Queueing Syst (2007) 57: 19–28
DOI 10.1007/s11134-007-9043-8

Processor sharing for two queues with vastly different rates

John A. Morrison

Received: 4 August 2006 / Revised: 9 February 2007 / Published online: 17 August 2007

Abstract We consider a 2-class queueing system, operating under a generalized processor-sharing discipline, in an asymptotic regime where the arrival and service rates of the two classes are vastly different. We use regular and singular perturbation analyses in a small parameter measuring this difference in rates. It is assumed that the system is stable, and not close to instability. Three different regimes are analyzed, corresponding to an underloaded, an overloaded and a critically loaded fast queue, respectively. In the first two regimes the lowest order approximation to the joint stationary distribution of the queue lengths is derived. For a critically loaded fast queue only the mean queue lengths are investigated, and the asymptotic matching, to lowest order, with the results for an underloaded and an overloaded fast queue is established.

Keywords Asymptotics · Matching · Processor sharing · Singular perturbations

Mathematics Subject Classification (2000) 60K30 · 90B22

1 Introduction and summary

In this paper we consider a 2-class queueing system, operating under a generalized processor-sharing discipline, in an asymptotic regime where the dynamics of the two classes occur on vastly different time scales. This was motivated by the work of van Kessel et al. [9]. The disparity in service characteristics reflects the extreme heterogeneity in job sizes observed in the Internet, with a vast majority of short jobs ('mice') and a tiny fraction of large jobs ('elephants'). The jobs arrive as Poisson processes of rates λ (the 'fast' class) and $\epsilon\lambda$ (the 'slow' class), where $0 < \epsilon \ll 1$. The fast and slow jobs have exponentially distributed service requirements with parameters 1 and $\epsilon\nu$, respectively. The server works at unit rate, and if neither queue is empty it devotes fractions $1-\gamma$ and γ of its effort to the fast and slow jobs, respectively, where $0 \le \gamma \le 1$. If one queue is empty the server works at unit rate on the other queue. It is assumed that the system is stable, and not close to instability.

This problem has been analyzed, without ϵ being necessarily small, in a pioneering paper by Fayolle and Iasnogorodski [6], and more recently by Guillemin and Pinchon [8]. They consider the generating function $F(x, y)$ for the joint stationary distribution of the number of jobs in the two queues, which leads to a functional equation involving the unknown functions $F(x, 0)$ and $F(0, y)$. The determination of these boundary functions is reduced, by means of a complicated analysis, to a Dirichlet problem on a circle. Related references are given in [6, 8].

In this paper, we use perturbation analyses in the small parameter ϵ. Three different regimes are analyzed. First, if $0 \le \gamma < 1-\lambda$, so that the fast queue is underloaded, we use a direct perturbation analysis of the balance equations for the joint stationary distribution of the queue lengths. Such an approach has been adopted by Altman et al. [2] for denumerable Markov chains, who establish rigorous results, and provide expressions for the higher-order terms in the asymptotic expansion. Asymptotically, to lowest order in ϵ, it turns out that for $0 \le \gamma < 1-\lambda$ the mean number of jobs in the slow queue is independent of γ, but the mean number of jobs in the fast queue, and hence the mean waiting time,

J.A. Morrison (✉)
Consultant, Alcatel-Lucent, Bell Laboratories, Murray Hill, NJ 07974, USA
e-mail: johnmorrison@alcatel-lucent.com

increases with γ. Hence there is no significant advantage to taking $0 < \gamma < 1 - \lambda$. An intuitive interpretation of the results, due to Borst [3], is presented. We have verified that the results in [6] (after correction of some misprints) and [8] lead to our lowest order asymptotic results. However, this took significant effort, and the direct approach in this paper is much simpler when $0 < \epsilon \ll 1$.

We have not attempted to investigate the conditions in [2] that establish, for singularly perturbed Markov chains, the existence of a unique solution which is an analytic function of ϵ, since it appears difficult to do so. However, for the problem under consideration, the results in [6, 8] give integral representations with limits and integrands which are analytic functions of ϵ.

We next consider the regime $1 - \lambda < \gamma \le 1$, so that the fast queue is overloaded when the slow queue is not empty, and it needs the idle periods of the slow queue to recover stability. The motivation for doing so is to allow for some priority for the slow jobs. It now turns out, asymptotically to lowest order in ϵ, that the mean number of jobs in the slow queue decreases as γ increases with $1 - \lambda < \gamma \le 1$, but the mean number of jobs in the fast queue, and hence the mean waiting time, is $O(1/\epsilon)$. We perform a singular perturbation analysis of the functional equation for $F(x, y)$ when $x = 1 - \epsilon\xi$ with $\xi = 0(1)$. We have verified, after significant effort, that the results in [8] lead to our lowest order approximations to $F(0, y)$ for $|y| < \sqrt{\gamma\nu/\lambda}$, $F(x, 0)$ for $|x| < \sqrt{(1-\gamma)/\lambda}$ and $F(1-\epsilon\xi, 0)$. (An explicit expression for $F(1-\epsilon\xi, 0)$ is not given in [6].) Again, the results in [8] give integral representations with limits and integrands that are analytic functions of ϵ. We derive asymptotic approximations to the joint stationary distribution of the number of jobs in the two queues.

Finally, we consider the regime $\gamma = 1 - \lambda + \sqrt{\epsilon}\delta$, where $\delta = O(1)$ may have either sign, so that the fast queue is critically loaded. We restrict our attention to the mean queue lengths. In particular, we are interested in how the mean fast queue length changes from $O(1)$ to $O(1/\epsilon)$ as δ changes from a large negative value to a large positive one. We now consider the functional equation for $F(x, y)$ when $x = 1 - \sqrt{\epsilon}u$ with $u = O(1)$. In order to calculate an asymptotic approximation to the mean number in the fast queue, we need to evaluate the correction term $f_1(0, y)$ in the expansion

$$F(0, y) = (\nu - \lambda - \lambda\nu)/\nu + \sqrt{\epsilon} f_1(0, y) + O(\epsilon). \tag{1.1}$$

The derivation of this boundary function requires the solution of a Dirichlet problem, so we use the result in [6]. We show, with somewhat more effort, that the result in [8] leads to an equivalent expression for $f_1(0, y)$. We establish the matching, i.e., agreement to lowest order, of the asymptotic approximation to the mean number in the fast queue with the results for an underloaded and for an overloaded fast queue, by analyzing the regimes $\sqrt{\epsilon}|\delta| \ll 1$, and $-\delta \gg 1$ and $\delta \gg 1$, respectively.

The remainder of the paper is organized as follows. In Sect. 2 we formulate the problem. In Sects. 3, 4 and 5 we analyze the underloaded, overloaded and critically loaded fast queue, respectively. The derivation of the correction term in (1.1) is given in Appendix 1 using the result in [6], and in Appendix 2 using the result in [8].

2 Formulation

We consider two parallel infinite capacity queues for different traffic classes, operating under a generalized processor-sharing discipline. The jobs arrive as Poisson processes of rates λ (the 'fast' class) and $\epsilon\lambda$ (the 'slow' class), where $0 < \epsilon \ll 1$. The fast and slow jobs have exponentially distributed service requirements with parameters 1 and $\epsilon\nu$, respectively. The server works at unit rate, and if neither queue is empty it devotes fractions $1-\gamma$ and γ to the fast and slow jobs, respectively, where $0 \le \gamma \le 1$. The corresponding instantaneous service rates are $1-\gamma$ and $\epsilon\gamma\nu$. If one queue is empty the server works at unit rate on the other queue, so it is work conserving, and the stability condition is

$$\lambda\left(1 + \frac{1}{\nu}\right) < 1, \tag{2.1}$$

which we assume holds.

Let $p(m, n)$ denote the stationary probability that there are m jobs in the fast queue and n jobs in the slow queue. Let $I(\cdot)$ be the indicator function. Then the balance equations satisfied by $p(m, n)$ are

$$\begin{aligned}
&\{\lambda + \epsilon\lambda + I(m \ge 1)[1 - \gamma I(n \ge 1)] \\
&\quad + \epsilon\nu I(n \ge 1)[1 - (1-\gamma) I(m \ge 1)]\} p(m, n) \\
&\qquad = \lambda I(m \ge 1) p(m-1, n) + \epsilon\lambda I(n \ge 1) p(m, n-1) \\
&\qquad\quad + [1 - \gamma I(n \ge 1)] p(m+1, n) \\
&\qquad\quad + \epsilon\nu[1 - (1-\gamma) I(m \ge 1)] p(m, n+1).
\end{aligned} \tag{2.2}$$

The normalization condition is

$$\sum_{m=0}^{\infty}\sum_{n=0}^{\infty} p(m, n) = 1. \tag{2.3}$$

The mean numbers of jobs in the fast and slow queues are

$$\begin{aligned}
E(m) &= \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} m p(m, n), \\
E(n) &= \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} n p(m, n).
\end{aligned} \tag{2.4}$$

From Little's result, the corresponding mean waiting times are $E(m)/\lambda$ and $E(n)/(\epsilon\lambda)$. From [6], the conservation law of Kleinrock implies that

$$E(m) + \frac{E(n)}{\epsilon\nu} = \frac{\lambda\nu}{(\nu-\lambda-\lambda\nu)}\left(1+\frac{1}{\epsilon\nu^2}\right). \tag{2.5}$$

This relationship between the mean numbers of jobs is useful both as a check on the analysis, and for understanding the results. In particular, if $0 \le \gamma < 1-\lambda$, so that the fast queue is underloaded, $E(n) = \lambda/(\nu-\lambda-\lambda\nu) + O(\epsilon)$ and $E(m) = O(1)$. If $1-\lambda < \gamma \le 1$, so that the fast queue is overloaded when the slow queue is not empty, $E(n) = O(1)$ and $E(m) = O(1/\epsilon)$. In the transition region where $\gamma + \lambda - 1 = O(\sqrt{\epsilon})$, so that the fast queue is critically loaded, $E(n) = \lambda/(\nu-\lambda-\lambda\nu) + O(\sqrt{\epsilon})$ and $E(m) = O(1/\sqrt{\epsilon})$.

If we introduce the generating function, as in [6, 8],

$$F(x,y) = \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} p(m,n)x^m y^n, \quad |x|\le 1,\ |y|\le 1, \tag{2.6}$$

then, from (2.2)

$$\begin{aligned}&\left[\lambda(1-x) + (1-\gamma)\left(1-\frac{1}{x}\right)\right.\\ &\quad\left. + \epsilon\lambda(1-y) + \epsilon\gamma\nu\left(1-\frac{1}{y}\right)\right]F(x,y)\\ &\quad = \left[(1-\gamma)\left(1-\frac{1}{x}\right) - \epsilon\nu(1-\gamma)\left(1-\frac{1}{y}\right)\right]F(0,y)\\ &\qquad + \left[\epsilon\gamma\nu\left(1-\frac{1}{y}\right) - \gamma\left(1-\frac{1}{x}\right)\right]F(x,0)\\ &\qquad + \left[\gamma\left(1-\frac{1}{x}\right) + \epsilon\nu(1-\gamma)\left(1-\frac{1}{y}\right)\right]F(0,0).\end{aligned} \tag{2.7}$$

Here

$$F(0,0) = 1-\lambda\left(1+\frac{1}{\nu}\right), \tag{2.8}$$

i.e., $1-\rho$, where ρ is the load.

3 Underloaded fast queue

We here consider the case in which the fast queue is underloaded, corresponding to $\lambda < 1-\gamma$. We seek a solution of (2.2) and (2.3) in powers of ϵ,

$$p(m,n) = p_0(m,n) + \epsilon p_1(m,n) + O(\epsilon^2). \tag{3.1}$$

It follows from (2.2) that

$$\begin{aligned}&\{\lambda + I(m\ge 1)[1-\gamma I(n\ge 1)]\}p_0(m,n)\\ &\quad = \lambda I(m\ge 1)p_0(m-1,n)\\ &\qquad + [1-\gamma I(n\ge 1)]p_0(m+1,n).\end{aligned} \tag{3.2}$$

Hence

$$p_0(m,n) = \left\{\frac{\lambda}{[1-\gamma I(n\ge 1)]}\right\}^m p_0(0,n), \quad m\ge 0,\ n\ge 0. \tag{3.3}$$

If we sum (2.2) on m, and expand in powers of ϵ, we obtain

$$\begin{aligned}&\sum_{m=0}^{\infty}\{\lambda + \nu I(n\ge 1)[1-(1-\gamma)I(m\ge 1)]\}p_0(m,n)\\ &\quad = \lambda I(n\ge 1)\sum_{m=0}^{\infty} p_0(m,n-1)\\ &\qquad + \nu\sum_{m=0}^{\infty}[1-(1-\gamma)I(m\ge 1)]p_0(m,n+1).\end{aligned} \tag{3.4}$$

It follows that

$$\begin{aligned}&\lambda\sum_{m=0}^{\infty} p_0(m,n)\\ &\quad = \nu\sum_{m=0}^{\infty}[1-(1-\gamma)I(m\ge 1)]p_0(m,n+1), \quad n\ge 0.\end{aligned} \tag{3.5}$$

From (3.3), the lowest order asymptotic approximation to the stationary probability that there are n jobs in the slow queue is

$$\begin{aligned}P(n) &= \sum_{m=0}^{\infty} p_0(m,n)\\ &= \frac{[1-\gamma I(n\ge 1)]}{[1-\gamma I(n\ge 1)-\lambda]}p_0(0,n), \quad n\ge 0,\end{aligned} \tag{3.6}$$

and

$$\begin{aligned}&\sum_{m=0}^{\infty}[1-(1-\gamma)I(m\ge 1)]p_0(m,n+1)\\ &\quad = \frac{(1-\gamma)(1-\lambda)}{(1-\gamma-\lambda)}p_0(0,n+1) = (1-\lambda)P(n+1).\end{aligned} \tag{3.7}$$

From (3.5–3.7) we obtain

$$\lambda P(n) = \nu(1-\lambda)P(n+1), \quad n\ge 0. \tag{3.8}$$

But, from (2.3) and (3.1),

$$\sum_{n=0}^{\infty} P(n) = 1. \tag{3.9}$$

Hence,

$$P(n) = \frac{(\nu-\lambda-\lambda\nu)}{\nu(1-\lambda)}\left[\frac{\lambda}{\nu(1-\lambda)}\right]^n, \quad n\ge 0. \tag{3.10}$$

An expression for $p_0(m,n)$ follows from (3.3), (3.6) and (3.10). The lowest order asymptotic approximations to the mean numbers of jobs in the fast and slow queues are

$$E(m) \sim \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} m p_0(m,n) = \frac{\lambda}{\nu(1-\lambda)}\left[\frac{(\nu-\lambda-\lambda\nu)}{(1-\lambda)} + \frac{\lambda}{(1-\gamma-\lambda)}\right], \quad (3.11)$$

and

$$E(n) \sim \sum_{n=0}^{\infty} n P(n) = \frac{\lambda}{(\nu-\lambda-\lambda\nu)}. \quad (3.12)$$

We note that the approximation for $E(m)$ increases with γ, while that for $E(n)$ is independent of γ.

An intuitive explanation of the above results has been given by Borst [3], in a more general setting where the arrival and service rates are state-dependent. We repeat his explanations here for the problem under consideration. As the value of ϵ becomes smaller, the dynamics of the fast queue evolve on an increasingly fast time scale compared to the slow queue. In the limit $\epsilon \to 0$, a complete time scale decomposition occurs, and the dynamics of the fast queue will completely average out on the relevant time scale for the slow queue. Thus the fast jobs take away a constant service rate equal to the traffic intensity λ, and the slow jobs behave as in an isolated system with service rate $1-\lambda$. Hence, the marginal distribution of the number of jobs in the slow queue is geometric with parametric $\lambda/(\nu(1-\lambda))$, as in (3.10). In particular, the marginal distribution is independent of γ.

Conversely, the dynamics of the slow jobs will nearly vanish as $\epsilon \to 0$ on the relevant time scale for the fast jobs, which will reach some sort of statistical equilibrium for a given number of slow jobs. Thus, when there are n jobs in the slow queue, the fast jobs behave as in an isolated system with n permanent customers and service rate $1-\gamma I(n \geq 1)$. The conditional number of fast jobs then follows the distribution of a birth-death process with birth rates λ and death rates $1-\gamma I(n\geq 1)$, as in (3.3).

If we expand the generating function in (2.6) in powers of ϵ,

$$F(x,y) = F_0(x,y) + \epsilon F_1(x,y) + O(\epsilon^2), \quad (3.13)$$

we obtain from (2.7) and (2.8),

$$(1-\gamma-\lambda x)F_0(x,y) = (1-\gamma)F_0(0,y) - \gamma F_0(x,0) + \gamma F_0(0,0), \quad (3.14)$$

and

$$F_0(0,0) = 1-\lambda\left(1+\frac{1}{\nu}\right). \quad (3.15)$$

It follows that

$$F_0(x,0) = \frac{(\nu-\lambda-\lambda\nu)}{\nu(1-\lambda x)}. \quad (3.16)$$

This is consistent with (3.3), with $n=0$.

There are some unfortunate misprints in [6]. Specifically, the factor $F(0,0)\sqrt{\mu_1/\lambda_1}$ should appear in the numerator of the expression for $u(\rho)$ in (6.3), and the factor $F(0,0)\sqrt{\mu_2/\lambda_2}$ should appear in the numerator of the expression for $v(\theta)$ in (6.6). Also, there should be a factor 2 multiplying the integrals in (6.4) and (6.5). The subscripts under the square roots in the second expression for $K(\theta)$ at the top of page 340 should be 1 instead of 2. The last definition below (1.2) should be $q=\mu_2-\mu_2^*$. With these corrections we obtained, after some straightforward but tedious calculations, (3.16) and

$$F_0(0,y) = \left[1-\lambda\left(1+\frac{1}{\nu}\right)\right] \times \left\{1 + \frac{\lambda(1-\gamma-\lambda)y}{(1-\gamma)(1-\lambda)[\nu(1-\lambda)-\lambda y]}\right\}. \quad (3.17)$$

The expressions for $F(x,0)$ and $F(0,y)$ in [8] are in a more compact form. Surprisingly, however, the evaluation of $F_0(x,0)$ and $F_0(0,y)$ took significantly more effort, but it did lead to (3.16) and (3.17). It follows from (3.14–3.17) that

$$F_0(x,y) = \left[1-\lambda\left(1+\frac{1}{\nu}\right)\right] \times \left\{\frac{1}{(1-\lambda x)} + \frac{\lambda(1-\gamma-\lambda)y}{(1-\lambda)(1-\gamma-\lambda x)[\nu(1-\lambda)-\lambda y]}\right\}. \quad (3.18)$$

Then, from (2.6), (3.1) and (3.13), we obtain an expression for $p_0(m,n)$ which agrees with (3.3), where $p_0(0,n)$ is given by (3.6) and (3.10). While it was satisfying to obtain this agreement with the results in [6, 8], it is evident that the direct approach in this paper is much simpler when $0<\epsilon\ll 1$.

4 Overloaded fast queue

We now consider the case in which the fast queue is overloaded when the slow queue is not empty, corresponding to $\lambda > 1-\gamma$. It follows from (3.3) and (3.15) that

$$p_0(m,0) = \lambda^m p_0(0,0) = \lambda^m\left[1-\lambda\left(1+\frac{1}{\nu}\right)\right], \quad m\geq 0, \quad (4.1)$$

and that $p_0(m,n)=0$ for $m\ge 0$, $n\ge 1$. Hence, from (2.6),

$$F_0(x,y)=\frac{(\nu-\lambda-\lambda\nu)}{\nu(1-\lambda x)}. \tag{4.2}$$

This cannot hold for $x=1$, since $F_0(1,1)<1$. Consequently, we consider $x=1-\epsilon\xi$, where $\xi=0(1)$, and let

$$F(1-\epsilon\xi,y)=\Phi_0(\xi,y)+\epsilon\Phi_1(\xi,y)+O(\epsilon^2). \tag{4.3}$$

The determination of $\Phi_0(\xi,y)$ will enable us to obtain an asymptotic approximation to $p(\zeta/\epsilon,n)$. From (2.7) and (2.8), since $F_0(0,y)=(\nu-\lambda-\lambda\nu)/\nu$,

$$\begin{aligned}&\left[(\gamma-1+\lambda)\xi+(\gamma\nu-\lambda y)\left(1-\frac{1}{y}\right)\right]\Phi_0(\xi,y)\\&\quad=\gamma\left[\nu\left(1-\frac{1}{y}\right)+\xi\right]\Phi_0(\xi,0)-\left[1-\lambda\left(1+\frac{1}{\nu}\right)\right]\xi.\end{aligned} \tag{4.4}$$

In particular, since $\Phi_0(0,1)=1$,

$$\Phi_0(0,y)=\frac{(\gamma\nu-\lambda)}{(\gamma\nu-\lambda y)}. \tag{4.5}$$

Note that $\gamma\nu>\nu(1-\lambda)>\lambda$, from (2.1). Hence,

$$E(n)\sim\frac{\partial\Phi_0}{\partial y}(0,1)=\frac{\lambda}{(\gamma\nu-\lambda)}, \tag{4.6}$$

which decreases as $\gamma>1-\lambda$ increases, to the exact value $\lambda/(\nu-\lambda)$ when $\gamma=1$. From (2.5) it follows that

$$E(m)\sim\frac{\lambda(\gamma-1+\lambda)}{\epsilon(\nu-\lambda-\lambda\nu)(\gamma\nu-\lambda)}, \tag{4.7}$$

so that the mean number of jobs in the fast queue, and the corresponding mean waiting time, are $O(1/\epsilon)$.

We will determine $\Phi_0(\xi,0)$ by requiring $\Phi_0(\xi,y)$ to be analytic for $|y|<1$. We define

$$\omega=\gamma-1+\lambda>0, \tag{4.8}$$

and

$$\Delta(y)=\lambda y^2-(\lambda+\gamma\nu+\omega\xi)y+\gamma\nu, \tag{4.9}$$

and note that the left-hand side of (4.4) vanishes if $\Delta(y)=0$. We next define

$$\begin{aligned}\eta(\xi)=\frac{1}{2\lambda}&\left[(\lambda+\gamma\nu+\omega\xi)\right.\\&\left.-\sqrt{(\lambda+\gamma\nu+\omega\xi)^2-4\lambda\gamma\nu}\right],\end{aligned} \tag{4.10}$$

where the positive square root is to be taken for $\xi\ge 0$. Then $\Delta(\eta(\xi))=0$ and $\Delta(\gamma\nu/\lambda\eta(\xi))=0$. If $\eta(\xi)=e^{i\theta}$, where θ is real, then $|\eta(\xi)|=1$ and, from (4.9)

$$\lambda e^{2i\theta}-(\lambda+\gamma\nu+\omega\xi)e^{i\theta}+\gamma\nu=0. \tag{4.11}$$

Hence,

$$\omega\xi=-(\lambda+\gamma\nu)(1-\cos\theta)+i(\lambda-\gamma\nu)\sin\theta, \tag{4.12}$$

so that $\mathrm{Re}\,\xi\le 0$, since $\omega>0$. It follows that $|\eta(\xi)|<1$ and $\gamma\nu/\lambda|\eta|(\xi)|>1$ for $\mathrm{Re}\,\xi>0$, and so $\Delta(y)=0$ has the unique solution $y=\eta(\xi)$ in $|y|<1$.

From (4.4), and the analyticity of $\Phi_0(\xi,y)$ for $|y|<1$, we obtain

$$\gamma\left\{\nu\left[1-\frac{1}{\eta(\xi)}\right]+\xi\right\}\Phi_0(\xi,0)=\left[1-\lambda\left(1+\frac{1}{\nu}\right)\right]\xi. \tag{4.13}$$

Although the analyticity of the asymptotic approximation is not ensured by that of the exact solution, (4.13) shows that its requirement determines $\Phi_0(\xi,0)$. Since $\Delta(\eta(\xi))=0$, it follows from (4.8) and (4.9) that

$$\begin{aligned}&\left[(\xi+\nu)-\frac{\nu}{\eta(\xi)}\right][\gamma(\xi+\nu)-\lambda\eta(\xi)]\\&\quad=\xi\left[(1-\lambda)\xi+(\nu-\lambda-\lambda\nu)\right].\end{aligned} \tag{4.14}$$

Hence, from (4.13),

$$\Phi_0(\xi,0)=\frac{(\nu-\lambda-\lambda\nu)[\gamma(\xi+\nu)-\lambda\eta(\xi)]}{\gamma\nu[(1-\lambda)\xi+(\nu-\lambda-\lambda\nu)]}. \tag{4.15}$$

From (22) in [8], we have verified, with some effort, that $F(x,0)\sim F_0(x,0)$ for $|x|<\sqrt{(1-\gamma)/\lambda}$ and $F(0,y)\sim F_0(0,y)$ for $|y|<\sqrt{\gamma\nu/\lambda}$, where $F_0(x,y)$ is given by (4.2). As noted before, $\gamma\nu>\lambda$. From (23) in [8], after significantly more effort, we have verified that $\Phi(\xi,0)\sim\Phi_0(\xi,0)$, as given by (4.15). Since $\Delta(\eta(\xi))=0$ and $\Delta(\gamma\nu/\lambda\eta(\xi))=0$, (4.8) and (4.9) imply that

$$\begin{aligned}&(\gamma\nu-\lambda y)(1-y)-(\gamma-1+\lambda)\xi y\\&\quad=\lambda[y-\eta(\xi)]\left[y-\frac{\gamma\nu}{\lambda\eta(\xi)}\right].\end{aligned} \tag{4.16}$$

Hence, from (4.4) and (4.13), we obtain

$$\Phi_0(\xi,y)=\frac{\gamma\nu\Phi_0(\xi,0)}{[\gamma\nu-\lambda\eta(\xi)y]}. \tag{4.17}$$

In particular, from (4.15) and (4.17),

$$\begin{aligned}\Phi_0(\xi,1)=&\frac{(\nu-\lambda-\lambda\nu)}{[(1-\lambda)\xi+(\nu-\lambda-\lambda\nu)]}\\&\times\left\{1+\frac{\gamma\xi}{[\gamma\nu-\lambda\eta(\xi)]}\right\}.\end{aligned} \tag{4.18}$$

But,

$$E(m)\sim-\frac{1}{\epsilon}\frac{\partial\Phi_0}{\partial\xi}(0,1), \tag{4.19}$$

and the approximation in (4.7) readily follows from (4.18), since $\eta(0)=1$. Generally, from (4.17),

$$\Phi_0(\xi, y) = \Phi_0(\xi, 0)\sum_{n=0}^{\infty}\left[\frac{\lambda\eta(\xi)y}{\gamma\nu}\right]^n. \tag{4.20}$$

But, from (4.10) and (4.15),

$$\lim_{\xi\to\infty}\eta(\xi)=0, \qquad \lim_{\xi\to\infty}\Phi_0(\xi,0)=\frac{(\nu-\lambda-\lambda\nu)}{\nu(1-\lambda)}. \tag{4.21}$$

Hence, from (4.2), $F_0(1-\epsilon\xi, y)$ matches with $F_0(x,y)$ for $\epsilon\xi \ll 1$, $\xi \gg 1$ and we form a composite asymptotic approximation by adding the two approximations and subtracting their common part,

$$F(x,y) \sim \frac{(\nu-\lambda-\lambda\nu)}{\nu(1-\lambda x)} + \Phi_0\left(\frac{1-x}{\epsilon}, y\right) - \frac{(\nu-\lambda-\lambda\nu)}{\nu(1-\lambda)}. \tag{4.22}$$

For $\epsilon m = \zeta = O(1)$,

$$p\left(\frac{\zeta}{\epsilon}, n\right) = \epsilon\pi_n(\zeta) = \epsilon[\pi_n^{(0)}(\zeta) + \epsilon\pi_n^{(1)}(\zeta) + O(\epsilon^2)], \tag{4.23}$$

and, from the Euler-Maclaurin summation formula [1]

$$\sum_{m=0}^{\infty}\sum_{n=0}^{\infty} p(m,n)(1-\epsilon\xi)^m y^n - \frac{(\nu-\lambda-\lambda\nu)}{\nu[1-\lambda(1-\epsilon\xi)]} \sim \sum_{n=0}^{\infty}\int_0^\infty e^{-\xi\zeta}\pi_n^{(0)}(\zeta)d\zeta y^n = \Phi_0(\xi,y) - \frac{(\nu-\lambda-\lambda\nu)}{\nu(1-\lambda)}. \tag{4.24}$$

Hence, from (4.20),

$$\int_0^\infty e^{-\xi\zeta}\pi_0^{(0)}(\zeta)d\zeta = \Phi_0(\xi,0) - \frac{(\nu-\lambda-\lambda\nu)}{\nu(1-\lambda)}, \tag{4.25}$$

and

$$\int_0^\infty e^{-\xi\zeta}\pi_n^{(0)}(\zeta)d\zeta = \Phi_0(\xi,0)\left[\frac{\lambda\eta(\xi)}{\gamma\nu}\right]^n, \quad n\ge 1. \tag{4.26}$$

Also, from (4.15),

$$\Phi_0(\xi,0) = \frac{\lambda(\nu-\lambda-\lambda\nu)[\gamma-(1-\lambda)\eta(\xi)]}{\gamma\nu(1-\lambda)[(1-\lambda)\xi+(\nu-\lambda-\lambda\nu)]} + \frac{(\nu-\lambda-\lambda\nu)}{\nu(1-\lambda)}. \tag{4.27}$$

It remains to invert the Laplace transforms in (4.25) and (4.26), to obtain expressions for $\pi_n^{(0)}(\zeta)$, $n\ge 0$.

From (4.10), we have

$$\eta(\xi) = 2\gamma\nu\Big[(\lambda+\gamma\nu+\omega\xi) + \sqrt{(\lambda+\gamma\nu+\omega\xi)^2 - 4\lambda\gamma\nu}\Big]^{-1}. \tag{4.28}$$

But [5], for $l=1,2,\ldots$,

$$\left(p+\sqrt{p^2-a^2}\right)^{-l} = \frac{l}{a^l}\int_0^\infty e^{-pt} I_l(at)\frac{dt}{t}, \tag{4.29}$$

where $I_l()$ is a modified Bessel function. Hence,

$$\left[\frac{\lambda\eta(\xi)}{\gamma\nu}\right]^l = l\left(\frac{\lambda}{\gamma\nu}\right)^{l/2} \times \int_0^\infty e^{-\xi\zeta}e^{-(\lambda+\gamma\nu)\zeta/\omega} I_l(2\sqrt{\lambda\gamma\nu}\zeta/\omega)\frac{d\zeta}{\zeta}, \tag{4.30}$$

for $l=1,2,\ldots$. Also,

$$[(1-\lambda)\xi + (\nu-\lambda-\lambda\nu)]^{-1} = \frac{1}{(1-\lambda)}\int_0^\infty e^{-\xi\zeta}\exp\left[-\frac{(\nu-\lambda-\lambda\nu)}{(1-\lambda)}\zeta\right]d\zeta. \tag{4.31}$$

We define the convolution integral

$$V_l(\zeta) = l\left(\frac{\lambda}{\gamma\nu}\right)^{l/2}\int_0^\zeta \exp\left[-\frac{(\nu-\lambda-\lambda\nu)}{(1-\lambda)}(\zeta-z)\right] \times e^{-(\lambda+\gamma\nu)z/\omega} I_l(2\sqrt{\lambda\gamma\nu}z/\omega)\frac{dz}{z}. \tag{4.32}$$

Then, for $l=1,2,\ldots$,

$$\left[\frac{\lambda\eta(\xi)}{\gamma\nu}\right]^l [(1-\lambda)\xi+(\nu-\lambda-\lambda\nu)]^{-1} = \frac{1}{(1-\lambda)}\int_0^\infty e^{-\xi\zeta}V_l(\zeta)d\zeta. \tag{4.33}$$

It follows, from (4.25–4.27), (4.30) and (4.32), that

$$\pi_0^{(0)}(\zeta) = \frac{(\nu-\lambda-\lambda\nu)}{\nu(1-\lambda)} \times \left\{\frac{\lambda}{(1-\lambda)}\exp\left[-\frac{(\nu-\lambda-\lambda\nu)}{(1-\lambda)}\zeta\right] - \nu V_1(\zeta)\right\}, \tag{4.34}$$

and

$$\pi_n^{(0)}(\zeta) = \frac{(\nu-\lambda-\lambda\nu)}{\nu(1-\lambda)}\left[\frac{\lambda}{(1-\lambda)}V_n(\zeta) - \nu V_{n+1}(\zeta) + \frac{n}{\zeta}\left(\frac{\lambda}{\gamma\nu}\right)^{n/2} e^{-(\lambda+\gamma\nu)\zeta/\omega} I_n(2\sqrt{\lambda\gamma\nu}\zeta/\omega)\right], \quad n\ge 1. \tag{4.35}$$

From (4.23), we have the asymptotic approximation to the joint stationary probabilities

$$p(m,n) \sim \epsilon\pi_n^{(0)}(\epsilon m), \quad m = O(1/\epsilon),\ n \geq 0. \tag{4.36}$$

5 Critically loaded fast queue

Finally we consider the case in which the fast queue is critically loaded, corresponding to

$$\gamma = 1 - \lambda + \sqrt{\epsilon}\,\delta, \tag{5.1}$$

where $\delta = O(1)$ may have either sign. We will restrict our attention to the calculation of the mean queue lengths. In particular, we are interested in how the mean fast queue length changes from $O(1)$ to $O(1/\epsilon)$ as δ changes from a large negative value to a large positive one. We now expand in powers of $\sqrt{\epsilon}$, and let

$$F(0,y) = f_0(0,y) + \sqrt{\epsilon}\, f_1(0,y) + O(\epsilon), \tag{5.2}$$

and, with $x = 1 - \sqrt{\epsilon}u$, where $u = O(1)$,

$$F(1-\sqrt{\epsilon}u, y) = \Psi_0(u,y) + \sqrt{\epsilon}\Psi_1(u,y) + O(\epsilon). \tag{5.3}$$

Then, if we equate terms of $O(\sqrt{\epsilon})$ in (2.7), we obtain

$$\lambda f_0(0,y) = (1-\lambda)[\Psi_0(u,0) - f_0(0,0)], \quad u \neq 0. \tag{5.4}$$

It follows that

$$\begin{aligned}\Psi_0(u,0) &= \frac{f_0(0,0)}{(1-\lambda)},\\ f_0(0,y) &= f_0(0,0) = (\nu - \lambda - \lambda\nu)/\nu,\end{aligned} \tag{5.5}$$

from (2.8).

If we set $x = 1$ in (2.7) we then obtain

$$\begin{aligned}&[(1-\lambda+\sqrt{\epsilon}\delta) - \lambda y/\nu][\Psi_0(0,y) + \sqrt{\epsilon}\,\Psi_1(0,y) + O(\epsilon)]\\ &\quad = (1-\lambda+\sqrt{\epsilon}\delta)[\Psi_0(0,0) + \sqrt{\epsilon}\Psi_1(0,0) + O(\epsilon)]\\ &\qquad + (\lambda - \sqrt{\epsilon}\delta)\{\sqrt{\epsilon}[f_1(0,0) - f_1(0,y)] + O(\epsilon)\}.\end{aligned} \tag{5.6}$$

Hence

$$\begin{aligned}&(1-\lambda-\lambda y/\nu)\Psi_0(0,y)\\ &\quad = (1-\lambda)\Psi_0(0,0) = (\nu-\lambda-\lambda\nu)/\nu,\end{aligned} \tag{5.7}$$

from (5.5), and

$$\begin{aligned}&(1-\lambda-\lambda y/\nu)\Psi_1(0,y)\\ &\quad = (1-\lambda)\Psi_1(0,0) + \delta[\Psi_0(0,0) - \Psi_0(0,y)]\\ &\qquad + \lambda[f_1(0,0) - f_1(0,y)].\end{aligned} \tag{5.8}$$

But $\Psi_1(0,1) = 0$, since $F(1,1) = 1$. It follows from (5.7) and (5.8) that

$$\begin{aligned}\Psi_1(0,y) = \frac{\lambda\nu}{(\nu-\lambda\nu-\lambda y)}\bigg\{&\frac{\delta(1-y)}{(\nu-\lambda\nu-\lambda y)}\\ &+ f_1(0,1) - f_1(0,y)\bigg\}.\end{aligned} \tag{5.9}$$

From (5.7),

$$\frac{\partial\Psi_0}{\partial y}(0,1) = \frac{\lambda}{(\nu-\lambda-\lambda\nu)}. \tag{5.10}$$

Hence,

$$E(n) = \frac{\lambda}{(\nu-\lambda-\lambda\nu)} + \sqrt{\epsilon}\frac{\partial\Psi_1}{\partial y}(0,1) + O(\epsilon), \tag{5.11}$$

and, from (2.5),

$$E(m) = -\frac{1}{\sqrt{\epsilon}\nu}\frac{\partial\Psi_1}{\partial y}(0,1) + O(1). \tag{5.12}$$

To evaluate $\partial\Psi_1/\partial y(0,1)$ we need to know $f_1(0,y)$, and it appears that its determination requires complex analysis. We will make use of the (corrected) results of Fayolle and Iasnogorodski [6], and the results of Guillemin and Pinchon [8], to determine this quantity.

With $\gamma_0 = 1 - \lambda$, we define

$$z(\theta) = \lambda + \gamma_0\nu - 2\sqrt{\lambda\gamma_0\nu}\cos\theta, \tag{5.13}$$

and

$$\begin{aligned}J(y) = 4\lambda\gamma_0\nu&\\ \times \int_0^{\pi}&\frac{\sin^2\theta\sqrt{\delta^2 + 4\lambda z(\theta)}}{z(\theta)(\gamma_0\nu + \lambda y^2 - 2y\sqrt{\lambda\gamma_0\nu}\cos\theta)}d\theta.\end{aligned} \tag{5.14}$$

It is shown in Appendix 1 that, for $|y| < \sqrt{\gamma_0\nu/\lambda}$,

$$\begin{aligned}f_1(0,1) - f_1(0,y) = &\frac{(\nu-\lambda-\lambda\nu)}{4\pi\lambda\nu\gamma_0}[J(1) - yJ(y)]\\ &- \frac{\delta(1-y)}{2(\nu-\lambda\nu-\lambda y)}.\end{aligned} \tag{5.15}$$

We note that $\gamma\nu \sim \nu(1-\lambda) > \lambda$. It follows from (5.9) and (5.15) that

$$\begin{aligned}\Psi_1(0,y) = \frac{1}{(\nu-\lambda\nu-\lambda y)}\bigg\{&\frac{\delta\lambda\nu(1-y)}{2(\nu-\lambda\nu-\lambda y)}\\ &+ \frac{(\nu-\lambda-\lambda\nu)}{4\pi(1-\lambda)}[J(1) - yJ(y)]\bigg\}.\end{aligned} \tag{5.16}$$

Hence,

$$\frac{\partial\Psi_1}{\partial y}(0,1) = -\bigg\{\frac{\delta\lambda\nu}{2(\nu-\lambda-\lambda\nu)^2}$$

$$+\frac{1}{4\pi(1-\lambda)}[J(1)+J'(1)]\bigg\}, \tag{5.17}$$

where the prime denotes derivative. From (5.13) and (5.14) we obtain

$$J(1)+J'(1) = 4\lambda\gamma_0\nu(\gamma_0\nu-\lambda) \times \int_0^\pi \frac{\sin^2\theta\sqrt{\delta^2+4\lambda(\gamma_0\nu+\lambda-2\sqrt{\lambda\gamma_0\nu}\cos\theta)}}{(\gamma_0\nu+\lambda-2\sqrt{\lambda\gamma_0\nu}\cos\theta)^3}d\theta. \tag{5.18}$$

We first consider the matching, i.e., agreement to lowest order, with the results in the previous sections for an underloaded and for an overloaded fast queue, by analyzing the regimes $\sqrt{\epsilon}|\delta| \ll 1$, and $-\delta \gg 1$ and $\delta \gg 1$, respectively. Now,

$$\sqrt{\delta^2+4\lambda z(\theta)} = |\delta|\left[1+\frac{2\lambda}{\delta^2}z(\theta)+O\left(\frac{1}{\delta^4}\right)\right], \quad |\delta| \gg 1. \tag{5.19}$$

But [7],

$$\int_0^\pi \frac{\sin^2\theta}{(a^{-1}+a-2\cos\theta)}d\theta = \frac{\pi a}{2}, \quad |a|<1. \tag{5.20}$$

If we differentiate with respect to a we obtain

$$\int_0^\pi \frac{\sin^2\theta}{(a^{-1}+a-2\cos\theta)^2}d\theta = \frac{\pi a^2}{2(1-a^2)}, \quad \int_0^\pi \frac{\sin^2\theta}{(a^{-1}+a-2\cos\theta)^3}d\theta = \frac{\pi a^3}{2(1-a^2)^3}, \quad |a|<1. \tag{5.21}$$

It follows from (5.18–5.21), since $\gamma_0\nu = \nu(1-\lambda) > \lambda$, that

$$J(1)+J'(1) = \frac{2\pi\lambda\nu(1-\lambda)|\delta|}{(\nu-\lambda-\lambda\nu)^2}+\frac{4\lambda^2}{|\delta|}+O\left(\frac{1}{|\delta|^3}\right). \tag{5.22}$$

From (5.17) and (5.22) we obtain

$$\frac{\partial\Psi_1}{\partial y}(0,1) = -\frac{\lambda\nu(\delta+|\delta|)}{2(\nu-\lambda-\lambda\nu)^2}-\frac{\lambda^2}{(1-\lambda)|\delta|}+O\left(\frac{1}{|\delta|^3}\right). \tag{5.23}$$

It follows from (5.12) that

$$E(m) \sim \frac{-\lambda^2}{\sqrt{\epsilon}\nu(1-\lambda)\delta}, \quad -\delta \gg 1. \tag{5.24}$$

This matches the result in (3.11) for an underloaded fast queue, since $1-\gamma-\lambda = -\sqrt{\epsilon}\delta \ll 1$. On the other hand, from (5.12) and (5.23),

$$E(m) \sim \frac{\lambda\delta}{\sqrt{\epsilon}(\nu-\lambda-\lambda\nu)^2}, \quad \delta \gg 1. \tag{5.25}$$

This matches with the result in (4.7) for an overloaded fast queue, since $\gamma-1+\lambda = \sqrt{\epsilon}\delta \ll 1$.

Generally, if we replace the variable θ in (5.18) by z, where $z(\theta)$ is given by (5.13), we obtain

$$J(1)+J'(1) = (\gamma_0\nu-\lambda) \times \int_{(\sqrt{\gamma_0\nu}-\sqrt{\lambda})^2}^{(\sqrt{\gamma_0\nu}+\sqrt{\lambda})^2} \sqrt{[z-(\sqrt{\gamma_0\nu}-\sqrt{\lambda})^2][(\sqrt{\gamma_0\nu}+\sqrt{\lambda})^2-z](\delta^2+4\lambda z)}\frac{dz}{z^3}, \tag{5.26}$$

which is an elliptic integral. With the help of the transformation at the top of page 77 in [4], the integral in (5.26) may be expressed in terms of one involving Jacobian elliptic functions. This, in turn, may be evaluated by 336.00–336.03 in [4], in terms of complete elliptic integrals of the first, second and third kinds. We omit the result since it is not very illuminating. The integral in (5.26) may be evaluated numerically for prescribed parameter values, and the asymptotic approximation to $E(m)$ is then given by (5.12) and (5.17).

Acknowledgements The author is grateful to Sem Borst for several helpful discussions, and for providing an interpretation of the results for an underloaded fast queue, in a more general setting [3]. He is also grateful to the referees for their helpful suggestions for improving the presentation. Above all, the author is deeply indebted to John A. Chabot, M.D., F.A.C.S., Chief, Division of General Surgery, Columbia University Medical Center, whose skill enabled him to survive to revise this paper, the first draft of which was completed shortly before his surgery on March 10, 2006.

Appendix 1

We here use the (corrected) results in Sect. VI of Fayolle and Iasnogorodski [6] to determine $f_1(0,y)$. In their notation,

$$\lambda_1=\lambda, \quad \mu_1 = 1-\gamma, \quad \mu_1^*=1; \qquad \lambda_2=\epsilon\lambda, \quad \mu_2=\epsilon\gamma\nu, \quad \mu_2^*=\epsilon\nu. \tag{6.1}$$

Also,

$$\xi = 1-\gamma, \quad \rho_1^*=\lambda, \quad \rho_2^*=\lambda/\nu, \tag{6.2}$$

$F(0,0)$ is given by (2.8) and, since $\gamma=\gamma_0+\sqrt{\epsilon}\delta$ and $\gamma_0 = 1-\lambda$,

$$\beta = \epsilon(\lambda+\gamma\nu-2\sqrt{\lambda\gamma\nu}\cos\theta) = \epsilon[z(\theta)+O(\sqrt{\epsilon})], \tag{6.3}$$

where $z(\theta)$ is given by (5.13). It follows that

$$2\lambda K(\theta) = \lambda_1+\mu_1+\beta - \sqrt{[(\sqrt{\lambda_1}+\sqrt{\mu_1})^2+\beta][(\sqrt{\lambda_1}-\sqrt{\mu_1})^2+\beta]} = 2\lambda-\sqrt{\epsilon}[\delta+\sqrt{\delta^2+4\lambda z(\theta)}\,]+O(\epsilon). \tag{6.4}$$

Hence,

$$\begin{aligned}&\rho_1^*(\mu_2^*-\mu_1^*)[K(\theta)]^2+(\mu_1^*-\mu_2^*+\lambda_1+\lambda_2)K(\theta)-\mu_1^*\\&\quad=-\frac{\sqrt{\epsilon}\gamma_0}{2\lambda}[\delta+\sqrt{\delta^2+4\lambda z(\theta)}\,]+O(\epsilon).\end{aligned}\tag{6.5}$$

Consequently,

$$\begin{aligned}v(\theta)&=\frac{-F(0,0)\sqrt{\lambda_2\mu_2}\sin\theta K(\theta)}{\{\rho_1^*(\mu_2^*-\mu_1^*)[K(\theta)]^2+(\mu_1^*-\mu_2^*+\lambda_1+\lambda_2)K(\theta)-\mu_1^*\}\xi}\\&=\frac{\sqrt{\epsilon}(\nu-\lambda-\lambda\nu)}{2\sqrt{\lambda\gamma_0\nu}z(\theta)}\sin\theta[\sqrt{\delta^2+4\lambda z(\theta)}-\delta]+O(\epsilon).\end{aligned}\tag{6.6}$$

Hence, for $|y|<\sqrt{\gamma\nu/\lambda}$,

$$\begin{aligned}&F(0,y)-F(0,0)\\&\quad=\frac{2}{\pi}\sqrt{\lambda\gamma\nu}y\int_0^\pi\frac{\sin\theta v(\theta)}{(\gamma\nu+\lambda y^2-2y\sqrt{\lambda\gamma\nu}\cos\theta)}d\theta\\&\quad=\frac{\sqrt{\epsilon}}{\pi}(\nu-\lambda-\lambda\nu)y\\&\qquad\times\int_0^\pi\frac{\sin^2\theta[\sqrt{\delta^2+4\lambda z(\theta)}-\delta]}{z(\theta)(\gamma_0\nu+\lambda y^2-2y\sqrt{\lambda\gamma_0\nu}\cos\theta)}d\theta+O(\epsilon).\end{aligned}\tag{6.7}$$

But, from (5.13), (5.20) with the help of partial fractions for $y\neq 1$, and (5.21) for $y=1$,

$$\begin{aligned}&\int_0^\pi\frac{\sin^2\theta}{z(\theta)(\gamma_0\nu+\lambda y^2-2\sqrt{\lambda\gamma_0\nu}\cos\theta)}d\theta\\&\quad=\frac{\pi}{2\gamma_0\nu(\gamma_0\nu-\lambda y)},\qquad|y|<\sqrt{\gamma_0\nu/\lambda}.\end{aligned}\tag{6.8}$$

Hence, from (5.2), (5.14), (6.7) and (6.8),

$$\begin{aligned}&f_1(0,y)=\frac{(\nu-\lambda-\lambda\nu)}{4\pi\lambda\nu\gamma_0}y\left[J(y)-\frac{2\pi\lambda\delta}{(\gamma_0\nu-\lambda y)}\right],\\&\quad|y|<\sqrt{\gamma_0\nu/\lambda}.\end{aligned}\tag{6.9}$$

The result in (5.15) follows, since $\gamma_0=1-\lambda$.

We show in Appendix 2 that the expression for $F(0,y)$ in [8] leads to the asymptotic approximation in (6.7), although it takes somewhat more effort than the above derivation does.

Appendix 2

We here show that the expression for $F(0,y)$ in [8] leads to the asymptotic approximation in (6.7). This requires the evaluation of several quantities. In the notation of [8],

$$\begin{aligned}&\lambda_0=\lambda,\qquad\mu_0=1,\qquad\phi_0=1-\gamma;\\&\lambda_1=\epsilon\lambda,\qquad\mu_1=\epsilon\nu,\qquad\phi_1=\gamma.\end{aligned}\tag{7.1}$$

Hence, with

$$D=\lambda+1-\gamma+\epsilon(\lambda+\gamma\nu),\tag{7.2}$$

we have

$$\begin{aligned}&p_0=\lambda/D,\qquad q_0=(1-\gamma)/D,\\&p_1=\epsilon\lambda/D,\qquad q_1=\epsilon\gamma\nu/D.\end{aligned}\tag{7.3}$$

It follows from (5.1), after some algebra, that

$$\begin{aligned}s_-&=\frac{1}{2p_0}\left[1+2\sqrt{p_1q_1}-\sqrt{(1+2\sqrt{p_1q_1})^2-4p_0q_0}\,\right]\\&=1-\frac{\sqrt{\epsilon}}{2\lambda}\left[\delta+\sqrt{\delta^2+4\lambda(\sqrt{\gamma_0\nu}+\sqrt{\lambda})^2}\,\right]+O(\epsilon),\end{aligned}\tag{7.4}$$

where, as before $\gamma_0=1-\lambda$, and

$$\begin{aligned}\sigma_-&=\frac{1}{2p_0}\left[1-2\sqrt{p_1q_1}-\sqrt{(1-2\sqrt{p_1q_1})^2-4p_0q_0}\,\right]\\&=1-\frac{\sqrt{\epsilon}}{2\lambda}\left[\delta+\sqrt{\delta^2+4\lambda(\sqrt{\gamma_0\nu}-\sqrt{\lambda})^2}\,\right]+O(\epsilon).\end{aligned}\tag{7.5}$$

Next, from (7.2) and (7.3),

$$D(p_0x^2-x+q_0)=(1-\gamma-\lambda x)(1-x)-\epsilon(\lambda+\gamma\nu)x.\tag{7.6}$$

But [8],

$$\Delta(x;1)=(p_0x^2-x+q_0)^2-4p_1q_1x^2.\tag{7.7}$$

After some further algebra it is found that

$$\begin{aligned}&\sqrt{-\Delta(1-\sqrt{\epsilon}u;1)}\\&\quad=\frac{\epsilon}{2\lambda}\sqrt{[(\sqrt{\gamma_0\nu}+\sqrt{\lambda})^2-u(\lambda u-\delta)][u(\lambda u-\delta)-(\sqrt{\gamma_0\nu}-\sqrt{\lambda})^2]}\\&\qquad+O(\epsilon^{3/2}),\end{aligned}\tag{7.8}$$

and

$$p_0(1-\sqrt{\epsilon}u)^2-q_0=\frac{\sqrt{\epsilon}}{2\lambda}(\delta-2\lambda u)+O(\epsilon).\tag{7.9}$$

Also [8],

$$-K_1(x,y)=x(p_1y^2+q_1)+y(p_0x^2-x+q_0),\tag{7.10}$$

and it is found that

$$\begin{aligned}&K_1(1-\sqrt{\epsilon}u,y)\\&\quad=-\frac{\epsilon}{2\lambda}[(1-y)(\gamma_0\nu-\lambda y)+yu(\lambda u-\delta)]+O(\epsilon^{3/2}).\end{aligned}\tag{7.11}$$

Next [8],

$$\mathcal{Q}_0(x) = p_0\left(1-\frac{\phi_0 q_1}{\phi_1 q_0}\right)x^2 - \left(1-q_1+\frac{\phi_1}{\phi_0}q_0-\frac{q_1}{\phi_1}\right)x+\frac{q_0}{\phi_0}, \tag{7.12}$$

which after some more algebra leads to

$$\mathcal{Q}_0(1-\sqrt{\epsilon}u) = \sqrt{\epsilon}\frac{(1-\lambda)}{2\lambda}u + O(\epsilon). \tag{7.13}$$

Finally [8], for $|y| < \sqrt{\gamma\nu/\lambda}$,

$$F(0,y)-F(0,0) = \frac{(\nu-\lambda-\lambda\nu)}{2\pi(1-\gamma)\nu}y\int_{s_-}^{\sigma_-}\frac{(p_0x^2-q_0)\sqrt{-\Delta(x;1)}}{x\mathcal{Q}_0(x)K_1(x,y)}dx. \tag{7.14}$$

If we let $x = 1-\sqrt{\epsilon}u$ in (7.14), and use (7.4), (7.5), (7.8), (7.9), (7.11) and (7.13), we obtain

$$F(0,y)-F(0,0) = \frac{\sqrt{\epsilon}(\nu-\lambda-\lambda\nu)}{2\pi\lambda\gamma_0\nu}y \times \int_{\frac{1}{2\lambda}[\delta+\sqrt{\delta^2+4\lambda(\sqrt{\gamma_0\nu}-\sqrt{\lambda})^2}]}^{\frac{1}{2\lambda}[\delta+\sqrt{\delta^2+4\lambda(\sqrt{\gamma_0\nu}+\sqrt{\lambda})^2}]} I(u,y)du + O(\epsilon), \tag{7.15}$$

where

$$I(u,y) = \frac{(2\lambda u-\delta)\sqrt{[u(\lambda u-\delta)-(\sqrt{\gamma_0\nu}-\sqrt{\lambda})^2][(\sqrt{\gamma_0\nu}+\sqrt{\lambda})^2-u(\lambda u-\delta)]}}{u[(1-y)(\gamma_0\nu-\lambda y)+yu(\lambda u-\delta)]}. \tag{7.16}$$

The transformation

$$2\lambda u = \delta+\sqrt{\delta^2+4\lambda z} \tag{7.17}$$

leads to

$$F(0,y)-F(0,0) = \frac{\sqrt{\epsilon}(\nu-\lambda-\lambda\nu)}{\pi\gamma_0\nu}y \times \int_{(\sqrt{\gamma_0\nu}-\sqrt{\lambda})^2}^{(\sqrt{\gamma_0\nu}+\sqrt{\lambda})^2}\frac{\sqrt{[z-(\sqrt{\gamma_0\nu}-\sqrt{\lambda})^2][(\sqrt{\gamma_0\nu}+\sqrt{\lambda})^2-z]}}{(\delta+\sqrt{\delta^2+4\lambda z})[(1-y)(\gamma_0\nu-\lambda y)+yz]}dz + O(\epsilon). \tag{7.18}$$

If we replace the variable θ in (6.7) by z, where $z(\theta)$ is given by (5.13), we obtain (7.18).

References

1. Abramowitz, M., Stegun, I.A.: Handbook of Mathematical Functions. National Bureau of Standards, Washington (1964)
2. Altman, E., Avrachenkov, K.E., Núñez-Queija, R.: Perturbation analysis for denumerable Markov chains with application to queueing models. Adv. Appl. Probab. **36**, 839–853 (2004)
3. Borst, S.C., Morrison, J.A.: Bandwidth sharing with disparate flow classes and state-dependent service rates. Preprint
4. Byrd, P.F., Friedman, M.D.: Handbook of Elliptic Integrals for Engineers and Scientists, 2nd edn. Springer, New York (1971)
5. Erdélyi, A., Magnus, W., Oberhettinger, F., Tricomi, F.: Tables of Integral Transforms, vol. 1. McGraw-Hill, New York (1954)
6. Fayolle, G., Iasnogorodski, R.: Two coupled processors: the reduction to a Riemann–Hilbert problem. Z. Wahrsch. **47**, 325–351 (1979)
7. Gradshteyn, I.S., Ryzhik, I.M.: Tables of Integrals, Series and Products, 4th edn. Academic, New York (1965)
8. Guillemin, F., Pinchon, D.: Analysis of generalized processor-sharing systems with two classes of customers and exponential services. J. Appl. Probab. **41**, 832–858 (2004)
9. van Kessel, G., Núñez-Queija, R., Borst, S.C.: Differentiated bandwidth sharing with disparate flow sizes. In: Proc. of INFOCOM, 2005