ASYMPTOTICS FOR STEADY-STATE TAIL PROBABILITIES IN STRUCTURED MARKOV QUEUEING MODELS

by

Joseph Abate,[1] *Gagan L. Choudhury,*[2] *and Ward Whitt*[3]

AT&T Bell Laboratories

September 15, 1992

Revision: October 2, 1992

[1] AT&T Bell Laboratories Retired: 900 Hammond Road, Ridgewood, NJ 07450-2908
[2] AT&T Bell Laboratories, Room 3K-603, Holmdel, NJ 07733
[3] AT&T Bell Laboratories, Room 2C-178, Murray Hill, NJ 07974-0636

Abstract

In this paper we establish asymptotics for the basic steady-state distributions in a large class of single-server queues. We consider the waiting time, the workload (virtual waiting time) and the steady-state queue lengths at an arbitrary time, just before an arrival and just after a departure. We start by establishing asymptotics for steady-state distributions of Markov chains of M/GI/1 type. Then we treat steady-state distributions in the BMAP/GI/1 queue, which has a batch Markovian arrival process (BMAP). The BMAP is equivalent to the versatile Markovian point process or Neuts (N) process; it generalizes the Markovian arrival process (MAP) by allowing batch arrivals. The MAP includes the Markov-modulated Poisson process (MMPP), the phase-type renewal process (PH) and independent superpositions of these as special cases. We also establish asymptotics for steady-state distributions in the MAP/MSP/1 queue, which has a Markovian service process (MSP). The MSP is a MAP independent of the arrival process generating service completions during the time the server is busy. In great generality (but not always), the basic steady-state distributions have asymptotically exponential tails in all these models. When they do, the asymptotic parameters of the different distributions are closely related.

1. Introduction

One of the major accomplishments in queueing theory in recent years has been the identification and investigation of a large class of models that can be analyzed via structured Markov chains, in particular, the Markov chains of GI/M/1 type and M/GI/1 type in Neuts [17], [20]. This structure facilitates detailed analysis of many models of interest. One kind of analysis in this framework that has been very successful is asymptotic analysis of steady-state tail probabilities. Important asymptotic results for queues of GI/M/1 type, including the GI/PH/s and MMPP/PH/1 queues, have been established by Neuts [18], [19], Takahashi [25] and Neuts and Takahashi [21].

In this paper we continue this line of research. We start in §2 by establishing asymptotics for Markov chains of M/GI/1 type, exploiting the final-value theorem for generating functions together with the theory for Markov chains of M/GI/1 type in Neuts [20]. Then in §3 and §4 we establish asymptotics for steady-state tail probabilities in the BMAP/GI/1 and MAP/MSP/1 queues. The queue length at departures in the BMAP/GI/1 queue (together with an auxiliary phase state) is a Markov chain of M/GI/1 type, while the queue length at an arbitrary time in the MAP/MSP/1 queue (together with an auxiliary phase state) is a quasi-birth-and-death (QBD) process, which can be regarded as of both GI/M/1 and M/GI/1 type.

The BMAP/GI/1 model has a single server, the first-come first-served service discipline, unlimited waiting room and i.i.d. (independent and identically distributed) service times that are independent of a batch Markovian arrival process (BMAP). The BMAP and the BMAP/GI/1 queue were introduced and investigated by Lucantoni [13]. The BMAP is equivalent to the versatile Markovian point process in Neuts [16], [20] or the Neuts (N) process in Ramaswami [22], with an appealing simple notation. Thus the BMAP/GI/1 queue is the same, except for notation, as the N/G/1 queue analyzed by Ramaswami [22]; see also Ramaswami [23]. The

BMAP is a generalization of the Markovian arrival process (MAP) introduced by Lucantoni, Meier-Hellstern and Neuts [14] allowing batch arrivals. The MAP includes Markov modulated Poisson processes (MMPPs), phase-type (PH) renewal processes and independent superpositions of these processes as special cases. The MMPP models changes of phase without arrivals and arrivals without changes of phase; the BMAP models these as well as arrivals and changes of phase occurring simultaneously.

To model service times that are not necessarily i.i.d., in §4 we also consider the MAP/MSP/1 queue, which has a Markovian service process (MSP) as well as a MAP. The MSP is a MAP independent of the arrival process that operates (generates service completions) when the server is busy. The MAP/MSP/1 queue can be represented as a quasi-birth-and-death (QBD) process, so that its asymptotic behavior is determined by previous results of Neuts [17], [19].

Let W be the steady-state *waiting time* (experienced by an arriving customer before beginning service); let L be the steady-state *workload* (at an arbitrary time, i.e., the virtual waiting time); and let Q, Q^a and Q^d be the steady-state *queue lengths* (number in system) at an arbitrary time, just before an arrival and just after a departure, respectively. We show under quite general conditions that there exist positive constants $\eta, \sigma, \alpha_L, \alpha_W, \beta$, β^a and β^d such that

$$e^{\eta x} P(L > x) \to \alpha_L \text{ and } e^{\eta x} P(W > x) \to \alpha_W \text{ as } x \to \infty \qquad (1)$$

and

$$\sigma^{-k} P(Q > k) \to \beta\,,\ \ \sigma^{-k} P(Q^a > k) \to \beta^a \text{ and } \sigma^{-k} P(Q^d > k) \to \beta^d \text{ as } k \to \infty\,. \qquad (2)$$

We call η and σ the *asymptotic decay rates* and $\alpha_L, \alpha_W, \beta, \beta^a$ and β^d the associated *asymptotic constants.* Instead of (2), we actually establish related results for probability mass functions, such as

$$\sigma^{-k} P(Q = k) \to \beta(1 - \sigma)/\sigma \text{ as } k \to \infty\,. \qquad (3)$$

The convergence in (3) and (2) are equivalent, but (2) is usually of greater applied interest. We also establish limits for the joint distribution of these variables and the auxiliary phase state. Related results for the MAP/GI/1/K queue ($K \leq \infty$) have also recently been established by Baiocchi [4].

We also establish important relations among the asymptotic parameters in (1)–(3) in the BMAP/GI/1 model, extending previous results in this direction by Neuts [18], [19]. For Example, if V is a generic service time, then the asymptotic decay rates η and σ are related by $Ee^{\eta V} = \sigma^{-1}$. There are also simple relations among the asymptotic constants: $\beta^a = \alpha_W$, $\beta = \alpha_L$ and $\alpha_L / \alpha_W = \rho(1-\sigma)/\eta\sigma$. Hence, if you know the asymptotic decay rates and one asymptotic constant, then you know the other asymptotic constants. Indeed, assuming that we know the service-time transform Ee^{sV}, we only need to know one of the asymptotic rates. Additional results of this kind appear in [2].

We illustrate and confirm our results for the BMAP/GI/1 queue by analyzing an $\mathrm{MMPP}_2/\mathrm{D}_2/1$ example in §5. The MMPP_2 is an MMPP with a two-state environment Markov chain, while D_2 is a two-point service-time distribution. As in [1] and [2], the exact numerical results are based on the algorithms in Lucantoni [13], using numerical transform inversion algorithms in Abate and Whitt [3], as implemented by Choudhury [5]. The asymptotic parameters are calculated by a moment-based numerical inversion algorithm in Choudhury and Lucantoni [6]. The algorithm in [6] also calculates moments of all desired orders.

We are motivated to establish our asymptotic results because we believe that they often provide excellent approximations. The remarkable quality of exponential approximations for tail probabilities in GI/GI/s queues was pointed out in §1.9 and Chapter 4 of Tijms [26]. Further evidence is provided in §9 of Abate and Whitt [3] and in [1] and [2]. The associated approximations for higher percentiles are especially good; e.g., if $w_p = \inf\{x : P(W > x) < 1-p\}$, then $w_p \approx \log(\alpha_W/(1-\rho))/\eta$. We also develop and evaluate relatively simple approximations for the asymptotic parameters in [1], [2].

The BMAP/GI/1 queue is especially attractive because it includes superposition arrival processes. Whitt [27] applies the asymptotic decay rates in the BMAP/GI/1 and MAP/MSP/1 queues to develop effective bandwidths for independent sources to use for admission control in multi-service networks. This paper provides theoretical support for the procedures in [27]. For related work in this direction, see Elwalid and Mitra [8], [9] and references therein.

Our first goal here is to establish the validity of the limits in (1)–(3). Our second goal is to determine relatively simple expressions for the asymptotic decay rates η and σ. Our third, and least important, goal is to find expressions for the asymptotic constants. The expressions for the asymptotic constants are relatively complicated. Thus, it often will be convenient to actually compute or approximate the asymptotic constants in other ways; e.g., by [6].

2. Structured Markov Chains

For Markov chains of the GI/M/1 type the asymptotics of steady-state distributions is discussed at length in Neuts [19]. These GI/M/1-type Markov chains have matrix-geometric steady-state distributions, with a steady-state probability vector of the form $x_0, x_1, x_1 R, x_1 R^2, \ldots$ where R is the rate matrix. The decay rate corresponding to σ in (2) and (3) is the Perron-Frobenius eigenvalue of the nonnegative matrix R, which coincides with its spectral radius; see Chapter 1 and § 2.3 of Seneta [24] for background on the Perron-Frobenius theory. Moreover, the steady-state vector has the asymptotic form

$$x_1 R^i = \sigma^i (x_1 r) l + o(\sigma^i) \text{ as } i \to \infty , \tag{4}$$

where l and r are left and right eigenvectors associated with σ normalized so that $le = lr = 1$, with e being a vector of 1's; see (10) on p. 224 of Neuts [19]. (Our σ and η are η and ξ in [19].) Neuts points out that the Perron-Frobenius eigenvalue σ of R is relatively easy to obtain, even for large models, but the asymptotic constant (vector) $x_1 (r) l$ in (4) often is not.

We now show how to determine the asymptotics for the steady-state distributions of Markov chains of M/G/1 type. Let the transition matrix be $Q(\infty)$ as in (2.1.9) of Neuts [20] with component $m \times m$ matrices $A_k \equiv A_k(\infty)$; $m_1 \times m$ matrices $B_k \equiv B_k(\infty)$ and $m \times m_1$ matrix $C_0 \equiv C_0(\infty)$. We assume that $Q(\infty)$ is irreducible and positive recurrent. In the queueing models, positive recurrence primarily corresponds to $\rho < 1$, where ρ is the traffic intensity. Let $(x_0, x_1, \ldots)$ be the steady-state vector, with $x_i = (x_{i1}, \ldots, x_{im})$ and let $X(z) = \sum_{i=1}^{\infty} x_i z^i$ be its (vector) generating function. By (3.3.2) on p. 143 of [20], the generating function satisfies the equation

$$X(z)[zI - A(z)] = z x_0 B(z) - z x_1 A_0 , \tag{5}$$

where $A(z) = \sum_{k=0}^{\infty} A_k z^k$ and $B(z) = \sum_{k=1}^{\infty} B_k z^k$.

We start by directly establishing a solidarity result, showing that the asymptotic decay rate is independent of the auxiliary phase state. Motivated by Example 4 of [1], we allow asymptotic behavior of the form $\beta i^{-p} \sigma^i$ as $i \to \infty$ for $p \neq 0$ as well as $p = 0$. Let $\underline{\lim}_{i \to \infty}$ and $\overline{\lim}_{i \to \infty}$ denote the limit inferior and limit supremum, respectively.

Theorem 1. *Consider an irreducible positive-recurrent Markov chain of M/GI/1 type. Suppose that* $\underline{\lim}_{i \to \infty} \sigma^{-i} i^p x_{ij} = l_j$ *and* $\overline{\lim}_{i \to \infty} \sigma^{-i} i^p x_{ij} = u_j$ *for all* j, *where* $0 \leq l_j \leq u_j \leq \infty$ *and* $-\infty < p < \infty$. *If* $A \equiv \sum_{k=0}^{\infty} A_k$ *is irreducible, then* $l_j > 0$ ($u_j < \infty$) *holds for one j if and only if it holds for all j.*

Proof. Consider an initial state (i, j) where i is the level and j is the phase, with $i \geq m + 1$. Let j' be a designated alternative phase state. Since A is irreducible and there are only m phase states, it is possible to go from (i, j) to $(i + k, j')$ in at most m steps for some k. Since the chain can go down at most one level at each transition, we can have $k \geq -m$. In particular, there is a finite product of at most m of the $m \times m$ submatrices A_l that produce this transition. It is significant that this bounding probability is independent of i, provided that $i \geq m + 1$, since it is impossible to reach the lower boundary level from level above level $m + 1$ in m steps. As a consequence, there is a constant ε as well as the constant k such that

$$x_{ij} \geq \varepsilon \, x_{i+k,j'} \tag{6}$$

for all i. Since the states j and j' are arbitrary, formula (6) implies that

$$\underline{\lim}_{i \to \infty} \sigma^{-i} i^p x_{ij} \geq \varepsilon \, \underline{\lim}_{i \to \infty} \sigma^{-i} i^p x_{i+k,j'} = \varepsilon \sigma^k \underline{\lim}_{i \to \infty} \sigma^{-(i+k)} (i+k)^p x_{i+k,j'} \tag{7}$$

and

$$\overline{\lim}_{i \to \infty} \sigma^{-(i+k)} (i+k)^p x_{i+k,j'} \leq \varepsilon^{-1} \overline{\lim}_{i \to \infty} \sigma^{-(i+k)} (i+k)^p x_{ij} = \varepsilon^{-1} \sigma^{-k} \overline{\lim}_{i \to \infty} \sigma^{-i} i^p x_{ij} \; . \tag{8}$$

The inequalities (7) and (8) imply the desired conclusion. ■

We apply (5) to establish the asymptotic behavior of x_i as $i \to \infty$. For this purpose, we apply (5) with $z > 1$, which is valid provided that everything is finite. We directly assume that $B(z)$ is finite for the z of interest. From (5) it is evident that the radius of convergence of the generating function $X(z)$ is the minimal $z > 1$ such that $p(z) = z$ where $p(z) \equiv pf(A(z))$ is the Perron-Frobenius eigenvalue of $A(z)$. (It is understood that $p(z) = \infty$ if all elements of $A(z)$ are not finite.) Properties of $p(z)$ are discussed in Chapter 1 and §2.3 of Seneta [24] as well as in §2.3 and the Appendix of Neuts [20]. We will need the following basic matrix convexity result due to Kingman [12], which is reviewed in the Appendix of [20].

Lemma 1. *If the elements of a nonnegative square matrix $A(s)$ are log-convex functions of s, then the Perron-Frobenius eigenvalue $pf(A(s))$ is log-convex.*

We apply Lemma 1 to transforms and related functions, using the following consequence of Hölder's inequality; see p. 284 of Kingman [12].

Lemma 2. *Given any random variable X, the transform Ee^{sX} is log-convex in s.*

Proposition 3. *Let $p(z) \equiv pf(A(z))$ be the Perron-Frobenius eigenvalue of $A(z)$, with*

$p(z) = \infty$ if $A(z)$ is not finite. Then $\log p(z)$ *and $p(z)$ are increasing convex functions of z for $z > 0$, so that when $A(1)$ is irreducible and positive recurrent the equation $p(z) = z$ has at most one root for $z \neq 1$, and, if there is a root, this root must satisfy $z > 1$.*

Proof. The argument is a variant of the proof of Lemma 2.3.4 on p. 94 of Neuts [20]. As done there, start by making the change of variables $z = e^{-s}$. Then the elements of the matrix $A(e^{-s})$ are log-convex functions of s by Lemma 2. (The elements are constant multiples of transforms.) Consequently, $\log p(e^{-s})$ is a decreasing convex function of s on $(-\infty,\infty)$, by Lemma 1. (The relevant domain for s in [20] is $[0,\infty)$, but we are interested in $(-\infty,0)$; we can let the domain be $(-\infty,\infty)$, allowing infinite values. Alternatively, we can conclude that $\log p(e^{-s})$ is convex on the interval it is finite.) By taking the composition of this decreasing convex function with the decreasing convex function $-\log z$, we see that $\log p(z)$ is an increasing convex function of z for $z > 0$. Trivially, by taking the composition with the exponential function, we see that $p(z)$ itself is also an increasing convex function of z. Since $A(1)$ is stochastic, $p(1) = 1$. Since $A(1)$ is irreducible and positive recurrent, $p'(1) < 1$; see p. 481 of [20]. Hence, there is at most one other root of the equation $p(z) = z$. By the convexity, if there is a root, then it must satisfy $z > 1$. ■

Remark 1. The change of variables in the proof of Proposition 3 plays an important role. It is easy to see that the elements of $A(z)$ are convex functions of z, but as noted by Kingman [12] convexity of the elements does *not* directly imply convexity of the Perron-Frobenius eigenvalue $pf(A(z))$. ■

Example 4 in [1] involving the simple $M/GI/1$ queue shows that the equation $p(z) = z$ may actually *not* have a root for $z > 1$, even when the service-time distribution has a finite moment generating function. Then W does not satisfy (1). Similarly, in that model Q^d (which has the steady-state distribution of a Markov chain of M/GI/1 type) does not satisfy (2) or (3). In that example, $P(Q > k) \sim \tilde{\beta} x^{-3/2} \sigma^k$ as $k \to \infty$, but the asymptotic form is not a good approximation until extremely large k.

We now want to provide conditions for the singularity of $X(z)$ at $z = \sigma^{-1}$ to be a simple pole, from which it follows that $\sigma^{-i} x_{ij}$ converges to a nondegenerate limit. (It is easy to see that if $\sigma^{-i} x_{ij} \to l_j$ as $i \to \infty$, then $\sigma^{-i} \sum_{k=i+1}^{\infty} x_{kj} \to \sigma l_j / (1-\sigma)$ as $i \to \infty$.) For this purpose, it is convenient to rewrite (5) as

$$X(z)(I - \bar{A}(z)) = x_0 B(z) - x_1 A_0 \tag{9}$$

where $\bar{A}(z) = A(z)/z$. Let $\bar{p}(z)$ be the Perron-Frobenius eigenvalue of $\bar{A}(z)$. By Proposition 3, we are interested in the z such that $\bar{p}(z) = 1$. We next introduce a generalization of the fundamental matrix of Kemeny and Snell [11], which is a convenient form of a generalized inverse; see Hunter [10]. Variants of this argument are used in Neuts [20]; e.g., see Theorem 3.3.1 on p. 144. A similar argument is also used by Baiocchi [4] to obtain the corresponding result for the MAP/GI/1 special case.

Consider an irreducible nonnegative matrix P with Perron-Frobenius eigenvalue $p \leq 1$ and let l and r be left and right eigenvectors associated with p normalized so that $le = lr = 1$, where e is a vector of 1's. Let the *fundamental matrix* associated with P be $Z = (I - P + prl)^{-1}$.

We now show that our fundamental matrix Z has the proper properties. First, note that Z is the familiar fundamental matrix on pp. 75 and 100 of Kemeny and Snell [11] when P is stochastic. (Then l is the steady-state vector π and $r = e$, the vector of 1's.) Next, we relate the spectral radiuses of P and $P - prl$. The spectral radius of P is the Perron-Frobenius eigenvalue since P is nonnegative; this is not the case for $P - prl$.

Lemma 3. *If P is an irreducible nonnegative matrix with positive Perron-Frobenius eigenvalue p and associated left and right eigenvectors l and r normalized so that $le = lr = 1$, then $P - prl$ is a matrix with spectral radius strictly less than p.*

Proof. Note, by the orthogonality of eigenvectors, that the matrices P and $P - prl$ have the same eigenvectors, with the eigenvalue of $P - prl$ associated with l and r being 0. Moreover, the remaining eigenvalues of the matrices P and $P - prl$ coincide. Since p is strictly greater than the modulus of any other eigenvalue of P, the result is established. ■

Proposition 4. *If P is an irreducible nonnegative matrix with Perron-Frobenius eigenvalue $p \le 1$ with associated left and right eigenvectors l and r normalized so that $le = lr = 1$, then $(I - P + prl)$ is nonsingular,*

$$Z \equiv (I - P + prl)^{-1} = I + \sum_{n=1}^{\infty} (P - prl)^n = I + \sum_{n=1}^{\infty} (P^n - p^n rl) , \qquad (10)$$

$Zr = r > 0$ and $lZ = l > 0$.

Proof. Let sp be the spectral radius. By Lemma 3, $sp(P - prl) < 1$. Hence, $(P - prl)^n \to 0$ and (10) is valid; see p. 22 of Kemeny and Snell [11] and p. 252 of Seneta [24]. Then, as on p. 75 of [11],

$$\begin{aligned}(P - prl)^n &= \sum_{i=0}^{n} \binom{n}{i} (-1)^{n-i} P^i (prl)^{n-i} \\ &= P^n + \sum_{i=0}^{n-1} \binom{n}{i} (-1)^{n-i} p^n rl = P^n - p^n rl .\end{aligned}$$

Finally,

$$Zr = (I + \sum_{n=1}^{\infty} P^n - p^n rl) r = r + \sum_{n=1}^{\infty} (p^n r - p^n r) = r ,$$

where r is known to be strictly positive. A similar argument applies to lZ. ■

We are now ready to show that the analog of (3) is valid in great generality for Markov chains of the M/GI/1 type.

Theorem 5. *Consider an irreducible positive-recurrent Markov chain of M/GI/1 type in which the matrix $A \equiv A(1)$ is irreducible and positive recurrent. If the equation $p(z) \equiv pf(A(z)) = z$ has a root σ^{-1} for $z > 1$, $p(\sigma^{-1} + \varepsilon) < \infty$ for some positive ε, $B(\sigma^{-1})$ is finite and $(x_0 B(\sigma^{-1}) - x_1 A_0) r(\sigma^{-1}) > 0$, then the radius of convergence of $X(z)$ in (5) and (9) is σ^{-1},* *the derivative $p'(\sigma^{-1})$ exists with $p'(\sigma^{-1}) > 1$ and the singularity σ^{-1} is a simple pole, so that*

$$\sigma^{-i} x_i \rightarrow l \equiv \frac{(x_0 B(\sigma^{-1}) - x_1 A_0) r(\sigma^{-1}) l(\sigma^{-1})}{p'(\sigma^{-1}) - 1} \text{ as } i \rightarrow \infty , \tag{11}$$

with all components of the limit in (11) being strictly positive and finite.

Proof. Since the matrix A is irreducible, so is $A(z)$ for all $z > 0$. Hence, $pf(A(z))$ is a simple eigenvalue. By Proposition 3, there is at most one root for $z > 1$. From (9), it is evident that $(I - \bar{A}(z))^{-1}$ exists and equals $\sum_{k=0}^{\infty} \bar{A}(z)^k$ provided that $\bar{p}(z) < 1$. Let $l(z)$ and $r(z)$ be left and right eigenvectors associated with $\bar{p}(z)$. Multiplying by $r(z)$ on the right in (9), we see that

$$X(z) r(z)(1 - \bar{p}(z)) = (x_0 B(z) - x_1 A_0) r(z) \equiv H(z) r(z) , \tag{12}$$

so that

$$X(z)(I - \bar{A}(z) + \bar{p}(z) r(z) l(z)) = H(z) + \frac{H(z)\bar{p}(z) r(z) l(z)}{1 - \bar{p}(z)}$$

and, by Proposition 4,

$$X(z) = \left[H(z) + \frac{H(z)\bar{p}(z) r(z) l(z)}{1 - \bar{p}(z)} \right] (I - \bar{A}(z) + \bar{p}(z) r(z) l(z)^{-1}$$

Now we can apply the final-value theorem for probability generating functions (see §5.2 of Wilf [28]) to get

$$\lim_{i \rightarrow \infty} \sigma^{-1} x_i = \lim_{z \rightarrow \sigma^{-1}} (1 - z\sigma) X(z)$$

$$= \lim_{z \rightarrow \sigma^{-1}} (1 - z\sigma) \left[H(z) + \frac{H(z)\bar{p}(z) r(z) l(z)}{1 - \bar{p}(z)} \right] (I - \bar{A}(\sigma^{-1}) + \bar{p}(\sigma^{-1}) r(\sigma^{-1}) l(\sigma^{-1}))^{-1}$$

$$= \frac{H(\sigma^{-1})\bar{p}(\sigma^{-1}) r(\sigma^{-1}) l(\sigma^{-1})}{\sigma^{-1} \bar{p}'(\sigma^{-1})} (I - \bar{A}(\sigma^{-1}) + \bar{p}(\sigma^{-1}) r(\sigma^{-1}) l((\sigma^{-1}))^{-1}$$

$$= \frac{H(\sigma^{-1}) r(\sigma^{-1}) l(\sigma^{-1})}{p'(\sigma^{-1}) - 1} (I - \bar{A}(\sigma^{-1}) + r(\sigma^{-1}) l(\sigma^{-1}))^{-1} ,$$

because

$$\lim_{z \to \sigma^{-1}} \frac{\bar{p}(z) - 1}{z\sigma - 1} = \frac{\bar{p}'(\sigma^{-1})}{\sigma} = p'(\sigma^{-1}) - 1 \ .$$

The derivative $p'(\sigma^{-1})$ exists because $p(\sigma^{-1} + \varepsilon) < \infty$ by assumption, and $p(z)$ is an analytic function of z away from its singularity; see the Appendix of [20]. Since $A(1)$ is positive recurrent, $p'(1) < 0$, as noted in the proof of Theorem 3. Since $p(z)$ is convex, we thus must have $p'(\sigma^{-1}) > 1$. Next, apply Proposition 4 to obtain

$$l(\sigma^{-1})(I - \bar{A}(\sigma^{-1}) + r(\sigma^{-1})l(\sigma^{-1}))^{-1} = l(\sigma^{-1}) \ .$$

Finally, we need to show that all components of the limit in (11) is strictly positive. By assumption $(x_0 B(\sigma^{-1}) - x_1 A_0) r(\sigma^{-1})$ is positive, and it is known that all components of $l(\sigma^{-1})$ are strictly positive. Hence, the vectors $\sigma^{-i} x_i$ converge as $i \to \infty$ to a finite positive limit. ■

Remark 2. The unappetizing condition that $(x_0 B(\sigma^{-1}) - x_1 A_0) r(\sigma^{-1}) > 0$ needs to be verified in applications. This can often be done easily in specific contexts, as we illustrate for the BMAP/GI/1 queue in the next section. In general, we know that $r(1) = e$ since $A \equiv A(1)$ is stochastic and that $(x_0 B(1) - x_1 A_0) e = 0$; see p. 145 of [20]. By (12), $(x_0 B(z) - x_1 A_0) r(z) > 0$ for all z, $1 < z < \sigma^{-1}$. For the M/GI/1 queue, $x_1 = A_0^{-1}(1 - B_0) x_0$ by p. 16 of [20], so that

$$x_0 B(\sigma^{-1}) - x_1 A_0 = x_0 (B(\sigma^{-1}) - (1 - B_0)) \ , \qquad (13)$$

which is strictly positive for $\sigma^{-1} > 1$ because $B(z) > 0$ for $z > 0$. ■

Several related limits follow immediately from Theorem 5. Results later in the paper have similar corollaries.

Corollary. *Under the conditions of Theorem 5,*

$$\sigma^{-i} x_i e \to le = \frac{(x_0 B(\sigma^{-1}) - x_1 A_0) r(\sigma^{-1})}{p'(\sigma^{-1}) - 1} \ ,$$

$$\frac{x_{i+1,j}}{x_{ij}} \to \sigma\ , \ \frac{x_{i+1}e}{x_i e} \to \sigma \text{ and } \frac{x_{ij}}{x_i e} \to l(\sigma^{-1})$$

as $i \to \infty$.

3. The BMAP/GI/1 Queue

The queue length at departure epochs in the BMAP/GI/1 queue is a Markov chain of the M/GI/1 type, so that we can apply Theorem 5 to analyze the BMAP/GI/1 queue. Explicit transform results for the steady-state distributions are given in Lucantoni [13], but the transform results we will apply primarily follow from Ramaswami [22]. As we have indicated above, results related to ours for Q^d in the MAP/GI/1/K queue ($K \leq \infty$) have also recently been obtained by Baiocchi [4].

A BMAP can be defined in terms of two processes $N(t)$ and $J(t)$: $N(t)$ counts the number of arrivals in $[0,t]$ and $J(t)$ is an auxiliary state variable. The pair $(N(t), J(t))$ is a continuous-time countably-infinite-state Markov chain with infinitesimal generator $\tilde{Q}$ in block-partitioned form, i.e.,

$$\frac{\tilde{Q}}{\rho} = \begin{bmatrix} D_0 & D_1 & D_2 & D_3 \ldots \\ & D_0 & D_1 & D_2 \ldots \\ & & D_0 & D_1 \ldots \\ & & & D_0 \ldots \end{bmatrix} \tag{14}$$

where ρ is the overall arrival rate, D_k, $k \geq 0$, are $m \times m$ matrices, D_0 has negative diagonal elements and nonnegative off-diagonal elements, D_k is nonnegative for $k \geq 1$, and $D \equiv \sum_{k=0}^{\infty} D_k$ is an irreducible infinitesimal generator matrix for an m-state continuous-time Markov chain. (Our expressions differ slightly from Lucantoni's (1991), because we have factored out the overall arrival rate ρ in (14).)

Let V denote a generic service time. We assume that $\phi(s) \equiv Ee^{sV} < \infty$ for some positive s. This is a necessary, but not sufficient, condition for the asymptotics in (1)–(3). Then

$\phi'(s) = EVe^{sV} < \infty$ too. Key quantities are $\phi(\eta)$ and $\phi'(\eta)$ where η is the asymptotic decay rate in (1).

The generating functions of Q^d and Q are given in (20) and (35) of [14], while the Laplace transform of L is given in (44) there. In particular, let x_{ij}^d be the steady-state probability that the queue length is i and the auxiliary state is j just after a departure. Let $x_i^d = (x_{i1}^d, \ldots, x_{im}^d)$ and let $X^d(z) = \sum_{i=0}^{\infty} x_i^d z^i$. Let $D(z) = \sum_{k=0}^{\infty} D_k z^k$ and

$$A(z) = E[e^{\rho D(z)V}] = \int_0^\infty e^{\rho D(z)y} dP(V \le y) \ . \tag{15}$$

Then

$$X^d(z)[zI - A(z)] = (-x_0^d D_0^{-1}) D(z) A(z) \ , \tag{16}$$

where $(-x_0^d D_0^{-1})$ is a positive vector. The generating function of Q^d itself is $Q^d(z) = X^d(z)e$, where again e is a vector of 1's.

As we have indicated, (16) is a consequence of (5). As with (5), we apply (16) for $z > 1$. For this, we thus need the matrices $D(z)$ and $A(z)$ to be finite for the z of interest. To analyze (16), we use the Perron-Frobenius eigenvalues of $D(z)$ and $A(z)$, which are well defined since $A(z)$ is nonnegative and $D(z)$ has nonnegative off-diagonal elements; see Chapter 1 and §2.3 of Seneta [24]. As with (5), we will be interested in the z such that $pf(A(z)) = z$. By Proposition 3, we know that there is at most one root to the equation $pf(A(z)) = z$ arising in (16). Moreover, a necessary condition for such a root to exist is for $D(z)$ and $A(z)$ to be finite, so we do not need to assume anything extra about finiteness.

We now find alternative conditions for the root to exist in the special case of a BMAP/GI/1 queue. These conditions are consistent with Neuts [19].

Proposition 6. *In the BMAP/GI/1 queue, there is at most one root with* $z > 1$ *to the equation* $pf(A(z)) = z$. *One exists if and only if there are solutions* σ *and* η *with* $0 < \sigma < 1$ *and*

$0 < \eta < \infty$ *to the two equations*

$$pf(D(1/\sigma)) = \frac{\eta}{\rho} \quad \text{and} \quad Ee^{\eta V} = 1/\sigma \ . \tag{17}$$

Proof. By §2.3 of Seneta [24], the Perron-Frobenius eigenvalue of $D(z)$ is well defined for all $z > 0$, provided that $D(z)$ is finite. Moreover, the matrices $D(z)$ and $A(z)$ have a common associated positive real right eigenvector $r(z)$ by (15):

$$\begin{aligned} A(z)r(z) &= \int_0^\infty e^{\rho D(z)y} r(z)\, dP(V \le y) \\ &= r(z) \int_0^\infty e^{\rho y pf(D(z))}\, dP(V \le y) \ , \end{aligned}$$

so that

$$pf(A(z)) = Ee^{\rho pf(D(z))V} \ . \tag{18}$$

From (18), it is easy to see that (17) is equivalent to $pf(A(z)) = z$. ■

We will show that the two equations in (17) determine the asymptotic decay rates σ and η in (1)–(3). The representation of the single equation $pf(A(z)) = z$ in terms of the two equations in (17) allows us to identify and separate the effects of the arrival process and the service-time distribution. The arrival process alone and service-time distribution alone each determine a relation between the asymptotic decay rates σ and η. For example, if the service-time distribution is deterministic (D), then $\sigma = e^{-\eta}$; if the service-time distribution is exponential (M), then $\sigma = 1 - \eta$. As the service-time distribution gets more variable in the convex stochastic order (see §8 of [1]), then $\sigma(\eta)$ increases, so that for any service-time distribution with mean 1, $e^{-\eta} \le \sigma < 1$. For further discussion of this decomposition of independent arrival and service processes, see Whitt [27] and Elwalid and Mitra [8], [9].

Next we apply Proposition 3 to deduce some properties of the Perron-Frobenius eigenvalue of $D(z)$.

Proposition 7. *The Perron-Frobenius eigenvalue of* $D(z)$ *is a strictly increasing convex function* *of z for* $z \geq 0$ *with* $pf(D(1)) = 0$.

Proof. Since $D(z)$ is irreducible, $e^{D(z)}$ is a nonnegative matrix with $pf(e^{D(z)}) = e^{pf(D(z))}$; see (18) above and Theorem 2.7 of Seneta [24]. By Proposition 3, $pf(D(z)) = \log(e^{pf(D(z))})$ is increasing and convex function of z. Since $e^{D(1)}$ is stochastic, $pf(e^{D(1)}) = 1$, which implies that $pf(D(1)) = 0$. ■

We can now obtain large-time asymptotics results for Q^d in the BMAP/GI/1 queue directly from Theorem 5 and Proposition 6.

Theorem 8. *In the BMAP/GI/1 queue, if* $\rho < 1$, *if* $D \equiv D(1)$ *is irreducible, if the equations in (17) have solutions with* $0 < \sigma < 1$ *and* $0 < \eta < \infty$, *and if* $Ee^{(\varepsilon + \eta)V} < \infty$ *and* $pf(D(\sigma^{-1} + \varepsilon)) < \infty$ *for some positive* ε, *then the radius of convergence of the generating function* $X^d(z)$ *in (16) is* σ^{-1} *and this singularity is a simple pole, so that*

$$\sigma^{-i} x_i^d \rightarrow l^d = \frac{\eta(-x_0^d D_0^{-1}) r(\sigma^{-1}) l(\sigma^{-1})}{\rho(p'(\sigma^{-1}) - 1)} \quad \text{as } i \rightarrow \infty , \tag{19}$$

where $p(z) \equiv pf(A(z))$ *and all components of the limit* l^d *in (19) are positive and finite. In (19),* $p'(\sigma^{-1}) = \rho\phi'(\eta) p'_D(\sigma^{-1}) > 1$, *where* $\phi(\eta) = Ee^{\eta V}$ *and* $p_D(z) = pf(D(z))$.

Proof. The conditions here imply the conditions of Theorem 5, since (16) is a special case of (5). Note, however, that the summation in $X(z)$ and $X^d(z)$ start at 1 and 0, respectively. Thus, we can apply the proof of Theorem 5. First, the Markov chain is irreducible and positive recurrent because D is irreducible and $\rho < 1$. Here $(-x_0^d D_0^{-1}) D(z) \bar{A}(z)$ plays the role of $x_0 B(z) - x_0 B(z) - x_1 A_0$ in Theorem 5, and

$$(-x_0^d D_0^{-1}) D(z) \bar{A}(z) r(z) = (-x_0^d D_0^{-1}) pf(D(z)) pf(\bar{A}(z)) r(z) , \tag{20}$$

which is strictly positive for all $z > 1$, because $pf(D(1)) = 0$, $pf(D(z))$ is strictly increasing in z, and all other components are strictly positive. The right side of (20) simplifies at σ^{-1}; by (17) and (18), $pf(D(\sigma^{-1})) = \eta/\rho$ and $pf(A(\sigma^{-1})) = 1$. From (18), we see that $p'(\sigma^{-1})$ is as given. ■

Remark 3. Let g be the steady-state probability vector associated with the stochastic matrix G in (23) of Lucantoni [13]. By (53) of [13], $(-x_0^d D_0^{-1})$ in (19) can be expressed as $-(x_0^d D_0^{-1}) = (1 - \rho) g$. As noted before, (19) implies a limit $\sigma l^d / (1 - \sigma)$ for the tail probabilities $\sigma^{-k} \sum_{k=i+1}^{\infty} x_i^d$. This expression agrees with the formula for the MAP/GI/1 queue below (26) in Baiocchi [4]. ■

We also have expressions for the transforms of the other steady-state distributions in the BMAP/GI/1 queue. Let x_{ij}^t be the steady-state probability that the queue length is i and the auxiliary state is j at an arbitrary time. Let $x_i^t = (x_{i1}^t , \ldots , x_{im}^t)$ and $X^t(z) = \sum_{i=0}^{\infty} x_i^t z^i$. Then, by (35) of Lucantoni [13],

$$X^t(z) D(z) = (z-1) X^d(z) \; . \tag{21}$$

Relation (21) is established in Theorem 3.3.18 of Ramaswami [22] for the case $|z| < 1$, but it is valid more generally provided that everything is finite. It is convenient to use the following alternative expression for $X^t(z)$ from (3.3.20) of [22]:

$$X^t(z) - x_0^t = \rho x_0^d (D(z) - I) + \rho X^d(z) E[e^{\rho D(z) V_e}] \tag{22}$$

where V_e is a random variable with the service-time stationary-excess distribution, i.e.,

$$P(V_e \le x) = (EV)^{-1} \int_0^x P(V > u) \, du \; . \tag{23}$$

Lemma 4. *If* $pf(A(z)) = z$ *has a finite root* $\sigma^{-1} > 1$, *then* $D(\sigma^{-1})$ *and* $Ee^{\rho D(\sigma^{-1}) V_e}$ *are finite.*

Proof. By Proposition 6, $pf(D(\sigma^{-1})) = \eta / \rho < \infty$, so that $D(\sigma^{-1})$ must be finite. Also, by Proposition 6, $Ee^{\eta V} = \sigma^{-1} < \infty$. Hence, using integration by parts, $Ee^{\eta V_e} = E(e^{\eta V} - 1) / \eta EV < \infty$. Finally, as in Proposition 6,

$$pf(Ee^{\rho D(\sigma^{-1}) V_e}) = Ee^{\rho pf(D(\sigma^{-1})) V_e} = Ee^{\eta V_e} < \infty \; . \; ■$$

Similarly, let x_{ij}^a and $X^a(z)$ be the corresponding probabilities and generating function seen by the first customer in a batch upon arrival. Let $D(j,k)$ be the $(j,k)^{\text{th}}$ element of D. Then, by the covariance formula in (8) of Melamed and Whitt [15] for instance,

$$X_j^a(z) = X_j^t(z) \sum_{k=1}^{m} (D-D_0)(j,k) \ , \tag{24}$$

so that the generating function of Q^a is $Q^a(z) = X^a(z)e = X^t(z)(D-D_0)e$.

Remark 4. In the MAP/GI/1 queue, where customers arrive and depart one at a time, Q^a has the same distribution as Q^d, but we need not have $X^a(z) = X^d(z)$. What we do have is $X^a(z)e = X^d(z)e$. To see that this is consistent with (21) and (24), note that in the MAP/GI/1 case (21) becomes $X^t(z)(D_0 + D_1 z) = (z-1)X^d(z)$ and (24) yields $X^a(z)e = X^t(z)D_1 e$. Then note that $D_1 e = -D_0 e$, so that $X(z)D_1 e = X^d(z)e$. By (34) of Lucantoni [14], in the BMAP/GI/1 queue

$$x_0^t = -x_0^d D_0^{-1} \text{ or } x_0^d = -x_0^t D_0 \ ,$$

whereas by the reasoning in (24) $x_0^a e = x_0^t(D - D_0)e$. ■

From (21), (22), (24) and Theorem 8, we have the following asymptotic results for x_i^t and x_i^a.

Theorem 9. *Under the conditions of Theorem 8,*

$$\sigma^{-i}x_i^t \to l^t \text{ and } \sigma^{-i}x_i^a \to l^a \ \textit{as} \ i \to \infty \ , \tag{25}$$

where

$$l^t = \frac{\rho(1-\sigma)}{\sigma\eta} l^d \tag{26}$$

and

$$l^a = l^t \sum_{k=1}^{m} (D-D_0)(j,k) \tag{27}$$

for l^d in (19), so that $l^a e = l^t(D-D_0)e$.

Proof. Apply the final-value theorem for generating functions, first with (22) and (24). Note that

$$l^d \rho E(e^{\rho D(\sigma^{-1}) V_e}) = l^d \rho (1-\sigma)/\eta\sigma ,$$

because

$$l(\sigma^{-1}) E(e^{\rho D(\sigma^{-1} V_e}) = l(\sigma^{-1}) E(e^{p_D(\sigma^{-1}) V_e}) = l(\sigma^{-1}) \frac{(e^{\eta V} - 1)}{\eta} = \frac{(1 - \sigma) l(\sigma^{-1})}{\eta\sigma}$$

using Lemma 1 of [2]. Alternatively, from (19) and (21), since $l(\sigma^{-1}) D(\sigma^{-1}) = l(\sigma^{-1}) p_D(\sigma^{-1})$,

$$l^d = \frac{\sigma}{1-\sigma} l^t D(\sigma^{-1}) = \frac{\sigma}{1-\sigma} l^t p_D(\sigma^{-1}) = \frac{\sigma\eta l^t}{\rho(1-\sigma)} . \quad \blacksquare$$

Remark 5. In the MAP/GI/1 queue, (25)–(27) yield $l^d e = l^a e = l^t D_1 e = l^t e \sigma\eta/\rho(1-\sigma)$. As a quick check on (25)–(27), note that in the M/GI/1 queue $D - D_0 = D_1 = - D_0 = 1$, and $p_D(\sigma^{-1}) = \sigma^{-1} - 1 = \eta/\rho$, so that $\eta\sigma/\rho(1-\rho) = 1$. $\blacksquare$

Let $F_j(x)$ be the joint probability that the workload is less than or equal to x and the auxiliary state variable is j at an arbitrary time in steady state. Let $F(x) = (F_1(x), \ldots, F_m(x))$ and $\hat{F}(s) = \int_0^\infty e^{-sx} dF(x)$. Then, by (44) of Lucantoni [13],

$$\hat{F}(s) \left[I + \rho \frac{D(\hat{V}(s))}{s} \right] = x_0^t . \tag{28}$$

The Laplace transform of L is then $\hat{F}(s) e$.

Let $F^a(x) = F_1^a(x), \ldots, F_m^a(x)$ be the joint probability that the waiting time of the first customer in a batch is less than or equal to x and the auxiliary state variable upon arrival is j. Let $\hat{F}^a(s) = \int_0^\infty e^{-sx} dF^a(x)$. Then

$$\hat{F}_j^a(s) = \hat{F}(s) \sum_{k=1}^m (D - D_0)(j,k) \tag{29}$$

for $\hat{F}(s)$ in (28), by the same argument as for (24). The Laplace transform of W^b, where W^b is the waiting time of the first customer in a batch, is then $\hat{F}^a(s) e = \hat{F}(s)(D - D_0) e$.

To treat the workload and waiting time of the first customer in a batch in (28) and (29), let $\tilde{l}(s)$ and $\tilde{r}(s)$ be left and right eigenvectors of $\rho D(\hat{V}(s))/s$ associated with its Perron-Frobenius eigenvalue, which we denote by $f(s)$, normalized so that $\tilde{l}(s)e = \tilde{l}(s)\tilde{r}(s) = 1$. Note that $\tilde{l}(-\eta) = l(\sigma^{-1})$ and $\tilde{r}(-\eta) = r(\sigma^{-1})$. From (17), it is evident that the critical singularity is at $s = -\eta$. From (17), $f(-\eta) = -1$. By Theorem 7, $0 > f(s) \geq -1$ for $0 > s \geq -\eta$. We need the following analog of Proposition 4.

Lemma 5. *The matrix* $Y(\sigma^{-1}) \equiv I - \rho\eta^{-1}D(\sigma^{-1}) + r(\sigma^{-1})l(\sigma^{-1})$ *is nonsingular and* $l(\sigma^{-1})Y(\sigma^{-1}) = l(\sigma^{-1})$.

Proof. As noted above, (17) implies that $pf(\rho\eta^{-1}D(\sigma^{-1})) = 1$. Consider a vector u such that $u(I - \rho\eta^{-1}D(\sigma^{-1}) + r(\sigma^{-1})l(\sigma^{-1})) = 0$. If we multiply on the right by $r(\sigma^{-1})$, we see that $u\,r(\sigma^{-1}) = 0$, but this implies that u is a left eigenvector of $D(\sigma^{-1})$. Since $ur(\sigma^{-1}) = 0$, we must have $u = 0$. ■

Let S and S^a denote the auxiliary state at an arbitrary time and at an arrival epoch, respectively, in steady-state. As in (4) of [13], let π be the probability vector of S, i.e., satisfying $\pi D = 0$ and $\pi e = 1$. Let π^a be the probability vector of S^a, which by the argument for (24) satisfies $\pi_j^a = \pi_j \sum_{k=1}^{m} (D - D_0)(j,k)$. Then $F(\infty) = \pi$ and the Laplace transform for the tail probabilities $\pi - F(x)$ is $(\pi - \hat{F}(s))/s$. As before, let $\phi(\eta) = Ee^{\eta V}$ and $p_D(z) = pf(D(z))$.

Theorem 10. *Under the conditions of Theorem 8, the critical singularity of* $[\pi - \hat{F}(s)]/s$ *is at* $s = -\eta$ *and it is a simple pole, so that*

$$e^{\eta x}P(L > x,\ S = j) = e^{\eta x}(\pi_j - F_j(x)) \to l_j^L = \frac{x_0^t r(\sigma^{-1})\,l(\sigma^{-1})_j}{\rho p_D'(\sigma^{-1})\phi'(\eta) - 1} = \frac{\sigma l_j^t}{(1-\sigma)}\ , \tag{30}$$

$$e^{\eta x}P(W > x,\ S^a = j) = e^{\eta x}(\pi_j^a - F_j^a(x)) \to l^{W^b} = l_j^L \sum_{k=1}^{m} (D - D_0)(j,k) = \frac{\sigma l_j^a}{(1-\sigma)}, \tag{31}$$

$$e^{\eta x}P(L > x) \to \alpha_L \equiv l^L e \tag{32}$$

and

$$e^{\eta x} P(W^b > x) \to \alpha_{W^b} \equiv l^{W^b} e \ as \ x \to \infty \ , \tag{33}$$

where l^l and l^{W^b} are finite with all positive components.

Proof. Our approach is to apply the final-value theorem for Laplace transforms; e.g., p. 254 of Doetsch [7]. For this purpose, apply (27) to obtain

$$\frac{(\pi - F(s))}{s} (I + \rho D(\hat{V}(s))/s) = M(s) \equiv \frac{\pi (I + \rho D(\hat{V}(s))/s)}{s} - \frac{x_0^t}{s} \ . \tag{34}$$

Multiplying on the right in (34) by $\tilde{r}(s)$, we obtain for $s < 0$

$$\frac{(\pi - F(s))}{s} \tilde{r}(s)(1 - f(s)) = \frac{\pi (1 - F(s)) \tilde{r}(s) - x_0^t \tilde{r}(s)}{s} \ . \tag{35}$$

Immediately, we can write

$$\lim_{s \to -\eta} (s + \eta) \frac{(\pi - F(s))}{s} \tilde{r}(s) = \lim_{s \to -\eta} \left[(s + \eta) \frac{\pi \tilde{r}(s)}{s} - \frac{(s + \eta) x_0^t \tilde{r}(s)}{(1 - f(s)) s} \right]$$

$$= \frac{x_0^t \tilde{r}(-\eta)}{\eta f'(-\eta)} \ , \tag{36}$$

where, with $p_D(z) \equiv pf(D(z))$,

$$f'(s) = \frac{\rho}{s} p'_D(\hat{V}(s)) \hat{V}'(s) - \frac{\rho p_D(\hat{V}(s))}{s^2} \ ,$$

so that

$$f'(\eta) = \frac{\rho p'_D(\sigma^{-1}) \phi'(\eta) - 1}{\eta} \ . \tag{37}$$

As in Theorem 8,

$$p'(\sigma^{-1}) \equiv \frac{d}{dz} pf(A(z))\big|_{z = \sigma^{-1}} = \rho p'_D(\sigma^{-1}) E[V e^{\eta V}] = \rho p'(\sigma^{-1}) \phi'(\eta) \tag{38}$$

and the denominator of (30) is strictly positive. Next, by (34) and (35),

$$(s+\eta)\frac{(\pi - F(s))}{s}\left[I + \rho\frac{D(\hat{V}(s))}{s} - f(s)\tilde{r}(s)\tilde{l}(s)\right]$$

$$= (s+\eta)\left[M(s) - \frac{(\pi(1-f(s)) - x_0^t)\tilde{r}(s)\tilde{l}(s)f(s)}{s(1-f(s))}\right] \quad (39)$$

Taking the limit as $s \to -\eta$ in (39), we obtain

$$\left[\lim_{s \to -\eta}(s+\eta)\frac{(\pi - F(s))}{s}\right]\left[I - \rho\frac{D(\sigma^{-1})}{\eta} + \tilde{r}(-\eta)\tilde{l}(-\eta)\right] = \frac{x_0^t\tilde{r}(-\eta)\tilde{l}(-\eta)}{\eta f'(-\eta)}\ , \quad (40)$$

where the matrix on the left is nonsingular by Lemma 5 and the right side is strictly positive and finite as in (36). We then note that $\tilde{l}(-\eta) = l(\sigma^{-1})$ and $\tilde{r}(-\eta) = r(\sigma^{-1})$. We take the inverse in (40) to obtain (30). By Lemma 5, $l(\sigma^{-1})Y(\sigma^{-1}) = l(\sigma^{-1})$. Hence $l(\sigma^{-1}) = l(\sigma^{-1})Y(\sigma^{-1})^{-1}$, and we can delete the $Y(\sigma^{-1})$. We apply (29) and (30) to obtain (31)–(33). ■

We conclude this section by stating some relations among the asymptotic constants that follow from theorems 8–10.

Theorem 11. *(a) In the BMAP/GI/1 queue, if the limits exist, then*

$$\beta = \alpha_L\ ,\ \beta^a = \alpha_W \quad (41)$$

and

$$\frac{\beta^a}{\beta} = \frac{l^t(D-D_0)e}{l^te} = \frac{l^L(D-D_0)e}{l^Le} = \frac{l^{W^b}e}{l^Le} = l(\sigma^{-1})(D-D_0)e\ . \quad (42)$$

(b) In the MAP/GI/1 queue,

$$\frac{\beta^a}{\beta} = \frac{\beta^d}{\beta} = \frac{l^tD_1e}{l^te} = \frac{l^LD_1e}{l^Le} = \frac{l^We}{l^Le} = \frac{\alpha_W}{\alpha_L} = l(\sigma^{-1})D_1e = \frac{\eta\sigma}{\rho(1-\sigma)}\ . \quad (43)$$

Proof. (a). Apply Theorems 8–10, recalling that $\beta = \sigma l^te/(1-\sigma)$ and $\beta^a = \sigma l^ae/(1-\sigma)$. (b) In the MAP/GI/1 queue, customers arrive one at a time, so that $\beta^a = \beta^d$ and $l^We = \alpha_W$. Then $D - D_0 = D_1$. Finally, by (26), $\beta^d/\beta = \eta\sigma/\rho(1-\sigma)$. Theorem 2 of [2] also shows that

$\alpha_W / \alpha_L = \eta\sigma / \rho(1-\sigma)$ in any G/GI/1 queue. ∎

Corollary. *(a) In the MAP/M/1 queue,* $\sigma = 1 - \eta$ *so that*

$$\frac{\beta^a}{\beta} = \frac{\alpha_W}{\alpha_L} = \frac{\sigma}{\rho} = \frac{1-\eta}{\rho} \ . \qquad (44)$$

(b) In the MAP/D/1 queue, $\sigma = e^{-\eta}$*, so that*

$$\frac{\beta^a}{\beta} = \frac{\alpha_W}{\alpha_L} = \frac{\eta e^{-\eta}}{\rho(1-e^{-\eta})} = \frac{-\sigma\log\sigma}{\rho(1-\sigma)} \ . \qquad (45)$$

Remark 6. In applications we are often interested in the steady-state sojourn time or response time T, i.e., the waiting time W plus the service time V. Theorem 1 of [2] shows that in any G/GI/1 queue if $e^{\eta x} P(W > x) \rightarrow \alpha_W$ as $x \rightarrow \infty$, then $e^{\eta x} P(T > x) \rightarrow \alpha_W / \sigma$. Moreover, the exponential approximation for the sojourn-time distribution is also remarkably good, even when the service-time distribution is not nearly exponential. ∎.

4. The MAP/MSP/1 Queue

In this section we briefly consider a related class of Markovian queueing models that allows dependence among the service times. In particular, we let the number of service completions during the first t units of time that the server is busy be a MAP that is independent of the arrival process. We call this service process a *Markovian service process* (MSP).

In the MAP/MSP/1 queue, it is easy to see that the queue length at an arbitrary time, together with the auxiliary phase state of both the arrival and the service process, can be represented as a *quasi-birth-and-death* (QBD) process, which is a specially structured continuous-time Markov chain, as in Chapter 3 of Neuts [17]. In particular, a QBD process has an infinitesimal generator of the form

$$\tilde{Q} = \begin{bmatrix} \tilde{B}_0 & \tilde{A}_0 & & & \\ \tilde{B}_1 & \tilde{A}_1 & \tilde{A}_0 & & \\ & \tilde{A}_2 & \tilde{A}_1 & \tilde{A}_0 & \\ & & \tilde{A}_2 & \tilde{A}_1 & \tilde{A}_0 \ldots \\ & & & \vdots & \end{bmatrix} \tag{46}$$

Let $D_0^{\uparrow}$ and $D_1^{\uparrow}$ be the matrices characterizing the MAP and let $D_0^{\downarrow}$ and $D_1^{\downarrow}$ be the matrices characterizing the MSP, as in (14), assuming that they each have overall rate 1. Let the overall service rate and arrival rate be 1 and ρ, respectively. Then, as in Whitt [27], it is easy to see that the matrices in (46) can be expressed using the Kronecker product $\otimes$ and sum $\oplus$ operations as

$$\tilde{A}_0 = I_1 \otimes \rho D_1^{\uparrow} \ , \ \tilde{A}_1 = D_0^{\downarrow} \oplus \rho D_0^{\uparrow} \text{ and } \tilde{A}_2 = D_1^{\downarrow} \otimes I_2 \ , \tag{47}$$

where I_1 and I_2 are identity matrices. By our assumption that the MSP operates only when the server is busy,

$$\tilde{B}_0 = I_1 \oplus \rho D_0^{\uparrow} \text{ and } \tilde{B}_1 = \tilde{A}_2 \ . \tag{48}$$

We can obtain alternative models by changing the definition of B_0. For example, we could let the phase process run without generating any real service completions. Then we would have $\tilde{B}_0 = D^{\downarrow} \oplus \rho D_0^{\uparrow}$.

Remark 7. Special cases of the MAP/MSP/1 are the M/M/1 and PH/PH/1 queues. For the M/M/1 queue, $\tilde{Q}$ in (46) reduces to the familiar infinitesimal generator of the continuous-time queue-length process, without any auxiliary states. The PH/PH/1 queue is a familiar QBD process; see §3.7 of Neuts [17]. The number of auxiliary states is $m_1 m_2$ where m_1 and m_2 are the number of phases in the interarrival time and service-time distributions.

Remark 8. The MAP/MSP/1 model has the MSP and MAP stochastically independent. Dependence can easily be introduced within the QBD framework. We would then typically not have the Kronecker structure in (47) and (48). ■

Since the MAP/MSP/1 queue produces a Markov chain of GI/M/1 type, it has the asymptotic behavior in (4), where in this case the rate matrix R is the minimal nonnegative solution to the equation

$$R^2 \tilde{A}_2 + R\tilde{A}_1 + \tilde{A}_0 = 0 \ . \tag{49}$$

The asymptotic decay rate σ is then the Perron-Frobenius eigenvalue of R, $pf(R)$, which upon multiplying by the associated eigenvector in (49) we see satisfies the equation

$$pf(\tilde{A}(\sigma)) = 0 \ , \tag{50}$$

where $\tilde{A}(z) = x^2 A_2 + zA_1 + A_0$. As with $D(z)$ in §3, the Perron-Frobenius eigenvalue of $A(z)$ is well defined since $A(z)$ has nonnegative off-diagonal elements. By the same argument as for Proposition 7, which follows Lemma 2.3.4 of Neuts [20], we have the following result.

Proposition 12. *In a QBD process if $\tilde{A}(1)$ is irreducible, then the Perron-Frobenius eigenvalue $pf(\tilde{A}(z))$ is a strictly increasing convex function of z with $pf(A(1)) = 0$, so that the equation $pf(A(z)) = 0$ has at most one root σ with $0 < \sigma < 1$.*

We apply Proposition 12 to establish asymptotics for the MAP/MSP/1 queue.

Theorem 13. *In the MAP/MSP/1 queue, if $\tilde{Q}$ is irreducible and positive recurrent, and $D^{\uparrow}(1)$ and $D^{\downarrow}(1)$ are irreducible, then*

$$x_i = x_0 R^i = \sigma^i (x_0 \hat{r}(\sigma)) \hat{l}(\sigma) + o(\sigma^i) \ \ as \ i \to \infty \ , \tag{51}$$

where the asymptotic decay rate σ is the unique root with $0 < \sigma < 1$ of the equation $pf(\tilde{A}(z)) = 0$ or, equivalently, the equation

$$pf(D^{\downarrow}(z)) = -pf(\rho D^{\uparrow}(1/z)) \ , \tag{52}$$

and $\hat{l}(\sigma)$ and $\hat{r}(\sigma)$ are left and right eigenvectors of $A(\sigma)$ associated with $pf(A(\sigma))$ such that $\hat{l}(\sigma) r(\sigma) = \hat{l}(\sigma) e = 1$.

Proof. It is easy to see that $\tilde{A}(1)$ is irreducible if $D^{\uparrow}(1)$ and $D^{\downarrow}(1)$ are. Thus, by Proposition

12, the equation $pf(\tilde{A}(z)) = 0$ has at most one root with $0 < z < 1$. By Theorem 3.1.1 of [17], this root σ exists and satisfies $0 < \sigma < 1$, and (51) is valid. As shown in [27], from basic properties of the Kronecker operations, we obtain

$$
\begin{aligned}
pf(\tilde{A}(z)) &= pf\left[z(I_1 \times \rho D_1^{\uparrow}) + (I_1 \oplus \rho D_0^{\uparrow}) + (D_0^{\downarrow} \otimes I_2) + \frac{(D_1^{\downarrow} \otimes I_2)}{z} \right] \\
&= pf((I_1 \otimes (\rho z D_1^{\uparrow} + \rho D_0^{\uparrow})) + ((D_0^{\downarrow} + \frac{D_1^{\downarrow}}{z}) \otimes I_2)) \\
&= pf(z \rho D_1^{\uparrow} + \rho D_0^{\uparrow}) + pf(D_0^{\downarrow} + \frac{D_1^{\downarrow}}{z})) \\
&= pf(D^{\uparrow}(z)) + pf(D^{\downarrow}(1/z)) \ . \qquad (53)
\end{aligned}
$$

By Proposition 7, note that $pf(D^{\uparrow}(z))$ and $pf(D^{\downarrow}(z))$ are increasing convex functions of z. It then follows directly that $pf(D^{\downarrow}(1/z))$ is a convex function of z as well. ■

We can easily obtain related results for the steady-state distributions just before arrivals and just after departures by applying [15]. For this purpose, let x_{ij}^t denote the steady-state probability that the queue length is i and the (joint arrival and service) phase is j at an arbitrary time, and similarly for x_{ij}^a and x_{ij}^d. Then, paralleling (24), we obtain

$$
x_{ij}^a = x_{ij}^t \sum_k \tilde{A}_0(j,k) \ \text{ and } \ x_{ij}^d = x_{ij}^t \sum_k \tilde{A}_2(j,k) \ , \qquad (54)
$$

from [15], using time reversal in §4 there for x_{ij}^d. Hence,

$$
x_i^a e = x_i^t \tilde{A}_0 e \ \text{ and } \ x_i^d e = x_i^t \tilde{A}_2 e \ , \qquad (55)
$$

from which the asymptotics for x_i^a and x_i^d follow. Moreover, since the long-run flow rate from level i up (down) is $x_i^t \tilde{A}_0 e$ ($x_i^t \tilde{A}_2 e$), we have $x_i^a e = x_i^d e$, as we should.

5. A Numerical Example

To illustrate and confirm our BMAP/GI/1 results in §3, we conclude with a numerical example. We consider a relatively simple MAP, in particular, a two-state Markov modulated

Poisson process ($MMPP_2$). As on p. 37 of Lucantoni [13], an MMPP is a BMAP with $D_0 = M - \Lambda$, $D_1 = \Lambda$ and $D_j = 0$ for $j \geq 2$, where M is the infinitesimal-generator matrix of the Markovian environment process and Λ is the associated Poisson rate matrix. With two phases,

$$M = \begin{bmatrix} -m_0 & m_0 \\ m_1 & -m_1 \end{bmatrix} \text{ and } \Lambda = \begin{bmatrix} \lambda_0 & 0 \\ 0 & \lambda_1 \end{bmatrix} , \tag{56}$$

where m_0, m_1, λ_0 and λ_1 are positive constants and $m_0 \lambda_1 + m_1 \lambda_0 = m_0 + m_1$, so that the arrival rate is 1. The overall arrival rate is then ρ where ρ is specified separately. Then

$$D(z) = \begin{bmatrix} -m_0 + \lambda_0 (z-1) & m_0 \\ m_1 & -m_1 + \lambda_1 (z-1) \end{bmatrix} . \tag{57}$$

As usual, the Perron-Frobenius eigenvalue $pf(D(z)) \equiv p_D(z)$ can be found by solving the characteristic equation. Let $x = z - 1$, $m = m_0 + m_1$, $s = \lambda_0 + \lambda_1$ and $d = \lambda_0 - \lambda_1$. Then the characteristic equation is

$$\det(D(z) - \gamma I) = \gamma^2 - \gamma(sx - m) + \lambda_0 \lambda_1 x^2 - mx = 0 , \tag{58}$$

yielding

$$p_D(z) = \gamma(x) = \frac{sx - m + \sqrt{d^2 x^2 + 2mx(2-s) + m^2}}{2} . \tag{59}$$

In this context, the key equations in (17) are

$$\gamma(\sigma^{-1} - 1) = \frac{\eta}{\rho} \text{ and } Ee^{\eta V} = \sigma^{-1} \tag{60}$$

for $\gamma(x)$ in (59).

We now consider a concrete example. We verify the equations in this paper by independently calculating the asymptotic parameters by the moment-based numerical inversion procedure in Choudhury and Lucantoni [6]. We also calculate the probability distributions themselves using the algorithms in Lucantoni [13], together with the numerical transform inversion in Abate and Whitt [3], as implemented by Choudhury [5].

For our concrete example, consider the two-point service-time distribution:

$$P(V = 11) = 0.01960784 = 1 - P(V = 0.8) .$$

This distribution has first two moments $EV = 1$ and $EV^2 = 3$. Let $\rho = 0.7$, $\rho \lambda_0 = 1.1$, $\rho \lambda_1 = 0.3$ and $\rho m_0 = \rho m_1 = 0.1$, so that the overall arrival rate is 0.7, but the instantaneous arrival rate in phase 0 is 1.1, which is greater than the overall service rate of 1. Then $m = 0.285714$, $s = 2.0$ and $d = 1.142857$, so that (59) becomes

$$\gamma(x) = x - 0.142857 + 0.5\sqrt{1.306122x^2 + 0.0816326} . \qquad (61)$$

In this case the solution of (60) is $\eta = 0.1115972$ and $\sigma = 0.878066$, as is easy to check. Thus $p_D(\sigma^{-1}) = 0.159424$, $l(\sigma^{-1}) = (0.629544,\ 0.370456)$,

$$D(\sigma^{-1}) = \begin{bmatrix} 0.075362 & 0.142857 \\ 0.142857 & -0.083342 \end{bmatrix} \text{ and } r(\sigma^{-1}) = \begin{bmatrix} 1.179891 \\ 0.694306 \end{bmatrix} \qquad (62)$$

with $l(\sigma^{-1})$ and $r(\sigma^{-1})$ being chosen so that $l(\sigma^{-1}) r(\sigma^{-1}) = l(\sigma^{-1}) e = 1$.

To calculate l^d in (19), we also need $\phi'(\eta)$, $p_D'(\sigma^{-1})$ and g. (Recall that $(-x_0 D_0^{-1}) = (1-\rho) g$). First,

$$\phi'(\eta) = EVe^{\eta V} = 1.593677 . \qquad (63)$$

From (59),

$$p_D'(z) = \gamma'(x) = \frac{s}{2} + \frac{(d^2 x^2 + 2mx(2-s) + m^2)^{-1/2}}{4} (2d^2 x + 2m(2-s)) , \qquad (64)$$

so that here

$$p_D'(\sigma^{-1}) = \gamma'(0.138866) = 1.277475 . \qquad (65)$$

From [5], we obtain $g = (0.241185, 0.758815)$, so that $g\, r(\sigma^{-1}) = 0.811421$. Thus, we obtain from (19)

$$l^d = \frac{\eta(1-\rho) g\, r(\sigma^{-1}) l(\sigma^{-1})}{\rho(\rho p_D'(\sigma^{-1})\phi'(\eta) - 1)} = 0.091288 l(\sigma^{-1}) \tag{66}$$

and $l^a e = l^d e = 0.091287$. From (26), we obtain

$$l^t = \frac{\rho(1-\sigma)}{\sigma\eta} l^d = 0.0795156 l(\sigma^{-1}) \tag{67}$$

and $l^t e = 0.0795156$. We also see that (66) and (67) are consistent with (27), i.e., $l^a e = l^t D_1 e = l^d e$ as it should.

Next,

$$l^L = \frac{\sigma}{1-\sigma} l^t = 0.572608 l(\sigma^{-1}) \tag{68}$$

and

$$l^W e = \frac{\sigma l^a e}{1-\sigma} = 0.657379\ , \tag{69}$$

so that $\alpha_L = \beta$, $\alpha_W = \beta^a$ and

$$\frac{\alpha_L}{\alpha_W} = \frac{\beta}{\beta^a} = \frac{l^t e}{l^d e} = \frac{\rho(1-\sigma)}{\eta\sigma} = 0.871049 \tag{70}$$

as in Theorem 11.

Finally, Table 1 compares the exponential approximations $\beta\sigma^k$ and $\beta^a\sigma^k$ with the exact values $P(Q > k)$ and $P(Q^a > k)$ in this model. From Table 1 it is apparent that the exponential approximations are excellent at the 80th percentile and beyond. (Of course, in general, the point where the exponential approximations become good depends on the model.)

Acknowledgment. We are grateful to Dave Lucantoni for assistance.

References

[1] ABATE, J., G. L. CHOUDHURY and W. WHITT. 1992a. Exponential Approximations for Tail Probabilities in Queues, I: Waiting Times. submitted for publication.

[2] ABATE, J., G. L. CHOUDHURY and W. WHITT. 1992b. Exponential Approximations for Tail Probabilities in Queues, II: Sojourn Time, Workload and Queue Lengths, in preparation.

[3] ABATE, J. and W. WHITT. 1992. The Fourier-Series Method for Inverting Transforms of Probability Distributions. *Queueing Systems* 10, 5-88.

[4] BAIOCCHI, A. 1992. Asymptotic Behavior of the Loss Probability of the MAP/G/1/K Queue, Part I: Theory. INFOCOM Dept., University of Rome ‘‘La Sapienza.’’

[5] CHOUDHURY, G. L. 1992. An Algorithm for a Large Class of G/G/1 Queues, in preparation.

[6] CHOUDHURY, G. L. and D. M. LUCANTONI. 1992. Numerical Computation of the Moments of a Probability Distribution From Its Transforms, in preparation.

[7] DOETSCH, G. 1974. *Introduction to the Theory and Application of the Laplace Transformation,* Springer-Verlag, New York.

[8] ELWALID, A. I. and D. MITRA. 1992a. Effective Bandwidth of General Markovian Traffic Sources and Admission Control of High Speed Networks, to appear.

[9] ELWALID, A. I. and D. MITRA. 1992b. Markovian Arrival and Service Communication Systems: Spectral Expansions, Separability and Kronecker-Product Forms, in preparation.

[10] HUNTER, J. J. 1982. Generalized Inverses and Their Applications to Applied Probability Problems. *Lin. Alg. Appl.* 45, 157-198.

[11] KEMENY, J. G. and J. L. SNELL. 1960. *Finite Markov Chains,* Van Nostrand, Princeton.

[12] KINGMAN, J. F. C. 1961. A Convexity Property of Positive Matrices. *Quart. J. Math.* 12, 283-284.

[13] LUCANTONI, D. M. 1991. New Results on the Single Server Queue with a Batch Markovian Arrival Process. *Stochastic Models* 7, 1-46.

[14] LUCANTONI, D. M., K. S. MEIER-HELLSTERN and M. F. NEUTS. 1990. A Single-Server Queue with Server Vacations and a Class of Non-Renewal Arrival Processes. *Adv. Appl. Prob.* 22, 676-705.

[15] MELAMED, B. and W. WHITT. 1990. On Arrivals That See Time Averages. *Oper. Res.* 38, 156-179.

[16] NEUTS, M. F. 1979. A Versatile Markovian Point Process. *J. Appl. Prob.* 16, 764-779.

[17] NEUTS, M. F. 1981a. *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach.* The Johns Hopkins Press, Baltimore.

[18] NEUTS, M. F. 1981b. Stationary Waiting-Time Distributions in the GI/PH/1 Queue. *J. Appl. Prob.* 18, 901-912.

[19] NEUTS, M. F. 1986. The Caudal Characteristic Curve of Queues. *Adv. Appl. Prob.* 18, 221-254.

[20] NEUTS, M. F. 1989. *Structured Stochastic Matrices of M/G/1 Type and Their Applications,* Marcel Dekker, New York.

[21] NEUTS, M. F and Y. TAKAHASHI. 1981. Asymptotic Behavior of the Stationary Distributions in the GI/PH/C Queue with Heterogeneous Servers. *Z. Wahrscheinlichkeitsth.* 57, 441-452.

[22] RAMASWAMI, V. 1980. The N/G/1 Queue and its Detailed Analysis. *Adv. Appl. Prob.* 12, 222-261.

[23] RAMASWAMI, V. 1990. From the Matrix-Geometric to the Matrix-Exponential. *Queueing Systems* 6, 229-260.

[24] SENETA, E. 1981. *Nonnegative Matrices and Markov Chains,* second ed., Springer-Verlag, New York.

[25] TAKAHASHI, Y. 1981. Asymptotic Exponentiality of the Tail of the Waiting-Time Distribution in a PH/PH/c Queue. *Adv. Appl. Prob.* 13, 619-630.

[26] TIJMS, H. C. 1986. *Stochastic Modeling and Analysis: A Computational Approach.* John Wiley, New York.

[27] WHITT, W. 1992. Tail Probabilities with Statistical Multiplexing and Effective Bandwidths for Multi-Class Queues, submitted for publication.

[28] WILF, H. S. 1990. Generatingfunctionology. Academic Press, New York.

k	$P(Q^a > k)$		$P(Q > k)$	
	exact	approx.	exact	approx.
2	0.5039	0.5068	0.4366	0.4415
4	0.3660	0.3908	0.3223	0.3404
8	0.2267	0.2323	0.1989	0.2023
10	0.1790	0.1791	0.1563	0.1560
12	0.1395	0.1381	0.1214	0.1203
14	0.1076	0.1065	0.09348	0.09273
16	0.08249	0.08208	0.07174	0.07150
20	0.04869	0.04879	0.04242	0.04250
24	0.02898	0.02901	0.02525	0.02526
28	0.017250	0.017242	0.015024	0.015019
36	0.0060921	0.0060927	0.0053066	0.0053070
48	0.0012798	0.0012798	0.0011148	0.0011148

Table 1. A comparison of approximations with exact values for the steady-state queue-length tail probabilities in the $MMPP_2/D_2/1$ example in §5.

x	exact	$\alpha_L e^{-\eta x}$	$\hat{\alpha}_L(x)$	$\hat{\eta}(x)$
3.0	0.4278	0.4243	0.5931	9.178
6.0	0.31441	0.31431	0.5740	9.966
9.0	0.232846	0.232844	0.57279	9.9984
12.0	0.17249290	0.17249283	0.5727272	9.999554
18.0	0.094663802	0.094663802	0.572723848	9.99960019
24.0	0.051951350	0.051951350	0.572723841	9.99960026

Table 1. A comparison of exponential approximations with exact values of the workload tail probabilities, $P(L > x)$, in the $M/H_2^b/1$ queue with $\rho = 0.7$ and $c_s^2 = 4.0$ in Example 1. Also included are the local linear regression estimates of the asymptotic parameters.

x	exact	$\alpha_T e^{-\eta x}$	$\hat{\alpha}_T(x)$	$\hat{\eta}(x)$
3.0	0.4943	0.4849	0.7107	8.263
6.0	0.35947	0.35921	0.6581	9.921
9.0	0.266115	0.266108	0.6547	9.99667
12.0	0.1971358	0.1971356	0.6545	9.99949
18.0	0.108187743	0.108187743	0.654544812	9.9996010
24.0	0.059373268	0.059373268	0.654544803	9.99960026

Table 2. A comparison of exponential approximations with exact values of the sojourn-time tail probabilities, $P(T > x)$, in the $M/H_2^b/1$ queue with $\rho = 0.7$ and $c_s^2 = 4.0$ in Example 1. Also included are the local linear regression estimates of the asymptotic parameters.

x	workload exact	workload approx.	workload percent error	sojourn exact	sojourn approx.	sojourn percent error
3.0	0.3765	0.4097	8.8	0.4801	0.5356	11.6
6.0	0.2900	0.2931	1.0	0.3564	0.3833	7.6
9.0	0.2230	0.2097	−6.0	0.2884	0.2742	−4.9
12.0	0.1506	0.1501	−0.3	0.2033	0.1962	3.5
15.0	0.1049	0.1074	2.4	0.1355	0.1403	3.5
18.0	0.0771	0.0768	−0.4	0.0997	0.1004	0.7
21.0	0.0557	0.0550	−1.3	0.0733	0.0719	−1.9
24.0	0.03913	0.03932	0.5	0.05137	0.05142	0.1
27.0	0.02800	0.02814	0.5	0.03644	0.03678	0.9
30.0	0.02020	0.02013	−.03	0.02638	0.02632	−0.2
36.0	0.010304	0.01030	0.2	0.01344	0.01347	0.2
42.0	0.005282	0.005275	−0.1	0.006908	0.006898	−0.1
48.0	0.002699	0.002701	0.7	0.003528	0.003521	0.1
54.0	0.001383	0.001383	0.0	0.001808	0.001808	0.0
60.0	0.000708	0.000708	0.0	0.000925	0.000925	0.0

Table 3. A comparison of exponential approximations for the steady-state workload and sojourn-time tail probabilities with exact values in the $MMPP_2/D_2/1$ queue in Example 2.

k	exact	$(1-\sigma)\beta\sigma^{k-1}$	$(1-\sigma_{ap})\beta^*_{ap}\sigma^k_{ap}$
1	0.1384	0.0954	0.0787
2	0.0986	0.0835	0.0693
3	0.0780	0.0731	0.0611
4	0.0652	0.0639	0.0538
8	0.03694	0.03748	0.0324
12	0.02163	0.02197	0.0195
16	0.01268	0.01287	0.0118
20	0.00743	0.00755	0.00710
40	0.000514	0.000522	0.000565
50	0.000135	0.000137	0.000159

Table 4. A comparison of approximations with exact values of $P(Q = k)$ in the $M/H_2^b/1$ queue with $\rho = 0.7$ and $c_s^2 = 4.0$ in Example 1 revisited. Here $\sigma = 0.874996$, $\beta = 0.763625$ and $\beta(1-\sigma)/\sigma = 0.109095$.

k	$P(Q = k)$ exact	$P(Q = k)$ approx.	$P(Q > k)$ exact	$P(Q > k)$ approx.
2	0.06561	0.05383	0.4366	0.4415
4	0.03856	0.03644	0.3223	0.3404
8	0.02236	0.02467	0.1989	0.2023
12	0.01479	0.01467	0.1214	0.1203
16	0.008884	0.008718	0.07174	0.01715
20	0.005172	0.005182	0.04242	0.04250
30	0.001413	0.001412	0.01158	0.01158
40	0.0004380	0.0004381	0.003155	0.003155

Table 5. A comparison of approximations with exact values of $P(Q = k)$ and $P(Q > k)$ in the MMPP/D_2/1 queue with $\rho = 0.7$ in Example 2 revisited in §6.

k	exact	$\beta\sigma^k$	$\beta_{ap}\sigma_{ap}^k$
0	0.69138	0.6988	0.6979
1	0.60513	0.6077	0.6069
3	0.45968	0.45963	0.45901
7	0.26323	0.26288	0.26252
15	0.08612	0.08600	0.08587
23	0.028175	0.028134	0.028087
29	0.012188	0.012170	0.012148
40	0.002624	0.002618	0.002613
46	0.001136	0.001132	0.001130

Table 6. A comparison of approximations with exact values from pp. 24, 74 of Hillier and Yu (1981) of the tail probabilities $P(Q > k)$ in the M/E_2/4 model with $\rho = 0.9$ in Example 4. Here $\sigma = 0.869646$, $\beta = 0.69879$, $EQ = 5.3542$, $\beta_{ap} = EQ(1-\sigma) = 0.69794$ and $\sigma_{ap} = 0.86963$.

k	exact	$\beta\sigma^k$	$\beta_{ap}\sigma_{ap}^k$
0	0.36963	0.4031	0.3681
1	0.25167	0.2763	0.2522
2	0.17438	0.1894	0.1728
4	0.08264	0.08900	0.08112
8	0.01830	0.01965	0.01788
10	0.008600	0.00923	0.008394
12	0.004042	0.00434	0.003941
14	0.001900	0.00204	0.001850
16	0.000893	0.000958	0.000869

Table 7. A comparison of approximations with exact values from pp. 24, 74 of Hillier and Yu (1981) of the tail probabilities $P(Q > k)$ in the $M/E_2/4$ model with $\rho = 0.75$ in Example 4. Here $\sigma = 0.68548$, $\beta = 0.4031$, $EQ = 1.1691$, $\beta_{ap} = EQ(1-\sigma) = 0.36805$ and $\sigma_{ap} = 0.68519$.

k	exact	$\beta(1-\sigma)\sigma^k$
1	0.086259	0.0912
2	0.077288	0.07933
4	0.059657	0.06000
6	0.045318	0.045374
8	0.034306	0.034316
10	0.025951	0.025952
12	0.019626	0.019627
28	0.0021006	0.0021006
44	0.00022482	0.00022481
66	0.000010408	0.000010408

Table 8. A comparison of approximations and exact values from p. 74 of Hillier and Yu (1981) of the probability mass function values $P(Q = k)$ in the $M/E_2/4$ queue with $\rho = 0.9$ in Example 4.

random variable	asymptotic decay rate	asymptotic constant	supporting theory
W	η	α_W	part I
T	η	$\alpha_T = \alpha_W / \sigma$	Theorem 1
L	η	$\alpha_L = \rho(1-\sigma)\alpha_W / \eta\sigma$	Theorem 2
Q^a	$\sigma = Ee^{\eta V}$	$\beta^a = \eta\sigma^2\phi'(\eta)\alpha_W/(1-\sigma)$	Theorem 3
Q	σ	$\beta = \beta^a \alpha_L / \alpha_W = \rho\sigma\phi'(\eta)\alpha_W$	Corollary to Theorem 17
Q^d	σ	$\beta^d = \beta^a$ if no batches	Theorems 3 and 15
N^a	σ	β^a/σ	Remark 7
N	σ	β/σ	Remark 7
N^d	σ	β^d/σ	Remark 7

Table 9. A summary of the asymptotic parameters of the steady-state random variables in the G/GI/1 queue.