Euro. Jnl of Applied Mathematics (1999), *vol.* 10, *pp.* 497–509. Printed in the United Kingdom 497

A new heavy traffic limit for the asymmetric shortest queue problem

CHARLES KNESSL

Department of Mathematics, Statistics and Computer Science (M/C 249)
University of Illinois at Chicago, Chicago, IL 60607-7045

(*Received 13 March 1998; revised 31 March 1999*)

We consider the classic shortest queue problem in the heavy traffic limit. We assume that the second server works slowly and that the service rate of the first server is nearly equal to the arrival rate. Solving for the (asymptotic) joint steady state queue length distribution involves analyzing a backward parabolic partial differential equation, together with appropriate side conditions. We explicitly solve this problem. We thus obtain a two-dimensional approximation for the steady state queue length probabilities.

1 Introduction

A classic model in queueing theory is the shortest queue problem. Here we have two queues in parallel with exponential servers, with respective rates μ_1 and μ_2. Customers arrive according to a Poisson process with rate λ, and they join the shorter of the two queues. If the queue lengths are identical, we assume that the customer joins either queue with probability $1/2$. Such problems arise naturally in applications such as computer network communications, packet switched data networks (see Foschini [1], [2]), airports with two runways (Gertsbakh [3]) and waiting in a line in a bank with two tellers.

This model is easy to formulate, but its analysis is far from trivial. The symmetric case ($\mu_1 = \mu_2$) was analysed by Flatto and McKean [4], who used generating functions and function-theoretic arguments. They characterized the (two-dimensional) generating function of the steady state queue length probabilities in terms of a meromorphic function of a single complex variable. This problem is also discussed in the book of Cohen and Boxma [5]. There the authors also use generating functions and complex variable arguments to convert the problem to a standard boundary value problem. Its solution, however, involves the conformal mapping of an ellipse onto a circle using Jacobi elliptic functions, and the form of the final result is not particularly explicit. An elementary solution of the shortest queue problem was obtained by Adan, Wessels, and Zijm, first [6] for the symmetric case and later [7] for the asymmetric case, where $\mu_1 \neq \mu_2$. The solution is an infinite sum of product form exponentials. By the latter we mean $a^{n_1} b^{n_2}$, where n_j denotes the number of customers in the jth queue and a, b are constants. The technique of deriving the solution is sometimes referred to as the "compensation approach". Asymptotically (for n_1, $n_2 \to \infty$) only two or three of these exponentials are important; these were obtained by Knessl, Matkowsky, Schuss and Tier [8] using singular perturbation methods. Some asymptotic results are also given in [4] and by Kingman [9].

The stability condition for this model is $\mu_1 + \mu_2 > \lambda$. The heavy traffic limit is defined as $\mu_1 + \mu_2 \downarrow \lambda$. Then the system is close to being unstable, and large queue lengths tend to develop. The shortest queue problem in heavy traffic is discussed in Knessl, et. al [10] and Reiman [11]. These investigations allow for general interarrival and/or service time distributions. The main feature is that for $(\mu_1 + \mu_2)/\lambda = 1 + \epsilon$ with $\epsilon \to 0^+$, the probability distribution becomes concentrated in the range where $n_1 + n_2$ is large (of the order $O(\epsilon^{-1})$), but with $n_1 - n_2 = O(1)$. Thus the mass becomes concentrated near the main diagonal $n_1 = n_2$ and the two-dimensional queue length process may be approximated by a one-dimension diffusion process along $n_1 = n_2$.

Recently, Turner [18] considered a shortest queue model where, in addition to the arrival stream of rate λ that is routed to the shorter queue, the two parallel queues are fed by independent arrival streams, with respective rates λ_1 and λ_2. Assuming that the service rates $\mu_1 = \mu_2 = 1$ and scaling $\lambda_j - 1 = O(N^{-1/2})$, $j = 1, 2$, with $\lambda = O(N^{-1/2})$, Turner obtains, in the limit $N \to \infty$, a two-dimensional diffusion model. This consists of a $2 - d$ reflected Brownian motion, which is shown to have essentially a product form stationary distribution.

The purpose of this note is to identify a new heavy traffic limit for the asymmetric case. We assume (cf. (2.1)) that $(\mu_1 + \mu_2)/\lambda \downarrow 1$ and simultaneously $\mu_2/\lambda \to 0$. This means that the second server works slowly compared to the first. The new scaling allows for large queue lengths to develop in both queues, and we shall show that the probability mass is now spread out over the octant $0 < n_1 < n_2$ in the (n_1, n_2) plane. Computing the steady state queue length probabilities in this limit involves solving a backward parabolic partial differential equation (PDE), with an appropriate boundary condition (BC) along $n_1 = 0$ and an "interface condition" along $n_1 = n_2$. We explicitly solve this boundary value problem, as an infinite sum of exponentials. We believe that the basic approach can be extended to the shortest queue problem with general interarrival and/or service distributions, though we only discuss the exponential case here. An alternate approach to the asymptotics might be to analyze the exact results in [7]. However, the present approach, which consists of deriving a PDE with appropriate side conditions and solving it, has the advantage of being generalizable to non-exponential arrivals and/or service, and (perhaps) to problems with more than two queues.

We also mention other work on models (either discrete or diffusion) that admit steady state distributions consisting of a finite or infinite sum of product form exponentials. A survey of discrete models analyzed by the compensation approach is given in van Houtum [12]. Newell [13] considers a diffusion model for tandem queues and, in one particular case, obtains an exact solution to the steady state density as an infinite sum of exponentials. A diffusion model for tandem queues with exponential service and deterministic arrivals has a solution consisting of a sum of three exponentials (see Knessl and Tier [14]), and similar solutions were obtained by Foschini [15] for other models. It should be noted that such solutions are the exception rather than the rule. For most models, either discrete or diffusion, the steady state solutions are much more complicated.

In § 2 we summarize the main results. The detailed calculations are given in § 3.

2 Main results

We let λ be the arrival rate and μ_1 (resp. μ_2) be the service rate of the first (resp. second) server. We define the heavy traffic limit by $\epsilon \to 0^+$ where

$$\frac{\mu_1 + \mu_2}{\lambda} = 1 + \epsilon, \qquad \frac{\mu_2}{\lambda} = \epsilon b \tag{2.1}$$

with $b > 0$. Thus, the second server works slowly and $\mu_1 \approx \lambda$. With (2.1), it is easy to see that large queue lengths tend to develop in both queues. Once a large backlog develops in the second queue, the customers are all routed to the first queue. But, since λ is nearly equal to μ_1, the first queue is in heavy traffic and large queue lengths also develop, until the number of customers n_1 in the first queue equals or exceeds the number n_2 in the second queue.

Under the stability condition $\mu_1 + \mu_2 > \lambda$ (i.e. $\epsilon > 0$) we define the steady state distribution by

$$p(n_1, n_2) = \lim_{t\to\infty} \text{Prob}\left[N_1(t) = n_1, N_2(t) = n_2\right], \tag{2.2}$$

where $N_j(t)$ is the number of customers in the jth queue at time t. The marginal probabilities will be denoted by

$$\overline{p}(n_1) = \sum_{n_2=0}^{\infty} p(n_1, n_2), \quad p(n_2) = \sum_{n_1=0}^{\infty} p(n_1, n_2). \tag{2.3}$$

We summarize below the main results of the paper.

Theorem 2.1 *With the scaling (2.1) and for* $b > 0$ *we have*

$$\left.\begin{aligned} p(n_1, n_2) &\sim \epsilon^2 P(x, y), \quad n_2 > n_1, \\ p(n_1, n_2) &\sim \epsilon^2 \frac{2}{3} P(x, x), \quad n_2 = n_1, \\ p(n_1, n_2) &\sim \epsilon^2 \epsilon^{n_1 - n_2 - 1} \left(\frac{b}{2}\right)^{n_1 - n_2 - 1} \frac{1}{6} P(x, x), \quad n_2 < n_1, \end{aligned}\right\} \tag{2.4}$$

where $(x, y) = \epsilon(n_1, n_2)$. *The function* $P(x, y)$ *satisfies the boundary value problem*

$$P_{xx} + (1 - b)P_x + bP_y = 0, \;\; y > x > 0 \tag{2.5}$$

$$P_x(0, y) + (1 - b)P(0, y) = 0, \;\; y > 0 \tag{2.6}$$

$$3P_x(x, x) + P_y(x, x) + (2 - 4b)P(x, x) = 0, \;\; x > 0 \tag{2.7}$$

$$\int_0^{\infty} \int_x^{\infty} P(x, y) dy\, dx = 1.$$

The solution of (2.5)–(2.7) is

$$P(x, y) = Ce^{-2(b+1)y} \sum_{\ell=0}^{\infty} [K_{2\ell} e^{(\ell+2)bx} + K_{2\ell+1} e^{-(b+1+b\ell)x}] \times \exp\left[-b\ell^2 y - (3b + 1)\ell y\right], \tag{2.8}$$

where

$$C = \frac{2(b + 1)(2b + 1)}{b}, \tag{2.9}$$

https://doi.org/10.1017/S0956792599003861 Published online by Cambridge University Press

$$K_{2\ell} = \frac{(-1)^\ell}{2}\frac{\ell+2}{\ell!}\frac{\Gamma(\ell+3+1/b)}{\Gamma(3+1/b)}, \quad K_{2\ell+1} = \frac{(-1)^\ell}{2}\frac{\ell+1+1/b}{\ell!}\frac{\Gamma(\ell+3+1/b)}{\Gamma(3+1/b)}. \tag{2.10}$$

The marginals in (2.3) are given by

$$\overline{p}(n_1) \sim \epsilon\overline{P}(x), \;\; p(n_2) \sim \epsilon P(y),$$

where

$$\overline{P}(x) = C\left[\frac{e^{-2x}}{2(b+1)} - \sum_{\ell=0}^{\infty}\frac{K_{2\ell+1}(2b\ell+4b+1)}{(\ell+1)(\ell+2)(b\ell+b+1)(b\ell+2b+1)}e^{-(\ell+3)(b\ell+b+1)x}\right], \tag{2.11}$$

$$P(y) = C\left[\frac{e^{-2y}}{2b} - \sum_{\ell=0}^{\infty}\frac{K_{2\ell+1}(2b\ell+4b+1)}{b(\ell+1)(b\ell+b+1)}e^{-(\ell+3)(b\ell+b+1)y}\right]. \tag{2.12}$$

We note that (2.5) is a parabolic PDE. A solution of the type (2.8) is only possible if $b > 0$, in which case (2.5) is backward parabolic. Note that, in view of (2.1), $b > 0$ implies that $\epsilon > 0$, which is needed for stability. The structure of the solution is similar to that of the discrete model (see [4], [6], [7]), though we do get some simplication in the diffusion limit. We believe that a similar boundary value problem can be obtained for models with general arrival and/or service distributions. For these problems it is unlikely that the discrete model can be solved exactly. We also observe that $p(n_1, n_2)$ is $O(\epsilon^2)$ for $n_2 > n_1$, $n_2 = n_1$ and $n_1 = n_2 + 1$. However, the total mass in these respective portions of the state space is (as $\epsilon \to 0^+$) $O(1)$, $O(\epsilon)$ and $O(\epsilon)$. Along diagonals where $n_1 - n_2 = \ell \geqslant 2$, the probabilities become negligible, both in magnitude and in total mass. Using the approach in section 3, it is possible to obtain further terms (e.g. the $O(\epsilon^3)$ correction for $n_2 > n_1$) in the asymptotic series (2.4).

We next show how to obtain alternate representations for $P(x, y)$, and also study the corner singularity of this function (i.e., its behavior as $(x, y) \to (0, 0)$). When $b = 1$ (2.8) becomes

$$\begin{aligned} P(x,y) &= e^{-4y}\sum_{\ell=0}^{\infty}(-1)^\ell(\ell+3)(\ell+2)^2(\ell+1)[e^{(\ell+2)x} + e^{-(\ell+2)x}]e^{-(\ell^2+4\ell)y} \\ &= \sum_{j=-\infty}^{\infty}(-1)^j j^2(j^2-1)e^{jx}e^{-j^2 y} \\ &= (\partial_y^2 - \partial_x^2)\sum_{j=-\infty}^{\infty}(-1)^j e^{jx}e^{-j^2 y} \\ &= (\partial_y^2 - \partial_x^2)\prod_{n=1}^{\infty}(1-e^{-2ny})(1-e^{-2ny}e^{x+y})(1-e^{-2ny}e^{y-x}). \end{aligned} \tag{2.13}$$

To obtain the last equality we have used properties of elliptic theta functions. The infinite product representation in (2.13) is useful for studying the asymptotic behavior of P as $(x, y) \to (0, 0)$. This is extremely difficult to obtain from (2.8), due to the alternating sum.

In the appendix we show that (for $b = 1$) as $x, y \to 0$

$$P(x,y) \sim \frac{\pi^{9/2}}{8}y^{-9/2}\exp\left(\frac{-\pi^2}{4y}\right)\cos\left(\frac{\pi x}{2y}\right), \qquad 0 \leqslant x/y < 1 \tag{2.14}$$

https://doi.org/10.1017/S0956792599003861 Published online by Cambridge University Press

and

$$P(y,y) \sim \frac{\pi^{7/2}}{2} y^{-7/2} \exp\left(\frac{-\pi^2}{4y}\right). \tag{2.15}$$

This shows that P is exponentially small in y as we approach the corner along any direction, and that P is smallest as we approach along $y = x$, where the boundary condition (2.7) applies. Note also that P is integrable and positive near the origin. It becomes negative for $x > y$, but there (2.4) shows that we must use a different approximation for the discrete probabilities $p(n_1, n_2)$.

Now consider general $b > 0$. When $b = 1/N$ where N is a positive integer, we can obtain a representation for P as a partial differential operator acting on a theta function, such as that in (2.13). In particular if $b = 1/2$ we get

$$P(x,y) = 2\partial_x\partial_y(\partial_y + 1)\left\{ e^{x-3y} \prod_{n=1}^{\infty} (1 - e^{-ny})(1 - e^{-ny} e^{x/2-2y})(1 - e^{-ny} e^{3y-x/2}) \right\}.$$

Then we can obtain the asymptotic behavior for $(x, y) \to (0, 0)$ using calculations similar to those in the appendix. For general $b > 0$ we develop a perturbation method (see the appendix) that yields the corner behavior up to a multiplicative constant. This uses only the equations (2.5)-(2.7) and the final result is

$$\left.\begin{aligned} P(x,y) &\sim (\text{const.}) y^{-7/2-1/b} \exp\left(\frac{-\pi^2}{4by}\right) \cos\left(\frac{\pi x}{2y}\right), \quad 0 \leqslant x/y < 1 \\ P(y,y) &\sim (\text{const.}) y^{-5/2-1/b} \exp\left(\frac{-\pi^2}{4by}\right) \frac{4b}{\pi}. \end{aligned}\right\} \tag{2.16}$$

The constant can presumably be determined from (2.8). When $b = 1$ (2.15) shows that const.$= \pi^{9/2}/8$. The expression in (2.16) again shows that P is positive and integrable near the origin, as it must be.

We conclude by noting that parabolic PDEs also arise in the study of heavy traffic diffusion approximations to priority queues [16]. There, however, the problems are forward rather than backward parabolic and the solutions are much different than that in (2.8)-(2.12). Heavy traffic analysis of stochastic fluid models [17] also lead to parabolic PDEs. These tend to have linear rather than constant coefficients, and may be forward parabolic in a part of the domain and backward parabolic in the remainder. The structure of the solutions, including the nature of the corner singularities, can be quite different for the forward, backward and forward/backward PDEs.

3 Calculations

The stationary probabilities satisfy the balance equations

$$(\lambda + \mu_1 + \mu_2) p(n_1, n_2) = \lambda p(n_1 - 1, n_2) + \mu_1 p(n_1 + 1, n_2) + \mu_2 p(n_1, n_2 + 1), \ n_2 > n_1 + 1, \tag{3.1}$$

$$\begin{aligned}(\lambda + \mu_1 + \mu_2) p(n_1, n_2) = {}& \lambda \left[p(n_1, n_2 - 1) + p(n_1 - 1, n_2) \right] + \mu_1 p(n_1 + 1, n_2) \\ & + \mu_2 p(n_1, n_2 + 1), \ n_2 = n_1,\end{aligned} \tag{3.2}$$

$$\begin{aligned}(\lambda + \mu_1 + \mu_2) p(n_1, n_2) = {}& \frac{\lambda}{2} p(n_1, n_2 - 1) + \lambda p(n_1 - 1, n_2) + \mu_1 p(n_1 + 1, n_2) \\ & + \mu_2 p(n_1, n_2 + 1), \ n_2 = n_1 + 1,\end{aligned} \tag{3.3}$$

https://doi.org/10.1017/S0956792599003861 Published online by Cambridge University Press

$$(\lambda+\mu_1+\mu_2)p(n_1,n_2) = \lambda p(n_1,n_2-1)+\frac{\lambda}{2}p(n_1-1,n_2)+\mu_1 p(n_1+1,n_2) + \mu_2 p(n_1,n_2+1),\ n_2=n_1-1, \quad (3.4)$$

$$(\lambda+\mu_1+\mu_2)p(n_1,n_2) = \lambda p(n_1,n_2-1)+\mu_1 p(n_1+1,n_2)+\mu_2 p(n_1,n_2+1),\ n_2<n_1-1 \quad (3.5)$$

and the boundary conditions

$$(\lambda+\mu_1)p(n_1,0) = \mu_1 p(n_1+1,0)+\mu_2 p(n_1,1),\ n_1 \geqslant 2, \quad (3.6)$$

$$(\lambda+\mu_2)p(0,n_2) = \mu_1 p(1,n_2)+\mu_2 p(0,n_2+1),\ n_2 \geqslant 2. \quad (3.7)$$

We do not give the corner equations (that apply for $n_1, n_2 = 0, 1$), as the heavy traffic analysis does not require these. The structure of $p(n_1, n_2)$ will be different for $n_1 > n_2$ and $n_1 < n_2$. We view (3.2)–(3.4) as 'interface' conditions, that couple the two parts of the state space. We also have $\sum_{n_1,n_2\geqslant 0} p(n_1,n_2) = 1$.

We first consider (3.1), use the scaling (2.1), set $(x, y) = \epsilon(n_1, n_2)$ and expand p as

$$p(n_1,n_2) = \epsilon^2 P(x,y;\epsilon) = \epsilon^2\left[P_0(x,y)+\epsilon P_1(x,y)+\epsilon^2 P_2(x,y)+\cdots\right]. \quad (3.8)$$

Then (3.1) becomes

$$\left[1+\epsilon(1-b)\right]\left[P(x+\epsilon,y)-P(x,y)\right]+\epsilon b\left[P(x,y+\epsilon)-P(x,y)\right]+P(x-\epsilon,y)-P(x,y)=0.$$

Using (3.8) the above becomes, as $\epsilon \to 0^+$,

$$P_{0,xx}+(1-b)P_{0,x}+bP_{0,y}=0,$$

which is the same as (2.5). The boundary condition (3.7) becomes

$$P(\epsilon,y)-P(0,y)+\epsilon b\left[P(0,y+\epsilon)-P(0,y)\right]+\epsilon(1-b)P(\epsilon,y)=0.$$

Letting $\epsilon \to 0$ we obtain $P_{0,x}(0,y)+(1-b)P_0(0,y)=0$, which is (2.6).

It remains to analyze the region $n_2 < n_1$ and the vicinity of the interface. By examining (3.2)-(3.4), we see that $p(n_1, n_1)$ and $p(n_1, n_1-1)$ must be of the same order of magnitude as $p(n_1, n_2)$ for $n_2 > n_1$. Thus we set

$$\left.\begin{aligned} p(n_1,n_1-1) &= \overline{R}(n_1)=\epsilon^2 R(x)=\epsilon^2\left[R_0(x)+\epsilon R_1(x)+\epsilon^2 R_2(x)+\cdots\right] \\ p(n_1,n_1) &= \overline{Q}(n_1)=\epsilon^2 Q(x)=\epsilon^2\left[Q_0(x)+\epsilon Q_1(x)+\epsilon^2 Q_2(x)+\cdots\right]. \end{aligned}\right\} \quad (3.9)$$

Equation (3.3) may be rewritten as

$$(2+\epsilon)P(x,x+\epsilon)=\frac{1}{2}Q(x)+P(x-\epsilon,x+\epsilon)+\left[1+\epsilon(1-b)\right]Q(x+\epsilon)+\epsilon bP(x,x+2\epsilon). \quad (3.10)$$

Using (3.8) and (3.9) we obtain from (3.10) as $\epsilon \to 0$

$$P_0(x,x)=\frac{3}{2}Q_0(x) \quad (3.11)$$

and at the next order

$$P_1(x,x)=\frac{3}{2}Q_1(x)+(1-b)Q_0(x)+Q_0'(x)+(b-1)P_0(x,x)-P_{0,x}(x,x)-P_{0,y}(x,x). \quad (3.12)$$

Similarly, the scaled version of (3.2) is

$$(2+\epsilon)Q(x)=R(x)+P(x-\epsilon,x)+\epsilon bP(x,x+\epsilon)+\left[1+\epsilon(1-b)\right]R(x+\epsilon). \quad (3.13)$$

Letting $\epsilon \to 0$ in (3.13) yields

$$2Q_0(x) = 2R_0(x) + P_0(x,x) \tag{3.14}$$

and then

$$2Q_1(x) + Q_0(x) = 2R_1(x) + P_1(x,x) + R_0'(x) + (1-b)R_0(x) + bP_0(x,x) - P_{0,x}(x,x). \tag{3.15}$$

Next we consider (3.5) for $n_1 - n_2 > 2$. If we set $p(n_1,n_2) \sim \epsilon^2 \widetilde{P}(x,y)$ and use (2.1) in (3.5), we obtain to leading order $\widetilde{P}_x - \widetilde{P}_y = 0$, so that $\widetilde{P}(x,y) = F\big((x+y)/2\big)$. Then we let $\epsilon \to 0$ in (3.4) and obtain

$$2R_0(x) = F(x) + \frac{1}{2}Q_0(x) + F(x). \tag{3.16}$$

But, in view of (3.11) and (3.14), we have $4R_0(x) = Q_0(x)$ so that (3.16) forces $F(x) = 0$. We conclude that $p(n_1,n_2) = o(\epsilon^2)$ for $n_1 - n_2 \geqslant 2$. We set $n_2 = n_1 - 2$ in (3.5) to find that $p(n_1, n_1 - 2) \sim \epsilon^3 bR_0(x)/2$. Then we can easily show that (3.5) is satisfied asymptotically by

$$p(n_1,n_2) \sim \epsilon^{n_1-n_2+1}\left(\frac{b}{2}\right)^{n_1-n_2-1} R_0(x) \tag{3.17}$$

and this also satisfies (asymptotically) the boundary condition (3.6).

Now consider (3.4) with the scaling (2.1), and with (3.17) used to approximate $p(n_1, n_1-2)$ and $p(n_1+1, n_1-1)$. This equation becomes

$$\begin{aligned}(2+\epsilon)R(x) = {}& \frac{1}{2}Q(x-\epsilon) + \epsilon b Q(x) + \frac{\epsilon b}{2}\big[R_0(x) + O(\epsilon)\big] \\ & + \epsilon\big[1 + \epsilon(1-b)\big]\left[\frac{b}{2}R_0(x+\epsilon) + O(\epsilon)\right].\end{aligned} \tag{3.18}$$

In view of (3.9), (3.18) yields

$$2R_0(x) = \frac{1}{2}Q_0(x), \tag{3.19}$$

$$2R_1(x) + R_0(x) = b\big[Q_0(x) + R_0(x)\big] + \frac{1}{2}Q_1(x) - \frac{1}{2}Q_0'(x). \tag{3.20}$$

We consider the equations (3.11), (3.12), (3.14), (3.15), (3.19) and (3.20). While (3.19) is consistent with (3.11) and (3.14), it yields no new information, as we already found that $Q_0(x) = \frac{2}{3}P_0(x,x)$ and $R_0(x) = \frac{1}{6}P_0(x,x)$. Using this result in (3.12) we obtain

$$P_1(x,x) - \frac{3}{2}Q_1(x) = \frac{1}{3}\big[(b-1)P_0(x,x) - P_{0,x}(x,x) - P_{0,y}(x,x)\big]. \tag{3.21}$$

Adding (3.15) and (3.21) we are led to

$$\frac{1}{2}Q_1(x) - 2R_1(x) = \frac{7b-5}{6}P_0(x,x) - \frac{7}{6}P_{0,x}(x,x) - \frac{1}{6}P_{0,y}(x,x) \tag{3.22}$$

since $P_1(x,x)$ drops out. Equation (3.20) may also be written as

$$\frac{1}{6}P_0(x,x) = \frac{1}{2}Q_1(x) - 2R_1(x) + \frac{5b}{6}P_0(x,x) - \frac{1}{3}\big[P_{0,x}(x,x) + P_{0,y}(x,x)\big]. \tag{3.23}$$

By comparing (3.22) and (3.23) we find that

$$-\frac{3}{2}P_{0,x}(x,x) - \frac{1}{2}P_{0,y}(x,x) + (2b-1)P_0(x,x) = 0. \tag{3.24}$$

https://doi.org/10.1017/S0956792599003861 Published online by Cambridge University Press

This yields the necessary boundary condition along $y = x$ for the PDE (2.5), which applies in the range $y > x > 0$. By solving (2.5)-(2.7) we determine $P_0(x, y)$ and then (3.9) and (3.17) give the leading term for $p(n_1, n_2)$ for $n_1 \geqslant n_2$.

An alternate derivation of the problem (2.5)-(2.7) can be made by writing the generator for the Markov chain, introducing the scaling (2.1), letting $\epsilon \to 0$, and then determining the adjoint PDE and boundary conditions. We have verified that this yields the same results as above. While the alternate approach may give an easier way of deriving the boundary conditions (especially (3.24)), the longer approach presented here has the advantage of carefully relating the discrete probabilities along the interface, to the boundary values of the diffusion approximation (i.e. $P_0(x, x)$). Using (2.4) allows us to estimate the probability mass along diagonals where $n_1 - n_2 = \ell \geqslant 0$.

To solve (2.5)–(2.7), we first set $y = x + \xi$. In terms of (x, ξ), the problem becomes

$$P_{xx} - 2P_{x\xi} + P_{\xi\xi} + (1-b)P_x + (2b-1)P_\xi = 0; \quad x, \xi > 0 \tag{3.25}$$

$$P_x - P_\xi + (1-b)P = 0, \quad x = 0 \tag{3.26}$$

$$3P_x - 2P_\xi + (2-4b)P = 0, \quad \xi = 0. \tag{3.27}$$

Furthermore, setting

$$P(x, \xi) = \exp\left[\frac{b-1}{2}x + \frac{(b-1)^2}{4b}(x+\xi)\right] Q(x, \xi), \tag{3.28}$$

the problem (3.25)-(3.27) becomes

$$Q_{xx} - 2Q_{x\xi} + Q_{\xi\xi} + bQ_\xi = 0; \quad x, \xi > 0 \tag{3.29}$$

$$Q_x - Q_\xi + \frac{1}{2}(1-b)Q = 0, \quad x = 0 \tag{3.30}$$

$$3Q_x - 2Q_\xi + \left(\frac{1}{4b} - \frac{9b}{4}\right)Q = 0, \quad \xi = 0. \tag{3.31}$$

We seek solutions to (3.29)-(3.31) in the form

$$Q = C\left[e^{A_0 x + B_0 \xi} + K_1 e^{A_1 x + B_0 \xi} + K_2 e^{A_1 x + B_1 \xi} + K_3 e^{A_2 x + B_1 \xi} + K_4 e^{A_2 x + B_2 \xi} + \cdots\right]. \tag{3.32}$$

We require that the first term in (3.32) satisfies the PDE and the BC along $\xi = 0$, so that

$$A_0^2 + B_0^2 - 2A_0 B_0 + bB_0 = 0,$$
$$3A_0 - 2B_0 + \frac{1}{4b} - \frac{9b}{4} = 0,$$

and hence

$$A_0 = -\frac{1}{4b} - 1 - \frac{3b}{4}, \quad B_0 = -\frac{1}{4b} - \frac{3}{2} - \frac{9b}{4}. \tag{3.33}$$

Each term in (3.32) will satisfy the PDE, provided that

$$\left.\begin{aligned} A_n^2 + B_n^2 - 2A_n B_n + bB_n = 0, \quad n \geqslant 0 \\ A_{n+1}^2 + B_n^2 - 2A_{n+1} B_n + bB_n = 0, \quad n \geqslant 0. \end{aligned}\right\} \tag{3.34}$$

Since we may assume that $A_{n+1} \neq A_n$, the above can be replaced by the linear recurrences

$$A_{n+1} + A_n = 2B_n, \quad B_{n+1} + B_n = 2A_{n+1} - b. \tag{3.35}$$

To obtain the latter we have replaced n by $n+1$ in the first equation in (3.34), and then subtracted the second equation. The difference equations (3.35) are easily solved (subject to the initial conditions in (3.33)) to give

$$A_n = -bn^2 - (2b+1)n - \left(\frac{1}{4b} + 1 + \frac{3b}{4}\right) \tag{3.36}$$

and

$$B_n = -bn^2 - (3b+1)n - \left(\frac{1}{4b} + \frac{3}{2} + \frac{9b}{4}\right). \tag{3.37}$$

The constants K_j in (3.32) are chosen so that the BC (3.30) and (3.31) are satisfied. In view of the form of the series (3.32), we see that the BC at $x = 0$ is satisfied if

$$\left[A_{\ell+1} - B_\ell + \frac{1}{2}(1-b)\right]K_{2\ell+1} + \left[A_\ell - B_\ell + \frac{1}{2}(1-b)\right]K_{2\ell} = 0, \quad \ell \geqslant 0 \tag{3.38}$$

and that at $\xi = 0$ holds if

$$\left[3A_{\ell+1} - 2B_{\ell+1} + \frac{1}{4b} - \frac{9b}{4}\right]K_{2\ell+2} + \left[3A_{\ell+1} - 2B_\ell + \frac{1}{4b} - \frac{9b}{4}\right]K_{2\ell+1} = 0, \quad \ell \geqslant 0. \tag{3.39}$$

Here we define $K_0 = 1$. Using (3.36) and (3.37) we find that $A_\ell - B_\ell + \frac{1}{2}(1-b) = b\ell + b + 1$, $A_{\ell+1} - B_\ell + \frac{1}{2}(1-b) = -b(\ell+2)$, $3A_{\ell+1} - 2B_\ell + (4b)^{-1} - 9b/4 = -b(\ell+3)(\ell+3+1/b)$ and $3A_{\ell+1} - 2B_{\ell+1} + (4b)^{-1} - 9b/4 = -(\ell+1)(b\ell+b+1)$. It follows from (3.38) and (3.39) that

$$\frac{K_{2\ell+1}}{K_{2\ell}} = \frac{\ell+1+1/b}{\ell+2} \quad \text{and} \quad \frac{K_{2\ell+2}}{K_{2\ell+1}} = -\frac{(\ell+3)(\ell+3+1/b)}{(\ell+1)(\ell+1+1/b)}.$$

Hence, using $K_0 = 1$, we obtain

$$\begin{pmatrix} K_{2\ell} \\ K_{2\ell+1} \end{pmatrix} = \frac{(-1)^\ell}{2\ell!}\frac{\Gamma(\ell+3+1/b)}{\Gamma(3+1/b)}\begin{pmatrix} \ell+2 \\ \ell+1+1/b \end{pmatrix}, \quad \ell \geqslant 0, \tag{3.40}$$

where $\Gamma(\cdot)$ is the gamma function. By combining (3.40), (3.36), (3.37), (3.32) and (3.28) we obtain precisely the formula in (2.8) (with (2.10)). The marginals in (2.11) and (2.12) are obtained by noting that

$$\overline{p}(n_1) = \sum_{n_2=0}^{\infty} p(n_1,n_2) \sim \sum_{n_2=n_1+1}^{\infty} p(n_1,n_2) \sim \epsilon^2 \sum_{n_2=x/\epsilon}^{\infty} P(x,\epsilon n_2) \sim \epsilon \int_x^\infty P(x,y)\,dy$$

and

$$p(n_2) = \sum_{n_1=0}^{\infty} p(n_1,n_2) \sim \sum_{n_1=0}^{n_2-1} p(n_1,n_2) \sim \epsilon^2 \sum_{n_1=0}^{y/\epsilon} P(\epsilon n_1, y) \sim \epsilon \int_0^y P(x,y)\,dx.$$

Using (2.8) to evaluate these two integrals we obtain, after some rearranging, the expressions in (2.11) and (2.12).

It remains only to determine the constant C in (3.32) by normalization. Using (2.8) we obtain

$$\frac{1}{C} = \frac{1}{C}\int_0^\infty \int_x^\infty P(x,y)\,dy\,dx = \sum_{\ell=0}^{\infty}\left[\frac{K_{2\ell}}{b(\ell+2)^2(b\ell+1)} - \frac{K_{2\ell+1}}{(\ell+3)(b\ell+b+1)^2}\right]. \tag{3.41}$$

For large ℓ the summand in (3.41) is of order $O\big((-1)^\ell \ell^{1/b-1}\big)$. Thus the series converges

https://doi.org/10.1017/S0956792599003861 Published online by Cambridge University Press

only for $b > 1$. To evaluate the sum we use

$$\Gamma(\ell + 3 + 1/b) = \int_0^\infty t^{\ell+2+1/b} e^{-t}\, dt$$

and

$$\sum_{m=0}^{\infty} x^{m+1} \frac{(-1)^m}{m!} \frac{\Gamma(m+v)}{\Gamma(v)} = x(1+x)^{-v}. \tag{3.42}$$

We integrate (3.42) over the range $[0, x]$, then multiply the result by $x^{\alpha-3}$ and again integrate over $[0, x]$. We thus have

$$\begin{aligned} \sum_{m=0}^{\infty} \frac{x^{m+\alpha}(-1)^m \Gamma(m+v)}{(m+\alpha)(m+2)m!\Gamma(v)} &= \int_0^x u^{\alpha-3} \int_0^u t(1+t)^{-v}\, dt\, du \\ &= \frac{1}{\alpha-2} \int_0^x \frac{x^{\alpha-2}t - t^{\alpha-1}}{(1+t)^v}\, dt. \end{aligned} \tag{3.43}$$

Using (3.43) with $x = 1$, $\alpha = 1/b$ and $v = 3 + 1/b$ we find that

$$\sum_{\ell=0}^{\infty} \frac{K_{2\ell}}{b(\ell+2)^2(b\ell+1)} = \frac{1}{2b(1-2b)} \int_0^1 (t - t^{1/b-1})(1+t)^{-3-1/b}\, dt. \tag{3.44}$$

A completely analogous calculation can be used to represent the second sum in (3.41) as an integral. When combined with (3.44), (3.41) then becomes

$$\frac{1}{C} = \frac{1}{2b(1-2b)} \int_0^1 (1-t)(t - t^{1/b-1})(1+t)^{-3-1/b}\, dt. \tag{3.45}$$

Note that the expression remains finite if $b = 1/2$. Upon changing integration variables with $u = t/(1+t)$ we obtain

$$\begin{aligned} I &\equiv \int_0^1 (1-t)t(1+t)^{-3-1/b}\, dt \\ &= -\frac{2b}{2b+1}\left[1 - \left(\frac{1}{2}\right)^{2+1/b}\right] + \frac{3b}{b+1}\left[1 - \left(\frac{1}{2}\right)^{1+1/b}\right] - b\left[1 - \left(\frac{1}{2}\right)^{1/b}\right] \end{aligned} \tag{3.46}$$

and

$$\begin{aligned} J &\equiv \int_0^1 (1-t)t^{1/b-1}(1+t)^{-3-1/b}\, dt = \int_0^{1/2} \left(1 - \frac{u}{1-u}\right)(1-u)^2 u^{1/b-1} du \\ &= b\left(\frac{1}{2}\right)^{1/b} - \frac{3b}{b+1}\left(\frac{1}{2}\right)^{1+1/b} + \frac{2b}{2b+1}\left(\frac{1}{2}\right)^{2+1/b}. \end{aligned} \tag{3.47}$$

Finally, we combine (3.45)–(3.47) and obtain

$$C = \frac{2b(1-2b)}{I-J} = \frac{2(b+1)(2b+1)}{b}. \tag{3.48}$$

We have assumed that $b > 1$. However, (3.48) can be used to define C for all $b > 0$. For $0 < b \leqslant 1$ we interpret the sums in (3.41) in a generalized sense, using

$$\sum_{\ell=0}^{\infty} (-1)^\ell = \lim_{x\to 1^-} \sum_{\ell=0}^{\infty} (-x)^\ell = \frac{1}{2}, \quad \sum_{\ell=0}^{\infty} \ell(-1)^\ell = \lim_{x\to 1}\left[\frac{-x}{(1+x)^2}\right] = -\frac{1}{4},$$

and so on. This completes the analysis.

4 Conclusion

We have identified a new heavy traffic limit for the asymmetric shortest queue problem. Computing the approximate joint steady state queue length distribution led to a backward parabolic PDE, which we explicitly solved. Thus, for some queueing models, the heavy traffic (or diffusion) limit leads to solving parabolic PDEs (the present model, [16] and [17]), while for others we must solve elliptic PDEs (cf. [13]-[15]), which are generally more difficult. It would be interesting to find other models that lead to parabolic PDEs. Other areas for future investigation are (1) computing the transient probability distribution for the present model (we could then drop the condition that $b > 0$) and (2) extending the analysis to three or more queues.

Appendix A

We first evaluate the product in (2.13) as $(x, y) \to (0, 0)$. We begin by considering the sum

$$\sum_{n=1}^{\infty} \log(1 - \mathrm{e}^{-\delta n} \mathrm{e}^{a\delta}) \tag{A 1}$$

for $\delta \to 0^+$ with $a < 1$. We write (A 1) as

$$\lim_{N\to\infty} \left\{ \sum_{n=1}^{N} \log\left(\frac{1 - \mathrm{e}^{-\delta(n-a)}}{\delta(n-a)} \right) + N \log \delta + \log\left(\frac{\Gamma(N - a + 1)}{\Gamma(1-a)} \right) \right\} \tag{A 2}$$

where we have used $\sum_{n=1}^{N} \log(n-a) = \log\big[(1-a)(2-a)\cdots(N-a)\big]$. Applying the Euler-MacLaurin summation formula to the sum in (A 2) yields

$$\sum_{n=1}^{N} \log\left(\frac{1 - \mathrm{e}^{-\delta(n-a)}}{\delta(n-a)} \right) = \frac{1}{\delta} \int_{\delta}^{\delta N} \log\left(\frac{1 - e^{-(t-\delta a)}}{t - \delta a} \right) \mathrm{d}t \tag{A 3}$$
$$+ \frac{1}{2} \log\left(\frac{1 - e^{-\delta(N-a)}}{\delta(N-a)} \right) + O(\delta).$$

We can clearly choose N large enough so that $\delta N \to \infty$; then (A 3) becomes

$$\frac{1}{\delta} \int_0^{\infty} \log(1 - e^{-t}) dt - (N - a) \log\big[(N-a)\delta\big] + N - a - \frac{1}{2} \log \delta - \frac{1}{2} \log N + O(\delta, N^{-1}). \tag{A 4}$$

The above integral evaluates to $-\pi^2/6$. We then use (A 4) in (A 2) and expand $\log\big[\Gamma(N - a + 1)\big]$ by Stirling's formula as $N \to \infty$. Then we evaluate the limit in (A 2) and exponentiate (A 1) to obtain

$$\prod_{n=1}^{\infty} (1 - e^{-\delta n} e^{a\delta}) \sim \frac{\sqrt{2\pi} \delta^{a-1/2}}{\Gamma(1-a)} \exp\left(\frac{-\pi^2}{6\delta} \right), \quad \delta \to 0^+. \tag{A 5}$$

We apply (A 5) to the product in (2.13) with $\delta = 2y$; $a = 0$, $a = (x + y)/2y$ and $a = (y - x)/2y$. The three cases of a correspond to the three factors in the infinite product. Recalling that $\Gamma\big((1+u)/2\big)\Gamma\big((1-u)/2\big) = \pi \sec(\pi u/2)$, we find that as $x, y \to 0$ with

https://doi.org/10.1017/S0956792599003861 Published online by Cambridge University Press

$0 \leqslant x/y < 1$ the product in (2.13) becomes

$$2\sqrt{\pi}y^{-1/2}\exp\left(\frac{-\pi^2}{4y}\right)\cos\left(\frac{\pi x}{2y}\right).$$

Applying the operator $\partial_y^2 - \partial_x^2$ to the above expression and retaining only leading order terms yields (2.14). The result in (2.14) ceases to be valid for $y = x$, since the leading term vanishes. To obtain $P(y, y)$ as $y \to 0$ we write the product in (2.13) as

$$(1 - e^{x-y})\prod_{n=1}^{\infty}(1 - e^{-2ny})(1 - e^{-2ny}e^{x-y})(1 - e^{-2ny}e^{y-x}). \tag{A 6}$$

We again apply (A 5) to the product in (A 6) and note that when $y = x$,

$$(\partial_y^2 - \partial_x^2)(1 - e^{x-y})\prod_{n=1}^{\infty}(\cdots) \sim 2\partial_y\left\{\prod_{n=1}^{\infty}(1 - e^{-2ny})^3\right\}.$$

We thus obtain (2.15).

We next show how to obtain the corner behavior of P directly from (2.5)-(2.7). Since the normalization condition below (2.7) will not be used, this can only yield P up to an arbitrary multiplicative constant. We assume an expansion of the form

$$P(x, y) \sim e^{-a/y}y^{\nu}\left[f\left(\frac{x}{y}\right) + yg\left(\frac{x}{y}\right) + y^2h\left(\frac{x}{y}\right) + \cdots\right] \tag{A 7}$$

where a and ν are constants. Using the ansatz (A 7) in the PDE (2.5) and setting $z = x/y$, we obtain at the first two orders in y the ODEs

$$f''(z) + abf(z) = 0, \qquad 0 < z < 1 \tag{A 8}$$

and

$$g''(z) + abg(z) = b\left[zf''(z) - \nu f(z)\right] + (b-1)f'(z). \tag{A 9}$$

The boundary condition (2.6) yields

$$f'(0) = 0 \ \text{ and } \ g'(0) = (b-1)f(0) \tag{A 10}$$

while (2.7) gives

$$f(1) = 0 \ \text{ and } \ ag(1) + 2f'(1) = 0. \tag{A 11}$$

The problem for $f(z)$ in (A 8), (A 10) and (A 11) constitutes an elementary eigenvalue problem. We have $f(z) = (\text{const.})\cos(\sqrt{ab}\,z)$ and (A 11) then forces $a = (M + 1/2)^2\pi^2/b$; $M = 0, 1, \ldots$. But, if P is to be positive for $0 \leqslant x/y < 1$ we must choose $M = 0$. We thus let $f(z) = \cos(\pi z/2)$ and $a = \pi^2/(4b)$.

Now consider (A 9), together with the second equations in (A 10) and (A 11). We multiply (A 9) by $f(z)$ and integrate over $[0, 1]$. This yields the solvability condition for $g(z)$. Using the BC in (A 10) and (A 11) we find that

$$\nu = \frac{-7}{2} - \frac{1}{b} \tag{A 12}$$

and then $g(1) = 4b/\pi$. With (A 12) we have determined the algebraic factor in (A 7) and $g(1)$ yields the behavior of $P(y, y)$ as $y \to 0$. We have thus derived (2.16).

References

[1] Foschini, G. (1977) On heavy traffic diffusion analysis and dynamic routing in packet switched networks. In K. Chandy and J. Reiser (eds.), *Computer Performance*, North-Holland, 499–513.

[2] Foschini, G. & Salz, J. (1978) A basic dynamic routing problem and diffusion. *IEEE Trans. Commun.* **26**, 320–327.

[3] Gertsbakh, I. (1984) The shorter queue problem: a numerical study using the matrix-geometric solution. *Eur. J. Oper. Res.* **15**, 374–381.

[4] Flatto, L. & McKean, H. (1977) Two queues in parallel. *Commun. Pure Appl. Math.* **30**, 255–263.

[5] Cohen, J. W. & Boxma, O. J. (1983) *Boundary Value Problems in Queueing System Analysis*. North-Holland.

[6] Adan, I. J. B. F., Wessels, J. & Zijm, W. H. M. (1990) Analysis of the symmetric shortest queue problem. *Commun. Statist.-Stoch. Models.* **6**, 691–713.

[7] Adan, I. J. B. F., Wessels, J. & Zijm, W. H. M. (1991) Analysis of the asymmetric shortest queue problem. *Queueing Systems*, **8**, 1–58.

[8] Knessl, C., Matkowsky, B. J., Schuss, Z. & Tier, C. (1986) Two parallel queues with dynamic routing. *IEEE Trans. Commun.* **34**, 1170–1175.

[9] Kingman, J. F. C. (1961) Two similar queues in parallel. *Ann. Math. Statist.* **32**, 1314–1323.

[10] Knessl, C., Matkowsky, B. J., Schuss, Z. & Tier, C. (1987) Two parallel $M/G/1$ queues where arrivals join the system with the smaller buffer content. *IEEE Trans. Commun.* **35**, 1153–1158.

[11] Reiman, M. (1984) Some diffusion approximations with state space collapse. In A. V. Balakreshnan and M. Thomas (eds.), *Modelling and Performance Evaluation Methodology*. Springer-Verlag, 209–240.

[12] van Houtum, G. J. (1985) New approaches for multi-dimensional queueing systems. *PhD Thesis*, Eindhoven University of Technology.

[13] Newell, G. F. (1979) *Approximate Behavior of Tandem Queues*, Springer-Verlag.

[14] Knessl, C. & Tier, C. (1999) A diffusion model for two tandem queues with general renewal input. *Commun. Statist.-Stoch. Models*, **15**, 299–343.

[15] Foschini, G. J. (1982) Equilibria for diffusion models for pairs of communicating computers-symmetric case. *IEEE Trans. Inform. Theory*, **28**, 273–284.

[16] Choi, D. I., Knessl, C. & Tier, C. A dynamic priority queue model for simultaneous service of voice and data traffic. Submitted.

[17] Knessl, C. & Morrison, J. A. (1991) Heavy traffic analysis of a data-handling system with many sources. *SIAM J. Appl. Math.* **51**, 187–213.

[18] Turner, S. R. E. A join the shorter queue model in heavy traffic, http://www.statslab.cam.ac.uk/~sret1/work/