ANALYSIS OF OVERLOAD PERFORMANCE FOR A CLASS OF M/D/1 PROCESSOR QUEUEING DISCIPLINES

by

H. Heffes
AT&T Bell Laboratories
Holmdel, NJ

ABSTRACT

We consider the overload performance of a class of queueing and service disciplines aimed at controlling the load offered to a processor in such a way as to keep delays, for served customers, small. In particular, we analyze one such control scheme, the M/D/1 LIFO queueing discipline, where the customer rejection mechanism corresponds to arrivals to a full buffer pushing out the oldest customer in the queue. Comparisons with the M/D/1 FIFO finite buffer scheme and related M/M/1 results show the significant effect of the queueing discipline, and the variability of unloading the input buffer, on system performance. We present results for the delay distribution of served customers and for the throughput-delay tradeoffs. We also present an approximation for the system performance. We present results for the delay distribution of served customers and for the throughput-delay tradeoffs. We also present an approximation for the system performance.

1. INTRODUCTION

When a call processing system is overloaded, long delays can result in customers abandoning or turning "bad" (e.g., dialing before receiving dial-tone). For the situation where customers in queue turn "bad" at a random time after their arrival we present results for the throughput of good customers. Here, the results show a strong dependence on the mechanism for customers turning bad.

In some overloaded call processing systems, long delays can result in customers abandoning or turning "bad" (e.g., dialing before receiving dial-tone). For the situation where customers in queue turn "bad" at a random time after their arrival we present results for the throughput of good customers. Here, the results show a strong dependence on the mechanism for customers turning bad.

In this paper we analyze one such control scheme, the M/D/1 LIFO queueing discipline where the customer rejection mechanism corresponds to arrivals to a full buffer pushing out the oldest customer in the queue. The LIFO-pushout scheme has been shown to have desirable performance, in overload, for the M/M/1 queue when compared to other queueing disciplines [4,5]. However, for some control mechanisms, e.g., a rate based control scheme [6], where the rate of new customers into the system is limited (e.g., one customer admitted every T seconds), a deterministic "service time" assumption is more appropriate.

For the M/D/1 LIFO-pushout scheme, we present results for the delay distribution of served customers and for the throughput-delay tradeoff operating characteristic, which is a useful basis for comparisons; improved schemes resulting in smaller delays for a given throughput. In an environment where a customer in queue, unknown to the system, can turn "bad" at a random time after its arrival, we present results for the throughput of successful services ("goodput") [1]. For the results in this paper we have used the same service time for good and bad customers and consider two distribution functions for the time a customer in queue remains good. The results exhibit a strong dependence on the mechanism for customers turning bad and thus indicate the importance in characterizing the mechanism in such an environment.

Comparisons are made with the FIFO; finite buffer, discipline which we show to be "equivalent" to an infinite buffer FIFO scheme where customers time out after spending an appropriately defined time in queue. These disciplines are defined as follows:

(i) LIFO-Pushout (LIFO-PO)
Last-in-first-out service, a finite buffer of size N - 1; a customer arriving to see a full buffer pushes out the oldest customer in the buffer. This is a component of the overload strategy discussed in [3].

(ii) FIFO-Blocking (FIFO-BL)
First-in-first-out (FIFO) service; a finite buffer of size N - 1; a customer arriving to see a full buffer leaves immediately. This is the M/D/1/N queue.

(iii) FIFO-Timeout (FIFO-TO)
FIFO discipline, infinite buffer; every arriving customer joins the queue but will leave at a time T₀ after arrival if it is still in the buffer at that time.

In addition to comparing the above schemes we make comparisons with results [5] for exponentially distributed service times to determine the sensitivity of performance to the service distribution. Based on the observed sensitivity an approximation for the M/G/1 case is suggested. These results can be used to estimate the effect of overload control strategies with varying degrees of regularity of unloading the input buffer (e.g., window based strategies [6]).

The M/D/1 LIFO-pushout discipline is analyzed by considering the partial differential-difference equations for the probability a tagged customer in a given queue position gets served with remaining delay not exceeding a given value, conditioned on the elapsed service time of the customer in service. The solution is in terms of boundary coefficients obtained by solving a system of linear equations. The throughput of good customers for the FIFO case, which we use for comparison, is obtained by recognizing its equivalence to an appropriately defined timeout problem, and using level crossing ideas [7].

We note that although the results are developed for a single server system they can be used to approximately analyze overload control schemes which control access to distributed systems (e.g., analysis of dial-tone delays for a pre-dial control of a distributed switching machine [6]). They can also be used to study the class of SPC overload controls treated in [13], under the LIFO-pushout discipline.
2. ANALYSIS OF CONTROL SCHEME

The positions in the buffer are numbered 2, 3, ..., N with the processor numbered 1. An arriving customer goes into service (position 1) if the system is empty, otherwise it goes into position 2. If an arrival occurs while a customer is waiting in position 2, the waiting customer moves into position 3 and the new arrival moves into position 2. Of course a service completion brings the positions of all waiting customers down by 1. If a customer is in position N and an arrival occurs then the waiting customer gets pushed out and the arrival joins position 2.

2.1 Delay Distribution of Served Customers

Let 
\[ P_r = P(\text{an arrival gets served}), \]
\[ p_0 = P(\text{an arrival sees an empty system}), \]
\[ f_r(t) = \text{density function for the waiting time of customers who get served}, \]
\[ G(j,x,t) = P(\text{a customer in position } j \text{ gets served and its remaining waiting time } \leq t \mid x = x), \]

where \( x \) denotes the elapsed service time of the customer in service, and

\[ g(j,x,t) = \frac{\partial}{\partial t} G(j,x,t). \]

Denoting \( \lambda \) as the arrival rate of the Poisson arrival process and \( T \) as the deterministic service time, we have

\[ f_r(t) = \frac{1-P_0}{P_r} \int_0^T g(2,x,t) \frac{dx}{T}, \quad t > 0. \]

We denote

\[ \tilde{g}(2,t) = \int_0^T g(2,x,t) \frac{dx}{T} \]

which has the interpretation as the marginal density of delay and being served for those customers arriving to a busy system. Using Little's Law we have

\[ \lambda T P_r = 1 - p_0, \]

and obtain

\[ f_r(t) = \lambda T \tilde{g}(2,t), \quad t > 0, \quad (2-1a) \]

with an atom at zero

\[ F_r(0) = \frac{P_0}{P_r}. \quad (2-1b) \]

The delay distribution determination reduces to the determination of \( g(2,x,t) \) since \( p_0 \) can be obtained from the M/D/1 finite buffer problem [8] or as a byproduct of the analysis presented here. To analyze the system we have

\[ G(j,x,t) = G(j,x+h,t-h) \quad (1-\lambda h) + G(j+1,x+h,t-h) \lambda h + o(h), \]

where we identify \( G(N+1,x,t) = 0 \), from which we obtain the partial differential-difference equations,

\[ \frac{\partial}{\partial t} G(j,x,t) - \frac{\partial}{\partial x} G(j,x,t) + \lambda \ G(j,x,t) = \lambda \ G(j+1,x,t), \]

\[ (1<j\leq N, 0<x<T) \quad (2-2) \]

where \( G(N+1,x,t) = 0 \).

From (2-2) we obtain

\[ \frac{\partial}{\partial x} g^*(j,x,t) = (\lambda + T) g^*(j+1,x,t) - \lambda g^*(j+1,x,t), \]

\[ (1<j\leq N, 0<x<T) \]

with \( g^*(N+1,x,t) = 0 \), and where the Laplace-Stieltjes transform

\[ g^*(j,x,t) = \int_0^\infty e^{-\xi t} dG(j,x,t). \]

This set of differential - difference equations can be solved backwards (with respect to \( t \), in terms of the boundary conditions, \( g^*(j,x,t) \), to yield

\[ g^*(j,x,t) = e^{\lambda T} \sum_{i=0}^{\infty} \frac{(-\lambda T)^{i-j}}{(i-j)!} g^*(i,0,t). \quad (2-3) \]

Using the boundary conditions

\[ g^*(j=1,0,t) = g^*(j,0,0), \quad 2 \leq j \leq N \]

with

\[ g^*(1,0,t) = 1, \]

we obtain the set of linear equations for the desired boundary conditions

\[ g^*(j=1,0,t) = e^{\lambda T} \sum_{i=2}^{\infty} \frac{(-\lambda T)^{i-2}}{(i-2)!} g^*(i,0,t) \]

\[ 2 \leq j \leq N \quad (2-4a) \]

and

\[ 1 = e^{\lambda T} \sum_{i=2}^{\infty} \frac{(-\lambda T)^{i-2}}{(i-2)!} g^*(i,0,t). \quad (2-4b) \]

For a given \( s \), these equations (2-4a) can be recursively solved backwards, using (2-4b) for normalizing the solution. Finally, we denote the Laplace transform of \( g^*(2,t) \)

\[ \bar{g}^*_{(2,s)} = \int_0^\infty g^*(2,x,s) \frac{dx}{T} = (\bar{g}(2,s)) \]

and use (2-3) to obtain

\[ \bar{g}^*_{(2,s)} = \int_0^\infty \bar{g}^*_{(2,x,s)} \frac{dx}{T} = (\bar{g}(2,s)) \]

Inversion of (2-5) yields \( \bar{g}(2,t) \) which, together with (2-1), gives the desired density function for the waiting time of served customers. To obtain the atom at zero we can use the normalization condition

\[ 1 = P_r + \int_0^\infty f_r(t) dt = P_0 + \lambda T \bar{g}^*(2,0). \]

which gives

\[ F_r(0) = 1 - \lambda T \bar{g}^*(2,0). \quad (2-6) \]
The throughput is then given by
\[ \lambda P_e = \frac{\lambda}{1 + \lambda T (1 - g(2,0))}. \] (2-7)

### 2.2 Mean Delay of Served Customers

To obtain the mean delay of served customers,
\[ M = \lambda T \int_0^s \bar{g}(2,s) \, ds = -\lambda T \int_0^s \frac{\partial}{\partial s} g(2,s) \, ds. \]
we differentiate (2-5) which gives
\[ g'(s) = \frac{1}{\lambda} \sum_{0 \leq j \leq r} \left( \frac{(-\lambda T)^j}{(j-k)!} \frac{e^{\lambda T}}{\lambda} \right) g'(i,0,0) + \frac{1}{\lambda^2} \sum_{0 \leq j \leq r} \left( \frac{(-\lambda T)^j}{(j-k)!} \frac{e^{\lambda T}}{\lambda} \right) \left( 1 - \frac{k-j+1}{\lambda} \right) g'(i,0,0) \]
\[ + \left( 1 - \frac{k-j+1}{\lambda} \right) g'(i,0,0) \] (2-8)
The quantities \( g'(i,0,0) \) are obtained by solving the linear equations (2-4) at \( s = 0 \) and the quantities \( g'(i,0,0) \) are obtained from the following linear equations obtained by differentiating (2-4):
\[ -T e^{-\lambda T} g'(i-1,0,0) = -e^{-\lambda T} g'(i-j,0,0) \]
\[ + \sum_{0 \leq j \leq r} \frac{(-\lambda T)^j}{(i-k)!} \frac{e^{\lambda T}}{\lambda} g'(i,0,0); \] (2-9a)
for \( 2 < j = N \) and
\[ -T e^{-\lambda T} = \sum_{0 \leq j \leq r} \frac{(-\lambda T)^j}{(i-j)!} \frac{e^{\lambda T}}{\lambda} g'(i,0,0). \] (2-9b)

### 3. 'GOODPUT' RESULTS

The next set of results correspond to the situation where a customer in queue, unknown to the system, turns "bad" at a random time after its arrival [1]. Thus serving a customer with delay in excess of this random time results in an unsuccessful service. For the results in this paper we have used the same service time for good and bad customers. Clearly the delay distribution of served customers and the distribution of the time at which a customer turns bad determine the rate at which the system serves good customers (goodput). Defining
\[ P(t) = Pr \left[ \text{customer in queue for } t \text{ seconds is good} \right] \]
we consider two cases:

**Case I:** \[ P(t) = e^{-\alpha t} \]
and

**Case II:** \[ P(t) = \begin{cases} 1 & t \leq \tau \\ 0 & t > \tau \end{cases} \]

The goodput, \( V \), is given by
\[ V = \lambda P_e \int_0^T P(t) \, dF_e(t) \]
which results in
\[ V_1 = \lambda p_e + \lambda (1 - p_e) g(2,\alpha). \]
From (2-7) and, \( p_e = 1 - \lambda \, T \, P_e \), we obtain
\[ V_1 = \frac{\lambda}{1 + \lambda T (1 - g(2,0))} \left( 1 + T \left( g(2,\alpha) - g(2,0) \right) \right) \] (3-1)
For Case II we clearly have
\[ V_{II} = \frac{\lambda}{1 + \lambda T (1 - g(2,0))} F_e(t). \] (3-2)

### 4. NUMERICAL RESULTS

In this section we present system performance measures for the queuing discipline studied, in overload, and compare the results with the M/D/1 FIFO-BL scheme. We also make comparisons with the M/M/1 FIFO-PO queue to see the sensitivity of the performance measures to the service distribution. Specifically we present numerical results for the tails of the delay distribution, throughputs, mean delay tradeoffs and the effect of customers turning "bad".

Figure 1 shows the conditional mean delay results as a function of the offered load (\( \rho = \lambda T \)), ranging into overload conditions. The unit of time in all the numerical results is the service time. The FIFO-PO scheme as well as the FIFO-BL scheme (obtained from tabulated results in [9]) are shown for three values of \( N \). We observe the significant range of results with the FIFO-PO scheme being more than an order of magnitude smaller than the FIFO-BL scheme for \( N = 11 \) and \( \rho = 2 \). We also note the decreasing behavior as a function of \( \rho \), for the pushout scheme, as \( \rho \) becomes large. While the FIFO results clearly approach \( (N-1)T \), the limiting pushout behavior is explained as follows. As \( \rho \) becomes large, an arriving customer enters the first waiting position and either goes into service, if the next arrival occurs after a service completion, or gets pushed back to the second waiting position, where his chances of being served become vanishingly small. As \( \rho \) gets large it is the last arrival during a service time that gets held of the server and the mean remaining service is the mean delay experienced by this last arrival. Thus
\[ E[W_{\text{served}}] = \frac{1}{\lambda \rho}. \]
This limiting behavior has been observed numerically, requiring up to \( \rho = 5 \), for some cases. We thus have
\[ E[W_{\text{served}}]_{\text{FIFO-PO}} = \frac{2}{(N-1)\rho} \]
as \( \rho \) becomes large.

In Figure 2 we look at the results for a given value of \( \rho = 1.6 \) and plot the throughput as a function of the conditional mean delay. These throughput-delay tradeoff curves, generated by varying \( N \), can be used to compare the performance of the schemes for a given maximum allowable throughput or processor utilization. Conversely for a given maximum mean delay e.g., \( E[W_{\text{served}}] = 1 \), the throughput of the FIFO-PO scheme is more than 5 percent greater than that of the FIFO-BL scheme. Figure 3 shows the throughput-delay tradeoffs for \( \rho = 1.0 \). If the maximum allowable processor utilization is e.g., 0.95 then the corresponding mean delays are 2.6 for the FIFO-PO scheme as compared to 4.5 for the FIFO-BL scheme. When comparing Figures 2 and 3, it should be kept in mind that to achieve a given throughput one requires a larger \( N \) for the smaller \( \rho \).

The sensitivity of the throughput-mean delay tradeoff curves to the service distribution is shown in Figures 4 and 5 which show the results for the FIFO-PO schemes with deterministic service and exponential service [5] for \( \rho = 1.0 \) and 1.5 respectively. We note from Figure 4 that, for a given throughput, say 0.95, the mean delays even for the FIFO-PO scheme are quite sensitive to the service
distribution with $E[W|\text{served}] = 2.9$ for deterministic service and $E[W|\text{served}] = 6.0$ for exponentially distributed service. It is interesting to note that for a given throughput the mean delay for a deterministic service times are approximately $\frac{1}{2}$ the mean delays for exponential service times.

The factor $\frac{1}{2}$ is the ratio of the second moments of the service distribution. The approximate tradeoff curves for FIFO-PO (M/D/1), generated in this manner, are also shown in Figures 4 and 5 and suggest obtaining approximate M/G/1 FIFO-PO tradeoffs by scaling the M/M/1 mean results for a given throughput with the ratio of the second moments of the service distribution.

We note further that scaling the M/M/I FIFO-PO tradeoffs in the same manner results in a fairly good approximation to the M/D/I FIFO-PO tradeoffs. This is shown in Figure 6 and suggests using the same scaling to approximate the tradeoffs for other queuing disciplines studied in [5] for more general service distributions.

For a given buffer size, approximate mean delays can then be obtained by evaluating the throughput and reading off the mean delay for the approximate tradeoff. We note that the throughput can generally be simply evaluated since for a given service distribution and buffer size the throughput is often independent of the service discipline e.g., the throughputs for FIFO-BL, FIFO-PO, FIFO-CO can be evaluated using the simpler FIFO-BL discipline. To test out the approximation we draw the exact tradeoff curve for Erlangian service time with two stages, i.e., the M/E2/1 FIFO-BL system, using results in [9] and compare it with the approximate curve obtained by scaling the mean delays for the M/M/I FIFO-BL system at each throughput by the ratio of the second moments of the service distributions 0.75. Figure 7 shows the excellent degree of approximation.

The results have shown the significant advantages of using the FIFO-BL discipline in addition to the significant advantage of more regular unloading of the input queue, a fact that was observed when comparing average dial-tone delay in a distributed switching machine for a rate based and a window based overload control [6].

In Figures 8 and 9 we show the delay distribution tails for the overload case $p = 1.6$, and for $N = 11$ and $N = 2$ respectively. We note that the FIFO-BL scheme serves all accepted customers within $(N-1)$ service times and this scheme is exactly equivalent, with respect to the distribution of the waiting time of served customers, to a FIFO-Timeout scheme with the timeout interval $T_{\text{opt}} = (N-1)T$. We note that the shape of the distribution tail in Figure 8 is very similar to the FIFO-PO scheme in [5]. For the FIFO-PO scheme it is possible for a customer's position to oscillate and therefore we have the small tail for $t > 10$. As expected we see that most of the customers get served very quickly for the FIFO-PO scheme (approximately 80 percent get served within a service time), whereas less than 1 percent of the customers wait less than 5 service times for the FIFO-BL scheme. The corresponding mean delays are 9.0 and 0.97 for the FIFO-BL and FIFO-PO schemes respectively.

The complementary delay distributions, $F_{\text{FIFO-PO}}(t)$, for $N = 2$ can be shown to satisfy

$$F_{\text{FIFO-PO}}(t) - F_{\text{FIFO-BL}}(t) = e^{-t}, 0 \leq t < T.$$

In the previous examples we compared the delay and throughput performance. Here we look at the goodput performance for the FIFO-PO and FIFO-BL schemes, where we choose the control parameters (buffer size) to maximize the goodput, as a function of the severity of the overload. Figure 10 shows the results for Case I, for both deterministic and exponential service times, where the average time a customer remains good is $a^{-1} = 10$ service times. We see the superiority of the FIFO-PO scheme over FIFO-BL as well as the sensitivity to the service time distribution. The goodput results for the M/D/I FIFO-Bl scheme are obtained using the previously recognized equivalence with the FIFO-Timeout case, and level crossing arguments [7] to analyze the Timeout case. Details are in the Appendix.

In Figure 11 we show the corresponding maximum goodput results for Case II which corresponds to the case where a customer in queue turns bad exactly $r$ time units after its arrival. We choose $r = 10$, which gives the same average time for a customer in queue remaining good as the Case I examples. The maximum goodput results for the M/D/I FIFO-BL scheme are obtained by choosing $N = 11$, which ensures an accepted customer initiates service within 10 service times. We note however (see Figure 1) that the average time spent in queue by a customer is large (9.0 service times at $p = 1.6$). The M/D/I FIFO-PO scheme, which has excellent delay performance, has more limited goodput for this case. In heavy overload ($p = 1.6$) the small (0.85 percent) goodput reduction can be traded of by a 9:1 reduction in the average time a customer (good or bad) spends in queue for the FIFO-PO scheme. At a smaller overload ($p = 1.1$) the 7.1 percent goodput reduction is traded off by only a 2:1 reduction in the average time a customer (good or bad) spends in queue. We again note the sensitivity of the results to the service distribution, particularly for the FIFO-BL scheme. The M/M/I results are explainable by observing that crossing of the complimentary conditional waiting time distributions, for FIFO-PO and FIFO-BL, occurs further out on the tail, as $p$ increases. For deterministic service times, the crossing of the maximum goodput curves does not occur due to the finite support of the FIFO-BO distribution and the selection of the maximizing buffer size.

The numerical results have demonstrated some potential advantages of the overload control queueing discipline studied. We note that to reduce the time to practice, and realize its significant potential, one needs to consider implementation questions related to the application. Examples of such issues are the effect of retries, methods for rewarding "patient customers", the possibility that good and bad customers have different service times and the use of appropriate distributions for when customers turn bad. These issues are discussed in [3], for a given application, and similar approaches may be applicable to address some of these issues here. We note, however, that whether one is faced with effectively different service times for good and bad customers depends on the application. If this scheme is used to model a rate based overload control mechanism [6] where the "service time" corresponds to the reciprocal of the controlled rate over a time interval, say 10 seconds, then lower real time requirements for unsuccessful calls would serve to slightly lower the "service times" for all customers, good and bad, over the next 10 second time interval.

5. ACKNOWLEDGMENT

The author wishes to thank B. T. Doshi for his helpful discussions.

REFERENCES


We evaluate the Case I goodput results for the M/D/1 FIFO-BL scheme. For the probability of getting served, the throughput and the waiting time for customers who get served the FIFO-BL scheme, with buffer size $N - 1$, is equivalent to the FIFO-TO scheme with timeout interval $T_o = (N-1)T$. This in turn is equivalent to the following scheme:

Let $x_t$ denote the work in the system (processor and buffer) at time $t$. If an arrival occurs at time $t$ and $x_t < T_o$, then it joins the buffer. It will be served and its waiting time will be $x_t$. If, on the other hand, an arrival at time $t$ sees $x_t > T_o$, then it will leave.

Let $J$ denote the distribution of work in the system. Then

$P_i = P(x < T_o) , \quad F_i(o) = J(o) F \overset{p}{=} P_i , \quad \text{and} \quad f_i(t) = J'(t) F = \frac{dJ}{dt} F , \quad 0 < t < T_o .

We note that $P_i$ and $F_i(o)$ (and therefore $J(o)$) are identical with the corresponding quantities for the LIFO-PO scheme and are thus known. It suffices to obtain $j(t)$. By level crossing arguments [7] we obtain

$$ j(t) = \lambda \int_{0}^{t} \bar{G}(t-y) dy + \lambda J(o) \bar{G}(t) $$

with normalization condition

$$ J(0) + \int_{0}^{T_o} j(y) dy = 1, \quad (A-1b) $$

where $\bar{G}$ is the complementary service time distribution,

$$ \bar{G}(x) = \begin{cases} 1 & x < T \\ 0 & x > T. \end{cases} $$

In general one can solve the integral equation (A-1) by converting it to a first order differential equation which can then be solved over successive intervals of length $T$ using the solution from the previous interval. Our concern here is to evaluate the Case I goodput, given by

$$ V_i = \lambda F_i \int e^{-\alpha t} dF_i(t) .$$

Using arguments similar to those in [10] we define

$$ h(x) = \lambda \bar{G}(x) $$

and let $m(x)$ be the renewal density for $h(x)$. Then $m(x)$ satisfies [11]

$$ m(x) = h(x) + \int_{0}^{x} h(x-y) m(y) dy . \quad (A-2) $$

Equation (A-1) can now be written as, for $0 < t < T_o$,

$$ j(t) = J(o) h(t) + \int_{0}^{t} j(y) h(t-y) dy . \quad (A-3) $$

This is a renewal equation whose solution is

$$ j(t) = J(o) h(t) + \int_{0}^{t} h(t-x) m(x) dx , \quad 0 < t < T_o $$

which simplifies, upon using (A-2), to

$$ j(t) = J(o) m(t) , \quad 0 < t < T_o . \quad (A-4) $$

We note that

$$ \int_{0}^{T_o} h(x) dx = \lambda T = \rho $$

and the function $m(t)$ is well defined for any finite $x$ irrespective of the value of $0 < \rho < \infty$. Using (A-4) in the goodput expression gives

$$ V_i = \lambda J(o) + \lambda J(o) \int_{0}^{T_o} e^{-\alpha t} m(t) dt . \quad (A-5) $$

Since $J(o)$ is known we only need to evaluate the integral in (A-5). To evaluate this we define

$$ y(t) = \int_{0}^{t} e^{-\alpha s} m(x) dx $$

with Laplace transform

$$ y(s) = \tau [y(t)] = \frac{1}{s} \theta (s + \alpha) \quad (A-6) $$

where $\theta(s)$, the Laplace transform of $m(t)$, is given by

$$ \theta(s) = \frac{\theta'(s)}{1 - \theta(s)} = \frac{\lambda(1-e^{-\alpha T})}{\lambda + \lambda e^{-\alpha T}} .$$

We evaluate the integral by numerically inverting the transform [12] in (A-6) at the point $t = T_o$. 

---


FIGURE 1  MEAN DELAY CHARACTERISTICS
(Q DISCIPLINE COMPARISONS)

FIGURE 2  THROUGHPUT - MEAN DELAY TRADEOFFS (p = 1.6)
(Q DISCIPLINE COMPARISONS)

FIGURE 3  THROUGHPUT - MEAN DELAY TRADEOFFS (p = 1.0)
(Q DISCIPLINE COMPARISONS)

FIGURE 4  THROUGHPUT - MEAN DELAY TRADEOFFS (p = 1.0)
(SERVICE DISTRIBUTION COMPARISONS)

FIGURE 5  THROUGHPUT - MEAN DELAY TRADEOFFS (p = 1.5)
(SERVICE DISTRIBUTION COMPARISONS)

FIGURE 6  THROUGHPUT - MEAN DELAY TRADEOFFS (p = 1.6)
(SERVICE DISTRIBUTION COMPARISONS)
FIGURE 7 THROUGHPUT-MEAN DELAY TRADEOFF ($\rho=1.6$) (SERVICE DISTRIBUTION COMPARISONS)

FIGURE 8 DELAY DISTRIBUTION COMPARISONS ($\rho=1.6$)

FIGURE 9 DELAY DISTRIBUTION COMPARISONS ($\rho=1.6$)

FIGURE 10 MAXIMUM GOODPUT COMPARISONS-CASE I

FIGURE 11 MAXIMUM GOODPUT COMPARISONS-CASE II