ASYMPTOTIC TRANSIENT BEHAVIOUR OF THE BULK SERVICE QUEUE

B. D. CRAVEN

(received 14 November 1962)

1. Introduction

Various authors have studied the transient behaviour of single-server queues. Notably, Takacs [13], [14] has analysed a queue with recurrent input and exponential service time distributions, Keilson and Kooharian [9], [10] and Finch [5] have considered a queue with general independent input and service times, Finch [6] has analysed a queue with non-recurrent input and Erlang service, and Jaiswal [8] has considered the bulk-service queue with Poisson input and Erlang service. Except for Finch [6], these authors have made the usual assumptions of independence for successive intervals between arrivals and for service times, as formulated by Lindley [12]. The transient distributions for queue length or waiting time have been expressed, usually, as Laplace transforms of generating functions. These distributions are shown to tend, as time increases, to the known distributions for the steady state of the queue, but the rate of this convergence has not, apparently, been investigated.

More detailed results are available for the single-server queue with a random (Poisson) input process and negative exponential distribution of service times. In particular, Bailey [2][3] has obtained explicitly the queue length distribution as a function of time, and has obtained an asymptotic expression showing the rate at which the steady state is approached. References to earlier work are given in Bailey's paper.

This paper obtains, for the more general case of the bulk service queue introduced by Bailey [1], an asymptotic expression for the queue length distribution for large times. The results given are not restricted to negative exponential distributions of service times.

2. Bulk service queue

In Kendall's [11] queue model, customers arrive at random (Poisson process) at a single serving station, and are served in order of arrival, successive service intervals being independent with the same distribution. In

https://doi.org/10.1017/S1446788700039136 Published online by Cambridge University Press

the bulk service queue, conditions are the same, except that during each service interval, a batch of N customers, or the whole queue if less than N, is served. The queue lengths (excluding customers being served) at instants immediately before a service interval begins form a Markov chain, with matrix Q_{ij} of transition probabilities, where

$$\begin{aligned} Q_{ij} &= k_j && (i = 0, 1, \cdots, N) \\ Q_{ij} &= k_{j-i+N} && (i > N, j \geqq i-N) \\ Q_{ij} &= 0 && (i > N, j < i-N) \end{aligned}$$

and k_r denotes the probability of r arrivals in a service time. If p_r is the probability of a queue of r in the steady state, and

$$P(z) = \sum_{r=0}^{\infty} p_r z^r \tag{1}$$

then

$$P(z) = \frac{\sum_{j=0}^{N-1} p_j(z^N - z^j)}{[z^N/K(z)]-1} \tag{2}$$

in which

$$K(z) = \sum_{r=0}^{\infty} k_r z^r. \tag{3}$$

Since $P(z)$ is a probability generating function, it can have no singularity in $|z| < 1$. Bailey shows, that for a choice of $K(z)$ corresponding to a gamma distribution of service times, the N values $p_0, p_1, \cdots, p_{N-1}$ can be chosen so that the zeros in numerator and denominator of (2) cancel for $|z| < 1$. This leads to an explicit solution for $P(z)$.

For the transient case, let $p_r(n)$ denote the probability of a queue length of r immediately before service interval n has commenced. Define

$$P(z, n) = \sum_{r=0}^{\infty} p_r(n) \cdot z^n.$$

Then the analysis leading to (2) leads at once to

$$P(z, n+1) = z^{-N} K(z) \cdot \left\{ P(z, n) + \sum_{j=0}^{N-1} p_j(n) \cdot (z^N - z^j) \right\} \tag{4}$$

valid for $n = 0, 1, 2, \cdots$

For later analysis, it is convenient to express (4) in terms of a continuous variable t, replacing the integer variable n. Define then

$$\begin{aligned} F(z, t) &= P(z, n) && \text{for} \quad n \leqq t < n+1 \\ &= 0 && \text{for} \quad t < 0 \end{aligned} \tag{5}$$

$$f_j(t) = p_j(n) \qquad \text{for} \quad n \leqq t < n+1. \tag{6}$$

Then, for all $t \geqq 0$

$$F(z, t+1) = z^{-N}K(z)\left\{F(z, t) + \sum_{j=0}^{N-1} f_j(t) \cdot (z^N - z^j)\right\}. \tag{7}$$

If the initial queue length is a, then

$$F(z, t) = z^a \quad \text{for} \quad 0 \leqq t < 1.$$

From (7), writing $W(z) = z^N/K(z)$.

$$W(z)\int_0^\infty F(z, t+1)e^{-st}dt = \int_0^\infty F(z, t)e^{-st}dt + \int_0^\infty e^{-st}\left\{\sum_{j=0}^{N-1} f_j(t)(Z^N - z^j)\right\} dt.$$

But since $F(z, t) = 0$ for $t < 0$,

$$\int_{t=0-}^\infty e^{-st}dF(z, t) = s\int_0^\infty e^{-st}F(z, t)dt = P^*(z, s), \quad \text{say.} \tag{8}$$

A similar equation applies to each $f_j(t)$. Then

$$\begin{aligned}\int_0^\infty e^{-st}F(z, t+1)dt &= e^s\int_1^\infty e^{-st}F(z, t)dt \\ &= e^s s^{-1}P^*(z, s) - e^s s^{-1}z^a(1-e^{-s}).\end{aligned}$$

From these results,

$$P^*(z, s) = \{z^a W(z)(1-e^{-s}) + e^{-s}X(z)\}/\{W(z) - e^{-s}\} \tag{9}$$

in which

$$X(z) = \sum_{j=0}^{N-1}(z^N - z^j) \cdot p_j^*(s)$$

$$p_j^*(s) = \int_{0-}^\infty e^{-st}df_j(t).$$

If, following Bailey [1], $K(z)$ is taken as

$$K(z) = \{1 + m(1-z)/p\}^{-p} \tag{10}$$

which results from Poisson arrivals at mean rate λ and a service time distribution

$$dB(v) = [\alpha^p/\Gamma(p)]v^{p-1}e^{-\alpha v}dv \qquad (m = \lambda E(v) = \lambda p/\alpha) \tag{11}$$

then (9) gives

$$P^*(z, s) = z^a(1-e^{-s}) + [e^{-s}\{X(z) + z^a(1-e^{-s})\}]/[W(z) - e^{-s}]. \tag{12}$$

The denominator of (12) is a polynomial in z of degree $N+p$, and the numerator is of degree N if $a \leqq N$. This restriction on a will be assumed henceforth. Since $P(z, n)$ is a generating function, $F(z, t)$ and $P^*(z, t)$ can have no singularity in $|z| < 1$. The factors of the characteristic equation

$$W(z)-e^{-s}=0 \tag{13}$$

which vanish in $|z|<1$ must therefore cancel with corresponding factors of the numerator.

3. Roots of characteristic equation

Bailey [1] showed that $W(z)-1$ has exactly N distinct zeros in $|z|\leqq 1$, and p zeros in $|z|>1$. Consider

$$W(z)\equiv z^N\{1+m(1-z)/p\}^p=e^{-s}\equiv u. \tag{14}$$

For $m<N$, there are no multiple zeros in $|z|\leqq 1$, otherwise

$$z^N=u\{1+m(1-z)/p\}^{-p}$$

and

$$Nz^{N-1}=um\{1+m(1-z)/p\}^{-p-1}$$

which implies $m\geqq N$. Thus the solutions are distinct.

For $u=1$, Bailey [1] applied Rouché's theorem to show that (14) has exactly N roots in $|z|\leqq 1$. To extend this to the present case, let $z=\zeta(s)$ be any one solution of (14) satisfying $|\zeta(s)|\leqq 1$. Then $\zeta(s)$ is a regular function, except at the branch points given by $W'(z)=0$, i.e. at

$$\begin{aligned} z&=[N(p+m)]/[m(p+N)]\equiv g\\ u&=\left(\frac{N}{m}\right)^N\left(\frac{p+m}{p+N}\right)^{N+p}\equiv b\\ s&=-\beta+2\pi\nu i \qquad\qquad (\nu=0,\pm1,\pm2,\cdots)\\ \beta&=\log b, \end{aligned} \tag{15}$$

and

$$\begin{aligned} u&=0\\ s&=\infty. \end{aligned} \tag{16}$$

Therefore $\zeta(s)$ is continuous for finite s in $\mathrm{Re}(s)>0$. If any $\zeta(s)$ has $|\zeta(s)|=1$ for some $u=e^{-s}$ with $|u|<1$, then

$$1>|u|=|1+m(1-\zeta)/p|^p\geqq 1 \quad \text{for} \quad |\zeta|=1.$$

This contradiction, together with the continuity of $\zeta(s)$, shows that for Re $(s)>0$, no solution $\zeta(s)$ with $|\zeta(0)\leqq 1$ can have $|\zeta(s)|>1$ for any s in Re $(s)>0$, i.e. the N solutions with $|\zeta(0)|\leqq 1$ satisfy $|\zeta(s)|<1$ for all s in $\mathrm{Re}(s)>0$. A similar argument shows that the p solutions $\zeta(s)$ with $|\zeta(0)|>1$ satisfy $|\zeta(s)|>1$ for all s in $\mathrm{Re}(s)>0$.

Therefore cancelling the N factors common to numerator and denominator of (12)

https://doi.org/10.1017/S1446788700039136 Published online by Cambridge University Press

$$P^*(z, s) = z^a(1-e^{-s})+\phi(s)/\prod[z-\zeta(s)] \tag{17}$$

where the product is over the p solutions $\zeta(s)$ with $|\zeta(s)| > 1$, and $\phi(s)$ is of degree zero in z. The cancellation (17) assumes N linear conditions on the N undetermined $p_j^*(s)$. For $s = 0$, Bailey [1] showed that the determinant of the N equations is non-zero, given that the solutions $\zeta(0)$ are distinct, so that the equations are consistent. The determinant is a rational function of the $\zeta(0)$. Since each $\zeta(s)$ is an analytic function of s, the determinant for general values of s can only vanish for isolated values of s, so that the cancellation (17) is generally valid.

Since $P(z, n)$ is a generating function, $F(1-0, t) = 1$, whence $P^*(1-0, s) = 1$. This determines $\phi(s)$, so that

$$P^*(s) = z^a(1-e^{-s})+e^{-s}\prod_{j=1}^{p} [1-\zeta_j(s)]/[z-\zeta_j(s)] \tag{18}$$

where $\zeta_j(s)$ for $j = 1, 2, \cdots, p$ are the p zeros of $W(z)-e^{-s}$ for which $|\zeta_j(s)| > 1$.

4. Inversion procedure

Since $P^*(z, s)$ is a Laplace transform, a standard Tauberian theorem gives the stationary state distribution of queue length (provided $m < N$) as

$$P(z) = \lim_{t\to\infty} F(z, t) = \lim_{s\to\infty} P^*(z, s). \tag{19}$$

For $\mathrm{Re}(s) > 0$, the $\zeta_j(s)$ are regular, and no factor of the denominator of (18) vanishes. Hence $P^*(z, s)$ is a regular function of u for $|u| < 1$ and $|z| < 1$, and thus can be expanded in a power series

$$P^*(z, s) = \sum_{j=0}^{\infty} c_j(z) \cdot e^{-js}.$$

Applying the inversion formula

$$\begin{aligned} &\tfrac{1}{2}\{F(z, t+0)+F(z, t-0)\}-\tfrac{1}{2}\{F(z, 0+0)+F(z, 0-0)\} \\ &\qquad = \frac{1}{2\pi i}\lim_{c\to\infty}\int_{\gamma-ic}^{\gamma+ic} \frac{e^{ts}-1}{s}\cdot P^*(z, s)ds \end{aligned} \tag{20}$$

in which $\gamma > 0$ and $\frac{1}{2}\{F(z, 0+0)+F(z, 0-0)\} = \frac{1}{2}z^a$, an expansion of the form

$$F(z, t) = \sum_{j=0}^{\infty} c_j(z) \cdot U(t-j) \tag{21}$$

is obtained, where $U(t)$ is the Heaviside unit function. This series gives the transient distribution for small values of t, if the functions $\zeta_j(s)$ have been evaluated.

https://doi.org/10.1017/S1446788700039136 Published online by Cambridge University Press

For large values of t, the series (21) is not useful. What is required is an asymptotic inversion, valid for large t, of

$$T^*(z, s) = \prod_{j=1}^{p} [1-\zeta_j(s)]/[z-\zeta_j(s)]. \tag{22}$$

This expression differs from (18) by the omission of the multiplier e^{-s}, representing a unit time delay.

5. Asymptotic distribution

The following argument shows that no $\zeta_j(s)$ in (22) can assume a real value z in $0 < z < 1$. For $\mathrm{Re}(s) \geqq 0$, $|\zeta_j(s)| > 1$. If $\zeta_j(s) = z$, $0 < z < 1$, for $R(s) < 0$, then by (14) there is a real (negative) s with this property. Since $0 < z < 1$, $W'(z) > 0$, $d\zeta_j(s)/ds < 0$, so that $\zeta_j(0) < 1$. By (14), no $\zeta_j(0)$ has a real value less than -1, so that $|\zeta_j(0)| < 1$, contradicting $|\zeta_j(0)| > 1$.

The only singularities of the integrand of (20), when $0 < z < 1$, are therefore the branch points of the $\zeta_j(s)$. The integrand is therefore single-valued in the s-plane cut from each branch point $-\beta+2\pi\nu i$ to $-\infty+2\pi\nu i$. The integral (20) then equals the sum of integrals round each cut, plus a component which is zero if $T^*(z, s)$ is bounded for sufficiently large $|s|$. (This is the case, since from (14), $\zeta_j(s) = O\,(\exp[-s/(N+p)])$ as $|s| \to \infty$).

The integral (20) equals the stationary distribution (19), plus a function of t, obtained by inverting

$$\psi(z, s) = T^*(z, s) - T^*(z, 0).$$

The integral around the cut at $-\beta+2\pi\nu i$ contributes (for integer t)

$$\frac{1}{2\pi i} \oint_{-\infty}^{(0+)} \frac{e^{t(y-\beta)}}{y-\beta+2\pi\nu i} \cdot \psi(z, t-\beta+2\pi\nu i)dy. \tag{24}$$

Since

$$\lim_{r\to\infty} \sum_{\nu=-r}^{r} 2(y-\beta+2\pi\nu i)^{-1} = \frac{e^y+e^\beta}{e^y-e^\beta}$$

and $\psi(z, s)$ is a function of s only through e^{-s}, the integrals (24) sum to

$$\begin{aligned} &\frac{e^{-\beta t}}{4\pi i} \oint_{-\infty}^{(0+)} e^{ty} \cdot \frac{e^y+e^\beta}{e^y-e^\beta} \cdot \psi(z, t-\beta)dy \\ &\qquad = -\frac{e^{-\beta t}}{2\pi} \int_0^\infty e^{-tx} \cdot \frac{e^\beta+e^{-x}}{e^\beta-e^{-x}} \cdot G(x)dx \end{aligned} \tag{25}$$

where $x = -y$ and

https://doi.org/10.1017/S1446788700039136 Published online by Cambridge University Press

(26) $$2iG(x) = \psi(z, -x-\beta-0i)-\psi(z, -x-\beta+0i).$$

From (14), $u = W(z)$ may be expanded by Taylor's theorem about $z = g$, where g is defined in equation (15):

$$u = e^{\beta}+\tfrac{1}{2}(z-g)^2 W''(g)+\cdots$$

since $W'(g) = 0$. For small values of $v = s+\beta$, this series may be inverted to give an expansion of $z = \zeta(s)$ in half-integer powers of v, with leading terms

(27) $$\zeta(s) \sim g\pm[-2/W''(g)]^{\frac{1}{2}}[e^{\beta}-e^{-s}]^{\frac{1}{2}} \sim g\pm Kv^{\frac{1}{2}}$$

where

$$K = e^{\beta/2}[-2/W''(g)]^{\frac{1}{2}} = \left[\frac{2p}{N(N+p)}\right]^{\frac{1}{2}} g.$$

Only the positive sign in (27) corresponds to a $\zeta(s)$ with $|\zeta(0)| > 1$. Denote this function by $\zeta_1(s) = \eta(s)$.

Then $[1-\eta(s)]/[z-\eta(s)]-[1-\eta(0)]/[z-\eta(0)]$ has an expansion in half-integer powers of v, with leading term (apart from the constant) of

$$(1-z)(z-g)^{-2}Kv^{\frac{1}{2}}.$$

The other $\zeta_j(s)$ ($j = 2, 3, \cdots, p$) have no branch point at $s = -\beta$, so can be expanded in integer powers of v. Therefore $\psi(z, s)$ has an expansion in half-integer powers of v (with non-zero radius of convergence) in which the leading term (apart from the constant) is

(28) $$A(z)\cdot(1-z)(z-g)^{-2}Kv^{\frac{1}{2}}$$

where

$$A(z) = \prod_{j=2}^{p}[1-\zeta_j(0)]/[z-\zeta_j(0)].$$

Therefore $G(x)$ has a similar expansion, with leading term

(29) $$A(z)\cdot(z-1)(z-g)^{-2}\cdot Kx^{\frac{1}{2}}.$$

Since $\psi(z, s)$ is bounded, for large $|s|$, the conditions of Watson's lemma (e.g. Copson [4]) apply, so that a valid asymptotic expansion is obtained by integrating this expansion of the integrand term by term. The generating function for queue length distribution is therefore given, asymptotically for large n, by

(30) $$P(z, n) = \left\{\prod_{j=2}^{p}\frac{1-\zeta_j(0)}{z-\zeta_j(0)}\right\}\cdot\left\{\frac{1-\eta(0)}{z-\eta(0)}\right. \\ \left.-\frac{K(z-1)}{(z-g)^2}\cdot\frac{e^{\beta}+1}{e^{\beta}-1}\cdot\frac{e^{-\beta n}n^{-3/2}}{4\pi^{1/2}}\right\}.$$

https://doi.org/10.1017/S1446788700039136 Published online by Cambridge University Press

From (30), the asymptotic moments of the distribution are readily obtained. In particular, the mean queue length is

$$(31)\qquad \left[\frac{\partial P(z,n)}{\partial z}\right]_{z=1} \sim \left\{\sum_{j=1}^{p}[\zeta_j(0)-1]^{-1}\right\} - \frac{K(e^{\beta}+1)}{(e^{\beta}-1)(1-g)^2}\cdot\frac{e^{-\beta n}n^{-3/2}}{4\pi^{1/2}}.$$

6. Special cases

For the queue with Poisson arrivals and negative exponential service time distribution, and unit batch size ($p = N = 1$), the $\zeta(s)$ can be obtained explicitly. Assuming that the traffic intensity (m/N) is less than one, only one $\zeta(s)$ has $|\zeta(0)| > 1$, namely

$$(32)\qquad \eta(s) = \{(1+m)+[(1+m)^2-4me^{-s}]^{\frac{1}{2}}\}/(2m).$$

The stationary state distribution is then

$$\begin{aligned} P(z) &= [1-\eta(0)]/[z-\eta(0)] = (1-m)/(1-mz) \\ &= \sum_{j=0}^{\infty}(1-m)m^j z^j \end{aligned}$$

which is the known (Erlang) distribution. The mean queue length is obtained, asymptotically for large n, as

$$(33)\qquad \frac{m}{1-m} - \frac{2m(1+m)(1+6m+m^2)}{4\pi^{1/2}(1-m)^4}\cdot\left[\frac{(1+m)^2}{4m}\right]^{-n} n^{-3/2}.$$

For example, if the traffic intensity is 0.8, then mean queue length

$$\begin{aligned} &\sim 4.00-1634n^{-3/2}e^{-.0124n} \\ &= 3.5 \quad \text{at} \quad n = 100. \end{aligned}$$

This illustrates the slow approach of the mean queue length from an initial value of zero to the stationary value of 4.0.

Equation (33) may be compared with Bailey's result [3] for the simple queue ($N = p = 1$) in continuous time. Taking the mean service time as the time unit, the leading time-dependent term in Bailey's asymptotic expression is proportional to

$$c^{-2}e^{-ct}t^{-3/2} \quad \text{for} \quad c = [1-m^{\frac{1}{2}}]^2.$$

Both c and $\log[(1+m)^2/(4m)]$ from (33) are equal to

$$\tfrac{1}{4}(1-m)^2+0\{(1-m)^3\}$$

so that the two results agree, to this order, in their functional dependence on time.

As an example of a bulk service queue, the case of $N = 2$, $p = 2$ may be considered. For traffic intensity $m/N = 0.8$,

$$W(z) = 0.64z^4 - 2.88z^3 + 3.24z^2$$
$$g = 1.125$$
$$\beta = 0.0248.$$

The solutions $\zeta(s)$ having $|\zeta(0)| > 1$ are

$$\eta(s) = \zeta_1(s) \text{ for which } \eta(0) = 1.2498,\ \eta(-\beta) = 1.125$$

and $$\zeta_2(s) \text{ for which } \zeta_1(0) = 2.7111,\ \zeta_2(-\beta) = 2.7159.$$

The mean queue length is asymptotically

$$4.58 - 578n^{-3/2}e^{-.0248n}.$$

The queue length is therefore 4.11 at $n = 50$ and 4.53 at $n = 100$.

7. Comments

The steady state queue length distribution for the bulk service queue is determined by the p roots $\zeta_j(0)$. Equation (30) shows that the leading time-dependent term is the asymptotic transient distribution is completely determined by the $\zeta_j(0)$ and by quantities given explicitly by equations (15) and (30). It is of interest that, to obtain this result, only one of the roots $\zeta(s)$ of (14) had to be expanded as a function of s, and for this the first two terms of the series sufficed. The method should be applicable to other queue models, whose transient distribution has been obtained as a Laplace transform.

An asymptotic expansion such as (30) is particularly relevant when the traffic intensity in the queue is close to one. In this case, the stationary state may be approached very slowly, and, in a practical application, the stationary solution may be quite inapplicable, particularly if there is any change in the system parameters with time.

References

[1] Bailey, N. T. J., On queueing problems with bulk service, *J. Roy. Statist. Soc. B.*, **16** (1954), 80–86.
[2] Bailey, N. T. J., A continuous time treatment of a simple queue using generating functions, *J. Roy. Statist. Soc. B*, **16** (1954), 288–291.
[3] Bailey, N. T. J., Some further results in the non-equilibrium theory of a simple queue, *J. Roy. Statist., Soc. B*, **19** (1957) 326–333.
[4] Copson, E. T., *Theory of functions of a complex variable*, O.U.P. (1935).
[5] Finch, P. D., On the busy period in the queueing system *GI/G/1*, This *Journal* **2** (1961–2) 217–228.

[6] Finch, P. D., The single-server queueing problem with non-recurrent input process and Erlang service time, This *Journal* (to appear).
[7] Gani, J., Problems in the probability theory of storage systems, *J. Roy. Statist. Soc. B*, **19** (1957), 181–206.
[8] Jaiswal, N. K., Time-dependent solution of the bulk service queueing problem, *Opns. Res.*, **8** (1960), 773–781.
[9] Keilson, J. and Kooharian, A., On time-dependent queueing processes, *Ann. Math. Statist.*, **31** (1960), 104–112.
[10] Keilson, J. and Kooharian, A., On the general time-dependent queue with a single server, *Ann. Math. Statist.*, **33** (1962), 767–791.
[11] Kendall, D. G., Some problems in the theory of queues, *J. Roy. Statist. Soc. B*, **13** (1951), 151–175.
[12] Lindley, D. V., The theory of queues with a single server, *Proc. Camb. Phil. Soc.*, **48** (1952), 277–289.
[13] Takacs, L., Transient behaviour of single-server queueing processes with recurrent input and exponentially distributed service times, *Opns. Res.*, **8** (1960), 231–245.
[14] Takacs, L., A single-server queue with recurrent input and exponentially distributed service times, *Opns. Res.*, **10** (1962), 395–399.

Department of Mathematics,
University of Melbourne

https://doi.org/10.1017/S1446788700039136 Published online by Cambridge University Press