CrossMark

Transient analysis of the Erlang A model

Charles Knessl[1] · **Johan S. H. van Leeuwaarden**[2]

Abstract We consider the Erlang A model, or $M/M/m + M$ queue, with Poisson arrivals, exponential service times, and m parallel servers, and the property that waiting customers abandon the queue after an exponential time. The queue length process is in this case a birth–death process, for which we obtain explicit expressions for the Laplace transforms of the time-dependent distribution and the first passage time. These two transient characteristics were generally presumed to be intractable. Solving for the Laplace transforms involves using Green's functions and contour integrals related to hypergeometric functions. Our results are specialized to the $M/M/\infty$ queue, the $M/M/m$ queue, and the $M/M/m/m$ loss model. We also obtain some corresponding results for diffusion approximations to these models.

Keywords Erlang A model · Queueing theory · Complex analysis · QED regime · Time-dependent analysis

1 Introduction

In many real-world systems customers that are waiting for service may decide to abandon the system before entering service. In the process of designing systems, it is

✉ Johan S. H. van Leeuwaarden
  j.s.h.v.leeuwaarden@tue.nl

  Charles Knessl
  knessl@uic.edu

1 Department of Mathematics, Statistics and Computer Science, University of Illinois at Chicago, 815 South Morgan Street, Chicago, IL 60607-7045, USA

2 Department of Mathematics and Computer Science, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

⚛ Springer

important to understand the effect of this abandonment phenomenon on the system's behavior. There has been a huge effort in developing models for systems that incorporate the effect of abandonments, also referred to as reneging or impatience (see, e.g., Dai et al. 2010; Garnett et al. 2002; Ward and Glynn 2005; Whitt 2004, 2006a, b; Kang and Ramanan 2010; Zeltyn and Mandelbaum 2004, 2005). The simplest yet widely used model is the completely Markovian $M/M/m + M$ model, also known as the Erlang A model. Its performance analysis has been an important subject of study in the literature (see for example Garnett et al. 2002; Whitt 2006c), not only because the Erlang A model is being used in practice (Mandelbaum and Zeltyn 2007), but also because it delivers valuable approximations for more general abandonment models (Whitt 2005).

The Erlang A model assumes Poisson arrivals with rate λ, exponential service times with mean $1/\mu$, m parallel servers, and most importantly, it incorporates the feature that waiting customers abandon the system after exponentially distributed times with mean $1/\eta$. Let $N(t)$ denote the queue length at time t. Assuming independence across the interarrival, service and reneging times, the queue length process is a birth–death process $(N(t))_{t \geq 0}$. The stationary distribution of this process, and associated performance measures like delay or abandonment probabilities, are easy to obtain (Garnett et al. 2002; Mandelbaum and Zeltyn 2007). In contrast, studying the time-dependent behavior of $(N(t))_{t \geq 0}$ is generally judged to be prohibitively difficult (Fralix 2013; Ward 2012) because, among other things, the Kolmogorov forward equations do not seem to allow for a tractable solution. The main contributions of this paper are the exact solutions of both the forward and backward Kolmogorov equations, leading to exact expressions for the Laplace transforms of the time-dependent queue length distribution in Sect. 2 and first-passage times in Sect. 3.

The birth–death process describing the Erlang A model has birth rates, conditioned on $N(t) = j$, $\lambda_j = \lambda$ and death rates $\mu_j = j\mu$ for $j \leq m$ and $\mu_j = m\mu + (j - m)\eta$ for $j > m$. There are available general results for the time-dependent behavior of birth–death processes. Karlin and McGregor (1957a, b, 1958) have shown that the backward and forward Kolmogorov equations satisfied by the transition probabilities of a birth–death process can be solved via the introduction of a system of orthogonal polynomials and a spectral measure. For each set of birth and death rates (λ_j, μ_j) there is an associated family of orthogonal polynomials. In some cases, when the set (λ_j, μ_j) is assumed to have a special structure, these orthogonal polynomials can be identified. One such special case is the $M/M/m$ queue, with $\lambda_j = \lambda$ and $\mu_j = \min\{j, m\}\mu$. Notice that the Erlang A model incorporates the $M/M/m$ queue as the special case $\eta \to 0^+$. Karlin and McGregor (1958) have shown for the $M/M/m$ queue that the relevant orthogonal polynomials are the Poisson–Charlier polynomials. Determining the spectral measure, though, is rather complicated, which is why van Doorn (1979) made a separate study of determining the spectral properties of the $M/M/m$ queue, starting from the general expression for the spectral measure in Karlin and McGregor (1958) in terms of the Stieltjes transform. For the same $M/M/m$ queue, Saaty (1960) derived the Laplace transform of Prob$[N(t) = n]$ over time, in terms of hypergeometric functions. As in Saaty (1960), we do not resort to the approach in Karlin and McGregor (1957a, b, 1958) for solving the Erlang A model, but instead opt to derive the explicit solution for the Laplace transform of Prob$[N(t) = n]$ in a direct

manner. The inverse transform then gives the desired solution for the time-dependent distribution, and we can also obtain the time-dependent moments. Mathematically, we shall use discrete Green's functions, contour integrals, and special functions related to hypergeometric functions. Having explicit expressions for the Laplace transforms is useful for ultimately obtaining various asymptotic formulas, which would likely be simpler than the full solution and yield insight into model behavior.

Due to the cumbersome expressions for some of the stationary characteristics, and the presumed intractability of the time-dependent distribution, simpler analytically tractable processes $(D(t))_{t \geq 0}$ have been constructed that have similar time-dependent and stationary behaviors as $(N(t))_{t \geq 0}$. This can be done by imposing limiting regimes in which such approximating processes naturally arise as stochastic-process limits. Ward and Glynn (2002) make precise when the sample paths of the Erlang A model (and extensions using more general assumptions Ward and Glynn 2005) can be approximated by a diffusion process, where the type of diffusion process depends on the heavy-traffic regime. The diffusion process $(D(t))_{t \geq 0}$ is generally easier to study than the birth–death process $(N(t))_{t \geq 0}$, and can thus be employed to obtain simple approximations for both the stationary and the time-dependent system behavior. In Ward and Glynn (2002, 2003, 2005) the limiting diffusion process is a reflected Ornstein–Uhlenbeck process, whose properties are well understood (Fricker et al. 1999; Linetsky 2005; Ward and Glynn 2003). Garnett et al. (2002) proved a diffusion limit for the Erlang A model in another heavy-traffic regime, known as the Halfin–Whitt or QED regime. In this regime, the diffusion process $(D(t))_{t \geq 0}$ is a combination of two Ornstein–Uhlenbeck processes with different restraining forces, depending on whether the process is below or above zero. Both the stationary behavior (Garnett et al. 2002) and the time-dependent behavior (Leeuwaarden and Knessl 2012) of this process are well understood. From our general result for the Laplace transform of Prob$[N(t) = n]$ we show how the results obtained in Leeuwaarden and Knessl (2012) for the above diffusion processes can be recovered. See the survey paper Ward (2012) for a comprehensive overview of diffusion approximations for many-server systems with abandonments.

The paper is structured as follows. In Sect. 2 we work towards Theorems 1 and 2 that provide explicit expressions for the Laplace transform of the time-dependent distribution of $(N(t))_{t \geq 0}$. In order to do so, we first reformulate the forward Kolmogorov equations in terms of difference equations in Sect. 2.1 and Laplace transforms in Sect. 2.2. In these steps we identify several key special functions of which some properties are proved in Sect. 2.3. Then, in Sect. 2.4 we prove Theorems 1 and 2 using all preliminary results in Sects. 2.1–2.3. In Sect. 2.5 we consider the limiting steady state case, and in Sect. 2.6 we treat the special cases $\eta = \mu = 1$ ($M/M/\infty$ queue), $\eta \to 0^+$ ($M/M/m$ queue) and $\eta \to \infty$ (the $M/M/m/m$ loss model).

In Sect. 3 we follow a similar approach to obtain in Theorem 3 the Laplace transform of the distribution of the first time that $(N(t))_{t \geq 0}$ reaches some level $n_* > m$. This result is established in Sect. 3.1. We then again specialize the general results in Theorem 3 to several limiting cases. In particular, we derive in Sect. 3.2 results for the Halfin–Whitt regime. Finally, we give in Sect. 3.3 an expression for the mean first passage time.

2 Transient distribution

2.1 Expressions in terms of difference equations

We let $N(t)$ be the number of customers in the system and set

$$p_n(t) = \text{Prob}[N(t) = n \mid N(0) = n_0], \tag{2.1}$$

so that $p_n(t)$ depends parametrically on the initial condition n_0, as well as the model parameters m, η and $\rho = \lambda/\mu$. Since $N(t)$ is a birth–death process with birth rate λ, death rate (setting $\mu = 1$) $N(t)$, for $N(t) \leqslant m$, and death rate $m + [N(t) - m]\eta$, for $N(t) \geqslant m$, the forward Kolmogorov equations are

$$p_0'(t) = p_1(t) - \rho p_0(t) \tag{2.2}$$

$$p_n'(t) = \rho[p_{n-1}(t) - p_n(t)] + (n+1)p_{n+1}(t) - np_n(t), \quad 1 \leqslant n \leqslant m-1, \tag{2.3}$$

$$p_m'(t) = \rho[p_{m-1}(t) - p_m(t)] + (m+\eta)p_{m+1}(t) - mp_m(t), \tag{2.4}$$

and for $n \geqslant m+1$,

$$p_n'(t) = \rho[p_{n-1}(t) - p_n(t)] + [m + (n-m+1)\eta]p_{n+1}(t) - [m + (n-m)\eta]p_n(t) \tag{2.5}$$

with the initial condition

$$p_n(0) = \delta(n, n_0), \tag{2.6}$$

with $\delta(n, n_0) = 1$ for $n = n_0$ and $\delta(n, n_0) = 0$ for $n \neq n_0$. Setting

$$\widehat{P}_n(\theta) = \int_0^\infty e^{-\theta t} p_n(t)\, dt \tag{2.7}$$

and assuming that $0 < n_0 < m$ we obtain from (2.2) to (2.6)

$$\widehat{P}_1(\theta) - (\rho + \theta)\widehat{P}_0(\theta) = 0 \tag{2.8}$$

$$(n+1)\widehat{P}_{n+1}(\theta) + \rho\widehat{P}_{n-1}(\theta) - (\rho + \theta + n)\widehat{P}_n(\theta) = -\delta(n, n_0), \quad 0 < n < m, \tag{2.9}$$

$$[m + (n-m+1)\eta]\widehat{P}_{n+1}(\theta) + \rho\widehat{P}_{n-1}(\theta)$$
$$-[\rho + \theta + m + (n-m)\eta]\widehat{P}_n(\theta) = 0, \quad n \geqslant m. \tag{2.10}$$

If $n_0 = 0$ the right side of (2.8) must be replaced by -1, while if $n_0 \geqslant m$ the right side of (2.10) must be replaced by $-\delta(n, n_0)$, and then the right side of (2.9) is zero. We proceed to explicitly solve (2.8)–(2.10), distinguishing the cases $0 < n_0 < m$ and $n_0 > m$, and then we show that the results also apply for $n_0 = m$ and $n_0 = 0$.

2.2 Special function solutions to the difference equations

Since the coefficients in the difference equations in (2.9) and (2.10) are linear functions of n, we can solve these explicitly with the help of contour integrals. First consider the integral

$$F_n(\theta) \equiv \frac{1}{2\pi i} \int_{C_0} \frac{e^{\rho z}}{z^{n+1}(1-z)^\theta} \, dz, \tag{2.11}$$

where C_0 is a small circle in the z-plane, on which $|z| < 1$. The integrand in (2.11) is analytic inside the unit circle, if we define

$$(1-z)^\theta = |1-z|^\theta e^{i\theta \arg(1-z)}$$

with $|\arg(1-z)| < \pi$, so that for z real and $z < 1$, $\arg(1-z) = 0$. By expanding

$$(1-z)^{-\theta} = 1 + \theta z + \theta(\theta+1)z^2/2! + \dots \tag{2.12}$$

as a binomial series, we obtain the alternate form

$$\begin{aligned}
F_n(\theta) &= \sum_{\ell=0}^{n} \frac{\rho^{n-\ell}}{(n-\ell)!} \frac{\theta(\theta+1)\dots(\theta+\ell-1)}{\ell!} \\
&= \sum_{\ell=0}^{n} \frac{\rho^{n-\ell}}{(n-\ell)!\ell!} \frac{\Gamma(\theta+\ell)}{\Gamma(\theta)},
\end{aligned} \tag{2.13}$$

where $\Gamma(\cdot)$ is the Gamma function. It follows that $F_{-1}(\theta) = 0$, $F_0(\theta) = 1$ and $F_1(\theta) = \rho + \theta$, and hence $F_n(\theta)$ satisfies Eq. (2.8). Furthermore, from (2.11) we have

$$\begin{aligned}
&\rho(F_{n-1} - F_n) + (n+1)F_{n+1} - (n+\theta)F_n \\
&= \frac{1}{2\pi i} \int_{C_0} \frac{e^{\rho z}}{z^{n+1}} \frac{1}{(1-z)^\theta} \left[\rho(z-1) + \frac{n+1}{z} - n - \theta \right] dz \\
&= -\frac{1}{2\pi i} \int_{C_0} \frac{d}{dz} \left[\frac{e^{\rho z}}{z^{n+1}(1-z)^{\theta-1}} \right] dz = 0, \tag{2.14}
\end{aligned}$$

as the contour C_0 is closed and the integrand in (2.14) is a perfect derivative. Thus $F_n(\theta)$ provides a solution to the homogeneous version of (2.9) (with the right side replaced by zero). We shall solve (2.8)–(2.10) using a discrete Green's function approach, and this will require a second, linearly independent, solution to (2.9). Such a solution may be obtained by using the same integrand as in (2.11) but integrating over a different contour. Thus we let

$$G_n(\theta) = \frac{1}{2\pi i} \int_{C_1} \frac{e^{\rho z}}{z^{n+1}(z-1)^\theta} \, dz, \tag{2.15}$$

where C_1 goes from $-\infty - i\varepsilon$ to $-\infty + i\varepsilon$ ($\varepsilon > 0$), encircling $z = 1$ in the counter-clockwise sense (see Fig. 1). In (2.15) we use the branch $(z-1)^\theta = |z-1|^\theta e^{i\theta \arg(z-1)}$, where $|\arg(z - 1)| < \pi$, so the integrand is analytic in $\mathbb{C} - \{\text{Im}(z) = 0, \text{Re}(z) \leqslant 1\}$. By a calculation completely analogous to (2.14), and noting that C_1 begins and ends at $z = -\infty$, where the integrand in (2.15) decays exponentially to zero, we see that $G_n(\theta)$ satisfies the homogeneous form of (2.9). However, G_n does not satisfy the boundary equation in (2.8), and we now have

$$G_{-1}(\theta) = \frac{1}{2\pi i} \int_{C_1} \frac{e^{\rho z}}{(z-1)^\theta} \, dz = \frac{e^\rho \rho^{\theta-1}}{\Gamma(\theta)}. \qquad (2.16)$$

Now consider (2.10). We shall again construct two independent solutions to this difference equation. Let f_n satisfy $[\rho + \theta + m + (n - m)\eta]f_n = \rho f_{n-1} + [m + (n - m + 1)\eta]f_{n+1}$ and represent f_n as a contour integral, with

$$f_n = \int_C z^{-n-1} \mathcal{F}(z) \, dz, \qquad (2.17)$$

for some function $\mathcal{F}(\cdot)$ and contour C. Then we have

$$\int_C \frac{1}{z^{n+1}} \left[\rho + \theta + m + (n - m)\eta - \rho z - \frac{m}{z} - \frac{(n - m + 1)\eta}{z} \right] \mathcal{F}(z) \, dz = 0. \qquad (2.18)$$

We use integration by parts in (2.18) with

$$\int_C \frac{n}{z^{n+1}} \mathcal{F}(z) \, dz = \int_C \frac{z\mathcal{F}'(z)}{z^{n+1}} \, dz \qquad (2.19)$$

and for now assume that C is such that there are no boundary contributions arising in (2.19), from endpoints of C. Using (2.19) in (2.18) we can rewrite (2.18) as a contour integral of z^{-n-1} times a function of z only, and if (2.18) is to hold for all n we argue that this function must vanish. We thus obtain the following differential equation for $\mathcal{F}(z)$:

$$\eta\mathcal{F}'(z)(z - 1) + \mathcal{F}(z)\left[\rho + \theta + m(1 - \eta) - \rho z - \frac{m}{z}(1 - \eta) \right] = 0, \qquad (2.20)$$

whose solution is, up to a multiplicative constant,

$$\mathcal{F}(z) = e^{\rho z/\eta}(z - 1)^{-\theta/\eta} z^m z^{-m/\eta}. \qquad (2.21)$$

Now we use (2.21) in (2.17) and make two different choices of C and different branches of (2.21), to obtain two independent solutions to (2.10). Note that now (2.21)
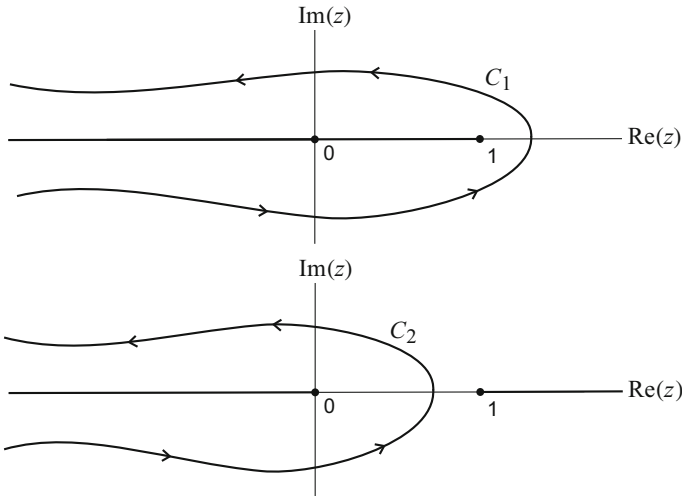
Fig. 1 A sketch of the branch cuts and the contours C_1 and C_2

has branch points both at $z = 1$ and $z = 0$. We let

$$H_n(\theta; m) = \frac{1}{2\pi i} \int_{C_1} \frac{e^{\rho z/\eta}}{(z-1)^{\theta/\eta} z^{n+1-m} z^{m/\eta}}\, dz, \tag{2.22}$$

where C_1 is as in (2.15) (or Fig. 1) and the branch of $(z-1)^{\theta/\eta}$ is

$$|z-1|^{\theta/\eta} \exp[i\theta \arg(z-1)/\eta] \tag{2.23}$$

with $|\arg(z-1)| < \pi$, and $z^{m/\eta} = |z|^{m/\eta} \exp[im(\arg z)/\eta]$ with $|\arg(z)| < \pi$.

Then the integrand in (2.22) is analytic in $\mathbb{C} - \{\mathrm{Im}(z) = 0, \mathrm{Re}(z) \leqslant 1\}$. For the second solution we set

$$I_n(\theta; m) = \frac{1}{2\pi i} \int_{C_2} \frac{e^{\rho z/\eta}}{(1-z)^{\theta/\eta} z^{n+1-m} z^{m/\eta}}\, dz, \tag{2.24}$$

where C_2 goes from $-\infty - i\varepsilon$ to $-\infty + i\varepsilon$, encircling $z = 0$ in the counterclockwise sense, and $(1-z)^{\theta/\eta}$ is defined to be analytic exterior to the branch cut where $\mathrm{Im}(z) = 0$ and $\mathrm{Re}(z) \geqslant 1$, similarly to (2.11). Also, $z^{m/\eta}$ is defined as below (2.22), so the integrand in (2.24) is analytic exterior to the branch cuts where $\mathrm{Im}(z) = 0$ and $\mathrm{Re}(z) \leqslant 0$ or $\mathrm{Re}(z) \geqslant 1$, and in particular on the contour C_2 (see again Fig. 1).

We have thus shown that the general solution to (2.10) is a linear combination of H_n and I_n, while that of (the homogeneous version of) (2.9) is a combination of F_n and G_n.

2.3 Properties of the key special functions

We now establish several useful properties of these functions. The integrals in (2.11), (2.15), (2.22) and (2.24) may all be expressed in terms of generalized hypergeometric functions, but we shall not use this fact. First, we note that if $\eta = 1$ then $F_n = I_n$ and $G_n = H_n$. The latter is obvious since (2.15) and (2.22) have the same contour C_1, while if $\eta = 1$ in (2.24) the branch point at $z = 0$ disappears and C_2 may be deformed to the loop C_0 in (2.11).

The functions H_n and I_n have very different asymptotic behaviors as $n \to \infty$. For n large, standard singularity analysis shows that the asymptotics of I_n are governed by the singularity at $z = 1$, and then setting $z = 1 - \xi/n$ and letting $n \to \infty$ in (2.24) yields

$$I_n \sim e^{\rho/\eta} \frac{1}{2\pi i} \int_{C_\xi} n^{\theta/\eta-1} e^\xi \xi^{-\theta/\eta} \, d\xi = n^{\theta/\eta-1} \frac{e^{\rho/\eta}}{\Gamma(\theta/\eta)}, \quad n \to \infty. \tag{2.25}$$

Here C_ξ goes from $-\infty - i\varepsilon$ to $-\infty + i\varepsilon$, with $\varepsilon > 0$, and encircles $\xi = 0$. Thus I_n has an algebraic dependence on n for n large. To expand H_n we can simply dilate the contour C_1 in (2.22) to the range $|z| \gg 1$ and then expand $(z - 1)^{\theta/\eta} = z^{\theta/\eta}[1 - \theta/(\eta z) + O(z^{-2})]$. We thus obtain

$$H_n(\theta) \sim \left(\frac{\rho}{\eta}\right)^{n-m+\frac{\theta+m}{\eta}} \frac{1}{\Gamma\left(n + 1 + \frac{m+\theta}{\eta} - m\right)}$$

$$\sim \frac{1}{n!} \left(\frac{\rho}{\eta}\right)^{n-m+\frac{\theta+m}{\eta}} n^{m-\frac{m+\theta}{\eta}}, \quad n \to \infty, \tag{2.26}$$

and hence H_n decays roughly as $1/n!$. Here we also used $\Gamma(n + x) \sim \Gamma(n)n^x$, which holds for $n \to \infty$ and x fixed. Next we consider the discrete Wronskian

$$W_n = W_n(\theta; m) = H_n(\theta; m)I_{n+1}(\theta; m) - H_{n+1}(\theta; m)I_n(\theta; m). \tag{2.27}$$

Using the fact that H_n and I_n satisfy (2.10) we find that

$$\rho(I_n H_{n-1} - I_{n-1}H_n) + [m + (n - m + 1)\eta][I_n H_{n+1} - H_n I_{n+1}] = 0 \tag{2.28}$$

and thus $W_n[m + (n - m + 1)\eta] = \rho W_{n-1}$. Solving this simple recurrence leads to

$$W_n(\theta; m) = \omega_*(\theta; m) \left(\frac{\rho}{\eta}\right)^n \frac{1}{\Gamma\left(n + \frac{m}{\eta} - m + 2\right)}. \tag{2.29}$$

To determine $\omega_*(\cdot)$ we let $n \to \infty$ in (2.27), and use (2.25) and (2.26). Then $H_{n+1} \ll H_n$ with $I_{n+1} = O(I_n)$, so that

$$W_n \sim H_n I_{n+1} \sim \frac{1}{n!} \frac{e^{\rho/\eta}}{\Gamma\left(\frac{\theta}{\eta}\right)} \left(\frac{\rho}{\eta}\right)^{n-m+\frac{\theta+m}{\eta}} n^{m-1-\frac{m}{\eta}}, \quad n \to \infty. \tag{2.30}$$

Comparing (2.29) with (2.30) we determine $\omega_*(\cdot)$, and thus

$$W_n = \frac{e^{\rho/\eta}}{\Gamma\left(\frac{\theta}{\eta}\right)} \left(\frac{\rho}{\eta}\right)^{n-m+\frac{\theta+m}{\eta}} \frac{1}{\Gamma\left(n-m+2+\frac{m}{\eta}\right)}. \tag{2.31}$$

A completely analogous calculation shows that

$$\widetilde{W}_n = -G_{n+1} F_n + G_n F_{n+1} = \frac{e^\rho}{\Gamma(\theta)} \frac{\rho^{n+\theta}}{(n+1)!}, \tag{2.32}$$

which also follows by setting $\eta = 1$ in (2.31).

2.4 Main results: solutions of the difference equations

We solve (2.8)–(2.10) for $0 < n_0 < m$, writing the solution as

$$\widehat{P}_n(\theta) = \begin{cases} AF_n(\theta), & 0 \leqslant n \leqslant n_0 \\ BF_n(\theta) + CG_n(\theta), & n_0 \leqslant n \leqslant m \\ DH_n(\theta; m), & n \geqslant m. \end{cases} \tag{2.33}$$

Then $\widehat{P}_n$ will decay faster than exponentially as $n \to \infty$, in view of (2.26), and satisfy the boundary equation in (2.8), since F_n does but G_n does not. It remains only to determine A, B, C, D; these functions will depend only on θ and the parameters m, η, ρ. By continuity at $n = n_0$ and $n = m$, we have

$$\begin{aligned} AF_{n_0} &= BF_{n_0} + CG_{n_0} \\ BF_m + CG_m &= DH_m. \end{aligned} \tag{2.34}$$

Then using (2.10) with $n = m$ leads to

$$(m + \theta + \rho) D H_m = \rho(B F_{m-1} + C G_{m-1}) + (m + n) D H_{m+1}, \tag{2.35}$$

and using (2.9) with $n = n_0$ (then $\delta(n, n_0) = 1$) leads to

$$\rho A F_{n_0-1} + (n_0 + 1)(B F_{n_0+1} + D G_{n_0+1}) - (\rho + \theta + n_0) A F_{n_0} = -1. \tag{2.36}$$

If we introduce a and α by setting

$$A = a[F_{n_0} + \alpha G_{n_0}] H_m \tag{2.37}$$

$$B = a F_{n_0} H_m, \quad C = \alpha a F_{n_0} H_m \tag{2.38}$$

$$D = a[F_m + \alpha G_m] F_{n_0}, \tag{2.39}$$

then both equations in (2.34) are satisfied. Using the fact that

$$\rho F_{n_0-1} + (n_0 + 1) F_{n_0+1} = (\rho + \theta + n_0) F_{n_0},$$

and $B - A = C\,G_{n_0}/F_{n_0}$, we obtain from (2.36)

$$C\left(-\frac{G_{n_0}}{F_{n_0}} F_{n_0+1} + G_{n_0+1}\right) = -\frac{1}{n_0 + 1}. \tag{2.40}$$

Then using the Wronskian identity in (2.32), with $n = n_0$, and (2.38) we see that

$$\alpha a H_m = \frac{n_0! \Gamma(\theta) e^{-\rho}}{\rho^{n_0+\theta}}. \tag{2.41}$$

Using the fact that H_n satisfies (2.10) with $n = m$, (2.35) is equivalent to $B F_{m-1} + C G_{m-1} = D H_{m-1}$ and using (2.38) and (2.39) leads to

$$H_m\, F_{m-1} + \alpha H_m\, G_{m-1} = F_m\, H_{m-1} + \alpha G_m\, H_{m-1},$$

and thus

$$\alpha = \frac{F_m\, H_{m-1} - H_m\, F_{m-1}}{H_m\, G_{m-1} - G_m\, H_{m-1}}. \tag{2.42}$$

Using (2.41) and (2.42) in (2.37)–(2.39), and then in (2.32) we have thus solved for $\widehat{P}_n(\theta)$, which we summarize below.

Theorem 1 *For initial conditions $0 \leqslant n_0 \leqslant m$, the Laplace transform $\widehat{P}_n(\theta) = \int_0^\infty e^{-\theta t} p_n(t)\, dt$ of the time dependent distribution of $N(t)$ is given by*

$$\widehat{P}_n(\theta) = \frac{n_0!}{m!} \rho^{m-n_0-1} \frac{F_{n_0}(\theta)\, H_n(\theta; m)}{F_m(\theta)\, H_{m-1}(\theta; m) - H_m(\theta; m)\, F_{m-1}(\theta)}, \quad n \geqslant m; \tag{2.43}$$

$$\widehat{P}_n(\theta) = \frac{n_0! \Gamma(\theta) e^{-\rho}}{\rho^{n_0+\theta}} F_{n_0}(\theta)$$
$$\times \left[G_n(\theta) + \frac{H_m(\theta; m)\, G_{m-1}(\theta) - G_m(\theta)\, H_{m-1}(\theta; m)}{F_m(\theta)\, H_{m-1}(\theta; m) - H_m(\theta; m)\, F_{m-1}(\theta)} F_n(\theta) \right], \quad n_0 \leqslant n \leqslant m; \tag{2.44}$$

$$\widehat{P}_n(\theta) = \frac{n_0! \Gamma(\theta) e^{-\rho}}{\rho^{n_0+\theta}} F_n(\theta)$$
$$\times \left[G_{n_0}(\theta) + \frac{H_m(\theta; m)\, G_{m-1}(\theta) - G_m(\theta)\, H_{m-1}(\theta; m)}{F_m(\theta)\, H_{m-1}(\theta; m) - H_m(\theta; m)\, F_{m-1}(\theta)} F_{n_0}(\theta) \right], \quad 0 \leqslant n \leqslant n_0. \tag{2.45}$$

Here $F_n\, G_n$, and H_n are given by the contour integrals in (2.13), (2.15) and (2.22).

Thus far we have established this result only for $0 < n_0 < m$. However, it holds also if $n_0 = 0$. We need only verify that (2.44) satisfies the boundary equation (2.8), which becomes $\widehat{P}_1(\theta) - (\rho + \theta)\widehat{P}_0(\theta) = -1$ if $n_0 = 0$. But $G_1 - (\rho + \theta)G_0$ can be computed from (2.15) as

$$
\begin{aligned}
G_1 - (\rho + \theta)G_0 &- \frac{1}{2\pi i}\int_{C_1}\frac{e^{\rho z}}{z^2(z-1)^\theta}[1-(\rho+\theta)z]\,dz \\
&= \frac{1}{2\pi i}\int_{C_1}\left\{-\rho\frac{e^{\rho z}}{(z-1)^\theta} + \frac{d}{dz}\left[\frac{e^{\rho z}}{z(z-1)^{\theta-1}}\right]\right\}\,dz \\
&= -\frac{e^\rho \rho^\theta}{\Gamma(\theta)}.
\end{aligned}
\tag{2.46}
$$

Since $F_1 = (\rho + \theta)F_0$, using (2.46) and (2.44) with $n_0 = 0$ yields

$$
\widehat{P}_1(\theta) - (\rho + \theta)\widehat{P}_0(\theta) = -1.
$$

We can also show that if $n_0 = m$, the expressions in Theorem 1 satisfy $(m + \eta)\widehat{P}_{m+1} + \rho\widehat{P}_{m-1} - (\rho + \theta + m)\widehat{P}_m = -1$, corresponding to initial conditions $N(0) = m$, i.e., starting with all m servers occupied but no one in the queue.

We note that if $n_0 = m$, (2.43)–(2.45) somewhat simplify, to

$$
\widehat{P}_n(\theta) = \frac{\rho^{-1}}{F_m\,H_{m-1} - H_m\,F_{m-1}}
\begin{cases}
F_m\,H_n, & n \geqslant m \\
H_m\,F_n, & 0 \leqslant n \leqslant m.
\end{cases}
\tag{2.47}
$$

Next we assume that $N(0) = n_0 > m$. Now we must solve the homogeneous form of (2.9) with the boundary condition in (2.8), and these imply that $\widehat{P}_n(\theta)$ must be proportional to F_n for all $0 \leqslant n \leqslant m$. Thus now G_n will not enter the analysis. For n large, $\widehat{P}_n(\theta)$ must again be proportional to H_n, which has the appropriate decay as $n \to \infty$. Now we write

$$
\widehat{P}_n(\theta) =
\begin{cases}
\widetilde{A}F_n(\theta), & 0 \leqslant n \leqslant m \\
\widetilde{B}H_n(\theta; m) + \widetilde{C}I_n(\theta; m), & m \leqslant n \leqslant n_0 \\
\widetilde{D}\,H_n(\theta; m), & n \geqslant n_0.
\end{cases}
\tag{2.48}
$$

Imposing the continuity conditions at $n = n_0$ and $n = m$ yields

$$
\widetilde{A}F_m = \widetilde{B}H_m + \widetilde{C}I_m
\tag{2.49}
$$
$$
\widetilde{D}H_{n_0} = \widetilde{B}H_{n_0} + \widetilde{C}I_{n_0}.
\tag{2.50}
$$

Setting $n = m$ in (2.10) then yields

$$
\begin{aligned}
(m + \eta)&\left[\widetilde{B}H_{m+1} + \widetilde{C}I_{m+1}\right] + \rho\widetilde{A}F_{m-1} \\
&= [\rho + \theta + m + (n-m)\eta]\left[\widetilde{B}H_m + \widetilde{C}I_m\right]
\end{aligned}
\tag{2.51}
$$

and (2.10) with $n = n_0$ and the right side replaced by $-\delta(n, n_0) = -1$ leads to

$$[m + (n_0 - m + 1)\eta]\widetilde{D}H_{n_0+1} + \rho(\widetilde{B}H_{n_0+1} + \widetilde{C}I_{n_0-1})$$
$$- [\rho + \theta + m + (n_0 - m)\eta]\widetilde{D}H_{n_0} = -1. \tag{2.52}$$

Thus (2.49)–(2.52) yields four equations for the four unknowns $\widetilde{A}$, $\widetilde{B}$, $\widetilde{C}$, $\widetilde{D}$. They can be solved similarly to (2.37)–(2.39), and we give below only the final result.

Theorem 2 *For initial conditions $n_0 \geqslant m$, $\widehat{P}_n(\theta)$ is given by*

$$\widehat{P}_n(\theta) = \frac{1}{\rho}e^{-\rho/\eta}\left(\frac{\eta}{\rho}\right)^{n_0-m-1+\frac{\theta+m}{\eta}}\Gamma\left(\frac{\theta}{\eta}\right)\Gamma\left(n_0 - m + 1 + \frac{m}{\eta}\right)$$
$$\times\left[I_{n_0} + \frac{I_m F_{m-1} - I_{m-1} F_m}{F_m H_{m-1} - H_m F_{m-1}}H_{n_0}\right]H_n, \quad n \geqslant n_0; \tag{2.53}$$

$$\widehat{P}_n(\theta) = \frac{1}{\rho}e^{-\rho/\eta}\left(\frac{\eta}{\rho}\right)^{n_0-m-1+\frac{\theta+m}{\eta}}\Gamma\left(\frac{\theta}{\eta}\right)\Gamma\left(n_0 - m + 1 + \frac{m}{\eta}\right)$$
$$\times\left[I_n + \frac{I_m F_{m-1} - I_{m-1} F_m}{F_m H_{m-1} - H_m F_{m-1}}H_n\right]H_{n_0}, \quad m \leqslant n \leqslant n_0; \tag{2.54}$$

$$\widehat{P}_n(\theta) = \frac{1}{\rho}\left(\frac{\rho}{\eta}\right)^{m-n_o}\frac{\Gamma\left(n_0 - m + 1 + \frac{m}{\eta}\right)}{\Gamma\left(1 + \frac{m}{\eta}\right)}\frac{H_{n_0}F_n}{F_m H_{m-1} - H_m F_{m-1}}, \quad 0 \leqslant n \leqslant m. \tag{2.55}$$

Here F_n, H_n and I_n are given by the contour integrals in (2.11), (2.22) and (2.24).

We note that if $n_0 = m$, expression (2.54) is not needed, and then (2.53) and (2.55) agree with the expression(s) in (2.47). Setting $n_0 = m$ in (2.53) and using

$$I_m H_{m-1} - I_{m-1} H_m = \frac{e^{\rho/\eta}}{\Gamma\left(\frac{\theta}{\eta}\right)\Gamma\left(\frac{m}{\eta}+1\right)}\left(\frac{\rho}{\eta}\right)^{\frac{\theta+m}{\eta}-1}, \tag{2.56}$$

which follows from (2.31), we obtain (2.47) for $n \geqslant m$.

2.5 Limiting case: steady state

We proceed to examine some limiting cases of Theorems 1 and 2, where the expressions simplify, sometimes considerably. First we consider the steady state limit of $p_n(t)$ as $t \to \infty$, which corresponds to the limit of $\theta\widehat{P}_n(\theta)$ as $\theta \to 0$. First, observe that as $\theta \to 0$, (2.11) and (2.15) yields

$$F_n(0) = \frac{\rho^n}{n!} = G_n(0), \tag{2.57}$$

while (2.22) and (2.24) lead to

$$H_n(0; m) = \left(\frac{\rho}{\eta}\right)^{n-m+m/\eta} \frac{1}{\Gamma\left(n-m+1+\frac{m}{\eta}\right)} = I_n(0; m). \qquad (2.58)$$

Now consider $n_0 \geqslant m$, where Theorem 2 applies. At $\theta = 0$,

$$F_m(0)\, H_{m-1}(0; m) = F_{m-1}(0)\, H_m(0; m) \qquad (2.59)$$

and $I_m(0; m)\, F_{m-1}(0) = I_{m-1}(0; m)\, F_m(0)$, in view of (2.57) and (2.58). To estimate the various terms in (2.53)–(2.55) as $\theta \to 0$, we first compute

$$\begin{aligned}
\Delta_1 &\equiv \frac{d}{d\theta}[F_m\, H_{m-1} - H_m\, F_{m-1}]\Big|_{\theta=0} \\
&= F_m'(0)\, H_{m-1}(0; m) + F_m(0)\, H_{m-1}'(0; m) \\
&\quad - H_m'(0; m)\, F_{m-1}(0) - H_m(0; m)\, F_{m-1}'(0).
\end{aligned} \qquad (2.60)$$

Using (2.57) and (2.22) we have

$$\begin{aligned}
F_m(0)\, &H_{m-1}'(0; m) - F_{m-1}(0)\, H_m'(0; m) \\
&= \frac{\rho^{m-1}}{m!} \left[\rho H_{m-1}'(0; m) - m H_m'(0; m)\right] \\
&= \frac{\rho^{m-1}}{m!} \frac{1}{2\pi i} \int_{C_1} \left[\frac{-\log(z-1)}{\eta} \frac{\rho e^{\rho z/\eta}}{z^{m/\eta}} + \frac{m}{\eta} \frac{\log(z-1)}{z^{1+m/\eta}} e^{\rho z/\eta}\right] dz \\
&= \frac{\rho^{m-1}}{m!} \frac{1}{2\pi i} \int_{C_1} \log(z-1) \frac{d}{dz}\left[-\frac{e^{\rho z/\eta}}{z^{m/\eta}}\right] dz \\
&= \frac{\rho^{m-1}}{m!} \frac{1}{2\pi i} \int_{C_1} \frac{e^{\rho z/\eta}}{z^{m/\eta}} \frac{1}{z-1} dz \\
&= \frac{\rho^{m-1}}{m!} \sum_{\ell=0}^{\infty} \left[\frac{1}{2\pi i} \int_{C_1} \frac{e^{\rho z/\eta}}{z^{\ell+1+m/\eta}} dz\right] \\
&= \frac{\rho^{m-1}}{m!} \sum_{\ell=0}^{\infty} \left(\frac{\rho}{\eta}\right)^{\ell+\frac{m}{\eta}} \frac{1}{\Gamma\left(\ell+1+\frac{m}{\eta}\right)}.
\end{aligned} \qquad (2.61)$$

We can take $|z| > 1$ on C_1, and then expand $(z-1)^{-1}$ as a Laurent series on C_1. Using (2.58) and (2.11) we have

$$\begin{aligned}
H_{m-1}(0; m)\, &F_m'(0) - H_m(0; m)\, F_{m-1}'(0) \\
&= \frac{H_m(0; m)}{\rho} \left[m F_m'(0) - \rho F_{m-1}'(0)\right] \\
&= \frac{H_m(0)}{\rho} \frac{1}{2\pi i} \int_{C_0} -\log(1-z) \left[\frac{m e^{\rho z}}{z^{m+1}} - \frac{\rho e^{\rho z}}{z^m}\right] dz
\end{aligned}$$

$$
\begin{aligned}
&= \frac{H_m(0)}{\rho} \frac{1}{2\pi i} \int_{C_0} \log(1-z) \frac{d}{dz}\left(\frac{e^{\rho z}}{z^m}\right) dz \\
&= \frac{H_m(0)}{\rho} \frac{1}{2\pi i} \int_{C_0} \frac{e^{\rho z}}{1-z} \frac{1}{z^m} dz \\
&= \rho^{-1} \left(\frac{\rho}{\eta}\right)^{m/\eta} \frac{1}{\Gamma\left(1+\frac{m}{\eta}\right)} \sum_{J=0}^{m-1} \frac{\rho^J}{J!},
\end{aligned}
\tag{2.62}
$$

where now on C_0 we can expand $(1-z)^{-1}$ as $\sum_{\ell=0}^{\infty} z^\ell$, since $|z| < 1$. Combining (2.61) with (2.62), (2.60) then yields

$$
\Delta_1 = \left(\frac{\rho}{\eta}\right)^{m/\eta} \frac{1}{\rho} \left[\frac{1}{\Gamma\left(1+\frac{m}{\eta}\right)} \sum_{J=0}^{m-1} \frac{\rho^J}{J!} + \frac{\rho^m}{m!} \sum_{\ell=0}^{\infty} \frac{(\rho/\eta)^\ell}{\Gamma\left(\ell+1+\frac{m}{\eta}\right)} \right]. \tag{2.63}
$$

Now let $\Delta_2 = \frac{d}{d\theta}[F_m I_{m-1} - I_m F_{m-1}\big|_{\theta=0}$. Since $I_m = H_m$ when $\theta = 0$, the difference between Δ_1 and Δ_2 is

$$
\begin{aligned}
\Delta_1 - \Delta_2 &= F_m(0)\left[H'_{m-1}(0;m) - I'_{m-1}(0;m)\right] \\
&\quad - F_{m-1}(0)\left[H'_m(0;m) - I'_m(0;m)\right] \\
&= \frac{\rho^{m-1}}{m!} \frac{1}{2\pi i} \left(\int_{C_1} - \int_{C_2}\right)\left(\frac{1}{z-1}\frac{e^{\rho z/\eta}}{z^{m/\eta}}\right) dz, \tag{2.64}
\end{aligned}
$$

when we used (2.24) and calculations similar to those in (2.61). But the difference between the contour integrals over C_1 and over C_2 is simply the residue from the pole at $z = -1$, and thus

$$
\Delta_1 - \Delta_2 = \frac{\rho^{m-1}}{m!} e^{\rho/\eta}. \tag{2.65}
$$

Using (2.63)–(2.65) we thus have

$$
\lim_{\theta \to 0}\left\{\left[I_{n_0} + \frac{I_m F_{m-1} - I_{m-1} F_m}{F_m H_{m-1} - H_m F_{m-1}} H_{n_0}\right] H_n\right\} = H_{n_0}(0;m) H_n(0;m)\left[1 - \frac{\Delta_2}{\Delta_1}\right],
\tag{2.66}
$$

and (2.66) can be used in view of (2.53) and (2.54), for both $n \in [m, n_0]$ and $n \geqslant n_0$. Then $\theta\Gamma(\theta/\eta) \to \eta$ as $\theta \to 0$ and $F_m H_{m-1} - H_m F_{m-1} = \theta\Delta_1 + O(\theta^2)$. We have thus obtained the steady state limit from Theorem 2 as stated below (see e.g., Garnett et al. 2002; Ward 2012).

Corollary 1 *The steady state distribution is*

$$p_n(\infty) = K \frac{\rho^m}{m!} \left(\frac{\rho}{\eta}\right)^{n-m} \frac{\Gamma\left(1+\frac{m}{\eta}\right)}{\Gamma\left(n-m+1+\frac{m}{\eta}\right)}, \quad n \geqslant m, \tag{2.67}$$

$$p_n(\infty) = K \frac{\rho^n}{n!}, \quad 0 \leqslant n \leqslant m, \tag{2.68}$$

with

$$K = \left[\sum_{J=0}^{m-1} \frac{\rho^J}{J!} + \frac{\rho^m}{m!} \sum_{\ell=0}^{\infty} \left(\frac{\rho}{\eta}\right)^\ell \frac{\Gamma\left(1+\frac{m}{\eta}\right)}{\Gamma\left(\ell+1+\frac{m}{\eta}\right)}\right]^{-1}. \tag{2.69}$$

Note that K and Δ_1 are related by $\rho\Delta_1\Gamma(1+m/\eta)(\rho/\eta)^{-m/\eta} K = 1$. While we obtained Corollary 1 from Theorem 2, which applies for $n_0 \geqslant m$, the result is independent of n_0 and Corollary 1 will also follow from Theorem 1 using very similar calculations to those in (2.60)–(2.66), which we omit. Of course, $p_n(\infty)$ is more easily obtained by letting $t \to \infty$ in (2.2)–(2.5) and solving the resulting elementary difference equations.

2.6 Limiting cases: extreme abandonments rates

Next we evaluate Theorems 1 and 2 for the special cases $\eta = 1$, $\eta \to 0^+$ (vanishing abandonment effects) and $\eta \to \infty$. For $\eta = 1$ the model reduces to the standard infinite server $M/M/\infty$ queue, and from Theorems 1 and 2 we obtain the following.

Corollary 2 *When $\eta = 1$ the Laplace transform of $p_n(t)$ is given by*

$$\widehat{P}_n(\theta) = \frac{\Gamma(\theta)n!e^{-\rho}}{\rho^{n_0+\theta}} \begin{cases} F_{n_0}(\theta)\,G_n(\theta), & n \geqslant n_0 \\ G_{n_0}(\theta)\,F_n(\theta), & 0 \leqslant n \leqslant n_0. \end{cases} \tag{2.70}$$

A spectral representation of $p_n(t)$ is then

$$p_n(t) = \frac{n_0!e^{-\rho}}{\rho^{n_0}} \sum_{k=0}^{\infty} \frac{\rho^k}{k!} e^{-kt} F_{n_0}(-k)\,F_n(-k) \tag{2.71}$$

where

$$F_n(-k) = \frac{1}{2\pi i} \int_{C_0} \frac{(1-z)^k}{z^{n+1}} e^{\rho z}\,dz = \sum_{j=0}^{\min\{k,n\}} \binom{k}{j} (-1)^j \frac{\rho^{n-j}}{(n-j)!}, \tag{2.72}$$

and an alternate form is given by

$$p_n(t) = \rho^n \left(1 - e^{-t}\right)^n \exp\left[-\rho\left(1 - e^{-t}\right)\right]$$

$$\times \sum_{j=0}^{\min\{n,n_0\}} \binom{n_0}{j} \rho^{-j} e^{-jt} \left(1 - e^{-t}\right)^{n_0 - 2j} \frac{1}{(n-j)!}, \qquad (2.73)$$

and then $p_n(\infty) = e^{-\rho} \rho^n / n!$.

We have already seen that when $\eta = 1$, $H_n = G_n$ and $I_n = F_n$ and we have the Wronskian identities in (2.31) and (2.32). Then both Theorems 1 and 2 reduce to (2.70), and we need not distinguish the cases $n_0 \gtrless m$, as m disappears altogether from the expressions. Now, from (2.11) and (2.15) it is clear that $F_n(\theta)$ and $G_n(\theta)$ are entire functions of θ, for every n. Thus the only singularities of (2.70) are the poles of $\Gamma(\theta)$, which occur at $\theta = -N$, $N = 0, 1, 2, \ldots$ and the corresponding residues are $(-1)^N / N!$. When $\theta = -N$, G_n and F_n are no longer linearly independent, and in fact $G_n(-N) = (-1)^N F_n(-N)$, which follows by comparing (2.15) with (2.11). Thus evaluating the contour integral $p_n(t) = (2\pi i)^{-1} \int_{\mathrm{Br}} e^{\theta t} \widehat{P}_n(\theta) \, d\theta$ (where $\mathrm{Re}(\theta) > 0$ on the vertical Bromwich contour) as a residue series we obtain precisely (2.71), with (2.72). To obtain the expression in (2.73) we represent the $F_n(-k)$ in (2.71) as contour integrals, yielding

$$p_n(t) = \frac{n_0!}{\rho^{n_0}} e^{-\rho(1-e^{-t})} \frac{1}{(2\pi i)^2} \int_{C_0} \int_{C_0} \frac{\exp\left(\rho z w e^{-t}\right)}{z^{n_0+1} w^{n+1}}$$

$$\times \exp\left[\rho\left(1 - e^{-t}\right) z + \rho\left(1 - e^{-t}\right) w\right] dz \, dw$$

$$= e^{-\rho(1-e^{-t})} \frac{1}{2\pi i} \int_{C_0} \frac{e^{\rho(1-e^{-t})w}}{w^{n+1}} \left[w e^{-t} + 1 - e^{-t}\right]^{n_0} dw. \qquad (2.74)$$

Then expanding $\left[w e^{-t} + 1 - e^{-t}\right]^{n_0}$ using the binomial theorem leads to (2.73). Note that as $t \to 0$ $\left(1 - e^{-t}\right)^{n+n_0-2j} \to 0$ unless $n = n_0$ and $j = n$, so that $p_n(0) = \delta(n, n_0)$. As $t \to \infty$ only the term with $j = 0$ in (2.73) remains and we obtain the steady state Poisson distribution. If $\eta = 1$ it is easier to solve (2.2)–(2.6) using the generating function $\mathcal{G}(t, u) = \sum_{n=0}^{\infty} p_n(t) u^n$ which leads to the first order PDE

$$\frac{\partial \mathcal{G}}{\partial t} + (u - 1) \frac{\partial \mathcal{G}}{\partial u} = \rho(u - 1)\mathcal{G}, \quad \mathcal{G}(0, u) = u^{n_0}, \qquad (2.75)$$

whose solution is

$$\mathcal{G}(t, u) = \exp\left[\rho\left(1 - e^{-t}\right)(u - 1)\right] \left[1 + (u - 1)e^{-t}\right]^{n_0}. \qquad (2.76)$$

Inverting the generating function then regains (2.73).

Next we let $\eta \to 0^+$, so that the model reduces to the m-server $M/M/m$ queue. Then we obtain the following.

Corollary 3 *When $\eta = 0$ the Laplace transform of $p_n(t)$ is given by, for $0 \leqslant n_0 \leqslant m$,*

$$\widehat{P}_n(\theta) = \frac{n_0!}{m!} \rho^{m-n_0} \frac{F_{n_0}(\theta)[A(\theta)]^{n-m}}{(m+1)F_{m+1}(\theta) - A(\theta)m F_m(\theta)}, \quad n \geqslant m, \tag{2.77}$$

$$A(\theta) = \frac{1}{2m} \left[m + \rho + \theta - \sqrt{(m+\rho+\theta)^2 - 4m\rho} \right], \tag{2.78}$$

$$\widehat{P}_n(\theta) = \frac{n_0!\Gamma(\theta)e^{-\rho}}{\rho^{n_0+\theta}} \left[G_n + \frac{mA G_m - (m+1)G_{m+1}}{(m+1)F_{m+1} - mA F_m} F_n \right] F_{n_0}, \quad n_0 \leqslant n \leqslant m, \tag{2.79}$$

$$\widehat{P}_n(\theta) = \frac{n_0!\Gamma(\theta)e^{-\rho}}{\rho^{n_0+\theta}} \left[G_{n_0} + \frac{mA G_m - (m+1)G_{m+1}}{(m+1)F_{m+1} - mA F_m} F_{n_0} \right] F_n, \quad 0 \leqslant n \leqslant n_0. \tag{2.80}$$

For $n_0 \geqslant m$ we have

$$\widehat{P}_n(\theta) = B^{m-n_0} \frac{F_n}{(m+1)F_{m+1} - Am F_m}, \quad 0 \leqslant n \leqslant m, \tag{2.81}$$

$$B(\theta) = \frac{1}{2m} \left[\rho + m + \theta + \sqrt{(m+\rho+\theta)^2 - 4m\rho} \right], \tag{2.82}$$

$$\widehat{P}_n(\theta) = \frac{1}{\sqrt{(m+\rho+\theta)^2 - 4m\rho}}$$
$$\times \left[B^{n-n_0} + A^{n-m} B^{m-n_0} \frac{(m+1)F_{m+1} - Bm F_m}{Am F_m - (m+1)F_{m+1}} \right], \quad m \leqslant n \leqslant n_0, \tag{2.83}$$

$$\widehat{P}_n(\theta) = \frac{1}{\sqrt{(m+\rho+\theta)^2 - 4m\rho}}$$
$$\times \left[A^{n-n_0} + B^{m-n_0} A^{n-m} \frac{(m+1)F_{m+1} - Bm F_m}{Am F_m - (m+1)F_{m+1}} \right], \quad n \geqslant n_0. \tag{2.84}$$

We also note that the transient distribution for the $M/M/m$ model was previously obtained, in different forms, by Saaty (1960) and van Doorn (1979). In van Doorn (1979) spectral methods are used, while in Saaty (1960) the Laplace transform is expressed in terms of hypergeometric functions.

To establish (2.77)–(2.84) we need to evaluate $H_n(\theta; m)$ and $I_n(\theta; m)$ for $\eta \to 0^+$. We write H_n in (2.22) as

$$H_n = \frac{1}{2\pi i} \int_{C_1} \frac{1}{z^{n+1-m}} \exp\left[\frac{1}{\eta} f(\theta, z) \right] dz \tag{2.85}$$

where

$$f(\theta, z) = \rho z - m \log z - \theta \log(z - 1), \tag{2.86}$$

so that the integrand has saddle points where $\partial f / \partial z = 0$, and this occurs at

$$z = Z_\pm(\theta) \equiv \frac{1}{2\rho} \left[\rho + \theta + m \pm \sqrt{(\rho + \theta + m)^2 - 4\rho m} \right]. \qquad (2.87)$$

We can take $|z| > 1$ on C_1 and then the saddle at Z_+ determines the asymptotic behavior of H_n as

$$H_n \sim \sqrt{\frac{\eta}{2\pi}} Z_+^{m-n-1} \left[\frac{m}{Z_+^2} + \frac{\theta}{(Z_+ - 1)^2} \right]^{-1/2}$$

$$\times \exp \left\{ \frac{1}{\eta} \left[\rho Z_+ - m \log Z_+ - \theta \log(Z_+ - 1) \right] \right\}, \quad \eta \to 0^+. \quad (2.88)$$

It follows that $H_{n-1}/H_n \sim Z_+$ in this limit,

$$\frac{H_m \, G_{m-1} - G_m \, H_{m-1}}{F_m \, H_{m-1} - H_m \, F_{m-1}} \to \frac{G_{m-1} - Z_+ \, G_m}{F_m \, F_+ - F_{m-1}}, \quad \eta \to 0^+, \qquad (2.89)$$

and

$$\frac{H_n}{F_m \, H_{m-1} - H_m \, F_{m-1}} \to \frac{Z_+^{m-n}}{F_m \, Z_+ - F_{m-1}}, \quad n \to 0^+. \qquad (2.90)$$

But, $\rho F_{m-1} + (m+1)F_{m+1} = (m + \rho + \theta)F_m$ so that

$$\rho Z_+ F_m - \rho F_{m-1} = \rho F_m (Z_+ - 1) - (m + \theta) F_m + (m+1)F_{m+1}$$
$$= (m+1)F_{m+1} - Am \, F_m$$

as $A = 1/Z_+$ and $Z_\pm$ satisfy the quadratic equation

$$\rho Z_\pm^2 - (\rho + m + \theta)Z_\pm + m = 0.$$

Hence (2.43) reduces to (2.77) as $\eta \to 0^+$. Also, (2.79) and (2.80) follow from (2.44) and (2.45), in view of (2.89) and the fact that

$$\frac{G_{m-1} - G_m \, Z_+}{F_m \, Z_+ - F_{m-1}} \cdot \frac{\rho}{\rho} = \frac{Am \, G_m - (m+1)G_{m+1}}{(m+1)F_{m+1} - Am \, f_m}.$$

Now consider $n_0 \geqslant m$. We shall obtain (2.81)–(2.84) from Theorem 2. We must then expand $I_n(\theta; m)$ for $\eta \to 0^+$. Using the saddle point method we find that

$$I_n \sim \sqrt{\frac{\eta}{2\pi}} Z_-^{m-n-1} \left[\frac{m}{Z_-^2} + \frac{\theta}{(1 - Z_-)^2} \right]^{-1/2}$$

$$\times \exp \left\{ \frac{1}{\eta} \left[\rho Z_- - m \log Z_- - \theta \log(1 - Z_-) \right] \right\}, \qquad (2.91)$$

as the expansion of (2.24), which involves the contour C_2, is determined by the other saddle point in (2.87). Using (2.88) with n replaced by n_0, (2.91), and Stirling's formula we obtain, after a lengthy calculation, the following limit (as $\eta \to 0^+$):

$$I_n(\theta; m) H_{n_0}(\theta; m) \Gamma\left(\frac{\theta}{\eta}\right) \Gamma\left(n_0 + 1 - m + \frac{m}{\eta}\right)$$

$$\times \frac{1}{\rho} e^{-\rho/\eta} \left(\frac{\eta}{\rho}\right)^{n_0 - m - 1 + \frac{\theta + m}{\eta}} \to Z_-^{n_0 - n} \frac{1}{\sqrt{(\rho + \theta + m)^2 - 4m\rho}}. \quad (2.92)$$

After factoring out I_n, the bracketed factor in (2.54) becomes

$$1 + \frac{H_n}{I_n} \frac{I_m F_{m-1} - I_{m-1} F_m}{F_m H_{m-1} - H_m F_{m-1}}$$

$$\to 1 + \frac{Z_-^{n-m} F_{m-1} - Z_-^{n+1-m} F_m}{Z_+^{n+1-m} F_m - Z_+^{n-m} F_{m-1}}$$

$$= 1 + Z_+^{m-n} Z_-^{n-m} \frac{F_{m-1} - Z_- F_m}{Z_+ F_m - F_{m-1}}$$

$$= 1 + A^{n-m} B^{m-n} \frac{(m+1) F_{m+1} - Bm F_m}{Am F_m - (m+1) F_{m+1}}, \quad (2.93)$$

where we again used $AZ_+ = 1$, $BZ_- = 1$ and the quadratic equation satisfied by $Z_\pm$. With (2.92) and (2.93) the expression in (2.54) becomes that in (2.83). A completely analogous calculation shows that (2.53) leads to (2.84) as $\eta \to 0^+$. Now consider (2.55). As $\eta \to 0^+$, by Stirling's formula,

$$\left(\frac{\rho}{\eta}\right)^{m-n_0} \frac{\Gamma\left(n_0 + 1 - m + \frac{m}{\eta}\right)}{\Gamma\left(1 + \frac{m}{\eta}\right)} \to m^{n_0 - m} \rho^{m - n_0}$$

and we also use (2.90) with n replaced by n_0, and

$$\rho[F_m Z_+ - F_{m-1}] = (m+1) F_{m+1} - Am F_m.$$

Then (2.55) goes to the limit in (2.81), since $\rho Z_+/m = B$. This completes the proof of Corollary 3.

In the limit $\eta \to \infty$, we expect our results to reduce to the Erlang loss model, or the $M/M/m/m$ queue. We then obtain the following.

Corollary 4 *As $\eta \to \infty$ the Laplace transform of $p_n(t)$, for $0 \leqslant n_0 \leqslant m$, approaches the limit*

$$\widehat{P}_n(\theta) = \frac{n_0! \Gamma(\theta) e^{-\rho}}{\rho^{n_0 + \theta}} F_{n_0}[G_n + \omega F_n], \quad n_0 \leqslant n \leqslant m \quad (2.94)$$

$$\widehat{P}_n(\theta) = \frac{n_0! \Gamma(\theta) e^{-\rho}}{\rho^{n_0 + \theta}} F_n[G_{n_0} + \omega F_{n_0}], \quad 0 \leqslant n \leqslant n_0, \quad (2.95)$$

where

$$\omega = \frac{(m+1)G_{m-1} - \rho G_m}{\rho F_m - (m+1)F_{m+1}}.$$

In particular the blocking probability $p_m(t)$ has the Laplace transform

$$\widehat{P}_m(\theta) = \frac{n_0!}{m!} \rho^{m-n_0} \frac{F_{n_0}(\theta)}{(m+1)F_{m+1}(\theta) - \rho F_m(\theta)}. \tag{2.96}$$

Note that (2.35) follows by setting $n = m$ in (2.94) and using the Wronskian $\widetilde{W}_m$ in (2.32). To establish Corollary 4, we note that by expanding the integrand in (2.22) for $\eta \to \infty$ we obtain

$$H_n = \delta(n, m) + \frac{1}{\eta}\left[\rho\delta(n, m+1) - \frac{1}{2\pi i}\int_{C_1} \frac{m\log z + \theta\log(z-1)}{z^{n+1-m}}\, dz\right] + O(\eta^{-2}) \tag{2.97}$$

and thus $H_m(\theta; m) = 1 + O(\eta^{-1})$ and $\eta H_{m+1}(\theta; m) \to \rho$ as $\eta \to \infty$. Since

$$\rho H_{m-1} + (m+\eta)H_{m+1} = (\rho + m + \theta)H_m$$

we have $H_{m-1} \to (\theta + m)/\rho$ as $\eta \to \infty$. Thus, as $\eta \to \infty$,

$$\frac{H_m\, G_{m-1} - G_m\, H_{m-1}}{F_m\, H_{m-1} - H_m\, F_{m-1}} \to \frac{\rho G_{m-1} - (\theta + m)G_m}{(\theta + m)F_m - \rho F_{m-1}}. \tag{2.98}$$

But

$$(m+1)G_{m+1} - \rho G_m = -\rho G_{m-1} + (\theta + m)G_m$$

and

$$(m+1)F_{m+1} - \rho F_m = -\rho F_{m-1} + (\theta + m)F_m,$$

so with (2.98), (2.47) and (2.45) yields (2.94) and (2.95) in the limit $\eta \to \infty$.

The blocking probability in (2.96) may also be written as

$$\widehat{P}_m(\theta) = \frac{n_0!}{m!} \rho^{m-n_0} \frac{F_{n_0}(\theta)}{\theta F_m(\theta+1)}, \tag{2.99}$$

since $\theta F_m(\theta + 1) = (m+1)F_{m+1}(\theta) - \rho F_m(\theta)$, which follows from (2.11) with $n = m$ and an integration by parts. If $N(0) = 0$ (starting with an empty system) we obtain from (2.99)

$$\widehat{P}_m(\theta) = \frac{\Gamma(\theta)}{\sum_{\ell=0}^{m} \binom{m}{\ell} \rho^{-\ell} \Gamma(\theta + \ell + 1)}. \tag{2.100}$$

Previously expressions for the Laplace transform of the blocking probability were obtained by Jagerman (2000), who showed that (if $n_0 = 0$)

$$\widehat{P}_m(\theta) = \frac{\Gamma(\theta)}{\int_0^\infty e^{-\xi} \xi^\theta \left(1 + \frac{\xi}{m\rho}\right)^m d\xi}, \tag{2.101}$$

and this can easily be shown to agree with both (2.100) and (2.99), as

$$\frac{\rho^m}{m!} \int_0^\infty e^{-\xi} \xi^\theta \left(1 + \frac{\xi}{m\rho}\right)^m d\xi = \frac{\Gamma(\theta + 1)}{2\pi i} \int_{C_0} \frac{e^{\rho z}}{z^{m+1}} (1 - z)^{-\theta - 1} dz \tag{2.102}$$

follows by expanding both integrands using the binomial theorem.

For general $n_0 \in [0, m]$ the blocking probability is given by

$$\widehat{P}_m(\theta) = \frac{\sum_{\ell=0}^{n_0} \binom{n_0}{\ell} \rho^{-\ell} \Gamma(\theta + \ell)}{\sum_{\ell=0}^{m} \binom{m}{\ell} \rho^{-\ell} \Gamma(\theta + \ell + 1)}. \tag{2.103}$$

Since the expressions in Theorems 1 and 2 and even Corollaries 3 and 4, are quite complicated, it is useful to expand these in various asymptotic limits. One such limit would have $m \to \infty$, $\rho \to \infty$ with $m/\rho \to 1$ and $m - \rho = O(\sqrt{m})$. This is a diffusion limit, sometimes referred to as the Halfin–Whitt regime. Here we would scale n, n_0 and ρ, for $m \to \infty$, as

$$\rho = m - \sqrt{m}\beta, \quad n = m + \sqrt{m}x, \quad n_0 = m + \sqrt{m}x_0, \tag{2.104}$$

and x, x_0 and β are $O(1)$. In this limit we can approximate the contour integrals F_n, G_n, H_n and I_n by simpler special functions, namely parabolic cylinder functions. We discuss this limit in detail in van Leeuwaarden and Knessl (2011) for the $M/M/m$ model with $\eta = 0$, and in Leeuwaarden and Knessl (2012) for the $M/M/m + M$ model with $\eta > 0$. We can obtain then $p_n(t) \sim m^{-1/2} P(x, t)$ where P will satisfy a parabolic PDE, which we explicitly solved in van Leeuwaarden and Knessl (2011), Leeuwaarden and Knessl (2012). An alternate approach is to evaluate Theorems 1 and 2, or Corollary 3 in the limit in (2.104), and thus identify $P(x, t)$ directly. We shall discuss in more detail the limit in (2.104) for the first passage distributions. We also comment that the transient behavior of the $M/M/m/m$ model was analyzed thoroughly in Knessl (1990) and Xie and Knessl (1993), for $m \to \infty$ and various cases of ρ, including the scaling in (2.104). There we used mostly singular perturbation methods, but equivalent results could be obtained using Corollary 4 and methods for asymptotically expanding integrals.

3 First passage times

3.1 Main result: Laplace transform of the first passage time

Here we compute the distribution of the time for the number $N(t)$ of customers to reach some level n_*, which may be viewed as a measure of congestion. We take $n_* > m$, for otherwise the problem reduces to that of the $M/M/\infty$ or $M/M/m/m$ models. Thus we define the stopping time

$$\tau(n_*) = \min\{t : N(t) = n_*\}, \tag{3.1}$$

and its conditional distribution is

$$Q_n(t)dt = \text{Prob}\left[\tau(n_*) \in (t, t + dt) \mid N(0) = n\right]. \tag{3.2}$$

When $n = n_*$ we clearly have

$$Q_{n_*}(t) = \delta(t) \tag{3.3}$$

and for $n < n_*$, $Q_n(t)$ satisfies the backward Kolmogorov equation(s)

$$Q_0'(t) = \rho Q_1(t) - \rho Q_0(t) \tag{3.4}$$

$$Q_n'(t) = \rho Q_{n+1}(t) + n Q_{n-1}(t) - (\rho + n)Q_n(t), \quad 1 \leqslant n \leqslant m, \tag{3.5}$$

$$Q_n'(t) = \rho Q_{n+1}(t) + [m + (n - m)\eta]Q_{n-1}(t)$$
$$- [\rho + m + (n - m)\eta]Q_n(t), \quad m \leqslant n \leqslant n_*. \tag{3.6}$$

To analyze (3.3)–(3.6) we first introduce the Laplace transform

$$\widehat{Q}_n(\theta) = \int_0^\infty e^{-\theta t} Q_n(t) \, dt$$

and, expecting that $Q_n(0) = 0$ for $n < n_*$, we obtain

$$\widehat{Q}_{n_*}(\theta) = 1 \tag{3.7}$$

$$\rho \widehat{Q}_1(\theta) = (\rho + \theta)\widehat{Q}_0(\theta) \tag{3.8}$$

$$(\rho + n + \theta)\widehat{Q}_n(\theta) = \rho \widehat{Q}_{n+1}(\theta) + n\widehat{Q}_{n-1}(\theta), \quad 1 \leqslant n \leqslant m, \tag{3.9}$$

$$[\rho + m + (n - m)\eta + \theta]\widehat{Q}_n(\theta) = \rho \widehat{Q}_{n+1}(\theta)$$
$$+ [m + (n - m)\eta]\widehat{Q}_{n-1}(\theta), \quad m \leqslant n \leqslant n_* - 1. \tag{3.10}$$

The recurrences in (3.9) and (3.10) are similar to those in (2.9) and (2.10), and indeed we can convert the former to the latter by setting

$$\widehat{Q}_n(\theta) = \rho^{-n} \frac{n!}{m!} R_n(\theta), \quad 0 \leqslant n \leqslant m \tag{3.11}$$

$$\widehat{Q}_n(\theta) = \rho^{-n}\eta^{n-m}\frac{\Gamma\left(n-m+1+\frac{m}{\eta}\right)}{\Gamma\left(1+\frac{m}{\eta}\right)}R_n(\theta), \quad m \leqslant n \leqslant n_*. \tag{3.12}$$

Then from (3.7) and (3.12) we have

$$R_{n_*}(\theta) = \eta^{m-n_*}\rho^{n_*}\frac{\Gamma\left(1+\frac{m}{\eta}\right)}{\Gamma\left(n_*-m+1+\frac{m}{\eta}\right)}, \tag{3.13}$$

and $R_n(\theta)$ will satisfy

$$(\rho+n+\theta)R_n = (n+1)R_{n+1} + \rho R_{n-1}$$

for $0 < n < m$, which is just the homogeneous version of (2.9), while for $n > m$, $R_n(\theta)$ will satisfy (2.10). Also, $R_1(\theta) = (\rho+\theta)R_0(\theta)$, so that $R_n(\theta)$ will satisfy the boundary equation in (2.8). We can thus write R_n in terms of the special functions F_n, G_n, H_n, I_n that we introduced in Sect. 2, and since F_n satisfies (2.8) we write

$$R_n(\theta) = c_1 F_n(\theta), \quad 0 \leqslant n \leqslant m \tag{3.14}$$

and

$$R_n(\theta) = c_2 H_n(\theta; m) + c_3 I_n(\theta; m), \quad m \leqslant n \leqslant n_*. \tag{3.15}$$

In view of (3.15) and (3.13) we have

$$c_2 H_{n_*} + c_3 I_{n_*} = \eta^{m-n_*}\rho^{n_*}\frac{\Gamma\left(1+\frac{m}{\eta}\right)}{\Gamma\left(n_*-m+1+\frac{m}{\eta}\right)} \tag{3.16}$$

and if both (3.14) and (3.15) apply for $n = m$ we have the continuity equation

$$c_1 F_m = c_2 H_m + c_3 I_m. \tag{3.17}$$

Finally, using (3.5) with $n = m$ and noting that, in view of (3.11) and (3.12),

$$\begin{aligned} \widehat{Q}_m - \widehat{Q}_{m-1} &= \rho^{-m}\left[R_m - \frac{\rho}{m}R_{m-1}\right], \\ \widehat{Q}_{m+1} - \widehat{Q}_m &= \rho^{-m-1}\left[(m+\eta)R_{m-1} - \rho R_m\right], \end{aligned} \tag{3.18}$$

we find that $(m+\eta)R_{m+1} + \rho R_{m-1} = (\theta+\rho+m)R_m$ and thus

$$(m+\eta)\left[c_2 H_{m+1} + c_3 I_{m+1}\right] + \rho c_1 F_{m-1} = (\theta+\rho+m)c_1 F_m. \tag{3.19}$$

Then (3.16), (3.17) and (3.19) yield three equations for the unknowns c_1, c_2, c_3. After some algebra and use of (2.31) with $n = m$ we obtain R_n, and then $\widehat{Q}_n$ follows from (3.11) and (3.12). We summarize below the final results.

Theorem 3 *The distribution of the first passage time to a level $n_*(> m)$ has the Laplace transform $\widehat{Q}_n(\theta) = E\left[e^{-\theta\tau(n_*)} \mid N(0) = n\right]$:*

$$\widehat{Q}_n(\theta) = \rho^{n_*-n}\frac{n!}{m!}\eta^{m-n_*+1}\left(\frac{\rho}{\eta}\right)^{\frac{m+\theta}{\eta}}\frac{e^{\rho/\eta}}{\Gamma\left(\frac{\theta}{\eta}\right)\Gamma\left(n_*-m+1+\frac{m}{\eta}\right)}$$

$$\times\frac{F_n}{(m+\eta)(H_{n_*}I_{m+1}-I_{n_*}H_{m+1})F_m+(m+1)(H_m I_{n_*}-H_{n_*}I_m)F_{m+1}},$$
$$0\leqslant n\leqslant m, \tag{3.20}$$

$$\widehat{Q}_n(\theta) = \rho^{n_*-n}\eta^{n-n_*}\frac{\Gamma\left(n-m+1+\frac{m}{\eta}\right)}{\Gamma\left(n_*-m+1+\frac{m}{\eta}\right)}$$

$$\times\frac{(m+\eta)(H_n I_{m+1}-I_n H_{m+1})F_m+(m+1)(H_m I_n-H_n I_m)F_{m+1}}{(m+\eta)(H_{n_*}I_{m+1}-I_{n_*}H_{m+1})F_m+(m+1)(H_m I_{n_*}-H_{n_*}I_m)F_{m+1}},$$
$$m\leqslant n\leqslant n_*. \tag{3.21}$$

Note that actually (3.20) can be used even if $n = m + 1$ and it then agrees with (3.21). Similarly, (3.21) holds even if $n = m - 1$. If $\eta = 1$ we have $F_n = I_n$ and then both (3.20) and (3.21) reduce to

$$\widehat{Q}_n(\theta) = \frac{n!}{n_*!}\rho^{n_*-n}\frac{F_n(\theta)}{F_{n_*}(\theta)}, \quad 0\leqslant n\leqslant n_* \tag{3.22}$$

which is the result for the $M/M/\infty$ model. We can again get results for the standard $M/M/m$ model by letting $\eta \to 0^+$ in Theorem 3. Using the asymptotic results in (2.88) and (2.91), after some calculations that we omit we obtain the following.

Corollary 5 *For the $M/M/m$ model the first passage distribution to a level $n_*(> m)$ is given by*

$$\widehat{Q}_n(\theta) = \rho^{m-n}\frac{n!}{m!}\sqrt{(\theta+m+\rho)^2-4m\rho}$$

$$\times\frac{F_n(\theta)}{\rho F_m(Z_+Z_-^{n_*-m}-Z_-Z_+^{n_*-m})+(m+1)F_{m+1}(Z_+^{n_*-m}-Z_-^{n_*-m})},$$
$$0\leqslant n\leqslant m \tag{3.23}$$

and

$$\widehat{Q}_n(\theta) = \frac{\rho F_m(Z_+Z_-^{n-m}-Z_-Z_+^{n-m})+(m+1)F_{m+1}(Z_+^{n-m}-Z_-^{n-m})}{\rho F_m(Z_+Z_-^{n_*-m}-Z_-Z_+^{n_*-m})+(m+1)F_{m+1}(Z_+^{n_*-m}-Z_-^{n_*-m})},$$
$$m\leqslant n\leqslant n_*. \tag{3.24}$$

Here $Z_\pm$ are as in (2.87).

Using the fact that $F_n(0) = \rho^n/n!$ and $Z_\pm(0) = [m + \rho \pm |m - \rho|]/(2\rho)$ we can easily verify that $\widehat{Q}_n(0) = 1$ for all n, so that the density is properly normalized. We shall discuss later the mean first passage time, which is equal to $-\widehat{Q}'_n(0)$.

3.2 Halfin–Whitt regime

We next consider the limit in (2.104) in Corollary 5, also scaling the exit point n_* as

$$n_* = m + \sqrt{m}b, \quad 0 < b < \infty. \tag{3.25}$$

From (2.87) we obtain, using (2.104),

$$Z_\pm = 1 + \frac{1}{2\sqrt{m}}\left[\beta \pm \sqrt{\beta^2 + 4\theta}\right] + O(m^{-1}), \quad m \to \infty$$

and hence

$$Z_\pm^{n-m} \sim \exp\left[\frac{1}{2}\left(\beta \pm \sqrt{\beta^2 + 4\theta}\right)x\right]. \tag{3.26}$$

By scaling $z = 1 - \xi/\sqrt{m}$ in (2.11) and noting that $\rho z - n \log z = \rho + (x + \beta)\xi + \frac{1}{2}\xi^2 + o(1)$ with the Halfin–Whitt scaling in (2.104), the integral in (2.11) can be approximated by

$$F_n(\theta) \sim \frac{1}{2\pi i}\frac{m^{\theta/2}e^\rho}{\sqrt{m}}\int_{\mathrm{Br}_+}\xi^{-\theta}e^{(x+\beta)\xi}e^{\xi^2/2}\,d\xi$$
$$= \frac{m^{\theta/2}e^\rho}{\sqrt{2\pi m}}e^{-(x+\beta)^2/4}D_{-\theta}(-x-\beta), \tag{3.27}$$

where $D_p(z)$ is the parabolic cylinder function of index p and argument z. In (3.27) the approximating contour Br_+ is a vertical contour in the ξ-plane, on which $\mathrm{Re}(\xi) > 0$, and $\xi^{-\theta}$ is defined to be analytic for $\mathrm{Re}(\xi) > 0$ and real and positive for ξ real and positive. In view of (2.104), setting $n = m$ corresponds to $x = 0$ and thus

$$F_m(\theta) \sim \frac{m^{\theta/2}e^\rho}{\sqrt{2\pi m}}e^{-\beta^2/4}D_{-\theta}(-\beta), \quad m \to \infty. \tag{3.28}$$

A similar calculation shows that

$$F_{m+1}(\theta) - F_m(\theta) \sim \frac{m^{\theta/2}e^\rho}{\sqrt{2\pi m}}e^{-\beta^2/4}D_{1-\theta}(-\beta), \quad m \to \infty \tag{3.29}$$

and we note that the difference $F_{m+1} - F_m$ is smaller than F_m by a factor of $m^{-1/2}$.

We write the denominator in (3.23) and (3.24) as

$$
\rho F_m \left[Z_+ Z_-^{n_*-m} - Z_- Z_+^{n_*-m} \right] - (m+1) F_{m+1} \left[Z_-^{n_*-m} - Z_+^{n_*-m} \right]
$$

$$
= -(m+1)(F_{m+1} - F_m) \left(Z_-^{n_*-m} - Z_+^{n_*-m} \right)
$$

$$
+ Z_-^{n_*-m} F_m (\rho Z_+ - m - 1) + Z_+^{n_*-m} F_m (-\rho Z_- + m + 1)
$$

$$
\sim \frac{m^{\theta/2}}{\sqrt{2\pi}} e^{-\beta^2/4} e^{b\beta/2} \left\{ 2 D_{1-\theta}(-\beta) \sinh \left(\frac{b}{2} \sqrt{\beta^2 + 4\theta} \right) \right.
$$

$$
+ e^{-\sqrt{\beta^2+4\theta} b/2} \frac{1}{2} \left[-\beta + \sqrt{\beta^2 + 4\theta} \right] D_{-\theta}(-\beta)
$$

$$
\left. + e^{\sqrt{\beta^2+4\theta} b/2} \frac{1}{2} \left[\beta + \sqrt{\beta^2 + 4\theta} \right] D_{-\theta}(-\beta) \right\}. \tag{3.30}
$$

Here we used (3.28), (3.29), (3.26), and also

$$
\rho Z_\pm - m - 1 \sim \frac{1}{2} \sqrt{m} \left[-\beta \pm \sqrt{\beta^2 + 4\theta} \right].
$$

The expansion of the numerator in (3.24) follows by replacing b by x in (3.30). In the limit in (2.104) we also have, using Stirling's formula,

$$
\rho^{m-n} \frac{n!}{m!} \sqrt{(\theta + m + \rho)^2 - 4m\rho} \sim e^{x\beta} e^{x^2/2} \sqrt{m} \sqrt{\beta^2 + 4\theta}. \tag{3.31}
$$

We summarize below our final results.

Corollary 6 *In the limit $m \to \infty$, with the scaling in (2.104) and (3.25), the transform of the first passage distribution $\widehat{Q}_n(\theta)$ for the $M/M/m$ model has the limit $\widehat{\mathcal{P}}(x, \theta)$ where*

$$
\widehat{\mathcal{P}}(x, \theta) = e^{x\beta/2} e^{x^2/4} \sqrt{\beta^2 + 4\theta} e^{-\beta b/2} \frac{D_{-\theta}(-\beta - x)}{\Lambda(\theta; b, \beta)}, \quad -\infty < x \leqslant 0 \tag{3.32}
$$

with

$$
\Lambda(\theta; b, \beta) = \sqrt{\beta^2 + 4\theta} \cosh \left(\frac{b}{2} \sqrt{\beta^2 + 4\theta} \right) D_{-\theta}(-\beta)
$$

$$
+ \sinh \left(\frac{b}{2} \sqrt{\beta^2 + 4\theta} \right) \left[2 D_{1-\theta}(-\beta) + \beta D_{-\theta}(-\beta) \right] \tag{3.33}
$$

and

$$
\widehat{\mathcal{P}}(x, \theta) = \frac{\Lambda(\theta; x, \beta)}{\Lambda(\theta; b, \beta)}, \quad 0 \leqslant x \leqslant b. \tag{3.34}
$$

We have previously obtained these results in Fralix et al. (2014), by directly solving the parabolic PDE satisfied by the diffusion approximation. Since $2D_{1-\theta}(-\beta) + \beta D_{-\theta}(-\beta) = -2D'_{-\theta}(-\beta)$, Corollary 6 agrees with Theorems 1 and 2 in Fralix et al. (2014).

Now, we can also consider the Halfin–Whitt limit for the first passage distribution in the $M/M/m + M$ model (with a fixed $\eta > 0$), and then Theorem 3 reduces to the following.

Corollary 7 *For $m \to \infty$ with the scaling in (2.104) and (3.25), $\widehat{Q}(\theta)$ in the $M/M/m + M$ model has the limit $\widehat{\mathcal{P}}(x, \theta)$ where*

$$\widehat{\mathcal{P}}(x, \theta) = \frac{e^{\beta(x-b)/2} e^{(x^2-\eta b^2)/4} \sqrt{2\pi} D_{-\theta}(-\beta - x)}{\Gamma\left(\frac{\theta}{\eta}\right) \left[D_{-\theta/\eta}\left(\frac{\beta+\eta b}{\sqrt{\eta}}\right) \Lambda_1 + D_{-\theta/\eta}\left(\frac{-\beta-\eta b}{\sqrt{\eta}}\right) \Lambda_2\right]}, \quad -\infty < x \leqslant 0,$$

$$(3.35)$$

$$\Lambda_1 = -\sqrt{\eta} D'_{-\theta/\eta}\left(\frac{-\beta}{\sqrt{\eta}}\right) D_{-\theta}(-\beta) + D_{-\theta/\eta}\left(\frac{-\beta}{\sqrt{\eta}}\right) D'_{-\theta}(-\beta), \quad (3.36)$$

$$\Lambda_2 = -\sqrt{\eta} D'_{-\theta/\eta}\left(\frac{\beta}{\sqrt{\eta}}\right) D_{-\theta}(-\beta) - D_{-\theta/\eta}\left(\frac{\beta}{\sqrt{\eta}}\right) D'_{-\theta}(-\beta), \quad (3.37)$$

$$\widehat{\mathcal{P}}(x, \theta) = e^{\beta(x-b)/2} e^{\eta(x^2-b^2)/4}$$

$$\times \frac{D_{-\theta/\eta}\left(\frac{\beta+\eta x}{\sqrt{\eta}}\right) \Lambda_1 + D_{-\theta/\eta}\left(\frac{-\beta-\eta x}{\sqrt{\eta}}\right) \Lambda_2}{D_{-\theta/\eta}\left(\frac{\beta+\eta b}{\sqrt{\eta}}\right) \Lambda_1 + D_{-\theta/\eta}\left(\frac{-\beta-\eta b}{\sqrt{\eta}}\right) \Lambda_2}, \quad 0 \leqslant x < b.$$

$$(3.38)$$

We can show that as $\eta \to 0^+$, Corollary 7 reduces to Corollary 6, so that the order of the limits of small η and that in (2.104) may be, in this case, interchanged. While we can obtain Corollary 7 from Theorem 3 by expanding H_n and I_n in the limit in (2.104), where

$$H_n \sim \sqrt{\frac{\eta}{2\pi m}} e^{\rho/\eta} \left(\frac{m}{\eta}\right)^{\frac{\theta}{2\eta}} e^{-(\eta x+\beta)^2/(4\eta)} D_{-\theta/\eta}\left(\frac{\eta x+\beta}{\sqrt{\eta}}\right),$$

and a similar expression holds for I_n, it is easier to simply obtain a limiting PDE from (3.5) and (3.6) (or limiting ODE from (3.9) and (3.10)) and solve it. If $\sqrt{m}\widehat{Q}_n(\theta) \to \widehat{\mathcal{P}}(x, \theta)$ then $\widehat{\mathcal{P}}$ must satisfy

$$\theta\widehat{\mathcal{P}} = \widehat{\mathcal{P}}_{xx} - (\beta + \eta x)\widehat{\mathcal{P}}_x, \quad x < 0, \quad (3.39)$$

$$\theta\widehat{\mathcal{P}} = \widehat{\mathcal{P}}_{xx} - (\beta + x)\widehat{\mathcal{P}}_x, \quad 0 < x < b, \quad (3.40)$$

and the boundary condition is $\widehat{\mathcal{P}}(b, \theta) = 1$. We also have the interface conditions $\widehat{\mathcal{P}}(0^-, \theta) = \widehat{\mathcal{P}}(0^+, \theta)$ and $\widehat{\mathcal{P}}_x(0^-, \theta) = \widehat{\mathcal{P}}_x(0^+, \theta)$, where subscripts denote partial derivatives. Setting

$$\widehat{\mathcal{P}}(x, \theta) = e^{x^2/4} e^{\beta x/2} \widetilde{\mathcal{P}}(x, \theta), \quad x < 0 \tag{3.41}$$

$$\widehat{\mathcal{P}}(x, \theta) = e^{\eta x^2/4} e^{\beta x/2} \widetilde{\mathcal{P}}(x, \theta), \quad 0 < x < b \tag{3.42}$$

we obtain from (3.39) and (3.40)

$$\widetilde{\mathcal{P}}_{xx} + \left[\frac{1}{2} - \theta - \frac{1}{4}(\beta + x)^2 \right] \widetilde{\mathcal{P}} = 0, \quad x < 0 \tag{3.43}$$

$$\widetilde{\mathcal{P}}_{xx} + \left[\frac{\eta}{2} - \theta - \frac{1}{4}(\beta + \eta x)^2 \right] \widetilde{\mathcal{P}} = 0, \quad 0 < x < b, \tag{3.44}$$

and $\widetilde{\mathcal{P}}$ and $\widetilde{\mathcal{P}}_x$ must also be continuous at $x = 0$, in view of (3.41) and (3.42) and the continuity of $\widehat{\mathcal{P}}$ and $\widehat{\mathcal{P}}_x$. Also, the boundary condition is

$$\widetilde{\mathcal{P}}(b, \theta) = \exp\left[-\frac{1}{4}\eta b^2 - \frac{1}{2}\beta b \right]. \tag{3.45}$$

Equation (3.43) is the parabolic cylinder equation of index $-\theta$, and its two linearly independent solution are $D_{-\theta}(\beta + x)$ and $D_{-\theta}(-\beta - x)$, for $-\theta \neq 0, 1, 2, \ldots$. But as $x \to -\infty$ $D_{-\theta}(\beta + x)$ has Gaussian growth in x, which would lead to $\widehat{\mathcal{P}}$ in (3.41) being roughly $O(e^{x^2/2})$ as $x \to -\infty$. Thus for $x < 0$ the solution must be proportional to $D_{-\theta}(-\beta - x)$, hence we write

$$\widetilde{\mathcal{P}}(x, \theta) = a(\theta) D_{-\theta}(-\beta - x), \quad x < 0. \tag{3.46}$$

The equation in (3.44) may be transformed, by the substitution

$$y = (\beta + \eta x)/\sqrt{\eta},$$

into a parabolic cylinder equation of index $-\theta/\eta$, and thus for $x > 0$ we have

$$\widetilde{\mathcal{P}}(x, \theta) = b(\theta) D_{-\theta/\eta}\left(\frac{\beta + \eta x}{\sqrt{\eta}} \right) + c(\theta) D_{-\theta/\eta}\left(\frac{-\beta - \eta x}{\sqrt{\eta}} \right). \tag{3.47}$$

The continuity conditions at $x = 0$ then yield

$$a(\theta) D_{-\theta}(-\beta) = b(\theta) D_{-\theta/\eta}\left(\frac{\beta}{\sqrt{\eta}} \right) + c(\theta) D_{-\theta/\eta}\left(\frac{-\beta}{\sqrt{\eta}} \right) \tag{3.48}$$

and

$$-a(\theta) D'_{-\theta}(-\beta) = b(\theta) \sqrt{\eta} D_{-\theta/\eta}\left(\frac{\beta}{\sqrt{\eta}} \right) - c(\theta) \sqrt{\eta} D'_{-\theta/\eta}\left(\frac{-\beta}{\sqrt{\eta}} \right). \tag{3.49}$$

Using the Wronskian identity

$$-D_{-\theta/\eta}\left(\frac{\beta}{\sqrt{\eta}}\right)D'_{-\theta/\eta}\left(\frac{-\beta}{\sqrt{\eta}}\right) - D_{-\theta/\eta}\left(\frac{-\beta}{\sqrt{\eta}}\right)D'_{-\theta/\eta}\left(\frac{\beta}{\sqrt{\eta}}\right) = \frac{\sqrt{2\pi}}{\Gamma\left(\frac{\theta}{\eta}\right)} \qquad (3.50)$$

we solve the system (3.45), (3.48) and (3.49), for the unknowns $a(\theta)$, $b(\theta)$, $c(\theta)$. We thus find that

$$b(\theta) = \frac{a(\theta)}{\sqrt{2\pi}}\Gamma\left(\frac{\theta}{\eta}\right)\Lambda_1, \quad c(\theta) = \frac{a(\theta)}{\sqrt{2\pi}}\Gamma\left(\frac{\theta}{\eta}\right)\Lambda_2, \qquad (3.51)$$

where the Λ_j are as in (3.36) and (3.37), and

$$a(\theta) = \frac{\sqrt{2\pi}}{\Gamma\left(\frac{\theta}{\eta}\right)} \frac{e^{-\eta b^2/4}e^{-\beta b/2}}{D_{-\theta/\eta}\left(\frac{-\beta-\eta b}{\sqrt{\eta}}\right)\Lambda_2 + D_{-\theta/\eta}\left(\frac{\beta+\eta b}{\sqrt{\eta}}\right)\Lambda_1}. \qquad (3.52)$$

Using (3.51) and (3.52) in (3.46), (3.47), (3.41) and (3.42) gives the result in Corollary 7.

3.3 Mean first passage time

Finally, we give below the mean first passage time,

$$q_n = E\left[\tau(n_*) \mid N(0) = n\right] = \int_0^\infty t\,Q_n(t)\,dt = -\widehat{Q}'_n(0). \qquad (3.53)$$

Corollary 8 *The conditional mean time to reach $N(t) = n_*$ starting from $N(0) = n \leqslant n_*$ is*

$$q_n = q_m + \sum_{j=n}^{m-1} j!\rho^{-j}\left[\sum_{\ell=0}^{j}\frac{\rho^{\ell-1}}{\ell!}\right], \quad 0 \leqslant n \leqslant m, \qquad (3.54)$$

$$q_m = \frac{1}{\rho}\sum_{J=m}^{n_*-1}\left[\left(\frac{\rho}{\eta}\right)^{m-J}\frac{\Gamma\left(J-m+1+\frac{m}{\eta}\right)}{\Gamma\left(1+\frac{m}{\eta}\right)}\sum_{\ell=0}^{m}\frac{m!}{\ell!}\rho^{\ell-m}\right.$$

$$\left. + \sum_{\ell=m+1}^{J}\left(\frac{\rho}{\eta}\right)^{\ell-J}\frac{\Gamma\left(J-m+1+\frac{m}{\eta}\right)}{\Gamma\left(\ell-m+1+\frac{m}{\eta}\right)}\right], \qquad (3.55)$$

$$q_n = \frac{1}{\rho} \sum_{J=n}^{n_*-1} \sum_{\ell=m+1}^{J} \left(\frac{\rho}{\eta}\right)^{\ell-J} \frac{\Gamma\left(J - m + 1 + \frac{m}{\eta}\right)}{\Gamma\left(\ell - m + 1 + \frac{m}{\eta}\right)}$$

$$+ \frac{1}{\rho} \left(\frac{\rho}{\eta}\right)^m \frac{1}{\Gamma\left(1 + \frac{m}{\eta}\right)} \left[\sum_{\ell=0}^{m} \frac{m!}{\ell} \rho^{\ell-m}\right]$$

$$\times \left[\sum_{J=n}^{n_*-1} \left(\frac{\eta}{\rho}\right)^J \Gamma\left(J - m + 1 + \frac{m}{\eta}\right)\right], \quad m \leqslant n < n_*, \tag{3.56}$$

with $q_{n_*} = 0$.

We note that using (2.22) we have

$$H_n(0) = \left(\frac{\rho}{\eta}\right)^{n-m+\frac{m}{\eta}} \frac{1}{\Gamma\left(n - m + 1 + \frac{m}{\eta}\right)} \tag{3.57}$$

and the expression in (3.50) may also be written as

$$q_n = \frac{1}{\rho} \left[\sum_{J=n}^{n_*-1} \sum_{\ell=m+1}^{J} \frac{H_\ell(0)}{H_J(0)} + \sum_{J=n}^{n_*-1} \frac{H_m(0)}{H_J(0)} \sum_{\ell=0}^{m} \frac{m!}{\ell!} \rho^{\ell-m}\right], \quad m \leqslant n < n_* \tag{3.58}$$

By multiplying (3.4)–(3.6) by t and integrating from $t = 0$ to $t = \infty$ we see that q_n satisfies the recurrence(s)

$$\rho(q_{n+1} - q_n) + n(q_{n-1} - q_n) = -1, \quad 0 \leqslant n \leqslant m, \tag{3.59}$$

$$\rho(q_{n+1} - q_n) + [m + (n - m)\eta](q_{n-1} - q_n) = -1, \quad m \leqslant n \leqslant n_* - 1, \tag{3.60}$$

with $q_{n_*} = 0$. Solving the difference equations in (3.59) and (3.60) by elementary methods leads to Corollary 8. The same results can be obtained by computing $-\widehat{Q}_n'(0)$ using the expressions in Theorem 3.

References

Dai JG, He S, Tezcan T (2010) Many-server diffusion limits for $G/Ph/n + M$ queues. Ann Appl Probab 20:1854–1890

Fralix B (2013) On the time-dependent moments of Markovian queues with reneging. Queueing Syst 75:149–168

Fralix B, Knessl C, van Leeuwaarden JSH (2014) First passage times to congested states of many-server systems in the Halfin–Whitt regime. Stoch Models 30:162–186

Fricker C, Robert Ph, Tibi D (1999) On the rates of convergence of Erlang's model. J Appl Probab 36:1167–1184

Garnett O, Mandelbaum A, Reiman M (2002) Designing a call center with impatient customers. Manuf Serv Oper Manag 4:208–227

Jagerman DL (2000) Difference equations with applications to queues. Monographs and textbooks in pure and applied mathematics. Marcel Dekker, New York

Kang W, Ramanan K (2010) Fluid limits of many-server queues with reneging. Ann Appl Probab 20:2204–2260

Karlin S, McGregor JL (1957) The differential equation of birth and death processes, and the Stieltjes moment problem. Trans Am Math Soc 85:489–546

Karlin S, McGregor JL (1957) The classification of birth and death processes. Pac J Math Trans Am Math Soc 86:366–401

Karlin S, McGregor JL (1958) Many server queueing processes with Poisson input and exponential service times. Pac J Math 8:87–118

Knessl C (1990) On the transient behavior of the Erlang loss model. Commun Stat Stoch Models 6:749–776

Linetsky V (2005) On the transition densities for reflected diffusions. Adv Appl Probab 37:435–460

Mandelbaum A, Zeltyn S (2007) Service engineering in action: the Palm/Erlang-A queue, with applications to call centers. Advances in services innovations. Springer, New York, pp 17–48

Saaty TL (1960) Time-dependent solution of the many-server Poisson queue. Oper Res 8:755–772

van Doorn EA (1979) Stochastic monotonicity and birth–death processes. Ph.D. thesis, Twente

van Leeuwaarden JSH, Knessl C (2011) Transient analysis of the Halfin–Whitt diffusion. Stoch Process Appl 121:1524–1545

van Leeuwaarden JSH, Knessl C (2012) Spectral gap of the Erlang A model in the Halfin–Whitt regime. Stoch Models 2:149–207

Ward A, Glynn P (2002) A diffusion approximation for a Markovian queue with reneging. Queueing Syst 43:103–128

Ward A, Glynn P (2003) Properties of the reflected Ornstein–Uhlenbeck process. Queueing Syst 44:109–123

Ward A, Glynn P (2005) A diffusion approximation for a GI/GI/1 queue with balking or reneging. Queueing Syst 50:371–400

Ward A (2012) Asymptotic analysis of queueing systems with reneging: a survey of results for FIFO, single class models. Surv Oper Res Manag Sci 17:1–14

Whitt W (2006) Fluid limits for many-server queues with abandonments. Oper Res 54:363–372

Whitt W (2006) Heavy-traffic limits for the $G/H_2^*/n/m$ queue. Math Oper Res 30:1–27

Whitt W (2006) Sensitivity of performance in the Erlang-A queueing model to changes in the model parameters. Oper Res 54:247–260

Whitt W (2004) Efficiency-driven heavy-traffic approximations for many-server queues with abandonments. Manag Sci 50:1449–1461

Whitt W (2005) Engineering solution of a basic call-center model. Manag Sci 51:221–235

Xie S, Knessl C (1993) On the transient behavior of the Erlang loss model: heavy usage asymptotics. SIAM J Appl Math 53:555–599

Zeltyn S, Mandelbaum A (2004) The impact of customers patience on delay and abandonment: some empirically-driven experiments with the $M/M/n + G$ queue. OR Spectr 26:377–411

Zeltyn S, Mandelbaum A (2005) Call centers with impatient customers: many-server asymptotics of the $M/M/n + G$ queue. Queueing Syst 51:361–402