Baltzer Journals *Queueing Systems* 25 (1997) 173–223

ASYMPTOTICS FOR M/G/1 LOW-PRIORITY WAITING-TIME TAIL PROBABILITIES

Joseph Abate[1], and Ward Whitt[2],

[1] *900 Hammond Road, Ridgewood, NJ 07450-2908*

[2] *AT&T Laboratories, Room 2C-178, Murray Hill, NJ 07974-0636*
E-mail: wow@research.att.com

We consider the classical M/G/1 queue with two priority classes and the nonpreemptive and preemptive-resume disciplines. We show that the low-priority steady-state waiting-time can be expressed as a geometric random sum of i.i.d. random variables, just like the M/G/1 FIFO waiting-time distribution. We exploit this structures to determine the asymptotic behavior of the tail probabilities. Unlike the FIFO case, there is routinely a region of the parameters such that the tail probabilities have non-exponential asymptotics. This phenomenon even occurs when both service-time distributions are exponential. When non-exponential asymptotics holds, the asymptotic form tends to be determined by the non-exponential asymptotics for the high-priority busy-period distribution. We obtain asymptotic expansions for the low-priority waiting-time distribution by obtaining an asymptotic expansion for the busy-period transform from Kendall's functional equation. We identify the boundary between the exponential and non-exponential asymptotic regions. For the special cases of an exponential high-priority service-time distribution and of common general service-time distributions, we obtain convenient explicit forms for the low-priority waiting-time transform. We also establish asymptotic results for cases with long-tail service-time distributions. As with FIFO, the exponential asymptotics tend to provide excellent approximations, while the non-exponential asymptotics do not, but the asymptotic relations indicate the general form. In all cases, exact results can be obtained by numerically inverting the waiting-time transform.

Keywords: priority queues, M/G/1 queue, low-priority waiting time, tail probabilities, asymptotics, non-exponential asymptotics, asymptotic expansions, Laplace transforms, algebraic singularities

1 Introduction

In this paper we study the low-priority steady-state waiting-time distribution in the classical M/G/1 queue with two priority classes and the nonpreemptive and preemptive-resume disciplines. The priority structure tends to make the low-priority waiting-time distribution have a relatively long tail. We quantify this effect.

The Laplace transform of the low-priority waiting-time distribution and the first few moments are well known, e.g., see Section III.3.6 of Cohen [30], Section 11.5 of Heyman and Sobel [41] and Theorem 2.1 below, but the distribution itself is somewhat complicated, depending on the busy-period distribution of the high-priority customers, which in turn is characterized implicitly via the Kendall functional equation for its transform, (2.4) below. In fact, the low-priority waiting-time distribution is easily computed numerically by numerical transform inversion, e.g., by the Fourier-series method or the Laguerre series method; see Abate and Whitt [4] and Abate, Choudhury and Whitt [15]. The required high-priority busy-period transform values for complex arguments are easily obtained by iterating the Kendall functional equation for the high-priority busy-period transform, as shown in Abate and Whitt [5]. This iterative calculation of the busy-period transform was also used to calculate transient performance measures in the M/G/1 queue with the first-in first-out (FIFO) discipline in Choudhury, Lucantoni and Whitt [27].

Here we aim to obtain a better understanding of the low-priority waiting-time distribution by doing asymptotic analysis to determine the limiting form as $t \to \infty$. As in the FIFO case, there is often an exponential tail; i.e., if W_2 is the steady-state low-priority waiting-time cumulative distribution function, then we often have

$$1 - W_2(t) \sim \alpha e^{-\eta t} \quad \text{as} \quad t \to \infty \tag{1.1}$$

for positive constants α and η, where $f(t) \sim g(t)$ as $t \to \infty$ means that $f(t)/g(t) \to 1$ as $t \to \infty$; see Abate, Choudhury and Whitt [12], [14]. Moreover, as in the FIFO case, the asymptote in (1.1) usually provides a good approximation for t not too small.

However, unlike the FIFO case, there also is routinely asymptotics of the form

$$1 - W_2(t) \sim \alpha t^{-3/2} e^{-\eta t} \quad \text{as} \quad t \to \infty \tag{1.2}$$

for positive constants α and η (different from the parameters in (1.1)). This observation evidently has not been made before in the literature except in a special case considered by Washburn [52], which we discuss in Section 13.

With FIFO we can also have asymptotics of the general form (1.2), but only for special service-time distributions, called class II in [11], for which the service-time Laplace transform $\hat{g}(s)$ has rightmost singularity $-s^* < 0$ and $\hat{g}(-s^*) < \infty$. (A Laplace transform of a probability density is analytic in the right halfplane, so any singularity s must have $Re(s) \leq 0$. See Section 3 for a classification of probability distributions.) Unlike FIFO, asymptotics of the form (1.2) occurs for the low-priority waiting-time distribution for well-behaved service-time distributions, e.g., when both service-time distributions are exponential. Indeed, for well behaved service-time distributions, we have exponential asymptotics as in (1.1) when the low-priority arrival rate is above a critical threshold (with the requirement that the model still be stable), and the non-exponential asymptotics in (1.2) when the low-priority arrival rate is below this threshold. Moreover, when (1.2) holds, the asymptotic decay rate η in (1.2) does not increase (corresponding to smaller tail probabilities) when the low-priority arrival rate decreases.

Upon reflection, this can be understood by observing that the low-priority waiting time should be dominated by the remaining high-priority busy period when the low-priority arrival rate is sufficiently small. Thus, as the low-priority arrival rate approaches 0, the low-priority waiting-time distribution approaches a nondegenerate limit, which corresponds to the high-priority equilibrium time to emptiness, which in turn is closely related to the equilibrium excess of the busy-period distribution (see (2.13) below). Thus, the asymptotic form (1.2) can be anticipated if one realizes that it is the asymptotic form of the M/G/1 busy-period distribution, first determined by Cox and Smith [31]; see Theorems 7.1 and 8.1 below.

As in the FIFO case, it is also possible to have asymptotics different from (1.1) and (1.2). However, such asymptotics can occur only with the class II and III service-time distributions in [11]. (Class III distributions are long-tail service-time distributions, having 0 as the rightmost singularity of their Laplace transforms; see Section 3 for more discussion.) We obtain such asymptotic results here, although some of our procedures remain to be fully justified. (We will point out the gaps.)

Our experience is that the non-exponential asymptotes, (1.2) and others, tend to be not nearly as accurate approximations as the exponential asymptote in (1.1), so that in applications it may be desirable to calculate exact values in those cases, but the asymptotes indicate the general tendency. We give numerical examples here, which support previous examples [3], [8], [11], [12], [13], [14]. The reader could elect to go directly to the examples in Section 14.

The quality of the different asymptotic approximations can be explained theoretically by the very different rates of convergence that prevail. Usually the rate of convergence is exponential in (1.1) and only linear in (1.2); i.e., the typical refinements of (1.1) and (1.2) are, respectively,

$$1 - W_2(t) \sim \alpha e^{-\eta t}(1 + o(e^{-\epsilon t})) \quad \text{as} \quad t \to \infty \tag{1.3}$$

for $\epsilon > 0$ and

$$1 - W_2(t) \sim \alpha t^{-3/2} e^{-\eta t}(1 + O(1/t)) \quad \text{as} \quad t \to \infty\ . \tag{1.4}$$

(Recall that an asymptotic expansion

$$f(t) \sim \sum_{k=0}^{\infty} g_k(t) \quad \text{as} \quad t \to \infty \tag{1.5}$$

means that

$$f(t) - \sum_{k=0}^{n} g_k(t) \sim g_{n+1}(t) \quad \text{as} \quad t \to \infty\ , \tag{1.6}$$

for all $n \geq 0$; see Erdélyi [37] or Olver [46].

The exponential asymptotics with the exponential rate of convergence in (1.3) occurs when the rightmost singularity of the low-priority waiting-time Laplace transform $\hat{w}_2(s)$ is a simple pole at $-\eta$ with $Re(\phi) < -(\eta + \epsilon)$ for all other singularities ϕ. The non-exponential asymptotics in (1.4) occurs when the rightmost singularity of $\hat{w}_2(s)$ is a branch point singularity at $-\eta$, such that Heaviside's asymptotic expansion applies in the form

$$\hat{w}_2(s) \equiv \int_0^{\infty} e^{-st} dW_2(t) \sim \sum_{k=0}^{\infty} a_k (s + \eta)^{k/2} \quad \text{as} \quad s \to -\eta\ ; \tag{1.7}$$

see p. 254 of Doetsch [34] or p. 139 of Van der Pol and Bremmer [51]. The whole integer powers in (1.7) correspond to analytic functions and thus have no influence upon the asymptotics as $t \to \infty$ in the time domain. We establish results of the form (1.3) and (1.4) in Sections 8 and 12 here. We provide a conceptually satisfying derivation by obtaining a corresponding asymptotic expansion for the high-priority busy-period transform from

Kendall's functional equation. The asymptotic expansion (1.7) for the transform $\hat{w}_2(s)$ follows quite directly from the asymptotic expansion for the busy-period transform by exploiting simple relations among asymptotic expansions, as on p. 19 of Olver [46].

We also determine the different asymptotic behavior at the boundary between (1.1) and (1.2). At this boundary point,

$$1 - W_2(t) \sim \alpha t^{-1/2} e^{-\eta t} \quad \text{as} \quad t \to \infty \ , \tag{1.8}$$

where α and η are positive constants (different from those in (1.1) and (1.2)). The boundary asymptote in (1.8) can provide a much better approximation than the other two approximations in the neighborhood of the boundary. When (1.1) applies near the boundary point, the asymptotic constant α in (1.1) becomes very small; when (1.2) applies near the boundary point, the asymptotic constant α in (1.2) becomes very large. (See Remark 7.4 and Theorem 7.5.)

The work here complements other recent work on asymptotics for waiting-time distributions with non-FIFO disciplines. Related asymptotics for the LIFO discipline appears in Abate, Choudhury and Whitt [12] and Abate and Whitt [1], for polling models in Choudhury and Whitt [26] and Duffield [35], and for the random-order-of-service (ROS) discipline in Flatto [39]. For the LIFO discipline, the asymptotic form is as in (1.2). For polling models, the asymptotics is of the more general form $W^c(t) \sim \alpha t^{-\beta} e^{-\eta t}$ as $t \to \infty$. For ROS, even for the M/M/1 model, the asymptotics is not of this form; Flatto shows for the M/M/1 model that $W^c(t) \sim \alpha t^{-5/6} e^{-\eta t - \gamma t^{1/3}}$ as $t \to \infty$ for positive constants α, η and γ.

Here is how this paper is organized. In Section 2 we introduce our notation and establish convenient representations for the low-priority waiting-time distribution. A geometric random sum representation enables us to directly apply previous asymptotic results. In Section 3 we discuss the three classes of distributions on the positive halfline mentioned above and establish preliminary results. In Section 4 we show that the low-priority waiting-time distribution has a long tail if and only if at least one of the two service-time distributions has a long tail. In Section 5 we establish the main asymptotic result for $1-W_2(t)$. Throughout the paper we give arguments in both the time domain and the transform domain. In Section 6 we establish heavy-traffic limits (as the low-priority traffic intensity increases). We show that the two iterated limits agree.

A more detailed description of the asymptotics for $1 - W_2(t)$ depends on asymptotics for the high-priority busy-period distribution. In Section 7 we review the asymptotics for the busy-period distribution obtained by saddle point methods by Cox and Smith [31] and apply the results to support (1.2). In Section 8 we develop the new asymptotic expansion for the busy-period transform and apply it to support (1.3)–(1.8). In Section 9 we establish results for the long-tail case. We obtain the asymptotes in considerable generality there, but several of the arguments remain to be made rigorous. In Section 10 we establish associated asymptotic results for the low-priority sojourn time, which is the low-priority waiting time plus the low-priority service time (completion time) for the nonpreemptive (preemptive-resume) discipline.

The next three sections consider special cases. In Section 11 we discuss the special case in which the high-priority service-time distribution is exponential. We also discuss the special case in which both service-time distributions are exponential. In Section 12 we discuss the special case in which the two service distributions coincide, but may be general. In Section 13 we discuss the special case, first considered by Washburn [52], in which the low-priority arrival rate is high and service times are short, such that the low-priority traffic intensity is nonnegligible, making the low priority class correspond to fluid input. We also consider the case in which the high-priority input corresponds to a fluid. In all these special cases we are able to find more explicit expressions.

In Section 14 we give numerical examples. The first two numerical examples are for the case in which the two classes have a common exponential service-time distribution. The first example focuses on the two cases in (1.1) and (1.2); the second example focuses on the boundary case in (1.8). The third example is for the case in which the two classes have a common long-tail service-time distribution.

We conclude this introduction by mentioning that results in this paper are applied in Berger and Whitt [21] to investigate the concept of effective bandwidths for waiting times in high-speed communication networks using priorities; see de Veciana, Kesidis and Walrand [33] and Chang and Thomas [24] for overviews of previous work with the FIFO discipline. The asymptotic results here imply that modifications in the concept of effective bandwidths are needed when they are applied to waiting times with priorities.

2 The Geometric Random Sum Representation

In this section we express the classical results for the M/G/1 low-priority waiting-time transform in a convenient form. We consider two priorities, with high priority indexed by 1 and low priority indexed by 2. We consider the waiting time before beginning service in the non-preemptive case, which is the same as for preemptive resume; e.g., see (3.63) on p. 450 and (3.76) on p. 455 of Cohen [30]. Results for the waiting time to complete service – the overall sojourn time – will be obtained as a corollaries in Section 10, just as was done for the FIFO discipline in [14].

For any *cumulative distribution function* (cdf) $G(t)$ on the nonnegative real line, let $g(t)$ be its *probability density function* (pdf), $G^c(t) = 1 - G(t)$ its *complementary cdf* (ccdf), g_k its k^{th} *moment*,

$$\hat{g}(s) = \int_0^\infty e^{-st} dG(t) = \int_0^\infty e^{-st} g(t) dt \tag{2.1}$$

its *Laplace-Stieltjes transform* (the Laplace transform of $g(t)$) and

$$\hat{G}_e(t) = \frac{1}{g_1} \int_0^t G^c(u) du \tag{2.2}$$

its associated *equilibrium excess cdf*, which has Laplace-Stieltjes transform $\hat{g}_e(s) = (1 - \hat{g}(s))/sg_1$.

For class i ($i = 1, 2$), let the *arrival rate* be λ_i, the *service-time cdf* be $G_i(t)$ and the *traffic intensity* be $\rho_i \equiv \lambda_i g_{i1}$. To have proper steady-state distributions, we assume that

$$\rho \equiv \rho_1 + \rho_2 < 1 \ . \tag{2.3}$$

Without loss of generality (by choosing units), let $g_{11} = 1$.

The distribution of the low-priority waiting time depends on the busy-period distribution of the high priority class and related quantities. Let $B_1(t)$ be the class-1 busy-period cdf, $b_1(t)$ its density and $\hat{b}_1(s)$ the Laplace transform of $b_1(t)$, defined as in (2.1). The busy-period transform is characterized as the solution to the *Kendall functional equation*

$$\hat{b}_1(s) = \hat{g}_1(s + \rho_1 - \rho_1 \hat{b}_1(s)) \ . \tag{2.4}$$

Let

$$H_0^{(1)}(t) = (1 - P_{00}^{(1)}(t))/\rho \tag{2.5}$$

be the high-priority *server-occupancy cdf*, where $P_{00}^{(1)}(t)$ is the high-priority *emptiness probability*, i.e., the probability that the system has no class-1 customers at time t given that it had none at time 0, which is known to be monotone, so that $H_0^{(1)}(t)$ in (2.5) is a bonafide cdf; see Section 3 of [6]. Let $\hat{h}_0^{(1)}(s)$ be the Laplace-Stieltjes transform of $H_0^{(1)}(t)$, again defined as in (2.1). From Theorem 4 of [6],

$$\hat{h}_0^{(1)}(s) = \frac{\hat{b}_{1e}(s)}{1 - \rho_1 + \rho_1 \hat{b}_{1e}(s)} = \frac{1 - \hat{b}_1(s)}{s + \rho_1 - \rho_1 \hat{b}_1(s)} \ . \tag{2.6}$$

Let $F_{x0}^{(1)}(t)$ be the cdf of the *first passage time* to 0 for class 1 starting from an initial level of class-1 work x. The Laplace transform of its density $f_{x0}^{(1)}(t)$ is

$$\hat{f}_{x0}^{(1)}(s) = e^{-xz_1(s)} \ , \tag{2.7}$$

where

$$z_1(s) = s + \rho_1 - \rho_1 \hat{b}_1(s) \ ; \tag{2.8}$$

see (33) of [6]. The first passage time starting from a random level with density $g(t)$ having Laplace transform $\hat{g}(s)$ thus has Laplace transform $\hat{g}(z_1(s))$. The busy-period transform can thus be expressed as $\hat{b}_1(s) = \hat{g}_1(z_1(s))$ as in (2.4).

A convenient representation for the low-priority virtual waiting time until beginning service, which holds in much more general models, is the class-1 first passage time to 0 initialized by the steady-state workload of both classes. (The class-i workload is the remaining service time of all class-i customers in the system at an arbitrary time in steady state or, by the Poisson Arrivals See Time Averages (PASTA) property, just before an arrival; see Wolff [54].) If a low-priority customer were to arrive in steady state, he finds the steady-state workload of both classes. He must then wait for this work and all subsequent class-1 input to clear until he can begin service. By the PASTA property, the actual waiting-time distribution coincides with the virtual waiting-time distribution with Poisson arrivals.

Thus, let V be the steady-state workload cdf for both classes and let $\hat{v}(s)$ be its Laplace-Stieltjes transform. We have just observed that

$$\hat{w}_2(s) = \hat{v}(z_1(s)) \ . \tag{2.9}$$

and

$$W_2(t) = \int_0^\infty F_{x0}^{(1)}(t) dV(x) \tag{2.10}$$

A key concept in analyzing the M/G/1 priority queue introduced by Gaver [40], is the *completion time*, which is the time from when one low-priority customer can begin service until the next could begin if he were present. The completion time transform is

$$\hat{c}(s) = \hat{g}_2(z_1(s)) \ . \tag{2.11}$$

Let $\hat{y}_2(s)$ be the Pollaczek-Khintchine transform with the completion time as service time and arrival rate ω/c_1, i.e.,

$$\hat{y}_2(s) = \frac{1-\omega}{1-\omega\hat{c}_e(s)} \ . \tag{2.12}$$

(Formula (2.12) is not the class-2 workload.)

Another useful quantity is the *equilibrium time to emptiness* for class-1, which has transform

$$\hat{f}_{e0}^{(1)}(s) = \hat{v}_1(z_1(s)) = 1-\rho_1+\rho_1\hat{b}_{1e}(s) \ , \tag{2.13}$$

where $\hat{v}_1(s) = \hat{w}_1(s)$ is the Laplace transform of the steady-state workload for class-1; see Theorem 3 of [6]. In the introduction, we noted that $\hat{f}_{e0}^{(1)}(s)$ is the light-traffic limit of $\hat{w}_2(s)$, i.e., the limit as $\rho_2 \to 0$ with ρ_1 fixed. (This result follows easily from the second representation below.)

We now give three convenient expressions for the low-priority steady-state waiting-time transform.

Theorem 2.1 *The steady-state waiting time until beginning service for the low-priority class has Laplace transform*

$$\begin{aligned}
\hat{w}_2(s) \equiv \int_0^\infty e^{-st}dW_2(t) &= \hat{v}(z_1(s)) \\
&= \frac{(1-\omega)\hat{f}_{e0}^{(1)}(s)}{1-\omega\hat{c}_e(s)} = \hat{y}_2(s)\hat{v}_1(z_1(s)) \\
&= \frac{1-\rho}{1-\rho\hat{f}(s)} \ ,
\end{aligned} \tag{2.14}$$

where

$$\hat{f}(s) = \frac{\rho_1}{\rho_1+\rho_2}\hat{h}_0^{(1)}(s) + \frac{\rho_2}{\rho_1+\rho_2}\hat{g}_{2e}(z_1(s)) \tag{2.15}$$

for $\omega = \rho_2/(1-\rho_1)$, $\hat{h}_0^{(1)}(s)$ in (2.6), $z_1(s)$ in (2.8), $\hat{g}_{2e}(s) = (1-\hat{g}_2(s))/g_{21}s$, $\hat{f}_{\epsilon0}^{(1)}(s)$ in (2.13), $\hat{c}_e(s) = (1-\hat{c}(s))/c_1$ for $\hat{c}(s)$ in (2.11) and $c_1 = g_{21}/(1-\rho_1)$, $\hat{y}_2(s)$ in (2.12), and $\hat{v}(s)$ is the transform of the steady-state workload for both classes.

The first formula in Theorem 2.1 is given in (2.9) above. The second formula is a variant of a familiar decomposition; see p. 441 of Heyman and Sobel [41]. It is also elementary to show that the last formula is equivalent to the transform formula in Cohen [30].

The final representation (2.14) is attractive because it is in the same form as the Pollaczek-Khintchine transform for the FIFO waiting time with $\hat{f}(s)$ in (2.14) and (2.15) substituted for the transform of the equilibrium excess of the service-time distribution. The transform $\hat{f}(s)$ in (2.15), being a mixture of the two probability transforms $\hat{h}_0^{(1)}(s)$ and $\hat{g}_{2e}(z_1(s))$, is itself a transform of a probability distribution. Let $F(t)$ be the cdf associated with $\hat{f}(s)$ in (2.15), i.e.,

$$F(t) = \frac{\rho_1}{\rho_1+\rho_2}H_0^{(1)}(t) + \frac{\rho_2}{\rho_1+\rho_2}\int_0^\infty F_{x0}^{(1)}(t)dG_{2e}(x) \ . \tag{2.16}$$

Paralleling the classic Beneš formula for the M/G/1 FIFO waiting time, e.g., (4.82) on p. 255 of [30], we obtain a time-domain formula directly from (2.14). It is a geometric mixture of convolutions of the cdf F in (2.16).

Corollary. *The ccdf of the low-priority waiting time is*

$$W_2^c(t) = \sum_{n=1}^{\infty} (1-\rho)\rho^n F_n^c(t) \tag{2.17}$$

where $F_n^c(t)$ is the ccdf of the n-fold convolution of $F(t)$ in (2.16).

The geometric random sum representation in (2.14) and (2.17) also holds for the GI/G/1 FIFO steady-state waiting-time, there achieved by the ladder variable representation; see Chapters VII and VIII of Asmussen [18] and (14) of [12].

The moments of the low-priority waiting time are easily computed from (2.14), (2.16) and the moments of $H_0^{(1)}(t)$ and $F_{x0}^{(1)}(t)$, given in Theorems 6 and 7 of [11]. For example, the well-known formula for the mean is

$$w_{21} = \frac{\rho_1(g_{12}/g_{11}) + \rho_2(g_{22}/g_{21})}{2(1-\rho_1)(1-\rho)} = \frac{v_1}{1-\rho_1} , \tag{2.18}$$

where v_1 is the mean steady-state workload for both classes, i.e.,

$$v_1 = \frac{\rho g_1(c_s^2+1)}{2(1-\rho)} , \tag{2.19}$$

with G being the mixture of the cdf's G_1 and G_2, i.e.,

$$G(t) = \frac{\lambda_1 G_1(t)}{\lambda_1+\lambda_2} + \frac{\lambda_2 G_2(t)}{\lambda_1+\lambda_2} , \tag{2.20}$$

g_k is the k^{th} moment of G and $c_s^2 = (g_2 - g_1^2)/g_1^2$. To have (2.18) finite we require that $g_{12} < \infty$ and $g_{22} < \infty$. It turns out that the mean in (2.18) describes the performance remarkably well under heavy loads (when ρ is suitably large; see Section 6). When ρ is large and ρ_2 is small (so that $\rho_1 \approx \rho$), we see that $w_{21} = O((1-\rho)^{-2})$.

Remark 2.1. The first formula in (2.14) suggests an approximation. We can approximate the distribution of V by an exponential cdf with mean $v_1 = (1-\rho_1)w_{21}$ (for motivation, see Section 3, Theorem 4.1(a) and Remark 4.1) and we can approximate $f_{x0}^{(1)}(t)$ by an inverse Gaussian distribution. Then we obtain an *exponential mixture of inverse Gaussian distributions* (EMIG) as an approximation for the cdf $W_2(t)$. As a refinement, we can let V have its known atom of $1-\rho$ at the origin and let $(V|V>0)$ be exponential with mean v_1/ρ. For related previous work, see [3], [8] and [9]. ∎

Unfortunately, the explicit formula for the ccdf $W_2^c(t)$ in (2.17) is not so convenient for calculation. However, it is straightforward to calculate the ccdf $W_2^c(t)$ numerically by inverting its transform

$$\hat{W}_2^c(s) = \int_0^{\infty} e^{-st} W_2^c(t)dt = \frac{1-\hat{w}_2(s)}{s} \tag{2.21}$$

for $\hat{w}_2(s)$ in (2.14). For example, we can apply the Fourier-series method in [4]. We can obtain the required busy-period transform values $\hat{b}_1(s)$ for complex s by iterating the Kendall equation (2.4), as indicated in [5].

We conclude this section by pointing out that the two-dimensional transform of the joint steady-state distribution of the high-priority and low-priority workloads, say $\hat{v}(s_1, s_2)$, is given in Kella [44]. It is thus also possible to calculate this joint distribution by applying two-dimensional inversion algorithms, e.g., as in Choudhury, Lucantoni and Whitt [27] or Abate, Choudhury and Whitt [16]. An algorithm to compute the joint distribution of the queue lengths when both classes have exponential service times was developed by Miller [45].

3 Three Classes of Distributions

As indicated in Section 5 of [11], it is useful to divide probability distributions on the positive halfline into three classes according to the rightmost singularity of the Laplace transform and the value of the Laplace transform at that singularity. When we classify the service-time distributions in this way, we determine the relevant cases for waiting-time asymptotics.

Let $g(t)$ be a pdf with Laplace transform $\hat{g}(s)$ (defined as in (2.1)). Let $-s^*$ be the rightmost singularity of $\hat{g}(s)$, with $-s^* = -\infty$ if $\hat{g}(s)$ is analytic everywhere, as in the case $g(t)$ has bounded support. (Note that s^* is the radius of convergence of the moment generating function, which equals $\hat{g}(-s)$ for real s with $s < s^*$. However, the Laplace transform $\hat{g}(s)$ might be defined for complex s with $Re(s) < -s^*$ even though the moment generating function is infinite.) Since $g(t)$ is a pdf, we always have $s^* \geq 0$. In this setting the pdf $g(t)$ and its transform $\hat{g}(s)$ are classified as follows:

$$\begin{array}{ll} \text{class I:} & s^* > 0 \text{ and } \hat{g}(-s^*) = \infty \ , \\ \text{class II:} & s^* > 0 \text{ and } 1 < g(-s^*) < \infty \ ; \\ \text{class III:} & s^* = 0 \text{ and } \hat{g}(-s^*) = 1 \ . \end{array} \tag{3.1}$$

Class-I distributions are the "well behaved" distributions. All pdf's with rational Laplace transforms are class I; e.g., this includes all phase-type distributions. Whenever the rightmost singularity $-s^*$ of $\hat{g}(s)$ is a pole (single or multiple), $\hat{g}(-s^*) = \infty$. However, even if $-s^*$ is a branch point singularity, we may have $g^*(-s^*) = \infty$; e.g., this is the case for the gamma (1/2) transform $1/\sqrt{1+2s}$. The steady-state waiting time in an M/G/1 queue with a service-time distribution having mean 1 has an exponential tail for all arrival rates ρ, $0 < \rho < 1$, if and only if the service-time distribution is class I; e.g., see [12].

Class-III distributions are the long-tail distributions, which fail to have proper moment generating function. They are often called *subexponential distributions* because they tend to decay more slowly than any exponential. Any pdf with only finitely many finite moments is in Class III. A familiar example is the Pareto distribution. Another example is the Pareto mixture of exponential (PME) distributions in [11]. There are also class-III distributions with moments of all orders, such as the Weibull and log-normal distributions. In particular, the Weibull ccdf $G^c(t) = e^{-(at)^c}$ is class III when $c < 1$. The steady-state waiting time in an M/G/1 queue with a service-time distribution having mean 1 has a class III distribution for all arrival rates ρ, $0 < \rho < 1$, if and only if the service-time distribution is class III. (This is a consequence of Theorem 4.2 below; related results are reviewed in [11].)

In [11] we called class-II distributions *semi-exponential distributions*, because they are dominated by an exponential, i.e.,

$$\lim_{t\to\infty} e^{\gamma t} G^c(t) = 0 \tag{3.2}$$

for all $\gamma < s^*$, but they do not have a pure exponential tail. Since $\hat{g}(-s^*) < \infty$, the rightmost singularity $-s^*$ is necessarily a branch point singularity, not a pole.

Class-II distributions are less familiar, but they are important for us because the M/G/1 busy-period distribution is usually class II. Indeed, the M/G/1 busy-period distribution is always class II when the service-time distribution is class I. We will elaborate on this point in Sections 7 and 8.

Class-II distributions are intimately related to class-III distributions, because one can be converted into the other by exponential damping. Given any class-III pdf $g(t)$, and any $\gamma > 0$, we can construct an associated class-II pdf

$$g_\gamma(t) = \frac{e^{-\gamma t} g(t)}{\hat{g}(\gamma)}, \quad t \ge 0 . \tag{3.3}$$

Similarly, given any class-II pdf $g_\gamma(t)$ with rightmost singularity at $-\gamma$ (and $\hat{g}_\gamma(-\gamma) < \infty$), we can construct an associated class-III pdf

$$g(t) = \frac{e^{\gamma t} g_\gamma(t)}{\hat{g}_\gamma(-\gamma)}, \quad t \ge 0 . \tag{3.4}$$

Since the distributions are classified according to the rightmost singularities of the Laplace transform, it is useful to be able to draw conclusions about these singularities. For this purpose, we give an elementary comparison lemma which is based on ordinary stochastic order; see Section 1.2 of Stoyan [49].

Lemma 3.1. *If X_1 and X_2 are two random variables with $P(X_1 > t) \le P(X_2 > t)$ for all t, then $Ee^{\gamma X_1} \le Ee^{\gamma X_2}$ for all positive real γ. Hence the rightmost singularities of the Laplace transforms $\hat{g}_i(s) = Ee^{-sX_i}$ are ordered by $-s_1^* \le -s_2^*$.*

We now want to relate the rightmost singularity of $\hat{g}(s)$ to the tail behavior of the ccdf $G^c(t)$.

Lemma 3.2. *Let X be a random variable with ccdf $G^c(t)$. Then for positive real γ,*

$$Ee^{\gamma X} \equiv \int_0^\infty e^{\gamma t} dG(t) < \infty \textit{ if and only if } \int_0^\infty e^{\gamma t} G^c(t) dt < \infty , \tag{3.5}$$

in which case

$$\lim_{t\to\infty} e^{\gamma t} G^c(t) = 0 . \tag{3.6}$$

As a consequence, the rightmost singularities of $\hat{g}(s)$ and $\hat{G}^c(s)$ coincide.

Proof. Apply integration by parts as on p. 150 of Feller [38], obtaining

$$\int_0^b e^{\gamma t} dG(t) = 1 - e^{\gamma b} G^c(b) + \gamma \int_0^b e^{\gamma t} G^c(t) dt . \tag{3.7}$$

First suppose that the integral on the left in (3.7) converges as $b \to \infty$. Since the contribution over (b, ∞) dominates $e^{\gamma b} G^c(b)$, we must have $e^{\gamma b} G^c(b) \to 0$ as $b \to \infty$ and the integral on the right in (3.7) must converge. Next suppose that the integral on the right in (3.7) converges. Since one plus the integral on the right is greater than the integral on the left, the integral on the left must converge. ∎

Lemma 3.2 implies that if $-s^*$ is the rightmost singularity of $\hat{g}(s)$, then s^* is the supremum of those γ for which $e^{\gamma t}G^c(t)$ is integrable. Moreover, if $\gamma < s^*$, then $e^{\gamma t}G^c(t) \to 0$ as $t \to \infty$. Under extra regularity conditions, we have $e^{\gamma t}G^c(t) \to \infty$ for all $\gamma > s^*$.

Lemma 3.3. *Assume that*

$$e^{\gamma t}G^c(t) \to C(\gamma) \quad \text{as } t \to \infty \ , \tag{3.8}$$

where $C(\gamma)$ is some constant with $0 \le C(\gamma) \le \infty$, for all but at most one real γ. Then we must have $C(\gamma) = 0$ for $\gamma < s^*$ and $C(\gamma) = \infty$ for $\gamma > s^*$, where $-s^*$ is the rightmost singularity of the Laplace-Stieltjes transform of $G(t)$.

Proof. The result for $\gamma < s^*$ holds by (3.6) in Lemma 3.2 without condition (3.8). Suppose that $C(\gamma) < \infty$ for some γ with $\gamma > s^*$. Then $e^{\eta t}G^c(t) = O(e^{-(\gamma-\eta)t})$ as $t \to \infty$ for any η with $\gamma > \eta > s^*$. This implies that $e^{\eta t}G^c(t)$ must be integrable, which by Lemma 3.2, implies that $s^* \ge \eta$, which is a contradiction. Hence, we must have $C(\gamma) = \infty$ for all $\gamma > s^*$. The one value of γ for which (3.8) need not hold must be s^*. ∎

Example 3.1. Lemma 3.2 implies that $e^{\gamma t}G^c(t)$ is not integrable if $\gamma > s^*$, but that does not imply that $e^{\gamma t}G^c(t) \to \infty$ as $t \to \infty$. To understand the need for regularity condition (3.8), let

$$G^c(t) = e^{-s^* 2^k} \ , \quad 2^{k-1} \le t < 2^k \ ,$$

for $k \ge 1$, with $G^c(t) = 1$ for $t < 1$. Then

$$\limsup_{t\to\infty} e^{s^* t}G^c(t) = 1 \ ,$$

but

$$\liminf_{t\to\infty} e^{2s^* t}G^c(t) = 1 \ . \quad ∎$$

The three classes of distributions can also be identified by comparing the tail behavior of the ccdf of the two-fold convolution with the tail behavior of the ccdf itself. For this purpose, we assume the *convolution asymptotic property*

$$\lim_{t\to\infty} G_2^c(t)/G^c(t) = \gamma \tag{3.9}$$

for some constant γ, $0 \le \gamma \le \infty$, where $G_2^c(t)$ is the ccdf of the two-fold convolution of G with itself. From an applied point of view, condition (3.9) is a regularity condition that we should be very willing to assume. The idea of characterizing the tail behavior of the distribution $G(t)$ via the convolution asymptotic property is due to Chistyakov [25] and was further developed by Chover, Ney and Wainger [28], [29]. Further discussion appears in Section IV.4 of Athreya and Ney [20] and Appendix 4 of Bingham, Goldie and Teugels [22].

Assuming that the convolution asymptotic property (3.9) holds, the distributions can be classified according to the constant γ appearing in (3.9). This is intimately related to our previous classification in (3.1) because it can be shown that necessarily

$$\gamma = 2\hat{g}(-s^*) \ ; \tag{3.10}$$

see [20], [28], [29]. As a consequence, $\gamma \geq 2$. Moreover, given (3.9), it can be shown that the n-fold convolutions satisfy the related asymptotic relation

$$\lim_{t \to \infty} G_n^c(t)/G^c(t) = n\hat{g}(-s^*)^{n-1} \ . \tag{3.11}$$

The convolution asymptotic property (3.9) has been established in many special cases, but it remains somewhat elusive in general. We give some illustrative examples and supporting results here.

We first give an example showing the need for the regularity condition (3.9).

Example 3.2. To see the need for condition (3.9), let $\{t_k : k \geq 1\}$ be a sequence with $t_{k+1} > 2t_k$ for all k. Let

$$G^c(t) = e^{-s^* t_{k-1}} \text{ for } \ t_{k-1} \leq t < t_k$$

for $k \geq 1$, with $G^c(t) = 1$ for $t < t_0 = 1$. Then

$$G_2^c(t_k) = G^c(t_k)(2 - G^c(t_k))$$

but, for any fixed m and all k suitably large,

$$G_2^c(t_k + t_m) = 2G^c(t_k)G^c(t_m) - G^c(t_k)^2 = 2G^c(t_k + t_m)G^c(t_m) \ - \ G^c(t_k + t_m)^2 \ ,$$

so that

$$\limsup_{t \to \infty} \frac{G_2^c(t)}{G^c(t)} \geq 2$$

but

$$\liminf_{t \to \infty} \frac{G_2^c(t)}{G^c(t)} \leq 2G^c(t_m) < 2 \ . \ \blacksquare$$

We next give an example covering many class-I distributions. To relate the result about pdf's to ccdf's, we use the following basic lemma; see p. 17 of Erdélyi [37].

Lemma 3.4. *If $f(t) \sim g(t)$ as $t \to \infty$, then $\int_t^\infty f(u)du \sim \int_t^\infty g(u)du$ as $t \to \infty$.*

Example 3.3. Consider two pdf's $g_i(t)$ for $i = 1, 2$, with Laplace transforms $\hat{g}_i(s)$. Suppose that the rightmost singularity of $\hat{g}_i(s)$ is an isolated *simple pole* at $-s^*$ for each i, so that

$$\lim_{s \to 0} s\hat{g}_i(s - s^*) = \alpha_i \tag{3.12}$$

and

$$\lim_{t \to \infty} e^{s^* t} g_i(t) = \alpha_i \tag{3.13}$$

for $0 < \alpha_i < \infty$. Then the convolution pdf

$$(g_1 * g_2)(t) = \int_0^t g_1(t - y)g_2(y)dy \ , \quad t \geq 0 \ , \tag{3.14}$$

has Laplace transform $\hat{g}_1(s)\hat{g}_2(s)$ whose rightmost singularity is an isolated *multiple pole* at $-s^*$, so that

$$s^2 g_1(s - s^*)g_2(s - s^*) \to \alpha_1\alpha_2 \quad \text{as} \ \ s \to 0 \tag{3.15}$$

and

$$\lim_{t\to\infty} e^{s^* t} g_i(t) = \infty \ . \tag{3.16}$$

If the two pdf's $g_1(t)$ and $g_2(t)$ coincide, then we obtain (3.9) with $\gamma = 2\hat{g}(-s^*) = \infty$. By induction on n, we also obtain (3.11) with $g(-s^*) = \infty$. ∎

We next establish (3.9) and (3.11) for the non-exponential asymptotics of primary interest in the priority model. For this purpose, we apply the following lemma.

Lemma 3.5. *If $f(t) \sim \theta(t) \equiv \alpha t^{-\beta} e^{-\eta t}$ as $t \to \infty$, then $\int_t^\infty f(u)du \sim \theta(t)/\eta$ as $t \to \infty$.*

Proof. Given Lemma 3.4 and the condition, it suffices to differentiate $\theta(t)$ and observe that $-\theta'(t)\eta^{-1} \sim \theta(t)$ as $t \to \infty$. ∎

Lemma 3.6. *Let X_1 and X_2 be independent random variables with densities $g_i(t)$, $i = 1, 2$, such that*

$$g_i(t) \sim A_i t^{-\beta} e^{-\eta t} \quad as \quad t \to \infty \tag{3.17}$$

for constants $\eta > 0$, $\beta > 1$ and $A_i > 0$. Then

$$P(X_1 + X_2 > t) \sim \frac{(A_1 d_2 + A_2 d_1)}{\eta} t^{-\beta} e^{-\eta t} \quad as \quad t \to \infty \ . \tag{3.18}$$

where

$$d_i = E e^{\eta X_i} = \hat{g}_i(-\eta) < \infty \ . \tag{3.19}$$

Proof. The expression for convolution yields

$$t^\beta e^{\eta t} P(X_1 + X_2 > t) = t^\beta e^{\eta t} G_2^c(t) + \int_0^t t^\beta e^{\eta(t-y)} G_1^c(t-y) e^{\eta y} g_2(y) dy \ . \tag{3.20}$$

By the assumption and Lemma 3.2, the first term in (3.20) approaches A_2/η, so we focus on the second term of (3.20). We break up the integral in (3.20) into two pieces, above and below $(1-\epsilon)t$ for small positive ϵ. First

$$\int_0^{(1-\epsilon)t} t^\beta e^{\eta(t-y)} G_1^c(t-y) e^{\eta y} g_2(y) dy \to \frac{A_1}{\eta} \int_0^\infty e^{\eta y} g_2(y) dy$$

by the dominated convergence theorem and Lemma 3.2, since $(t/(t-y))^\beta \le (1-\epsilon)^{-\beta}$ for all t and $y \le (1-\epsilon)t$. Second, we do a change of variables in the second integral with $z = t - y$, writing

$$\int_{(1-\epsilon)t}^t t^\beta e^{\eta(t-y)} G_1^c(t-y) e^{\eta y} g_2(y) dy = \int_0^{\epsilon t} e^{\eta z} G_1^c(z) \left(\frac{t}{t-z}\right)^\beta (t-z)^\beta e^{\eta(t-z)} g_2(t-z) dz \ .$$

We now let $t \to \infty$ and apply the dominated convergence theorem plus Lemma 3.2 again to obtain the limit

$$A_2 \int_0^\infty e^{\eta y} G_1^c(y) dy = A_2 \left(\frac{e^{\eta y}}{\eta} G_1^c(y)|_0^\infty + \frac{1}{\eta} \int_0^\infty e^{\eta y} dG_1(y) \right) = \frac{A_2(d_1 - 1)}{\eta} \ ,$$

with the last step following from integration by parts. Combining the pieces yields (3.18). ∎

An important subclass of class III is the set of the ccdf's with *regularly varying tails*; i.e., where

$$G^c(t) \sim t^{-c}L(t) \quad \text{as } t \to \infty \ ,$$

for $c \geq 1$ and $L(t)$ slowly varying; see p. 275 of Feller [38] and Bingham, Goldie and Teugels [22]. (An important special case is when $L(t)$ is a constant.) Feller establishes the following result on p. 278 of [38], which implies (3.9) with $\gamma = 1$ and (3.11) with $g(-s^*) = 1$.

Lemma 3.7. *Let X_1 and X_2 be independent random variables with ccdf's $G_i^c(t)$ for $i = 1, 2$, having regularly varying tails, i.e.,*

$$G_i^c(t) \sim t^{-c}L_i(t) \quad \text{as} \ \ t \to \infty$$

for $c \geq 1$ and $L_i(t)$ slowly varying. Then

$$P(X_1 + X_2 > t) \sim t^{-c}(L_1(t) + L_2(t)) \quad \text{as} \ \ t \to \infty \ .$$

We now show how to establish the convolution asymptotic properties for class II and III distributions using operational principles for Laplace transforms. These operational principles can be rigorously justified in some cases, but remain to be in others; see p. 139 of Van Der Pol and Bremmer [51], p. 254 of Doetsch [34] and Sections 3 and 5 of [11] for more discussion.

Heaviside Operational Principle. *Suppose that a function $g(t)$ has Laplace transform $\hat{g}(s)$ with rightmost singularity $-s^*$ and asymptotic expansion*

$$\hat{g}(s) \sim \sum_{k=0}^{\infty} a_k(s + s^*)^k + \hat{\theta}(s) \quad \text{as} \ \ s \to -s^* \ , \tag{3.21}$$

where $\hat{\theta}(s)$ is the transform (possibly a pseudofunction, see p. 61 of Doetsch [34]) of a function $\theta(t)$. Then $g(t) \sim \theta(t)$ as $t \to \infty$.

The idea is that since the terms $a_k(s + s^*)^k$ are analytic functions, they should not contribute to the asymptotics in the time domain. This operational principle is established as Heaviside's theorem in the case that

$$\hat{\theta}(s) \sim A(s + s^*)^{\gamma} \quad \text{as} \ \ s \to -s^* \tag{3.22}$$

and

$$\theta(t) \sim \frac{Ae^{-s^*t}}{t^{\gamma+1}\Gamma(-\gamma)} \quad \text{as} \ \ t \to \infty \ , \tag{3.23}$$

for some positive non-integer power γ, where $\Gamma(x)$ is the gamma function, on p. 254 of Doetsch [34]. Note that $(s + s^*)^{\gamma}$ is then *not* an analytic function. Then the series only has positive coefficients a_k for $k < \gamma$. Then a condition in the theorem is that $\hat{g}(s)$ have no other singularity besides $-s^*$ for $Re(s) > -s^* - \epsilon$ for some $\epsilon > 0$. Doetsch [34] actually has a slightly stronger singularity condition; the form stated is in Sutton [50]. The application to *Pareto mixture of exponential* (PME) distributions in [11] illustrates an application beyond Heaviside's theorem with $\hat{\theta}(s)$ in (3.22) in Doetsch [34] and Sutton [50].

We can apply Heaviside's operational principle to obtain a convolution operational principle supporting (3.9) and (3.11).

Convolution Operational Principle. *For $i = 1, 2$, let $g_i(t)$ be a function with Laplace transform $\hat{g}_i(s)$ with rightmost singularity $-s^*$ (independent of i) and asymptotic expansion*

$$\hat{g}_i(s) \sim \sum_{k=0}^{\infty} a_{ik}(s+s^*)^k + \hat{\theta}_i(s) \quad \text{as} \quad s \to -s^* \ , \tag{3.24}$$

as in (3.21), where $a_{i0} \neq 0$ for $i = 1, 2$. Let $(g_1 * g_2)(t)$ be the convolution of $g_1(t)$ and $g_2(t)$. Then

$$(g_1 * g_2)(t) \equiv \int_0^t g_1(t-y)g_2(y)dy \sim a_{10}\theta_2(t) + a_{20}\theta_1(t) \quad \text{as} \quad t \to \infty \ . \tag{3.25}$$

Supporting Argument. Note that the transform of $(g_1 * g_2)(t)$ is

$$\hat{g}_1(s)\hat{g}_2(s) \sim \sum_{k=0}^{\infty} b_k(s+s^*)^k + a_{10}\hat{\theta}_2(s) + o(\hat{\theta}_1(s)) + a_{20}\theta_1(s) + o(\theta_2(s)) \quad \text{as} \quad s \to -s^*$$

for suitable constants b_k and apply the Heaviside operational principle. ∎

Of course, if $\theta_1(t)$ dominates $\theta_2(t)$ or vice versa, then only one term appears in the right of (3.25).

Example 3.4. To illustrate the Heaviside operational principle and the convolution operational principle applied to establish the convolution asymptotic formulas (3.9) and (3.11), consider the inverse Gaussian pdf

$$g(t) = \sqrt{\frac{e}{4\pi t^3}} \exp\left(-\frac{t}{4} - \frac{1}{4t}\right) \ , \quad t \geq 0 \ , \tag{3.26}$$

with Laplace transform

$$\hat{g}(s) = \sqrt{e}\exp(-\sqrt{1+4s}/2) \ , \tag{3.27}$$

which has rightmost singularity $-s^* = -1/4$ and $\hat{g}(-1/4) = \sqrt{e}$. From (3.26), we directly obtain the asymptotic relation

$$g(t) \sim \theta(t) \equiv \sqrt{\frac{e}{4\pi t^3}} e^{-t/4} \quad \text{as} \quad t \to \infty \ , \tag{3.28}$$

so that in this case we can verify the results we get by operations on the transforms.

If we expand the exponential in (3.27) in a power series, then we obtain

$$\hat{g}(s) = d + \hat{\psi}(s) \tag{3.29}$$

where $d = \sqrt{e}$, $\hat{\psi}(s) = \sqrt{e}\sum_{n=1}^{\infty} \frac{\hat{\theta}(s)^n}{n!} e^{-n/2}$ and

$$\hat{\theta}(s) = -\sqrt{e}\sqrt{1+4s}/2 \ . \tag{3.30}$$

The function $\hat{\theta}(s)$ in (3.30) is the pseudofunction transform of $\theta(t)$ in (3.28) (see p. 61 of Doetsch [34]). (Note that $\hat{\psi}(s) \to \hat{\theta}(s)$ as $s \to -s^*$.) From Heaviside's operational principle, we deduce that $g(t) \sim \theta(t)$ as $t \to \infty$. Moreover,

$$g(s)^n = (d + \hat{\psi}(s))^n \sim d^n + nd^{n-1}\hat{\psi}(s) \quad \text{as} \quad s \to -s^* = -1/4 \ , \tag{3.31}$$

so that we should have

$$g_n(t) \sim n d^{n-1} g(t) \quad \text{as} \quad t \to \infty \ , \tag{3.32}$$

where $g_n(t)$ is the n-fold convolution of $g(t)$ with itself, just as in (5.13). In this case we can also directly invert $\hat{g}_2(s) = g(s)^2$ to obtain

$$g_2(t) = 2\sqrt{e}\theta(t)e^{-1/t} \ , \quad t \geq 0 \ , \tag{3.33}$$

which agrees with the asymptotic result (3.32). Indeed, in this case we can actually apply Heaviside's theorem as in Doetsch [34] to prove that, not only does $g(t)$ have the asymptotic form (3.28), but it also has an asymptotic expansion of the form

$$g(t) \sim \theta(t)(1 + \sum_{i=1}^{\infty} a_i t^{-i}) \quad \text{as} \quad t \to \infty \tag{3.34}$$

for constants a_i. ∎

4 Long Tail or Not

Focusing on the two service-time distributions in the priority model, we obtain nine cases in which the service-time distributions each can be class I, II or III as described in Section 3. In five of these cases, the low-priority waiting-time distribution is class III; in the other four it is class I or class II. We will show that if either service-time distribution is class III, then so is the low-priority waiting-time distribution. We will also show that if both service-time distributions are class I or II, then so is the low-priority waiting-time distribution.

We first determine when the high-priority busy-period distribution is class III.

Theorem 4.1. *The high-priority busy-period transform $\hat{b}_1(s)$ has 0 as its rightmost singularity if and only if the high-priority service-time transform $\hat{g}_1(s)$ has 0 as its rightmost singularity.*

Proof. First, since a busy period is at least as long as one service time, we have $B_1^c(t) \geq G_1^c(t)$ for all t, so that we can apply Lemma 3.1 to deduce that the rightmost singularities are ordered: If the rightmost singularity of $\hat{g}_1(s)$ is 0, then so is the rightmost singularity of $\hat{b}_1(s)$. On the other hand, if the rightmost singularity of $\hat{g}_1(s)$ is less than 0, then $\hat{g}_1(s)$ is analytic in a neighborhood of 0. If $\hat{g}_1(s)$ is analytic in a neighborhood of 0, then we can apply the implicit function theorem, e.g., p. 269 of Hille [43]. The Kendall functional equation (2.4) is known to have a unique solution, e.g., see [5]. We can expand $\hat{g}(s)$ in a power series and obtain a power series for $\hat{b}_1(s)$ from the Kendall functional equation (2.4) by reversion of power series, as on p. 147 of Cox and Smith [31], which implies that $\hat{b}_1(s)$ is analytic in a neighborhood of 0 as well. ∎

We now turn to the low priority waiting-time distribution.

Theorem 4.2. *The low-priority waiting-time transform $\hat{w}_2(s)$ has 0 as its rightmost singularity if and only if at least one of the service-time transforms $\hat{g}_1(s)$ and $\hat{g}_2(s)$ has 0 as its rightmost singularity.*

Proof. First we bound $W_2^c(t)$ below by

$$W_2^c(t) \geq \rho_1 G_{13}^c(t) + \rho_2 G_{2e}^c(t) \ , \quad t \geq 0 \ , \tag{4.1}$$

so that we can apply Lemma 3.1 to conclude that the rightmost singularity of $\hat{w}_2(s)$ is at least as large as the rightmost singularities of $\hat{g}_{1e}(s)$ and $\hat{g}_{2e}(s)$. By Lemma 3.2, the rightmost singularities of $\hat{g}_i(s)$ and $\hat{g}_{ie}(s)$ coincide for each i. Hence, if either $\hat{g}_1(s)$ or $\hat{g}_2(s)$ has 0 as its rightmost singularity, then so does $\hat{w}_2(s)$.

We can write (4.1) because the steady-state low-priority waiting time exceeds the steady-state workload of the two classes. The steady-state workload in turn exceeds the remaining service time of the customer in service, if any. With probability $1-\rho$ the server is idle, with probability ρ_i the server is serving a class-i customer. Conditional on serving a class-i customer, the remaining service time is distributed as $G_{ie}^c(t)$. Hence we have (4.1) as claimed.

On the other hand, if $\hat{g}_1(s)$ is analytic in a neighborhood of 0, then so is $\hat{b}_1(s)$ by Theorem 4.1. Then also is $\hat{h}_0^{(1)}(s)$ and $z_1(s)$ by (2.6) and (2.8). For $\hat{h}_0^{(1)}(s)$, we can rewrite (2.6) as

$$\hat{b}_{1e}(s) = \frac{(1-\rho_1)\hat{h}_0^{(1)}(s)}{1-\rho_1\hat{h}_0^{(1)}(s)} \ . \tag{4.2}$$

Since the rightmost singularity of $\hat{b}_1(s)$ is inherited by $\hat{b}_{1e}(s)$, by Lemma 3.2, the rightmost singularity of $h_0^{(1)}(s)$ must also be less than 0. (Note that the dominator in (4.2) can only have a root for $s<0$.) By (2.8),

$$\begin{aligned} z_1(s) &= s+\rho_1(1-\hat{b}_1(s)) \\ &= s+\rho_1(b_{11}s+o(s)) \\ &= s+\frac{\rho_1}{1-\rho_1}s+o(s) \\ &= \frac{s}{1-\rho_1}+o(s) \quad \text{as} \quad s\to 0 \ . \end{aligned} \tag{4.3}$$

Hence, if $\hat{g}_2(s)$ is analytic in a neighborhood of 0, so is $\hat{g}_{2e}(s)$ by Lemma 3.2 and $\hat{g}_{2e}(z_1(s))$. Consequently, if $\hat{g}_1(s)$ and $\hat{g}_2(s)$ are both analytic at $s=0$, then so is $\hat{f}(s)$ in (2.15). Since $1-\rho\hat{f}(s)$ can only have a zero for $s<0$, $\hat{w}_2(s)$ in (2.14) is analytic at $s=0$ too. ∎

5 The Main Theorem

We now establish asymptotics for the low-priority waiting-time ccdf $W_2^c(t)$. Because of Theorem 2.1, we can draw on previous theory in the time domain, just as Pakes [47] did to treat non-exponential asymptotics for the GI/G/1 FIFO model; see Athreya and Ney [20], Chover, Ney and Wainger [28], [29] and Abate et al. [11], [12]. We also give alternative transform arguments, which are remarkably simple, but which either require extra conditions or which remain to be fully justified.

As a technical regularity condition needed for treating the cases with non-exponential tails, we will assume that the cdf $F(t)$ in (2.16) has the *convolution asymptotic property*

$$F_2^c(t)/F^c(t)\to\gamma \qquad \text{as} \quad t\to\infty \tag{5.1}$$

for some constant γ, $2\le\gamma\le\infty$, where $F_2^c(t)$ is the ccdf of the two-fold convolution of $F(t)$ with itself, as in (3.9). Lemma 3.6 implies that (5.1) holds in the principal case of non-exponential asymptotics considered here leading to (1.2). Lemma 3.7 treats the special case of a regularly varying tail, leading to (5.1) with $\gamma = 2$.

We will also assume that

$$\frac{F^c(t-b)}{F^c(t)} \to \psi(b) \quad \text{as} \quad t \to \infty \tag{5.2}$$

for each real b. It follows that $\psi(b) = e^{s^* b}$, where $-s^*$ is the rightmost singularity of the Laplace transform $\hat{f}(s)$. When (5.1) holds with $\gamma = 2$, (5.2) also holds with $\psi(b) = 1$; see Lemma 3 on p. 148 of Athreya and Ney [20].

It turns out the asymptotics is determined by the *low-priority waiting-time root equation*

$$\hat{f}(s) = \frac{1}{\rho} \; ; \tag{5.3}$$

a root of (5.3) is a zero of the denominator of the low-priority waiting-time transform $\hat{w}_2(s)$ in (2.14). Clearly, any root of equation (5.3) is a singularity of $\hat{w}_2(s)$. Moreover, since $\hat{f}(s)$ is monotone in real s where it is analytic, equation (5.3) has at most one root to the right of the rightmost singularity of $\hat{f}(s)$.

Theorem 5.1. *Let $-s^*$ be the rightmost singularity of $\hat{f}(s)$ in (2.15).*

(a) If equation (5.3) has a negative real root $-\eta$ with $\eta < s^*$, then

$$W_2^c(t) \sim \alpha e^{-\eta t} \quad \text{as} \quad t \to \infty \; , \tag{5.4}$$

where

$$\alpha = \frac{1-\rho}{-\rho\eta \hat{f}'(-\eta)} \; . \tag{5.5}$$

(b). If (5.1) and (5.2) hold and $\hat{f}(-s^*) < \rho^{-1}$, so that (5.3) has no root in $(-s^*, \infty)$, then

$$W_2^c(t) \sim \frac{\rho(1-\rho)}{(1-\rho\hat{f}(-s^*))^2} F^c(t) \quad \text{as} \quad t \to \infty \tag{5.6}$$

for F in (2.16).

Proof. (a) We give both a probabilistic proof and a transform proof of (5.4). The transform proof requires an extra technical condition, but the proof is very simple. The probabilistic proof is a renewal-theory argument, as on p. 411 of Feller [38], supplemented by a derivation of direct Riemann integrability (a technical condition). In particular, by (2.17),

$$\begin{aligned} W_2(t) &= 1 - \rho + \rho \sum_{n=1}^{\infty} (1-\rho)\rho^{n-1} \int_0^t F_{n-1}(t-y) dF(y) \\ &= 1 - \rho + \rho \int_0^t W_2(t-y) dF(y) \; , \end{aligned}$$

so that $W_2^c(t)$ satisfies the *defective* renewal equation

$$W_2^c(t) = \rho F^c(t) + \rho \int_0^t W_2^c(t-y) dF(y) \; . \tag{5.7}$$

Multiplying (5.7) through by $e^{\eta t}$, we see that $e^{\eta t}W_2^c(t)$ satisfies the *proper* renewal equation

$$e^{\eta t}W_2^c(t) = e^{\eta t}\rho F^c(t) + \rho \int_0^t e^{\eta(t-y)}W_2^c(t-y)e^{\eta y}dF(y) \ , \tag{5.8}$$

provided that $e^{\eta y}\rho dF(y)$ corresponds to a proper probability distribution, i.e., provided that

$$1 = \int_0^\infty e^{\eta y}\rho dF(y) = \rho \hat{f}(-\eta) \ , \tag{5.9}$$

which is the stated condition. Assuming that $e^{\eta x}F^c(x)$ is directly Riemann integrable (dRi), which we verify below, we can apply the key renewal theorem (p. 363 of Feller [38] or p. 118 of Asmussen [18]) to deduce that

$$\lim_{t\to\infty} e^{\eta t}W_2^c(t) = \frac{\int_0^\infty e^{\eta y}F^c(y)dy}{\int_0^\infty ye^{\eta y}dF(y)} \ . \tag{5.10}$$

Using integration by parts in the numerator and denominator of (5.10), we see that the numerator is $\eta^{-1}(\hat{f}(-\eta)-1) = (1-\rho)/\rho\eta$ and the denominator is $-\hat{f}'(-\eta) = (-\hat{f}'(-\eta)\eta - (\eta/\rho))/\eta^2$, which implies (5.5).

Finally, we need to verify that $e^{\eta t}F^c(t)$ is dRi for the application of the renewal theorem. We can apply condition (iv) on p. 119 of Asmussen [18]. Since $\hat{f}(s)$ is analytic for $Re(s) > -s^*$, $Ee^{(\eta+\epsilon)X} < \infty$ for X distributed as F and some $\epsilon > 0$. Hence, $E[e^{(\eta+\epsilon)X};\ X > t] \to 0$ as $t \to \infty$, but $E[e^{(\eta+\epsilon)X}; X > t] \geq e^{(\eta+\epsilon)t}F^c(t)$, which implies that $e^{(\eta+\epsilon)t}F^c(t) \to 0$ as $t \to \infty$. Hence, there is a constant C such that $e^{\eta t}F^c(t) \leq Ce^{\epsilon t}$ for all $t \geq 0$. Since $e^{\eta t}F^c(t)$ is also continuous and bounded, it is dRi by (iv) on p. 119 of [18].

Now we give the transform proof of (5.4). For this, we need an extra technical regularity condition, which seems no serious imposition for practical purposes. We assume that

$$e^{ct}W_2^c(t) \to d(c) \quad \text{as} \quad t \to \infty \tag{5.11}$$

for all positive c, where $d(c)$ is some constant depending on c with $0 \leq d(c) \leq \infty$. (Clearly for all but at most one choice of c we must have $d(c) = 0$ or ∞.) Given that (5.11) holds,

$$s\hat{W}_2^c(s-c) \to d(c) \quad \text{as} \quad s \to 0 \tag{5.12}$$

for each positive c by the final value theorem for Laplace transforms. As a special case of (5.12), we have

$$sW_2^c(s-\eta) \to \alpha \quad \text{as} \quad s \to 0$$

for the η satisfying (5.9) and some α, $0 \leq \alpha \leq \infty$. However, by L'Hospital's rule,

$$s\hat{W}_2^c(s-\eta) = s\frac{(1-\hat{w}_2(s-\eta))}{s-\eta} = \frac{s\rho(1-\hat{f}(s-\eta))}{(s-\eta)(1-\rho\hat{f}(s-\eta))} \to \alpha$$

for α in (5.5). Since $\eta < s^*$, $\hat{f}(s)$ is analytic at $-\eta$, so that the derivative $\hat{f}'(-\eta)$ is finite. With transforms, in a specific application we could verify condition (5.11) by examining the singularities of $\hat{w}_2(s)$ and showing that the rightmost singularity is an isolated pole.

(b) Again we give both probabilistic and transform arguments. In both cases, the limit (5.6) follows by a term-by-term limit in (2.17), with

$$F_n^c(t) \sim n\hat{f}(-s^*)^{n-1} F^c(t) \quad \text{as} \quad t \to \infty \tag{5.13}$$

by (5.1), just as in (3.11). The supporting technical details for the case $s^* = 0$ are given in Section IV.4 of Athreya and Ney [20]; then $\hat{f}(-s^*) = 1$ in (5.6) and the constant becomes $\rho/(1-\rho)$. The supporting technical details for the case $-s^* < 0$ are given in Chover, Ney and Wainger [28], [29]. Conditions (5.1) and (5.2) are assumptions there. The term-by-term limit is justified in these sources.

We now turn to the transform proof of part (b). Since the root equation (5.3) has no root to the right of $-s^*$, the rightmost singularity of $\hat{w}_2(s)$ is the rightmost singularity of $\hat{f}(s)$, which is $-s^*$. Thus, the desired limiting behavior of both $F^c(t)$ and $W_2^c(t)$ as $t \to \infty$ is determined by the limiting behavior of $\hat{f}(s)$ and $\hat{w}_2(s)$ as $s \downarrow -s^*$. Moreover, we must have $\hat{f}(-s^*) < \infty$, so that $-s^*$ is a branch point singularity. Hence, we are in the domain of Heaviside's operational principle. Indeed, the convolution operational principle in Section 3 implies (5.13) by induction on n. However, both the convolution operational principle and the term-by-term limit in (2.17) remain to be justified with the transform argument.

Remark 5.1. Note that the root equation (5.3) has a solution $-\eta$ for all ρ if and only if $\hat{f}(s)$ is class I, while it never has a solution if $\hat{f}(s)$ and $\hat{w}(s)$ are class III. There are two regions when $\hat{f}(s)$ is of class II. In Section 7 we will show that $\hat{f}(s)$ is typically of class II.

6 The Heavy-Traffic Limit

We now discuss the heavy-traffic limit for the steady-state low-priority waiting-time distribution obtained by letting $\rho \to 1$, which we do by letting $\rho_2 \to 1 - \rho_1$ with ρ_1 fixed. It lends extra support to the exponential asymptotics in Theorem 5.1(a). We omit the dependence on ρ_2 in our notation. Our limit here complements previous heavy-traffic limits for the stochastic processes in this model; see Harrison [42] and Whitt [53].

Theorem 6.1 *Assume that $g_{12} < \infty$ and $g_{22} < \infty$. If $\rho_2 \to 1 - \rho_1$ then*

$$W_2^c(t/w_{21}) \to e^{-t}, \; t \geq 0 \; , \tag{6.1}$$

for w_{21} in (2.18).

Proof. We use Laplace transforms, exploiting the continuity theorem; p. 431 of Feller [38]. We show that $\hat{w}_2(s/w_{21}) \to (1+s)^{-1}$ as $\rho \to 1$. Rewrite the transform (2.14) as

$$\hat{w}_2(s) = \frac{1}{1 + \dfrac{\rho(1-\hat{f}(s))}{1-\rho}} \; .$$

Then we show that

$$\frac{\rho}{1-\rho}(1 - \hat{f}((1-\rho)s) \to s \left(\frac{\rho_1(g_{12}/2g_{11})}{1-\rho_1} + (g_{22}/2g_{21}) \right) \; .$$

This follows from (2.6) and (2.8), since $h_{01}^{(1)} = g_{12}/2(1-\rho_1)$ by Theorem 6(a) of [6], $b_{11} = 1/(1-\rho_1)$,

$$\hat{h}_0^{(1)}((1-\rho)s) = 1 - (1-\rho)sh_{01}^{(1)} + O((1-\rho)^2)$$

$$z((1-\rho)s) = (1+\rho_1 b_{11})(1-\rho)s + O((1-\rho)^2)$$

$$\hat{g}_{2e}(z((1-\rho)s)) = 1 - g_{2e1}(1+\rho_1 b_{11})(1-\rho)s + O((1-\rho)^2)$$

as $\rho_2 \to 1-\rho_1$. ■

The approximation suggested by Theorem 6.1 is

$$W_2^c(t) \approx \rho e^{-\rho t/w_{21}} \ , \quad t > 0 \ , \tag{6.2}$$

for w_{21} in (2.18). In (6.2) we have refined the heavy-traffic approximation obtained directly from (6.1) by accounting for the known probability of emptiness of $1-\rho$. However, unlike FIFO, the heavy-traffic approximation (6.2) is not exact for the usual exponential special cases. (We will give exact results for exponential service times later.)

Remark 6.1. We note that the heavy-traffic approximation can be extended to very general single-server queues with non-Poisson arrivals. The key observation is that the low-priority steady-state virtual waiting time satisfies the first representation in Theorem 2.1. In heavy traffic, the actual waiting time will be close to the virtual waiting time. In heavy traffic, the total workload of both classes will tend to be large. When the initial level of work, say x, is very large, then the high-priority first passage time to 0 is approximately $x/(1-\rho_1)$ by the law of large numbers. (For the M/G/1 model this is always the expected value; see Theorem 7 of [6].) Hence, in heavy traffic, we have the approximation

$$W_2^c(x) \approx V^c((1-\rho_1)x) \tag{6.3}$$

where V is the steady-state workload cdf for both classes. Thus, standard heavy-traffic approximations for V (ignoring the priority structure) translate into heavy-traffic approximations for W_2. Theorem 6.1 and approximation (6.2) can be regarded as special cases of this argument.

Remark 6.2. By (6.2), ρ/w_{21} is a natural candidate to serve as an initial guess for the asymptotic parameter η in the Newton-Raphson method for finding a root of the equation $1-\rho\hat{f}(s) = 0$ in (5.3).

As in [7] and Section 3 of [12], we can develop a heavy-traffic expansion in powers of $1-\rho$ for the asymptotic parameters η and α in the exponential asymptotics in (5.4) by doing a Taylor series expansion of $\hat{f}(s)$. We thus obtain the following result by a minor modification of the proof of Theorem 6.1.

Theorem 6.2. *Assume that the condition of Theorem 5.1(a) holds. Then, as $\rho_2 \to 1-\rho_1$,*

$$\eta = w_{21}^{-1} + O((1-\rho)^2) \quad \text{as} \quad \rho \to 1 \tag{6.4}$$

for w_{21} in (2.18) and

$$\alpha = 1 + O((1-\rho)) \quad \text{as} \quad \rho \to 1 \ . \tag{6.5}$$

Theorems 6.1 and 6.2 imply that the two iterated limits involving $\rho_2 \to 1-\rho_1$ and $t \to \infty$ agree.

7 The Busy-Period Asymptotics Equation

Further detail about the asymptotics for $W_2^c(t)$ in Theorem 5.1 primarily depends upon asymptotics for the high-priority first-passage-time and busy-period distributions. We draw upon asymptotic results established by saddle point methods by Cox and Smith [31]. We restate their result in the form of an asymptotic expansion and give additional details about the proof. (An asymptotic expansion is defined in (1.5) and (1.6).)

Note that the reciprocal of the asymptotic decay rate for the busy period is the *relaxation time* for the total workload process; see III.7.3 of Cohen [30]. We let τ_1 represent this relaxation time.

Theorem 7.1. (Cox and Smith [31]) *If there is a negative real root $-\zeta_1$ of the busy-period asymptotics equation*

$$-\hat{g}_1'(s) = \frac{1}{\rho_1} \ , \tag{7.1}$$

where $-\zeta_1$ is to the right of all singularities of $\hat{g}_1(s)$, then $f_{x0}^1(t)$ and $b_1(t)$ have asymptotic expansions

$$f_{x0}^{(1)}(t) \sim xe^{\zeta x}\rho_1\alpha_1(\pi t^3)^{-1/2}e^{-t/\tau_1}(1+\sum_{i=1}^{\infty}\beta_i t^{-i}) \quad \text{as} \quad t\to\infty \tag{7.2}$$

for all $x > 0$ and

$$b_1(t) \sim \alpha_1(\pi t^3)^{-1/2}e^{-t/\tau_1}(1+\sum_{i=1}^{\infty}\beta_i t^{-i}) \quad \text{as} \quad t\to\infty \ , \tag{7.3}$$

where

$$\tau_1^{-1} = \rho_1 + \zeta_1 - \rho_1\hat{g}_1(-\zeta_1) \ , \tag{7.4}$$

$$\alpha_1 = [2\rho_1^3\hat{g}_1''(-\zeta_1)]^{-1/2} \tag{7.5}$$

β_i is a constant for each i with

$$\beta_1 = -\alpha_1^4\rho_1^5\left[\frac{5}{3}\alpha_1^2\rho_1^3\hat{g}_1^{(3)}(-\zeta_1)^2 - \frac{\hat{g}_1^{(4)}(-\zeta_1)}{2}\right] \ . \tag{7.6}$$

Proof. The asymptotic expansions (7.2) and (7.3) are obtained in the same way from integral representations, so we only discuss (7.3). Note that 7.3 is obtained formally from (7.2) by integrating with respect to G_1, i.e.,

$$b_1(t) = \int_0^\infty f_{x0}^{(1)}(t)dG_1(x) \ , \tag{7.7}$$

using condition (7.1). Cox and Smith [31] established a contour integral representation for $b_1(t)$, namely

$$b_1(t) = \frac{1}{2\pi\rho_1 ti}\int_{a-i\infty}^{a+i\infty} e^{t\phi_1(s)}ds \ , \tag{7.8}$$

where

$$\phi_1(s) = s - \rho_1 + \rho_1\hat{g}_1(s) \ . \tag{7.9}$$

(We do not elaborate on the derivation of (7.8) because it is explained fully in [31] and reviewed in [13].) The contour in (7.8) is a vertical line with $Re(s) = a$ such that $\hat{g}_1(s)$ has no singularities on or to the right of it. To establish (7.3), we apply Laplace's method or the saddle point method as on pages 80, 121–127 of Olver [46]. We elaborate on this point, because it is not explained in [31]. Since the service-time transform $\hat{g}_1(s)$ is generic, it is not possible to exploit specific structure beyond the fact that it is a probability transform, defined as in (2.1). Assuming that (7.1) holds, we see that $\phi(-\zeta) < 0$, $\phi'(-\zeta) = 0$ and $\phi_1''(-\zeta) > 0$, so that for large t the integral in (7.8) is dominated by its behavior in the neighborhood of $-\zeta$. Since $-\zeta$ is to the right of all singularities of $\hat{g}_1(s)$ and thus of $\phi_1(s)$, we can change the contour without changing the value of the integral. It is convenient to replace the contour by three pieces: two vertical lines $-\zeta - \epsilon - \omega i$ and $-\zeta + \epsilon + \omega i$ for $\omega > 0$ and the closed interval $[-\zeta - \epsilon\ ,\ -\zeta + \epsilon]$ for suitably small positive ϵ (proceeding continuously starting from $-\zeta - \epsilon - i\infty$). It turns out that we can neglect the integral over the two vertical lines, because

$$|e^{t\phi_1(s)}| = e^{tRe(\phi_1(s))} \le e^{t\phi_1(Re(s))}\ , \tag{7.10}$$

for $\phi_1(s)$ in (7.9). To establish the inequality in (7.10), let $s = \sigma + i\omega$ and note that

$$\begin{aligned} Re(\phi_1(s)) &= \sigma - \rho + \rho \int_0^\infty e^{-\sigma x} \cos(\omega x) dG_1(x) \\ &\le \sigma - \rho + \rho \int_0^\infty e^{-\sigma x} dG(x) = \phi_1(Re(s))\ . \end{aligned} \tag{7.11}$$

Hence, for this problem, the saddle point problem in the complex plane reduces to Laplace's method on the real line, as discussed on pp. 80–86 of Olver [46]. Then (7.3) follows from taking a Taylor series expansion of $\phi_1(s)$, i.e., $\phi_1(s) \approx \phi_1(-\zeta) + \phi_1''(-\zeta)\frac{(s+\zeta)^2}{2}$ for real s; see p. 80 of [18] for details. See pp. 86 and 127 of [46] for the asymptotic expansion. ∎

Corollary 7.1. *Under the conditions of Theorem 7.1, the rightmost singularity of* $\hat{b}_1(s)$ *is* $-\tau_1^{-1}$ *and*

$$\hat{b}_1(-\tau_1^{-1}) = 1 + \rho_1^{-1}(\zeta_1 - \tau_1^{-1}) = \hat{g}_1(-\zeta_1)\ ,$$

so that $b_1(t)$ *is a class II distribution.*

Remark 7.1. Note that a root $-\zeta_1$ exists for the busy-period equation (7.1) whenever the service-time cdf $G_1(t)$ is class I. When $G_1(t)$ class II, a root $-\zeta_1$ for (7.1) exists if and only if $-\hat{g}_1'(-s_1^*) > 1/\rho_1$, where $\hat{g}_1'(-s_1^*)$ is the right derivative at the singularity $-s_1^*$.

We combine Theorem 7.1 and Lemmas 3.4 and 3.5 to obtain the following corollary. (This corollary is not stated correctly in [31]. Recall that we have assumed that $g_{11} = 1$.)

Corollary 7.2. *Under the assumptions of Theorem 7.1,*

$$B_1^c(t) \sim \tau_1 b_1(t) \quad \text{as} \quad t \to \infty \tag{7.12}$$

and

$$B_{1e}^c(t) \sim (1-\rho_1)\tau_1 B_1^c(t) \sim (1-\rho_1)\tau_1^2 b_1(t) \quad \text{as} \quad t \to \infty\ . \tag{7.13}$$

For the non-exponential asymptotics in (5.6), we need to determine the asymptotics for the ccdf $F^c(t)$ in (2.16). However, we have already determined the asymptotics for the component $H_0^{(1)c}(t)$ in our previous treatment of M/G/1 LIFO waiting times, assuming that the busy-period root equation (7.1) has the zero, which is by essentially the same argument as in Theorem 7.1. By (37) of [12] or by (2.7) and Theorem 3.1 of [10],

$$h_0^{(1)}(t) \sim (\alpha_1/\tau_1\zeta_1^2)(\pi t^3)^{-1/2}e^{-t/\tau_1} \quad \text{as} \quad t \to \infty \tag{7.14}$$

and, by Lemma 3.5,

$$H_0^{(1)c}(t) \sim (\alpha_1/\zeta_1^2)(\pi t^3)^{-1/2}e^{-t/\tau_1} \quad \text{as} \quad t \to \infty\ . \tag{7.15}$$

Since $H_0^{(1)c}(t)$ is related to the probability of emptiness by (2.5), (7.15) also follows from asymptotic results for the probability of emptiness in the GI/G/1 queue on p. 609 of Cohen [30].

More generally, we can relate the asymptotics of $H_0^{(1)c}(t)$ to the asymptotics of $B_e^c(t)$, by the same reasoning used in Theorem 5.1(b).

Theorem 7.2. *Let $-s^*$ be the rightmost singularity of $\hat{h}_0^{(1)}(s)$. (a) If $1-\rho_1\hat{h}_0^{(1)}(s)$ has no real root in $(-s^*,0)$ and if $H_0^{(1)c}(t)$ satisfies (5.1) and (5.2), then*

$$B_{1e}^c(t) \sim \frac{1-\rho_1}{(1-\rho_1\hat{h}_0^{(1)}(-s^*))^2}H_0^{(1)c}(t) \quad \text{as} \quad t \to \infty\ . \tag{7.16}$$

If in addition $s^*=0$, then

$$H_0^{(1)c}(t) \sim (1-\rho_1)B_{1e}^c(t) \quad \text{as} \quad t \to \infty\ . \tag{7.17}$$

Proof. Note that we can rewrite (2.6) as

$$\hat{b}_{1e}(s) = \frac{(1-\rho_1)\hat{h}_0^{(1)}(s)}{1-\rho_1\hat{h}_0^{(1)}(s)} = (1-\rho_1)\sum_{n=0}^{\infty}\rho_1^n h_0^{(1)}(s)^{n+1}\ , \tag{7.18}$$

so that we have a minor modification of the geometric random sum representation and we can apply the argument in Theorem 5.1(b). ∎

We now characterize the rightmost singularity of $\hat{h}_0^{(1)}(s)$ and $\hat{f}(s)$ under the assumption of Theorem 7.1 and determine when the exponential asymptotics holds. In our statement we use the inverse function of $z_1(s)$ for negative real s, denoted by $z_1^{-1}(s)$, which can be expressed as

$$z_1^{-1}(s) = s - \rho_1 + \rho_1\hat{g}_1(s)\ . \tag{7.19}$$

By (2.4), $z_1^{-1}(z_1(s)) = s$ for any complex s; see (29) and (30) of [6].

Theorem 7.3. *Assume the condition of Theorem 7.1.*

(a). *Then $-\tau_1^{-1}$ is the rightmost singularity of $\hat{h}_0^{(1)}(s)$,*

$$\hat{h}_0^{(1)}(-\tau_1^{-1}) = \frac{1}{\rho_1}(1-(\zeta_1\tau_1)^{-1}) \tag{7.20}$$

and $1-\rho_1\hat{h}_0^{(1)}(s)$ has no root in $(-\tau_1^{-1},0)$.

(b). *Let $-s_2^*$ and $-s^*$ be the rightmost singularities of $\hat{g}_2(s)$ and $\hat{f}(s)$, respectively. Then*

$$s^* = \begin{cases} \tau_1^{-1} , & \zeta_1 < s_2^* \\ -z_1^{-1}(-s_2^*) > 0 , & \zeta_1 > s_2^* > 0 \\ 0, & s_2^* = 0 \end{cases} \tag{7.21}$$

(c) *If in addition $s_2^* > 0$, then the root equation (5.3) has a negative real zero $-\eta$ with $\eta < s^*$, so that the exponential asymptotics in (5.4) holds for all $\rho > \rho_2^*$, where*

$$\rho_2^* = \begin{cases} \frac{g_{21}\tau_1^{-1}}{\hat{g}_2(-\zeta_1)-1} , & s_2^* > \zeta_1 \\ 0 , & s_2^* < \zeta_1 \text{ and } \hat{g}_2(-s_2^*) = \infty \\ \frac{1-\rho_1\hat{h}_0^{(1)}(z_1^{-1}(-s_2^*))}{\hat{g}_{2e}(-s_2^*)} , & s_2^* < \zeta_1 \text{ and } \hat{g}_2(-s_2^*) < \infty . \end{cases} \tag{7.22}$$

Proof. (a) By (7.15), $H_0^{(1)c}(t) \sim (1/\tau_1\zeta_1^2)B_1^c(t)$ as $t \to \infty$, so that $-\tau_1^{-1}$ is the rightmost singularity of $\hat{h}_0^{(1)}(s)$, just like of $\hat{b}_1(s)$. Moreover, $1-\rho_1\hat{h}_0(s)$ must not have a zero in the interval or else $\hat{b}_{1e}(s)$ would have exponential asymptotics, by virtue of (7.18), using the reasoning of Theorem 5.1. By (7.16),

$$\frac{1}{\tau_1\zeta_1^2} = \tau_1(1-\rho_1\hat{h}_0^{(1)}(-\tau_1^{-1}))^2 ,$$

which implies (7.20). By (7.20), $\rho_1\hat{h}_0^{(1)}(-\tau^{-1}) < 1$. Since $\hat{h}_0^{(1)}(s)$ is monotone, we have a second direct derivation showing that $\rho_1\hat{h}_0^{(1)}(s) - 1$ has no root in the interval $(-\tau_1^{-1}, 0)$.

(b) By (2.4), (2.8) and (7.19),

$$\hat{b}_1(-\tau_1^{-1}) = 1 + \rho_1^{-1}(\zeta_1 - \tau_1^{-1}) \tag{7.23}$$

and

$$z_1(-\tau_1^{-1}) = -\zeta_1 , \tag{7.24}$$

from which we obtain (7.21).

(c) First suppose that $s_2^* > \zeta_1$, so that the rightmost singularity of $\hat{f}(s)$ is $-\tau_1^{-1}$. To find the boundary point ρ_2^* in (7.22) we solve for ρ_2^* in the equation

$$1 - \rho_1\hat{h}_0^{(1)}(-\tau_1^{-1}) - \rho_2^*\hat{g}_{2e}(z_1(-\tau_1^{-1})) = 0 . \tag{7.25}$$

Combining (7.24), (7.25) and (7.25) yields the first case of (7.22).

Since $\hat{g}_{2e}(z_1(-s))$ is increasing in real s, if $s_2^* < \zeta_1$, then the rightmost singularity of $\hat{f}(s)$ is $z_1^{-1}(-s_2^*)$. Then the equation $1-\rho\hat{f}(s) = 0$ will have a root $-\eta$ with $\eta < \tau_1^{-1}$ for all ρ_2 provided that $\hat{g}_{2e}(-s_2^*) = \infty$, which occurs if $\hat{g}_2(-s_2^*) = \infty$. If $\hat{g}_2(-s_2^*) < \infty$, then $\hat{g}_{2e}(-s^*) < \infty$ and the boundary for the existence of the root η is the equation

$$1 - \rho_1 h_0^{(1)}(z_1^{-1}(-s_2^*)) - \rho_2^*\hat{g}_{2e}(-s_2^*) = 0 ,$$

which yields the third case of (7.22). ∎

Remark 7.2. Note that exponential asymptotics holds for all ρ_2 (i.e., $\rho_2^* = 0$) when $s_2^* < \zeta_1$ and $\hat{g}_2(-s_2^*) = \infty$. However, if $\hat{g}_2(s)$ has no singularities (as with a finite mixture of atoms), then ρ_2^* is never 0. ∎

We now establish the non-exponential asymptotics in Theorem 6.1(b) in the main case. The next three results are used in our proof of Theorem 7.4 below.

Lemma 7.1. *Under the conditions of Theorem 7.1,*

$$\tau_1^{-1} < (1-\rho_1)\zeta_1 \ .$$

Proof. Since $\hat{g}_1(-s)$ is strictly convex in s,

$$\frac{\hat{g}_1(-\zeta_1) - \hat{g}(0)}{\zeta_1} > -\hat{g}_1'(0) = 1 \ .$$

Since $\hat{g}(0) = 1$, combining this with (7.4) completes the proof. ∎

Our next lemma establishes a regularity property about $F_{x0}^{(1)c}(t)$ for large x. We can apply the law of large numbers to show that the first passage time is likely to occur near time $x/(1-\rho_1)$ for large x. We use a large deviations argument to establish an exponential bound refinement.

Lemma 7.2. *Under the conditions of Theorem 7.1, for all positive ϵ,*

$$F^{(1)c}_{(1-\rho_1)(1-\epsilon)t,0}(t) \le e^{-(1-\rho_1)(1-\epsilon)t\ell(\epsilon)} \ , \quad t > 0 \ , \tag{7.26}$$

and

$$F^{(1)}_{(1-\rho_1)(1+\epsilon)t,0}(t) \le e^{-(1-\rho_1)(1+\epsilon)t\ell(\epsilon)} \ , \quad t > 0 \ , \tag{7.27}$$

where

$$\ell(\epsilon) = \sup_s \{s\epsilon - z_1(s)\} \to \zeta_1 \text{ as } \epsilon \downarrow 0 \tag{7.28}$$

for $z_1(s)$ in (2.8), so that for all positive ϵ sufficiently small

$$F^{(1)c}_{(1-\rho_1)(1-\epsilon)t,0}(t) = o(F_{x0}^{(1)c}(t)) = o(e^{-t/\tau_1}) \text{ as } t \to \infty \tag{7.29}$$

and

$$F^{(1)}_{(1-\rho_1)(1+\epsilon)t,0}(t) = O(e^{-t/\tau_1}) \quad \text{as} \quad t \to \infty \ . \tag{7.30}$$

Proof. Recall that $F^{(1)}_{nx,0}(t)$ is the n-fold convolution of $F^{(1)}_{x,0}(t)$ for all positive x and positive integers n, that the mean of $F^{(1)}_{x0}(t)$ is $x/(1-\rho_1)$ (e.g., see Theorem 7 of [6]), and that the Laplace transform of $f^{(1)}_{x0}(t)$ is given in (2.7). Hence, we can apply Chernoff's bound, e.g., (1.6a) on p. 14 of Shwartz and Weiss [48], to obtain (7.26) and (7.27). By (29) and (30) of [6] and (7.23) and (7.24), we see that the root of $z_1'(-s) = 0$ is ζ_1, which implies that $\ell(\epsilon) \to \zeta_1$ as $\epsilon \downarrow 0$. We apply Lemma 7.1 to obtain (7.29) and (7.30). ∎

We now obtain control of the asymptotics of the second term of $F^c(t)$ in (2.16) under regularity conditions.

Lemma 7.3. *Under the conditions of Theorem 7.1, if G is any cdf on the nonnegative real line such that*

$$G^c(ct) = o(e^{-t/\tau_1}) \quad \text{as} \quad t \to \infty \tag{7.31}$$

for $c < 1 - \rho_1$, then

$$\begin{aligned}\int_0^\infty F_{x0}^{(1)c}(t)dG(x) &\sim -\hat{g}'(-\zeta_1)F_{10}^c(t) \\ &\sim -\hat{g}'(-\zeta_1)\rho_1\tau_1\alpha_1(\pi t^3)^{-1/2}e^{-t/\tau_1} \text{ as } t \to \infty\ . \end{aligned} \tag{7.32}$$

Proof. By Fatou's lemma, (7.2) and Lemma 3.5,

$$\liminf_{t\to\infty} F_{10}^{(1)c}(t)^{-1}\int_0^\infty F_{x0}^{(1)c}(t)dG(x) \geq \int_0^\infty xe^{\zeta_1 x}dG(x) = -\hat{g}'(-\zeta_1)\ . \tag{7.33}$$

Thus it suffices to establish an upper bound for the lim sup. For this purpose, we break up the integral into three parts, as

$$\int_0^\infty = \int_0^N + \int_{N\vee ct}^\infty + \int_N^{N\vee ct} \tag{7.34}$$

letting $c = (1-\rho_1)(1-\epsilon)$ for small positive ϵ. We bound $F_{x0}^{(1)c}(t)$ above by 1 in the second integral in (7.34) and we apply Lemma 7.2 to show that the third integral in (7.34) is negligible. In particular, for the third integral, $F_{x0}^{(1)c}(t) \leq F_{ct,0}^{(1)c}(t)$ for all x and $\int_N^{N\vee cd} dG(x) \leq 1$. Thus,

$$\begin{aligned}\limsup_{t\to\infty} F_{10}^{(1)c}(t)^{-1}\int_0^\infty F_{x0}^{(1)c}(t)dG(x) &\leq \int_0^N xe^{\zeta x}dG(x) + \limsup_{t\to\infty}\left\{\frac{G^c(ct)}{F_{10}^{(1)c}(t)} + \frac{F_{ct,0}^{(1)c}(t)}{F_{10}^{(1)c}(t)}\right\} \\ &\leq \int_0^\infty xe^{\zeta x}dG(x)\ . \ \blacksquare\end{aligned}$$

Theorem 7.4. *If the conditions of Theorem 7.1 hold and if $s_2^* > \zeta_1$, where $-s_2^*$ is the rightmost singularity of the low-priority service-time transform $\hat{g}_2(s)$, then*

$$F^c(t) \sim \left(\frac{\rho_1\zeta_1^{-2} - \rho_2\rho_1\tau_1\hat{g}_{2e}'(-\zeta_1)}{\rho_1+\rho_2}\right)\alpha_1(\pi t^3)^{-1/2}e^{-t/\tau_1} \quad \text{as} \quad t\to\infty \tag{7.35}$$

and, if $\rho_2 < \rho_2^*$ for $\rho_2^* = g_{21}/\tau_1(\hat{g}_2(-\zeta_1)-1)$ as in (7.22), then

$$W_2^c(t) \sim \frac{(1-\rho)\rho_1\tau_1^2[1-\rho_2\tau_1\zeta_1\hat{g}_2'(-\zeta_1)g_{21}^{-1} - (\rho_2/\rho_2^*)]}{[1-(\rho_2/\rho_2^*)]^2}\ \frac{\alpha_1 e^{-t/\tau_1}}{\sqrt{\pi t^3}} \quad \text{as} \quad t\to\infty \tag{7.36}$$

for ζ_1 in (7.1), τ_1^{-1} in (7.4) and α_1 in (7.5).
Proof. By Theorem 7.3(b), the rightmost singularity of $\hat{f}(s)$ is $-\tau_1^{-1}$. The asymptotics for $H_0^{(1)c}(t)$ is given in (7.15). The asymptotics for the second term of F in (2.16) is given by Lemma 7.3. Since $s_2^* > \zeta_1$, by assumption, and $(1-\rho_1)\zeta_1 > \tau_1^{-1}$, by Lemma 7.1, condition (7.31) holds for $G = G_{2e}$. Since $\hat{g}_{2e}(s) = (1-\hat{g}_2(s))/g_{21}s$,

$$-\hat{g}_{2e}'(-\zeta_1) = \frac{1-\hat{g}_2(-\zeta_1) - \zeta_1\hat{g}_2'(-\zeta_1)}{g_{21}\zeta_1^2}\ . \tag{7.37}$$

Finally, we apply Theorem 5.1(b), (7.20) and (7.22) to obtain (7.36). In this case we can directly verify conditions (5.1) and (5.2); condition (5.1) follows from Lemma 3.6. $\blacksquare$

Remark 7.3. Note that the asymptotic decay rate τ_1^{-1} in (7.36) does not depend on ρ_2 provided that $\rho_2 < \rho_2^*$.
Remark 7.4. Note that the asymptotic constant in (7.36) explodes as $\rho_2 \uparrow \rho_2^*$.

We have just noted that the non-exponential asymptote in (7.36) becomes useless as $\rho_2 \uparrow \rho_2^*$, because the asymptotic constant explodes. Our next result shows that the exponential asymptote in (5.4) also becomes useless as $\rho_2 \downarrow \rho_2^*$. We need the following lemma in our proof of Theorem 7.5 below.
Lemma 7.4. *If the conditions of Theorem 7.1 hold, then* $\hat{b}_1'(-\tau_1^{-1}) = -\infty$.
Proof. By Kendall's equation (2.4),

$$\hat{b}_1'(s) = \hat{g}_1'(z_1(s))(1 - \rho_1 \hat{b}_1'(s)) \tag{7.38}$$

Since $z(-\tau_1^{-1}) = -\zeta_1$ and $\hat{g}_1'(-\zeta_1) = -\rho_1^{-1}$, equation (7.38) approaches

$$-\rho_1 \hat{b}_1'(-\tau_1^{-1}) = 1 - \rho_1 \hat{b}_1'(-\tau_1^{-1})$$

as $s \downarrow -\tau_1^{-1}$. However, $x = \infty$ is the only solution to the equation $x = x + 1$. ∎

Theorem 7.5. *Assume that the conditions of Theorem 7.1 hold with* $s_2^* > \zeta_1$, *so that* $-\tau_1^{-1}$ *is the rightmost singularity of* $\hat{f}(s)$ *as in (7.21). If* $\rho_2 \downarrow \rho_2^*$, *then* $\alpha \to 0$ *for* α *in (5.5).*
Proof. As $\rho_2 \downarrow \rho_2^*$, $\eta \downarrow \tau_1^{-1}$. Hence, by (5.5), it suffices to show that $\hat{f}'(-\tau_1^{-1}) = -\infty$, where all derivatives at $-\tau_1^{-1}$ are understood to be right derivatives. First, by Lemma 7.4, $\hat{b}_1'(-\tau_1^{-1}) = -\infty$. Then, since

$$\hat{h}_0^{(1)}(s) = \frac{1}{\rho_1} - \frac{s}{\rho_1 z_1(s)} \ ,$$

we have

$$\frac{d}{ds}\hat{h}_0^{(1)}(s) = -\frac{-1}{\rho_1 z_1(s)} + \frac{s}{\rho_1 z_1(s)^2}(1 - \rho_1 \hat{b}_1'(s)) \ .$$

Since $z_1(-\tau_1^{-1}) = -\zeta_1$ and $\hat{b}_1'(-\tau_1^{-1}) = -\infty$, $\hat{h}_0^{(1)'}(-\tau_1^{-1}) = -\infty$, which implies that $\hat{f}'(-\tau_1^{-1}) = -\infty$. ∎

We conclude this section by treating the third case in Theorem 7.3 in which $s_2^* < \zeta_1$, $\hat{g}_2(-s_2^*) < \infty$ and $\rho_2 < \rho_2^*$.
Theorem 7.6. *Assume that the condition of Theorem 7.1 holds, but that* $-\zeta_1 < -s_2^*$, *where* $-s_2^*$ *is the rightmost singularity of* $\hat{g}_2(s)$ *and* $\hat{g}_{2e}(-s_2^*) < \infty$, *so that*

$$\hat{g}_{2e}(s) \sim \hat{g}_{2e}(-s_2^*) + \hat{\theta}_{2e}(s) \quad \text{as} \quad s \to -s_2^* \ . \tag{7.39}$$

then

$$\hat{w}_2(s) \sim \frac{1-\rho}{A_0 - \rho_2 \hat{\theta}(s)} \sim \frac{1-\rho}{A_0} + \frac{(1-\rho)\rho_2 \hat{\theta}(s)}{A_0^2} \quad \text{as} \quad s \to -s_2^* \tag{7.40}$$

for

$$\hat{\theta}(s) = \hat{\theta}_{2e}(z_1'(z_1^{-1}(-s_2^*))) \ , \tag{7.41}$$

$$A_0 = 1 - \rho_1 \hat{h}_0^{(1)}(z_1^{(-1)}(-s_2^*)) - \rho_2 \hat{g}_{2e}(-s_2^*) = \hat{g}_{2e}(-s_2^*)(\rho_2^* - \rho_2) \ , \tag{7.42}$$

$z_1^{-1}(s)$ in (7.19) and ρ_2^* in (7.22). If, in addition,

$$\hat{\theta}_{2e}(s) \sim A(s+s_2^*)^{\gamma} \quad as \quad s \to -s_2^* \tag{7.43}$$

for some positive non-integer γ and $\hat{g}_{2e}(s)$ has no singularities other than $-s_2^*$ with $Re(s) > -s_2^* - \epsilon$ for some $\epsilon > 0$, then

$$W_2^c(t) \sim \frac{(1-\rho)\rho_2}{A_0^2} G_{2e}^c \left(\frac{t}{z_1'(z_1^{-1}(-s_2^*))} \right) \quad as \quad t \to \infty \ . \tag{7.44}$$

If, in addition $-s_2^* = 0$, then

$$W_2^c(t) \sim \frac{\rho_2}{1-\rho} G_{2e}^c((1-\rho_1)t) \quad \text{as} \quad t \to \infty \ . \tag{7.45}$$

Proof. Expand $\hat{g}_{2e}(z_1(s))$ around $z_1^{-1}(-s_2^*)$. Then

$$\hat{g}_{2e}(z_1(s)) \sim \hat{g}_{2e}(-s_2^*) + \hat{\theta}(s) \quad \text{as} \quad s \to -s_2^*$$

and

$$1 - \rho \hat{f}(s) \sim A_0 - \rho_2 \hat{\theta}(s) \quad \text{as} \quad s \to -s_2^*$$

for $\hat{\theta}(s)$ in (7.41) and A_0 in (7.42), from which we obtain (7.40). Under the extra conditions on $\hat{\theta}_{2e}(s)$ in (7.43) and on the singularities of $\hat{g}_{2e}(s)$, we can apply Heaviside's theorem on p. 254 of Doetsch [34] and Sutton [50] to obtain the asymptotic relation for the density in the time domain, which implies (7.44) by Lemma 3.4. Finally, for (7.45), note that $z_1^{-1}(0) = 0$ and

$$z_1'(0) = 1 - \rho_1 b_1'(0) = 1 - \rho_1 \left(\frac{-1}{1-\rho_1} \right) = \frac{1}{1-\rho_1} \ .$$

Also note that $A_0 = 1 - \rho$ for A_0 in (7.42) when $-s_2^* = 0$. ∎

Remark 7.5. By the Heaviside operational principle in Section 3, we conclude (7.44) and (7.45) without imposing the condition (7.43), but this step remains to be justified.

8 Asymptotic Expansions from Kendall's Functional Equation

In Theorem 7.1 we obtained an asymptotic expansion for the high-priority busy-period pdf $b_1(t)$ by using saddle point methods with an integral representation. This enabled us to obtain the asymptote for the low-priority waiting-time ccdf $W_2^c(t)$, but this required rather difficult arguments in the time domain, such as Lemma 7.2. In this section we show that asymptotics for $W_2^c(t)$ can be developed more easily starting from an asymptotic expansion for the high-priority busy-period transform $\hat{b}_1(s)$.

We obtain our asymptotic expansion for $\hat{b}_1(s)$ from Kendall's functional equation (2.4). If we assume that the rightmost singularity of $\hat{b}_1(s)$ is a branch point singularity at $s = -\tau_1^{-1}$ and that $b_1(s)$ no other singularities for $Re(s) \geq -\tau_1^{-1} - \epsilon$ for some positive ϵ, and if we also assume that $\hat{b}_1(s)$ has an asymptotic expansion in powers of $(s + \tau_1^{-1})$, i.e.,

$$\hat{b}_1(s) \sim \sum_{k=0}^{\infty} a_k (s + \tau_1^{-1})^{\gamma_k} \quad \text{as} \quad s \to -\tau_1^{-1} \ , \tag{8.1}$$

where $\gamma_k < \gamma_{k+1}$ for all k, then we can apply Theorem 7.1 and Heaviside's theorem on p. 254 of Doetsch [34] and Sutton [50] to conclude that $\hat{b}_1(s)$ must have an asymptotic expansion in half powers, i.e.,

$$\hat{b}_1(s) \sim \sum_{k=0}^{\infty} a_k (s + \tau_1^{-1})^{k/2} \quad \text{as} \quad s \to -\tau_1^{-1} \ , \tag{8.2}$$

when the busy-period root equation (7.1) has a root to the right of all singularities to the high-priority service-time transform $\hat{g}_1(s)$. The terms in (8.1) with non-integer powers must have counterparts in the time-domain expansion in Theorem 7.1. As indicated in Doetsch [34], the terms $a_{2k}(s + \tau_1^{-1})^k$ in (8.1) and (8.2) with whole integer powers are analytic functions and thus do not contribute to the asymptotic expansion in the time domain. Since the terms with integer powers in (8.2) have no counterpart in the time domain, evidently it is not possible to obtain the transform asymptotic expansion (8.2) directly from the time-domain asymptotic expansion established in Theorem 7.1, but that is an unresolved point.

We now directly establish the asymptotic expansion in half powers in (8.2). Thus, we obtain an asymptotic expansion for $\hat{b}_1(s)$ and an alternative proof of Theorem 7.1.

To carry out one step of the proof of Theorem 8.1 below and to understand how Theorems 7.1 and 8.1 are related, we need to know how Heaviside's theorem converts asymptotics for transforms into asymptotics in the time domain. Let $\mathcal{L}^{-1}(\hat{f}(s))$ denote the inverse of the transform $\hat{f}(s)$. Then

$$\mathcal{L}^{-1}(\sqrt{s + \tau_1^{-1}}) \sim \frac{-e^{-t/\tau_1}}{2\sqrt{\pi t^3}} \quad \text{as} \quad t \to \infty \tag{8.3}$$

and

$$\mathcal{L}^{-1}((s + \tau_1^{-1})^{3/2}) \sim \frac{3}{4} \frac{e^{-t/\tau_1}}{\sqrt{\pi t^5}} \quad \text{as} \quad t \to \infty \ ; \tag{8.4}$$

see p. 254 of Doetsch [34]. In particular, note the sign change in (8.3).

Theorem 8.1. *If the busy-period equation $-\rho_1 \hat{g}_1'(s) = 1$ in (7.1) has a root $-\zeta_1$ with $-s_1^* < -\zeta_1$, where $-s_1^*$ is the rightmost singularity of $\hat{g}_1(s)$, then the half-power asymptotic expansion (8.2) is valid and the first four coefficients in (8.2) are*

$$a_0 = \hat{g}_1(-\zeta_1) \ , \tag{8.5}$$

$$a_1 = \frac{-2}{\sqrt{2\rho_1^3 \hat{g}_1^{(2)}(-\zeta_1)}} = -2\alpha_1 \ , \tag{8.6}$$

$$a_2 = \frac{1}{\rho_1} + \frac{(a_1\rho_1)^4}{12} \hat{g}_1^{(3)}(-\zeta_1) \ , \tag{8.7}$$

and

$$a_3 = \frac{-(a_1\rho_1)^5}{48} \left[\hat{g}_1^{(4)}(-\zeta_1) - \frac{5}{6} a_1^2 \rho_1^3 \hat{g}_1^{(3)}(-\zeta_1)^2 \right] \tag{8.8}$$

for α_1 in (7.5).

Proof. Given Theorem 7.1, we anticipate that the transform $\hat{b}_1(s)$ should have asymptotic expansion (8.2). To justify the asymptotic expansion, we expand the transforms in power series in the region where they are analytic, and then let the expansion point approach the singularity. For $\epsilon > 0$, perform the change of variables $s = -\tau_1^{-1} + \epsilon + u^2$ for $Re(u) > 0$. By (2.8) and (7.4), the Kendall functional equation (2.4) becomes

$$\begin{aligned}\hat{b}_1(-\tau_1^{-1} + \epsilon + u^2) &= \hat{g}_1(z_1(-\tau_1^{-1} + \epsilon + u^2)) \\ &= \hat{g}_1(-\tau_1^{-1} + \epsilon + u^2 + \rho_1 - \rho_1 \hat{b}_1(-\tau_1^{-1} + \epsilon + u^2)) \\ &= \hat{g}_1(\epsilon + u^2 - \rho_1 \hat{b}_1(-\tau_1 + \epsilon + u^2) + \rho_1 \hat{b}_1(-\tau_1^{-1}) - \zeta_1) \ .\end{aligned} \tag{8.9}$$

Since $\hat{b}_1(s)$ is analytic at $-\tau_1^{-1} + \epsilon$ and $\hat{g}_1(s)$ is analytic at $-\zeta_1$, we can expand the functions in power series about $-\tau_1^{-1} + \epsilon$ and $-\zeta_1$, respectively, obtaining

$$\begin{aligned}\hat{b}_1(-\tau_1^{-1} + \epsilon + u^2) &= \textstyle\sum_{k=0}^{\infty} B_k(\epsilon) u^k \\ &= \textstyle\sum_{k=0}^{\infty} \dfrac{\hat{g}_1^{(k)}(-\zeta_1)}{k!} \delta(\epsilon, u)^k \ ,\end{aligned} \tag{8.10}$$

where

$$\delta(\epsilon, u) = \tilde{B}_0(\epsilon) - \rho_1 B_1(\epsilon) u + (1 - \rho_1 B_2(\epsilon)) u^2 - \sum_{j=3}^{\infty} \rho_1 B_j(\epsilon) u^j \ , \tag{8.11}$$

$B_k(\epsilon)$ are unknown coefficients depending upon ϵ and

$$\tilde{B}_0(\epsilon) = \epsilon - \rho_1 B_0(\epsilon) + \rho_1 \hat{b}_1(-\tau_1^{-1}) \ . \tag{8.12}$$

Since $\hat{b}_1(s) \to \hat{b}_1(-\tau_1^{-1})$ as $s \downarrow -\tau_1^{-1}$,

$$B_0(\epsilon) \to \hat{b}_1(-\tau_1^{-1}) \quad \text{and} \quad \tilde{B}_0(\epsilon) \to 0 \quad \text{as} \quad \epsilon \to 0 \ .$$

We now match coefficients of u^k in (8.10) for each k successively starting with $k = 0$. Considering u^0, we first obtain

$$B_0(\epsilon) = \hat{g}_1(-\zeta_1) + \sum_{k=1}^{\infty} \frac{\hat{g}_1^{(k)}(-\zeta_1)}{k!} \tilde{B}_0(\epsilon)^k = \hat{g}_1(-\zeta_1 + \tilde{B}_0(\epsilon)) \ . \tag{8.13}$$

Since $\hat{g}_1(s)$ is analytic at $-\zeta_1$, it is continuous. Since $\tilde{B}_0(\epsilon) \to 0$ as $\epsilon \to 0$ we see that $B_0(\epsilon) \to \hat{g}_1(-\zeta_1)$ as $\epsilon \to 0$. This implies that $\hat{b}_1(-\tau_1^{-1}) = \hat{g}_1(-\zeta_1)$, which is consistent with (7.4), (7.23) and (7.24).

Moreover, from (8.13), we see that $B_0(\epsilon)$ is the solution to the equation

$$x = \hat{g}_1(A - \rho_1 x) \tag{8.14}$$

for $A = -\zeta_1 + \epsilon + \rho_1 b_1(-\tau_1^{-1})$. Differentiating with respect to x in (8.14), we see that we must have

$$1 + \rho_1 \hat{g}_1^{(1)}(-\zeta_1 + \tilde{B}_0(\epsilon)) = 0 \tag{8.15}$$

for all ϵ, consistent with the assumed equation (7.1), which is the case $\epsilon = 0$.

Next, considering the coefficients of u^1 in (8.10), we obtain

$$B_1(\epsilon) = g_1^{(1)}(-\zeta_1)(-\rho_1 B_1(\epsilon)) + \sum_{k=2}^{\infty} \frac{g_1^{(k)}(-\zeta_1)}{k!} \binom{k}{1} \tilde{B}_0(\epsilon)^{k-1}(-\rho_1 B_1(\epsilon)) \ . \tag{8.16}$$

By (8.15) for $\epsilon = 0$, $-\rho_1 \hat{g}_1^{(1)}(-\zeta_1) = 1$, so that (8.16) reduces to

$$\begin{aligned} 0 &= \sum_{k=2}^{\infty} \hat{g}_1^{(k)} \frac{(-\zeta_1)}{k!} k \tilde{B}_0(\epsilon)^{k-1}(-\rho B_1(\epsilon)) \\ &= -\hat{g}_1^{(1)}(-\zeta_1 + \tilde{B}_0(\epsilon)) - \hat{g}_1^{(1)}(-\zeta_1)](-\rho B_1(\epsilon)) \ , \end{aligned} \tag{8.17}$$

By (8.15), using ϵ and 0,

$$g_1^{(1)}(-\zeta_1 + \tilde{B}_0(\epsilon)) - \hat{g}_1^{(1)}(-\zeta_1) = 0 \quad \text{for all } \epsilon \ .$$

Hence equation (8.17) places no constraint upon $B_1(\epsilon)$.

Next, considering the coefficients of u^2 in (8.10), we obtain

$$\begin{aligned} B_2(\epsilon) = {} & \hat{g}_1^{(1)}(-\zeta_1)(1 - \rho_1 B_2(\epsilon)) + \hat{g}_2^{(2)}(-\zeta_1)\frac{\rho_1^2 B_1(\epsilon)^2}{2} + \tilde{B}_0(\epsilon)\hat{g}_1^{(2)}(-\zeta_1)(1 - \rho_1 B_2(\epsilon)) \\ & + \sum_{k=3}^{\infty} \frac{g_1^{(k)}(-\zeta_1)}{k!} \left[\binom{k}{1} \tilde{B}_0(\epsilon)^{k-1}(1 - \rho_1 B_2(\epsilon)) + \binom{k}{2} \tilde{B}_0(\epsilon)^{k-2}(-\rho_1 B_1(\epsilon))^2 \right] \ . \end{aligned} \tag{8.18}$$

Applying (8.15) again, we see that (8.17) is equivalent to

$$\begin{aligned} 0 &= \hat{g}_1^{(1)}(-\zeta_1) + \hat{g}_1^{(2)}(-\zeta_1)\frac{\rho_1^2 B_1(\epsilon)^2}{2} + [\hat{g}_1^{(1)}(-\zeta_1 + \tilde{B}_0(\epsilon)) - \hat{g}_1^{(1)}(-\zeta_1)](1 - \rho_1 B_2(\epsilon)) \\ &\quad + [\hat{g}_1^{(2)}(-\zeta_1 + \tilde{B}_0(\epsilon)) - \hat{g}_1^{(2)}(-\zeta_1)](\rho_1^2 B_1(\epsilon)^2) \\ &= \hat{g}_1^{(1)}(-\zeta_1) + \hat{g}_1^{(2)}(-\zeta_1)\frac{\rho_1^2 B_1(\epsilon)^2}{2} + [g_1^{(2)}(-\zeta_1 + \tilde{B}_0(\epsilon)) - \hat{g}_1^{(2)}(-\zeta_1)]\rho_1^2 B_1(\epsilon)^2 \end{aligned} \tag{8.19}$$

Since $\hat{g}_1^{(2)}(-\zeta_1 + \tilde{B}_0(\epsilon)) \to \hat{g}_1^{(2)}(-\zeta_1)$ as $\epsilon \to 0$, (8.19) implies that

$$B_1(\epsilon)^2 \to B_1(0)^2 = a_1^2 = \frac{-2\hat{g}_1^{(1)}(-\zeta_1)}{\hat{g}_1^{(2)}(-\zeta_1)\rho_1^2} = \frac{2}{\rho_1^3 \hat{g}_1^{(2)}(-\zeta_1)} \tag{8.20}$$

as $\epsilon \to 0$.

Considering the coefficient of u^k in (8.10), we see that two applications of (8.15) eliminates the variable $B_k(\epsilon)$ and we can solve for $B_{k-1}(\epsilon)$ recursively. Given that $B_j(\epsilon) \to B_j(0)$ as $\epsilon \to 0$ for all $j \le k-2$ and $\tilde{B}_0(\epsilon) \to 0$, we obtain $B_{k-1}(\epsilon) \to B_{k-1}(0)$ by induction on k. ∎

Remark 8.1. Using (8.3) and (8.4), we can see that the coefficients a_1 and a_3 in (8.5) and (8.8) are consistent with the two calculated coefficients in Theorem 7.1. In Theorem 7.1 there is no counterpart of a_2 in (8.7) though. ∎

We now can apply the asymptotic expansion for $\hat{b}_1(s)$ in (8.2) to obtain asymptotic expansions for the other quantities of interest by using elementary operations on asymptotic expansions; e.g., see p. 19 of Olver [46]. We now treat $\hat{h}_0^{(1)}(s)$, the Laplace-Stieltjes transform of $H_0^{(1)}(t)$, which is related to the emptiness probability $P_{00}^{(1)}(t)$ via (2.5). The corresponding asymptotic expansion for $P_{00}^{(1)}(t)$ extends the previous asymptotic result (the first term) in III.7.3 of Cohen [30].

Theorem 8.2. *Under the assumptions of Theorem 8.1,*

$$\hat{h}_0^{(1)}(s) \sim \sum_{k=0}^{\infty} c_k (s+\tau_1^{-1})^{k/2} \quad \text{as} \quad s \to -\tau_1^{-1} , \tag{8.21}$$

where

$$c_0 = \frac{1}{\rho_1}\left(1 - \frac{1}{\zeta_1 \tau_1}\right) , \quad c_1 = \frac{a_1}{\tau_1 \zeta_1^2} , \tag{8.22}$$

$$c_2 = \frac{1}{\rho_1 \tau_1 \zeta_1^3}(\tau_1 \zeta_1^2 - a_1^2 \rho_1^2 - \zeta_1(1-\rho_1 a_2)) , \tag{8.23}$$

and

$$c_3 = \frac{1}{\tau_1 \zeta_1^4}\left(a_1^3 \rho_1^2 + 2a_1 \zeta_1 (1-\rho_1 a_2) + \zeta_1^2 (a_3 + a_1 \tau_1)\right) . \tag{8.24}$$

for a_1, a_2 and a_3 in Theorem 8.1.

Proof. By (2.6), we can write

$$\hat{h}_0^{(1)}(s) = \frac{1-\hat{b}_1(s)}{s+\rho_1-\rho_1 \hat{b}_1(s)} = \frac{1}{\rho_1} - \frac{s}{\rho_1 z_1(s)} \quad = \frac{1}{\rho_1} + \frac{\tau_1^{-1}-\sigma}{\rho_1 z_1(\sigma)} \tag{8.25}$$

where $\sigma = s+\tau_1^{-1}$ and $z_1(\sigma) = -\zeta_1 + \delta(\sigma)$ for $\delta \equiv \delta(0,\sigma)$ in (8.11). Hence, we can expand and match coefficients of $\sigma^{k/2}$ as in Theorem 8.1. ∎

Theorem 8.3. *If, in addition to the assumptions of Theorem 8.1, $-s_2^* < \zeta_1$, where $-s_2^*$ is the rightmost singularity of $\hat{g}_2(s)$, then*

$$\hat{g}_{2e}(z_1(s)) \sim \sum_{k=0}^{\infty} \frac{\hat{g}_{2e}^{(k)}(-\zeta_1)\delta^k}{k!} \sim \sum_{k=0}^{\infty} d_k (s+\tau_1^{-1})^{k/2} \quad \text{as} \quad s \to -\tau_1^{-1} , \tag{8.26}$$

where $\delta \equiv \delta(0,\sigma)$ is given in (8.11), $\sigma = s+\tau_1^{-1}$ and

$$d_0 = \hat{g}_{2e}(-\zeta_1) , \quad d_1 = \frac{-2}{\sqrt{2\rho_1^3 \hat{g}_{2e}^{(2)}(-\zeta_1)}} , \tag{8.27}$$

$$d_2 = \frac{1}{\rho_1} + \frac{(d_1\rho_1)^4}{12}\hat{g}_{2e}^{(3)}(-\zeta_1) \tag{8.28}$$

and

$$d_3 = \frac{(d_1\rho_1)^5}{48}[\hat{g}_1^{(4)}(-\zeta_1) - \frac{5}{6} d_1^2 \sigma_1^3 \hat{g}_1^{(3)}(-\zeta_1)^2] . \tag{8.29}$$

Combining Theorems 8.2 and 8.3, we obtain an asymptotic expansion for $1 - \rho f(s)$, the denominator of $\hat{w}_2(s)$ in (2.14).

Theorem 8.4. *Under the conditions of Theorem 8.3,*

$$1 - \rho \hat{f}(s) = 1 - \rho_1 \hat{h}_0^{(1)}(s) - \rho_2 \hat{g}_{2e}(z_1(s)) \sim \sum_{k=0}^{\infty} e_k (s + \tau_1^{-1})^{k/2} \quad \text{as } s \to -\tau_1^{-1} , \tag{8.30}$$

where the first two coefficients are

$$e_0 = \frac{1}{\zeta_1 \tau_1} [1 - (\rho_2/\rho_2^*)] \tag{8.31}$$

and

$$e_1 = \frac{2\alpha_1 \rho_1}{\tau_1 \zeta_1^2} (1 - \rho_2 \tau_1 \zeta_1^2 \hat{g}_{2e}'(-\zeta_1)) \tag{8.32}$$

for

$$\rho_2^* = \frac{g_{21} \tau_1^{-1}}{\hat{g}_2(-\zeta_1) - 1} \ . \tag{8.33}$$

Now we obtain asymptotic expansions for $\hat{w}_2(s)$ and corresponding asymptotic expansions for the pdf $w_2(t)$. There are three cases, depending on whether ρ_2 is greater than, equal to or less than ρ_2^* for ρ_2^* in (8.33).

Theorem 8.5 *Assume that the busy-period equation (7.1) has a root $-\zeta_1$ with $-s_1^* < -\zeta_1$ and $-s_2^* < -\zeta_1$, where $-s_i^*$ is the rightmost singularity of $\hat{g}_i(s)$. Assume that $\hat{b}_1(s)$ has no singularities besides $-\tau_1^{-1}$ with $Re(s) > -\tau_1^{-1} - \epsilon$. Let η, α, $\alpha_1 = -a_1/2$, e_0 and e_1 be as in (5.3), (5.5), and (7.5), (8.31) and (8.32), respectively.*

(a) *If $\rho_2 > \rho_2^*$ for ρ_2^* in (8.33), then*

$$\hat{w}_2(s) - \frac{\alpha\eta}{s + \eta} \sim \sum_{k=0}^{\infty} v_k (s + \tau_1^{-1})^{k/2} \quad \text{as} \quad s \to -\tau_1^{-1} \tag{8.34}$$

for constants v_k and, in particular,

$$v_0 = \frac{1 - \rho}{e_0} \quad \textit{and} \quad v_1 = \frac{-(1 - \rho) e_1}{e_0^2} \ , \tag{8.35}$$

so that

$$w_2(t) - \alpha\eta e^{-\eta t} \sim \frac{A_1 \alpha_1 e^{-t/\tau_1}}{\sqrt{\pi t^3}} (1 + O(1/t)) \quad \text{as} \quad t \to \infty \tag{8.36}$$

and

$$W_2^c(t) - \alpha e^{-\eta t} \sim \frac{A_1 \alpha_1 \tau_1 e^{-t/\tau_1}}{\sqrt{\pi t^3}} (1 + O(1/t)) \quad \textit{as} \quad t \to \infty \ , \tag{8.37}$$

where

$$A_1 = \frac{(1 - \rho) e_1}{2\alpha_1 e_0^2} = \frac{(1 - \rho)\rho_1 \tau_1 (1 - \rho_2 \tau_1 \zeta_1^2 \hat{g}_{2e}'(-\zeta_1))}{[1 - (\rho_2/\rho_2^*)]^2} \ . \tag{8.38}$$

(b) *If $\rho_2 = \rho_2^*$, then $e_0 = 0$ and*

$$\hat{w}_2(s) \sim \sum_{k=-1}^{\infty} v_k (s + \tau_1^{-1})^{k/2} \quad \textit{as} \quad s \to -\tau_1^{-1} \tag{8.39}$$

and, in particular,

$$\hat{w}_2(s) \sim \frac{1-\rho}{e_1\sqrt{s+\tau_1^{-1}}} \quad as \quad s \to -\tau_1^{-1} \ , \tag{8.40}$$

so that

$$w_2(t) \sim A_2 \frac{e^{-t/\tau_1}}{\sqrt{\pi t}}(1+O(1/t)) \quad as \quad t \to \infty \tag{8.41}$$

and

$$W_2^c(t) \sim A_2\tau_1 \frac{e^{-t/\tau_1}}{\sqrt{\pi t}}(1+O(1/t)) \quad as \quad t \to \infty \ , \tag{8.42}$$

where

$$A_2 = \frac{1-\rho}{e_1} = \frac{(1-\rho)\tau_1\zeta_1^2}{2\alpha_1\rho_1(1-\rho_2^*\tau_1\zeta_1^2\hat{g}_{2e}'(-\zeta_1))} \ . \tag{8.43}$$

(c) *If $\rho_2 < \rho_2^*$, then*

$$\hat{w}_2(s) \sim \sum_{k=0}^{\infty} v_k(s+\tau_1^{-1})^{k/2} \quad as \quad s \to -\tau_1^{-1} \ , \tag{8.44}$$

where v_0 and v_1 are given in (8.35), so that

$$w_2(t) \sim \frac{A_1\alpha_1 e^{-t/\tau_1}}{\sqrt{\pi t^3}}(1+O(1/t)) \quad as \quad t \to \infty \tag{8.45}$$

and

$$W_2^c(t) \sim A_1\alpha_1 \frac{\tau_1 e^{-t/\tau_1}}{\sqrt{\pi t^3}}(1+O(1/t)) \quad as \quad t \to \infty \ , \tag{8.46}$$

where A_1 is in (8.38).

Proof. The assumption on the singularities allows us to apply Heaviside's theorem on p. 254 of Doetsch [34] and Sutton [50] to obtain asymptotic expansions in the time domain from the transform asymptotic expansions. We apply Theorem 8.4 to get an asymptotic expansion for $1-\rho\hat{f}(s)$. For part (a), we write

$$\hat{w}_2(s) = \frac{1-\rho}{1-\rho\hat{f}(s)} = \frac{\alpha\eta}{\eta+s} + \frac{(1-\rho)-\eta\alpha\phi(s)}{1-\rho\hat{f}(s)} \tag{8.47}$$

where

$$\phi(s) = \frac{1-\rho\hat{f}(s)}{s+\eta} \ . \tag{8.48}$$

After we subtract the pole at $-\eta$, the rightmost singularity is at $-\tau_1^{-1}$. We can then exploit the asymptotic expansion for $1-\rho\hat{f}(s)$ at $s=-\tau_1^{-1}$. Using algebra, we see that (8.47) implies (8.34), where v_0 and v_1 are given by (8.35). We use the fact that

$$\phi(-\tau_1^{-1}) = \frac{1-\rho\hat{f}(-\tau_1^{-1})}{\eta-\tau_1^{-1}} = \frac{e_0}{\eta-\tau_1^{-1}} \ . \tag{8.49}$$

For (b), since $e_0 = 0$, the first term in $1-\rho\hat{f}(s)$ is $(s+\tau_1^{-1})^{1/2}$, which makes $(s+\tau_1^{-1})^{-1/2}$ the first term in $\hat{w}(s)$. By Lemma 3.5, we must multiply the asymptotic constant by τ_1

when going from the pdf $w_2(t)$ to the ccdf $W_2^c(t)$. ∎

Remark 8.2. The constant A_1 in (8.38) is consistent with the asymptotic constant for the ccdf $W_2^c(t)$ in (7.36).
Remark 8.3. By Theorem 7.5, $\alpha \to 0$ as $\rho_2 \downarrow \rho_2^*$ where α is the first asymptotic constant in (8.36). On the other hand, by (8.38), $e_0 \to 0$ and $A_1 \to \infty$ as $\rho_2 \downarrow \rho_2^*$. hence, even the two-term approximation based on (8.36) cannot be good when ρ_2 is close to ρ_2^*.

9 Long-Tail Distributions

In this section we obtain results for the case in which one of the service-time cdf's $G_1(t)$ or $G_2(t)$ has a long tail (class III). (Theorem 7.6 already includes a result for the case in which $G_2(t)$ is class III, but $G_1(t)$ is not.)

We are able to combine Theorem 7.2 with a result by De Meyer and Teugels [32] to obtain the asymptotics for $H_0^{(1)c}(t)$ when the service-time distribution has a long tail. They treat the special case in which $G_1^c(t)$ has a regularly varying tail, i.e., when

$$G_1^c(t) \sim t^{-c} L(t) \quad \text{as} \quad t \to \infty \ , \tag{9.1}$$

where $c \geq 1$ and $L(t)$ is a slowly varying function; see p. 275 of Feller [38] and Bingham, Goldie and Teugels [22]. (An important special case is when $L(t)$ is a constant.) We will apply this result in Section 12 to determine the asymptotics for $F^c(t)$, and thus $W_2^c(t)$, when the two classes have a common long-tail service-time distribution.
Theorem 9.1. (De Meyer and Teugels [32]) *If (9.1) holds, then*

$$B_1^c(t) \sim (1-\rho_1)^{-1} G_1^c((1-\rho_1)t) \sim (1-\rho_1)^{-c-1} t^{-c} L(t) \quad \text{as} \quad t \to \infty \ . \tag{9.2}$$

De Meyer and Teugels only state the final simple asymptotic form in (9.2), but in some cases the equivalent intermediate form can serve as a much better approximation. (The advantage of a well chosen equivalent asymptotic form is illustrated by the asymptotic normal approximation in Theorem 2 of [8].) We can combine Theorems 7.2(b) and 9.1 and Lemma 3.4 to obtain the following result. This result was also obtained in a different way by Asmussen and Teugels [19]; see (39) of [8].
Theorem 9.2. *If (9.1) holds for $c > 1$, then as $t \to \infty$*

$$\begin{aligned} H_0^{(1)c}(t) \quad &\sim \quad (1-\rho_1) B_{1e}^c(t) \sim (1-\rho_1)^2 \int_t^\infty B_1^c(u)du \\ &\sim \quad \int_{(1-\rho_1)t}^\infty G_1^c(u)du = G_{1e}^c((1-\rho_1)t) && (9.3) \\ &\sim \quad \frac{(1-\rho_1)^{-(c-1)} t^{-(c-1)} L(t)}{c-1} \quad \text{as} \quad t \to \infty \ . && (9.4) \end{aligned}$$

Proof. As indicated, we can combine Theorems 7.2(b) and 9.1 and Lemma 3.4 to obtain the first two lines. To obtain the final relation, apply Theorem 1 on p. 281 and (8.6) on p. 276 of Feller [38]. ∎

In fact, it appears that Theorems 9.1 and 9.2 can be extended to a much more general class of long-tail distributions. We call the following result a conjecture because it relies on the Heaviside operational principle in Section 3, which remain to be fully justified. Indeed, a variant of this reasoning is used by De Meyer and Teugels [32] to establish Theorem 9.1.

Conjecture 9.1. *Suppose that, the Laplace transform $\hat{g}_1(s)$ has 0 as its rightmost singularity and can be represented in the form (3.21). Then*

$$\begin{aligned} b_1(t) &\sim g_1((1-\rho_1)t) \ , & (9.5)\\ B_1^c(t) &\sim (1-\rho_1)^{-1}G_1^c((1-\rho_1)t) & (9.6) \end{aligned}$$

and

$$H_0^{(1)c}(t) \sim G_{1e}^c((1-\rho_1)t) \quad \textit{as} \quad t \to \infty \ . \tag{9.7}$$

Supporting Argument. By the Heaviside's operational principle, $g_1(t) \sim \theta(t)$ as $t \to \infty$. By (4.3), $z_1(s) \sim s/(1-\rho_1)$ as $s \to 0$. By Theorem 4.1, $\hat{b}_1(s)$ has 0 as its rightmost singularity. We assume that $\hat{b}_1(s)$ can also be represented in the form (3.21). We now apply the Kendall functional equation (2.4). Given

$$\hat{g}(z) \sim 1 - z + \sum_{i=2}^{\infty} a_i z(s)^i + \hat{\theta}_g(z) \quad \text{as} \quad z(s) \to 0 \ ,$$

we can write

$$\hat{b}_1(s) = \hat{g}(z_1(s)) \sim 1 - (s + \rho - \rho\hat{b}_1(s)) + \sum_{i=2}^{\infty} a_i z_i(s)^i + \hat{\theta}_g(z(s)) \quad \text{as} \quad s \to 0 \ ,$$

so that

$$\begin{aligned} (1-\rho)\hat{b}_1(s) &\sim 1 - \rho - s + \sum_{i=2}^{\infty} a_i z_i(s)^i + \hat{\theta}_g(z(s)) \\ &\sim 1 - \rho - s + \sum_{i=2}^{\infty} a_i z_i(s)^i + \hat{\theta}_g(\frac{s}{1-\rho}) \quad \text{as} \quad s \to 0 \end{aligned}$$

and

$$\hat{b}_1(s) \sim 1 - \frac{s}{1-\rho} + \frac{1}{1-\rho}\sum_{i=2}^{\infty} a_i z_i(s)^i + \frac{1}{1-\rho}\hat{\theta}_g(\frac{s}{1-\rho}) \quad \text{as} \quad s \to 0 \ .$$

Given that $z_i(s) \sim s/(1-\rho_1)$ as $s \to 0$, $\sum_{i=1}^{\infty} a_i z_i(s)^i$ should be asymptotically negligible as $s \to 0$. Hence, we should have

$$\hat{\theta}_b(s) \sim \frac{1}{1-\rho}\hat{\theta}_g(\frac{s}{1-\rho}) \quad \text{as} \quad s \to 0 \ . \tag{9.8}$$

Hence, with (9.8), the Heaviside operational principle implies (9.5). Then (9.6) follow from Lemma 3.4. Finally (9.7) follows from Lemma 3.4 and (7.17). ∎

We are able to treat the case in which the low-priority service-time cdf $G_2(t)$ is long-tail in the sense of (9.1). We make the assumption directly in terms of the associated

equilibrium-excess ccdf $G^c_{2e}(t)$. The following complements (7.45) in Theorem 7.6.
Theorem 9.3. *Suppose that the assumption of Theorem 7.1 holds and*

$$G^c_{2e}(t) \sim t^{-c}L(t) \quad \text{as} \quad t \to \infty \tag{9.9}$$

for $c \geq 1$ and a slowly varying function $L(t)$. Then

$$F^c(t) \sim \frac{\rho_2}{\rho} G^c_{2e}((1-\rho_1)t) \sim \frac{\rho_2 L(t)}{\rho(1-\rho_1)^c t^c} \qquad \textit{as} \quad t \to \infty \tag{9.10}$$

and

$$W^c_2(t) \sim \frac{\rho_2}{1-\rho} G^c_{2e}((1-\rho_1)t) \sim \frac{\rho_2 L(t)}{\rho(1-\rho_1)^c t^c} \quad \textit{as} \quad t \to \infty \ . \tag{9.11}$$

Proof. By (2.16), it suffices to show that

$$\int_0^\infty F^{(1)c}_{x0}(t) dG_{2e}(x) \to G^c_{2e}((1-\rho_1)t) \quad \text{as} \quad t \to \infty \ , \tag{9.12}$$

because the first term in $F^c(t)$ involving $H^{(1)c}_0(t)$ is clearly asymptotically negligible. The proof is a modification of the proof of Lemma 7.3. We establish an upper bound for the lim sup and a lower bound for the lim inf. For each, we divide the integral into three pieces as in (7.34). We use the fact that $L(ct)/L(t) \to 1$ for any positive c for a slowly varying function $L(t)$. For the upper (lower) bound, we let $c = (1-\rho)(1-\epsilon)$ $(c = (1-\rho)(1+\epsilon))$ for small positive ϵ and apply (7.26) and (7.27). First, by Lemma 7.2,

$$\limsup_{t\to\infty} G^c_{2e}((1-\rho_1)t)^{-1} \int_0^\infty F^{(1)c}_{x0}(t) dG_{2e}(x) \leq \limsup_{t\to\infty} \left\{ \frac{G^c_{2e}((1-\rho_1)(1-\epsilon)t)}{G^c_{2e}((1-\rho_1)t)} \right\} = \frac{1}{(1-\epsilon)^c} \ .$$

Second, by Lemma 7.2 again,

$$\liminf_{t\to\infty} G^c_{2e}((1-\rho_1)t)^{-1} \int_0^\infty F^{(1)c}_{x0}(t) dG_{2e}(x) \geq \liminf_{t\to\infty} \left\{ \frac{G^c_{2e}((1-\rho_1)(1+\epsilon)t)}{G^c_{2e}((1-\rho_1)t)} \right\} = \frac{1}{(1+\epsilon)^c} \ .$$

Since ϵ is arbitrary, we obtain (9.12). To obtain (9.11), we apply Theorem 5.1(b). Condition (5.1) in this case implied by Lemma 3.7. ∎

We now discuss the case in which the high-priority cdf $G_1(t)$ is long-tail. We already have a result for $H^{(1)c}_0(t)$ for the special case of regularly varying $G^c_1(t)$, as in (9.1), in Theorem 9.2 and the more general Conjecture 9.1. We now want to consider the second term of $F(t)$ and more general cdf's G_2. However, we do not yet have a proper proof for the formulas below. This conjecture is consistent with established results in Section 12 for the case in which both classes have a common long-tail service-time distribution.
Conjecture 9.2 *Assume that the rightmost singularity of $\hat{g}_1(s)$ is 0 and $F^{(1)c}_{x0}(t)$ satisfies (5.1) and (5.2) for all $x > 0$ (with $\gamma = 2$, the long tail case). Then*

$$F^{(1)c}_{x0}(t) \sim \frac{x\rho_1}{1-\rho_1} G^c_1((1-\rho_1)t) \quad \text{as} \quad t \to \infty \ , \tag{9.13}$$

for all positive x,

$$B^c_1(t) \sim \frac{1}{1-\rho_1} G^c_1((1-\rho_1)t) \quad \text{as} \quad t \to \infty \ , \tag{9.14}$$

$$\begin{aligned} F^c(t) &\sim \frac{\rho_1}{\rho_1+\rho_2} G_{1e}^c((1-\rho_1)t) + \frac{\rho_2}{(\rho_1+\rho_2)} \left[\frac{g_{22}\rho_1 G_1^c((1-\rho_1)t)}{2g_{21}(1-\rho_1)} + G_{2e}^c((1-\rho_1)t) \right] & (9.15) \\ &\sim \frac{\rho}{\rho_1+\rho_2} G_{1e}^c((1-\rho_1)t) + \frac{\rho_2}{\rho_1+\rho_2} G_{2e}^c((1-\rho_1)t) \quad as \quad t \to \infty & (9.16) \end{aligned}$$

and

$$W_2^c(t) \sim \frac{\rho_1}{1-\rho} G_{1e}^c((1-\rho_1)t) + \frac{\rho_2}{1-\rho} G_{2e}^c((1-\rho_1)t) \quad as \quad t \to \infty \ . \tag{9.17}$$

for F in (2.16).

Supporting Arguments. Since $F_{nx,0}^{(1)}(t)$ is the n-fold convolution of $F_{x0}^{(1)}(t)$ for all positive x and all positive integers n, and since (5.1) holds for $F_{x0}^{(1)c}(t)$ with $\gamma = 2$,

$$F_{x0}^{(1)c}(t) \sim x F_{10}^{(1)c}(t) \quad \text{as} \quad t \to \infty \tag{9.18}$$

for all x; see (3.11). Next note that

$$F_{\epsilon 0}^{(1)c}(t) = \rho_1 \epsilon (\int_0^\infty F_{x0}^c(t) dG_1(x)) + O(\epsilon^2) \quad \text{as} \quad \epsilon \to 0 \ , \tag{9.19}$$

where

$$\int_0^\infty F_{x0}^{(1)c}(t) dG_1(x) = \int_0^N F_{x0}^{(1)c}(t) dG_1(x) + \int_N^\infty F_{x0}^{(1)c}(t) dG_1(x) \ .$$

The argument of Theorems 7.4 and 16.1 suggest that

$$\int_N^\infty F_{x0}^{(1)c}(t) dG_1(x) \approx G_1^c((1-\rho_1)t)$$

for N suitably large. On the other hand,

$$\int_0^N F_{x0}^{(1)c}(t) dG_1(x) \sim g_{11}^{(N)} F_{10}^c(t) \quad \text{as} \quad t \to \infty \ ,$$

where $g_{11}^{(N)} \to g_{11} = 1$ as $N \to \infty$. Now divide through (9.19) by ϵ and note that, by (9.18),

$$\epsilon^{-1} F_{\epsilon 0}^{(1)c}(t) \sim F_{10}^{(1)c}(t) \quad \text{as} \quad t \to \infty \quad \text{for each} \quad \epsilon > 0 \ ,$$

so that we should have

$$F_{10}^{(1)c}(t) \sim \rho_1 F_{10}^{(1)c}(t) + \rho_1 G_1^c((1-\rho_1)t) \quad \text{as} \quad t \to \infty$$

and, thus,

$$F_{10}^{(1)c}(t) \sim \frac{\rho_1}{1-\rho_1} G_1^c((1-\rho_1)t) \quad \text{as } t \to \infty \ . \tag{9.20}$$

Combining (9.18) and (9.20) yields (9.13). Turning to (9.14), recall that

$$B_1^c(t) = \int_0^\infty F_{x0}^{(1)c}(t) dG_1(x) \ .$$

By the argument above,

$$B_1^c(t) \sim F_{10}^{(1)c}(t) + G_1^c((1-\rho_1)t) \quad \text{as} \quad t \to \infty \ ,$$

which gives (9.14). Finally, for (9.15), we treat the two pieces in (2.16). For $H_0^{(1)c}(t)$, we apply Conjecture 9.1. For the second piece, we repeat the argument above, obtaining

$$\int_0^\infty F_{x0}^{(1)c}(t)dG_{2e}(x) \sim g_{2e1}\frac{\rho_1}{1-\rho_1}G_1^c((1-\rho_1)t) + G_{2e}^c((1-\rho_1)t) \quad \text{as} \quad t \to \infty \ .$$

It seems that in this long tail case we should have

$$G_1^c(t) = o(G_{1e}^c(t)) \quad \text{as} \quad t \to \infty \ ; \tag{9.21}$$

e.g., this is clearly true under (9.1). That makes (9.15) asymptotically equivalent to (9.16). Indeed, (9.21) is needed to make (9.18) consistent with the special case in which both classes have the same long-tail service-time distribution in Section 13.

Remark 9.1. Of course one service-time ccdf may dominate the other in (9.16), in which case only one of the two terms in (9.16) and (9.17) will appear.

10 Sojourn Times

With the nonpreemptive priority discipline, the low-priority sojourn time is the low-priority waiting time plus the low-priority service time, which has transform $\hat{g}_2(s)$. With the preemptive-resume discipline, the low-priority sojourn time is the low-priority waiting time plus the low-priority completion time, which has transform $\hat{c}(s)$ in (2.11). Thus, the exponential asymptotic result for the low-priority waiting time in Section 5 extends quite directly to the low-priority sojourn time, as indicated for the FIFO discipline in [14].

Let $S_2^c(t)$ be the low-priority sojourn-time ccdf.

Theorem 10.1. *Assume that the condition of Theorem 7.1 holds and $\rho_2 > \rho_2^*$ for ρ_2^* in (7.22), so that the exponential asymptotics in (5.4) holds for $W_2^c(t)$.*

(a) With the nonpreemptive discipline,

$$S_2^c(t) \sim \hat{g}_2(-\eta)W_2^c(t) \quad \text{as} \quad t \to \infty \ . \tag{10.1}$$

(b) With the preemptive-resume discipline,

$$S_2^c(t) \sim \hat{c}(-\eta)W_2^c(t) \quad \text{as} \quad t \to \infty \ , \tag{10.2}$$

where

$$\hat{c}(-\eta) = \hat{g}_2(-\eta - \zeta_1 + \tau_1) > \hat{g}_2(-\eta) \ . \tag{10.3}$$

Proof. We apply Theorem 1 of [14]. By Theorem 7.3 (b) and Lemma 15.1,

$$\eta < s^* < \tau_1^{-1} < \zeta_1 < s_2^* \ . \tag{10.4}$$

For (a), we use $\eta < s_2^*$; for (b) we use $\eta < \tau_1^{-1}$, because τ_1^{-1} is the rightmost singularity of $\hat{c}(s)$ in (2.11). By (7.23), we have (10.3). ∎

We now consider the non-exponential asymptotics as in Theorem 7.5. For this purpose, we apply Lemma 3.6. We use Lemma 3.6 in part (b) below because both the waiting-time and the completion-time transforms have $-\tau_1^{-1}$ as their rightmost singularities.

Theorem 10.2. *Suppose that the conditions of Theorem 7.5 hold with $\rho_2 < \rho_2^*$, so that the non-exponential asymptotics for $W_2^c(t)$ in (7.36) holds.*

(a) *With the nonpreemptive discipline,*

$$S_2^c(t) \sim \hat{g}_2(-\tau_1^{-1}) W_2^c(t) \quad \text{as} \quad t \to \infty \; . \tag{10.5}$$

(b) *With the preemptive-resume discipline,*

$$S_2^c(t) \sim \hat{c}(-\tau_1^{-1}) W_2^c(t) + \hat{w}_2(-\tau_1^{-1}) C^c(t) \quad \text{as} \quad t \to \infty \; , \tag{10.6}$$

where the completion-time ccdf satisfies

$$C^c(t) \sim -\hat{g}_2'(-\zeta_1)\rho_1\tau_1 b_1(t) \quad \text{as} \quad t \to \infty \; . \tag{10.7}$$

Proof. Expressing $S_2^c(t)$ as a convolution, we have

$$t^{3/2} e^{\tau_1^{-1} t} S_2^c(t) = \int_0^t (t-y)^{3/2} e^{\tau_1^{-1}(t-y)} W_2^c(t-y) \left(\frac{t}{t-y}\right)^{3/2} e^{\tau_1^{-1} y} dG_2(y) \; . \tag{10.8}$$

For each y,

$$(t-y)^{3/2} e^{\tau_1^{-1}(t-y)} W_2^c(t-y) \left(\frac{t}{t-y}\right)^{3/2} \to A \equiv \lim_{t\to\infty} t^{3/2} e^{\tau_1^{-1} t} W_2^c(t) \; .$$

Hence, for part (a) we can apply Fatou's lemma to obtain

$$\liminf_{t\to\infty} t^{3/2} e^{\tau_1^{-1} t} S_2^c(t) \ge \int_0^\infty A \; e^{\tau_1^{-1} y} dG_2(y) = A\hat{g}(-\tau_1^{-1}) \; .$$

We bound the lim sup above by dividing the integral in (10.8) in two parts, over the intervals $(0,(1-\epsilon)t)$ and $((1-\epsilon t, t)$. We obtain

$$\begin{aligned}
\limsup_{t\to\infty} t^{3/2} e^{\tau_1^{-1} t} S_2^c(t) &\le \int_0^{(1-\epsilon)t} + \int_{(1-\epsilon)t}^t \\
&\le \int_0^\infty A e^{\tau_1^{-1} y} dG(y) + \int_{(1-\epsilon)t}^t t^{3/2} e^{\tau_1^{-1} t} W_2^c(0) dG_2(y) \\
&\le A\hat{g}(-\tau_1^{-1}) + \limsup_{t\to\infty} t^{3/2} e^{\tau_1^{-1} t} G_2^c((1-\epsilon)t) \le A\hat{g}(-\tau_1^{-1}) \; ,
\end{aligned}$$

because we can choose ϵ so that $\tau_1^{-1} < (1-\epsilon)s_2^*$.

(b) For part (b) we apply Lemma 3.6. We apply Lemma 7.3 with (2.11) to obtain (10.7). ∎

We conclude this section by conjecturing the results when G_1 is a long-tail cdf.

Conjecture 10.1. *Suppose that the rightmost singularity of $\hat{g}_1(s)$ is 0.*

(a) *With the nonpreemptive discipline,*

$$S_2^c(t) \sim W_2^c(t) + G_2^c(t) \quad \text{as} \quad t \to \infty \tag{10.9}$$

(b) *With the preemptive-resume discipline,*

$$S_2^c(t) \sim W_2^c(t) + C^c(t) \ , \tag{10.10}$$

where

$$C(t) \sim \frac{g_{21}\rho_1}{1-\rho_1} G_1^c((1-\rho_1)t) + G_2^c((1-\rho_1)t \quad \text{as} \quad t \to \infty \ . \tag{10.11}$$

Supporting Argument. Apply the convolution operational principle in Section 3. The case of regularly varying tails is covered by Lemma 3.7. Second, for (10.11), we apply Conjectures 9.1 and 9.2.

11 Exponential High-Priority Service Times

In this section we begin considering special cases for the low-priority waiting-time distribution. Here we consider the case in which the high-priority service-time distribution is exponential, denoted by $M/M, G/1$. This case is easier to analyze, because the Kendall functional equation for the busy-period transform (2.4) becomes a quadratic equation, yielding the explicit solution

$$\hat{b}_1(s) = \frac{1}{2\rho_1}(\sigma(s) - \sqrt{\sigma(s)^2 - 4\rho_1}) \tag{11.1}$$

for

$$\sigma(s) = 1 + \rho_1 + s \tag{11.2}$$

Moreover, for the M/M/1 model, $\hat{h}_0^{(1)}(s) = \hat{b}_1(s)$; see Corollary 4.2.3 of [1].

Thus, in this case the transform $\hat{f}(s)$ in (2.15) becomes

$$\hat{f}(s) = \frac{\rho_1}{\rho_1 + \rho_2}\hat{b}_1(s) + \frac{\rho_2}{\rho_1 + \rho_2}\hat{g}_{2e}(z_1(s)) \ . \tag{11.3}$$

Hence, the transform $\hat{w}_2(s)$ in (2.20) becomes explicit, no longer involving the Kendall functional equation.

When G_1 is exponential,

$$\tau_1^{-1} = (1 - \sqrt{\rho_1})^2 \ , \tag{11.4}$$

$$\hat{b}_1(-\tau_1^{-1}) = 1/\sqrt{\rho_1} \ , \tag{11.5}$$

and

$$-z_1(-\tau_1^{-1}) = \zeta_1 = 1 - \sqrt{\rho_1} \ , \tag{11.6}$$

so that, when $s_2^* > \zeta_1 = 1 - \sqrt{\rho_1}$,

$$\rho_2^* = \frac{1 - \sqrt{\rho_1}}{\hat{g}_{2e}(-1 + \sqrt{\rho_1})} = \frac{(1 - \sqrt{\rho_1})^2 g_{21}}{\hat{g}_2(-1 + \sqrt{\rho_1}) - 1} \ . \tag{11.7}$$

For $\rho_2 < \rho_2^*$,

$$W_2^c(t) \sim \alpha' \tau_1 b(t) \quad \text{as} \quad t \to \infty \tag{11.8}$$

where the asymptotics for $b_1(t)$ is given in (7.3) and

$$\alpha' = \frac{\rho_2^*(1-\rho)(\rho_2^*\rho_1 + \rho_2(\sqrt{\rho_1} - \rho_1))}{(1 - \sqrt{\rho_1})^2(\rho_2^* - \rho_2)^2} \ . \tag{11.9}$$

Remark 11.1 Note that the asymptotics for $W_2^c(t)$ when $\rho_2 < \rho_2^*$ depends strongly on the high-priority busy-period pdf $b_1(t)$, but not at all upon the low-priority service-time cdf G_2. Note that the asymptotic constant α' in (11.9) depends upon G_2 only through ρ_2 and ρ_2^*.
Remark 11.2 Roughly speaking, for any G_2, the exponential asymptotics takes over when $g_{21} > 2g_{11}$ and $\rho_2 > \rho_1/2$, which is the natural (good) operating regime for a two-priority system. This observation is supported by the following example. Consider three gamma distributions for G_2: $\Gamma(1/2)$, $\Gamma(1) \equiv M$ and $\Gamma(\infty) \equiv D$. Table 11.1 displays values of ρ_2^* as a function of the type of cdf G_2 and its mean g_{21} in the case $\rho_1 = 4/9$.

G_2	$g_{21} = 2$	$g_{21} = 0.2$
$\Gamma(1/2)$	0	0.30
M	0.11	0.31
D	0.23	0.32

Table 11.1 Values of ρ_2^* as a function of the low-priority service-time cdf G_2 and its mean g_{21} for the case with $\rho_1 = 4/9$.

We now consider the further special case in which the service-time distributions of both classes are exponential, denoted by $M/M_1, M_2/1$. Let the means for classes 1 and 2 be 1 and q^{-1}, respectively, with $0 < q < \infty$; i.e., the service-time Laplace transforms are

$$\hat{q}_1(s) = \frac{1}{1+s} \quad \text{and} \quad q_2(s) = \frac{q}{q+s} \; . \tag{11.10}$$

Theorem 11.1 *For the $M/M_1, M_2/1$ priority model, the Laplace transform of the low-priority steady-state waiting time before beginning service is*

$$\hat{w}_2(s) = \frac{(1-\rho_1-\rho_2)\left(\frac{\sigma}{2} + \sqrt{\left(\frac{\sigma}{2}\right)^2 - \rho_1} \; + q - 1\right)}{\frac{q}{2}\sqrt{\sigma^2 - 4\rho_1} + \sigma\left(1 - \frac{q}{2}\right) - r} \tag{11.11}$$

where

$$r = \rho_1 + \rho_2 - (q-1)(1-\rho_2) \tag{11.12}$$

and $\sigma \equiv \sigma(s)$ is in (11.2).
Proof. For the M/M/1 model, $\hat{h}_0^{(1)}(s) = \hat{b}_1(s)$, so that formula (2.14) becomes

$$\hat{w}_2(s) = \frac{1-\rho_1-\rho_2}{1-\rho_1\hat{b}_1(s) - \frac{\rho_2 q}{q+z_1(s)}} \; . \tag{11.13}$$

Since $z_1(s) = \hat{b}_1(s)^{-1} - 1$ for M/M/1 by (2.4), (11.13) is equivalent to

$$\hat{w}_2(s) = \frac{(1-\rho_1-\rho_2)(\hat{b}_1(s)^{-1} + q - 1)}{\hat{b}_1(s)^{-1} - r - (q-1)\rho_1\hat{b}_1(s))} \; . \tag{11.14}$$

for r in (11.12). Finally, (11.14) implies (11.11) because of (11.1) and

$$\hat{b}_1(s)^{-1} = \frac{1}{2}\left(\sigma + \sqrt{\sigma^2 - 4\rho_1}\right) \tag{11.15}$$

for σ in (11.2). ∎

Since the transform $\hat{w}_2(s)$ is available explicitly (in terms of σ in (11.2)), it can be directly inverted. We now identify the parameters for the exponential asymptotics explicitly.

Theorem 11.2 *In the $M/M_1, M_2/1$ model, if $\rho_2 > \rho_2^*$, where*

$$\rho_2^* = \begin{cases} \left(\dfrac{1-\sqrt{\rho_1}}{q}\right)(q-1+\sqrt{\rho_1}) & \text{if } q > 1-\sqrt{\rho_1} \\ 0 & \text{if } q \le 1-\sqrt{\rho_1}\ , \end{cases}$$

then the exponential asymptotics in (5.4) holds with

$$\eta = 1 + \rho_1 - \frac{r}{q-1}\left(\frac{q}{2}\left(1 + sgn(r)\sqrt{1 + 4\rho_1(q-1)/r^2}\right) - 1\right) \tag{11.16}$$

and

$$\alpha = \frac{(1-\rho_1-\rho_2)}{\eta}\left(\frac{\sigma(-\eta_1)^2 - 4\rho_1 + (\sigma(-\eta_1) + 2(q-1))\sqrt{\sigma(-\eta_1)^2 - 4\rho_1}}{q\sigma(-\eta_1) + (2-q)\sqrt{\sigma(-\eta_1)^2 - 4\rho_1}}\right) \tag{11.17}$$

for r in (11.12) and $\sigma(s)$ in (11.2).

Proof. If $\rho_2 > \rho_2^*$, the denominator of (11.11) has $s = -\eta$ in (11.16) as the rightmost negative zero. Then L'Hospital's rule yields

$$s\hat{w}(s-\eta) \to A(-\eta)/B'(-\eta) \tag{11.18}$$

where $\hat{w}(s) = A(s)/B(s)$ for A and B in (11.11). From (11.11) and (11.18), we obtain (11.17). ∎

We conclude this section by considering some further special cases:

> (a). If $q \to 0$, then $r \to 1 + \rho_1$ and $\eta \sim q$ as $q \to 0$. Recall that $\rho_2^* = 0$ for $q \le 1 - \sqrt{\rho_1}$.
>
> (b). If $q = 1$, then the two service-time distributions coincide, $r = \rho_1 + \rho_2$, $\rho_2^* = \sqrt{\rho_1} - \rho_1$ and $\eta = [\rho_2/(\rho_1+\rho_2)] - \rho_2$, which agrees with the special case considered in Section 13.
>
> (c). If $q = 2$, then $\rho_2^* = (1-\rho_1)/2$ and $\eta = 1+\rho_1-\sqrt{(1+\rho_1)^2 - 4\rho_2(1-\rho_1-\rho_2)}$.
>
> (d). If $q \to \infty$, then $r \sim -q(1-\rho_2)$ as $q \to \infty$, $\rho_2^* \to 1 - \sqrt{\rho_1}$ and $\eta \to \rho_2 - \rho_1\rho_2/(1-\rho_2)$, which agrees with the special case considered in Section 14.

12 One Common General Distribution

In this section we consider the special case in which G_1 is the service-time cdf of both classes, as occurs whenever the service-time distribution is not related to the priority structure. Let the common mean be $g_{11} = 1$. Then ρ_1 and ρ_2 represent the two arrival rates. Note that the two arrival rates need not coincide, however, so that we can have any values of ρ_1 and ρ_2, provided that $\rho \equiv \rho_1 + \rho_2 < 1$.

Theorem 12.1 *In the M/G/1 priority model with common service-time distributions having mean $g_{11} = 1$,*

$$\hat{w}_2(s) = \frac{1-\rho}{1-\rho\hat{h}_0^{(1)}(s)} \tag{12.1}$$

for $\hat{h}_0^{(1)}(s)$ in (2.6).

Proof. The general formula in (2.14) simplifies because

$$s + \rho_1 - z_1(s) = \rho_1\hat{g}_1(z_1(s)) = \rho_1\hat{g}_2(z_1(s))$$

by (31) of [6]. ∎

Theorem 12.2. *Assume that the condition of Theorem 7.1 holds for the common service-time cdf G_1 and let*

$$\rho_2^* = \frac{\rho_1}{\zeta_1\tau_1 - 1} \ . \tag{12.2}$$

(a) If $\rho_2 > \rho_2^*$, then

$$W_2(t) \sim \alpha e^{-\eta t} \quad \text{as} \quad t \to \infty \ ,$$

where η is the minimum positive real root of the equation

$$\rho\hat{h}_0^{(1)}(-s) = 1 \tag{12.3}$$

and

$$\alpha = \frac{1-\rho}{-\eta\rho\hat{h}_0^{(1)\prime}(-\eta)} \ . \tag{12.4}$$

(b) If $\rho_2 < \rho_2^*$, then

$$\begin{aligned} W_2^c(t) \quad &\sim \quad \frac{\rho(1-\rho)}{(1-\rho\hat{h}_0^{(1)}(-\tau_1^{-1}))^2} H_0^{(1)c}(t) \\ &\sim \quad \frac{\rho(1-\rho)[1+(\rho_1/\rho_2^*)]^2}{[1-(\rho_2/\rho_2^*)]^2} H_0^{(1)c}(t) \\ &\sim \quad \frac{\rho(1-\rho)[1+(\rho_1/\rho_2^*)]^2}{[1-(\rho_2/\rho_2^*)]^2} \frac{\alpha_1 e^{-t/\tau_1}}{\zeta_1^2\sqrt{\pi t^3}} \quad \text{as} \quad t \to \infty \end{aligned} \tag{12.5}$$

for ρ_2^* in (12.2) and for τ_1, α_1 and ζ_1 as in Theorem 7.1.

Proof. Because of the structure of (12.1), Theorem 5.1 can be applied, with $\hat{h}_0^{(1)}(s)$ playing the role of the transform $\hat{f}(s)$. The asymptotic behavior of $H_0^{(1)c}(t)$ was given in (7.15). The value of $\hat{h}_0^{(1)}(-\tau_1^{-1})$ is given in (7.20), yielding $1 - \rho_1\hat{h}_0^{(1)}(-\tau_1^{-1}) = (\zeta_1\tau_1)^{-1}$. ∎

Remark 12.1. As noted in Section 2, the low-priority steady-state waiting time coincides with the first passage time to 0 for the high priority alone (ignoring future low-priority arrivals) starting from the steady-state workload for both classes. Thus, when the two classes have the same service-time distribution, the low-priority steady-state waiting time is the same as the transient waiting time with the FIFO discipline when the arrival rate is shifted to λ_1 after the system has been in steady state with arrival rate $\lambda_1 + \lambda_2$. ∎

We next discuss the M/M,M/1 special case in which both classes have a common exponential distribution. In this case, $\hat{h}_0^{(1)}(s) = \hat{b}_1(s)$, so that

$$\hat{w}_2(s) = \frac{1-\rho}{1-\rho\hat{b}_1(s)} = 1 - \rho + \frac{\rho(1-\rho)\hat{b}_1(s)}{1-\rho\hat{b}_1(s)} \ , \tag{12.6}$$

where $\hat{b}_1(s)$ is given in (11.1) and can be reexpressed as

$$\hat{b}_1(s) = \frac{1}{2\rho_1}(1 + \rho_1 + s - \sqrt{(\tau_1^{-1} + s)(\gamma_1 + s)}) \tag{12.7}$$

for

$$\tau_1^{-1} = (1 - \sqrt{\rho_1})^2 \quad \text{and} \quad \gamma_1 = (1 + \sqrt{\rho_1})^2 \ . \tag{12.8}$$

In this case, equations (12.6) and (12.7) can be inverted analytically by using contour integrals, as in Chapter 3 of Duffy [36], yielding

$$w_2(t) = \begin{cases} (1-\rho)\delta(t) + \alpha\eta e^{-\eta t} + y(t) & \text{if} \quad \rho_2 > \rho_2^* \equiv \sqrt{\rho_1} - \rho_1 \\ \\ (1-\rho)\delta(t) + y(t) & \text{if} \quad \rho_2 \le \rho_2^* \ , \end{cases} \tag{12.9}$$

where

$$\alpha = \frac{\rho^2 - \rho_1}{\rho_2} = \frac{(\rho_2 - \rho_2^* + \sqrt{\rho_1})^2 - \rho_1}{\rho_2} \ , \tag{12.10}$$

$\delta(t)$ is the delta function corresponding to a unit point mass at the origin,

$$y(t) = \frac{1-\rho}{2\pi} \int_{\tau_1^{-1}}^{\gamma_1} \frac{e^{-xt}}{x - \eta} \sqrt{(x - \tau_1^{-1})(\gamma_1 - x)}dx \ , \tag{12.11}$$

$$\eta = \rho_2(1-\rho)/\rho \ , \tag{12.12}$$

and

$$\begin{aligned} b_1(t) &= \frac{1}{2\pi\rho_1} \int_{\tau_1^{-1}}^{\gamma_1} e^{-xt} \sqrt{(x - \tau_1^{-1})(\gamma_1 - x)}dx \\ &= \frac{e^{-t/\tau_1}}{2\pi\rho_1} \int_0^{4\sqrt{\rho_1}} e^{-\mu t} \sqrt{\mu(4\sqrt{\rho_1} - \mu)}d\mu \ . \end{aligned} \tag{12.13}$$

For (12.13), see Section 3 of [3].

The integrals in (12.11) and (12.13) have asymptotic expansions that can be determined by using Watson's lemma, e.g., see p. 71 of Olver [46]. It follows from (12.11) and (12.13) that

$$y(t) \sim \frac{\rho_1(1-\rho)b_1(t)}{(\tau_1^{-1} - \eta)} \quad \text{as} \quad t \to \infty \tag{12.14}$$

and

$$\rho(\tau^{-1} - \eta) = (\rho_2^* - \rho_2)^2$$

provided that $\rho_2 \neq \rho_2^*$ Watson's lemma enables us to determine an asymptotic expansion for the waiting-time pdf of the form

$$w_2(t) \sim \frac{\rho_1 \rho (1-\rho)}{(\rho_2^* - \rho_2)^2} \frac{e^{-t/\tau_1}}{2\rho_1^{3/4}\sqrt{\pi t^3}} \left(1 + \frac{a_1}{t} + \frac{a_2}{t^2} + \ldots\right) \quad \text{as} \quad t \to \infty \tag{12.15}$$

for $\rho_2 < \rho_2^*$ as indicated in (1.4). The first three terms for the busy period are given in (4.1) of [2]. This time domain analysis supports Section 8.

Note that the root equation $1-\rho\hat{b}_1(s) = 0$ in the denominator of (12.6) has no root for $\rho_2 < \rho_2^* = \sqrt{\rho_1} - \rho_1$, but $y(t)$ in (12.1) is valid for all ρ_2. At $\rho_2 = \rho_2^*$, $\eta = \tau_1^{-1} = (1-\sqrt{\rho_1})^2$, so that $y(t)$ has a different structure, namely

$$y(t) = \frac{1-\rho}{2\pi} \int_{\tau_1^{-1}}^{\gamma_1} e^{-xt} \sqrt{\frac{\gamma_1 - x}{x - \tau_1^{-1}}} dx \tag{12.16}$$

and therefore a different asymptotic expansion, consistent with Theorem 8.5(b).

We conclude this section by stating a result for long-tail distributions.

Theorem 12.3. *If the common service-time ccdf $G_1^c(t)$ has a regularly varying tail as in (9.1), then*

$$W_2^c(t) \sim \frac{\rho}{1-\rho} G_{1e}^c((1-\rho_1)t) \quad \text{as} \quad t \to \infty\ . \tag{12.17}$$

Proof. Combine Theorems 5.1(b) and 9.2. Condition (5.1) holds by Lemma 3.7. Condition (5.2) follows easily too. ∎

As in Conjectures 9.1 and 9.2, we conjecture that (12.17) holds for more general long-tail distributions (when 0 is the rightmost singularity of $\hat{g}_1(s)$).

Remark 12.2. It is instructive to compare (12.17) with the corresponding result for the FIFO discipline, see [11] or Pakes [47], which is

$$W_{\mathrm{FIFO}}^c(t) \sim \frac{\rho}{1-\rho} G_e(t) \quad \text{as} \quad t \to \infty\ . \tag{12.18}$$

If $G^c(t) \sim t^{-c}$, as $t \to \infty$, then (12.17) is equivalent to

$$W_2^c(t) \sim \frac{\rho}{(1-\rho)^{c-1}(1-\rho_1)^{c-1} t^{c-1}} \quad \text{as} \quad t \to \infty\ , \tag{12.19}$$

whereas (12.18) is equivalent to

$$W_{\mathrm{FIFO}}^c(t) \sim \frac{\rho}{(1-\rho)^{c-1} t^{c-1}} \quad \text{as} \quad t \to \infty\ ; \tag{12.20}$$

i.e.,

$$W_2^c(t) \sim \frac{1}{(1-\rho_1)^{c-1}} W_{\mathrm{FIFO}}(t) \quad \text{as} \quad t \to \infty\ . \tag{12.21}$$

13 Fluid Inputs

Washburn [52] considered the interesting case in which the low-priority service times are deterministic but very small. In particular, he considered the limiting case as $g_{21} \to 0$ and $\lambda_2 \to \infty$ with $\rho_2 = \lambda_2 g_{21}$ held fixed. Since $\hat{g}_2(s) \approx 1 - g_{21}s + o(g_{21})$ as $g_{21} \to 0$,

$$\lambda_2 g_2(z_1(s)) - \lambda_2 \to -\rho_2 z_1(s) \quad \text{as} \quad g_{21} \to 0 \tag{13.1}$$

and

$$\hat{w}_2(s) \to \frac{1-\gamma}{1-\gamma \hat{h}_0^{(1)}(s)} \quad \text{as} \quad g_{21} \to 0 \tag{13.2}$$

for $\gamma = \rho_1/(1-\rho_2)$, so that this case has the same general form as analyzed in Section 12 (with γ in (13.2) replacing ρ in (12.1)).

Now suppose instead that the high-priority service times are deterministic but very small, i.e., suppose that $g_{11} \to 0$ and $\lambda_1 \to \infty$ where $\rho_1 = \lambda_1 g_{11}$ is held fixed. Then, from (2.4), we see that

$$\hat{b}_1(s) \to 1 \quad \text{as} \quad g_{11} \to 0 \tag{13.3}$$

corresponding to $B_1(t) \to 1$ for all $t > 0$; i.e., the high-priority busy periods become negligible.

In this situation $\hat{h}_0^{(1)}(s) \to 1$ and $\hat{z}_1(s) \to s$, so that (2.14) holds with

$$\hat{f}(s) = \frac{\rho_1}{\rho_1+\rho_2} + \frac{\rho_2}{\rho_1+\rho_2}\hat{g}_{2e}(s)\ , \tag{13.4}$$

which in turn is equivalent to

$$\hat{w}_2(s) = \frac{1-\rho_2/(1-\rho_1)}{1-(\rho_2/(1-\rho_1))\hat{g}_{2e}(s)}\ . \tag{13.5}$$

Formula (13.5) is the M/G/1 FIFO Pollaczek-Khintchine formula for the steady-state workload when the traffic intensity is inflated by $(1-\rho_1)^{-1}$ to account for the unavailability of the server.

Finally, if both classes had fluid inputs, there would be no steady-state workload.

14 Numerical Examples

In this section we consider three numerical examples. Our first numerical example has two classes with a common exponential service-time distribution having mean 1. The only parameters are the two arrival rates ρ_1 and ρ_2. This example is a special case of both Sections 11 and 12, so all results there apply. In this example the boundary between exponential and non-exponential asymptotics is $\rho_2^* = \sqrt{\rho_1} - \rho_1$. We let $\rho_1 = 0.5$ and consider two cases: $\rho_2 = 0.3$ and $\rho_2 = 0.1$. In the first case $\rho_2 > \rho_2^* = 0.2071$; in the second case $\rho_2 < \rho_2^*$.

In the first case we have exponential asymptotics as in (1.1) and (5.4) with $\eta = (1-\rho)\rho_2/\rho$ and $\alpha = (\rho^2 - \rho_1)/\rho_2$, so that $\eta^{-1} = 13.33$ and $\alpha = 0.4667$. The exponential asymptote is compared to exact values of $W_2^c(t)$ obtained by numerical transform inversion in Table 14.1.

The approximation is not as good as in the FIFO examples in [12], but it is quite good, giving reasonable accuracy by the 90$^{\text{th}}$ percentile. We can understand why the exponential asymptote is not as good an approximation as in most FIFO examples by looking at the singularities of $\hat{w}_2(s)$. The rightmost singularity is of course $\eta = 0.0750$. The next singularity of $\hat{w}_2(s)$ is the rightmost singularity of $\hat{f}(s)$, which is determined by the busy-period distribution. It is $\tau_1^{-1} = (1-\sqrt{\rho})^2 = 0.0858$. The fact that τ_1^{-1} is close to η explains why the exponential approximation is not better; i.e.,

time	$W_2^c(t)$	$\alpha e^{-\eta t}$
10^{-5}	.8000	.47
1	.6780	.43
5	.4330	.32
10	.2742	.22
20	.1195	.104
40	.02492	.0232
80	.001187	.001157
120	.00005824	.00005759
160	.000002883	.000002867

Table 14.1. A comparison of the exponential asymptote in case 1 of the M/M,M/1 example with $\rho_2 = 0.3$ to exact values obtained by numerical transform inversion.

$$W_2^c(t) \sim \alpha e^{-\eta t} + \frac{A_1 \alpha_1 \tau_1 e^{-t/\tau_1}}{\sqrt{\pi t^3}} \quad \text{as} \quad t \to \infty \tag{14.1}$$

by (8.37).

In the second case we have the non-exponential asymptotics

$$W_2^c(t) \sim \frac{A_1 \alpha_1 \tau_1 e^{-t/\tau_1}}{\sqrt{\pi t^3}} \quad \text{as} \quad t \to \infty \ , \tag{14.2}$$

as in (1.2), (8.46) and (11.8) with $\tau_1 = (1-\sqrt{\rho_1})^{-2} = 11.65685, \tau_1 \zeta_1^2 = 1$, $\alpha_1^{-1} = 2\rho_1^{3/4}$ and $A_1 = \rho_1 \rho (1-\rho)/(\rho_2^* - \rho_2)^2 = 10.460376$, so that $A_1 \tau_1 \alpha_1 = 102.53477$. (The asymptotic constant in (14.2) is consistent with (7.36) and (12.5).) The non-exponential asymptote in (14.2) is compared to exact values obtained by numerical transform inversion in Table 14.2. As in previous FIFO examples, the non-exponential asymptote is a much worse approximation. As explained previously, this is due to the slower rate of convergence in (1.4). Even for every small tail probabilities, it might well be judged necessary to rely on the numerical inversion in this case. Reasonable approximations for $W_2^c(t)$ in the region of primary interest can also be obtained by fitting a hyperexponential (H_2, mixture of two exponentials) by matching the first three moments, as was done for the M/M/1 busy period in [3].

time	$W_2^c(t)$	asymptote (14.2)
10^{-5}	.6000	
1	.4323	
5	.1897	
10	.08814	.78
20	.02415	.12
40	.002516	.0074
60	.0003119	.00072
80	.00004208	.000085
100	.00000598	.000011

Table 14.2. A comparison of the non-exponential asymptote in case 2 of the M/M , M / 1 example with $\rho_2 = 0.1$ to exact values obtained by numerical transform inversion.

Our second example focuses on the boundary case in which $\rho_2 = \rho_2^*$. Again let both service times be exponential with mean 1, but now let $\rho_1 = 4/9$ and $\rho_2 = \rho_2^* = \sqrt{\rho_1} - \rho_1 = 2/9$. In this case $\rho = 2/3$, $\tau_1 = (1 - \sqrt{\rho_1})^{-2} = 9$ and

$$\hat{w}_2(s) \sim \frac{1-\rho}{2\rho\alpha_1\sqrt{s + \tau_1^{-1}}} \quad \text{as} \quad s \to -\tau_1^{-1} \ , \tag{14.3}$$

so that

$$W_2^c(t) \sim \rho_1^{3/4}(\rho^{-1} - 1)\frac{\tau_1 e^{-t/\tau_1}}{\sqrt{\pi t}} = (2.44949\ldots)\frac{e^{-t/9}}{\sqrt{\pi t}} \quad \text{as} \quad t \to \infty \ . \tag{14.4}$$

We compare the boundary non-exponential asymptote in (14.4) to exact values obtained by numerical transform inversion in Table 14.3. From Table 14.3, we see that the quality of the approximation is much better than in Table 14.2.

time	$W_2^c(t)$	asymptote (14.4)
10^{-5}	.6667	
1	.5039	
5	.2358	.35
10	.1104	.14
20	.02849	.033
30	.008000	.0090
40	.002338	.0027
50	.0006995	.00076
60	.0002126	.0002271
80	.00002025	.00002131
100	.000001981	.000002065

Table 14.3. A comparison of the non-exponential asymptote in the boundary case of the second M/M, M/1 example having $\rho_1 = 4/9$ and $\rho_2 = 2/9$ to exact values obtained by numerical transform inversion.

Our third example has two classes with a common long-tail distribution having mean 1. For this example we use a *Pareto mixture of exponentials* (PME) distribution, as in [11], with density, ccdf and transform

$$g_1(t) = \frac{16}{3t^4}\left(1 - \left(1 + \frac{3}{2} + \frac{9t^2}{8} + \frac{9t^3}{16}\right)e^{-3t/2}\right) \tag{14.5}$$

$$G_1^c(t) = \frac{16}{9t^3}\left(1 - \left(1 + \frac{3t}{2} + \frac{9}{8}t^2\right)e^{-3t/2}\right) , \tag{14.6}$$

$$\hat{g}_1(s) \equiv \int_0^\infty e^{-st} g(t)dt = 1 - s + \frac{4s^2}{3} - \frac{8}{9}s^3 \log(1 + 3/2s) . \tag{14.7}$$

The moments of G_1 are $g_{11} = 1$, $g_{12} = 8/3$ and $g_{1n} = \infty$ for $n \geq 3$. Both classes have the same service time cdf G_1, but they can have different arrival rates.

The PME distribution has a regularly varying tail as in (9.1). Hence, we can apply Theorem 12.3 to conclude that

$$W_2^c(t) \sim \frac{\rho}{1-\rho}\int_{(1-\rho_1)}^\infty G^c(u)du \sim \left(\frac{\rho}{1-\rho}\right)\frac{8}{9((1-\rho_1)t)^2} \quad \text{as} \quad t \to \infty . \tag{14.8}$$

As t gets large, the exponential term $e^{-3t/2}$ in (14.5) and (14.6) becomes negligible, so nothing is lost here in replacing $G^c(t)$ by its asymptote $16/9t^3$, which is what we have done in (14.8).

time	exact	asymptote (14.8)
10^{-5}	.6000	
1	.4171	
10	.04574	.016
20	.009022	.0041
40	.001473	.0010
60	.0005656	.00045
80	.0002988	.00026
100	.0001848	.00016
200	.00004342	.000041
400	.00001056	.0000103
800	.000002605	.00000257
1600	.0000006470	.000000643

Table 14.4. A comparison of the power-tail asymptote for $W_2^c(t)$ in the long-tail example to exact values obtained by numerical transform inversion.

As a specific example, we let the arrival rates be $\rho_1 = 0.5$ and $\rho_2 = 0.1$. In this case the mean of the low-priority waiting time is $w_{21} = 4$. The asymptote in (14.8) is compared with exact values obtained by numerically inverting the Laplace transform $\hat{W}_2^c(s) = (1 - \hat{w}_2(s))/s$ for $\hat{w}_2(s)$ in (12.1) and $\hat{h}_0^{(1)}(s)$ in (2.6) in Table 14.4. From Table 14.4, we see that the asymptote is not very accurate by the 99th percentile (about $t = 20$), but the asymptote does much better than in Table 14.2 further out in the tail. Since the class III (long-tail) distributions do not have the exponential term, the probabilities are non-negligible for larger values of t. Note that the largest values of t in

Tables 14.2 and 14.4 are 100 and 1600, respectively. This phenomenon seems to make the class III asymptotes more useful than the class II asymptotes.

Acknowledgment. We are grateful to John A. Morrison for help with the saddle point proof of Theorem 7.1.

References

[1] J. Abate and W. Whitt, Transient behavior of the M/M/1 queue via Laplace transforms, *Adv. Appl. Prob.* 20 (1988) 145–178.

[2] J. Abate and W. Whitt, Simple spectral representations for the M/M/1 queue, *Queueing Systems* 3 (1988) 321–346.

[3] J. Abate and W. Whitt, Approximations for the M/M/1 busy-period distribution, pp. 149–191 in *Queueing Theory and Its Applications — Liber Amicorium for J. W. Cohen*, O. J. Boxma and R. Syski (eds.), North-Holland, Amsterdam, 1988.

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

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

[6] J. Abate and W. Whitt, Transient behavior of the M/G/1 workload process, *Opns. Res.* 42 (1994) 750–764.

[7] J. Abate and W. Whitt, A heavy-traffic expansion for the asymptotic decay rates of tail probabilities in multi-channel queues, *Opns. Res. Letters* 15 (1994) 223-230.

[8] J. Abate and W. Whitt, Limits and approximations for the busy-period distribution in single-server queues, *Prob. Eng. Inf. Sci.* 9 (1995) 581–602.

[9] J. Abate and W. Whitt, An operational calculus for probability distributions via Laplace transforms, *Adv. Appl. Prob.* 28 (1996) 75–113.

[10] J. Abate and W. Whitt, Limits and approximations for the M/G/1 LIFO waiting-time distribution, submitted.

[11] J. Abate, G. L. Choudhury and W. Whitt, Waiting time tail probabilities in queues with long-tail service-time distributions, *Queueing Systems* 16 (1994) 311–338.

[12] J. Abate, G. L. Choudhury and W. Whitt, Exponential approximations for tail probabilities in queues, I: waiting times, *Opns. Res.* 43 (1995) 885–901.

[13] J. Abate, G. L. Choudhury and W. Whitt, Calculating the M/G/1 busy-period density and LIFO waiting-time distribution by direct numerical transform inversion, *Opns. Res. Letters* 18 (1995) 113–119.

[14] J. Abate, G. L. Choudhury and W. Whitt, Exponential approximations for tail probabilities in queues, II: sojourn time and workload, *Opns. Res.*, to appear.

[15] J. Abate, G. L. Choudhury and W. Whitt, On the Laguerre method for numerically inverting Laplace transforms, *INFORMS J. Comput.*, to appear.

[16] J. Abate, G. L. Choudhury and W. Whitt, Numerical inversion of multidimensional Laplace transforms by the Laguerre method, *Perf. Eval.*, to appear.

[17] J. Abate, M. Kijima and W. Whitt, Decompositions of the M/M/1 transition function, *Queueing systems* 9 (1993) 323–336.

[18] S. Asmussen, *Applied Probability and Queues*, Wiley, New York, 1987.

[19] S. Asmussen and J. L. Teugels, Convergence rates for M/G/1 queues and ruin problems with heavy tails, *Adv. Appl. Prob.*, to appear.

[20] K. B. Athreya and P. E. Ney, *Branching Processes*, Springer-Verlag, Berlin, 1972.

[21] A. W. Berger and W. Whitt, Effective bandwidths for waiting times with priorities, in preparation.

[22] N. H. Bingham, C. M. Goldie and J. L. Teugels, *Regular Variation*, Cambridge University Press, Cambridge, England, 1989.

[23] A. A. Borovkov, *Stochastic Processes in Queueing Theory*, Springer-Verlag, New York, 1976.

[24] C. S. Chang and J. A. Thomas, Effective bandwidths in high-speed digital networks. *IEEE J. Sel. Areas Commun.* 13 (1995) 1091–1100.

[25] V. P. Chistyakov, A theorem on sums of independent positive random variables and its applications to branching processes, *Theor. Probability Appl.* 9 (1964) 640–648.

[26] G. L. Choudhury and W. Whitt, Computing distributions and moments in polling models by numerical transform inversion, *Perf. Eval.*, to appear.

[27] G. L. Choudhury, D. M. Lucantoni and W. Whitt, Multidimensional transform inversion with applications to the transient M/G/1 queue, *Ann. Appl. Prob.* 4 (1994) 719–740.

[28] J. Chover, P. Ney and S. Wainger, Functions of probability measures, *J. d'Analyse Math.* 26 (1973) 255–302.

[29] J. Chover, P. Ney and S. Wainger, Degeneracy properties of subcritical branching process, *Ann. Probab.* 1 (1973) 663–673.

[30] J. W. Cohen, *The Single Server Queue*, Revised edition, North-Holland, Amsterdam, 1982.

[31] D. R. Cox and W. L. Smith, *Queues*, Methuen, London, 1961.

[32] A. De Meyer and J. L. Teugels, On the asymptotic behavior of the distribution of the busy period and the service time in M/G/1, *Adv. Appl. Prob.* 17 (1980) 802–813.

[33] G. de Veciana, G. Kesidis and J. Walrand, Resource management in wide-area ATM networks using effective bandwidths, *IEEE J. Sel. Areas Commun.* 13 (1995) 1081–1090.

[34] G. Doetsch, *Introduction to the Theory and Applications of the Laplace Transformation*, Springer, New York, 1974.

[35] N. Duffield, Exponents for the tails of distributions in some polling models, AT&T Research, 1996.

[36] D. G. Duffy, *Transform Methods for Solving Partial Differential Equations*, CRC Press, Boca Raton, FL, 1994.

[37] A. Erdélyi, *Asymptotic Expansions*, Dover, New York, 1956.

[38] W. Feller, *An Introduction to Probability Theory and its Applications*, volume II, second ed., Wiley, New York, 1971.

[39] L. Flatto, The waiting time distribution for random order of service, *Ann. Appl. Prob.*, to appear.

[40] D. P. Gaver, Jr., A waiting line with interrupted service, including priorities, *J. Roy. Stat. Soc.* B 25 (1962) 73–90.

[41] D. P. Heyman and M. J. Sobel, *Stochastic Models in Operations Research*, vol. I, McGraw-Hill, NY, 1982.

[42] J. M. Harrison, A limit theorem for priority queues in heavy traffic, *J. Appl. Prob.* 10 (1973) 907–912.

[43] E. Hille, *Analytic Function Theory*, vol. I, Blaisdell, Waltham MA, 1959.

[44] O. Kella, Parallel and tandem fluid networks with dependent Lévy inputs, *Ann. Appl. Prob.* 3 (1993) 682–695.

[45] D. R. Miller, Computation of steady-state probabilities for M/M/1 priority queues, *Opns. Res.* 29 (1981) 945–958.

[46] F. W. J. Olver, *Asymptotics and Special Functions*, Academic Press, New York, 1974.

[47] A. G. Pakes, On the tails of waiting-time distributions, *J. Appl. Prob.* 12 (1975) 555–564.

[48] A. Shwartz and A. Weiss, *Large Deviations for Performance Analysis*, Chapman and Hall, London, 1995.

[49] D. Stoyan, *Comparison Methods for Queues and Other Stochastic Models*, Wiley, Chichester, 1983.

[50] W. G. L. Sutton, The asymptotic expansion of a function whose operational equivalent is known, *J. London Math. Soc.* 9 (1934) 131–137

[51] B. Van Der Pol and H. Bremmer, *Operational Calculus*, Cambridge Press, 1955; reprinted, Chelsea, New York, 1987.

[52] A. Washburn, A particular two-priority queue, *Opns. Res.* 18 (1970) 1760–1764.

[53] W. Whitt, Weak convergence theorems for priority queues: preemptive-resume discipline, *J. Appl. Prob.* 8 (1971) 74–94.

[54] R. W. Wolff, *Stochastic Modeling and the Theory of Queues*, Prentice Hall, Englewood Cliffs, NJ, 1989.