ABSTRACT

In a queueing system, performance is often measured in terms of the output process. For example, the rate of departures from the system measures the rate at which work is done. If an analytic expression is available for the output process, the performance of the system can then be predicted under both normal and overload conditions. A prediction of the transient response may be very important under overload conditions since it may allow corrective action to be taken before a steady state is attained. An analytic model for the output process may also be useful in the analysis of queueing networks where the output from one queue often forms the input to another.

This paper considers an M/M/s queueing system for which a simple recursive differential equation relating the input and output streams under transient conditions is derived. More specifically, equations for the joint probability of k departures from, and N arrivals to, the queueing system in some interval of time (0,t) are derived. Solutions to these equations are given. The method is also applied to a system of two M/M/s queues which have a different number of servers and different service rates but which are supplied by the same input stream (i.e. each Poisson arrival supplies two customers, one to each queue). This can be considered as a system of two Markovian queues which receive highly correlated input streams, and as such may be used to derive a bound on the performance of systems with correlated arrivals.

1. INTRODUCTION

In a queueing system, the output process has importance in a number of applications. For example, the performance of a queueing system may be specified in terms of its output, since the rate of departures is a measure of the rate at which work is done. For many applications it is also important to know how the system operates under heavy load conditions. Both overload and nominal performance can be predicted from an analytic model which relates the input and the output of the queue. There exists a large amount of literature on the steady state properties of queueing output processes (see Burke [1] for a survey) but fewer results are available for transient queues. A prediction of transient response can be very important under overload conditions since corrective action may be taken before a steady state is attained.

An analytic model of the output process may also be useful in the analysis of queueing networks where the output stream from one queue often forms the input stream to another.

Results for the transient properties of some queueing systems have been obtained by other authors. For the M/G/\infty queue, results have been obtained by Mirasol [6] and Newell [7]. Both these methods use the fact that there is no interaction amongst the individual customers in the infinite server queue, and so the methods are not applicable to a queue with a finite number of servers except perhaps as a low traffic approximation. For the M/M/1 queue, Greenberg and Greenberg [4] obtain results for the output process by considering the infinitesimal transition probabilities for the queue, and laboriously solving the resulting set of differential equations.

For the M/M/s queue, Pack [8] obtains limited results by deriving a recurrence relation for \( \xi_n \), the time until the nth departure from the queue. Although this relationship is valid for the G/G/s queue, it is difficult to work with, and for the M/M/s queue, only the distribution of \( \xi \) is found. The method presented in this paper provides results for Markovian queues with any number of servers. Specifically, we find the joint probability of N arrivals to, and k departures from a non-steady state Markovian queue in some interval of time (0,t).

The method relies on the fact that the arrival epochs of a Poisson Process are distributed as the order statistics from a sample of independent, uniformly distributed r.v.'s, i.e. given that N arrivals from a Poisson Process have occurred in (0,t), then the arrival epochs \( T_1, T_2, \ldots, T_N \) are distributed as the order statistics from a sample of N independent r.v.'s which are identically uniformly distributed on (0,t) (see Cooper [2]). This fact was used by Mirasol [6] for the M/M/\infty queue and has also been used for the M/G/\infty queue considered as a delayed Poisson Process by Cox and Lewis [3] and Lawrance and Lewis [5].

The method is also applied to a system where two M/M/s queues receive the same input streams. For \( s = \infty \), this is the bivariate delayed Poisson Process of Cox and Lewis [3] and Lawrance and Lewis [5]. Since this system is one where two separate queueing systems receive maximally correlated arrivals, it may be used to provide a bound on the performance of systems with correlated arrivals.

2. OUTPUT OF THE M/M/s QUEUE

In this section we derive a set of simple differential-difference equations relating the input and output streams of an M/M/s queue. We give solutions to these equations and, where possible, compare these solutions with known results. For ease of presentation, we consider a queue which is empty at time zero.

For a queue non-empty at time zero, more complicated algebraic expressions result but, as is pointed out at the end of this section, the method is easily extended to this case.

Let \( P(k|N,t;\lambda) = \Pr\{k \text{ departures from an M/M/s queue in } (0,t) \mid N \text{ arrivals to the queue in } (0,t) \text{ and the system is in state } 0 \text{ at time } 0\} \)

These probabilities are conditioned on the arrival of N customers to the queue in (0,t).
If we denote the arrival epochs of these \( N \) customers by \( t_1, t_2, \ldots, t_N \) where \( 0 = t_1 < \ldots < t_N < t \), we can condition our calculations on the occurrence of a specific set of arrival times \( t_1 = T_1, t_2 = T_2, \ldots, t_N = T_N \). Since these times are (assumed) known, the probability of \( k \) departures, given that arrivals have occurred at these instants, can be calculated. From the properties of the Poisson process, these arrival epochs are distributed as the order statistics from a sample of \( N \) independent random variables, each of which is uniformly distributed on \( (0, t) \). Hence we have a p.d.f. for these arrival times, and can therefore remove the conditioning on \( T_1, \ldots, T_N \) to give the required result.

This is effectively the approach we use, although we go about it a little differently in order to simplify the mathematics. Hence in calculating \( P(k \mid n, t; s) \) we condition only on the arrival time \( T_N \) of the \( N \)th customer and therefore obtain a recursive formula for \( P(k \mid n, t; s) \). This time \( T_N \) is distributed as the \( N \)th order statistic from a sample of \( N \) independent identically distributed random variables uniformly distributed on \( (0, t) \) and therefore has a p.d.f. given by

\[
f(T_N) = N \frac{N-1}{N} / t^N \tag{1}
\]

Conditioning also on the state of the system at the arrival point \( T_N \) gives

\[
P(k \mid N, t; s) = \sum_{j=0}^{k} \binom{k}{j} \int_0^t P(j \mid N-1, T_N; s) f(T_N) dT_N
\]

where \( P(j \mid n, t; s) = \Pr \{ j \text{ departures in } (0, t) \text{ from an } M/M/s \text{ queue with } n \text{ customers present at time } 0 \mid \text{ no arrivals in } (0, t) \} \) and \( P(N-1, t; s) = 0 \).

Equation (2) relies on two facts.

1. Given \( T_N \) the times \( T_1, T_2, \ldots, T_{N-1} \) are distributed as order statistics from a sample of \( N-1 \) i.i.d. r.v.'s uniformly distributed on \( (0, T_N) \) and

2. The memoryless property of the exponential distribution, which implies that only the number of customers in the system at time \( T_N \) is important, and not the expended service times of those customers in service.

Let \( R(k \mid N, t; s) = \binom{N}{N} P(k \mid N, t; s) \) \tag{3}

This is simply a transformation to simplify the solution of the equations. Substituting (1) and (3) into (2) then gives

\[
R(k \mid N, t; s) = \sum_{j=0}^{k} \binom{k}{j} \int_0^t P(j \mid N-1, T_N; s) f(T_N) dT_N
\]

In order to facilitate the solution of (4) we differentiate with respect to \( t \) by the formula

\[
(3/\partial t) \int_0^t f(t, t) dT = f(t, t) + \int_0^t (3/\partial t) f(t, T) dT
\]

to obtain

\[
(3/\partial t) R(k \mid N, t; s) = \begin{cases} \frac{k}{k!} R(j \mid N-1, t; s) p(k-j \mid N-j, 0; s) & j=0 \\ \sum_{j=0}^{k-1} \int_0^t R(j \mid N-1, t; s) (3/\partial t) p(k-j \mid N-j, t-T; s) dT & j \geq 1 \end{cases}
\]

Now

\[
P(k \mid n, 0; s) = \begin{cases} 1 & k=0 \\ 0 & \text{otherwise} \end{cases}
\]

So

\[
(3/\partial t) R(k \mid N, t; s) = R(k \mid N-1, t; s)
\]

In appendix 1 the formulae for \( P(k \mid n, t; s) \) are derived, and it is shown that they obey the relationship

\[
(3/\partial t) P(k \mid n, t; s) + \min(s, n-k) P(k \mid n, t; s) = \min(s, n-k+1) P(k-1 \mid n, t; s)
\]

where \( P(-1 \mid n, t; s) = 0 \).

Substituting in (7) gives

\[
(3/\partial t) R(k \mid N, t; s) + \min(s, N-k) R(k \mid N, t; s) = R(k \mid N-1, t; s) + \min(s, N-k+1) R(k-1 \mid N, t; s)
\]

This is a differential-difference equation relating the variables of interest in the system, and its solutions contain the results we require.

**PARTICULAR CASES**

(i) \( M/M/\infty \) Queue

Here (9) becomes

\[
(3/\partial t) R(k \mid N, t; \infty) + \min(s, N-k) R(k \mid N, t; \infty) = R(k \mid N-1, t; \infty) + \min(s, N-k+1) R(k-1 \mid N, t; \infty)
\]

We try a solution of the form

\[
R(k \mid N, t; \infty) = \frac{a(t)^k}{k!} b(t)^{N-k}
\]

Substitution of (11) in (10) shows that we require that \( a(t) \) and \( b(t) \) satisfy

\[
a'(t) = u b(t) \quad (12a)
\]

\[
b'(t) + ub(t) = 1 \quad (12b)
\]

Solving with initial conditions \( a(0) = b(0) = 0 \) gives the solutions to (10) as

\[
R(k \mid N, t; \infty) = \frac{1}{k! (N-k)!} \left[ \frac{\mu t - (1-e^{-\mu t})}{\mu} \right] \left[ 1 - e^{-\mu t} \right]^{N-k}
\]

so

\[
P(k \mid N, t; \infty) = \frac{N!}{N^N} R(k \mid N, t; \infty)
\]
\[
E[k] = \frac{N}{k} \left\{ \frac{1 - (1 - e^{-\mu t})^k}{\mu t} \right\} \left\{ \frac{1 - (1 - e^{-\mu t})^{N-k}}{\mu t} \right\} N-k
\]  

which agrees with the previously known results.

(ii) M/M/1 Queue

It can be verified by substitution that the solution is given by

\[
P(k_{\text{IN}}, t) = e^{-\mu t} \left( \frac{1}{N!} \right) \sum_{j=0}^{\infty} \frac{\mu^j}{j!} \left( \frac{1 - e^{-\mu t}}{\mu t} \right)^j
\]

where

\[
\gamma(N, \mu t) = e^{-\mu t} \sum_{j=0}^{\infty} \frac{\mu^j}{j!}
\]

This solution agrees with the results from [4].

GENERAL SOLUTION

The general solution of (9) in the time domain presents some difficulties, so we approach the problem through the frequency domain.

Let

\[
R(k_{\text{IN}}, z; s) = \int_0^\infty e^{-zt} R(k_{\text{IN}}, t; s) dt
\]

Transforming equation (9) in this manner gives

\[
[z+\min(s, N-k)] R(k_{\text{IN}}, z; s) = R(k_{\text{IN}}-1, z; s) + [z+\min(s, N-k+1)] R(k_{\text{IN}}-1, z; s)
\]

(18) can be solved recursively starting with the initial condition

\[
R(0, 0, z; s) = \frac{1}{z+\min(s, 0)}
\]

corresponding to

\[
P(0, 0, t; s) = 1
\]

to give

\[
R(k_{\text{IN}}, z; s) = \frac{1}{z+\min(s, j)}
\]

for \( j \geq 0 \)

(19)

The inversion of these transforms presents some difficulties due to the multiple roots in the denominator of the transform. However, the transforms can be inverted relatively easily for the cases \( k=0 \) and \( k=1 \). Details of this inversion are given in appendix 2. Also, in equation (9), we note that for \( N<s \), the equation is the same as for the case \( s=\infty \).

For notational convenience, we suppress the \( s_1 \) and \( s_2 \) variables. Due to the independence of the service operations in the two queues, we can write

\[
P(k_1, k_2 | n_1, n_2, t) = p(k_1 | n_1, t) \cdot p(k_2 | n_2, t)
\]

Also, write \( R(k_1, k_2 | n_1, n_2, t) = (N!/t^N) p(k_1, k_2 | N, t) \)

Substituting in (20) and differentiating with respect to \( t \) gives

\[
(3/3t)R(k_1, k_2 | N, t)
\]

(21)

Finally, we note that for \( k > N \) or \( k < N \),

\[
P_m(k_{\text{IN}}, t; s) = \text{Pr} \{ k \text{ departures from an M/M/s queue in (0, t) } | N \text{ arrivals to the queue in (0, t) and the system is in state m at time 0} \}
\]

which can be solved as before although, as pointed out earlier, algebraically complex expressions are obtained. In this case, note that for \( k > N \) (which corresponds to the case when none of the arrivals during (0, t) may begin service) we also have

\[
P_m(k_{\text{IN}}, t; s) = p(k | m, t; s)
\]

3. PARALLEL QUEUES

In this section we consider two queueing systems in parallel, where each Poisson arrival to the system consists of two demands, one for each queue. The queues have \( s_1 \) and \( s_2 \) servers and have negative exponentially distributed service times with mean values \( 1/\mu_1 \) and \( 1/\mu_2 \) respectively. This can be considered as a system of two Markovian queues which receive input streams which are maximally correlated, and as such may be used to derive a bound on the performance of systems with correlated arrivals.

Let

\[
P(k_1, k_2 | N_{\text{IN}}, t; s_1, s_2) = \text{Pr} \{ k_1 \text{ departures from queue } #1 \text{ and } k_2 \text{ departures from queue } #2 \text{ in (0, t) } | N \text{ arrivals to the system in (0, t) and both queues in state } 0 \text{ at time 0} \}
\]

Conditioning on the state of the system at the time \( T_N \) of the Nth Poisson arrival gives

\[
P(k_1, k_2 | N_{\text{IN}}, t; s_1, s_2) = \sum_{i_1=0}^{N-1} \sum_{i_2=0}^{N-1} p(i_1, i_2 | N-1, T_N; s_1, s_2) f(T_N) \frac{dT_N}{N}
\]

where \( p(i_1, i_2 | n_1, n_2, t) = 0 \) for \( k_1 > N-1 \) or \( k_2 > N-1 \).

Then analogously to eqn (2) we have

\[
P_m(k_{\text{IN}}, t; s) = \frac{k_1}{k} \sum_{j=0}^{k_1} p(k-j | N-1, T_N; s) f(T_N) \frac{dT_N}{N}
\]

(22)
\[ R(k_1, k_2 | N-1, t) \]
\[
+ \sum_{i=0}^{N-1} \sum_{l=0}^{i} R(i, l | N-1, t) \]
\[
\times \frac{\partial}{\partial t} \left[ p_1(i, l | n_1, t) + p_2(l, n_2, t) \right] \]
\[ \text{(21)} \]

Now \( \frac{\partial}{\partial t} p(k_1, k_2 | n_1, n_2, t) \)
\[
= \frac{\partial}{\partial t} \left[ p_1(k_1 | n_1, t) \right] + \frac{\partial}{\partial t} \left[ p_2(k_2 | n_2, t) \right] \]
\[ \text{(21a)} \]

Using the results from appendix 1 gives
\[ \frac{\partial}{\partial t} p(k_1, k_2 | n_1, n_2, t) \]
\[= \sum_{i=0}^{n_1} \min(s_1, n_1-i+1) p(i, l_1, l_2 | n_1, n_2, t) \]
\[+ \sum_{j=0}^{n_2} \min(s_2, n_2-j+1) p(l_1, j_2 | n_1, n_2, t) \]
\[\text{(22)}\]

Substituting in (21) gives
\[ \frac{\partial}{\partial t} R(k_1, k_2 | N-1, t) \]
\[= \left[ \sum_{i=0}^{n_1} \min(s_1, n_1-i+1) \right] R(k_1, k_2 | N, t) \]
\[= \left[ \sum_{j=0}^{n_2} \min(s_2, n_2-j+1) \right] R(k_1, k_2 | N, t) \]
\[\text{(23)}\]

This can easily be generalized to \( K \) queues in parallel in the form
\[ \frac{\partial}{\partial t} R(k_1, \ldots, k_K | N, t) \]
\[= \left( K \sum_{i=1}^{\min(s_i, N-k_i)} \right) R(k_1, \ldots, k_K | N, t) \]
\[\text{(24)}\]

We convert equation (23) to the frequency domain as before, giving
\[ [\sum_{i=0}^{n_1} \min(s_1, n_1-i+1) + \sum_{j=0}^{n_2} \min(s_2, n_2-j+1)] R(k_1, k_2 | N, z) \]
\[= R(k_1, k_2 | N-1, z) \]
\[+ \sum_{i=0}^{n_1} \min(s_1, n_1-i+1) R(k_1, k_2 | N-1, z) \]
\[+ \sum_{j=0}^{n_2} \min(s_2, n_2-j+1) R(k_1, k_2 | N-1, z) \]
\[\text{(25)}\]

This has the recursive solution
\[ R(k_1, k_2 | N, z) \]
\[= \left[ \sum_{i=0}^{n_1} \min(s_1, n_1-i+1) \right] R(k_1, k_2 | N-1, z) \]
\[= \left[ \sum_{j=0}^{n_2} \min(s_2, n_2-j+1) \right] R(k_1, k_2 | N-1, z) \]
\[\text{(26)}\]

and
\[ \alpha(p) = \begin{cases} 1 & \text{if } m_p^2 - m_{p-1}^2 = 1 \\ 2 & \text{if } m_p^2 - m_{p-1}^2 = 1 \end{cases} \]
\[\beta(p) = \begin{cases} m_p \text{ if } m_p - m_{p-1} = 1 \\ 0 & \text{if } m_p - m_{p-1} = 1 \end{cases} \]

and summation over \( m_n \) is over values such that
\[ m_1 + n_1 = 1 \quad 0 \leq k_1 + k_2 \]
and
\[ m_1 - k_1, n_1 - k_2, n_1 \leq k_1 + k_2 \]

APPENDIX ONE

In Section 1 we required the probabilities
\[ p(t | n, z; s) = \text{Pr} \{ t \text{ departures from an initial } n \text{ customers} \} \]
\[ \text{in } (0, t) \text{ from an } M / M / s \text{ queue | no arrivals at} \]
\[\text{the queue in } (0, t) \} . \]

There are three distinct cases

Case 1 \( n < s \). In this case, all customers are served independently, giving
\[ p(t | n, z; s) = \begin{cases} \sum_{l=0}^{n} \left( 1 - e^{-sL} \right) \left( e^{-sL} \right)^{n-l} & 0 \leq t \leq n \\
0 & \text{otherwise} \end{cases} \]

Case 2 \( n = s + 1 \) and \( 0 \leq t \leq n-s \). In this case, the system is always completely occupied, so the departures form a Poisson process at rate \( s \), and
\[ p(t | n, z; s) = \begin{cases} \sum_{i=0}^{n} \left( s \right)^i \left( e^{-sL} \right)^{n-i} & 0 \leq t \leq n-s \\
0 & \text{otherwise} \end{cases} \]

Case 3 \( n > s + 1 \) and \( n > s + 1 \). Here the first \( n-s+1 \) departures occur according to a Poisson process at rate \( s \), and following the \( (n-s+1) \)th departure, the remaining customers act independently.

This gives
\[ p(t | n, z; s) = \int_0^{n-s+1} \left( s \right)^{n-s} e^{-sL} (t-T)^{n-s-1} (-1)^{t-n-s-1} \]
\[\text{dt} \]
then
\[ p(t | n, z; s) = \int_0^{n-s+1} \left( s \right)^{n-s} e^{-sL} (t-T)^{n-s-1} (-1)^{t-n-s-1} \]
\[\text{dt} \]
using the identity
\[ E \left( \frac{\left( -1 \right)^j \left( -s \right)^j}{j!} \sum_{j=0}^{n-s+1} \int_0^{n-s+1} z^{j} e^{-sz} \right) \]
\[\text{dt} \]
and
\[ E \left( \frac{\left( -1 \right)^j \left( -s \right)^j}{j!} \sum_{j=0}^{n-s+1} \int_0^{n-s+1} z^{j} e^{-sz} \right) \]
\[\text{dt} \]
Combining all three cases we have

\[ p(t|n,z;s) = \frac{(s-1)! (n-R.) n}{(n-k)! (z+su)_{n-k} j=n-k} \]

this can be written as

\[ p(t|n,z;s) = \begin{cases} \frac{n!^{|\min(s,j)|}}{\Pi (z+ju)}  & \text{otherwise} \\ \frac{n!^{|\min(s,j)|}}{(z+su)_{n-k} j=n-k} \end{cases} \]

Reverting to the time domain and noting we have

\[ p(t|n,0;s) = \begin{cases} 1 & \text{if } t=0 \\ 0 & \text{otherwise} \end{cases} \]

For \( k=1 \)

\[ R(1|n,z;s) = N \sum_{j=0}^{s-1} \frac{\mu_{i-1} \mu_{i}^{|\min(s,j)|}}{(z+su)_{n-k} j=n-k} \]

where \( \mu_{i-1} = e^{-\mu t} \sum_{k=0}^{i-1} (\mu t)^k \frac{1}{k!} \)

The unconditional probability of no departures in \((0,t)\) is then

\[ P(O|0,t;s) = \sum_{N=0}^{\infty} P(O|N,t;s) e^{-\lambda t} (\lambda t)^N/N! \]

which can easily be shown to yield the same results as [8], equations (8) and (9).

**APPENDIX TWO**

Pack [8] gives a formula for the marginal probability of no departures during \((0,t)\). Here we find similar formulae for the probability of no departures in \((0,t)\) given \( N \) arrivals. From eqn. (19) we have

\[ p(t|n,z;s) = \frac{(s-1)! (n-R.) n}{(n-k)! (z+su)_{n-k} j=n-k} \]

Inverting for \( N \)

\[ R(0|N,z;s) = \frac{1}{N!} \sum_{j=0}^{N} (-1)^j \binom{N}{j} (z+ju)_{N-j} \]

so \( P(0|N,z;s) = \left(1-e^{-\mu t}\right)^N \) for \( N \)

For \( N=s-1 \)

\[ R(0|s-1,z;s) = \frac{1}{(s-1)! (z+su)_{s-1}} \sum_{j=0}^{s-1} (-1)^j \binom{s-1}{j} (z+ju)_{s-1} \]

Taking partial fractions and inverting gives

\[ P(0|N,t;s) = \sum_{N=0}^{\infty} P(0|N,t;s) e^{-\lambda t} (\lambda t)^N/N! \]

We also note the identity

\[ N \sum_{j=0}^{(s-1)} (-1)^j \binom{N}{j} (z+ju)_{N-j} = \frac{(z+su)_{N-s}}{(s-1)! (z+su)_{s-1}} \]
Take partial fractions and invert, to give

\[
P(1|N,t;s) = \frac{N!}{(\mu t)^N(s-1)!} \left[ \sum_{j=0}^{s-1} \frac{(-1)^j{s-1 \choose j} \mu^j e^{-\mu t} \gamma(K,(s-j)\mu t)}{(s-j)^{K+1}} + \sum_{j=0}^{s(N-s)} \frac{(-1)^j{s-1 \choose j} \mu^j e^{-\mu t} \gamma(K,(s-j)\mu t)}{(s+j+1)^{K+1}} \right]
\]

where

\[
\beta_g(K,\mu t) = \sum_{j=0}^{s-1} (-1)^j{s-1 \choose j} e^{-\mu t} \frac{\gamma(K,(s-j)\mu t)}{(s-j)^{K+1}}
\]

and

\[
\beta_g^*(N-s,\mu t) = \sum_{j=0}^{s(N-s)} (-1)^j{s-1 \choose j} e^{-\mu t} \frac{\gamma(K,(s-j)\mu t)}{(s+j+1)^{K+1}}
\]

REFERENCES


