_2$ the total arrival rate). Note that in most figures in this section, we again use a two-dimensional binomial arrival process, with joint pgf $A(z_1, z_2) = (1 - \lambda_1(1 - z_1)/N - \lambda_2(1 - z_2)/N)^N$ (where $N = 16$). The service times are deterministically distributed, and equal to 1 slot. Furthermore, the system is influenced by a jumping process: the class-2 packets, which are initially stored in the low-priority queue, jump at the end of each slot with a probability β to the high-priority queue, in which arriving class-1 packets are queued. Packets in the high-priority queue, which are thus of class 1 or class 2, have obviously a higher priority than the packets in the low-priority queue. In other words, only when the high-priority queue is empty, packets of the low-priority queue can be served. This priority queueing system has been analyzed in [18] and the pgf of the class-2 packet delay is found to be

$$D_2(z) = \frac{\left\{ \begin{array}{l} (1-\lambda_T)z\Big[\beta(A_T(z) - A_1(z))(V_0(z) - A_T(V_0(z))) \\ +(1-\beta)(A_T(V_0(z)) - A_1(V_0(z)))(1 - A_1(z))(z - A_T(z))\Big] \end{array} \right\}}{\lambda_2(z - A_T(z))(1 - (1-\beta)A_1(z))(V_0(z) - A_T(V_0(z)))}. \tag{7}$$

The function $V_0(z)$ is a solution of $x - (1-\beta)zA_1(x) = 0$, $|x| < 1$ and $|z| < 1$, and is thus implicitly given by $(1-\beta)zA_1(V_0(z))$.

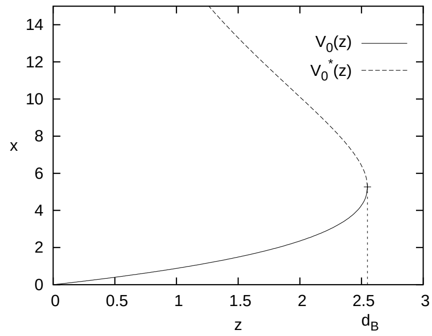


Figure 7: Solutions of $x - (1 - \beta)zA_1(x) = 0$

As for $U_T(z)$ in the previous section, it is not a priori clear what the dominant singularity is of $D_2(z)$. This is, in the first place, due to the occurrence of the function $V_0(z)$, which is only implicitly defined and which shows a similar behavior as $Y(z)$ (see subsection 2.2).

3.2 Singularity d_B

Specifically, in Figure 7, we see that the functional equation $x-(1-\beta)zA_1(x) = 0$ has two positive real solutions, namely $V_0(z)$ and $V_0^*(z)$ - where $V_0(z)$ is a strictly increasing and $V_0^*(z)$ a strictly decreasing function - and that those two solutions coincide for $z = d_B$. The figure shows further that $V_0(z)$ and $V_0^*(z)$ are no longer properly defined for values of z beyond d_B ($z > d_B$). Since $V_0(z)$ remains finite and $V_0'(z) \to \infty$ in d_B, we can easily determine the *branch point* d_B:

$$\begin{cases} V_0(d_B) = (1-\beta)d_B A_1(V_0(d_B)) \\ V_0'(d_B) \to \infty \end{cases} \Rightarrow \begin{cases} V_0(d_B) - (1-\beta)d_B A_1(V_0(d_B)) = 0 \\ (1-\beta)d_B A_1'(V_0(d_B)) = 1 \end{cases}. \tag{8}$$

In the same way as $Y(z)$, $V_0(z)$ is then approximated by

$$V_0(z) \approx V_0(d_B) - K_V(d_B - z)^{1/2}, \tag{9}$$

where K_V can be found by substituting $z = d_B$ in the latter expression, and by using the definition of $V_0(z)$:

$$K_V = \sqrt{\frac{2A_1(V_0(d_B))}{d_B A_1''(V_0(d_B))}}. \tag{10}$$

3.3 Singularity d_T

A second potential singularity d_T of $D_2(z)$ on the real positive axis is given by the zero of $z - A_T(z)$ larger than 1 (see Figure 8).

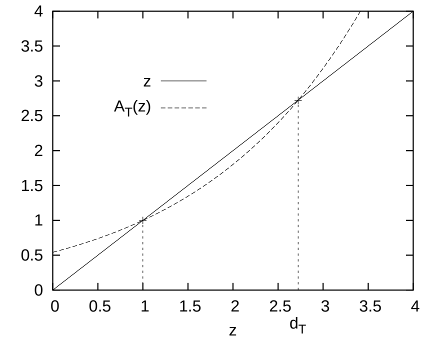
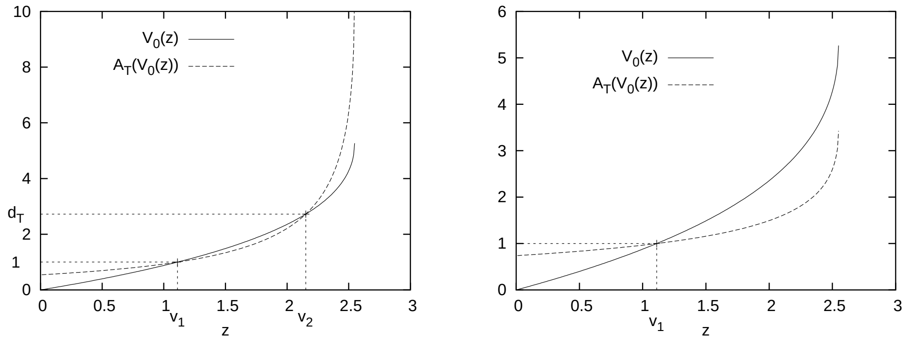


Figure 8: The functions z and $A_T(z)$ for z real and positive



a. d_V *exists* b. d_V *does not exist*

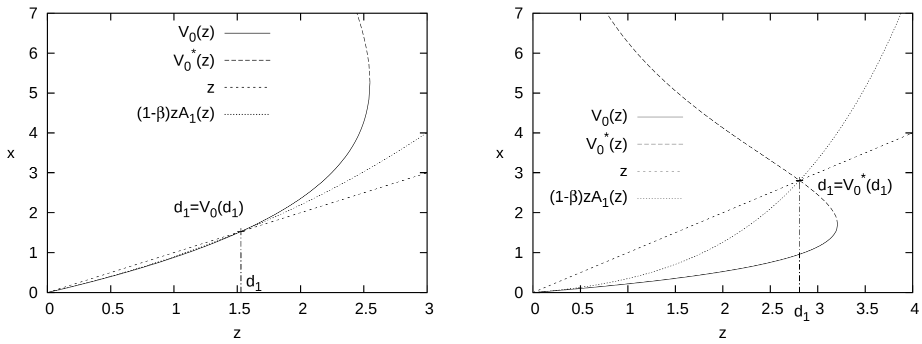
Figure 9: The singularity d_V

3.4 Singularity v_2

Thirdly, we look at the zeros of $V_0(z) - A_T(V_0(z))$, which may be singularities of (7), since $V_0(z) - A_T(V_0(z))$ is a factor of the denominator. We first rewrite $V_0(z) - A_T(V_0(z))$ as the following system of equations:

$$\begin{cases} x - A_T(x) = 0 \\ x = V_0(z) \end{cases} . \tag{11}$$

The equation $x - A_T(x) = 0$ has two positive real solutions, namely $x = 1$ and $x = d_T$ (see subsection 3.3). So $V_0(z) - A_T(V_0(z))$ may have two real positive solutions v_1 and v_2, satisfying $V_0(v_1) = 1$ and $V_0(v_2) = d_T$ respectively (see Figure 9a.). However, v_1 is never a singularity of $D_2(z)$ since the numerator of (7) is also zero for $V_0(v_1) = 1$. Secondly, v_2 does not always exist, since $V_0(z)$ ceases to exist for $z > d_B$, and thus the second solution is not always 'reached' before d_B (see Figure 9a.). Whether the singularity v_2 exists or not, depends on the values of all system parameters: the arrival process and the jumping probability β. When v_2 exists, $v_2 = \dfrac{d_T}{(1-\beta)A_1(d_T)}$, which is easily checked by substituting $V_0(z)$ by d_T in the definition of $V_0(z)$. So, in summary, three cases can occur for the potential singularity v_2: v_2 exists



a. $d_1 = V_0(d_1)$ b. $d_1 = V_0^*(d_1)$

Figure 10: The singularity d_1

and $v_2 < d_B$, v_2 exists and $v_2 = d_B$ or v_2 does not exist.

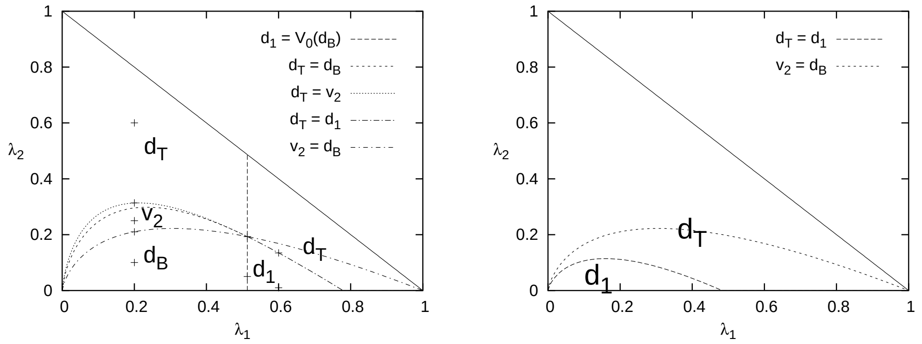
3.5 Singularity d_1

A fourth potential singularity of $D_2(z)$ on the real positive axis (> 1) - denoted by d_1 - is given by the zero of $1 - (1 - \beta)A_1(z)$. d_1 is however not always a singularity, as we will show in the remainder of this subsection. It is easily seen that $(x, z) = (d_1, d_1)$ is a solution of $x - (1 - \beta)zA_1(x) = 0$. This equation, which has been discussed in subsection 3.2, has no solutions for $z > d_B$, positive real. Hence, d_1 has to be smaller than d_B. For $z < d_B$, this equation has two positive real solutions (x, z), namely $(V_0(z), z)$ and $(V_0^*(z), z)$. Consequently, $d_1 = V_0(d_1)$ or $d_1 = V_0^*(d_1)$ (see Figure 10). We can now verify that when $d_1 = V_0(d_1)$, d_1 is also a zero of the numerator of $D_2(z)$ (see expression (7)), and thus not a singularity of $D_2(z)$. On the other hand, when $d_1 = V_0^*(d_1)$, d_1 is not a zero of the numerator, and is thus a singularity of $D_2(z)$. To conclude this subsection, we state the three possible cases for the potential singularity d_1: $d_1 = V_0(d_1) < V_0(d_B)$, $d_1 = V_0(d_B)$ (in which case the branch point d_B and d_1 coincide) and $d_1 = V_0^*(d_1) > V_0(d_B)$. d_1 is a singularity in the second and third case.

3.6 Determining the tail probabilities

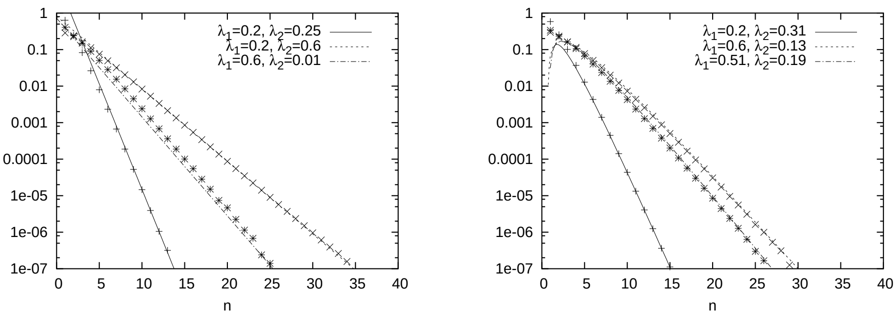
First note that it can be proven that in this case the radii of convergence of the generating functions appearing in (7) are never dominant. We can thus bring everything together: the singularities d_B, d_T, v_2 or d_1 - depending on which one is dominant - characterize the tail behavior of the class-2 delay. Their mutual behavior, illustrated in Figures 11a. and 11b. for a two-dimensional binomial arrival process, and for $\beta = 0.4$ and $\beta = 0.75$ respectively, can be determined in a similar way as in subsection 2.6: first we calculate for which combinations of class-1 and class-2 loads singularities coincide, and then for each region we determine which singularity is dominant.

Remark that in the area above the linear line in Figures 11a. and 11b. (defined by $\lambda_1 + \lambda_2 = 1$), the total



a. *when* $\beta = 0.4$ b. *when* $\beta = 0.75$

Figure 11: Tail behavior of $D_2(z)$ as a function of the arrival rates of both classes



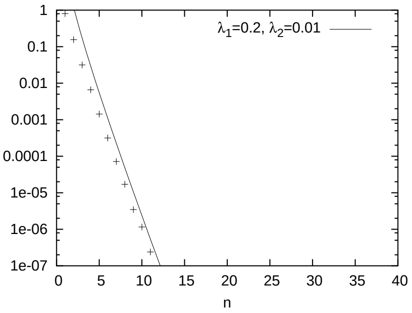
a. *behavior A* b. *behavior B*

Figure 12: Tail probabilities of the class-2 delay for some (λ_1, λ_2)-combinations (1)

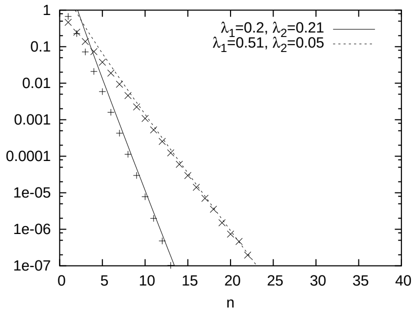
arrival rate is larger than 1, which results in an unstable system for these (λ_1, λ_2)-combinations. Figure 11a. shows that, when $\beta = 0.4$, all four singularities play a role in the tail behavior of $D_2(z)$. When $\beta = 0.75$ on the other hand, only two singularities play a role: d_T and d_1 (see Figure 11b.). These figures illustrate that again all system parameters - the arrival rates λ_1 and λ_2 and the jumping probability β - influence the existence and dominance of the possible singularities, and thus have an impact on the tail behavior of the class-2 delay in this HOL-PJ queue.

We can then formulate a similar procedure as in subsection 2.7 for approximating $D_2(z)$ in the neighbourhood of its dominant singularity, necessary for the calculation of the tail probabilities. In this way, we again encounter 4 distinct types of tail behavior.

Finally, by using Darboux's theorem on the approximations of $D_2(z)$ in their dominant singularities, the tail probabilities are calculated. Figures 12a., 12b., 13a. and 13b. show the tail probabilities of the class-2 delay for the (λ_1, λ_2)-combinations indicated by the marks in Figure 11a.. Note that the tail behavior of $D_2(z)$ depends on the type of the dominant singularity: a simple pole (behavior A), a pole of multiplicity 2 (behavior B), a branch point (behavior C) or a simple pole coexisting with a branch point (behavior D). The figures make clear that the approximate tail probabilities, compared to simulation results, are again more than satisfactory.



a. *behavior C*



b. *behavior D*

Figure 13: Tail probabilities of the class-2 delay for some (λ_1, λ_2)-combinations (2)

4 Conclusions

In this paper, we have analyzed the tail behavior and derived approximate expressions for tail probabilities of two particular variables in priority queueing systems, by means of the dominant singularity approximation. We have shown that several singularities may play a role. Indeed, depending on the values of the various system parameters, several singularities exist. Since each of them can be dominant, they all determine the tail probabilities. We have furthermore shown that the number of singularities varies from model to model. This makes studying the tail behavior much more complicated than in traditional queueing models, and even more complicated than in "basic" priority queueing models. Furthermore, we have proved that once the value of the dominant singularity is calculated, expressions of the tail probabilities are easy to evaluate, which makes the dominant singularity approximation extremely suitable to study these queueing systems. We have also compared our approximations with simulations, and the obtained results are excellent. Altogether, this makes the dominant singularity method a very powerful technique for deriving approximate expressions for tail probabilities in quite complicated queueing models.

Appendix A: Darboux's theorem

Theorem 1.1 *Suppose $X(z) = \sum_{n=0}^{\infty} x(n)z^n$ with positive real coefficients $x(n)$ is analytic near 0 and has only algebraic singularities α_k on its circle of convergence $|z| = R$, in other words, in a neighbourhood of α_k we have*

$$X(z) \sim (1 - \frac{z}{\alpha_k})^{-\omega_k} G_k(z), \tag{12}$$

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

$$x(n) = \sum_j \frac{G_j(\alpha_j)}{\Gamma(\omega_j)} n^{\omega_j - 1} \alpha_j^{-n} + o(n^{\omega-1} R^{-n}), \tag{13}$$

with the sum taken over all j with $Re(\omega_j) = \omega$ and $\Gamma(\omega)$ the Gamma-function of ω (with $\Gamma(n) = (n-1)!$ for n discrete).

Appendix B: Examples of approximating $U_T(z)$ in the neighbourhood of its dominant singularity

In this appendix, we give 4 examples, in which the tail behavior of $U_T(z)$ is determined applying the procedure described in subsection 2.7. In the first example, we consider the singularity s_T to be the only dominant singularity. Replacing factor $z - S_1(A_T(z))$ by its first-order Taylor-series approximation (since s_T is a zero of this factor with multiplicity 1) and the remaining factors by their 0^{th}-order Taylor-series approximation in expression (1), yields

$$U_T(z) \quad \approx \quad \frac{\left\{ \begin{array}{l} (1-\rho_T)(s_T - 1)s_T \Big[(A_T(s_T) - 1)(s_T - S_2(A(Y(s_T), s_T))) \\ +(A(Y(s_T), s_T) - 1)(S_2(A_T(s_T)) - S_1(A_T(s_T))) \Big] \end{array} \right\}}{(s_T - S_2(A(Y(s_T), s_T)))(A_T(s_T) - 1)(S_1'(A_T(s_T))A_T'(s_T) - 1)(s_T - z)},$$

or thus $U_T(z) \approx \dfrac{K_T^{(*)}}{(s_T - z)}$ for $z \to s_T$ (with $K_T^{(*)}$ easily obtained from the latter expression).

In the second example, s_T is co-dominant with s_L. Following the same procedure as in the previous example, we find

$$U_T(z) \quad \approx \quad \frac{(1-\rho_T)(s_L - 1)s_L(A(Y(s_L), s_L) - 1)(S_2(A_T(s_L)) - s_L)}{\left\{ \begin{array}{l} (A_T(s_L) - 1)(S_1'(A_T(s_L))A_T'(s_L) - 1)(s_L - z) \\ \times \Big(S_2'(A(Y(s_L), s_L) \left[A^{(1)}(Y(s_L), s_L)Y'(s_L) + A^{(2)}(Y(s_L), s_L) \right] - 1 \Big)(s_L - z) \end{array} \right\}},$$

for $z \to s_L = s_T$. This leads to $U_T(z) \approx \dfrac{K_T^{(*)}}{(s_L - z)^2}$ in the neighbourhood of $s_L = s_T$.

In the third example, the branch point s_B is the only dominant singularity of $U_T(z)$. Substituting expression (3) in (1) first produces

$$U_T(z) \approx \frac{\left\{\begin{array}{l} (1-\rho_T)(z-1)\Big[S_1(A_T(z))(A_T(z)-1)\Big(z-S_2(A(Y(s_B),s_B)) \\ +K_Y S_2'(A(Y(s_B),s_B))A^{(1)}(Y(s_B),s_B)(s_B-z)^{1/2}\Big) \\ +z\,(S_2(A_T(z))-S_1(A_T(z))) \\ \times\Big(A(Y(s_B),s_B)-K_Y A^{(1)}(Y(s_B),s_B)(s_B-z)^{1/2}-1\Big)\Big] \end{array}\right\}}{\left\{\begin{array}{l} (z-S_1(A_T(z)))(A_T(z)-1)\Big((z-S_2(A(Y(s_B),s_B))) \\ +K_Y S_2'(A(Y(s_B),s_B))A^{(1)}(Y(s_B),s_B)(s_B-z)^{1/2}\Big) \end{array}\right\}},$$

and after rationalising, we obtain

$$U_T(z) \approx \frac{\left\{\begin{array}{l} (1-\rho_T)(z-1)\Big[S_1(A_T(z))(A_T(z)-1)\Big(z-S_2(A(Y(s_B),s_B)) \\ +K_Y S_2'(A(Y(s_B),s_B))A^{(1)}(Y(s_B),s_B)(s_B-z)^{1/2}\Big) \\ +z\,(S_2(A_T(z))-S_1(A_T(z))) \\ \times\Big(A(Y(s_B),s_B)-K_Y A^{(1)}(Y(s_B),s_B)(s_B-z)^{1/2}-1\Big)\Big] \\ \times\Big(z-S_2(A(Y(s_B),s_B)) \\ -K_Y S_2'(A(Y(s_B),s_B))A^{(1)}(Y(s_B),s_B)(s_B-z)^{1/2}\Big) \end{array}\right\}}{\left\{\begin{array}{l} (z-S_1(A_T(z)))(A_T(z)-1)\Big((z-S_2(A(Y(s_B),s_B)))^2 \\ -K_Y^2 S_2'(A(Y(s_B),s_B))^2 A^{(1)}(Y(s_B),s_B)^2(s_B-z)\Big) \end{array}\right\}}.$$

We furthermore replace each factor of the denominator by the 0^{th}-order Taylor-series approximation in s_B, yielding

$$U_T(z) \approx \frac{\left\{\begin{array}{l} (1-\rho_T)(z-1)\Big[S_1(A_T(z))(A_T(z)-1)\Big(z-S_2(A(Y(s_B),s_B)) \\ +K_Y S_2'(A(Y(s_B),s_B))A^{(1)}(Y(s_B),s_B)(s_B-z)^{1/2}\Big) \\ +z\,(S_2(A_T(z))-S_1(A_T(z))) \\ \times\Big(A(Y(s_B),s_B)-K_Y A^{(1)}(Y(s_B),s_B)(s_B-z)^{1/2}-1\Big)\Big] \\ \times\Big(z-S_2(A(Y(s_B),s_B)) \\ -K_Y S_2'(A(Y(s_B),s_B))A^{(1)}(Y(s_B),s_B)(s_B-z)^{1/2}\Big) \end{array}\right\}}{\left\{\ (s_B-S_1(A_T(s_B)))(A_T(s_B)-1)(s_B-S_2(A(Y(s_B),s_B)))^2\ \right\}}.$$

We can now put $U_T(s_B)$ in front:

$$U_T(z) \approx \frac{\left\{ \begin{array}{l} (1-\rho_T)(s_B-1)\Big[S_1(A_T(s_B))(A_T(s_B)-1)(s_B - S_2(A(Y(s_B),s_B))) \\ +s_B(A(Y(s_B),s_B)-1)(S_2(A_T(s_B))-S_1(A_T(s_B)))\Big] \end{array} \right\}}{\left\{ (s_B - S_1(A_T(s_B)))(A_T(s_B)-1)(s_B - S_2(A(Y(s_B),s_B))) \right\}}$$
$$- \frac{\left\{ \begin{array}{l} (1-\rho_T)K_Y(s_B-1)s_B(S_2(A_T(s_B))-S_1(A_T(s_B)))A^{(1)}(Y(s_B),s_B) \\ \times\Big(s_B - S_2(A(Y(s_B),s_B)) + (A(Y(s_B),s_B)-1)S_2'(A(Y(s_B),s_B))\Big) \end{array} \right\}(s_B-z)^{1/2}}{(s_B - S_1(A_T(s_B)))(A_T(s_B)-1)(s_B - S_2(A(Y(s_B),s_B)))^2}$$
$$- \frac{\left\{ \begin{array}{l} (1-\rho_T)K_Y^2(s_B-1)S_2'(A(Y(s_B),s_B))A^{(1)}(Y(s_B),s_B)^2 \\ \times\Big(S_1(A_T(s_B))(A_T(s_B)-1)S_2'(A(Y(s_B),s_B)) - s_B(S_2(A_T(s_B))-S_1(A_T(s_B)))\Big) \end{array} \right\}(s_B-z)}{(s_B - S_1(A_T(s_B)))(A_T(s_B)-1)(s_B - S_2(A(Y(s_B),s_B)))^2}.$$

This finally leads to $U_T(z) \approx U_T(s_B) - K_T^{(*)}(s_B-z)^{1/2}$ for $z \to s_B$, since the last term tends faster to zero than the second term.

Finally, in the last example, the branch point s_B is co-dominant with s_L. We first replace the factors in which $Y(z)$ does not appear by their 0^{th}-order Taylor-series approximation in s_B, and then substitute expression (3) in (1). This produces

$$U_T(z) \quad \approx \quad \frac{\left\{ \begin{array}{l} (1-\rho_T)(s_B-1)\Big[S_1(A_T(s_B))(A_T(s_B)-1)\Big(z - S_2(A(Y(s_B),s_B)) \\ +K_Y S_2'(A(Y(s_B),s_B))A^{(1)}(Y(s_B),s_B)(s_B-z)^{1/2}\Big) \\ +s_B(S_2(A_T(s_B))-S_1(A_T(s_B))) \\ \times\Big(A(Y(s_B),s_B) - K_Y A^{(1)}(Y(s_B),s_B)(s_B-z)^{1/2} - 1\Big)\Big] \end{array} \right\}}{\left\{ \begin{array}{l} (s_B - S_1(A_T(s_B)))(A_T(s_B)-1)\Big((z - S_2(A(Y(s_B),s_B))) \\ +K_Y S_2'(A(Y(s_B),s_B))A^{(1)}(Y(s_B),s_B)(s_B-z)^{1/2}\Big) \end{array} \right\}}.$$

Reorganising the latter expression yields

$$
\begin{aligned}
U_T(z) \approx\ & \frac{\Big\{\ (1-\rho_T)(s_B-1)s_B(A(Y(s_B),s_B)-1)(S_2(A_T(s_B)))-S_1(A_T(s_B)))\ \Big\}}{\left\{\begin{array}{l}(s_B-S_1(A_T(s_B)))(A_T(s_B)-1)(z-S_2(A(Y(s_B),s_B)))\\ +(s_B-S_1(A_T(s_B)))(A_T(s_B)-1)K_Y S_2'(A(Y(s_B),s_B))A^{(1)}(Y(s_B),s_B)(s_B-z)^{1/2}\Big)\end{array}\right\}} \\
& + \frac{\left\{\begin{array}{l}(1-\rho_T)(s_B-1)K_Y A^{(1)}(Y(s_B),s_B)(s_B-z)^{1/2}\\ \times\Big[S_1(A_T(s_B))(A_T(s_B)-1)S_2'(A(Y(s_B),s_B))+s_B(S_2(A_T(s_B))-S_1(A_T(s_B)))\Big]\end{array}\right\}}{\left\{\begin{array}{l}(s_B-S_1(A_T(s_B)))(A_T(s_B)-1)(z-S_2(A(Y(s_B),s_B)))\\ +(s_B-S_1(A_T(s_B)))(A_T(s_B)-1)K_Y S_2'(A(Y(s_B),s_B))A^{(1)}(Y(s_B),s_B)(s_B-z)^{1/2}\Big)\end{array}\right\}} \\
& + \frac{\Big\{\ (1-\rho_T)(s_B-1)S_1(A_T(s_B))(A_T(s_B)-1)(z-S_2(A(Y(s_B),s_B)))\ \Big\}}{\left\{\begin{array}{l}(s_B-S_1(A_T(s_B)))(A_T(s_B)-1)(z-S_2(A(Y(s_B),s_B)))\\ +(s_B-S_1(A_T(s_B)))(A_T(s_B)-1)K_Y S_2'(A(Y(s_B),s_B))A^{(1)}(Y(s_B),s_B)(s_B-z)^{1/2}\Big)\end{array}\right\}}.
\end{aligned}
$$

Separately analysing the numerator and the denominator, and in both keeping the term that tends to zero the slowest, finally approximates $U_T(z)$ by $\dfrac{K_T^{(*)}}{(s_B-z)^{1/2}}$ in the neighbourhood of s_B.

References

[1] J. Abate, G. Choudhury, D. Lucantoni, and W. Whitt. Asymptotic analysis of tail probabilities based on the computation of moments. *The Annals of Applied Probability*, 5(4):983–1007, 1995.

[2] J. Abate, G. Choudhury, and W. Whitt. On the Laguerre method for numerically inverting Laplace transforms. *INFORMS Journal on Computing*, 8:413–427, 1996.

[3] J. Abate, G. Choudhury, and W. Whitt. Numerical inversion of multidimensional Laplace transforms by the Laguerre method. *Performance Evaluation*, 31(3-4):229–243, 1998.

[4] J. Abate and W. Whitt. The Fourier-series method for inverting transforms of probability distributions. *Queueing Systems*, 10:5–88, 1992.

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

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

[7] J. Abate and W. Whitt. Numerical inversion of Laplace transforms of probability distributions. *ORSA Journal on Computing*, 7(1):36–43, 1995.

[8] J. Abate and W. Whitt. Asymptotics for M/G/1 low-priority waiting-time tail probabilities. *Queueing Systems*, 25:173–233, 1997.

[9] J. C. P. Blanc. On the numerical inversion of busy-period related tranforms. *Operations Research Letters*, 30:33–42, 2002.

[10] H. Bruneel, B. Steyaert, E. Desmet, and G.H. Petit. Analytic derivation of tail probabilities for queue lengths and waiting times in ATM multiserver queues. *European Journal of Operational Research*, 76:563–572, 1994.

[11] G. Choudhury and D. Lucantoni. Numerical computation of the moments of a probability distribution from its transforms. *Operations Research*, 44(2):368–381, 1996.

[12] G. Choudhury and W. Whitt. Computing transient and steady-state distributions in polling models by numerical transform inversion. *Performance Evaluation*, 25:267–292, 1994.

[13] G. Choudhury and W. Whitt. Computing distributions and moments in polling models by numerical transform inversion. *Performance Evaluation*, 25(4):267–292, 1996.

[14] M. Drmota. Systems of functional equations. *Random Structures & Algorithms*, 10(1-2):103–124, 1997.

[15] A. Elwalid and D. Mitra. Analysis, approximations and admission control of a multi-service multiplexing system with priorities. In *Proceedings of the Fourteenth Annual Joint Conference of the IEEE Computer and Communication Societies*, pages 463–472, 1995.

[16] G.A. Frolov and M.Y. Kitaev. A problem of numerical inversion of implicitly defined Laplace transforms. *Computers Mathematics with Applications*, 36(5):35–44, 1998.

[17] K. Laevens and H. Bruneel. Discrete-time multiserver queues with priorities. *Performance Evaluation*, 33(4):249–275, 1998.

[18] T. Maertens, J. Walraevens, and H. Bruneel. On priority queues with priority jumps. *Performance Evaluation*. To be published.

[19] T. Sakurai. Numerical inversion for Laplace transforms of functions with discontinuities. *Advances in Applied Probability*, 36(2):616–642, 2004.

[20] S. Shakkottai and R. Srikant. Many-sources delay asymptotics with applications to priority queues. *Queueing Systems*, 39(2-3):183–200, 2001.

[21] R. L. Strawderman. Computing tail probabilities by numerical Fourier inversion: the absolutely continuous case. *Statistica Sinica*, 14:175–201, 2004.

[22] V. Subramanian and R. Srikant. Tail probabilities of low-priority waiting times and queue lengths in MAP/GI/1 queues. *Queueing Systems*, 34(1-4):215–236, 2000.

[23] A. Sughara, T. Takine, Y. Takahashi, and T. Hasegawa. Analysis of a non-preemptive priority queue with SPP arrivals of high class. *Performance Evaluation*, 21:215–238, 1995.

[24] J. Walraevens. *Discrete-time queueing models with priorities.* PhD thesis, Ghent University, 2004.

[25] J. Walraevens, B. Steyaert, and H. Bruneel. Performance analysis of the system contents in a discrete-time non-preemptive priority queue with general service times. *Belgian Journal of Operations Research, Statistics and Computer Science (JORBEL)*, 40(1-2), 2000.

[26] J. Walraevens, B. Steyaert, and H. Bruneel. Performance analysis of a single-server ATM queue with a priority scheduling. *Computers and Operations Research*, 30(12):1807–1829, 2003.