

ON A PROCESSOR SHARING QUEUE THAT MODELS BALKING

Qiang Zhen^{*} Johan S. H. van Leeuwen[†] Charles Knessl[‡]

March 09, 2009

Abstract

We consider the processor sharing $M/M/1$ -PS queue which also models balking. A customer that arrives and sees n others in the system “balks” (i.e., decides not to enter) with probability $1 - b_n$. If b_n is inversely proportional to $n + 1$, we obtain explicit expressions for a tagged customer’s sojourn time distribution. We consider both the conditional distribution, conditioned on the number of other customers present when the tagged customer arrives, as well as the unconditional distribution. We then evaluate the results in various asymptotic limits. These include large time (tail behavior) and/or large n , lightly loaded systems where the arrival rate $\lambda \rightarrow 0$, and heavily loaded systems where $\lambda \rightarrow \infty$. We find that the asymptotic structure for the problem with balking is much different from the standard $M/M/1$ -PS queue. We also discuss a perturbation method for deriving the asymptotics, which should apply to more general balking functions.

^{*}Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, 851 South Morgan (M/C 249), Chicago, IL 60607-7045, USA. Email: qzhen2@uic.edu

[†]Department of Mathematics and Computer Science, Eindhoven University of Technology, Room HG 9.13, P.O. Box 513, 5600 MB Eindhoven, The Netherlands. Email: j.s.h.v.leeuwen@tue.nl

[‡]Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, 851 South Morgan (M/C 249), Chicago, IL 60607-7045, USA. Email: knessl@uic.edu

Acknowledgements: Knessl was partly supported by NSF grant DMS 05-03745 and NSA grant H 98230-08-1-0102. Van Leeuwen was supported by a VENI grant from The Netherlands Organization for Scientific Research (NWO).

1 Introduction

Balking refers to the phenomenon that customers, when forced to wait for service, refuse to join the queue. First introduced by Haight [7], balking can be specified by a probability distribution associated with the system state. Specifically, a customer that finds n customers in the systems upon arrival, balks with probability $1 - b_n$ and joins the queue with probability b_n . Haight considered several examples of balking functions, including $b_n = 1/(n + 1)$, $b_n = 1\{n \leq K\}$ (with $1\{\cdot\}$ the indicator function), and $b_n = e^{-cn}$. The latter case was also studied in detail by Morse [12]. In this paper we shall investigate the effect of balking upon systems with processor sharing (PS).

A model for the round-robin scheduling mechanisms in time-shared computer systems, processor sharing was first introduced by Kleinrock [8], and refers to the service discipline under which every customer gets a fair share of the server. It is by now well known that PS is intimately related to random order of service (ROS), which refers to the discipline where customers are chosen for service at random. First studied by Vaultot [15] and Pollaczek [13], the ROS discipline has a long tradition in queueing theory. Pollaczek obtained the Laplace transform of the distribution of the steady-state waiting time \mathcal{W} in the $M/M/1$ -ROS queue, by solving a differential-difference equation. In fact, the latter was almost identical to the differential-difference equation studied by Coffman, Muntz and Trotter [2] for the $M/M/1$ -PS queue. Indeed, by comparing these differential-difference equations, it is readily established that (see Cohen [3])

$$\text{Prob}[\mathcal{V} > t] = C \cdot \text{Prob}[\mathcal{W} > t] \quad (1.1)$$

with \mathcal{V} the steady-state sojourn time in the $M/M/1$ -PS queue and C a constant. A probabilistic argument based on coupling was given in Borst, Boxma, Morrison and Núñez-Queija [1], and the equivalence result was shown to extend to other models as well, including finite capacity queues, repairman problems and networks. We shall show that for the $M/M/1$ queue with balking the equivalence result also holds.

The distribution of \mathcal{W} does not have a simple representation. Pollaczek [13] was able to invert the Laplace transform and obtained a rather intricate but explicit integral representation for the waiting time distribution. The integral, along with the method of steepest descent, allowed Pollaczek to derive an intriguing asymptotic expression for the tail distribution $\text{Prob}[\mathcal{W} >$

t]. This asymptotic expression was rediscovered by Flatto [5]. Morrison [11] considered the heavy-traffic limit, where the traffic intensity $\rho \rightarrow 1$, and derived asymptotic expansions in powers of $1 - \rho$ for the sojourn time distributions. The tail results of Pollaczek and Flatto were related to the heavy-traffic results in Morrison recently in Zhen and Knessl [18].

In this paper we consider the $M/M/1$ -PS (or ROS) queue with balking. The sojourn time distribution (or waiting time distribution in the ROS model) satisfies a differential-difference equation that differs only slightly from the one considered by Pollaczek for the systems without balking. However, the analysis, and also the system behavior, changes drastically. We shall assume that $b_n = 1/(n + 1)$, so that the non-balking probability exactly matches the share of a server that the customer gets upon arrival. For this choice of b_n the differential-difference equation allows for an exact and asymptotic analysis. We obtain the following results:

- (i) An exact spectral representation for the sojourn time density in terms of generalized Laguerre polynomials.
- (ii) An expression for the Laplace-Stieltjes transform of the sojourn time density.
- (iii) Asymptotic results for tail probabilities when ρ is fixed; asymptotics for the light-traffic case where $\rho \rightarrow 0$; and asymptotics for the heavy-traffic case where $\rho \rightarrow \infty$.

The heavy-traffic asymptotics are derived using a singular perturbation approach. It is also explained how this approach might be useful for analyzing models with more general balking functions.

1.1 Equivalence relation

We denote the sojourn time of a non-balking customer that arrives to a PS queue with n other customers competing for service by \mathcal{V}_n , and the waiting time of a non-balking customer that arrives to a ROS queue with n other customers waiting for service and one additional customer in service by \mathcal{W}_n . Then we let b_n and b_n^r be the non-balking probabilities in the PS system and the ROS system, respectively, when there are n customers in the system (including the customer in service).

Proposition 1. *If $b_0^r = 1$ and $b_n^r = b_{n-1}$, $n = 1, 2, \dots$, then*

$$\mathcal{V}_n \stackrel{d}{=} \mathcal{W}_n \quad (1.2)$$

and

$$\text{Prob}[\mathcal{V} > t] = C \cdot \text{Prob}[\mathcal{W} > t] \quad (1.3)$$

with

$$C = \frac{1}{\rho} \cdot \frac{1 + \sum_{n=1}^{\infty} \rho^n b_0 \cdots b_{n-1}}{1 + \sum_{n=1}^{\infty} \rho^n b_0 \cdots b_n}. \quad (1.4)$$

Proof. Borst, Boxma, Morrison and Núñez-Queija [1] made the observation that whenever a service completion occurs in the PS system, each of the customers present is equally likely to be the one that departs due to the memoryless property of the exponential distribution. In that respect, the pool of customers competing for service under PS behaves exactly as the pool of customers waiting for service under ROS. Note that the arrival processes in both systems can be coupled due to the assumption that $b_0^r = 1$ and $b_n^r = b_{n-1}$, $n = 1, 2, \dots$. A similar coupling argument as in [1] then yields (1.2).

Let \mathcal{N}_p and \mathcal{N}_r denote the number of customers at arrival epochs in the PS system and ROS system, respectively. Then, $\text{Prob}[\mathcal{N}_p = n] = \pi_0 \rho^n b_0 \cdots b_{n-1}$ and

$$\text{Prob}[\mathcal{V} > t] = \frac{\sum_{n=0}^{\infty} \rho^n b_0 \cdots b_n \text{Prob}[\mathcal{V}_n > t]}{\sum_{n=0}^{\infty} \rho^n b_0 \cdots b_n}. \quad (1.5)$$

Similarly, $\text{Prob}[\mathcal{N}_r = n] = \pi_0^r \rho^n b_0^r \cdots b_{n-1}^r$ and

$$\text{Prob}[\mathcal{W} > t] = \frac{\sum_{n=0}^{\infty} \rho^{n+1} b_0^r \cdots b_{n+1}^r \text{Prob}[\mathcal{W}_n > t]}{\sum_{n=0}^{\infty} \rho^n b_0^r \cdots b_n^r}. \quad (1.6)$$

Upon comparing (1.5) and (1.6), and using $b_0^r = 1$, $b_n^r = b_{n-1}$ for $n = 1, 2, \dots$, and $\mathcal{V}_n \stackrel{d}{=} \mathcal{W}_n$, the equivalence relation (1.3) follows. \square

The $M/M/1/K$ -PS queue can be viewed as a special case of the $M/M/1$ -PS queue with balking by choosing $b_n = 1$ if $n \leq K - 1$ and 0 otherwise. In that case the equivalence relation becomes

$$\text{Prob}[\mathcal{V} > t] = \frac{1}{\rho} \cdot \frac{1 - \rho^{K+1}}{1 - \rho^K} \cdot \text{Prob}[\mathcal{W} > t], \quad (1.7)$$

which was already obtained in [1]. For this $M/M/1/K$ -PS queue, Knessl [9] uses singular perturbation techniques to construct asymptotic approximations to the sojourn time distribution.

We shall consider the $M/M/1$ -PS queue with balking $b_n = \frac{1}{n+1}$, in which case

$$\text{Prob}[\mathcal{N}_p = n] = \frac{e^{-\rho} \rho^n}{n!}, \quad n = 0, 1, \dots \quad (1.8)$$

and

$$\text{Prob}[\mathcal{V} > t] = \frac{e^\rho}{e^\rho - 1} \cdot \text{Prob}[\mathcal{W} > t]. \quad (1.9)$$

2 Problem statement and summary of results

We consider a processor sharing $M/M/1$ queue, which also models balking. Customers arrive at rate λ and the service rate will be denoted by μ . We can clearly scale time so as to have $\mu = 1$, and then the traffic intensity is $\rho = \lambda/\mu = \lambda$. We let \mathcal{V}_n be the sojourn time of a tagged customer that finds n others in the system upon arrival. We then define

$$\mathbf{V}_n(t) = \text{Prob}[\mathcal{V}_n > t]. \quad (2.1)$$

With the PS discipline, each customer receives service at rate $1/n$ when there are n customers in service. When a tagged customer arrives we assume that he/she will enter the system with probability b_n , and “balk” with the remaining probability $1 - b_n$.

The function $\mathbf{V}_n(t)$ satisfies the differential-difference equation

$$\mathbf{V}'_n(t) = \frac{n}{n+1} \mathbf{V}_{n-1}(t) - \left(1 + \rho b_n\right) \mathbf{V}_n(t) + \rho b_n \mathbf{V}_{n+1}(t) \quad (2.2)$$

with $\mathbf{V}_n(0) = 1$. (There is a slight error in [14], where this equation was previously given.) It is reasonable to define $b_0 = 1$ and have b_n a decreasing function of n . Here we assume that

$$b_n = \frac{1}{n+1}, \quad n = 0, 1, 2, 3, \dots \quad (2.3)$$

Then we define the sojourn time density $p_n(t)$ by $p_n(t) = -\mathbf{V}'_n(t)$, which satisfies:

$$p'_n(t) = \frac{n}{n+1} p_{n-1}(t) - \left(1 + \frac{\rho}{n+1}\right) p_n(t) + \frac{\rho}{n+1} p_{n+1}(t), \quad t > 0 \quad (2.4)$$

with the initial condition

$$p_n(0) = \frac{1}{n+1}, \quad n \geq 0. \quad (2.5)$$

The above can be obtained by integrating (2.2) from $t = 0$ to $t = \infty$ and using $\mathbf{V}_n(0) = 1$.

Clearly (2.3) is a very special case of b_n . We can consider also $b_n = \alpha/(n+1)$, since α can be incorporated into the traffic intensity ρ . However, with even a slight change (such as taking $b_n = 1/(n+\beta)$ with $\beta \neq 1$) it seems that the problem is no longer amenable to exact solution. We shall discuss an asymptotic approach to solving (2.2) (cf. section 5), which should work also for more general b_n .

We give below various exact and asymptotic expressions for $p_n(t)$.

Theorem 2.1. *The conditional sojourn time density has the following exact expression (spectral representation):*

$$p_n(t) = \sum_{m=1}^{\infty} C_m(\nu_m) \phi_m(n, \nu_m) e^{\nu_m t} + \sum_{m=1}^{\infty} C_m(\tilde{\nu}_m) \phi_m(n, \tilde{\nu}_m) e^{\tilde{\nu}_m t}, \quad (2.6)$$

where

$$\nu_m = -1 + \frac{1}{2m} \left[-\rho + \sqrt{\rho^2 + 4m\rho} \right], \quad (2.7)$$

$$\tilde{\nu}_m = -1 + \frac{1}{2m} \left[-\rho - \sqrt{\rho^2 + 4m\rho} \right], \quad (2.8)$$

$$C_m(\nu) = \frac{m^{m-1}}{m!} \frac{\nu}{\nu-1} e^{-m}, \quad (2.9)$$

and

$$\phi_m(n, \nu) = n! \left(\frac{\nu+1}{-\rho} \right)^n L_n^{(m-1-n)} \left(\frac{\rho}{(\nu+1)^2} \right). \quad (2.10)$$

Here $L_n^{(\alpha)}(z)$ is the generalized Laguerre polynomial (see [10]).

If we take the Laplace transform of (2.4) and multiply by $n+1$, we have

$$\rho \widehat{p}_{n+1}(\theta) - [(n+1)(\theta+1) + \rho] \widehat{p}_n(\theta) + n \widehat{p}_{n-1}(\theta) = -1, \quad (2.11)$$

where $\widehat{p}_n(\theta) = \int_0^\infty p_n(t) e^{-\theta t} dt$. Solving the recurrence equation (2.11), we obtain another exact expression for $p_n(t)$, in terms of its Laplace transform.

Theorem 2.2. *The Laplace-Stieltjes transform of the conditional sojourn time density has the following form:*

$$\hat{p}_n(\theta) = M G_n \sum_{l=0}^n \frac{\rho^l}{l!} H_l + M H_n \sum_{l=n+1}^{\infty} \frac{\rho^l}{l!} G_l, \quad (2.12)$$

where

$$M = M(\theta) \equiv \frac{\rho^{r+1}}{(1+\theta)\Gamma(r+1)} e^{r/\theta}, \quad (2.13)$$

$$G_n = G_n(\theta) \equiv \int_0^{\frac{1}{(1+\theta)}} z^n \left(\frac{1}{1+\theta} - z \right)^r \exp\left(-\frac{\rho z}{1+\theta}\right) dz, \quad (2.14)$$

$$H_n = H_n(\theta) \equiv \int_{\frac{1}{(1+\theta)}}^{\infty} z^n \left(z - \frac{1}{1+\theta} \right)^r \exp\left(-\frac{\rho z}{1+\theta}\right) dz, \quad (2.15)$$

and

$$r = r(\theta) \equiv \frac{\rho\theta}{(1+\theta)^2}.$$

The first two conditional moments of the sojourn time are

$$\mathcal{M}_n = \int_0^{\infty} t p_n(t) dt = \frac{n+\rho}{2} + 1, \quad (2.16)$$

$$\mathcal{S}_n = \int_0^{\infty} t^2 p_n(t) dt = \frac{n^2}{3} + \left(2 + \frac{5}{6}\rho\right)n + \frac{5}{6}\rho^2 + 3\rho + 2. \quad (2.17)$$

Using (2.12), we obtain the following asymptotic expansions for $p_n(t)$, valid for $\rho > 0$ and n and/or $t \rightarrow \infty$.

Theorem 2.3. *For a fixed $\rho > 0$ with $n, t \rightarrow \infty$, the conditional sojourn time density has the following asymptotic expansions:*

1. $n \rightarrow \infty, n/t > 1,$

$$p_n(t) = \frac{1}{n} - \frac{\rho}{n(n-t)} + \frac{\rho-1}{n^2} + O(n^{-3}). \quad (2.18)$$

2. $n/t = 1 + \Delta t^{-1/2} = 1 + O(t^{-1/2}),$

$$p_n(t) \sim \frac{1}{2n} \operatorname{erfc}\left(-\frac{\Delta}{\sqrt{2}}\right) = \frac{1}{n\sqrt{\pi}} \int_{-\Delta/\sqrt{2}}^{\infty} e^{-u^2} du. \quad (2.19)$$

3. $\Lambda_0 < n/t < 1$ with $\Lambda_0 = (-\rho + \sqrt{\rho^2 + 4\rho})/2$,

$$p_n(t) \sim \frac{\Gamma(r_* + 1) e^{r_*}}{\sqrt{2\pi} (1 - n/t)^{r_*+1}} n^{-3/2 - \rho t/n + \rho t^2/n^2} \left(\frac{t}{n}\right)^n e^{n-t}, \quad (2.20)$$

where

$$r_* = r_*\left(\frac{n}{t}\right) \equiv \rho \frac{t^2}{n^2} \left(\frac{n}{t} - 1\right). \quad (2.21)$$

4. $n/t = \Lambda_0 + \Lambda/\sqrt{t}$, $\Lambda = O(1)$,

$$p_n(t) \sim \frac{\sqrt{\rho+4} - \sqrt{\rho}}{4\sqrt{\rho+4}} e^{-1} \Lambda_0^{-n} e^{-t+\Lambda_0 t} \operatorname{erfc}\left\{\frac{\Lambda}{\sqrt{2\Lambda_0}}\right\}. \quad (2.22)$$

5. $n/t < \Lambda_0$,

$$p_n(t) \sim \frac{\sqrt{\rho+4} - \sqrt{\rho}}{2\sqrt{\rho+4}} e^{-1} \Lambda_0^{-n} e^{-t+\Lambda_0 t}. \quad (2.23)$$

Expression (2.23) applies also to $t \rightarrow \infty$ with $n = O(1)$, and gives the exponential decay rate of the density $p_n(t)$. We note that the right side of (2.23) is precisely the $m = 1$ term in the first sum in (2.6), i.e., $C_1(\nu_1) \phi_1(n, \nu_1) e^{\nu_1 t}$. If we start with a fixed large n and increase time t from $t = 0$, we traverse cases 1-5 in Theorem 2.3 in the order given. The leading term in (2.18) is $p_n(t) \sim 1/n$ for $t < n$ which corresponds to a uniform distribution. The $O(n^{-2})$ correction term(s) have a singularity as $t \uparrow n$, which indicates that the asymptotics become invalid. We also note that if $t = 0$, (2.18) becomes $p_n(0) = 1/n - 1/n^2 + O(n^{-3})$, which is just the large n expansion of the initial condition $p_n(0) = 1/(n+1)$. As t/n increases through one, there is a transition region (cf. (2.19)) and then for $t/n > 1$ (but with $t/n < 1/\Lambda_0$) the density becomes exponentially small, with a rather intricate dependence on the space-time ratio, as given in (2.20). After another transition region where $t/n \approx 1/\Lambda_0$ (cf. (2.22)) the density becomes purely exponential in t , which corresponds to the dominant singularity in the Laplace transform $\widehat{p}_n(\theta)$, which occurs at $\theta = \nu_1 < 0$.

We next consider a small traffic intensity, $\rho \rightarrow 0^+$. We shall consider the time scales $t = O(\rho^{-1})$, $t = O(\rho^{-1/2})$ and $t = O(1)$.

Theorem 2.4. *For $\rho \rightarrow 0^+$, the conditional sojourn time density has the following asymptotic expansions:*

1. $t = \zeta/\rho = O(\rho^{-1})$,

(a) $n = x/\rho = O(\rho^{-1})$ with $x > \zeta$,

$$p_n(t) \sim 1/n. \quad (2.24)$$

(b) $n = x/\rho = O(\rho^{-1})$ with $x = \zeta + \Omega \sqrt{\rho}$, $\Omega = O(1)$

$$p_n(t) \sim \frac{1}{2n} \operatorname{erfc}\left(-\frac{\Omega}{\sqrt{2x}}\right). \quad (2.25)$$

(c) $n = x/\rho = O(\rho^{-1})$ with $x < \zeta$,

$$p_n(t) \sim \frac{\rho^{3/2} \zeta}{\sqrt{2\pi}(\zeta - x)} x^{-3/2} \exp\left\{\frac{1}{\rho}\left[x - \zeta + x \log(\zeta/x)\right]\right\}. \quad (2.26)$$

(d) $n = X/\sqrt{\rho} = O(\rho^{-1/2})$ with $X > \zeta$,

$$p_n(t) \sim \frac{1}{\sqrt{2\pi}} \Gamma\left(1 - \frac{\zeta^2}{X^2}\right) e^{-\zeta^2/X^2} n^{-3/2 + \zeta^2/X^2} \left(\frac{t}{n}\right)^n e^{n-t}. \quad (2.27)$$

(e) $n = X/\sqrt{\rho} = O(\rho^{-1/2})$ with $X = \zeta + \rho^{1/4} Y$,

$$p_n(t) \sim \frac{e^{-1}}{4} \rho^{-n/2} e^{-(1-\sqrt{\rho})t} \operatorname{erfc}\left(\frac{Y}{\sqrt{2\zeta}}\right). \quad (2.28)$$

(f) $n = X/\sqrt{\rho} = O(\rho^{-1/2})$ with $X < \zeta$,

$$p_n(t) \sim \frac{e^{-1}}{2} \rho^{-n/2} e^{X/2} e^{-\zeta/2} e^{-(1-\sqrt{\rho})t}. \quad (2.29)$$

(g) $n = O(1)$,

$$p_n(t) \sim \frac{e^{-1}}{2} \rho^{-n/2} e^{-\zeta/2} e^{-(1-\sqrt{\rho})t}. \quad (2.30)$$

2. $t = \omega/\sqrt{\rho} = O(\rho^{-1/2})$ and $n = O(1)$,

$$p_n(t) \sim e^{-t} \rho^{-n/2} Q_n(\omega), \quad (2.31)$$

where

$$\begin{aligned} Q_n(\omega) &= \sum_{m=1}^{\infty} (-1)^n \frac{n! m^{m-n/2-1}}{2m!} e^{-m} L_n^{(m-1-n)}(m) e^{\omega/\sqrt{m}} \\ &+ \sum_{m=1}^{\infty} \frac{n! m^{m-n/2-1}}{2m!} e^{-m} L_n^{(m-1-n)}(m) e^{-\omega/\sqrt{m}}. \end{aligned} \quad (2.32)$$

3. $t, n = O(1)$,

$$p_n(t) = p_n^{(0)}(t) + \rho p_n^{(1)}(t) + O(\rho^2), \quad (2.33)$$

where

$$p_n^{(0)}(t) = \frac{1}{n+1} \frac{1}{2\pi i} \int_{Br} \left[1 - \left(\frac{1}{1+\theta} \right)^{n+1} \right] \frac{e^{\theta t}}{\theta} dt = \frac{e^{-t}}{n+1} \sum_{l=0}^n \frac{t^l}{l!}$$

and

$$p_n^{(1)}(t) = \frac{e^{-t}}{n+1} \left[\frac{t^{n+2}}{(n+2)!} \sum_{l=0}^n \frac{1}{l+2} + \sum_{l=1}^{n+1} \frac{t^l}{l!} \left(\frac{1}{n+2} - \frac{1}{n+2-l} \right) \right].$$

Here Br is a vertical Bromwich contour in the θ -plane with $\Re(\theta) > 0$.

Some of the results in Theorem 2.4 for the time scale $t = O(\rho^{-1})$ can be derived as limiting cases of Theorem 2.3. However, this is not the case for the time ranges $t = O(\rho^{-1/2})$ and $t = O(1)$. For $n, t = O(1)$ the leading term in (2.33) corresponds to the tagged customer entering the system and no further arrivals entering during his/her sojourn time.

The result in (2.31) is obtained by letting $t = \omega/\sqrt{\rho}$ and taking $\rho \rightarrow 0$ in the exact expression (2.6). If we let $\omega \rightarrow \infty$, the $m = 1$ term in the first summation in (2.32) dominates and this verifies the asymptotic matching between the scales $t = O(\rho^{-1})$ and $t = O(\rho^{-1/2})$, for $n = O(1)$.

Finally, we consider a large traffic intensity, $\rho \rightarrow \infty$. The structure of $p_n(t)$ is different in two cases.

Theorem 2.5. *For $\rho \rightarrow \infty$, the conditional sojourn time density has the following asymptotic expansions:*

1. $t = T\rho = O(\rho)$ and $n = N\rho = O(\rho)$,

$$p_n(t) = \rho^{-1} P_0(N, T) + \rho^{-2} P_1(N, T) + O(\rho^{-3}), \quad (2.34)$$

where

$$P_0(N, T) = \frac{e^{U-T}}{N-U} = \frac{N-U-1}{(N-1)(N-U)} \quad (2.35)$$

and $U = U(N, T)$ is defined implicitly by

$$\frac{U}{N-1} = 1 - e^{U-T}. \quad (2.36)$$

If $N = 1$ we obtain the explicit form $P_0(1, T) = e^{-T}$.

2. $t = \tau/\rho = O(\rho^{-1})$ and $n = O(1)$,

$$p_n(t) \sim \int_0^1 (1-\xi)^n J_0\left(2\sqrt{\tau}\sqrt{-\xi - \log(1-\xi)}\right) d\xi, \quad (2.37)$$

where $J_0(\cdot)$ is the Bessel function of the first kind.

We shall compute the correction term $P_1(N, T)$ in (2.34) in section 5 and also give some alternate expressions for the leading term $P_0(N, T)$, as infinite series. The expression in (2.34) remains valid for $n = O(1)$ and $t = O(\rho)$, as well as $t = O(1)$ and $n = O(\rho)$. For $N/T \gg 1$ we have $U \sim T$ and then $P_0(N, T) \sim 1/N$ which is consistent with $p_n(0) = 1/(n+1) \sim \rho^{-1}/N$. For $T/N \gg 1$, $U \rightarrow -1$ and we obtain $P_0(N, T) \sim e^{-1} e^{-T}$, which is consistent with $C_1 \phi_1 e^{\nu_1 t}$ for $\rho \rightarrow \infty$ and $t = O(\rho)$. Note that $\nu_1 \sim -1/\rho$ from (2.7).

We remove the conditioning to get the unconditional sojourn time density for the PS model as

$$p_{PS}(t) = \sum_{n=0}^{\infty} \frac{\rho^n}{n!} e^{-\rho} p_n(t). \quad (2.38)$$

Then the density $p(t)$ for the ROS model follows from (1.9) as

$$p(t) = (1 - e^{-\rho}) p_{PS}(t).$$

Note also that the full density, $p_{ROS}(t)$, for the ROS model is $e^{-\rho} \delta(t) + p(t)$, since there is a non-zero probability that $\mathcal{W} = 0$. The exact representation for $p_{PS}(t)$ is as follows.

Theorem 2.6. *The unconditional sojourn time density has the exact expression*

$$p_{PS}(t) = \sum_{m=1}^{\infty} C_m(\nu_m) \Phi_m(\nu_m) e^{\nu_m t} + \sum_{m=1}^{\infty} C_m(\tilde{\nu}_m) \Phi_m(\tilde{\nu}_m) e^{\tilde{\nu}_m t}, \quad (2.39)$$

where

$$\Phi_m(\nu) = e^{-\rho} (-\nu)^{m-1} \exp\left(\frac{\rho}{\nu+1}\right).$$

We also give the asymptotic results for $p_{PS}(t)$ and $p(t)$ for the different scales of ρ and t .

Theorem 2.7. *The unconditional sojourn time density for the PS model and waiting time density for the ROS model have the following asymptotic expansions:*

1. ρ fixed with $t \rightarrow \infty$

$$p_{PS}(t) = \frac{p(t)}{1 - e^{-\rho}} \sim \frac{\sqrt{\rho+4} - \sqrt{\rho}}{2\sqrt{\rho+4}} e^{-1-\rho} e^{\rho/\Lambda_0} e^{-t} e^{\Lambda_0 t}. \quad (2.40)$$

2. $\rho \rightarrow 0$

(a) $t = \zeta/\rho = O(\rho^{-1})$

$$p_{PS}(t) \sim \rho^{-1} p(t) \sim \frac{1}{2} e^{-1} e^{-\zeta/2} e^{-(1-\sqrt{\rho})t}. \quad (2.41)$$

(b) $t = \omega/\sqrt{\rho} = O(\rho^{-1/2})$

$$p_{PS}(t) \sim \rho^{-1} p(t) \sim e^{-t} Q_0(\omega), \quad (2.42)$$

where

$$Q_0(\omega) = \sum_{m=1}^{\infty} \frac{m^{m-1}}{m!} e^{-m} \cosh(\omega/\sqrt{m}).$$

(c) $t = O(1)$

$$p_{PS}(t) = e^{-t} \left[1 + \frac{\rho}{4} (t^2 - 2) + O(\rho^2) \right]. \quad (2.43)$$

$$p(t) = \rho e^{-t} \left[1 + \frac{\rho}{4} (t^2 - 4) + O(\rho^2) \right]. \quad (2.44)$$

3. $\rho \rightarrow \infty$ with $t = T\rho = O(\rho)$

$$p_{PS}(t) \sim p(t) \sim \frac{1}{\rho} e^{-T}. \quad (2.45)$$

For fixed ρ and large t , we removed the condition by using the expansion in the region $t/n > 1/\Lambda_0$ (i.e., (2.23)) in (2.38), thus obtaining (2.40).

For a small traffic intensity ρ , (2.41) on the $t = O(\rho^{-1})$ scale is the limiting case of (2.40), as $\rho \rightarrow 0$. For the scale $t = O(\rho^{-1/2})$, we used (2.31) in (2.38).

Since ρ is small, the $n = 0$ term dominates, which leads to (2.42). When $t = O(1)$, using (2.33) in (2.38) and the fact that $e^{-\rho} \sim 1 - \rho$ leads to

$$p_{PS}(t) = (1 - \rho) \left[p_0^{(0)}(t) + \rho p_1^{(0)}(t) + \rho p_0^{(1)}(t) + O(\rho^2) \right],$$

which yields (2.43). We note that if we let $\omega \rightarrow 0$ in (2.42), (2.42) reduces to the leading term in (2.43). This indicates that the $t = O(1)$ scale is a special case of the $t = O(\rho^{-1/2})$ scale, for small ρ .

In the case $\rho \rightarrow \infty$ with $t = O(\rho)$, we used the leading term in (2.34) in (2.38) and noticed that the infinite sum concentrates near $n = \rho$ (i.e., $N = 1$), which led to (2.45). In fact this result is uniform on both the $t = O(\rho)$ and $t = O(\rho^{-1})$ time scales, for large ρ .

3 Derivations of the exact representations

We first derive the spectral representation (2.6) of the conditional sojourn time density. Consider the equation (2.4) and assume that $p_n(t)$ has the form $p_n(t) = e^{\nu t} \phi(n)$. Then $\phi(n)$ satisfies the recurrence equation

$$(\nu + 1)(n + 1)\phi(n) = n\phi(n - 1) - \rho\phi(n) + \rho\phi(n + 1). \quad (3.1)$$

We define the exponential generating function $G(z)$ by

$$G(z) = \sum_{n=0}^{\infty} \frac{z^n}{n!} \phi(n). \quad (3.2)$$

Then by (3.1), $G(z)$ satisfies

$$[(\nu + 1)z - \rho] G'(z) + (\nu + 1 + \rho - z) G(z) = 0. \quad (3.3)$$

Here we assumed that $n\phi(n - 1)$ is finite as $n \rightarrow 0$. Solving (3.3), we have

$$G(z) = C \left(1 - \frac{\nu + 1}{\rho} z \right)^{-R_0 - 1} \exp\left(\frac{z}{\nu + 1} \right), \quad R_0 = \frac{\rho\nu}{(1 + \nu)^2}, \quad (3.4)$$

where $C = G(0) = \phi(0)$. Without loss of generality, we let $\phi(0) = C = 1$.

To avoid $\phi(n)$ growing like $n!$ as $n \rightarrow \infty$, $G(z)$ must be an entire function of z , so that $-R_0 - 1$ must be a non-negative integer. The eigenvalues ν

thus satisfy the quadratic equation $R_0 = -m$, $m = 1, 2, \dots$, which leads to the two sets of eigenvalues given by (2.7) and (2.8). We denote by $\phi_m(n, \nu_m)$ and $\phi_m(n, \tilde{\nu}_m)$ the eigenfunctions corresponding to the eigenvalues ν_m and $\tilde{\nu}_m$, respectively. Then for any eigenvalue ν_m , using (3.4) we have

$$\begin{aligned} G_m(z) &= \left(1 - \frac{\nu_m + 1}{\rho} z\right)^{m-1} \exp\left(\frac{z}{\nu_m + 1}\right) \\ &= \sum_{l=0}^{m-1} \binom{m-1}{l} \left(-\frac{\nu_m + 1}{\rho}\right)^{m-1-l} z^{m-1-l} \sum_{k=0}^{\infty} \frac{z^k}{k! (\nu_m + 1)^k}. \end{aligned} \quad (3.5)$$

Thus, from (3.2) we obtain $\phi_m(n, \nu_m)$ as

$$\phi_m(n, \nu_m) = \sum_{l=0}^{\min(n, m-1)} \binom{m-1}{l} \binom{n}{l} l! \frac{(\nu_m + 1)^{2l-n}}{(-\rho)^l}.$$

We note that as $n \rightarrow \infty$, the $l = 0$ term dominates and $\phi_m(n, \nu_m)$ is asymptotically given by

$$\begin{aligned} \phi_m(n, \nu_m) &\sim \frac{n!}{(n-m+1)!} \frac{(\nu_m + 1)^{2(m-1)-n}}{(-\rho)^{m-1}} \\ &\sim \frac{n^{m-1} (\nu_m + 1)^{2m-n-2}}{(-\rho)^{m-1}}, \quad n \rightarrow \infty. \end{aligned} \quad (3.6)$$

Alternately, we can use the generating function of the generalized Laguerre polynomial (see [10])

$$(1+w)^\alpha e^{-wx} = \sum_{n=0}^{\infty} L_n^{(\alpha-n)}(x) w^n. \quad (3.7)$$

Comparing (3.5) to (3.7) we see that in our problem, $w = -(\nu_m + 1)z/\rho$, $x = \rho/(\nu_m + 1)^2$ and $\alpha = m - 1$. Thus, by the definition of G in (3.2), we have another representation for $\phi_m(n, \nu_m)$, as

$$\phi_m(n, \nu_m) = n! \left(-\frac{\nu_m + 1}{\rho}\right)^n L_n^{(m-1-n)}\left(\frac{\rho}{(\nu_m + 1)^2}\right), \quad (3.8)$$

which is (2.10) with $\nu = \nu_m$. By a similar calculation, we find that the eigenfunctions $\phi_m(n, \tilde{\nu}_m)$, which correspond to the eigenvalues $\tilde{\nu}_m$, also satisfy (2.10) with $\nu = \tilde{\nu}_m$.

Thus, we can express the conditional sojourn time density as the spectral representation in (2.6), with only the two coefficient sequences $C_m(\nu_m)$ and $C_m(\tilde{\nu}_m)$ to be determined.

To determine these coefficients, we first obtain an orthogonality relation for the eigenfunctions. Since all of the eigenfunctions satisfy (3.1), we consider any two eigenfunctions $\phi_m(n, \nu_m)$ and $\phi_{m'}(n, \nu_{m'})$ ($m \neq m'$), which satisfy

$$[(\nu_m + 1)(n + 1) - \rho] \phi_m(n, \nu_m) = n \phi_m(n - 1, \nu_m) + \rho \phi_m(n + 1, \nu_m) \quad (3.9)$$

and

$$[(\nu_{m'} + 1)(n + 1) - \rho] \phi_{m'}(n, \nu_{m'}) = n \phi_{m'}(n - 1, \nu_{m'}) + \rho \phi_{m'}(n + 1, \nu_{m'}). \quad (3.10)$$

We multiply (3.9) by $\rho^n \phi_{m'}(n, \nu_{m'})/n!$ and (3.10) by $\rho^n \phi_m(n, \nu_m)/n!$, subtract one equation from the other, and sum over $n \geq 0$. This leads to

$$(\nu_m - \nu_{m'}) \sum_{n=0}^{\infty} \frac{n+1}{n!} \rho^n \phi_m(n, \nu_m) \phi_{m'}(n, \nu_{m'}) = 0.$$

Since $\nu_m \neq \nu_{m'}$, we obtain the orthogonality relation

$$\sum_{n=0}^{\infty} \frac{n+1}{n!} \rho^n \phi_m(n, \nu_m) \phi_{m'}(n, \nu_{m'}) = 0. \quad (3.11)$$

By the spectral representation (2.6) and the initial condition (2.5), we must have

$$\sum_{m=1}^{\infty} C_m(\nu_m) \phi_m(n, \nu_m) + \sum_{m=1}^{\infty} C_m(\tilde{\nu}_m) \phi_m(n, \tilde{\nu}_m) = \frac{1}{n+1}.$$

Using (3.11), we can easily show that, for any eigenvalue ν_m or $\tilde{\nu}_m$,

$$C_m(\nu_m) = \frac{\sum_{n=0}^{\infty} \rho^n \phi_m(n, \nu_m)/n!}{\sum_{n=0}^{\infty} (n+1) \rho^n \phi_m^2(n, \nu_m)/n!}. \quad (3.12)$$

Using the generating function (3.5), the numerator in (3.12) is

$$\sum_{n=0}^{\infty} \frac{\rho^n}{n!} \phi_m(n, \nu_m) = G_m(\rho) = (-\nu_m)^{m-1} \exp\left(\frac{\rho}{\nu_m + 1}\right). \quad (3.13)$$

To determine the denominator in (3.12), we let $G(z, \mu)$ be a solution of (3.1) with $\nu = \mu$, whose generating function is given by

$$G(z, \mu) = \left(1 - \frac{\mu + 1}{\rho} z\right)^{-R-1} \exp\left(\frac{z}{\mu + 1}\right), \quad R = \frac{\rho \mu}{(1 + \mu)^2}.$$

Here μ is not necessarily an eigenvalue, and we assume that $R < -1$ and that $R + 1$ is not an integer. Thus, $\phi(n, \mu)$, which corresponds to the non-eigenvalue μ , can be represented as the Cauchy integral

$$\phi(n, \mu) = \frac{n!}{2\pi i} \oint_{\mathcal{C}} \left(1 - \frac{\mu + 1}{\rho} z\right)^{-R-1} \exp\left(\frac{z}{\mu + 1}\right) z^{-n-1} dz, \quad (3.14)$$

where the contour \mathcal{C} is a small circle in the complex z -plane centered at the origin. Since $-R - 1 > 0$ and R is not an integer, $z = \rho/(\mu + 1)$ is a branch point of the integrand in (3.14). We also have the binomial expansion

$$\left(1 - \frac{\mu + 1}{\rho} z\right)^{-R-1} = \sum_{k=0}^{\infty} \frac{\Gamma(R + 1 + k)}{\Gamma(R + 1) k!} \left(\frac{\mu + 1}{\rho}\right)^k z^k.$$

Then from (3.14), we find that as $n \rightarrow \infty$, $\phi(n, \mu)$ is asymptotically given by

$$\begin{aligned} \phi(n, \mu) &\sim \frac{\Gamma(R + 1 + n)}{\Gamma(R + 1)} \left(\frac{\mu + 1}{\rho}\right)^n \exp\left(\frac{\rho}{(\mu + 1)^2}\right) \\ &\sim \frac{n! n^R}{\Gamma(R + 1)} \left(\frac{\mu + 1}{\rho}\right)^n \exp\left(\frac{\rho}{(\mu + 1)^2}\right), \quad n \rightarrow \infty. \end{aligned} \quad (3.15)$$

Since $\phi(n, \mu)$ satisfies (3.1), we have

$$[(\mu + 1)(n + 1) - \rho] \phi(n, \mu) = n \phi(n - 1, \mu) + \rho \phi(n + 1, \mu). \quad (3.16)$$

We multiply (3.9) by $\rho^n \phi(n, \mu)/n!$ and (3.16) by $\rho^n \phi_m(n, \nu_m)/n!$, subtract one equation from the other, and sum over $0 \leq n \leq K$, which yields

$$\begin{aligned} &(\nu_m - \mu) \sum_{n=0}^K \frac{n + 1}{n!} \rho^n \phi_m(n, \nu_m) \phi(n, \mu) \\ &= \frac{\rho^{K+1}}{K!} [\phi_m(K + 1, \nu_m) \phi(K, \mu) - \phi_m(K, \nu_m) \phi(K + 1, \mu)]. \end{aligned} \quad (3.17)$$

We let $K \rightarrow \infty$ and use (3.6) and (3.15), which shows that

$$\phi_m(K+1, \nu_m) \phi(K, \mu) = O(K^{m+R-1}),$$

and

$$\phi_m(K, \nu_m) \phi(K+1, \mu) = O(K^{m+R}).$$

Thus, the second term inside the bracket in the right-hand side of (3.17) dominates the first and we obtain, after dividing both sides by $\nu_m - \mu$ and expanding for $K \rightarrow \infty$,

$$\begin{aligned} & \sum_{n=0}^K \frac{n+1}{n!} \rho^n \phi_m(n, \nu_m) \phi(n, \mu) \\ & \sim \frac{(-1)^m K^{m+R} (\mu+1)^{K+1} (\nu_m+1)^{2m-K-2}}{(\nu_m - \mu) \Gamma(R+1) \rho^{m-1}} \exp\left(\frac{\rho}{(\mu+1)^2}\right). \end{aligned} \quad (3.18)$$

Next we let $\mu \rightarrow \nu_m$ in (3.18), so that $R = R(\mu) \rightarrow R(\nu_m) = -m$. By using the Laurent expansion of $\Gamma(\cdot)$ near a pole

$$\Gamma(R+1) \sim \frac{(-1)^{m-1}}{(m-1)!(R+m)} \quad \text{as } R \rightarrow -m,$$

and then l'Hôpital's rule, we find that

$$\lim_{\mu \rightarrow \nu_m} [(\nu_m - \mu) \Gamma(R+1)] = \frac{(-1)^m}{(m-1)! R'(\nu_m)} = \frac{(-1)^m (1 + \nu_m)^3}{(m-1)! \rho (1 - \nu_m)}. \quad (3.19)$$

Thus, by using (3.18) and (3.19) and noting that $\rho \nu_m / (\nu_m + 1)^2 = -m$, we let $K \rightarrow \infty$ and obtain

$$\begin{aligned} & \sum_{n=0}^{\infty} \frac{(n+1)}{n!} \rho^n \phi_m^2(n, \nu_m) \\ & = \frac{m!}{m^{m-1}} (-\nu_m)^{m-2} (1 - \nu_m) \exp\left(\frac{\rho}{(\nu_m + 1)^2}\right), \end{aligned} \quad (3.20)$$

which determines the denominator in (3.12). Using (3.13) and (3.20) in (3.12), we obtain (2.9) with $\nu = \nu_m$. By the same argument, we find that (2.9) is also true for the eigenvalues $\tilde{\nu}_m$. This completes the derivation of Theorem 2.1.

In the remainder of this section, we use a discrete Green's function to derive (2.12). Consider the recurrence equation (2.11). The discrete Green's function $\mathcal{G}(\theta; n, l)$ satisfies

$$\begin{aligned} \rho \mathcal{G}(\theta; n+1, l) - [(n+1)(1+\theta) + \rho] \mathcal{G}(\theta; n, l) \\ + n \mathcal{G}(\theta; n-1, l) = -\delta(n, l), \quad (n, l \geq 0) \end{aligned} \quad (3.21)$$

where $\delta(n, l) = 1\{n = l\}$ is the Kronecker delta. To construct the Green's function we need two linearly independent solutions to

$$\rho G(\theta; n+1, l) - [(n+1)(1+\theta) + \rho] G(\theta; n, l) + n G(\theta; n-1, l) = 0, \quad (3.22)$$

which is the homogeneous version of (3.21).

We seek solutions of (3.22) of the form

$$G_n = \int_{\mathcal{D}} z^n g(z) dz,$$

where the function $g(z)$ and the path \mathcal{D} of integration in the complex z -plane are to be determined. Using the above form in (3.22) and integrating by parts yields

$$\begin{aligned} z^n g(z) [(1+\theta)z - 1] \Big|_{\mathcal{D}} \\ - \int_{\mathcal{D}} z^n \left\{ [(1+\theta)z - 1] g'(z) + \rho(z-1)g(z) \right\} dz = 0. \end{aligned} \quad (3.23)$$

The first term represents contributions from the endpoints of the contour \mathcal{D} .

If (3.23) is to hold for all n the integrand must vanish, so that $g(z)$ must satisfy the differential equation

$$[(1+\theta)z - 1] g'(z) + \rho(z-1)g(z) = 0, \quad (3.24)$$

and thus

$$g(z) = \left(z - \frac{1}{1+\theta} \right)^r \exp \left(-\frac{\rho}{1+\theta} z \right), \quad r = \frac{\rho\theta}{(1+\theta)^2}.$$

If the path of integration \mathcal{D} is chosen as the segment $[0, 1/(1+\theta)]$ of the real axis, then (3.23) is satisfied for $n \geq 1$. Thus, we obtain G_n as in (2.14).

We note that G_n decays as $n \rightarrow \infty$, and by scaling $z = (1 - y/n)/(1 + \theta)$ and using the Laplace method, we find that G_n is asymptotically given by

$$G_n \sim \frac{\Gamma(r+1)}{n^{r+1} (1+\theta)^{n+r+1}} e^{-r/\theta}, \quad n \rightarrow \infty. \quad (3.25)$$

However, G_n becomes infinite as $n \rightarrow -1$, and $n G_{n-1}$ goes to a nonzero limit as $n \rightarrow 0$. Thus G_n is not an acceptable solution to (3.22) at $n = 0$.

To construct a second solution to (3.22), we consider another path of the integration, the real interval $[1/(1+\theta), \infty)$. Then (3.23) is again satisfied. Thus, we have another solution of (3.22), H_n , which is given by (2.15). H_n is finite as $n \rightarrow -1$, but grows as $n \rightarrow \infty$. By scaling $z = n w = O(n)$ in the integrand of (2.15) we find that H_n grows roughly like $n!$ for n large; more precisely

$$H_n \sim n! n^r \left(\frac{1+\theta}{\rho} \right)^{n+r+1}, \quad n \rightarrow \infty. \quad (3.26)$$

Thus, the discrete Green's function can be represented by

$$\mathcal{G}(\theta; n, l) = \begin{cases} H_l G_n \mathcal{G}_0 & \text{if } n \geq l \\ G_l H_n \mathcal{G}_0 & \text{if } 0 \leq n < l, \end{cases} \quad (3.27)$$

which has acceptable behavior both at $n = 0$ and as $n \rightarrow \infty$. Here \mathcal{G}_0 depends only upon θ and l .

To determine \mathcal{G}_0 , we let $n = l$ in (3.22) and use the fact that both G_l and H_l satisfy (3.22) with $n = l$. Then we can infer a simple difference equation for the discrete Wronskian $G_l H_{l+1} - G_{l+1} H_l$, whose solution we write as

$$G_l H_{l+1} - G_{l+1} H_l = \frac{l!}{\rho^l \mathcal{G}_1}, \quad (3.28)$$

where $\mathcal{G}_1 = \mathcal{G}_1(\theta)$ depends upon θ only. Then using (3.27) in (3.22) with $n = l$ shows that \mathcal{G}_0 and \mathcal{G}_1 are related by $\mathcal{G}_0 = \rho^{l-1} \mathcal{G}_1 / l!$.

Letting $l \rightarrow \infty$ in (3.28) and using the asymptotic results in (3.25) and (3.26), we determine \mathcal{G}_1 and then obtain \mathcal{G}_0 as

$$\mathcal{G}_0 = \frac{\rho^{r+l+1}}{l! \Gamma(r+1) (1+\theta)} e^{r/\theta}.$$

Then, we multiply (3.21) by the solution $\widehat{p}_l(\theta)$ to (2.11) and sum over all $l \geq 0$. After some manipulation this yields

$$\widehat{p}_n(\theta) = \sum_{l=0}^{\infty} \mathcal{G}(\theta; n, l),$$

which is equivalent to (2.12). Taking the inverse Laplace transform gives the conditional sojourn time density $p_n(t)$ as the contour integral

$$p_n(t) = \frac{1}{2\pi i} \int_{Br} \widehat{p}_n(\theta) e^{\theta t} d\theta, \quad (3.29)$$

where Br is a vertical contour in the complex θ -plane, with $\Re(\theta) \geq 0$. The form in (3.29) is more useful than the spectral representation for obtaining asymptotic results in various limits, such as n, t simultaneously large.

To compute the first two moments \mathcal{M}_n and \mathcal{S}_n , we could expand $\widehat{p}_n(\theta)$ about $\theta = 0$, but it is easier to derive simple difference equations for the moments directly from (2.4). By multiplying (2.4) by t and integrating from $t = 0$ to $t = \infty$ we obtain

$$\rho \mathcal{M}_{n+1} - (n+1+\rho) \mathcal{M}_n + n \mathcal{M}_{n-1} = -(n+1).$$

This equation has the linear solution in (2.16). Similarly, for the second moment we obtain

$$\rho \mathcal{S}_{n+1} - (n+1+\rho) \mathcal{S}_n + n \mathcal{S}_{n-1} = -2(n+1) \mathcal{M}_n, \quad (3.30)$$

whose solution is given by (2.17).

4 Asymptotic results for fixed ρ and $\rho \rightarrow 0$

We first assume that the traffic intensity ρ is fixed. We sketch the main points in deriving Theorem 2.3. We first consider $n, t \rightarrow \infty$ with $n > t$ and use the result in (2.12). To obtain a two term approximation, we need the correction terms in the approximations in (3.25) and (3.26), which are given by

$$G_n = \frac{e^{-r/\theta}}{n^{r+1} (1+\theta)^{n+r+1}} \left[\Gamma(r+1) + \frac{1}{n} \left(\frac{r}{\theta} \Gamma(r+2) - \frac{1}{2} \Gamma(r+3) \right) + O(n^{-2}) \right] \quad (4.1)$$

and

$$H_n = n! n^r \left(\frac{1+\theta}{\rho} \right)^{n+r+1} \left[1 - \frac{r^2}{n\theta} + O(n^{-2}) \right]. \quad (4.2)$$

From (4.1) and (4.2), we note that the first term in (2.12) dominates the second, and thus the Laplace transform is asymptotically given by

$$\widehat{p}_n(t) \sim M G_n \sum_{l=0}^n \frac{\rho^l}{l!} H_l \quad (4.3)$$

$$\sim \frac{1+\theta}{\rho\theta} \sum_{l=0}^n (1+\theta)^{l-n} \left(\frac{l}{n} \right)^r \frac{1}{n} \left[r + \frac{A}{n} - \frac{r^2}{l\theta} \right] \quad (4.4)$$

$$\begin{aligned} &\sim \left[\frac{1}{1+\theta} + \frac{(1+\theta)A}{\rho\theta n} \right] \int_0^1 (1+\theta)^{-ny} (1-y)^r dy \\ &\quad - \frac{\rho}{(1+\theta)^3 n} \int_0^1 (1+\theta)^{-ny} (1-y)^{r-1} dy \\ &\equiv I_1 + I_2. \end{aligned} \quad (4.5)$$

Here

$$A = A(\theta) \equiv \frac{r(r+1)}{2\theta} [2r - \theta(r+2)]$$

and we used the Euler-Maclaurin summation formula to approximate the sums in (4.4) by integrals. By scaling $\theta = s/n = O(1/n)$ and noting that

$$\frac{(1+\theta)A}{\rho\theta} = \rho - 1 + \frac{s}{2n} (2\rho^2 - 9\rho + 2) + O(n^{-2})$$

and $r \sim \rho s/n$, the first term, I_1 , in (4.5) becomes

$$\begin{aligned} I_1 &\sim \left(1 + \frac{\rho-1}{n} \right) \int_0^1 e^{-sy} dy \\ &= \left(1 + \frac{\rho-1}{n} \right) \frac{1 - e^{-s}}{s}. \end{aligned}$$

Thus, taking the inverse Laplace transforms of I_1 and I_2 yields

$$\begin{aligned} \mathcal{L}^{-1}(I_1) &\sim \left(1 + \frac{\rho-1}{n} \right) \frac{1}{n} \text{Heaviside} \left(1 - \frac{t}{n} \right) \\ &= \frac{1}{n} + \frac{\rho-1}{n^2} \quad (n > t) \end{aligned} \quad (4.6)$$

and

$$\begin{aligned}
\mathcal{L}^{-1}(I_2) &\sim -\frac{\rho}{n^2} \frac{1}{2\pi i} \int_{Br} e^{st/n} \left[\int_0^1 \frac{e^{-sy}}{(1-y)^{1-\rho s/n}} dy \right] ds \\
&\sim -\frac{\rho}{n^2} \int_0^1 \frac{\delta(t/n - y)}{1-y} dy \\
&= -\frac{\rho}{n(n-t)} \quad (n > t). \tag{4.7}
\end{aligned}$$

We note that in deriving (4.7), we changed the order of integration and used $d\theta = n^{-1}ds$. Then (4.6) and (4.7) lead to (2.18).

This analysis suggests that $p_n(t)$ is approximately zero in the range $t/n > 1$. We shall show that in this sector the density is exponentially small. Before doing this, we first investigate the transition region, where $t \approx n$.

Thus, we consider $n, t \rightarrow \infty$ with $n/t = 1 + \Delta t^{-1/2} = 1 + O(t^{-1/2})$. We can still use (4.3) but now scale $l = y\sqrt{n} = O(\sqrt{n})$, and approximate the sum by

$$\hat{p}_n(\theta) \sim \frac{1}{(1+\theta)^{n+1} (\sqrt{n})^{r+1}} \int_0^\infty (1+\theta)^{\sqrt{n}y} y^r dy. \tag{4.8}$$

Scaling $\theta = \varpi/\sqrt{n} = O(1/\sqrt{n})$, and noting that

$$(1+\theta)^{-n-1} \sim e^{\varpi^2/2 - \sqrt{n}\varpi}$$

and $(1+\theta)^{\sqrt{n}y} \sim e^{\varpi y}$, the inverse Laplace transform leads to

$$\begin{aligned}
p_n(t) &\sim \frac{1}{n} \frac{1}{2\pi i} \int_{Br} e^{\varpi t/\sqrt{n}} e^{\varpi^2/2 - \sqrt{n}\varpi} \int_0^\infty e^{\varpi y} dy d\varpi \\
&= \frac{1}{n} \int_0^\infty \frac{1}{2\pi i} \int_{Br} e^{\varpi^2/2} e^{(-\sqrt{n}+y+t/\sqrt{n})\varpi} d\varpi dy. \tag{4.9}
\end{aligned}$$

Note that in this range of (n, t) ,

$$\frac{t}{\sqrt{n}} - \sqrt{n} = \left(1 - \frac{n}{t}\right) \frac{t}{\sqrt{n}} = -\Delta \frac{\sqrt{t}}{\sqrt{n}} \sim -\Delta = O(1).$$

Then by using the identity

$$\frac{1}{2\pi i} \int_{Br} e^{C_0\varpi^2 + C_1\varpi} d\varpi = \frac{1}{2\sqrt{\pi C_0}} \exp\left(-\frac{C_1^2}{4C_0}\right)$$

and noting that $\Delta = (n - t)/\sqrt{t}$, we explicitly evaluate the integral over ϖ in (4.9) and obtain (2.19).

Now we consider $n, t \rightarrow \infty$ with $t > n$. We rewrite (2.12) as

$$\widehat{p}_n(\theta) = M G_n \sum_{l=0}^{\infty} \frac{\rho^l}{l!} H_l + M \sum_{l=n+1}^{\infty} \frac{\rho^l}{l!} (H_n G_l - H_l G_n). \quad (4.10)$$

The first sum can be calculated exactly by using (2.15) and making the transformation $z = 1/(1 + \theta) + y$, which yields

$$\sum_{l=0}^{\infty} \frac{\rho^l}{l!} H_l = e^r \Gamma(r + 1) \left(-\frac{1 + \theta}{\rho \theta} \right)^{r+1}. \quad (4.11)$$

The result in (4.11) holds for $\theta < 0$ and $\theta > \theta_p = -1 + [-\rho + \sqrt{\rho^2 + 4\rho}]/2$, since $\Gamma(r + 1)$ has a simple pole at $\theta = \theta_p$. We note that $\theta_p = \nu_1$, which is the first eigenvalue in (2.7). The second sum in (4.10) is negligible in view of (3.25), (3.26) and (4.11), and the fact $\theta < 0$ on this scale. Using (2.13), (3.25) and (4.11) in the first sum of (4.10), then taking the inverse Laplace transform, we have

$$p_n(t) \sim \frac{1}{2\pi i} \int_{Br} h(\theta) e^{t f(\theta)} d\theta, \quad (4.12)$$

where $f(\theta) = \theta - \log(1 + \theta) n/t$ and

$$h(\theta) = \frac{\Gamma(r + 1) e^r}{(1 + \theta) (-\theta)^{r+1} n^{r+1}}.$$

For $t \rightarrow \infty$ and n/t fixed we evaluate (4.12) by the saddle point method. There is a saddle point at $\theta = \theta_s \equiv n/t - 1 < 0$, which satisfies $f'(\theta) = 0$. Hence, using the saddle point method gives

$$p_n(t) \sim \frac{h(\theta_s)}{\sqrt{2\pi t f''(\theta_s)}} e^{t f(\theta_s)}$$

and this leads to (2.20), where

$$r_* = r(\theta_s) = \frac{\rho \theta_s}{(1 + \theta_s)^2} = \rho \frac{t^2}{n^2} \left(\frac{n}{t} - 1 \right).$$

This analysis indicates that (2.20) only holds for n , $t \rightarrow \infty$ with $n/t < 1$ and $\theta_s > \theta_p$, so that $n/t = 1 + \theta_s > 1 + \theta_p = \Lambda_0$.

There is a transition region where $n/t = \Lambda_0 + \Lambda/\sqrt{t}$ with $\Lambda = O(1)$. We still use (4.12) and note that $h(\theta)$ has a simple pole at $\theta = \theta_p$, and the saddle point θ_s of $f(\theta)$ is now close to θ_p . We expand the integrand in (4.12) about $\theta = \theta_p$ using

$$h(\theta) \sim \frac{\sqrt{\rho+4} - \sqrt{\rho}}{2\sqrt{\rho+4}} e^{-1} \frac{1}{\theta - \theta_p}$$

and

$$\begin{aligned} f(\theta) &= f(\theta_p) + f'(\theta_p)(\theta - \theta_p) + \frac{1}{2}f''(\theta_p)(\theta - \theta_p)^2 + \dots \\ &\sim (\Lambda_0 - 1) - \frac{n}{t} \log(\Lambda_0) - \frac{\Lambda}{\Lambda_0} \frac{\theta - \theta_p}{\sqrt{t}} + \frac{1}{2\Lambda_0}(\theta - \theta_p)^2 + \dots \end{aligned}$$

Here we also used $1 + \theta_p = \Lambda_0$ and $n/t \sim \Lambda_0$. By scaling $\theta - \theta_p = S/\sqrt{t}$, (4.12) asymptotically becomes

$$p_n(t) \sim \Lambda_0^{-n} e^{-t+\Lambda_0 t} \frac{1}{2\pi i} \int_{Br} \frac{1}{S} \exp \left[-\frac{\Lambda}{\Lambda_0} S + \frac{1}{2\Lambda_0} S^2 \right] dS,$$

where $\Re(S) > 0$ on the contour Br . Then we use the identity

$$\frac{1}{2\pi i} \int_{Br} \frac{1}{S} e^{-AS+B S^2/2} dS = \frac{1}{\sqrt{2\pi}} \int_{A/\sqrt{B}}^{\infty} e^{-u^2/2} du,$$

with $A = \Lambda/\Lambda_0$ and $B = \Lambda_0^{-1}$, to eventually obtain (2.22).

Finally, for the scale n , $t \rightarrow \infty$ with $n/t < \Lambda_0$, the pole at $\theta = \theta_p$ dominates the asymptotic behavior of $p_n(t)$, and (2.23) is obtained by evaluating the residue at the dominant pole in (4.12). This concludes the derivation of Theorem 2.3.

If we consider a small traffic intensity, $\rho \rightarrow 0$, and scale $t = O(\rho^{-1})$, we obtain seven different asymptotic expressions based on different scalings of the space variable n , which are given in case 1 of Theorem 2.4. All seven results can be obtained as limiting cases of the results in Theorem 2.3, by letting $\rho \rightarrow 0$ and scaling n appropriately. We omit the derivations here. Note that as $\rho \rightarrow 0$ the transition line $n/t = \Lambda_0$ becomes close to the t -axis.

For the time scale $t = O(\rho^{-1/2})$, from the spectral representation we note that as $\rho \rightarrow 0$ the eigenvalues are

$$\nu_m = -1 + \frac{\sqrt{\rho}}{\sqrt{m}} + O(\rho), \quad \tilde{\nu}_m = -1 - \frac{\sqrt{\rho}}{\sqrt{m}} + O(\rho).$$

Then the eigenfunctions are asymptotically given by

$$\phi_m(n, \nu_m) \sim (-1)^n n! m^{-n/2} \rho^{-n/2} L_n^{(m-1-n)}(m)$$

and $\phi_m(n, \tilde{\nu}_m) \sim (-1)^n \phi_m(n, \nu_m)$. Thus, all the eigenvalues contribute to $p_n(t)$ on the scale $t = O(\rho^{-1/2})$ and $n = O(1)$, and we obtain (2.31).

Now we consider the scale n , $t = O(1)$ with $\rho \rightarrow 0$ and use the result in (2.12). Since $r = O(\rho)$, G_n in (2.14) becomes

$$G_n \sim \int_0^{\frac{1}{1+\theta}} z^n dz = \frac{1}{(n+1)(1+\theta)^{n+1}}. \quad (4.13)$$

By scaling $z = \frac{s}{(1+\theta)\rho}$ in (2.15), H_n is asymptotically given by

$$\begin{aligned} H_n &\sim \frac{1}{\rho^{n+1}(1+\theta)^{n+1}} \int_0^\infty s^n \exp\left(-\frac{s}{(1+\theta)^2}\right) ds \\ &= n! \frac{(1+\theta)^{n+1}}{\rho^{n+1}}. \end{aligned} \quad (4.14)$$

We also have $M \sim \rho/(1+\theta)$. Using (4.13) and (4.14) in (2.12), we find that the first sum dominates the second and we obtain

$$\begin{aligned} \hat{p}_n(\theta) &\sim M G_n \sum_{l=0}^n \frac{\rho^l}{l!} H_l \\ &\sim \frac{1}{(n+1)(1+\theta)^{n+2}} \sum_{l=0}^n (1+\theta)^{l+1} \\ &= \frac{1}{n+1} \frac{1}{\theta} \left[1 - \frac{1}{(1+\theta)^{n+1}} \right]. \end{aligned}$$

Then we invert the Laplace transform over time, letting $\theta = w - 1$, which gives

$$\begin{aligned} p_n(t) &\sim \frac{e^{-t}}{n+1} \frac{1}{2\pi i} \int_{Br} \frac{1 - w^{-n-1}}{w-1} e^{wt} dw \\ &= \frac{e^{-t}}{n+1} \sum_{l=0}^n \frac{t^l}{l!}. \end{aligned} \quad (4.15)$$

This is the leading term $p_n^{(0)}(t)$ in (2.33).

To obtain the second term $p_n^{(1)}(t)$, we need the correction terms of the asymptotic expansions in (4.13) and (4.14). It is much easier, however, to use a perturbation method to obtain $p_n^{(1)}(t)$. Assume that the conditional sojourn time density has an expansion in powers of ρ , as in (2.33). By using the recurrence equation (2.4), the leading term $p_n^{(0)}(t)$ satisfies

$$\frac{dp_n^{(0)}(t)}{dt} = \frac{n}{n+1} p_{n-1}^{(0)}(t) - p_n^{(0)}(t)$$

with the initial condition $p_n^{(0)}(t) = 1/(n+1)$. This can be easily solved to regain (4.15). The second term $p_n^{(1)}(t)$ satisfies

$$\frac{dp_n^{(1)}(t)}{dt} = \frac{n}{n+1} p_{n-1}^{(1)}(t) - p_n^{(1)}(t) + \frac{1}{n+1} p_{n+1}^{(0)}(t) - \frac{1}{n+1} p_n^{(0)}(t)$$

with the initial condition $p_n^{(1)}(0) = 0$. We set

$$p_n^{(1)}(t) = \frac{e^{-t}}{n+1} \mathcal{P}_n(t) \tag{4.16}$$

and take the Laplace transform of $\mathcal{P}_n(t)$ over time, with $\widehat{\mathcal{P}}_n(s) = \int_0^\infty \mathcal{P}_n(t) e^{-st} dt$. After some simplification, we find that $\widehat{\mathcal{P}}_n(s)$ satisfies the following difference equation:

$$s \widehat{\mathcal{P}}_n(s) - \widehat{\mathcal{P}}_{n-1}(s) = -\frac{(n+1)s^{-n-2} - (n+2)s^{-n-1} + 1}{(n+1)(n+2)(s-1)}, \quad n \geq 1$$

with

$$\widehat{\mathcal{P}}_0(s) = \frac{1-s}{2s^3}.$$

After some calculation, we obtain $\widehat{\mathcal{P}}_n(s)$ as

$$\begin{aligned} \widehat{\mathcal{P}}_n(s) &= -\frac{s^{-n-1}}{s-1} + \frac{1}{(n+2)s(s-1)} + \frac{(n+1)s^{-n}}{(n+2)s^2(s-1)} \\ &\quad + \frac{1}{s^{n+3}} \sum_{l=0}^n \frac{1}{l+2} - \sum_{l=0}^{n-1} \frac{s^{l-n-1}}{l+2}. \end{aligned}$$

Then by taking the inverse Laplace transform, and using the relation (4.16), we obtain $p_n^{(1)}(t)$ in (2.33).

5 Asymptotic results for $\rho \rightarrow \infty$

We shall use a singular perturbation approach to derive the asymptotic approximations for large traffic intensities, $\rho \rightarrow \infty$. This method should be useful for analyzing models with more general balking probabilities. We shall also sketch how the asymptotic results in Theorem 2.5 follow from the exact representations in Theorems 2.1 and 2.2.

We first consider the scale $t = T\rho = O(\rho)$ and $n = N\rho = O(\rho)$, and expand $p_n(t)$ in powers of ρ^{-1} , as in (2.34). Using (2.34) in the recurrence equation (2.4), the leading term $P_0(N, T)$ satisfies

$$\frac{\partial P_0}{\partial T} + \frac{N-1}{N} \frac{\partial P_0}{\partial N} = -\frac{1}{N} P_0 \quad (5.1)$$

with the initial condition $P_0(N, T) = 1/N$. We solve this first order PDE by the method of characteristics. The family of characteristics is given by

$$T = N + \log |1 - N| + \text{constant},$$

where the constant indexes the family. The characteristic $T = N + \log(1 - N)$ goes through the origin $(N, T) = (0, 0)$, along the parabola $T = -N^2/2$. The general solution to (5.1) is

$$P_0(N, T) = \frac{1}{N-1} \mathcal{F}((N-1) e^{N-T}). \quad (5.2)$$

Using the initial condition in (5.2), we determine the function $\mathcal{F}(\cdot)$ from

$$\mathcal{F}((N-1) e^N) = \frac{N-1}{N}.$$

If we denote by $N_* = N_*(N, T)$ the solution to

$$(N_* - 1) e^{N_*} = (N - 1) e^{N-T}, \quad (5.3)$$

P_0 in (5.2) becomes

$$P_0(N, T) = \frac{1}{N-1} \frac{N_* - 1}{N_*}. \quad (5.4)$$

Setting $N_* = N - U$ in (5.3) leads to (2.36). Thus, (5.4) can be rewritten as (2.35).

Alternately, we can rewrite (2.35) more explicitly, in terms of an infinite series. From (2.36), we let $U = N - 1 + U_0$, where $U_0 = U_0(N, T) = (1 - N) e^{U-T}$. Then U_0 can be expressed in terms of the Lambert W-function (see [4]), which satisfies

$$e^{-U_0} U_0 = (1 - N) e^{N-T-1} \equiv z.$$

We use the series expansion of the Lambert W-function to obtain U_0 as

$$U_0 = \sum_{m=1}^{\infty} \frac{(-m)^{m-1}}{m!} z^m,$$

where the series converges for $|z| < e^{-1}$. Thus, U has the following series expansion

$$U(N, T) = N - 1 + \sum_{m=1}^{\infty} \frac{m^{m-1}}{m!} (1 - N)^m e^{m(N-T-1)}, \quad (5.5)$$

which converges for $|1 - N| e^{N-T} < 1$. The series is always convergent for $N \leq 1$, but diverges for $N > 1$, if $T < N + \log(N - 1)$. For example, if $T = 0$ the series converges only for $N < N_c \doteq 1.2784$, where $(N_c - 1) e^{N_c} = 1$. Using (5.5) in (2.35), we have an alternate series expression for $P_0(N, T)$:

$$P_0(N, T) = \frac{\sum_{m=1}^{\infty} m^{m-1} (1 - N)^{m-1} e^{m(N-T-1)} / m!}{1 - \sum_{m=1}^{\infty} m^{m-1} (1 - N)^m e^{m(N-T-1)} / m!}. \quad (5.6)$$

Now we sketch how to compute the correction term $P_1(N, T)$, which satisfies the following PDE

$$\frac{\partial P_1}{\partial T} = \frac{1 - N}{N} \frac{\partial P_1}{\partial N} - \frac{1}{N} P_1 + \frac{N + 1}{2N} \frac{\partial^2 P_0}{\partial N^2} + \frac{N - 1}{N^2} \frac{\partial P_0}{\partial N} + \frac{1}{N^2} P_0, \quad (5.7)$$

with the initial condition $P_1(N, 0) = -1/N^2$. This follows from expanding $p_n(0) = 1/(n + 1) = 1/(N\rho + 1)$ in powers of ρ^{-1} . We make the substitution

$$P_1(N, T) = -\frac{1}{N} P_0(N, T) + P_1^*(N, T), \quad (5.8)$$

so that $P_1^*(N, 0) = 0$.

We introduce an operator \mathcal{D} , which is defined by

$$\mathcal{D} = \frac{\partial}{\partial T} + \frac{N-1}{N} \frac{\partial}{\partial N} + \frac{1}{N}. \quad (5.9)$$

Then by (5.1) we have $\mathcal{D}P_0 = 0$ and

$$\mathcal{D}\left(-\frac{P_0}{N}\right) = \frac{N-1}{N^3} P_0. \quad (5.10)$$

Applying the operator \mathcal{D} to (5.8), and using (5.7) and (5.10), we obtain

$$\mathcal{D}P_1^* = \frac{N+1}{2N} \frac{\partial^2 P_0}{\partial N^2} + \frac{N-1}{N^2} \frac{\partial P_0}{\partial N} + \frac{1}{N^3} P_0. \quad (5.11)$$

We change variables in (5.11) from (N, T) to (ξ, η) , where $N = \xi + \eta$ and $U = \eta$, with $U = U(N, T)$ given by (2.36). Thus, by the chain rule, we have

$$\frac{\partial}{\partial N} = \left(1 - \frac{\partial U}{\partial N}\right) \frac{\partial}{\partial \xi} + \frac{\partial U}{\partial N} \frac{\partial}{\partial \eta}$$

and

$$\frac{\partial}{\partial T} = -\frac{\partial U}{\partial T} \frac{\partial}{\partial \xi} + \frac{\partial U}{\partial T} \frac{\partial}{\partial \eta}.$$

By implicitly differentiating (2.36) with respect to N and T we obtain

$$\frac{\partial U}{\partial N} = \frac{\eta}{\xi(\xi + \eta - 1)}, \quad \frac{\partial U}{\partial T} = \frac{\xi - 1}{\xi}.$$

Then by (5.9), after changing variables, the operator \mathcal{D} can be rewritten as

$$\mathcal{D} = \frac{\xi + \eta - 1}{\xi + \eta} \frac{\partial}{\partial \eta} + \frac{1}{\xi + \eta}. \quad (5.12)$$

If we denote $P_1^*(N, T)$ by $F(\xi, \eta)$, then by the chain rule we have

$$\frac{\partial P_0}{\partial N} = \left(1 - \frac{\partial U}{\partial N}\right) \frac{\partial F}{\partial \xi} + \frac{\partial U}{\partial N} \frac{\partial F}{\partial \eta} \equiv F_1(\xi, \eta) \quad (5.13)$$

and

$$\frac{\partial^2 P_0}{\partial N^2} = \left(1 - \frac{\partial U}{\partial N}\right) \frac{\partial F_1}{\partial \xi} + \frac{\partial U}{\partial N} \frac{\partial F_1}{\partial \eta} \equiv F_2(\xi, \eta). \quad (5.14)$$

Thus, by using (5.11)-(5.14) and noting that $N = \xi + \eta$, we have

$$\begin{aligned} \mathcal{D}F &= \frac{\xi + \eta - 1}{\xi + \eta} \frac{\partial F}{\partial \eta} + \frac{1}{\xi + \eta} F \\ &= \frac{\xi + \eta + 1}{2(\xi + \eta)} F_2 + \frac{\xi + \eta - 1}{(\xi + \eta)^2} F_1 + \frac{1}{(\xi + \eta)^3} F. \end{aligned}$$

Multiplying the above by $\xi + \eta (= N)$ and using (5.12) we obtain

$$\frac{\partial}{\partial \eta} [(\xi + \eta - 1) F] = \frac{\xi + \eta + 1}{2} F_2 + \frac{\xi + \eta - 1}{\xi + \eta} F_1 + \frac{1}{(\xi + \eta)^2} F. \quad (5.15)$$

Solving (5.15) with the help of the symbolic computation program MAPLE, then imposing the initial condition $P_1^*(N, 0) = F(\xi, 0) = 0$ and using (5.8), we obtain the correction term $P_1(N, T)$ in (2.34) as follows

$$\begin{aligned} P_1(N, T) &= F(\xi, \eta) \\ &= -\frac{(\xi - 1)(2\xi - 3)}{2\xi^5} + \frac{3(\xi - 1)(2\xi^2 - 2\xi - 3)}{2\xi^5(\xi + \eta - 1)} \\ &\quad - \frac{(\xi - 1)^2(2\xi^3 + 2\xi^2 - 5\xi - 15)}{2\xi^5(\xi + \eta - 1)^2} - \frac{(\xi - 1)^3(2\xi^2 + 4\xi + 3)}{2\xi^5(\xi + \eta - 1)^3} \\ &\quad - \frac{2(\xi - 1)(2\xi - 3)}{\xi^5(\xi + \eta - 1)} \log \left| \frac{\xi + \eta - 1}{\xi - 1} \right|, \end{aligned} \quad (5.16)$$

where $\xi = N - U$, $\eta = U$.

Now we consider some special cases. If $N = 1$, then $U \rightarrow 0$ by (2.36). Thus, $\xi \sim 1$, $\eta \sim 0$ and

$$\frac{\xi - 1}{\xi + \eta - 1} \sim e^{-T}.$$

Then (5.16) reduces to the explicit result

$$P_1(1, T) = \left(2T - \frac{9}{2}\right) e^{-T} + 8e^{-2T} - \frac{9}{2}e^{-3T}.$$

We already showed that $P_0(1, T) = e^{-T}$. We relate the explicit result along $N = 1$ ($n = \rho$) to the spectral expansion (2.6). From (2.6) we can easily show that only the eigenvalues ν_1 , ν_2 and ν_3 are $O(\rho^{-1})$ or $O(\rho^{-2})$ in this limit (the others are $o(\rho^{-2})$). Expanding (2.6)-(2.10) for $\rho \rightarrow \infty$ and $N = 1$, we obtain

$$C_1(\nu_1) \phi_1(\rho, \nu_1) e^{\nu_1 t} \sim \rho^{-1} e^{-T} + \rho^{-2} \left(2T - \frac{9}{2}\right) e^{-T},$$

$$C_2(\nu_2) \phi_2(\rho, \nu_2) e^{\nu_2 t} \sim 8 \rho^{-2} e^{-2T},$$

and

$$C_3(\nu_3) \phi_3(\rho, \nu_3) e^{\nu_3 t} \sim -\frac{9}{2} \rho^{-2} e^{-3T}.$$

Thus (2.6) agrees precisely with $\rho^{-1} P_0(1, T) + \rho^{-2} P_1(1, T)$.

If $N = 0$ and $T \rightarrow \infty$, then $U \rightarrow -1$, $\eta \sim -1$, and $\xi \sim 1 - e^{-1} e^{-T}$. Hence, from (5.16), $P_0(0, T) \sim e^{-1} e^{-T}$ and $P_1(0, T) \sim (2T - 3) e^{-1} e^{-T}$. From (2.6) we obtain for $n = 0$, $t = T\rho \rightarrow \infty$

$$C_1(\nu_1) \phi_1(0, \nu_1) e^{\nu_1 t} \sim \rho^{-1} e^{-1} e^{-T} + \rho^{-2} (2T - 3) e^{-1} e^{-T}.$$

Again, this agrees with (5.16) and shows that for $N = 0$ and $T \gg 1$ ($t \gg \rho$) only the first eigenvalue ν_1 contributes to the expansion of $p_n(t)$.

Next, we consider short time scales, with $n = O(1)$ and $t = \tau/\rho = O(\rho^{-1})$. Expanding the conditional sojourn time density $p_n(t)$ in the form $p_n(t) = \mathcal{Q}_n(\tau) + O(\rho^{-1})$ and using equation (2.4), we have

$$\mathcal{Q}'_n(\tau) = \frac{1}{n+1} [\mathcal{Q}_{n+1}(\tau) - \mathcal{Q}_n(\tau)],$$

with the initial condition $\mathcal{Q}_n(0) = 1/(n+1)$. Taking the Laplace transform over the time variable τ with $\widehat{\mathcal{Q}}_n(s) = \int_0^\infty \mathcal{Q}_n(\tau) e^{-\tau s} d\tau$, we obtain the following difference equation for $\widehat{\mathcal{Q}}_n(s)$:

$$\widehat{\mathcal{Q}}_{n+1}(s) - [(n+1)s + 1] \widehat{\mathcal{Q}}_n(s) = -1. \quad (5.17)$$

Solving (5.17) yields

$$\widehat{\mathcal{Q}}_n(s) = \sum_{j=0}^{\infty} s^{-j-1} \frac{\Gamma(n+1+1/s)}{\Gamma(n+j+2+1/s)}. \quad (5.18)$$

By the inverse Laplace transform, we have

$$\mathcal{Q}_n(\tau) = \frac{1}{2\pi i} \int_{Br} \frac{e^{\tau s}}{s} \sum_{j=0}^{\infty} s^{-j} \frac{\Gamma(n+1+1/s)}{\Gamma(n+j+2+1/s)} ds. \quad (5.19)$$

Using the identity

$$\int_0^1 t^{x-1} (1-t)^{y-1} dt = \frac{\Gamma(x) \Gamma(y)}{\Gamma(x+y)}, \quad x, y > 0,$$

we can rewrite (5.19) as

$$\begin{aligned}
\mathcal{Q}_n(\tau) &= \frac{1}{2\pi i} \int_{Br} \frac{e^{\tau s}}{s} \sum_{j=0}^{\infty} \frac{s^{-j}}{j!} \int_0^1 z^j (1-z)^{n+1/s} dz ds \\
&= \frac{1}{2\pi i} \int_{Br} \frac{e^{\tau s}}{s} \int_0^1 e^{z/s} (1-z)^{n+1/s} dz ds \\
&= \int_0^1 (1-z)^n \frac{1}{2\pi i} \int_{Br} e^{\tau s} \frac{e^{z/s} (1-z)^{1/s}}{s} ds dz.
\end{aligned}$$

Then by the inverse Laplace transform

$$\mathcal{L}^{-1}\left(\frac{1}{s} e^{W/s}\right) = J_0(2\sqrt{\tau|W|}),$$

where $W = z + \log(1-z) < 0$ and $J_0(\cdot)$ is the Bessel function, we obtain

$$\mathcal{Q}_n(\tau) = \int_0^1 (1-z)^n J_0\left(2\sqrt{\tau}\sqrt{-z - \log(1-z)}\right) dz,$$

which leads to (2.37). By expanding $\mathcal{Q}_n(\tau)$ for $n \rightarrow \infty$ and $\tau \rightarrow \infty$, with $\tau = O(n^2)$ we obtain

$$\mathcal{Q}_n(\tau) \sim \int_0^{\infty} e^{-nz} J_0(z\sqrt{2\tau}) dz = \frac{1}{\sqrt{n^2 + 2\tau}} = \frac{1}{\rho\sqrt{N^2 + 2T}}. \quad (5.20)$$

Then we can easily show that $\rho^{-1}P_0(N, T)$, when expanded for $(N, T) \rightarrow (0, 0)$, gives the same result as in (5.20), which verifies the matching between the long time (T -scale) and short time (τ -scale) results.

We show how to obtain the results in Theorem 2.5 from the exact representations. First consider $n, t = O(\rho)$. Then from (2.7) and (2.8) we obtain $\nu_m \sim -m/\rho$ and $\tilde{\nu}_m = O(\rho)$. Thus on the large time scales $t = T\rho = O(\rho)$ we have $e^{\nu_m t} \sim e^{-mT}$, while $e^{\tilde{\nu}_m t}$ becomes exponentially small. Thus all of the terms in the first sum in (2.6) contribute to the leading term for $p_n(t)$. We furthermore scale $n = N\rho$ and use

$$C_m(\nu) \sim \frac{1}{\rho} \frac{m^m}{m!} e^{-m}, \quad \left(\frac{\nu+1}{-\rho}\right)^n \sim (-\rho)^{-n} e^{-mN}$$

and

$$n! L_n^{m-1-n} \left(\frac{\rho}{(\nu+1)^2}\right) \sim (-\rho)^n (1-N)^{m-1} e^{2mN}.$$

Thus from (2.6), on the (N, T) scale, we obtain

$$p_n(t) \sim \rho^{-1} \sum_{m=1}^{\infty} e^{m(N-1)} (1-N)^{m-1} \frac{m^m}{m!} e^{-mT}.$$

We can show that this is equal to $\rho^{-1}P_0(N, T)$ in (5.6).

Next we consider $n = O(1)$ and $t = \tau/\rho = O(\rho^{-1})$. We use the exact representation for $p_n(t)$ in Theorem 2.2 and scale $\theta = \rho s - 1 = O(\rho)$ ($s > 0$). Then $r \sim 1/s$ and from (2.13) we have

$$M \sim \frac{\rho^{1/s}}{s \Gamma(1 + 1/s)}. \quad (5.21)$$

By scaling $z = w/(\rho s) = O(\rho^{-1})$ in (2.14), G_n becomes

$$\begin{aligned} G_n &\sim \frac{1}{(\rho s)^{n+1+1/s}} \int_0^1 w^n (1-w)^{1/s} dw \\ &= \frac{n! \Gamma(1 + 1/s)}{\Gamma(n + 2 + 1/s) (\rho s)^{n+1+1/s}}. \end{aligned} \quad (5.22)$$

We also find, from (2.15), that H_n is asymptotically given by

$$H_n \sim \int_0^{\infty} z^{n+1/s} e^{-z/s} dz = s^{n+1+1/s} \Gamma(n + 1 + 1/s). \quad (5.23)$$

Now we rewrite (2.12) as

$$\widehat{p}_n(\theta) = M G_n \sum_{l=0}^{n-1} \frac{\rho^l}{l!} H_l + M H_n \sum_{l=n}^{\infty} \frac{\rho^l}{l!} G_l \quad (5.24)$$

and use (5.21)-(5.23) in (5.24). We find that the first term in the right-hand side of (5.24) is $O(\rho^{-2})$ and the second term is $O(\rho^{-1})$. Thus, the second term dominates the first and we have

$$\begin{aligned} \widehat{p}_n(\theta) &\sim M H_n \sum_{l=n}^{\infty} \frac{\rho^l}{l!} G_l \\ &\sim \frac{1}{\rho} \sum_{l=n}^{\infty} s^{n-l-1} \frac{\Gamma(n + 1 + 1/s)}{\Gamma(l + 2 + 1/s)} \\ &= \frac{1}{\rho} \sum_{j=0}^{\infty} s^{-j-1} \frac{\Gamma(n + 1 + 1/s)}{\Gamma(n + j + 2 + 1/s)}, \end{aligned}$$

which corresponds to $\rho^{-1}\widehat{\mathcal{Q}}_n(s)$ in (5.18). Inverting the transform using $d\theta = \rho ds$ leads to the same result we obtained by the perturbation method.

6 Discussion

To summarize, we have obtained both exact and asymptotic results for the $M/M/1$ -PS model with non-balking probability $b_n = 1/(n+1)$. We compare our results to the standard model, where $b_n = 1$. First, the spectral representation of $p_n(t)$ for the two models is very different as the standard model has a purely continuous spectrum (see also Guillemin and Boyer [6]) while the balking model has a purely discrete one.

We recently studied (see [17]) $p_n(t)$ for the standard model asymptotically, and found that if $\rho = \lambda/\mu < 1$ and $n, t \rightarrow \infty$ the asymptotic expansion is different according as $n/t > 1 - \rho$, $n/t \approx 1 - \rho$, $0 < n/t < 1 - \rho$, $n = O(t^{2/3})$, and $n = O(1)$. The scale $n = O(t^{2/3})$ is important in obtaining the tail of the unconditional density, which for the standard PS model has the form (see [13] and [3])

$$p_{PS}(t) \sim \alpha_2 t^{-5/6} e^{-\alpha_0 t} e^{-\alpha_1 t^{1/3}},$$

where $\alpha_0 = (1 - \sqrt{\rho})^2$ and α_1 and α_2 are constants. In contrast, for the model with balking Theorem 2.3 shows that the structure of $p_n(t)$ is different in three main sectors of the (n, t) plane ($n/t > 1$, $\Lambda_0 < n/t < 1$ and $0 < n/t < \Lambda_0$), with two transition regions connecting them. For $t \rightarrow \infty$ with $0 \leq n/t < \Lambda_0$ the asymptotics of $p_n(t)$ are governed by the eigenvalue with the largest real part and we obtain the purely exponential behavior in (2.23), which leads to (2.40) for the unconditional density $p_{PS}(t)$. Thus for the model with balking the scale $n = O(t^{2/3})$ is absent.

If $\rho > 1$ the standard PS model has an algebraic tail, with $p_{PS}(t) \sim \alpha_3 t^{-\rho/(\rho-1)}$, so that the mean sojourn time is finite for $\rho < 2$, the second moment is finite for $\rho < 3/2$, etc. Then the approximation

$$p_n(t) \sim \frac{1}{n} \left[1 + (\rho - 1) \frac{t}{n} \right]^{-\frac{\rho}{\rho-1}}$$

applies for n and/or $t \rightarrow \infty$. This situation is similar to the model with balking in the limit $\rho \rightarrow \infty$. Here the tail will be purely exponential, but for

n and/or $t \rightarrow \infty$ we have the approximation in (2.34), which is quite unlike the three sectors in Theorem 2.3.

We derived Theorem 2.5 by both a perturbation method and by using the exact representations. The former method should also be useful for general non-balking functions b_n , provided that we can write ρb_n in the form $\rho b_n = B(\varepsilon n)$, where ε is a small parameter. Thus ρb_n is a “slowly varying” function of n . For example, this would apply to $b_n = e^{-cn}$ (used by Morse [12]) if c is small. This limit would also apply to repairman problems (or finite populations queues) where $b_n = M - n$ and M is the customer population. Then $\rho b_n = \rho M(1 - n/M)$ and we would assume that $M \rightarrow \infty$ (thus $\varepsilon = M^{-1}$) and $\rho \rightarrow 0$, with $\rho M = O(1)$. It is likely that the asymptotic structure of all of these models is quite different, and the perturbation method should clearly show these differences.

References

- [1] S. C. Borst, O. J. Boxma, J. A. Morrison, and R. Núñez-Queija, The equivalence between processor sharing and service in random order, *Oper. Res. Letters* 31 (2003), 254-262.
- [2] E. G. Coffman, Jr., R. R. Muntz, and H. Trotter, Waiting time distributions for processor-sharing systems, *J. ACM* 17 (1970), 123-130.
- [3] J. W. Cohen, On processor sharing and random service (Letter to the editor), *J. Appl. Prob.* 21 (1984), 937-937.
- [4] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth, On the Lambert W function, *Adv. Comput. Math.* 5 (1996), 329-359.
- [5] L. Flatto, The waiting time distribution for the random order service $M/M/1$ queue, *Ann. Appl. Prob.* 7 (1997), 382-409.
- [6] F. Guillemin and J. Boyer, Analysis of the $M/M/1$ queue with processor sharing via spectral theory, *Queueing Syst.* 39 (2001), 377-397.
- [7] F. A. Haight, Queueing with balking, *Biometrika* 44 (1957), 360-369.

- [8] L. Kleinrock, Analysis of a time-shared processor, *Naval Research Logistics Quarterly* 11 (1964), 59-73.
- [9] C. Knessl, On the sojourn time distribution in a finite capacity processor shared queue, *J. Assoc. Comput. Mach.* 40 (1993), 1238-1301.
- [10] W. Magnus, F. Oberhettinger, and R. P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics*, Springer-Verlag, New York (1966).
- [11] J. A. Morrison, Response-time distribution for a processor-sharing system, *SIAM J. Appl. Math.* 45 (1985), 152-167.
- [12] P. M. Morse, *Queues, Inventories and Maintenance*, Wiley, New York (1958).
- [13] F. Pollaczek, La loi d'attente des appels téléphoniques, *C. R. Acad. Sci. Paris* 222 (1946), 353-355.
- [14] J. Riordan, *Stochastic Service Systems*, Wiley, New York (1962).
- [15] E. Vulot, Délais d'attente des appels téléphoniques traités au hasard, *C. R. Acad. Sci. Paris* 222 (1946), 268-269.
- [16] R. Wong, *Asymptotic Approximation of Integrals*, SIAM, Philadelphia (2001).
- [17] Q. Zhen and C. Knessl, On sojourn times in the $M/M/1$ -PS model, conditioned on the number of other users, submitted to *Queueing Syst.*
- [18] Q. Zhen and C. Knessl, Asymptotic expansions for the conditional sojourn time distribution in the $M/M/1$ -PS queue, *Queueing Syst.* 57 (2007), 157-168.