ORC 66-19
AUGUST 1966

SOME INEQUALITIES FOR SINGLE SERVER QUEUES

by
Kneale Thomas Marshall

OPERATIONS RESEARCH CENTER

COLLEGE OF ENGINEERING

UNIVERSITY OF CALIFORNIA - BERKELEY

SOME INEQUALITIES FOR SINGLE SERVER QUEUES

by

Kneale Thomas Marshall
Operations Research Center
University of California, Berkeley

AUGUST 1966 ORC 66-19

This research was supported by the Office of Naval Research under Contract Nonr-222(83). Reproduction in whole or in part is permitted for any purpose of the United States Government.

ABSTRACT

The expected wait in the GI/G/1 queue is related to the mean and variance of the idle time. For arrival distributions which are IFR or have mean residual life bounded by $\frac{1}{\lambda}$, simple bounds are obtained which give, for example, the expected number in queue to within at most one customer.

By equating input with output, relations between random variables are used to obtain expressions foı the moments of the waiting time in terms of moments of the inter-arrival, service, and idle time distributions. By bounding the idle time moments, bounds are obtained on the mean and variance of the waiting time, the mean length of a busy period, and the probability an arrival finds the system empty. Bounds on the mean wait lead to bounds on the expected virtual wait.

Similar results are obtained for some generalizations of the GI/G/1 queue, including batch arrivals, batch service and priority queues. Queues where the first customer in each busy period has some added delay are also considered.

Some preliminary results for tandem queues are given.

ACKNOWLEDGEMENTS

The author would like to express his appreciation for the many helpful comments and suggestions of Professor Ronald W. Wolff whose interest stimulated much of this paper. Acknowledgements are also made to Professors William S. Jewell and Gordon F. Newell for their continued interest and support.

CONTENTS

Chapter 1

SOME RESULTS AND BOUNDS FOR ALL GI/G/1 QUEUES

1. Introduction

Little work has been done on approximations in queueing. Emphasis has been on complex analytic results. Notable exceptions are papers by Kingman (1962 (a) and (b)) and recently by Newell (1965). The work of Newell is applied primarily to traffic light problems, whereas Kingman's is more closely related to this thesis.

In this chapter some new results are found for various indicators of performance in the GI/G/1 queue. Bounds which are easily calculable are found for such items as the expected wait in queue, expected length of an idle period and the variance of inter-output times.

We find a relationship between the idle time between busy periods and the waiting time of a customer in queue. The expected wait in queue is found in terms of the first two moments of the inter-arrival, service, and idle times. For Poisson arrivals the idle time distribution is exponential, and the expected wait is calculated easily. In general the moments of the idle distribution are difficult to calculate. However, an upper bound for all GI/G/1 queues is easily found in terms of the mean and variance of the arrival and service streams only (see also Kingman, 1962 (a) or (b)). A lower bound is found which requires knowledge of the arrival and service distributions, and not just the first two moments.

Only stationary queues are considered. No transient results are given.

2. Notation

The following notation is used throughout the paper. The sign $\sim$ is used to signify "with distribution function".

We shall deal exclusively with stationary queues in this paper , by which we shall mean that the queueing process either started at time zero with stationary conditions or that it started with some initial condition (such as the wait in queue of the first customer is zero) but that time was at $-\infty$.

By the subscript n (e.g., W_n) we shall be referring to the n-th customers in a stationary stream. When it is not required to note the order of the customers the subscript will be dropped.

T_n = time between n-th and (n+1)-th arrival, $T_n \sim A(t)$, $E[T_n] = 1/\lambda$

S_n = service time of n-th customer , $S_n \sim G(t)$, $E[S_n] = 1/\mu$

$U_n = S_n - T_n$, $U_n \sim K(t)$

τ_n = time between n-th and (n+1)-th departure

$\rho = \lambda/\mu$

W_n = wait in queue of n-th customer, $W_n \sim W(t)$

V = virtual wait in a stationary queue (see section 4)

I = length of idle period between busy periods, $I \sim H(t)$

B = length of busy period , $B \sim B(t)$

It is possible in some queueing situations that an arrival and service can take place together, leading to problems in defining what is an idle period for the queue. We shall define $P[I=0] = 0$, and thus if an arrival occurs at the instant the last customer present departs, the busy period continues, and ends only when the facility is empty for a positive length of time.

N_b = number served in a busy period

D = total delay in system = $W + S$

N_q = number in the queue at a random point in time

$\nu_f^{(n)}$ = n-th moment about origin of random variable with distribution F.

The superscript is dropped for n=1, e.g., $\nu_a = 1/\lambda$, $\nu_g = 1/\mu$, $\nu_h = E[I]$

σ_f^2 = variance of a random variable with distribution F

$c_f^2 = \sigma_f^2/(\nu_f)^2$, where c_f is the coefficient of variation

a_o = P[Arrival finds the system empty]

$F^c(t) = 1 - F(t)$ for any distribution F

3. The Wait in Queue and the Idle Period

By equating input times with output times relationships between the moments of the arrival, service, idle, and waiting time distributions are now found.

The following result is given in Riordan (1962), but because of its usefulness it is proved here.

Theorem 1.1: For all GI/G/1 queues with $\rho < 1$,

$$a_o \nu_h = \frac{1}{\lambda} - \frac{1}{\mu}. \tag{1}$$

Proof: Consider the time between the n-th and (n+1)-th departures

$$\tau_n = T_n - D_n + D_{n+1}$$

$$= S_{n+1} + X_n,$$

where $$X_n = 0 \text{ if } W_{n+1} > 0$$

$$= -(D_n - T_n) \text{ if } W_{n+1} = 0.$$

If $U_n = S_n - T_n$ we have the fundamental relationships

$$W_{n+1} = W_n + X_n + U_n \qquad (2)$$

$$= \text{Max}\,[0, W_n + U_n]. \qquad (3)$$

When $X_n > 0$, $X_n = I$. Hence, taking expectations in (2) and assuming stationarity

$$E[X_n] = a_o E[I] = -E[U_n] \text{ which proves (1).}$$

As examples, for Poisson arrivals $a_o = (1-\rho)$ and the idle distribution is exponential with mean $\frac{1}{\lambda}$. For the constant arrival, constant service case (D/D/1 queue) $a_o = 1$ and $I = \frac{1}{\lambda} - \frac{1}{\mu}$.

The following expression is now derived for the expected wait in queue.

Theorem 1.2: For all GI/G/1 queues with $\rho < 1$,

$$E[W] = \frac{E[U^2]}{-2E[U]} - \frac{E[I^2]}{2E[I]}$$

$$= \frac{\lambda^2(\sigma_a^2 + \sigma_q^2) + (1-\rho)^2}{2\lambda(1-\rho)} - \frac{\nu_h^{(2)}}{2\nu_h}. \qquad (4)$$

Proof: Write (2) as $W_{n+1} - X_n = W_n + U_n$. Square both sides and note that $W_{n+1}X_n = 0$, giving

$$W_{n+1}^2 + X_n^2 = W_n^2 + 2W_nU_n + U_n^2.$$

Taking expectations, since W_n and U_n are independent, and $E[X_n^2] = a_oE[I^2]$ we have

$$a_oE[I^2] = 2E[U_n]E[W_n] + E[U_n^2].$$

Using theorem 1.1 the result follows.

It is important to note the special way in which the moments of the idle distribution occur. $\frac{E[I^2]}{2E[I]}$ is the mean of an equilibrium excess idle distribution; that is, it is the mean of a random variable with distribution function $\int_0^t \frac{H^c(u)}{\nu_h} du$. This is a well known result in renewal theory.

Consider again our two examples. For Poisson arrivals $\frac{\nu_h^{(2)}}{2\nu_h} = \frac{1}{\lambda}$. In this case (4) reduces to $E[W] = \frac{\rho(1+c_a^2)}{2\mu(1-\rho)}$, a well known result. For the D/D/1 queue $\sigma_a^2 = \sigma_g^2 = 0$ and $\nu_n^{(2)} = (\nu_h)^2 = (\frac{1}{\lambda} - \frac{1}{\mu})^2$, in which case (4) reduces to $E[W] = 0$.

An expression for the variance of the wait is now found in a similar manner and is given by

Theorem 1.3: For all GI/G/1 queues with $\rho < 1$,

$$\sigma_w^2 = \frac{E[U^3]}{-3E[U]} + \left[\frac{E[U^2]}{-2E[U]}\right]^2 + \left[\frac{E[I^3]}{3E[I]} - \left[\frac{E[I^2]}{2E[I]}\right]^2\right],$$

or

$$\sigma_w^2 = \frac{\lambda(\nu_g^{(3)} - \nu_a^{(3)}) + 3(\rho\nu_a^{(2)} - \nu_g^{(2)})}{3(1-\rho)} + \left[\frac{\lambda^2(\sigma_a^2 + \sigma_g^2) + (1-\rho)^2}{2\lambda(1-\rho)}\right]^2 + \sigma_{e,h}^2 \qquad (5)$$

where $$\sigma_{e,h}^2 = \nu_{e,h}^{(2)} - \nu_{e,h}^2 = \frac{\nu_h^{(3)}}{3\nu_h} - \left[\frac{\nu_h^{(2)}}{2\nu_h}\right]^2.$$

<u>Proof</u>: Write (2) as $W_{n+1} - X_n = W_n + U_n$ and cube both sides. Note that $X_n^2 W_{n+1} = W_{n+1}^2 X_n = 0$. Using (1) and (4) after taking expectations the result follows.

The expression for the expected wait is of particular interest in queueing and it is seen to depend only on the first two moments of the interarrival, service, and idle distributions. In general these idle period moments are difficult to calculate but bounds will be obtained for them and this is the subject of section (7) of this chapter and sections (1-4) of Chapter 2.

4. The Virtual Wait

The virtual wait is defined here to be the time an arrival would have to wait in queue if he arrived at some random point in time into a stationary queue. The expected value of V is found easily in terms of the expected value of W, the actual wait in queue.

First we show that at a random point in time in a stationary queue, the probability the system is empty is $(1-\rho)$. The times between the starts of busy periods are independent and identically distributed, say with distribution $B^*(t)$, the distribution of $(B+I)$. Hence, the queue has an imbedded renewal process, and note that following the end of an idle period, the distribution of the next busy period is independent, and distributed as B(t). Let $M^*(t)$ be the expected number of busy cycles (i.e., from the start of one busy period to the start of the next) in (0,t] starting at t=0 with the start of a busy period.

$$P[\text{Busy at } T] = \int_0^T B^c(T-u)\,dM^*(u)$$

and applying the Key Renewal theorem, if $B^*(t)$ is non-lattice,

$$\lim_{T\to\infty} P[\text{Busy at } T] = \frac{E[B]}{E[B]+E[I]} = \rho .$$

In the lattice case, the Cesàro limit may be found by Tauberian arguments.

Now let V = X+Y, where X is the excess service time of the customer in service and Y is the sum of the service times of all those in queue when a random arrival occurs. Then

$$E[Y] = \frac{E[N_q]}{\mu} = \rho E[W].$$

Conditioning only on whether or not a random arrival finds the system busy, $E[X] = \frac{\nu_g^{(2)}}{2\nu_g}$.

Hence,
$$E[V] = \frac{\rho(c_g^2+1)}{2\mu} + \rho E[W]. \qquad (6)$$

It is interesting to note from (6) that E[V] = E[W] if and only if $E[W] = \frac{\rho(c_g^2+1)}{2\mu(1-\rho)}$ which is the case for Poisson arrivals. Using equation

(4) in (6) gives

$$E[V] = \frac{\rho(C_g^2 + C_a^2)}{2\mu(1-\rho)} + \frac{1+C_a^2}{2\mu} - \frac{\rho\nu_h^{(2)}}{2\nu_h}.$$

Using either (6) or (7) in the D/D/1 queue gives the Cesàro mean $E[V] = \frac{\rho}{2\mu}$.

5. The Variance of the Output

It is obvious that $E[\tau_n] = E[T_n] = \frac{1}{\lambda}$. The variance is found as follows. From the relationships in section 3,

$$\text{Var }[\tau_n] = \text{Var }[S_{n+1}] + \text{Var }[X_n] \quad (8)$$

and $$\text{Var }[W_{n+1}-X_n] = \text{Var }[D_n-T_n] = \sigma_a^2 + \sigma_g^2 + \sigma_w^2. \quad (8a)$$

But $$\text{Var }[W_{n+1}-X_n] = \sigma_w^2 + \text{Var }[X] - 2\text{Cov }(W_{n+1}X_n). \quad (8b)$$

Now $W_{n+1}X_n = 0$ and hence,

$$\text{Cov }(W_{n+1}X_n) = -E[W](\frac{1}{\lambda} - \frac{1}{\mu}).$$

Using this with (8), (8a), and (8b) gives

$$\text{Var }[\tau_n] = \sigma_a^2 + 2\sigma_g^2 - \frac{2}{\lambda}(1-\rho)E[W]. \quad (9)$$

Using equation (4) for E[W] we have finally,

$$\text{Var }(\tau_n) = \sigma_g^2 - \frac{(1-\rho)^2}{\lambda^2} + \frac{(1-\rho)}{\lambda}\frac{\nu_h^{(2)}}{\nu_h}.$$

For the M/G/1 queue this gives Var $(\tau_n) = \sigma_g^2 + \frac{1-\rho^2}{\lambda^2}$, so that the variance of the output of the M/G/1 queue is known exactly when the mean and variance of the service distribution are given.

If G(t) is exponential, $\sigma_g^2 = \frac{1}{\mu^2}$ and Var $(\tau_n) = \frac{1}{\lambda^2}$. In the case of constant arrivals, constant service, $\sigma_g^2 = 0$ and Var $(\tau_n) = 0$.

6. The Covariance of Consecutive Outputs

We now derive an expression for the covariance between two consecutive output times. This expression depends on the covariance between a service time and the wait of the next customer. However bounds can be obtained to show that it is bounded close to zero. This is done in section 7.

Theorem 1.4: For all GI/G/1 queues with $\rho < 1$,

$$\text{Cov } (\tau_n,\tau_{n+1}) = \text{Cov } (S_{n+1},W_{n+2}) + \frac{\alpha}{\lambda}(1-\rho) - \sigma_g^2, \tag{10}$$

where $\alpha = E\left[\text{Max } [0,U_n]\right] = \int_0^\infty x \, dK(x).$

Proof:

$$\tau_n = T_n - D_n + D_{n+1}$$

$$(\tau_n\tau_{n+1}) = (T_n - D_n + D_{n+1})(T_{n+1} - D_{n+1} + D_{n+2}),$$

and in a straight-forward but tedious manner we find that

$$\text{Cov}(\tau_n,\tau_{n+1}) = E[(D_n - T_n)D_{n+1}] - E[(D_n - T_n)D_{n+2}] - E[D_{n+1}^2] + E[D_{n+1}D_{n+2}]. \quad \text{(a)}$$

Since $D_{n+1} = W_{n+1} + S_{n+1}$ and $W_{n+1} = \text{Max}\ [0, D_n - T_n]$,

$$E[(D_n - T_n) D_{n+1}] = E[W_{n+1}^2] + (\tfrac{1}{\mu})(E[W] + \tfrac{1}{\mu} - \tfrac{1}{\lambda}) \qquad \text{(b)}$$

$$E[D_{n+1}^2] = E[W_{n+1}^2] + 2E[W_{n+1} S_{n+1}] + (\tfrac{1}{\mu})^2. \qquad \text{(c)}$$

Using (b) and (c) in (a), after some cancellation,

$$\text{Cov}(\tau_n, \tau_{n+1}) = E[(W_{n+1} + S_{n+1} - (D_n - T_n)) W_{n+2}] - \frac{E[W]}{\mu} - \sigma_g^2.$$

$$= -E[(\text{Min}(0, D_n - T_n)) W_{n+2}] + \text{Cov}(S_{n+1}, W_{n+2}) - \sigma_g^2. \qquad \text{(d)}$$

Now

$$W_{n+2}\ \text{Min}\ (0, D_n - T_n) = \text{Max}\ (0, W_{n+1} + S_{n+1} - T_{n+1})\ \text{Min}\ (0, D_n - T_n).$$

When $D_n - T_n < 0$, then $W_{n+1} = 0$,

$$\therefore W_{n+2}\ \text{Min}\ (0, D_n - T_n) = \text{Max}\ [0, S_{n+1} - T_{n+1}]\ \text{Min}\ (0, D_n - T_n)$$

which by independence gives

$$E[W_{n+2}\ \text{Min}\ (0, D_n - T_n)] = -\alpha E[\text{Max}\ (0, D_n - T_n) - (D_n - T_n)]$$

$$= -\alpha(\tfrac{1}{\lambda} - \tfrac{1}{\mu}).$$

Putting this in (d) gives

$$\text{Cov}(\tau_n, \tau_{n+1}) = \text{Cov}(S_{n+1}, W_{n+2}) + \frac{\alpha}{\lambda}(1-\rho) - \sigma_g^2,$$

which establishes (10).

7. Some Bounds for All GI/G/1 Queues

Using the results of the previous sections some simple bounds can be found for various factors in the GI/G/1 queue, such as the mean length of an idle period and the mean wait in queue.

a) The mean idle time. Since $a_o \leqq 1$, theorem 1.1 gives a lower bound on the length of an idle period,

$$E[I] \geqq \frac{1}{\lambda} - \frac{1}{\mu}. \tag{11}$$

The bound is tight for the D/D/1 queue.

b) The wait in queue. Recall that from theorem 1.2

$$E[W] = \frac{\lambda(\sigma_a^2 + \sigma_g^2) + (1-\rho)^2}{2\lambda(1-\rho)} - \frac{\nu_h^{(2)}}{2\nu_h}. \tag{12}$$

From (11) and Var $[I] \geqq 0$, it follows that

$$E[W] \leqq \frac{\lambda(\sigma_a^2 + \sigma_g^2)}{2(1-\rho)}.$$

This upper bound for all GI/G/1 queues is also found by Kingman (1962 (a) and (b)). Equality holds for the D/D/1 queue.

The importance of these bounds is that they involve at most only the first two moments of the arrival and service distributions and further knowledge of the distributions is not required. However, if K(t) is known (or alternatively if A(t) and G(t) are known) a lower bound on the wait in queue can be found as follows.

Theorem 1.5: Let ℓ be a solution of

$$x = \int_{-x}^{\infty} K^c(u)\,du, \; x \geqq 0, \text{ where } (S_n - T_n) \sim K(t)$$

which exists and is unique if and only if $\rho < 1$. Then for all GI/G/1 queues, $E[W] \geqq \ell$.

Proof: Recall the fundamental equation (3)

$$W_{n+1} = \text{Max } [0, W_n + U_n].$$

Then $[W_{n+1} | W_n = x] = [\text{Max } 0, x + U_n]$,

and $E[W_{n+1} | W_n = x] = \int_{-x}^{\infty} K^c(u)\,du$ all $x \geqq 0$. (13)

Now let $\int_{-x}^{\infty} K^c(u)\,du = g(x)$ which is a continuous convex function for $x \geqq 0$, with $g'(x) = K^c(-x)$, so $K^c(0^+) = g'(0^-) = P[U_n > 0]$ and $g'(x) \to 1$ as $x \to \infty$.

Let $-\beta = E[\text{Min } (0, U_n)] = \int_{-\infty}^{0} K(u)\,du$ and

$$\alpha = E[\text{Max } (0, U_n)] = \int_{0}^{\infty} K^c(u)\,du.$$

Then $\alpha - \beta = \frac{1}{\mu} - \frac{1}{\lambda}$.

From (13)

$$E[W_{n+1}] = \int_{0^-}^{\infty} g(x)\,d\,W_n(x)$$

or $$E[W_{n+1}] = E[g(W_n)].$$

Using Jensen's inequality for the expected value of a convex function of a non-negative random variable,

$$E[W_{n+1}] \geqq g(E[W_n]), \quad \text{so that}$$

$$E[W] \geqq \int_{-E[W]}^{\infty} K^c(u)du.$$

Consider the equation

$$x = \int_{-x}^{\infty} K^c(u)du, \qquad x \geqq 0 \tag{14}$$

This can be written

$$x = \alpha + \int_{-x}^{0} K^c(u)du \qquad x \geqq 0.$$

The situation is drawn in Figure 1. The equation has a solution if and only if the two curves cross. If $\alpha = 0$, x=0 is a solution; if $\alpha > 0$ the curves cross if and only if for x sufficiently large,

$x > \alpha + \int_{-x}^{0} K^c(u)du \Leftrightarrow \int_{-x}^{0} K(u)du > \alpha$ or if and only if $\beta > \alpha$.

But $\beta > \alpha$ if and only if $\frac{1}{\lambda} > \frac{1}{\mu}$. Uniqueness comes from convexity arguments. Uniqueness fails only when the two curves coincide over some range, $[a,b]$ say. This implies $g'(x) = K^c(-x) = 1$ on $[a,b] \Rightarrow g'(x) = 1$ on $[a,\infty) \Rightarrow$ curves don't cross. In the case $\rho \geqq 1$, either no solution exists or, for example in the case of the

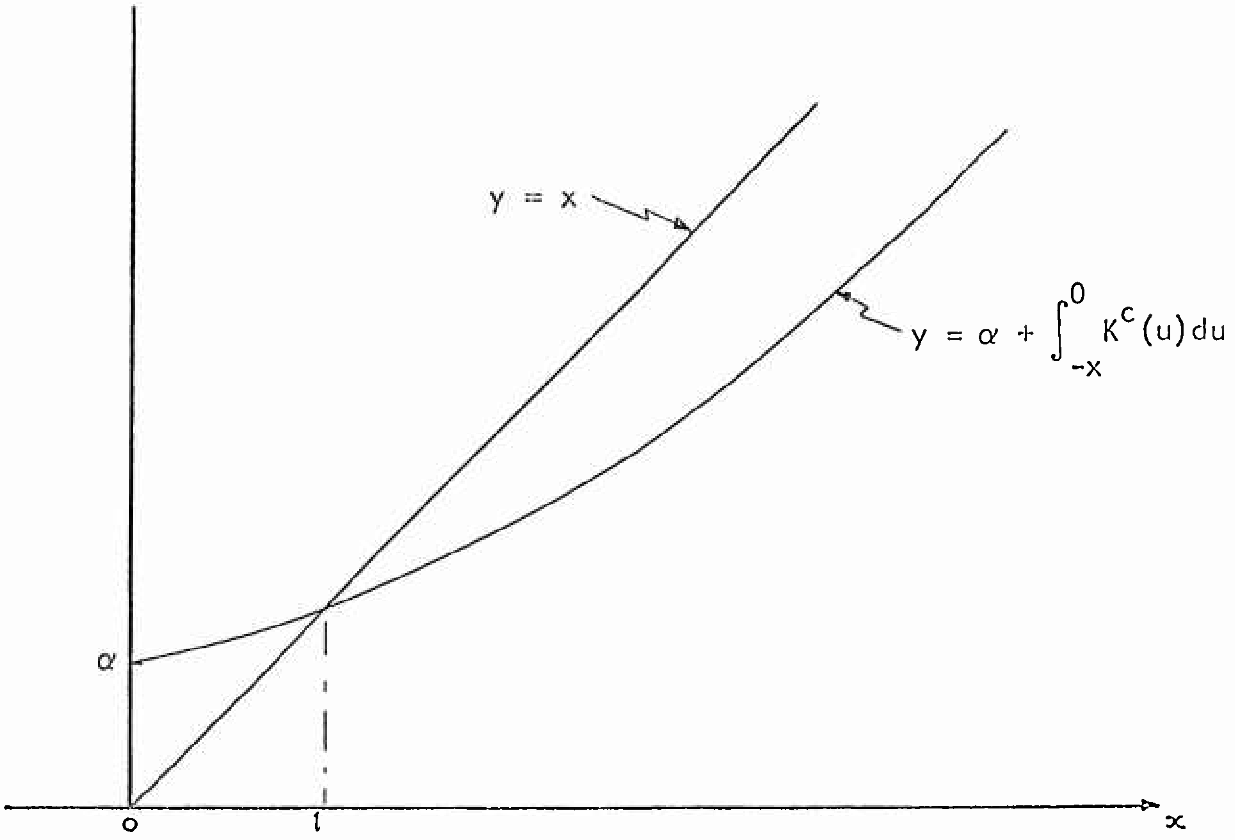

Figure 1: Determination of Lower Bound on the Wait in Queue, t.

D/D/1 queue, an infinite number of solutions exist with $\rho=1$.

So for $\rho < 1$, let t be the unique solution of (14). It is now shown that $t \leqq E[W]$. This is obvious from Figure 1 and equations (13) and (14). If $t=0$ the inequality is trivial. If $t > 0$ then $\alpha > 0$ and for all $0 \leqq x < t$, $x < \alpha + \int_{-x}^{0} K^c(u)du$ from the uniqueness property of t. Hence, if $E[W] < t$, then $E[W] < \int_{-E[W]}^{\infty} K^c(u)du$ which contradicts (14) and the theorem is proved.

Summarizing, we have shown that for all GI/G/1 queues with $\rho < 1$

$$t \leqq E[W] \leqq \frac{\lambda(\sigma_a^2 + \sigma_g^2)}{2(1-\rho)}. \qquad (15)$$

where t is the unique solution of (14). For $\sigma_a^2 + \sigma_g^2 > 0$ (i.e., all except the D/D/1 queue), both bounds tend to infinity as $\frac{1}{\lambda} \rightarrow \frac{1}{\mu} > 0$. However, their ratio may diverge in a particular case as is shown below for the case of M/M/1 queue.

For the Poisson arrival, Exponential service queue it is found that

$$k(t) = \frac{\lambda\mu}{\mu+\lambda}\, e^{-\mu t} \qquad t \geqq 0$$

$$= \frac{\lambda\mu}{\mu+\lambda}\, e^{\lambda t} \qquad t \leqq 0$$

which gives

$$K^c(t) = \frac{\rho}{1+\rho}\, e^{-\mu t} \qquad t \geqq 0$$

$$= 1 - \frac{1}{1+\rho}\, e^{\lambda t} \qquad t \leqq 0.$$

Using this in (14) it is found that the lower bound for this case is given by:

$$t = -(\tfrac{1}{\lambda}) \log_e (1-\rho^2) \quad \text{which} \rightarrow \infty \quad \text{as} \quad \rho \rightarrow 1^-.$$

However, it is easy to show that $\lim\limits_{\rho \rightarrow 1^-} (1-\rho) \log_e \frac{1}{1-\rho^2} = 0$ and hence, the bounds diverge. The upper and lower bounds and true value of E[W] are shown in Figure 2 for fixed $\lambda=1$ and varying μ.

c) <u>The variance of the output</u>.

The variance of the output distribution is given in equation (9). Using arguments similar to those in b) the following upper and lower bounds are found for all general arrival, general service single channel queues,

$$\sigma_g^2 \leq \text{Var}\,[\tau_n] \leq \sigma_a^2 + 2\sigma_g^2 - 2t\,(\frac{1}{\lambda} - \frac{1}{\mu}), \qquad (16)$$

where t is the solution of equation (14).

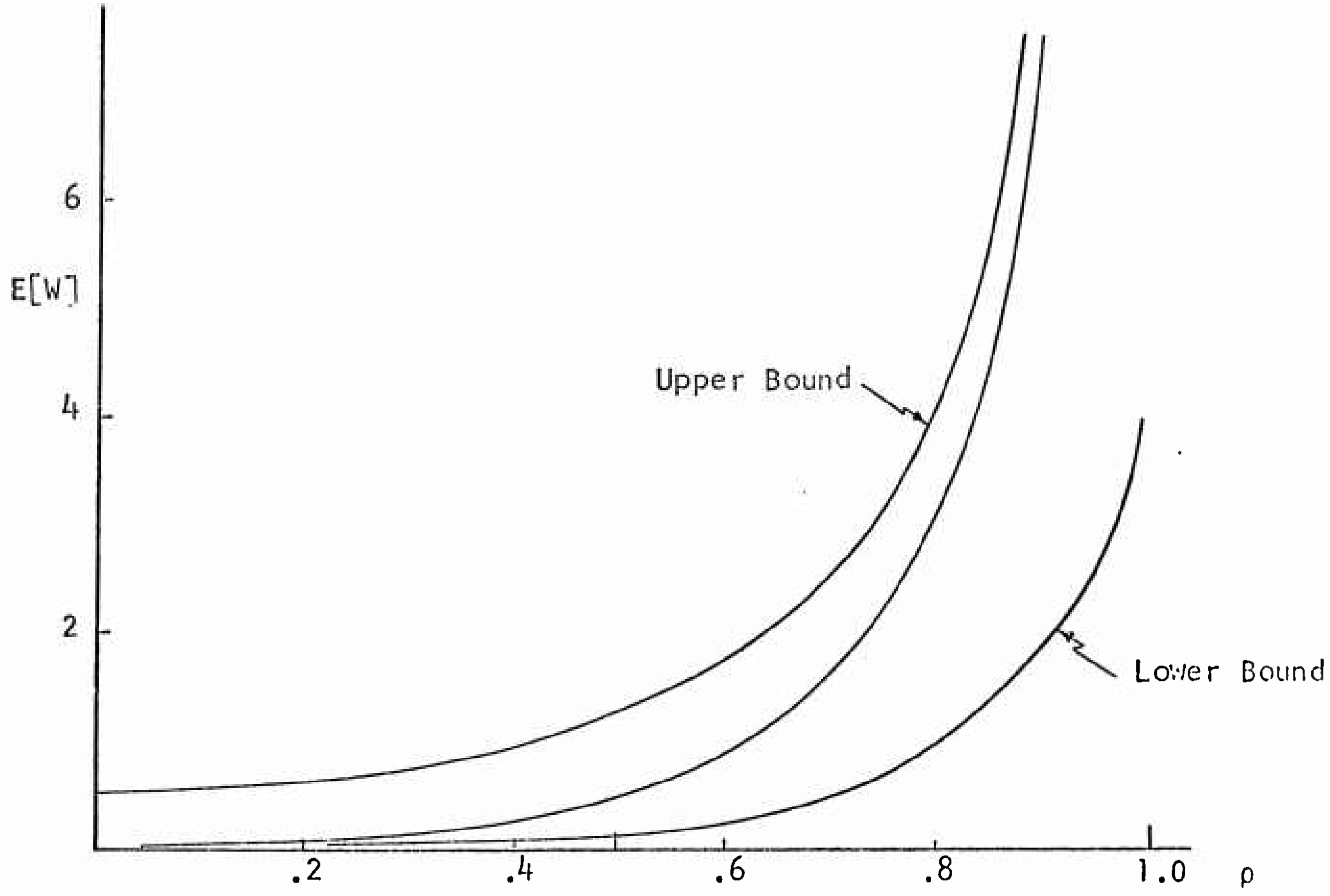

Figure 2: Bounds on the Expected Wait in the M/M/1 Queue.

d) The covariance of adjacent inter-output times.

Equation (10) gives the covariance of τ_n and τ_{n+1} in terms of the covariance of S_{n+1} and W_{n+2}.

Now $$S_{n+1}W_{n+2} = S_{n+1} \text{ Max } [0, W_{n+1}+S_{n+1}-T_{n+1}]$$

$$\geq S_{n+1}(W_{n+1}+S_{n+1}-T_{n+1}).$$

Hence, $$\text{Cov } (S_{n+1}W_{n+2}) \geqq E[S^2_{n+1}] - \frac{1}{\lambda\mu}$$

and using this in (10) gives

$$\text{Cov } (\tau_n,\tau_{n+1}) \geqq \frac{1}{\lambda}\,(1-\rho)\,(\alpha - \frac{1}{\mu}).$$

It is easy to show that $\alpha < \frac{1}{\mu}$ and thus that this expression is negative. An upper bound is found as follows:

$$\text{Max } [0,W_{n+1}+S_{n+1}-T_{n+1}] \leqq W_{n+1} + \text{Max } [0,S_{n+1}-T_{n+1}]$$

and using this gives

$$\text{Cov } (S_{n+1},W_{n+2}) \leqq E[S_{n+1} \text{ Max } [0,S_{n+1}-T_{n+1}]] \leqq E[S^2_{n+1}],$$

which leads to:

Theorem 1.6: For all GI/G/1 queues with $\rho < 1$

$$\alpha(\frac{1}{\lambda} - \frac{1}{\mu}) + \frac{1}{\mu^2} - \frac{1}{\lambda\mu} \leqq \text{Cov } (\tau_n,\tau_{n+1}) \leqq \alpha(\frac{1}{\lambda} - \frac{1}{\mu}) + \frac{1}{\mu^2}. \qquad (17)$$

The lower bound is negative and upper bound positive so in general no conclusion can be drawn as to the sign of the covariance. However, it has been bounded to within $\frac{1}{\lambda\mu}$.

e) The Virtual Wait.

The bounds obtained in part (b) and given in equation (15) can be used in equation (6) to show that, for all GI/G/1 queues with $\rho < 1$,

$$\rho\left[\frac{(1+c_g^2)}{2\mu} + t\right] \leqq E[V] \leqq \frac{\rho\,(c_g^2 + c_a^2)}{2\mu\,(1-\rho)} + \frac{\rho+c_a^2}{2\mu}. \tag{18}$$

Chapter 2

QUEUES WITH ARRIVAL DISTRIBUTIONS WHICH HAVE MONOTONE FAILURE RATES

1. Introduction

In the previous chapter it was seen that the moments of the idle time distribution occurred in many of the expressions. The idle time distribution is some complicated tail distribution of an inter-arrival time and it might be conjectured that by placing some restriction on the inter-arrival time distribution one might obtain some desirable properties of the oments of the idle period regardless of the service distribution. This indeed turns out to be true.

The first restriction to be placed on the arrival distribution is to restrict it to the class with decreasing mean residual life (DMRL). In this chapter the words decreasing and increasing are used in the weak sense and always should be read to mean non-increasing and non-decreasing respectively. The symbols ↓ and ↑ will be used respectively for decreasing and increasing in this weak sense.

<u>Definition 1</u>. A non-discrete distribution F has DMRL (IMRL) if and only if

$$\int_t^\infty \frac{F^c(x)dx}{F^c(t)} \quad \begin{matrix}\downarrow \\ (\uparrow)\end{matrix} \qquad \text{for all } t \geqq 0 \text{ when finite.}$$

The expressions and symbols in parenthesis should be read together.

A slightly stronger assumption on the arrival distribution will also be used which implies the above assumption.

<u>Definition 2</u>. A non-discrete distribution F has increasing failure rate (is IFR) if and only if

$$\frac{F(t+\Delta) - F(t)}{F^c(t)} \uparrow \qquad \text{for } \Delta > 0 \text{ and all } t \geqq 0 \text{ where finite.}$$

If F is discrete, then it is IFR if and only if

$$\frac{p_k}{\sum_{n=k}^{\infty} p_n} \uparrow \qquad \text{all } k \geqq 0 \text{ where finite.}$$

It is said to be DFR (have decreasing failure rate) when ↑ is replaced by ↓ in the above expressions.

These concepts are widely used in reliability theory where strong physical justifications can be given for their use in particular problems. In queueing an IFR arrival distribution would have the following physical interpretation. Given it has been a time t since the last customer arrived, the probability that a customer arrives in the next small interval Δ is increasing in t. Besides any physical justification many parametric families have this property; for example the gamma and Weibul distributions in certain parameter ranges, and the truncated normal and modified extreme value distributions. The degenerate distribution of the constant arrival queue also has the IFR property. For a fuller discussion on these properties the reader should consult Chapter 2 of R. E. Barlow and F. Proschan (1965).

For IFR/G/1 queues, (that is, the class of GI/G/1 queues whose arrival distributions have the IFR property) it is shown that simple expressions can be obtained to bound, for example, the expected number in the queue to within <u>at most</u> one customer. These bounds involve only the mean and variance of the arrival and service streams. For the special class of D/G/1, (constant arrival, general service), the expected number in the queue is bound to within at most one half.

2. Some Properties of the Idle Distribution

In this section two theorems are proved which give some of the useful properties of the idle distribution. These are then used in section 3 to obtain bounds for certain factors in the given class of queues. In what follows, symbols and expressions in parenthesis should be read together.

Theorem 2.1: For all GI/G/1 queues where A(t) is restricted to have DMRL (IMRL),

$$\int_t^\infty \frac{H^c(x)\,dx}{H^c(t)} \underset{(\geqq)}{\leqq} \int_t^\infty \frac{A^c(x)\,dx}{A^c(t)} \qquad \text{all } t \geqq 0.$$

The inequalities are tight for the exponential distribution.

Proof: Let X = time from the last arrival to enter a busy period to the end of the busy period and let $X \sim \Phi(x)$ (see Figure 3).

Note that by definition (see section 1, Chapter 1) $\Phi(0^+) = 0$.

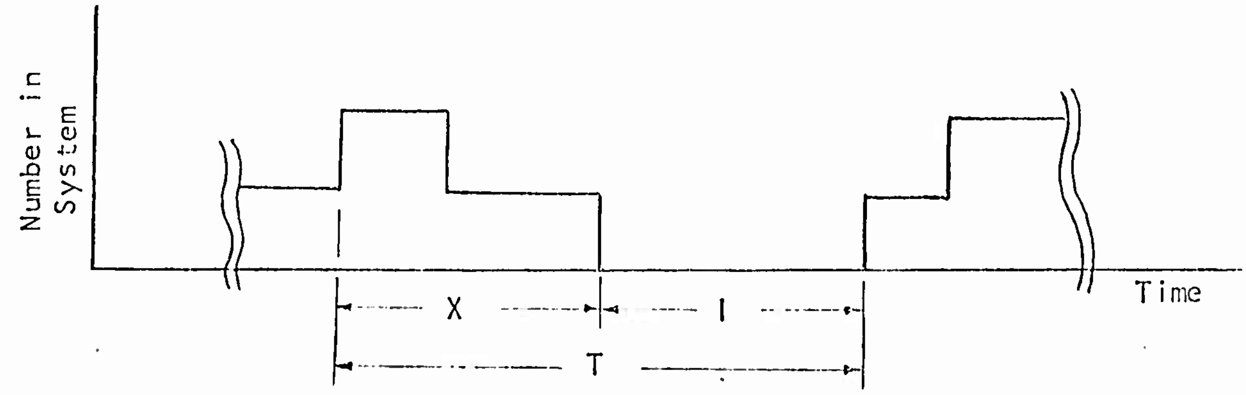

Figure 3: A Realization Showing the End of One Busy Period and the Start of the Next.

Let $\qquad H(t;\, x) = P[I \leqq t \mid X=x].$

It can be seen that $H^c(t;\, x) = \dfrac{A^c(t+x)}{A^c(x)}$ and $H(t) = \displaystyle\int_0^\infty H(t;\, x)\,d\Phi(x),$

where the integrals in this section are taken to be Lebesgue-Stieltjes integrals.

Conditioning on X,

$$\int_t^\infty \frac{H^c(u)\,du}{H^c(t)} = \int_t^\infty \int_0^\infty \frac{H^c(u;x)\,d\bar{\Phi}(x)}{H^c(t)} \cdot du$$

$$= \int_0^\infty \frac{H^c(t;x)}{H^c(t)} \int_t^\infty \frac{H^c(u;x)}{H^c(t;x)} \cdot du \cdot d\bar{\Phi}(x)$$

since the integral converge absolutely,

$$= \int_0^\infty \frac{H^c(t;x)}{H^c(t)} \int_{t+x}^\infty \frac{A^c(u)\,du}{A^c(t+x)} \cdot d\bar{\Phi}(x)$$

$$\underset{(\geqq)}{\leqq} \int_t^\infty \frac{A^c(u)\,du}{A^c(t)} \int_0^\infty \frac{H^c(t;x)}{H^c(t)}\, d\bar{\Phi}(x)$$

from the assumption of the theorem.

Hence,

$$\int_t^\infty \frac{H^c(u)\,du}{H^c(t)} \underset{(\geqq)}{\leqq} \int_t^\infty \frac{A^c(u)\,du}{A^c(t)}.$$

By letting t=0 one gets the following

<u>Corollary</u>: For (DMRL)/G/1 queues, ((IMRL)/G/1 queues),

$$\nu_h \underset{(\geqq)}{\leqq} \frac{1}{\lambda}. \qquad (19)$$

<u>Theorem 2.2</u>: For all GI/G/1 queues where A(t) is restricted to be IFR (DFR),

(i) $\dfrac{H(t+\Delta) - H(t)}{H^c(t)} \underset{(\leq)}{\geq} \dfrac{A(t+\Delta) - A(t)}{A^c(t)}$ for $\Delta > 0$, and all $t \geq 0$ where finite,

(ii) $\dfrac{H^c(t)}{A^c(t)} \underset{(\uparrow)}{\downarrow}$ all $t \geq 0$,

(iii) $\displaystyle\int_t^\infty \frac{H^c(u)du}{\nu_h} \underset{(\geq)}{\leq} \int_t^\infty \frac{A^c(u)du}{\nu_a}$ all $t \geq 0$.

The inequalities in (i)-(iii) are tight for the M/G/1 queue, and in this case the ratio in (ii) is constant and equal to 1.

Proof: Proofs are given for $A(t)$ not discrete.

(i) First, condition on X as in the proof of the previous theorem.

$$\frac{H(t+\Delta) - H(t)}{H^c(t)} = \int_0^\infty \frac{[H(t+\Delta;x) - H(t;x)]}{H^c(t)} \, d\Phi(x).$$

Now $H(t+\Delta;x) - H(t;x) = H^c(t;x) - H^c(t+\Delta;x)$

and substituting we get

$$\frac{H(t+\Delta) - H(t)}{H^c(t)} = \frac{1}{H^c(t)} \int_0^\infty \frac{A(t+\Delta+x) - A(t+x)}{A^c(x)} \cdot d\Phi(x)$$

$$= \frac{1}{H^c(t)} \int_0^\infty \frac{A(t+\Delta+x) - A(t+x)}{A^c(t+x)} \cdot \frac{A^c(t+x)}{A^c(x)} \, d\Phi(x)$$

$$\underset{(\leq)}{\geq} \frac{A(t+\Delta) - A(t)}{A^c(t)}$$ from the IFR, (DFR) assumption on $A(t)$.

(ii) Add and subtract 1 from both sides of (i),

$$1 - \frac{H^c(t+\Delta)}{H^c(t)} \underset{(\leqq)}{\geqq} 1 - \frac{A^c(t+\Delta)}{A^c(t)},$$

or

$$\frac{H^c(t)}{A^c(t)} \underset{(\leqq)}{\geqq} \frac{H^c(t+\Delta)}{A^c(t+\Delta)} \qquad \text{all } \Delta > 0, \quad t \geqq 0,$$

which proves part (ii). Notice that (i) and (ii) are equivalent as this argument is reversible.

(iii) From theorem 2.1

$$\frac{\int_t^\infty H^c(u)du}{\int_t^\infty A^c(u)du} \underset{(\geqq)}{\leqq} \frac{H^c(t)}{A^c(t)} \underset{(\geqq)}{\leqq} \frac{H^c(v)}{A^c(v)} \qquad \text{all } 0 \leqq v \leqq t$$

from (ii) and the fact that IFR (DFR) => DMRL (IMRL) (see Barlow and Proschan (1965)).

Putting this in determinant form,

$$\begin{vmatrix} \int_t^\infty H^c(u)du & H^c(v) \\ \int_t^\infty A^c(u)du & A^c(v) \end{vmatrix} \underset{(\geqq)}{\leqq} 0 \qquad \text{all } 0 \leqq v \leqq t.$$

Integrating v over (0,t)

$$\begin{vmatrix} \int_t^\infty H^c(u)du & \int_0^t H^c(u)du \\ \int_t^\infty A^c(u)du & \int_0^t A^c(u)du \end{vmatrix} \underset{(\geqq)}{\leqq} 0.$$

Adding the first column to the second gives

$$\begin{vmatrix} \int_t^\infty H^c(u)\,du & \nu_h \\ \int_t^\infty A^c(u)\,du & \nu_a \end{vmatrix} \underset{(\geqq)}{\leqq} 0$$

which proves (iii) and completes the proof of the theorem.

Part (iii) leads to the following

Corollary: For IFR/G/1 queues (DFR/G/1 queues)

$$\frac{\nu_h^{(2)}}{2\nu_h} \underset{(\geqq)}{\leqq} \frac{\nu_a^{(2)}}{2\nu_a} = \frac{\lambda}{2}\left(\sigma_a^2 + \left(\frac{1}{\lambda}\right)^2\right) = \frac{(c_a^2 + 1)}{2\lambda}. \tag{20}$$

Equality is taken on everywhere by the M/G/1 queue.

3. Bounds for Queues with Monotone Failure Rate Arrival Distributions.

a) The probability an arrival does not wait.

Recall from Chapter 1, theorem 1.1 that

$$a_o \nu_h = \left(\frac{1}{\lambda}\right)(1-\rho). \tag{21}$$

Using the corollary of theorem 2.1 above,

(i) For DMRL/G/1 queues

$$(1-\rho) \leqq a_o \leqq 1.$$

(ii) For IMRL/G/1 queues

$$0 \leqq a_o \leqq (1-\rho).$$

The lower bound in (i) and upper bound in (ii) are taken on by the Poisson arrival queue. The upper bound in (i) is taken on by the constant arrival, constant service queue.

b) The mean idle time.

Using equation 11 of Chapter 1 and the corollary to theorem 2.1 above gives

(i) For DMRL/G/1 queues

$$(\frac{1}{\lambda})(1-\rho) \leqq \nu_h \leqq \frac{1}{\lambda}.$$

(ii) For IMRL/G/1 queues

$$\frac{1}{\lambda} \leqq \nu_h.$$

The upper bound in (i) and lower bound in (ii) are taken on by the Poisson arrival queue. The lower bound in (i) is taken on by the D/D/1 queue.

c) The mean length and number served in busy period.

Using the relationships $E[B] = \frac{\rho}{(1-\rho)} E[I]$ and $E[N_b] = \mu E[B]$,

(i) For DMRL/G/1 queues,

$$\frac{1}{\mu} \leqq E[B] \leqq \frac{1}{\mu(1-\rho)},$$

$$1 \leqq E[N_b] \leqq \frac{1}{(1-\rho)},$$

the upper bounds being tight for the M/G/1 queue, and the lower ones for the D/D/1 queue.

(ii) For IMRL/G/1 queues,

$$\frac{1}{\mu(1-\rho)} \leqq E[B],$$

$$\frac{1}{(1-\rho)} \leqq E[N_b].$$

These bounds are tight and taken on by the M/G/1 queue.

d) The mean wait and number in queue.

Here use is made of the important queueing formula $E[N_q] = \lambda E[W]$ (see Little, (1961)). With this, equation (12) of Chapter 1 and the corollary to theorem 2.2 above, the following results are obtained.

(i) For all IFR/G/1 queues

$$J - \frac{(\rho+c_a^2)}{2\lambda} \leqq E[W] \leqq J, \tag{22}$$

and

$$\lambda J - \frac{(\rho+c_a^2)}{2} \leqq E[N_q] \leqq \lambda J, \tag{23}$$

where

$$J = \frac{c_a^2 + \lambda^2\sigma_g^2}{2\lambda(1-\rho)}.$$

The first expression shows that for this broad class of queues the expected wait has been bounded to within at most a mean inter-arrival time since $\rho < 1$ and for IFR arrivals $c_a^2 \leqq 1$ (see Barlow and Proschan (1965)). The second expression shows that the expected number in the queue has been bounded to within at most 1 customer.

An application of these to the important class of constant arrival, general service queues shows that the expected number in the queue has been bounded to within at most 1/2 customer, since in this case $c_a^2 = 0$.

The bounds in (22) and (23) are shown plotted in Figure 4 for $\lambda=1$, $c_a^2 = .9$, for different values of ρ, as a function of the variance of the service time. One might say that the expected wait increases "approximately linearly" with the variance of the service, with a slope of $\frac{\lambda}{2(1-\rho)}$. This is also true for the variance of the arrivals.

The lower bound is taken on by the M/G/1 queue and at least one member of the class takes on the upper bound, that being the constant arrival constant service queue.

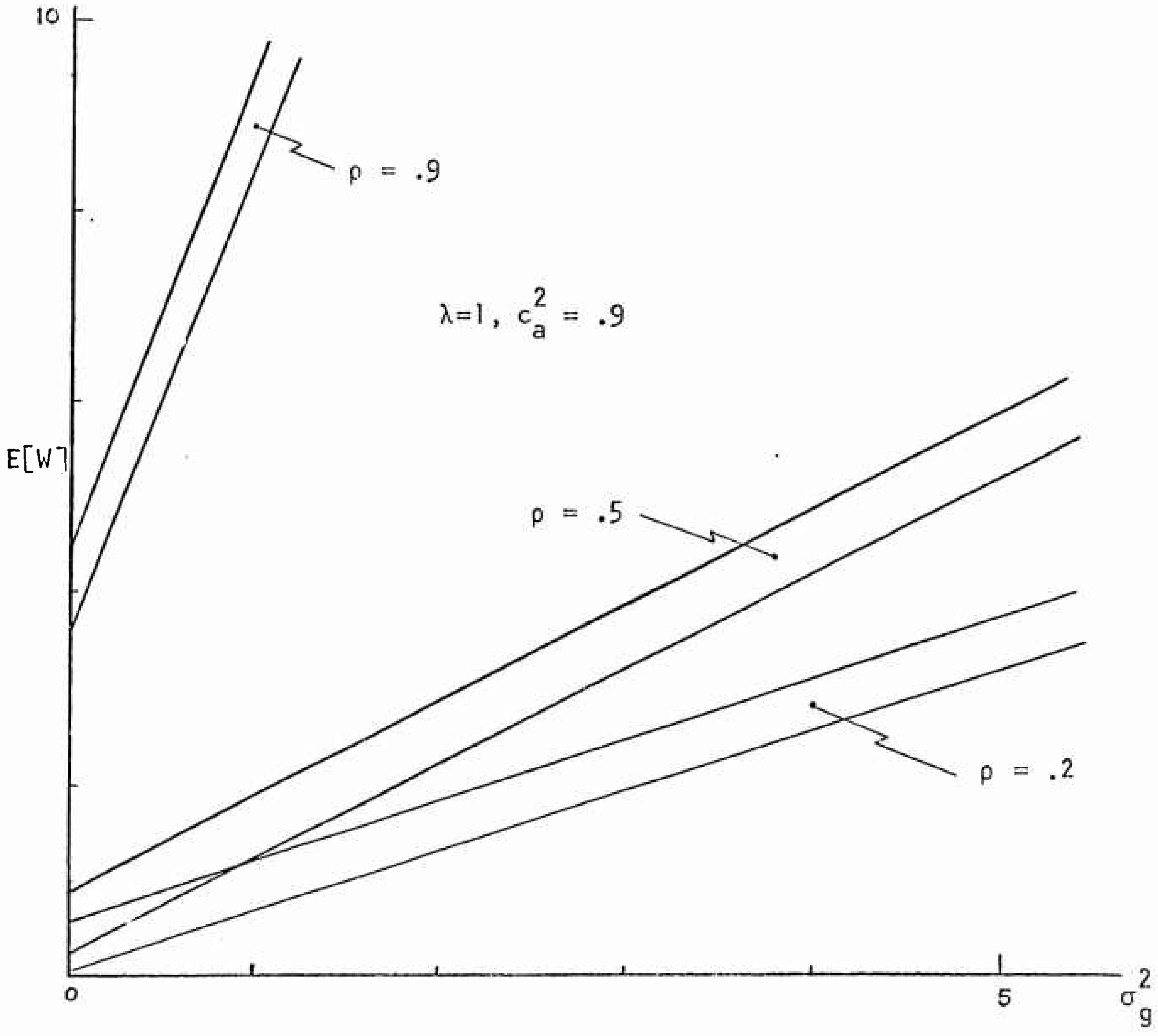

Figure 4. Bounds on E[W] for IFR/G/1 Queues.

(ii) For all DFR/G/1 queues

$$\ell \leqq E[W] \leqq \frac{c_a^2 + \lambda^2\sigma_g^2}{2\lambda(1-\rho)} - \frac{(\rho + c_a^2)}{2\lambda} \tag{24}$$

and

$$\lambda\ell \leqq E[N_q] \leqq \frac{c_a^2 + \lambda\sigma_g^2}{2(1-\lambda)} - \frac{\rho + c_a^2}{2}, \tag{25}$$

where ℓ solves equation (14). The upper bound is taken on by the M/G/1 queue. It has not been shown that the lower bound is taken on by any member in the class.

e) The variance of the wait in queue.

Theorem 2.2 allows bounds to be obtained on the variance of an equilibrium excess idle distribution, $\sigma^2_{e,h}$. If the first three moments of the arrival and service distributions are known this is the only unknown quantity in equation (5), which gives the variance of the wait. Clearly,

(i) For IFR/G/1 queues,

$$0 \leqq \sigma^2_{e,h} \leqq \frac{\nu_h^{(3)}}{3\nu_h} \leqq \frac{\nu_a^{(3)}}{3\nu_a},$$

and this with equation (5) gives upper and lower bounds on the variance of the wait which differ by no more than $\frac{2}{\lambda^2}$, since if A(t) is IFR

$$\frac{\nu_a^{(3)}}{3\nu_a} \leqq 2\nu_a^2 \leqq \frac{2}{\lambda^2}$$

(see Barlow and Proschan (1965)).

f) The variance of the output distribution.

The only unknown factor in the variance of the output is again

$\frac{\nu_h^{(2)}}{2\nu_h}$ and using theorem 2.2,

(i) For IFR/G/1 queues,

$$\sigma_g^2 \leqq \mathrm{Var}\,[\tau_n] \leqq \sigma_g^2 + \frac{1-\rho^2}{\lambda^2} ,$$

the upper bound being equality for the M/G/1 queue, and

(ii) For DFR/G/1 queues,

$$\sigma_g^2 + \frac{1-\rho^2}{\lambda^2} \leqq \mathrm{Var}\,[\tau_n] \leqq 2\sigma_g^2 + \sigma_a^2 - 2\iota\left(\frac{1}{\lambda} - \frac{1}{\mu}\right),$$

where ι solves equation (14).

g) The Virtual wait

Using the bounds in (22) and equation (6) give,

(i) For all IFR/G/1 queues,

$$\frac{\rho\,(c_g^2+c_a^2)}{2\mu\,(1-\rho)} \leqq E[V] \leqq \frac{\rho\,(c_g^2+c_a^2)}{2\mu\,(1-\rho)} + \frac{(\rho+c_a^2)}{2\mu}$$

and the expected virtual wait has been bounded to within at most $\frac{1}{\mu}$. The lower bound is equality for the M/G/1 queue. The upper bound is the Cesàro mean, or the mean wait of a random arrival in the D/D/1 queue.

(ii) For all DFR/G/1 queues,

$$\rho\left[\frac{(1+c_g^2)}{2\mu} + \iota\right] \leqq E[V] \leqq \frac{\rho\,(c_g^2+c_a^2)}{2\mu\,(1-\rho)}.$$

4. A Weaker Assumption on the Arrival Stream

Bounds on the expected number in the queue to within at most one customer can be obtained by weakening the IFR assumption on the arrival stream. We first prove the following

Theorem 2.3: For all GI/G/1 queues where $A(t)$ satisfies

$$\int_t^\infty \frac{A^c(u)du}{A^c(t)} \underset{(\geqq)}{\leqq} M < \infty \text{ for all } t \geqq 0,$$

then

$$\frac{\nu_h^{(2)}}{2\nu_h} \underset{(\geqq)}{\leqq} M.$$

Equality holds when $A(t)$ is exponential and $M = \frac{1}{\lambda}$.

Proof: The proof is similar to that of Theorem 2.1. Using the same notation as given there,

$$\int_t^\infty \frac{H^c(u)du}{H^c(t)} = \int_0^\infty \frac{H^c(t;x)}{H^c(t)} \int_{t+x}^\infty \frac{A^c(u)du}{A^c(t+x)} d\Phi(x)$$

$$\underset{(\geqq)}{\leqq} M \int_0^\infty \frac{H^c(t;x)}{H^c(t)} d\Phi(x) = M$$

from the assumption of the theorem.

Therefore,

$$\int_t^\infty \frac{H^c(u)du}{\nu_h} \underset{(\geqq)}{\leqq} M \frac{H^c(t)}{\nu_h} \quad \text{all } t \geqq 0.$$

Integrating over t gives

$$\frac{\nu_h^{(2)}}{2\nu_h} \underset{(\geqq)}{\leqq} M$$ and the theorem is proved.

Using this result and (12) in (4) gives,

For all GI/G/1 queues with $\rho < 1$ and

$$\int_t^\infty \frac{A^c(u)du}{A^c(t)} \leqq \frac{1}{\lambda} \quad \text{for all } t \geqq 0,$$

$$J - \frac{(1+\rho)}{2\lambda} \leqq E[W] \leqq J,$$

$$\lambda J - \frac{(1+\rho)}{2} \leqq E[N_q] \leqq \lambda J,$$

where
$$J = \frac{c_a^2 + \lambda^2\sigma_g^2}{2\lambda(1-\rho)} .$$

Example: Suppose we have m types of customers each requiring m different types of service, and we have one facility for each type with service distribution $G_i(t)$, mean $\frac{1}{\mu_i}$ and variance $\sigma_{g_i}^2$.

Customers arrive in a renewal process with inter-arrival time distribution A(t) which has DMRL (IMRL). The probability that an arrival is a type i customer is $p_i \geqq 0$, where $\sum_{i=1}^{m} p_i = 1$. The distribution of time between two type i arrivals does not in general have DMRL (IMRL). For example if in the DMRL case the original distribution A(t) is degenerate at $\frac{1}{\lambda}$, the distribution of time between type i arrivals is a step function which does not have DMRL. However, we show next that the assumption of Theorem 2.3 holds for each i, when $M = \frac{1}{\lambda p_i}$.

Consider a renewal process of all arrivals starting at t=0 with a type i arrival. Pick some arbitrary time $t_1 > 0$, and define

X_j = time between (j-1)-th and j-th arrival after t_1, j=1, 2, ...

X_{n,t_1} = time to n-th arrival after t_1.

Then
$$X_{n,t_1} = \sum_{j=1}^{n} X_j$$

$$E[X_{n,t_1}] = \sum_{j=1}^{n} E[X_j] = E[X_1] + \frac{(n-1)}{\lambda} \underset{(\geqq)}{\leqq} \frac{n}{\lambda}$$

from DMRL (IMRL) property. Now let X_{t_1} = time to first type i arrival after t_1.

Then
$$E[X_{t_1}] \underset{(\geqq)}{\leqq} \frac{1}{\lambda} \sum_{n=1}^{\infty} np_i(1-p_i)^{n-1} = \frac{1}{\lambda p_i},$$

independent of when the last type i arrived before t_1. Let T_i = time of last type i arrival before t_1. Then in particular

$$E[X_{t_1}|T_1=0] \underset{(\geqq)}{\leqq} \frac{1}{\lambda p_i}.$$

But
$$E[X_{t_1}|T_1=0] = \int_{t_1}^{\infty} \frac{A_i^c(\mu)\,d\mu}{A_i^c(t_1)}$$

and the result is at hand.

It follows immediately from theorem 2.3 that $\dfrac{\nu_{h_i}^{(2)}}{2\nu_{h_i}} \leqq \dfrac{1}{\lambda p_i}$ ∀ $i=1, 2, \ldots, m$

and for the i-th facility

$$J_i - \frac{(1+\rho_i)}{2\lambda p_i} \leqq E[W] \leqq J_i, \qquad (22a)$$

$$\lambda p_i J_i - \frac{(1+\rho_i)}{2} \leqq E[N_q] \leqq \lambda p_i J_i, \qquad (23a)$$

where
$$J_i = \frac{p_i c_a^2 + (\lambda p_i)^2 \sigma_{g_i}^2 + 1 - p_i}{2\lambda p_i(1-\rho_i)}, \quad \rho_i = \frac{\lambda p_i}{\mu_i}$$

5. Exponential Service Queues

The GI/M/1 queue has been studied at length by various authors and an important quantity to be determined in its analysis is δ, the unique non-zero root of the equation

$$z = \tilde{a}(\mu(1-z)) \tag{26}$$

where $$\tilde{a}(s) = \int_{0^-}^{\infty} e^{-st} dA(t).$$

It has been shown that (for example see Prabhu (1965 a))

$$a_o = P[W=0] = (1-\delta),$$

$$P[W \leqq x] = \delta(1-e^{-\mu(1-\delta)x}) \quad 0 < x,$$

$$P[\text{an arrival finds } j \text{ in system}] = (1-\delta)\delta^j \quad j=0,\ 1,\ 2,\ \ldots\ .$$

By restricting A(t) one can say something about the magnitude of δ in comparison to $\rho = \frac{\lambda}{\mu}$. In fact we prove here

Theorem 2.4: If A(t) is IFR (DFR) and if δ is the unique solution of

$$\{0 < z < 1,\ z = \tilde{a}(\mu(1-x))\} \text{ for } \frac{\lambda}{\mu} = \rho < 1, \text{ then } \delta \underset{(\geqq)}{\leqq} \rho.$$

Equality is taken on by the exponential distribution.

Proof: The proof is a direct application of the corollary 4.9 to Theorem 4.8 on page 33 of Barlow and Proschan (1965). There it is shown that if A(t) is IFR (DFR) with mean $\frac{1}{\lambda}$,

$$\tilde{a}(s) \underset{(\geqq)}{\leqq} \frac{\lambda}{\lambda+s} \qquad \text{for } s \text{ real.}$$

For $\tilde{a}(s) = \frac{\lambda}{\lambda+s}$, the solution of $\left\{0 < z < 1,\ z = \frac{\lambda}{\lambda+\mu(1-z)}\right\}$ is $z = \frac{\lambda}{\mu} = \rho$. For $\tilde{a}(s) \leqq \frac{\lambda}{\lambda+s}$ the solution must lie to the left of ρ and for $\tilde{a}(s) \geqq \frac{\lambda}{\lambda+s}$, it must lie to the right.

This leads to the following:

For the IFR (DFR)/M/1 queue

1) $$P[W > x] = \delta e^{-\mu(1-\delta)x} \underset{(\geqq)}{\leqq} \rho e^{-\mu(1-\rho)x} \qquad \forall\ x \geqq 0,$$

2) $$E[W] = \frac{\delta}{\lambda(1-\delta)} \underset{(\geqq)}{\leqq} \frac{\rho}{\lambda(1-\rho)},$$

3) $$E[N_q] = \frac{\delta}{1-\delta} \underset{(\geqq)}{\leqq} \frac{\rho}{(1-\rho)}.$$

Equality holds throughout for the M/M/1 queue.

6. Expected Number found by an Arrival

For the M/G/1 queue it can be shown that $E[N_q]$, the time average number in the queue is the same as the average number found by an arrival. Let L_a be the expected number in queue found by an arrival, and let $L_t = E[N_q]$. For the GI/G/1 queue one can show that

$$L_t = \rho L_a + (1-a_o)\lambda\gamma, \tag{27}$$

where γ = average excess service time of the customer in service when an arrival comes. The argument is similar to that leading to equation (6) for the virtual wait. For Poisson arrivals $a_o = (1-\rho)$,

$\gamma = \frac{\nu_g^{(2)}}{2\nu_g}$, the mean of an equilibrium excess service distribution, and hence

$$L_t = L_a = \frac{\rho^2(1+c_a^2)}{2(1-\rho)}.$$

If the service time distribution has DMRL (IMRL) it is easy to see that $\gamma \underset{(\geqq)}{\leqq} \frac{1}{\mu}$ with equality holding for exponential service, and (27) becomes

$$L_t \underset{(\geqq)}{\leqq} \rho L_a + (1-a_o)\rho.$$

Using section 3 part a) we have

(i) for DMRL/DMRL/1 queues

$$\frac{L_t}{\rho} -\rho \leqq L_a \leqq \frac{L_t}{\rho},$$

with the lower bound equality in the M/M/1 case,

(ii) for IMRL/IMRL/1 queues

$$0 \leqq L_a \leqq \frac{L_t}{\rho} - \rho,$$

with the upper bound an equality in the M/M/1 case.

Chapter 3

BOUNDS FOR BATCH ARRIVAL, BATCH SERVICE, AND INTERRUPTED SERVER QUEUES.

1. Introduction.

This chapter considers some generalizations of the GI/G/1 queue which preserve some important basic points in its underlying mathematical structure.

We first study a queue where customers arrive in batches of (possibly) random size. This is treated by redefining a service time. Queues with services occurring in batches of fixed size N are then analyzed, and solved by redefining an inter-arrival time. Finally a queue in which the service mechanism is subject to breakdown in a Poisson manner with a general repair time distribution is studied.

Each of these three generalizations of the basic GI/G/1 queue preserves the following structure. The (possibly) redefined inter-arrival or service times both form independent sequences of independent and identically distributed random variables. Also a sequence of waiting times can be found which satisfy the recursion formula

$$W_{n+1} = \text{Max}\ (0, W_n + S_n' - T_n'),$$

where the primes denote a possibly redefined service or inter-arrival time.

2. Queues with Batch Arrivals

Suppose "arrivals" appear at rate λ in a renewal process but that each "arrival" is now a batch of customers of random size η. Let $P(\eta = n) = \pi_n$,

$n=0, 1, 2, \ldots$. Customers in each batch are assumed to be numbered in some way to denote their order of service. The FIFO order of serving batches is maintained. Note that we allow a zero batch size, which is of significant importance as will be shown in an example.

Now let $S_{k,n}$ be the service time of the k-th customer in the n-th batch of a stationary queue, and $S_{k,n} \sim G(t)$ for all n and $k \geqq 1$, with $S_{0,n} \equiv 0$. Let $W_{b,n}$ be the waiting time (in queue) of the complete n-th batch, and define $S_n^* = S_{0,n} + S_{1,n} + \ldots + S_{\eta,n}$, $S_n^* \sim G^*(t)$. Then the sequence $\{W_{b,n}\}$ satisfies the relationship,

$$W_{b,n+1} = \text{Max}\,[0, W_{b,n}+S_n^*-T_n],$$

where T_n is the time between the arrival of the n-th and (n+1)-th batches, $T_n \sim A(t)$ for all n as before.

Noting that $\{S_n^*\}$ and $\{T_n\}$ form independent sequences of independent random variables, when $\lambda\nu_\pi < \mu$ the theorems in Chapter 1 are valid.

The expected wait of the first customer in each batch is thus given by

$$E[W_{b,n}] = \frac{\sigma_a^2+\sigma_{g^*}^2}{2(\frac{1}{\lambda} - \frac{\nu_\pi}{\mu})} + \frac{1}{2}(\frac{1}{\lambda} - \frac{\nu_\pi}{\mu}) - \frac{\nu_h^{(2)}}{2\nu_h}. \tag{30}$$

The variance $\sigma_{g^*}^2$ can be found easily by conditioning giving

$$\sigma_{g^*}^2 = \nu_\pi\sigma_g^2 + \frac{\sigma_\pi^2}{\mu^2}. \tag{31}$$

The average waiting time of an arbitrary customer is found as follows. The expected total additional wait of all customers in an average batch is found, and this is divided by the average number of customers per batch.

Let X_n be the total additional waiting time of all customers in some batch n. Then $X_n=0$ if $\eta=0$, 1, and for $\eta \geqq 2$,

$$X_n = S_{1,n} + (S_{1,n}+S_{2,n}) + \dots + (S_{1,n}+S_{2,n}+ \dots +S_{(\eta-1),n}).$$

Conditioning on η it follows that

$$E[X_n|\eta] = \frac{\eta(\eta-1)}{2\mu} \qquad \text{for all } \eta \geqq 0,$$

and hence

$$E[X_n] = \frac{\nu_\pi^{(2)}-\nu_\pi}{2\mu}.$$

The expected wait of any unspecified customer E[W] is then given in terms of the wait of the first one in each batch by

$$E[W] = E[W_{b,n}] + \left[\frac{\nu_\pi^{(2)}}{2\nu_\pi} - \frac{1}{2}\right]\frac{1}{\mu}.$$

Using this with (31) and (30) an expression is obtained for the expected wait which is bounded as follows under the assumption of IFR arrivals (between batches):

$$J_1+J_2 - \frac{(\rho+c_a^2)}{2\lambda} \leqq E[W] \leqq J_1+J_2 \tag{32}$$

$$\lambda\nu_\pi J_1 + \lambda\nu_\pi J_2 - \frac{\nu_\pi(\rho+c_a^2)}{2} \leqq E[N_q] \leqq \lambda\nu_\pi J_2 + \lambda\nu_\pi J_1 \tag{33}$$

where

$$J_2 = \frac{\lambda[\mu^2\sigma_a^2 + \mu^2\nu_\pi\sigma_g^2 + \sigma_\pi^2]}{2\mu^2(1-\rho)}, \quad \rho = \frac{\lambda\nu_\pi}{\mu}, \quad \text{and } J_1 = \left[\frac{\nu_\pi^{(2)}}{2\nu_\pi} - \frac{1}{2}\right]\frac{1}{\mu}.$$

Hence, the average number of customers in the queue is bounded to within at most an average batch size. For the compound Poisson input the lower bound is tight. For queues where $\int_t^\infty \frac{A^c(u)du}{A^c(t)} \leq \frac{1}{\lambda}$ for all $t \geq 0$, we can replace $\frac{\rho + c_a^2}{2\lambda}$ by $\frac{(\rho+1)}{2\lambda}$.

It is of interest to note that the bounds still hold even though the distribution of time between the arrival of customers is not IFR. For a random variable to have an IFR distribution, the distribution can have at most one jump, that being at the right hand end of its interval of support (for details see Barlow and Proschan (1965)). It is obvious that if there is a positive probability of having two or more customers in a single batch then there is a non-zero probability of a zero inter-arrival time between customers. In fact let T^* be the time between customer arrivals, $T^* \sim A^*(t)$, then considering only batches of non-zero size we find that

$$A^*(0^-) = 0 \quad \text{and} \quad A^*(0^+) = \frac{\nu_\pi - (1-\pi_o)}{\nu_\pi}.$$

These inter-arrival times are neither independent nor IFR.

As a special case of batch arrivals consider the example in section 4 of Chapter 2, where a customer from the original arrival stream stops at facility i with probability p_i. Since in our development we allow batches of zero size we can interpret this example as batch arrivals, where the batch size is 1 with probability p_i and 0 with probability $(1-p_i)$. Hence,

$\nu_\pi = p_i$ and $\sigma_\pi^2 = p_i(1-p_i)$. From (32) and (33), for the case where A(t) is assumed to have bounded mean residual life (as in section 4, Chapter 2),

$$J_i - \left[\frac{\rho_i+1}{2\lambda}\right] \leq E[W] \leq J_i, \tag{32a}$$

$$\lambda p_i J_i - \frac{p_i(\rho_i+1)}{2} \leq E[N_q] \leq \lambda p_i J_i, \tag{33a}$$

where $$J_i = \frac{p_i c_a^2 + (\lambda p_i)^2\sigma_g^2 + \rho_i^2(1-p_i)}{2\lambda p_i(1-\rho_i)}.$$

Comparing thes with (22a) and (23a) we see that this approach gives a better bound, as the expected number in queue is bounded to within at most p_i customers. It is easy to show the lower bounds are the same in both cases and are taken on by Poisson arrival queues.

It is also important to note that the batch arrival approach to this example gives even tighter bounds if A(t) is assumed to be IFR, whereas the approach in Chapter 2 breaks down in this case, since as shown above the distribution of time between customer arrivals is not IFR. The above approach gives the expected number in queue to within

$\frac{p_i(\rho_i+c_a^2)}{2}$ if A(t) is IFR.

All the DFR bounds of Chapter 2 hold when the correct means and variances are defined, since the idle time is still some excess of an inter-batch time.

3. Queues With Batch Services

In this section we assume a renewal input into a single channel queue with independent service times, but where service takes place only in batches of a fixed size of N customers. Service and inter-arrival times are again mutually independent.

Let $W_{N,n}$ be the wait in queue of the last (N-th) person who arrived and formed the n-th batch of N customers. By redefining an inter-arrival time to be

$$T_n^* = T_{0,n} + T_{1,n} + \dots + T_{(N-1),n}, \quad T_n^* \sim A^*(t),$$

where $T_{j,n}$ is the inter-arrival time between the j-th and (j+1)-th members of the n-th batch, we have

$$W_{N,n+1} = \text{Max}\,[0, W_{N,n} + S_n - T_n^*],$$

where S_n is the service time of the n-th batch.

The new inter-arrival time distribution is the N-th fold convolution of the original A(t). The IFR property is closed under convolutions and so with this assumption on A(t) the arguments of Chapter 2 can be applied. However, the DFR arguments fail since convolutions of DFR distributions are not necessarily DFR (for details see Barlow and Proschan, (1965)).

For the distribution $A^*(t)$ we see that

$$\nu_{a^*} = \frac{N}{\lambda}, \; \sigma_{a^*}^2 = N\sigma_a^2 \quad \text{and} \quad c_{a^*}^2 = \frac{c_a^2}{N}.$$

For any unspecified customer

$$E[W] = E[W_{N,n}] + \frac{1}{\lambda}\,\frac{(N-1)}{2},$$

and hence for IFR/G/1 queues with service in batches of fixed size N,

$$J - \frac{(N\rho + c_a^2)}{2\lambda} \leqq E[W] \leqq J, \qquad (34)$$

$$\lambda J - \frac{N}{2}\left[\rho + \frac{c_a^2}{N}\right] \leqq E[N_q] \leqq \lambda J, \qquad (35)$$

where $$J = \frac{Nc_a^2 + \lambda^2\sigma_g^2}{2N\lambda(1-\rho)}.$$

Thus the expected number in the queue is bounded to within at most $\frac{(N+1)}{2}$ customers.

It is well known that in generalizing the M/G/1 queue to batch services analytic problems arise and essentially the problem is the same as analyzing the E_N/G/1 queue. In order to calculate E[W] or $E[N_q]$ for this case it is necessary to determine the roots of the equation $(s+\lambda)^N = \lambda^N\tilde{g}(-s)$ for $Re(s) \geqq 0$, where $\tilde{g}(s) = \int_{0^-}^{\infty} e^{-st}dG(t)$ (for details see for example Prabhu 1965 (a)). Howev?r, all this can be avoided if the approximations in (34) and (35) give sufficient accuracy.

4. Queues with Server Breakdown

We now consider a GI/G/1 queue where the service mechanism can break down. When this occurs the normal queueing process is interrupted until the server is repaired, and it is assumed that these interruptions occur only in busy periods.

It will be assumed that

a) Times between successive breakdowns of the server are exponentially distributed and independent with mean $\frac{1}{\lambda_b}$.

b) Repair times are independent drawings $\{R_n\}$ from a distribution $R(t)$.

c) The customer in service when an interruption occurs resumes his service where he left off. This is called the Resume Rule.

To analyze such a queue we use a concept introduced by Gaver (1962) and dealt with in detail in his papers. This is the "completion time" of a customer, C_n for any customer n. It is the time the n-th customer spends actually occupying the service facility (his service time and all the repair time for breakdowns during his service). The assumption of exponential inter-breakdown times leads to $\{C_n\}$ being a sequence independent and identically distributed random variables. The sequence of waiting times (in queue) satisfy

$$W_{n+1} = \text{Max}\,[0,\ W_n+C_n-T_n]. \tag{36}$$

For the IFR or DFR assumption on the arrival stream all the bounds of Chapter 2 hold in terms of the mean and variance of the inter-arrival times and the mean and variance of a completion time C. These last two quantities are easily shown to be

$$E[C] = \frac{1}{\mu}(1 + \nu_r\lambda_b)$$

$$V[C] = (1 + \nu_r\lambda_b)^2\sigma_g^2 + \frac{\lambda_b}{\mu}(\sigma_r^2+\nu_r^2).$$

For a stationary queueing distribution to exist we must have $\frac{1}{\lambda} > E[C]$.

For any other distribution of time between breakdowns the completion times will not in general be independent or identically distributed and the method fails.

We can replace assumption c) by c_1),

c_1) The customer in service when an interruption occurs starts his service over again, but with the same service time. This is called the Repeat Identical Rule.

or by c_2)

c_2) The customer in service when ar. interruption occurs starts over again with a new independent service time. This is called the Repeat Different Rule.

For a detailed discussion of these see Gaver (1962).

In both the above cases the completion times form sequences of independent and identically distributed random variables which are independent of the arrival stream. Hence, all results for the GI/G/1 queue follow since consecutive waiting times satisfy the fundamental equation (36). Under the IFR or DFR arrival assumption the relevant bounds are thus obtained. Also under the assumption that $\int_t^\infty \frac{A^c(u)du}{A^c(t)} \underset{(\geq)}{\leq} \frac{1}{\lambda}$ the relevant bounds follow. All give bounds on the expected number in queue to within at most 1 customer.

Chapter 4

QUEUES WITH ADDED DELAY FOR THE FIRST CUSTOMER IN A BUSY PERIOD

1. Introduction

We now consider a single channel queue with renewal input at rate λ, independent service at rate μ, but where the first customer in every busy period suffers a random delay R before his service commences. This random variable R can be dependent on the preceding inter-arrival time and may affect the independence of the wait in queue of some customer n and the inter-arrival time of the n-th and the (n+1)-th customers. It is assumed that R is independent of the service times of all customers in the busy period generated by R.

In this chapter the random variable I will refer as before to the time from the end of a busy period until the next customer arrives. Hence, it is still some excess of an inter-arrival time. The idle time with respect to the server will be I+R.

A general expression is found for the expected waiting time in queue and is seen to be a generalization of equation (4) in Chapter 1. As an example we study the single server queue where the server waits until m people are present before starting on the first service in a busy period. He then continues to serve one customer at a time until the system is empty. We shall call this the $GI_m/G/1$ queue (see Heyman, 1966). When m=1 the subscript is dropped. For IFR arrivals bounds are obtained which give the expected number in queue or system to within $\frac{1}{2}$ customer. For arrivals with mean residual life bounded above by $\frac{1}{\lambda}$ the expected number is bounded to within $\left[\frac{m+1-c_a^2}{2m}\right]$.

2. General Results

By equating input with output we shall now find an expression for the expected waiting time in queue.

Let R be the delay before service commences for the first customer in a busy period in a stationary queue. This may be dependent on the previous inter-arrival time and hence on the idle time. It may also affect the independence of W_n and T_n as is seen in the example in section 3 (in this example the wait of the first customer in a busy period is in fact equal to the next (m-1) inter-arrival times). It is assumed that it has no effect on the independence of W_n and S_n. Writing the inter-output times as $\tau_n = S_{n+1} + X_n$, when customer n leaves the system busy, $X_n=0$, and when he leaves it empty, $X_n = I+R$, independent of S_{n+1}.

We now show that

$$E[W] = \frac{E[U^2]}{-2E[U]} + \frac{E[R^2] - E[I^2]}{2(E[R] + E[I])} + \frac{\text{Cov }(W_n, T_n)}{E[U]}. \tag{37}$$

Note that this is a generalization of equation (4) in Chapter 1.

To prove (37) we proceed as follows:

Equating input and output times

$$\tau_n = S_{n+1} + X_n$$

$$= T_n + D_{n+1} - D_n,$$

and hence,

$$W_{n+1} = X_n + (D_n - T_n), \tag{38}$$

where a) $$D_n - T_n \geq 0 \Rightarrow X_n = 0,$$

b) $$D_n - T_n < 0 \Rightarrow (D_n - T_n) = -I \text{ and } X_n = I+R,$$

c) $$E[X_n] = -E[U] \qquad \text{(from (38))}.$$

Squaring both sides of (38), taking expectations and assuming stationarity in the queue, we get

$$-2E[U]E[W] = E[U^2] + E[X_n(X_n+2(D_n-T_n)] - 2\text{Cov } (W_n,T_n). \tag{39}$$

Using a) and b) above we see that

$$E[X_n(X_n+2(D_n-T_n))] = E[(R+I)(R-I)]p \tag{40}$$

where $$p = P[(D_n-T_n) < 0].$$

But from c)

$$pE[R+I] = E[X_n] = -E[U]. \tag{41}$$

Substituting (40) and (41) in (39) gives (37).

Note that no assumption had to be made concerning the independence of R and I, and if R is independent of the arrival stream, Cov $(W_n,T_n) = 0$.

3. The $GI_m/G/1$ Queue

Suppose the server does rot start service until m customers are present. In this case the first customer served in a busy period waits for the next (m-1) customers to arrive before commencing service. Numbering the first customer in the busy period 1 for convenience, $R = T_1 + T_2 + \ldots + T_{m-1}$, so that

$$E[R] = \frac{(m-1)}{\lambda}, \; E[R^2] = \frac{(m-1)}{\lambda^2}\left[c_a^2 + m-1\right]. \tag{42}$$

For this case it is obvious that for the first (m-1) customers in each busy period the waiting time of a customer depends on the following inter-arrival time, and we must now calculate the non-zero covariance term.

The calculation of the covariance is achieved by conditioning on the position of service in a busy period. Obviously, a busy period serves at least m customers. We now use the covariance relationship for any three random variables X, Y, Z,

$$\mathrm{Cov}(X,Y) = E[\mathrm{Cov}(X,Y|Z)] + \mathrm{Cov}(E[X|Z], E[Y|Z]). \tag{43}$$

We interpret X to be a waiting time, Y the following inter-arrival time and Z the position of the customer's service in a busy period. Note that $E[Y|Z] = E[T_i|Z=i] = \frac{1}{\lambda}$, a constant. Hence in our case (43) reduces to

$$\mathrm{Cov}(W_n, T_n) = E[\mathrm{Cov}(W,T|Z)]. \tag{44}$$

If i indicates the position of service in a busy period, for i=1, 2, ...,m-1

$$W_i T_i = (T_i + T_{i+1} + \ldots + T_{m-1} + S_1 + \ldots + S_{i-1})T_i,$$

and so

$$\begin{aligned}\mathrm{Cov}(W,T|Z=i) &= \sigma_a^2 \qquad i=1,\ 2,\ \ldots,\ m-1\\ &= 0 \qquad i \geqq m.\end{aligned}$$

Averaging over busy periods,

$$P[Z=i] = \frac{1}{E[N_b]} \qquad i=1,\ 2,\ \ldots,\ m-1$$

and $$P[Z \geqq m] = \frac{E[N_b]-(m-1)}{E[N_b]}.$$

Using these results in (44) gives

$$\mathrm{Cov}(W_n,T_n) = \frac{(m-1)\sigma_a^2}{E[N_b]}.$$

But $$E[N_b] = \mu E[B] = \frac{\lambda(E[I] + E[R])}{(1-\rho)},$$ so finally

$$\mathrm{Cov}(W_n,T_n) = \frac{(m-1)(1-\rho)\sigma_a^2}{\lambda E[I] + m-1}. \tag{45}$$

Using (45) and (42) in (37) gives for the $GI_m/G/1$ queue with $\rho < 1$

$$E[W] = \frac{c_a^2+\lambda^2\sigma_g^2}{2\lambda(1-\rho)} + \frac{1}{2\lambda}(1-\rho) + \frac{(m-1)[m-1-c_a^2] - \lambda^2E[I^2]}{2\lambda(\lambda E[I] + m-1)}. \tag{46}$$

This expression depends only on the first two moments of the arrival, service, and idle distributions, and on m.

In general the distribution of I will depend on m (not the case for Poisson arrivals). For any stable queue ($\rho < 1$) I cannot be identically zero since busy periods end with probability 1. However, by using the apparently crude device of setting $E[I] = E[I^2] = 0$ in (46) we obtain an upper bound for all $GI_m/G/1$ queues with $\rho < 1$,

$$E[W] \leq \frac{c_a^2+\lambda^2\sigma_q^2}{2\lambda(1-\rho)} + \frac{(m-1)}{2\lambda} - \frac{(\rho+c_a^2)}{2\lambda} + \frac{1}{2\lambda}. \qquad (47)$$

We now show that this is actually a supremum for the class of queues considered by exhibiting a family of queues in the class with $E[I]$ and $E[I^2]$ arbitrarily small.

Consider a queue with deterministic arrivals at (normalized) rate 1, and deterministic service with service time for each customer $\frac{1-\varepsilon}{m}$, where $m \geq 2$ and $\varepsilon > 0$ is arbitrary. Then at stationarity it is easy to show that $I=\varepsilon$. For example take $m=3$, then $\frac{1}{\mu} = \frac{1}{3} - \frac{\varepsilon}{3}$. The true $E[W]$ is easily calculated to be $\frac{4}{3} - \frac{\varepsilon}{3}$. The above bound gives a value of $\frac{4}{3} + \frac{\varepsilon}{6}$. By letting $\varepsilon=0$ in this example the bound gives $\frac{4}{3}$ whereas the true $E[W] = 1$, since in this case $I = \frac{2}{3}$.

For IFR arrivals we now find a lower bound on $E[W]$. For any value of m it is easy to see that theorems 2.1 and 2.2 still hold for the random variable I. In the proofs of those theorems m would affect the distribution Φ, (of time from the last customer entering a BP until the end of the BP). However, the results were independent of the form of Φ, and we can say immediately that for IFR (DFR) arrivals and any $m \geq 1$,

$$E[I^n] \underset{(\geq)}{\leq} E[T^n] \text{ all } n \geq 1, \text{ and } \frac{E[I^2]}{2E[I]} \underset{(\geq)}{\leq} \frac{E[T^2]}{2E[T]}. \qquad (48)$$

We now prove a useful lemma.

Lemma 4.1: If (48) holds and a and b are any non-negative numbers,

$$\frac{a - E[I^2]}{b + E[I]} \underset{(\leqq)}{\geqq} \frac{a - E[T^2]}{b + E[T]}.$$

Proof: For $a \geqq 0$, using (48) we have

$$\frac{a}{E[I]} \underset{(\leqq)}{\geqq} \frac{a}{E[T]} \quad \text{and} \quad \frac{-E[I^2]}{E[I]} \underset{(\leqq)}{\geqq} \frac{-E[T^2]}{E[T]}.$$

Adding these and clearing the fraction gives

$$E[T](a-E[I^2]) \underset{(\leqq)}{\geqq} E[I](a-E[T^2]).$$

Also $\quad b(a-E[I^2]) \underset{(\leqq)}{\geqq} b(a-E[T^2]) \quad$ for $\; a,b \geqq 0$.

Adding these gives the desired result.

We apply the lemma to (46) with $a = (m-1-c_a^2)(m-1)$ for IFR arrivals since in this case $c_a^2 \leqq 1$. For DFR arrivals c_a^2 can be arbitrarily large, so to insure validity of the conditions of the lemma we let $a = (m-1)^2$ in this case.

Apply the lemma to (46) and using (47) we have shown for all IFR$_m$/G/1 queues with $\rho < 1$,

$$J \leqq E[W] < J + (\frac{1}{2\lambda}), \qquad (49)$$

$$\lambda J \leqq E[N_q] < \lambda J + \frac{1}{2},$$

where $$J = \frac{c_a^2+\lambda^2\sigma_g^2}{2\lambda(1-\rho)} + \frac{(m-1-\rho-c_a^2)}{2\lambda}.$$

The lower bound is taken on by Poisson arrivals, and as shown previously the upper bound is a supremum for at least one family of deterministic queues.

For the weaker assumption of bounded mean residual life, i.e., $\int_t^\infty \frac{A^c(u)du}{A^c(t)} \leqq \frac{1}{\lambda}$ it is easy to show that the bounds are

$$J - \frac{(1-c_a^2)}{2\lambda m} \leqq E[W] < J + \frac{1}{2\lambda},$$

$$\lambda J - \frac{(1-c_a^2)}{2m} \leqq E[N_q] < \lambda J + \frac{1}{2}. \tag{49a}$$

For DFR$_m$/G/1 queues with $\rho < 1$,

$$E[W] \leqq \frac{c_a^2+\lambda^2\sigma_g^2}{2\lambda(1-\rho)} + \frac{(m-1-\rho)}{2\lambda} - \frac{1}{2\lambda m}.$$

Equality occurs for Poisson arrivals only when m=1.

Chapter 5

QUEUES WITH 2 PRIORITY CLASSES

1. Arrivals in 2 Independent Renewal Processes

The following queueing situation is assumed in this section.

1. High priority customers (called type 1) arrive in a Poisson manner at rate λ_1. Their service times are drawn from some general service distribution $G_1(t)$, with mean $\frac{1}{\mu_1}$.

2. Low priority customers (called type 2) arrive as a renewal process with distribution of inter-arrival times $A_2(t)$, with mean $\frac{1}{\lambda_2}$. Their service times are drawn from a common distribution $G_2(t)$, with mean $\frac{1}{\mu_2}$.

3. The preemptive resume rule applies to all type 2 customers, (see Chapter 3, section 4, and Gaver (1962).

For convenience in this section we shall call the above 3 assumptions $\mathcal{A}$.

A busy period can start with the service of either a type 1 or a type 2 customer. It is important to note that those starting with a type 2 customer have the same structure as the busy periods (BP's) for the interrupted server case treated in section 4 of Chapter 3, with $\lambda_b = \lambda_1$ and $R(t)$ the distribution of a BP in an M/G/1 queue with the service distribution $G_1(t)$.

For the study of this queueing system we shall use the following notation. A BP started with a type i customer will be called a type i BP. The phrase "a type 1 service period" refers to the time the server is busy serving consecutive type 1 customers (i.e., a normal BP in the imbedded M/G/1 queue formed by the type 1 customers).

B = length of an arbitrary BP.

I = length of an arbitrary idle period.

$B^{(i)}$ = length of a type i BP, i=1, 2.

$N^{(i)}$ = total number of customers served in a type i BP, i=1, 2.

$N_j^{(i)}$ = number of type j customers served in a type i BP, i,j=1, 2.

C = completion time of a type 2 customer (see Chapter 3, section 4).

$\rho_1 = \frac{\lambda_1}{\mu_1}$, $\rho_2 = \frac{\lambda_2}{\mu_2}$, $\lambda = \lambda_1+\lambda_2$, $\frac{1}{\mu} = \left[\frac{\lambda_1}{\lambda_1+\lambda_2}\right]\frac{1}{\mu_1} + \left[\frac{\lambda_2}{\lambda_1+\lambda}\right]\frac{1}{\mu_2}$,

$\rho = \frac{\lambda}{\mu}$ and hence $\rho = \rho_1+\rho_2$.

Arbitrary Idle and Busy Periods

Using the above definitions and relating input to output, we have for $\rho < 1$,

$$E[B] = \frac{\rho}{1-\rho} E[I].$$

By assuming $A_2(t)$ has IFR or DFR we now obtain bounds on $E[I]$ and hence on $E[B]$.

Theorem 4.1: Under $\mathcal{A}$, if $A_2(t)$ has IFR (DFR) and $\rho < 1$, then

$$E[I] \underset{(\geqq)}{\leqq} \frac{1-\tilde{a}_2(\lambda_1)}{\lambda_1},$$

where $\tilde{a}_2(s) = \int_{0^-}^{\infty} e^{-st} dA_2(t)$. Equality is taken on when $A_2(t) = 1-e^{-\lambda_2 t}$.

Proof: Let $I \sim H(t)$, Y be the time from the end of the last BP to the next arrival of type 1, and Z be the time from the end of the last BP to the next arrival of type 2. Then $Y \sim$ Exponential with mean $\frac{1}{\lambda_1}$. Let $Z \sim F(t)$.

Under the IFR (DFR) assumption we know from Theorem 2.2 part (ii) that $F^c(t) \underset{(\geqq)}{\leqq} A_2^c(t)$.

Now $I = \text{Min}\ (Y,Z)$ and hence

$$H^c(t) = e^{-\lambda_1 t} F^c(t) \underset{(\geqq)}{\leqq} e^{-\lambda_1 t} A_2^c(t).$$

Integrating both sides gives the desired result.

Let a_o = P(an arbitrary arrival finds the system empty). Note that a_o is not the probability an arbitrary customer does not have to wait in queue. Now using theorem 1.1 with theorem 4.1 and $a_o \leqq 1$ gives:

For all 2 priority queues where $\mathcal{A}$ holds and

i) $A_2(t)$ if IFR,

$$\frac{1}{\lambda} - \frac{1}{\mu} \leqq E[I] \leqq \frac{1-\tilde{a}_2(\lambda_1)}{\lambda_1},$$

$$\frac{1}{\mu} \leqq E[B] \leqq \left[\frac{1-\tilde{a}_2(\lambda_1)}{\lambda_1}\right]\left[\frac{\rho}{1-\rho}\right].$$

The upper bounds are tight for Poisson type 2 arrivals.

ii) $A_2(t)$ is DFR,

$$\frac{1-\tilde{a}_2(\lambda_1)}{\lambda_1} \leqq E[I],$$

$$\left[\frac{\rho}{1-\rho}\right]\left[\frac{1-\tilde{a}_2(\lambda_1)}{\lambda_1}\right] \leqq E[B].$$

The bound is tight for Poisson type 2 arrivals.

Bounds for the expected number served in an arbitrary BP follow immediately.

Type 2 Busy Periods

As noted above these BP's have the same structure as those in the "server-breakdown" case dealt with in section 4 of Chapter 3. From the results there it is easily seen that

$$E[C] = \frac{1}{\mu_2(1-\rho_1)}.$$

Now if I_s is the idle time in the server breakdown case it is easy to show that

$$E[B^{(2)}] = \frac{\rho_2}{(1-\rho)} E[I_s],$$

and we immediately obtain the following bounds on $E[B^{(2)}]$.

i) For $A_2(t)$ IFR and under $\mathcal{A}$:

$$\frac{(1-\rho_2)}{(1-\rho)} \frac{1}{\mu_2} \leqq E[B^{(2)}] \leqq \frac{1}{\mu_2(1-\rho)}. \tag{50}$$

The upper bound is taken on by Poisson type 2 arrivals.

ii) For $A_2(t)$ DFR and under $\mathcal{A}$:

$$\frac{1}{\mu_2(1-\rho)} \leqq E[B^{(2)}],$$

with equality for $A_2(t) = 1-e^{-\lambda_2 t}$.

Since $E\left[N_2^{(2)}\right] = \frac{E[B^{(2)}]}{E[C]} = E[B^{(2)}]\mu_2(1-\rho_1)$ bounds on $E\left[N_2^{(2)}\right]$ follow immediately.

Each interruption of a type 2 service results in a type 1 service period. In each of these type 1 service periods the expected number served is $\frac{1}{(1-\rho_1)}$. Therefore, the expected number of type 1 customers served in each type 2 completion time is $\frac{\lambda_1}{\mu_2(1-\rho_1)}$. Hence $E\left[N_1^{(2)}\right] = \frac{E\left[N_2^{(2)}\right]\lambda_1}{\mu_2(1-\rho_1)}$, and finally we have

$$E[N^{(2)}] = (\mu_2(1-\rho_1) + \lambda_1)E[B^{(2)}].$$

Bounds now follow on $E[N^{(2)}]$ from the above inequalities on $E[B^{(2)}]$.

Waiting Times

The low priority customers have no effect on the high priority type when the pre-emptive rule applies. Hence, the waiting time of the type 1 is given by the Pollaczek-Khintchine formula. We can view type 2 customers in a similar way to Chapter 4. The type 2 customers form a general arrival, general service queue where service times are redefined to be completion times, and the first customer in a busy period has some random delay R, which is either zero or the remainder of a busy period of type 1 customers. Hence, formula (37) with Cov $(W_n, T_n) = 0$ gives the expected wait of a type 2 customer. Using this and the lemma in Chapter 4 we get for the type two customers if $A_2(t)$ is IFR,

$$\frac{E[(T^{(2)}-C)^2]}{2E[T^{(2)}-C]} + \frac{E[R^2] - E[T^{(2)^2}]}{2[E[R]+E[T^{(2)}]]} \leqq E[W] \leqq \frac{E[(T^{(2)}-C)^2]}{2E[T^{(2)}-C]} + \frac{E[R^2]}{2E[R]}. \qquad (51)$$

However, to bound E[R] and $E[R^2]$ seems to be very difficult. For Poisson type 2 arrivals it can be shown that, if $R \sim R(t)$, and has density $r(t)$, then

$$R^c(t) = B_1^c(t)\tilde{m}(\lambda_2) - \frac{1}{\lambda_2} r(t),$$

where $B_1(t)$ is the distribution of a busy period of type 1 customers, $\tilde{m}(s) = \int_0^\infty e^{-st} dM(t)$ and $M(t)$ is the renewal function for the modified renewal process $\{X_i\}$, where $X_1 \sim \exp(\lambda_1)$ and for $i > 1$, $X_i \sim$ Busy cycle of type 1 customers. Using this we get

$$\frac{E[R^2]-E[I^2]}{2[E[R]+E[I]]} = \frac{\nu_{b_1}^{(2)}}{2\nu_{b_1}} - \frac{1}{\lambda_2}.$$

Substituting this in (37) gives an exact expression for E[W] which checks with results in Wei Chang (1965).

2. Arrivals in a Single Renewal Process

In this section we assume that inter-arrival times form a single renewal process. With probability p the arrival is a type 1 (high priority), and with probability (1-p) it is a type 2. Assuming the pre-emptive resume rule applies the completion times of type 2 customers are no longer independent and identically distributed (unless the arrival stream is Poisson). However, idle periods (when the facility is empty) are excess inter-arrival times and under the IFR, DFR or bounded mean residual life assumption these can be bounded. As in section 1 we then get bounds on the expected length of an arbitrary busy period.

Waiting Times

Under the pre-emptive rule the average wait of a high priority customer can be bounded under the various assumptions on the arrival stream. For example, (32a) applies directly when the inter-arrival times have bounded mean residual life. This surprisingly gives a better bound than the approach in Chapter 2 (see 22a). In the non-preemptive case the type 1 customer who starts a type 1 service period must wait until any type 2 in service leaves the system. Hence, (51) holds where in this case R is the remaining service time of a type 2 if one is present, and R=0 if a type 1 arrival finds the system empty. If the type 2 customers have a service distribution which has bounded mean residual life, then $0 \leq \frac{E[R^2]}{2E[R]} \leq \frac{1}{\mu_2}$ and using this in (51) gives the expected wait of a type 1 customer to within $\frac{1}{\lambda_2} + \frac{1}{\mu_2}$, and hence the expected number of type 2 in the queue to within $(1+\rho_2)$.

Since in this model the completion times of the type 2 customers do not form a sequence of independent random variables, it is difficult to say much about their expected waiting time by the approaches used in this paper.

Chapter 6

MORE GENERAL QUEUES

1. Introduction

We now relax some of the independence assumptions of previous chapters while maintaining stationarity. Tandem queues are taken as a particular case of dependent input. In this case an expression is found for the total expected wait and is seen to depend on some unknown covariance terms.

2. Stationary Queues

When $E[U] < 0$ and the sequence $\{U_n\}$ is strictly stationary and metrically transitive, Loynes (1962 (a)) has shown that the queue has a stationary waiting time distribution. Also $E[W]$ and $E[I]$ exist and equation (1) still holds for this class of queues; that is,

$$a_o E[I] = -E[U].$$

Note that the proof of theorem 1.1 required stationarity but not independence.

3. Queues with Independent Services

If the service times form an independent sequence and the inter-arrival times form a stationary sequence, from equation (2) we can obtain a generalization of theorem 1.2. Using the same method of proof, but realizing that now W_n and T_n are not necessarily independent it is easy to show that

$$E[W] = \frac{E[U^2]}{-2E[U]} - \frac{E[I^2]}{2E[I]} + \frac{\mathrm{Cov}(W_n,T_n)}{E[U]}. \qquad (52)$$

Given the form of the dependence of the inter-arrival stream it may in a given case be possible to calculate the covariance term.

4. Tandem Queues

Let us consider now m single channel queues in tandem, where the output from one becomes the input into the next. We shall assume unlimited queueing space before each facility so that no blocking occurs, that the original input forms a renewal process, and that the service times in each facility are independent sequences of independent random variables. That is, if $S_{k,i}$ is the service time of customer k at stage i, $S_{k,i} \sim G_i$, i=1, 2, ..., m, and $\{S_{k,i}\}$ are mutually independent.

The input stream to all facilities after the first will not in general be a sequence of independent random variables (the case of Poisson arrivals with exponential service at each stage is the exception). Let

$\sigma^2_{a_i}$ = variance of input to stage i

$\sigma^2_{g_i}$ = variance of service at stage i

μ_i = service rate at stage i, $\rho_i = \frac{\lambda}{\mu_i}$

$\nu^{(n)}_{h_i}$ = n-th idle moment at stage i

$W^{(i)}_n$ = waiting time of n-th customer at stage i (in queue)

$T^{(i)}_n$ = inter-arrival time at stage i

$\tau^{(i)}_n$ = inter-output time at stage i

From (52) we immediately obtain

$$E[W^{(i)}] = \frac{\lambda(\sigma^2_{a_i}+\sigma^2_{g_i}) + (1-\rho_i)^2}{2(1-\rho_i)} - \frac{\nu^{(2)}_{h_i}}{2\nu_{h_i}} - \frac{\lambda_i \, \mathrm{Cov}(W_n^{(i)},T_n^{(i)})}{(1-\rho_i)}.$$

But $T_n^{(i-1)} = T_n^{(i)}$ and hence from section 5 of Chapter 1,

$$\sigma^2_{a_i} = \sigma^2_{g_{i-1}} - \frac{(1-\rho_{i-1})^2}{\lambda^2} + \frac{(1-\rho_{i-1})}{\lambda} \, \frac{\nu^{(2)}_{h_{i-1}}}{\nu_{h_{i-1}}} \qquad i=2,\ 3,\ \ldots,\ m.$$

Now if $\mu_i = \mu$, giving $\rho_i = \rho$ for all $i=1,\ 2,\ \ldots,\ m$, the expected <u>total</u> wait in queue is given by

$$E\left[\sum_{i=1}^{m} W^{(i)}\right] = \frac{\lambda(\sigma^2_a+2\sigma^2_{g_1}+\ldots+2\sigma^2_{g_{m-1}}+\sigma^2_{g_m}) + (1-\rho)^2}{2(1-\rho)} - \frac{\nu^{(2)}_{h_m}}{2\nu_{h_m}} - \frac{\lambda}{(1-\rho)} \sum_{i=1}^{m} \mathrm{Cov}(W_n^{(i)},T_n^{(i)}).$$

Note that there is cancellation of the idle time moments of all except the last facility (this happens only when ρ is the same for each facility).

The completely unknown quantities in the equation are the covariance terms. The author has had little success in determining their order of magnitude. It may be possible to find bounds on these quantities in some generality since each term is the covariance between the wait in queue of the <u>n-th</u> customer and the time until the <u>(n+1)-th</u> customer arrives. It seems to the author that in general this correlation would be small.

SUMMARY

The aim of this paper is to find simple expressions which approximate some of the measures of performance in the GI/G/1 queue. The large body of queueing literature shows that exact expressions for many of these measures are extremely complicated. Often they are implicit in nature, making them impractical for direct application. Notable exceptions are the papers of Kingman (1962 (a) and (b)) and Newell (1965). In one paper Kingman deals with the GI/G/1 queue, and in the other he deals with asymptotic properties (as $\rho \to 1^-$) of stationary queues. Newell's paper deals mainly with traffic light problems.

In Chapter 1, the moments of the waiting time and idle time are related by equating input with output in a stationary queue. New expressions are found for the mean and variance of the wait, and an expression for the mean queue length follows immediately. Upper bounds for all GI/G/1 queues are found easily from the non-negativity of the idle time variance and the bound of 1 on the probability a customer finds the system empty (Kingman (1962 (b)) finds the same upper bound for the expected wait). These bounds are in terms of the means and variances of the arrival and service streams. A non-trivial lower bound on the expected wait is found which requires knowledge of the arrival and service distributions.

In Chapter 2 we recognize that an idle time distribution is some complicated tail distribution of an inter-arrival time. Restrictions are placed on the arrival distribution which enable us to obtain "good" bounds on such measures as the mean wait and mean number in queue. When the mean residual life of an inter-arrival time is assumed bounded above by an ordinary mean inter-arrival time, that is, $\int_T^\infty \frac{A^c(u)du}{A^c(T)} \leqq \frac{1}{\lambda}$ all $T \geqq 0$, we

find bounds which give the mean queue length to within $\frac{(1+\rho)}{2}$. When the stronger assumption is made that the arrival distribution has increasing failure rate (see Chapter 2) the mean queue length is bounded to within $(\frac{c_a^2+\rho}{2})$, with $c_a^2 \leqq 1$ (for the D/G/1 queue this reduces to $\frac{\rho}{2}$). These bounds are in terms of <u>only</u> the means and variances of the arrival and service distributions. Bounds on the mean idle time give bounds on the mean length of a busy period, and hence on the mean number served in a busy period. Bounds on the mean <u>actual</u> wait give bounds on the mean <u>virtual</u> wait. Upper and lower bounds are also found when the mean residual life of an inter-arrival times is bounded below (that is, $\int_T^\infty \frac{A^c(u)du}{A^c(T)} \geqq \frac{1}{\lambda}$ all $T \geqq 0$), and when the arrival distribution has decreasing failure rate.

Chapter 3 deals with these generalizations of the GI/G/1 queue. Batch arrivals are treated by redefining service times, and when $A(t)$ has increasing failure rate or has mean residual life bounded above by $\frac{1}{\lambda}$, the mean queue length is bounded to within at most an average batch size. Queues with service in batches of fixed size N are treated by redefining inter-arrival times. For IFR arrivals the mean queue length is bounded to within $\frac{N}{2}(\rho + \frac{c_a^2}{N})$. Queues are also considered where the service breaks down in busy periods in a Poisson manner.

Queues where the first customer in each busy period has some added delay are dealt with in Chapter 4. The $GI_m/G/1$ queue (where m customers start a busy period) is used to illustrate the results. For $m > 1$ simple bounds are found which give the mean queue length to within $\frac{1}{2}$ when $A(t)$ is IFR.

Some two-priority queues are dealt with in Chapter 5, and two models are considered. In the first one the arrivals are assumed to generate two renewal processes, and high priority customers are Poisson. In the second one the arrivals are assumed to generate a single renewal process, and the probability that an arrival is a high priority customer is p ((1-p) that he is a low priority customer).

The last chapter deals with more general queues. Stationarity is retained but the independence assumptions are weakened. Some preliminary results on tandem queues are given, but the expression for the expected wait has some covariance terms of unknown order of magnitude.

There is much to be done in the are of approximations in queueing. The importance of the idle time distribution is clearly demonstrated in this paper. More work needs to be done in relating its properties to those of the arrival and service streams.

REFERENCES

1. R. E. Barlow and F. Proschan (1965), "Mathematical Theory of Reliability", Wiley.
2. W. Feller (1966), "An Introduction to Probability Theory and its Applications, Volume 2", Wiley.
3. D. P. Gaver (1962), "A Waiting Line with Interrupted Service, Including Priorities", Jour. Roy. Stat. Soc. series B, Vol. 24, No. 1.
4. D. P. Heyman (1966), "Optimal Operating Policies for Stochastic Service Systems", Ph.D. Thesis, College of Engineering, University of California, Berkeley.
5. J. F. C. Kingman (1962 (a)), "On Queues in Heavy Traffic", Jour. Roy. Stat. Soc. series B, Vol. 24.
6. J. F. C. Kingman (1962 (b)), "Some Inequalities for the GI/G/1 Queue", Biometrika, Vol. 49, Nos. 3 and 4.
7. J. D. C. Little (1961), "A Proof of the Queueing Formula L=λW", Jour. ORSA, Vol. 9, No. 3.
8. R. M. Loynes (1962 (a)), "The Stability of a Queue with Non-Independent Inter-arrival and Service Times", Proc. Comb. Phil. Soc., Vol. 58, No. 3.
9. R. M. Loynes (1962 (b)), "Stationary Waiting Time Distributions for Single-Server Queues", Ann. Math. Stat., Vol. 33.
10. G. F. Newell (1965), "Approximation Methods for Queues with Application to the Fixed-Cycle Traffic Light", SIAM Review, Vol. 7, No. 2.
11. N. U. Prahbu (1965 (a)), "Queues and Inventories", Wiley.
12. N. U. Prahbu (1965 (b)), "Stochastic Processes", MacMillan.
13. J. Riordan (1962), "Stochastic Service Systems", Wiley.

14. Wei Chang (1963), "Pre-emptive Priority Queues", Jour. ORSA, Vol. 13, No. 5.

Unclassified
Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)* University of California, Berkeley	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b GROUP

3. REPORT TITLE
SOME INEQUALITIES FOR SINGLE SERVER QUEUES

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*
Research Report

5. AUTHOR(S) *(Last name, first name, initial)*
Marshall, Kneale T.

6. REPORT DATE August 1966	7a. TOTAL NO. OF PAGES 73	7b. NO. OF REFS 14

8a. CONTRACT OR GRANT NO. Nonr-222(83) b. PROJECT NO. NR 047 033 c. Research Project No. RR 033-07-01 d.	9a. ORIGINATOR'S REPORT NUMBER(S) ORC 66-19 9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*

10. AVAILABILITY/LIMITATION NOTICES
Distribution of this document is unlimited

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Mathematical Science Division Office of Naval Research Washington, D. C. 20360

13. ABSTRACT The expected wait in the GI/G/1 queue is related to the mean and variance of the idle time. For arrival distributions which are IFR or have mean residual life bounded by $\frac{1}{\lambda}$, simple bounds are obtained which give, for example, the expected number in queue to within at most one customer.

By equating input with output, relations between random variables are used to obtain expressions for the moments of the waiting time in terms of moments of the inter-arrival, service, and idle time distributions. By bounding the idle time moments, bounds are obtained on the mean and variance of the waiting time, the mean length of a busy period, and the probability an arrival finds the system empty. Bounds on the mean wait lead to bounds on the expected virtual wait.

Similar results are obtained for some generalizations of the GI/G/1 queue, including batch arrivals, batch service and priority queues. Queues where the first customer in each busy period has some added delay are also considered.

Some preliminary results for tandem queues are given.

Unclassified

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
queuing						
approximations						
monotone failure rates						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2*a*. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2*b*. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7*a*. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7*b*. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8*a*. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8*b*, 8*c*, & 8*d*. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9*a*. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9*b*. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through

________________."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through

________________."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through

________________."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS)*, *(S)*, *(C)*, or *(U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, roles, and weights is optional.