THE TRANSIENT M/G/1/0 QUEUE: SOME BOUNDS AND APPROXIMATIONS FOR LIGHT TRAFFIC WITH APPLICATION TO RELIABILITY

J. BEN ATKINSON
University of North London
School of Mathematical Sciences
Holloway Road, London N7 8DB, UK

ABSTRACT

We consider the transient analysis of the M/G/1/0 queue, for which $P_n(t)$ denotes the probability that there are no customers in the system at time t, given that there are n $(n = 0, 1)$ customers in the system at time 0. The analysis, which is based upon coupling theory, leads to simple bounds on $P_n(t)$ for the M/G/1/0 and M/PH/1/0 queues and improved bounds for the special case M/E$_r$/1/0. Numerical results are presented for various values of the mean arrival rate λ to demonstrate the increasing accuracy of approximations based upon the above bounds in light traffic, i.e., as $\lambda \to 0$. An important area of application for the M/G/1/0 queue is as a reliability model for a single repairable component. Since most practical reliability problems have λ values that are small relative to the mean service rate, the approximations are potentially useful in that context. A duality relation between the M/G/1/0 and GI/M/1/0 queues is also described.

Key words: Queues, Reliability, M/G/1/0, GI/M/1/0, M/PH/1/0, M/E$_r$/1/0, Coupling Theory, Transient Analysis, Light Traffic, Bounds, Approximations, Duality.

AMS (MOS) subject classifications: 60K25, 90B25.

1. Introduction

In this paper, we consider the transient analysis of the M/G/1/0 queue; i.e., the single-server queue with a Poisson arrival process, a general distribution of service times and no buffer. We denote by $P_n(t)$ the probability that there are no customers in the system at time t, given that there are n $(n = 0, 1)$ customers in the system at time 0. Additionally, let

λ^{-1} = mean interarrival time of customers $(0 \leq \lambda < \infty)$,
μ^{-1} = mean service time of customers $(0 < \mu < \infty)$,
$B(t) = Pr(\text{service time} \leq t)$,
$\bar{B}(t) = 1 - B(t)$.

We also assume that the residual service time of a customer in service at time 0 has a Radon-Nikodym derivative. It is clearly of the form $\mu \bar{B}(t)$, $t \geq 0$.

It is well-known [5] that, for the M/G/1/0 queue:

$$P_0(t) \underset{t \to \infty}{\longrightarrow} \mu/(\lambda + \mu), \quad P_1(t) \underset{t \to \infty}{\longrightarrow} \mu/(\lambda + \mu), \tag{1}$$

and, for the M/M/1/0 queue,

$$P_0(t) = \mu/(\lambda + \mu) + \lambda/(\lambda + \mu)e^{-(\lambda + \mu)t}, \tag{2}$$

and

$$P_1(t) = \mu/(\lambda + \mu) - \mu/(\lambda + \mu)e^{-(\lambda + \mu)t}. \tag{3}$$

Closed-form solutions, such as (2) and (3), are difficult to obtain for any but the simplest transient queueing models. Much of the early work in this area concentrated on the derivation of exact analytical results [3, 10] while, more recently, there has been an increased emphasis on the use of approximate or exact numerical techniques [8] and the calculation of bounds. For the M/G/1/0 queue, it has been shown [7] that, in the case of service times having the property of increasing failure rate, the rate of convergence of $P_0(t)$ and $P_1(t)$ to their common limit $\mu/(\lambda + \mu)$ is $O(\exp(-(\lambda + \mu)t))$.

An important application of the M/G/1/0 queue is as a reliability model, involving a repairable component that can be in one of two states, either "working" ($n = 0$) or "under repair" ($n = 1$). Such a system can also be described by an alternating renewal process [4]. In the analysis below, we obtain approximations for $P_0(t)$ and $P_1(t)$ that become more exact as $\lambda \to 0$. Since many practical reliability problems have λ values that are small relative to μ, the results are potentially useful in that context. Alternatively, they can be viewed as a contribution to the study of queueing loss systems under conditions of light traffic.

In the next section, we use coupling theory to obtain simple bounds on the probabilities $P_0(t)$ and $P_1(t)$ for the M/PH/1/0 and M/G/1/0 queues; in the former case, the service time has a phase-type distribution, which can be used as a general model for approximating a wide variety of empirical distributions [6]. This is followed by a method to improve the bounds, which is illustrated for the special case of an Erlang distribution of service times. We then give a duality relation between the M/G/1/0 and GI/M/1/0 queues in terms of the method of analysis employed in this paper. Finally, some numerical results are presented for the M/E$_r$/1/0 queue to demonstrate the accuracy of approximations based upon the derived bounds for small values of λ.

2. The M/PH/1/0 and M/G/1/0 Queues

Our model will make use of coupling theory [1, 12]. By a coupling of two stochastic processes $\{X_t\}$ and $\{X_t'\}$ with the same state space and the same time parameter set, we understand a realization of $\{X_t\}$, $\{X_t'\}$ on a common probability space with an associated random time $\tau(\tau < \infty)$ and the property that $X_t = X_t'$ on $\{\tau \le t\}$. We now consider the coupled processes to be two M/G/1/0 queueing systems, as described in Section 1. The servers will be referred to as server 1 and server 2 and, at time 0, we assume that server 1 is busy and server 2 is idle. After some finite time τ, the two servers will both be idle for the first time. It then follows from coupling theory [1, 12] that, for the M/G/1/0 queue:

$$|P_0(t) - P_1(t)| \le Pr(\tau > t).$$

Hence, using the steady-state probabilities (1) it can easily be shown that

$$|P_0(t) - \mu/(\lambda + \mu)| \le \lambda/(\lambda + \mu)Pr(\tau > t), \tag{4}$$

and

$$| P_1(t) - \mu/(\lambda + \mu) | \le \mu/(\lambda + \mu)Pr(\tau > t). \tag{5}$$

For most of this section, we will assume that the service time distribution $B(t)$ is of the phase type: i.e.,

$$B(t) = \sum_{i=1}^{m} \sum_{k=1}^{r} p_{ik}B_E(t; \beta_i, k),$$

where

$$B_E(t; \beta_i, k) = \sum_{j=k}^{\infty} (\beta_i t)^j \exp(-\beta_i t)/j!,$$

and $\beta_i > 0$, β_i distinct $(i = 1, 2, ..., m)$. The maximum number of service phases is r, and so, for at least one i $(i = 1, 2, ..., m)$, $p_{ir} > 0$. A service can be considered to be in state (i, k) when there are k $(k = 1, ..., r)$ residual, exponentially distributed service phases, each having a mean duration β_i^{-1} $(i = 1, 2, ..., m)$. When a server is idle, we denote its state by $(0, 0)$, and when a service commences, it enters state (i, k) with probability p_{ik}. Therefore, at time 0 server 1 is in state (i, k) with probability q_{ik}, where

$$q_{ik} = (\mu/\beta_i) \sum_{j=k}^{r} p_{ij}, \tag{6}$$

while server 2 is in state $(0, 0)$. In general, when server 1 is in state (i, j) and server 2 is in state (k, l), the combined system state is denoted by (i, j, k, l) and the probability that the system is in this state at time t is denoted by $P_t(i, j, k, l)$. The coupling time τ is thus the time at which the system of queues first enters the state $(0, 0, 0, 0)$. We will assume in our analysis, that this is an absorbing state, and hence we have:

$$Pr(\tau > t) = 1 - P_t(0, 0, 0, 0). \tag{7}$$

Our strategy will be to obtain a lower bound for $P_t(0, 0, 0, 0)$ and hence an upper bound for $Pr(\tau > t)$ to use in (4) and (5). The basis of our analysis will be the following state equations, in which $P'(\cdot)$ denotes the derivative with respect to t. When no confusion can arise, for brevity we omit the parameter t from the probabilities.

$0 < j, l < r$:

$$P'(i, j, k, l) = -(\beta_i + \beta_k)P(i, j, k, l) + \beta_i P(i, j+1, k, l) + \beta_k P(i, j, k, l+1)$$

$$+ \lambda p_{ij}P(0, 0, k, l) + \lambda p_{kl}P(i, j, 0, 0), \tag{8}$$

$0 < l < r$:

$$P'(i, r, k, l) = -(\beta_i + \beta_k)P(i, r, k, l) + \beta_k P(i, r, k, l+1)$$

$$+ \lambda p_{ir}P(0, 0, k, l) + \lambda p_{kl}P(i, r, 0, 0), \tag{9}$$

$0 < j < r$:

$$P'(i, j, k, r) = -(\beta_i + \beta_k)P(i, j, k, r) + \beta_i P(i, j+1, k, r)$$

$$+ \lambda p_{ij}P(0,0,k,r) + \lambda p_{kr}P(i,j,0,0), \tag{10}$$

$$P'(i,r,k,r) = -(\beta_i + \beta_k)P(i,r,k,r) + \lambda p_{ir}P(0,0,k,r) + \lambda p_{kr}P(i,r,0,0), \tag{11}$$

$0 < l < r$:

$$P'(0,0,k,l) = -(\lambda + \beta_k)P(0,0,k,l) + \beta_k P(0,0,k,l+1) + \sum_{i=1}^{m} \beta_i P(i,1,k,l), \tag{12}$$

$$P'(0,0,k,r) = -(\lambda + \beta_k)P(0,0,k,r) + \sum_{i=1}^{m} \beta_i P(i,1,k,r), \tag{13}$$

$0 < j < r$:

$$P'(i,j,0,0) = -(\lambda + \beta_i)P(i,j,0,0) + \beta_i P(i,j+1,0,0) + \sum_{k=1}^{m} \beta_k P(i,j,k,1), \tag{14}$$

$$P'(i,r,0,0) = -(\lambda + \beta_i)P(i,r,0,0) + \sum_{k=1}^{m} \beta_k P(i,r,k,1), \tag{15}$$

$$P'(0,0,0,0) = \sum_{i=1}^{m} \beta_i P(i,1,0,0) + \sum_{k=1}^{m} \beta_k P(0,0,k,1), \tag{16}$$

$j > 0$ (initial conditions):

$$P_0(i,j,0,0) = q_{ij}. \tag{17}$$

We now introduce Laplace transforms as follows:

$$L(i,j,k,l) = \int_0^\infty e^{-st} P(i,j,k,l)dt, \quad \Re(s) > 0.$$

In addition, let

$$b_{ik} = (s + \beta_i + \beta_k)^{-1}, \text{ and } c_k = (s + \lambda + \beta_k)^{-1}.$$

Hence, equations (8) to (17) become:

$0 < j, l < r$:

$$b_{ik}^{-1}L(i,j,k,l) = \beta_i L(i,j+1,k,l) + \beta_k L(i,j,k,l+1)$$
$$+ \lambda p_{ij}L(0,0,k,l) + \lambda p_{kl}L(i,j,0,0), \tag{18}$$

$0 < l < r$:

$$b_{ik}^{-1}L(i,r,k,l) = \beta_k L(i,r,k,l+1) + \lambda p_{ir}L(0,0,k,l) + \lambda p_{kl}L(i,r,0,0), \tag{19}$$

$0 < j < r$:

$$b_{ik}^{-1}L(i,j,k,r) = \beta_i L(i,j+1,k,r) + \lambda p_{ij}L(0,0,k,r) + \lambda p_{kr}L(i,j,0,0), \tag{20}$$

$$b_{ik}^{-1}L(i,r,k,r) = \lambda p_{ir}L(0,0,k,r) + \lambda p_{kr}L(i,r,0,0), \tag{21}$$

$0 < l < r$:

$$c_k^{-1} L(0,0,k,l) = \beta_k L(0,0,k,l+1) + \sum_{i=1}^{m} \beta_i L(i,1,k,l),$$ (22)

$$c_k^{-1} L(0,0,k,r) = \sum_{i=1}^{m} \beta_i L(i,1,k,r),$$ (23)

$0 < j < r$:

$$c_i^{-1} L(i,j,0,0) - q_{ij} = \beta_i L(i,j+1,0,0) + \sum_{k=1}^{m} \beta_k L(i,j,k,1),$$ (24)

$$c_i^{-1} L(i,r,0,0) - q_{ir} = \sum_{k=1}^{m} \beta_k L(i,r,k,1),$$ (25)

$$sL(0,0,0,0) = \sum_{i=1}^{m} \beta_i L(i,1,0,0) + \sum_{k=1}^{m} \beta_k L(0,0,k,1).$$ (26)

Now, let x_{ij}, y_{ij}, z_{ij} and $h(i,j)$, $(i = 1,2,\ldots,m;\ j = 1,2,\ldots,r)$ be defined as follows:

$$x_{ij} = L(i,j,0,0),\quad y_{ij} = L(0,0,i,j),\quad z_{ij} = x_{ij} + y_{ij},$$

$$h(i,j) = \sum_{\nu=1}^{m} \beta_\nu [L(\nu,1,i,j) + L(i,j,\nu,1)].$$

Hence equations (22) to (25) give for $0 < j < r$:

$$c_i^{-1} z_{ij} = \beta_i z_{i,j+1} + h(i,j) + q_{ij},$$ (27)

and

$$c_i^{-1} z_{ir} = h(i,r) + q_{ir}.$$ (28)

Solving (27) and (28) for z_{ij} we get:

$$z_{ij} = \beta_i^{-1} \sum_{k=j}^{r} (c_i \beta_i)^{k-j+1} [h(i,k) + q_{ik}].$$ (29)

From (26) we have

$$sL(0,0,0,0) = \sum_{i=1}^{m} \beta_i z_{i1}.$$ (30)

We can then write

$$L(0,0,0,0) = g_0(s) + g_1(s),$$ (31)

where

$$g_0(s) = s^{-1} \sum_{i=1}^{m} \sum_{k=1}^{r} (c_i \beta_i)^k q_{ik},$$ (32)

and

$$g_1(s) = s^{-1} \sum_{i=1}^{m} \sum_{k=1}^{r} (c_i \beta_i)^k h(i,k).$$ (33)

Clearly, the functions $g_0(s)$ and $g_1(s)$ are Laplace transforms of bounded, nonnegative functions of λ and t, for which we shall use the notation $O_s(1)$. Hence, inverting $g_0(s)$ in (31), we get

$$P_t(0,0,0,0) \geq f_0(\lambda, t)$$ (34)

where

$$f_0(\lambda, t) = \sum_{i=1}^{m} \sum_{k=1}^{r} q_{ik} \left(\frac{\beta_i}{\beta_i + \lambda}\right)^k B_E(t; (\beta_i + \lambda), k).$$ (35)

Therefore, using (4), (5), (7) and (34), $P_0(t)$ and $P_1(t)$ can be bounded as follows:

$$| P_0(t) - \mu/(\lambda+\mu) | \leq \lambda/(\lambda+\mu)[1 - f_0(\lambda,t)], \tag{36}$$

and

$$| P_1(t) - \mu/(\lambda+\mu) | \leq \mu/(\lambda+\mu)[1 - f_0(\lambda,t)]. \tag{37}$$

We can give the following probabilistic interpretation of (35). Consider the more general system in which two M/G/1/0 queues are coupled as above. Let t^* be the departure time of the customer that was in service at time 0; let $n(t)$ be the number of customer arrivals during the interval $(0, t)$; and let $p(\lambda, t)$ be the probability $Pr(0 \leq t^* \leq t \cap n(t^*) = 0)$. By assumption, the residual service time of the customer in service at time 0 has the density $\mu \bar{B}(t)$, and so

$$p(\lambda, t) = \int_0^t \mu \exp(-\lambda t) d\bar{B}(t).$$

For the particular case of a phase-type distribution of service time, calculation shows that $p(\lambda, t) = f_0(\lambda, t)$. Clearly, for the M/G/1/0 queue, $Pr(\tau > t) \leq 1 - p(\lambda, t)$, and so, in this case, the bounds (36) and (37) apply if $f_0(\lambda, t)$ is replaced by $p(\lambda, t)$.

Returning to the M/PH/1/0 queue, we clearly expect $f_0(\lambda, t)$ to be a good approximation for $P_t(0,0,0,0)$ only when λ is small. Evidently, the error introduced as a consequence of using bounds (36) and (37) to approximate $P_0(t)$ and $P_1(t)$ respectively, instead of using the tighter bounds (4) and (5), tends to zero as $\lambda \to 0$. We can show that $f_0(\lambda, t)$ converges uniformly to $f_0(0, t)$ by showing that the function $g_1(s)$ in (33) can be written in the form

$$g_1(s) = \lambda O_s(1). \tag{38}$$

In establishing such a form for $g_1(s)$, we shall also obtain the basis for calculating improved bounds for $P_0(t)$ and $P_1(t)$ to be described in the next section.

We now examine the function $h(i, k)$. To this end, we partially solve equations (18)-(21) to obtain $L(i, j, k, l)$ in terms of the simpler transforms $x_{..}$ and $y_{..}$. First, using (19) and (21) and induction on l, we obtain

$0 < l < r$:

$$L(i, r, k, l) = \lambda b_{ik} \sum_{\nu = l}^{r} (\beta_k b_{ik})^{\nu - l} [p_{ir} y_{k\nu} + p_{k\nu} x_{ir}]. \tag{39}$$

Similarly, by using (20) and (21) and induction on j, or by applying an obvious symmetry argument to (39), we obtain

$0 < j < r$:

$$L(i, j, k, r) = \lambda b_{ik} \sum_{\nu = j}^{r} (\beta_i b_{ik})^{\nu - j} [p_{kr} x_{i\nu} + p_{i\nu} y_{kr}]. \tag{40}$$

We can then use (18) and induction on j to obtain

$0 < j, l < r$:

$$L(i, j, k, l) = \beta_k b_{ik} \sum_{\nu = 0}^{r - j - 1} (\beta_i b_{ik})^{\nu} L(i, j + \nu, k, l + 1) + (\beta_i b_{ik})^{r - j} L(i, r, k, l)$$

$$+ \lambda b_{ik} \sum_{\nu = 0}^{r - j - 1} (\beta_i b_{ik})^{\nu} [p_{i, j + \nu} y_{kl} + p_{kl} x_{i, j + \nu}]. \tag{41}$$

Inspection of (21) and (39)-(41) shows that, for $0 < j$, $l \le r$, we have $L(i, j, k, l) = \lambda O_s(1)$, and so, recalling the definitions of $h(i, k)$ and $g_1(s)$, it is clear that (38) is also established: i.e., $g_1(s) = \lambda O_s(1)$. We thus expect the simple bounds (36) and (37) to approximate $P_0(t)$ and $P_1(t)$ more closely as λ is reduced towards zero. However, before checking this numerically, we will make use of the above analysis to obtain improved bounds for $P_0(t)$ and $P_1(t)$, illustrating the approach for the particular case in which the service time has an Erlang-r distribution.

3. Improved Bounds for the $M/E_r/1/0$ Queue

For an Erlang-r distribution of service times, we can simplify our notation as follows:

$$m = 1, \beta_1 = \beta, b_{11} = b, c_1 = c;$$

$$p_{1r} = 1, p_{1k} = 0 \quad (k \ne r);$$

$$q_{1k} = 1/r \quad (k = 1, \ldots, r);$$

$$x_{1j} = x_j, \ y_{1j} = y_j, \ z_{1j} = z_j \quad (j = 1, \ldots, r);$$

$$L(1, j, 1, k) = L(j, k) \quad (j = 1, \ldots, r; \ k = 1, \ldots, r);$$

$$L(0, 0, 0, 0) = L(0, 0), \quad P_t(0, 0, 0, 0) = P_t(0, 0);$$

$$h(1, j) = h(j) \quad (j = 1, \ldots, r).$$

Equations (39)-(41) become:

$0 < l < r$:

$$L(r, l) = \lambda b \left\{ (b\beta)^{r - l} x_r + \sum_{\nu = l}^{r} (b\beta)^{\nu - l} y_{\nu} \right\}, \tag{42}$$

$0 < j < r$:

$$L(j, r) = \lambda b \left\{ (b\beta)^{r - j} y_r + \sum_{\nu = j}^{r} (b\beta)^{\nu - j} x_{\nu} \right\}, \tag{43}$$

$0 < j, l < r$:

$$L(j, l) = \sum_{\nu = 0}^{r - j - 1} (b\beta)^{\nu + 1} L(j + \nu, l + 1) + (b\beta)^{r - j} L(r, l). \tag{44}$$

Using (42)-(44) and induction on l, we can obtain the following partial solution for $L(j, l)$ in terms of the simpler transforms x_0 and y_0:

$0 < j, l < r$:

$$L(j, l) = \frac{\lambda}{\beta} \sum_{k = 0}^{r - l} \binom{2r - j - l - k}{r - j} (b\beta)^{2r - j - l - k + 1} y_{r - k}$$

$$+\frac{\lambda}{\beta}\sum_{k=0}^{r-j}\binom{2r-j-l-k}{r-l}(b\beta)^{2r-j-l-k+1}x_{r-k}. \tag{45}$$

From (31)-(33) it is clear that

$$sL(0,0)=\sum_{k=1}^{r}(c\beta)^{k}[h(k)+\tfrac{1}{r}], \tag{46}$$

where

$$h(k)=\beta[L(1,k)+L(k,1)]. \tag{47}$$

Using (45) and (47), and defining the function $Q_{kp}(z_p)$ as follows:

$$Q_{kp}(z_p)=\sum_{p=k}^{r}\binom{r+p-k-1}{r-1}(b\beta)^{r+p-k}z_p+\sum_{p=1}^{r}\binom{r+p-k-1}{r-k}(b\beta)^{r+p-k}z_p, \tag{48}$$

we can easily show that

$$h(k)=\lambda Q_{kp}(z_p). \tag{49}$$

Hence (29) and (46) become

$$z_j=\beta^{-1}\sum_{k=j}^{r}(c\beta)^{k-j+1}[\lambda Q_{kp}(z_p)+\tfrac{1}{r}], \tag{50}$$

and

$$sL(0,0)=\sum_{k=1}^{r}(c\beta)^{k}[\lambda Q_{kp}(z_p)+\tfrac{1}{r}]. \tag{51}$$

Then, using (50) and (51), we obtain

$$sL(0,0)=\frac{1}{r}\sum_{k=1}^{r}(c\beta)^{k}\left[1+\left(\frac{\lambda}{\beta}\right)Q_{kp}\left(\sum_{\nu=p}^{r}(c\beta)^{\nu-p+1}\right)+\left(\frac{\lambda}{\beta}\right)^{2}O_s(1)\right]. \tag{52}$$

Our improved bound for $P_t(0,0)$ is, therefore,

$$P_t(0,0)\geq f_0(\lambda,t)+(\lambda/\beta)f_1(\lambda,t), \tag{53}$$

where $f_1(\lambda,t)$ can be found by inverting the following Laplace transform:

$$\frac{1}{rs}\sum_{k=1}^{r}(c\beta)^{k}Q_{kp}\left(\sum_{\nu=p}^{r}(c\beta)^{\nu-p+1}\right). \tag{54}$$

Using (48) and carrying the inversion, we obtain, for the case $\beta\neq\lambda$:

$$f_1(\lambda,t)=\frac{1}{r}\sum_{k=1}^{r}\left[\sum_{p=r}^{2r-k}\binom{p-1}{r-1}\sum_{\nu=1+k}^{2r-p+1}R(p,\nu)+\sum_{p=r+1-k}^{2r-k}\binom{p-1}{r-k}\sum_{\nu=1+k}^{2r-p+1}R(p,\nu)\right] \tag{55}$$

where

$$R(p, \nu) = \left(\frac{\beta}{\lambda + \beta}\right)^{\nu} 2^{-p}$$

$$+ \sum_{i=0}^{\nu-1} \binom{\nu + p - 2 - i}{p - 1} P^*(i, (\lambda + \beta)t)(-1)^{\nu - i} \left(\frac{\lambda + \beta}{\beta}\right)^{-i-1} \left(\frac{\beta - \lambda}{\beta}\right)^{-\nu - p + 1 + i} \quad (56)$$

$$+ \sum_{i=0}^{p-1} \binom{\nu + p - 2 - i}{\nu - 1} P^*(i, 2\beta t)(-1)^{k-i}(2)^{-i-1}\left(\frac{\lambda - \beta}{\beta}\right)^{-\nu - p + 1 + i},$$

and

$$P^*(i, t) = \sum_{j=0}^{i} \frac{t^j}{j!} e^{-t}. \quad (57)$$

Thus we can write:

$$| P_0(t) - \mu/(\lambda + \mu) | \le \lambda/(\lambda + \mu)[1 - f_0(\lambda, t) - (\lambda/\beta)f_1(\lambda, t)] \quad (58)$$

and

$$| P_1(t) - \mu/(\lambda + \mu) | \le \mu/(\lambda + \mu)[1 - f_0(\lambda, t) - (\lambda/\beta)f_1(\lambda, t)]. \quad (59)$$

4. The Dual Queue GI/M/1/0

We can mention a duality relation between the M/G/1/0 and GI/M/1/0 queues. Consider the following changes to the two coupled queues described above: (*i*) interchange the distributions of interarrival time and service time so that the interarrival time now has the distribution $B(t)$ with mean μ^{-1} and the service time has an exponential distribution with mean λ^{-1}, and (*ii*) let the coupling time τ now be the first time at which both servers are busy. Under these changed conditions, it is nevertheless clear that the distribution of τ remains unchanged. Using steady-state results for the GI/M/1/0 queue [11] and coupling theory, the following bounds (analogues of (4) and (5)) apply for the GI/M/1/0 queue,

$$| P_0(t) - K | \le (1 - K)Pr(\tau > t), \quad (60)$$

and

$$| P_1(t) - K | \le K\,Pr(\tau > t), \quad (61)$$

where

$$K = 1 - (\mu/\lambda)(1 - \int_0^{\infty} \exp(-\lambda t)dB(t)).$$

Here, K is the steady-state probability that the server is idle. We could also write down analogues of the bounds (36)-(37) and (58)-(59) since $f_0(\lambda, t)$ and $f_1(\lambda, t)$ are also unchanged in going over to the dual system. With the roles of λ and μ reversed, clearly the bounds will now be asymptotically exact under conditions of heavy traffic, i.e., as $\lambda \to 0$. Approximations based upon such bounds are unlikely to have practical application in the reliability context.

5. Some Numerical Results

A FORTRAN program was written, using double-precision arithmetic, to test the use of upper bounds based upon (36) and (58) to calculate approximate values of $P_0(t)$ for small values of λ in the $M/E_r/1/0$ queue. Taking $\mu = 1$ and λ values in the range $[0.001, 0.5]$, the following estimates of $P_0(t)$ were obtained for a sufficiently large set of regularly-spaced, discrete time-points t, starting at $t = 0$.

(*i*) Numerically exact values were obtained by solving the appropriate state equations for the $M/E_r/1/0$ queue, using standard numerical software [9]. These values are referred to as EXACT.

(*ii*) Simple upper bounds were calculated using (36) and referred to as APPROX 1.

(*iii*) Improved upper bounds were calculated using (58) and referred to as APPROX 2.

(*iv*) Numerically exact values of $Pr(\tau > t)$ were obtained by solving the state equations (8)-(17) using standard numerical software [9], leading to calculation of an upper bound based upon (4), and this was referred to as APPROX 3.

(*v*) Exact values were calculated for the corresponding $M/M/1/0$ queue, using equation (2), and referred to as MM1.

It is clear that, for any value of $t(t \geq 0)$, the first four estimates above are ordered as follows:

$$\text{EXACT} \leq \text{APPROX 3} \leq \text{APPROX 2} \leq \text{APPROX 1}.$$

We are interested in how well the simple bound APPROX 1 performs, and how much improvement can be obtained by using the more complex bound APPROX 2. The estimate APPROX 3 gives an upper limit to the accuracy that could be obtained by using the basic inequality (4) and the particular method of analysis employed in this paper, for example by including higher order terms in (53). Finally, MM1 is an easily computed approximation which can serve as a benchmark for our comparisons. For all of the $M/E_r/1/0$ systems studied numerically, MM1 was also found to be an upper bound for $P_0(t)$.

An example of the above results is presented graphically in Figure 1, which shows the estimates of $P_0(t)$ for the case $r = 5$ (i.e., the $M/E_5/1/0$ queue) with $\lambda = 0.1$ and $\mu = 1$. For comparative purposes, it is useful to define an overall measure of accuracy that does not depend upon the time parameter t, and then to investigate numerically how this accuracy depends on λ. To do this, we use the maximum absolute error MAE, given by

$$\text{MAE} = \text{MAX}_t(e(t) - \text{EXACT}(t)),$$

where $e(t)$ represents the estimate under consideration (e.g., APPROX 1). MAE values, correct to five decimal places, are given in Table 1 for a range of r and λ values. As expected, the MAE values fall sharply as λ is reduced. For $\lambda \leq 0.1$, we see that APPROX 2 and APPROX 3 are, to the prescribed accuracy, almost equal in value, and so the benefit of trying to improve APPROX 2 by including additional terms of the form $f_i(\lambda, t)$ $(i > 1)$ in equation (53) would be negligible. In addition, for $\lambda \leq 0.1$, the MAE values for APPROX 1 and APPROX 2 are considerably better than the benchmark estimate MM1, the relative improvement increasing as λ decreases.
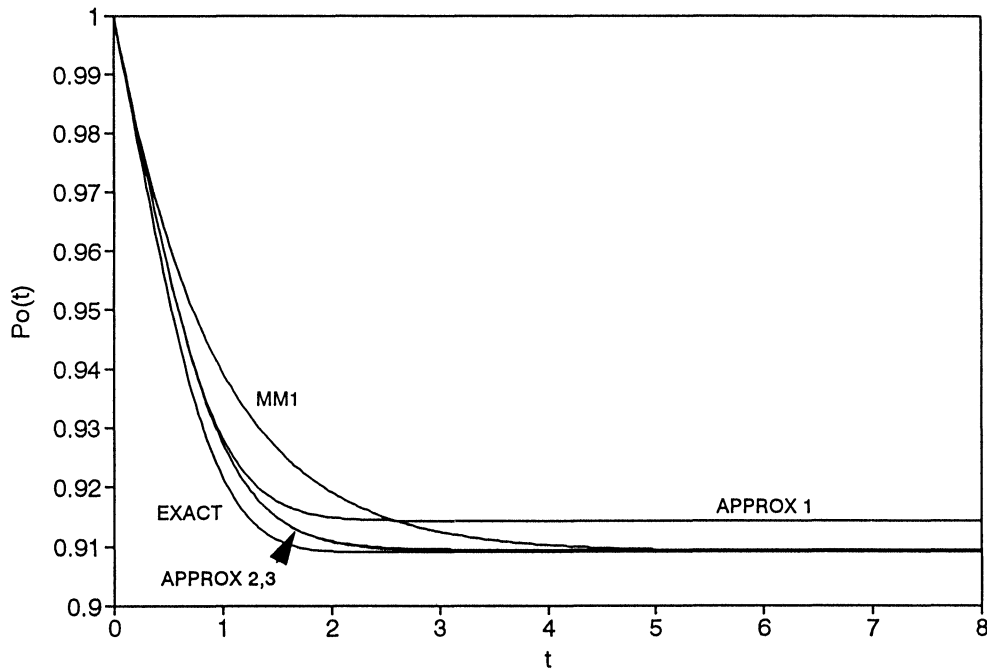
Figure 1. Approximate results for the $M/E_5/1/0$ queue with $\lambda = 0.1$, $\mu = 1.0$.

Table 1. MAE values for some $M/E_r/1/0$ queues.

r	λ	APPROX 1	APPROX 2	APPROX 3	MM1
1	0.500	0.11111	0.06963	0.06876	-
	0.100	0.00826	0.00446	0.00444	-
	0.010	0.00010	0.00005	0.00005	-
	0.001	0.00000	0.00000	0.00000	-
2	0.500	0.09831	0.07671	0.07645	0.03212
	0.100	0.00682	0.00494	0.00494	0.00916
	0.010	0.00008	0.00006	0.00006	0.00101
	0.001	0.00000	0.00000	0.00000	0.00010
5	0.500	0.10020	0.08879	0.08874	0.06405
	0.100	0.00675	0.00575	0.00575	0.01793
	0.010	0.00008	0.00007	0.00007	0.00196
	0.001	0.00000	0.00000	0.00000	0.00020

A more detailed analysis of the above results is given in reference [2].

6. Concluding Remarks

In this paper, we have presented some bounds for the state probabilities in the transient M/G/1/0 queue. We have obtained simple bounds on $P_n(t)$ for the M/G/1/0 and M/PH/1/0 queues and improved bounds for the special case $M/E_r/1/0$. Numerical results have been presented for various values of the mean arrival rate λ to demonstrate the increasing accuracy of approximations based upon the above bounds as $\lambda \to 0$. Such approximations are likely to be of practical use when the traffic intensity (λ/μ) is sufficiently small; i.e., less than 0.1. Our results are, therefore, applicable to the modelling of the reliability of a single repairable component, and to queueing loss systems under conditions of light traffic. In addition, the existence of comparable bounds for the dual system GI/M/1/0 was also noted.

Some possible extensions of the research described here could include the following:

(i) Obtain improved bounds of the form of (58) and (59) for other widely-used service-time distributions, such as the hyperexponential distribution.

(ii) Modify the model so that the coupling time τ becomes the first time that both servers occupy identical states; i.e., allowing all states of the form (i, j, i, j) to become absorbing states. This would, in principle, give rise to tighter bounds on $P_n(t)$, but the analysis is likely to be more complicated.

(iii) Carry out numerical comparisons of the bounds obtained here with similar bounds derived elsewhere (e.g., in reference [7]).

Acknowledgement

References

[1]   Asmussen, S., *Applied Probability and Queues*, John Wiley, Chichester 1987.

[2]   Atkinson, J.B., *Numerical approximation of the transient M/G/1/0 queue in light traffic*, Internal Report, School of Mathematical Sciences, University of North London 1995.

[3]   Cohen, J.W., *The Single-Server Queue*, John Wiley, New York 1969.

[4]   Cox, D.R., *Renewal Theory*, Methuen, London 1962.

[5]   Gross, D. and Harris, C.M., *Fundamentals of Queueing Theory*, 2nd Edition, John Wiley, New York 1985.

[6]   Johnson, M.A. and Taaffe, M.R., An investigation of phase-distribution moment-matching algorithms for use in queueing models, *Queueing Sys.* **8** (1991), 129-148.

[7]   Kovalenko, I.N. and Birolini, A., *Uniform exponential bounds for the pointwise availability of a repairable system*, A manuscript, School of Mathematical Sciences, University of North London 1995.

[8]   Odoni, A.R. and Roth, E., An empirical investigation of the transient behavior of stationary queueing systems, *Opns. Res.* **31** (1983), 432-455.

[9]   Press, W.H., et. al, *Numerical Recipes in FORTRAN*, 2nd Edition, Cambridge University Press 1992.

[10]   Takács, L., The time dependence of a single-server queue with Poisson input and general service times, *Ann. Math. Statist.* **33** (1962), 1340-1348.

[11]   Takács, L, *Introduction to the Theory of Queues*, Oxford University Press, New York 1962.

[12]   Thorisson, H., The coupling of regenerative processes, *Adv. Appl. Prob.* **15** (1983), 531-547.