TWO QUEUES IN SERIES WITH NON-OVERLAPPING SERVICE TIMES

O. J. BOXMA
I.B.M. Thomas J. Watson Research Center, Yorktown Heights, New York, U. S. A.

ABSTRACT

This paper considers a queuing system consisting of two single server queues in series, in which the service time of an arbitrary customer at the second queue is always longer than any service time at the first queue. Customers arrive at the first queue according to a Poisson process.

Some performance measures of this model will be studied. The results include expressions for the joint stationary distribution of the actual waiting times at both queues and their covariance, for the total actual waiting time distribution in the system and for the sojourn time distribution in the second queue. These results enable us to make a comparison with the behaviour of the model in which the order of the queues is reversed.

1. INTRODUCTION

We consider a queuing system consisting of two single server queues in series, with Poisson arrival process at the first queue and infinite waiting room at both queues. The service time distributions at the first queue $Q_1$ and second queue $Q_2$ have non-overlapping supports $(0,b)$ and $(b,\infty)$. Hence the service time of an arbitrary customer in $Q_2$ is at least as long as any service time in $Q_1$.

One can think of many practical examples of such a model; consider e.g. a service system in which the required amounts of service at both stages are lying between two fixed positive levels, while the first server is much faster than the second one.

The special case in which all service times in $Q_2$ are equal to a constant $b$, while all service times in $Q_1$ are at least equal to $b$, has been analysed in [1] and [5]. Friedman [5] has shown that the time spent in this system by each customer is independent of the order of the stages, and hence is equal to the time spent in an M/1 queue with arrival process that of $Q_1$ and service time distribution that of $Q_2$. A fundamental extension of part of Friedman's work is due to Tembe and Wolff [8]. They remark that in general the order of the queues does matter, proving that the total time spent by each customer in a system, in which each service time in $Q_2$ is at least as long as any service time in $Q_1$, is at least as long as the total time spent by each customer in the system with order of queues reversed (with equality in Friedman's constant service time case). Their proof is based on sample function relations between departure times, times in system, etc.; they do not derive expressions for the distributions of these variables.

Both Friedman and Tembe and Wolff admit a general arrival process. In the present study we restrict ourselves to the case of Tembe and Wolff's model with a Poisson arrival process. For this case we derive expressions for (a.o.) the Laplace-Stieltjes transforms of the joint distribution of the waiting times at both queues and of the distribution of the total waiting time in the system. These results can be compared with those for the case of reversed order of queues; this comparison may lead to a better insight into the influence of the order of queues on the behaviour of a tandem queuing system.

Before presenting an outline of the structure of the paper we now give a more detailed description of the model.

Consider a tandem queuing system consisting of two single server queues $Q_1$ and $Q_2$ in series, both with infinite waiting room. Customers enter the tandem system individually at $Q_1$. After completion of his service at $Q_1$ a customer enters immediately $Q_2$, and when his service at $Q_2$ is completed he leaves the tandem system. Customers are served individually, and at both counters the service discipline is the well-known first come - first served discipline.

Let $t_1$ denote the moment of arrival of the $n$-th customer $C_n$ at the system, with $t_1 = 0$; it will always be assumed that $C_1$ meets an empty system, i.e. neither at $Q_1$ nor at $Q_2$ customers are present at $t_1$. The interarrival times

$$
\gamma_n = t_n - t_{n-1}
$$

are assumed to be independent identically distributed (i.i.d.) stochastic variables (s.v.), with negative exponential distribution

$$
\Pr\{\gamma_n < t\} = 1 - e^{-\alpha t}, \quad t > 0,
$$

$$
\Pr\{\gamma_n \geq t\} = e^{-\alpha t}, \quad t \leq 0,
$$

so that the arrival process is a homogeneous Poisson process with rate $\alpha$.

Let $\tau_n$ be the service time of $C_n$ at $Q_1$, $i = 1,2$; $n = 1,2,\ldots$. The processes $\tau_{n-1} = (\tau_{n-1},n = 1,2,\ldots)$, $\tau_n = (\tau_{n-1},n = 1,2,\ldots)$ and $(\tau_n,n = 1,2,\ldots)$ are independent. The s.v. $\tau_n$, and also $\tau_{n-1} = (\tau_{n-1},n = 1,2,\ldots)$, $\tau_n = (\tau_{n-1},n = 1,2,\ldots)$ and $(\tau_n,n = 1,2,\ldots)$ have identical distributions.

The process $\tau_n$ is the well-known first come - first served discipline. The processes $\tau_{n-1} = (\tau_{n-1},n = 1,2,\ldots)$, $\tau_n = (\tau_{n-1},n = 1,2,\ldots)$ and $(\tau_n,n = 1,2,\ldots)$ are independent. The s.v. $\tau_n$, and also $\tau_{n-1} = (\tau_{n-1},n = 1,2,\ldots)$, $\tau_n = (\tau_{n-1},n = 1,2,\ldots)$ and $(\tau_n,n = 1,2,\ldots)$ have identical distributions.

Let $\tau_n$ be the service time of $C_n$ at $Q_1$, $i = 1,2$; $n = 1,2,\ldots$. The processes $\tau_{n-1} = (\tau_{n-1},n = 1,2,\ldots)$, $\tau_n = (\tau_{n-1},n = 1,2,\ldots)$ and $(\tau_n,n = 1,2,\ldots)$ are independent. The s.v. $\tau_n$, and also $\tau_{n-1} = (\tau_{n-1},n = 1,2,\ldots)$, $\tau_n = (\tau_{n-1},n = 1,2,\ldots)$ and $(\tau_n,n = 1,2,\ldots)$ have identical distributions.

The processes $\tau_{n-1} = (\tau_{n-1},n = 1,2,\ldots)$, $\tau_n = (\tau_{n-1},n = 1,2,\ldots)$ and $(\tau_n,n = 1,2,\ldots)$ have identical distributions.

We define the moments

$$
\beta_{(i)} = \mathbb{E}\tau_i \quad i = 1,2,\ldots.
$$

The processes $\tau_{n-1} = (\tau_{n-1},n = 1,2,\ldots)$, $\tau_n = (\tau_{n-1},n = 1,2,\ldots)$ and $(\tau_n,n = 1,2,\ldots)$ have identical distributions.

Let $\tau_n$ be the service time of $C_n$ at $Q_1$, $i = 1,2$; $n = 1,2,\ldots$. The processes $\tau_{n-1} = (\tau_{n-1},n = 1,2,\ldots)$, $\tau_n = (\tau_{n-1},n = 1,2,\ldots)$ and $(\tau_n,n = 1,2,\ldots)$ are independent. The s.v. $\tau_n$, and also $\tau_{n-1} = (\tau_{n-1},n = 1,2,\ldots)$, $\tau_n = (\tau_{n-1},n = 1,2,\ldots)$ and $(\tau_n,n = 1,2,\ldots)$ have identical distributions.

The processes $\tau_{n-1} = (\tau_{n-1},n = 1,2,\ldots)$, $\tau_n = (\tau_{n-1},n = 1,2,\ldots)$ and $(\tau_n,n = 1,2,\ldots)$ have identical distributions.

We define the moments

$$
\beta_{(i)} = \mathbb{E}\tau_i \quad i = 1,2,\ldots.
$$

The processes $\tau_{n-1} = (\tau_{n-1},n = 1,2,\ldots)$, $\tau_n = (\tau_{n-1},n = 1,2,\ldots)$ and $(\tau_n,n = 1,2,\ldots)$ have identical distributions.
We are now ready to proceed to the analysis of the model described above. The organization of the paper is as follows.

In Section 2 we study the joint distribution of the queue length \( L_n \) at \( Q_1 \) immediately after the epoch that \( C_n \) leaves \( Q_1 \), and the sojourn time \( s(2) \) of \( C_n \) at \( Q_2 \). From the resulting expression we derive in Section 3, for the stationary situation, the sojourn time distribution at \( Q_2 \), and the joint distribution of the waiting times in \( Q_1 \) and \( Q_2 \). This last expression leads to the distribution of the total waiting time in the system, and to an expression for the covariance of the waiting times in \( Q_1 \) and \( Q_2 \). The results are compared with those for the case of reversed order of queues.

An important role in the analysis of our model is played by the sum of the differences of the service times in \( Q_2 \) and \( Q_1 \) during a busy period of \( Q_1 \). This sum is studied in an appendix.

2. THE BASIC EQUATION

For the tandem model described in Section 1 we consider the two-dimensional imbedded Markov chain

\[(z_n^{(2)}, n = 1, 2, \ldots)\]. We can write for \( n = 1, 2, \ldots \):

\[ z_{n+1} = \begin{cases} z_{n-1}^+ + \gamma_{n+1}^+ & \text{if } z_n > 0, \\ z_{n-1}^+ - \gamma_{n+1}^+ + \tau_{n+1}^+ & \text{if } z_n = 0, \\ z_{n+1}^+ & \text{if } z_n < 0. \end{cases} \]  

(2.1)

It follows from (1.1), (2.1) and (2.2) for \( n = 1, 2, \ldots, |x| \leq 1, w > 0 \):

\[ K_{n+1}(r, w) = E\{ z_{n+1}^+ \} \]

\[ = E\{ z_{n-1}^+ + \gamma_{n+1}^+ [z_{n}^+(1) - \gamma_{n+1}^+ z_{n-1}^+ w, z_n > 0] + \] \[ + E\{ \gamma_{n+1}^+ [z_{n}^+(1) - \gamma_{n+1}^+ z_{n-1}^+ w, z_n = 0] \} + \] \[ + E\{ z_{n+1}^+ [z_{n}^+(1) - \gamma_{n+1}^+ z_{n-1}^+ w, z_n < 0] \} \]  

(2.2)

\[ = E\{ z_{n-1}^+ + \gamma_{n+1}^+ [z_{n}^+(1) - \gamma_{n+1}^+ z_{n-1}^+ w, z_n > 0] + \] \[ + E\{ \gamma_{n+1}^+ [z_{n}^+(1) - \gamma_{n+1}^+ z_{n-1}^+ w, z_n = 0] \} + \] \[ + E\{ z_{n+1}^+ [z_{n}^+(1) - \gamma_{n+1}^+ z_{n-1}^+ w, z_n < 0] \} \]  

(2.3)

Note that \( z_n^+ \) and \( z_{n-1}^+ \) are both independent of \( \gamma_{n+1}^+ \), \( \tau_{n+1}^+ \), and \( \epsilon_{n+1}^+ \), while \( \gamma_{n+1}^+ \) is independent of \( z_n^+ \) and \( \gamma_{n+1}^+ \). Furthermore, \( \epsilon_{n+1}^+ \) (when positive) is negative exponentially distributed with mean \( \alpha \), no matter what are the values of \( z_n^+ \) and \( z_{n-1}^+ \). Hence we can rewrite the two terms in the righthand side of (2.3) in the following way, using (1.1) (\( n = 1, 2, \ldots, |x| \leq 1, w > 0 \)):

\[ \frac{1}{x} E\{ z_{n-1}^+ \} \]

\[ = \frac{1}{x} E\{ z_{n-1}^+ \} \]

(2.4)

We now take Laplace-Stieltjes transforms (LST) in (2.6). Differentiation of the two terms in the righthand side of (2.6) with respect to \( w \) leads to four different terms, since \( w \) appears both in integral boundary and integrand. After some arithmetic we obtain for \( |p| < 1, |r| \leq 1, w > 0 \) (see (1.3)):

\[ K(p, r, w) = p E\{ z_1^+ \} \]

\[ = \frac{1}{x} E\{ z_{n-1}^+ \} \]

(2.5)

Now we take Laplace-Stieltjes transforms (LST) in (2.6). Differentiation of the two terms in the righthand side of (2.6) with respect to \( w \) leads to four different terms, since \( w \) appears both in integral boundary and integrand. After some arithmetic we obtain for \( |p| < 1, |r| \leq 1, w > 0 \) (see (1.3)):
Now we use the fact that (see Section 1) \( f(x) \), has the finite support \((0,b)\), while \( g(x) \) has the support \([b,\infty)\), thus causing \( K(p,r,z) \) and \( H(p,z) \) to be zero for \( z \leq b \).

Hence the first term in the righthand side of (2.7) disappears; the last integral in the second term can be replaced by \((k(p,r,s)-h(p,s))\); the third term becomes

\[
p \beta^{(2)}(s) \frac{\beta^{(1)}(-r/a) h(p,1/a)}{e^{-s} \int e^{-sw} d_H(p,sw) dw}.
\]

The fourth term can be rewritten in the following way,

\[
\frac{b}{a} \int e^{-(1-r)/a-s} t d_B(1)(t) = e^{-s} \int e^{-sw} d_H(p,sw) dw
\]

and using the transform \( x = \frac{t}{a} \), \( v = \frac{x}{s} \):

\[
\frac{b}{a} \int e^{-(1-r)/a-s} t d_B(1)(t) = \frac{a}{b} \int e^{-(1-r)/a-s} t d_B(1)(1/a).\]

Note that \( e^{-(1-r)/a-s} t d_B(1)(1/a) \) is finite for all values of \( v \).

Now \( e^{-(1-r)/a-s} t d_B(1)(1/a) \) is defined for all values of \( v \).

In Section 1 we have already mentioned the fact that a special case of the present model (viz. \( \beta_{1,2}^{(1)} = \beta \), \( \beta_{1,2}^{(2)} = \beta + a_n \in [\beta, \infty) \), \( n = 1,2, \ldots \), the s.v. \( a_n \) being i.i.d.) has been analysed in [1]. We can now proceed as in [1] to determine \( h(p,s) \) from the fact that \( k(p,r,s) \) is an analytic function of \( r \) for \( |r| \leq 1 \), \( p \) and \( s \) being fixed, \( |p| < 1 \), \( Re s \geq 0 \), whereas the denominator of the right-hand side of (2.10) has exactly one zero in the region \( |r| < 1 \) (see also the appendix), and from the fact that \( h(p,s) \) is an analytic function of \( s \) for \( Re s \geq 0 \). However, the calculations become rather lengthy, and the results are quite complicated. We therefore restrict ourselves in the following to the consideration of the stationary situation.

3. STATIONARY RESULTS

Let \( a_i \) be \( \frac{\beta^{(1)}}{a}, i = 1,2 \). We assume in the following that \( a_i < 1 \); hence also \( a_n < 1 \). Proceeding as in [1] it is not hard to prove that the stationary distributions of all relevant variables of the model (like waiting times, sojourn times and queue lengths) exist if \( a_i < 1 \).

See also Sacks [6] for a discussion of equilibrium conditions for queues in series.

Notation

The following convention is made: if \( X_n \) is i.i.d., \( \tau \) converges in distribution for \( n \rightarrow \infty \) to a distribution \( X(\tau) \) then by \( X_n \) we denote a s.v. having the distribution \( X(\tau) \), and by \( X(\tau) \) we denote the LST of \( X(\tau) \). The same holds for \( X_n \) if \( \tau = \tau_n \).

Introducing for \( |r| \leq 1, Re s \geq 0 \)

\[
\beta^{(1)}((1-r)/a-s) h(p,1/a) = \beta^{(1)}((1-r)/a) h(p,1/a),
\]

\( |p| < 1, |r| \leq 1, Re s \geq 0 \),

(2.8) finally becomes:

\[
k(p,r,s) = p \beta^{(1)}((1-r)/a-s) \beta^{(2)}(s) \]

\[
= \frac{b}{a} \frac{\beta^{(1)}((1-r)/a-s) \beta^{(2)}(s) [k(p,r,s) - h(p,s)] +}{e^{-s} \int e^{-sw} d_H(p,sw) dw}
\]

\[
= \frac{b}{a} \frac{\beta^{(1)}((1-r)/a-s) \beta^{(2)}(s) h(s) -}{e^{-s} \int e^{-sw} d_H(p,sw) dw}
\]

\[
= \frac{a}{b} \frac{[\beta^{(1)}((1-r)/a-s) h(p,1/a)] -}{e^{-s} \int e^{-sw} d_H(p,sw) dw}
\]

\[
\beta^{(1)}((1-r)/a-s) h(p,1/a) \]

\( |p| < 1, |r| \leq 1, Re s \geq 0 \),

(2.9) and hence for \( |p| < 1, |r| \leq 1, Re s \geq 0 \) we obtain the basic equation

\[
k(p,r,s) = [r - p \beta^{(1)}((1-r)/a-s) \beta^{(2)}(s)]^{-1} p \beta^{(2)}(s).
\]

\[
r \beta^{(1)}((1-r)/a-s) + \left[ \frac{1}{1-a} - 1 \right] \beta^{(1)}((1-r)/a-s) h(p,s) -
\]

\[
\frac{1}{1-a} \beta^{(1)}((1-r)/a-s) h(p,1/a) \]

\( |p| < 1, |r| \leq 1, Re s \geq 0 \),

(2.10) Note that \( \beta^{(1)}(v) \) is defined for all values of \( v \).
\[
\lim_{s \to 0} h(s) = \Pr\{\eta_1^{(1)} > s\},
\]

Letting \(s\) tend to zero from above in both sides of (3.4) then leads to \(h(1/a) = (1 - a^2)/a(1 - a/s)\), and hence for \(Re s > 0\)

\[
h(s) = (1 - a^2)/a(1 - a/s).
\]

Substitution of (3.5) in (3.2) gives

\[
k(r,s) = \frac{\beta(s)}{1 - \beta(1)(-r/a) - s} \frac{\beta(2)(s)}{1 - \beta(1)(-r/a) - s}.
\]

We are now able to determine the LST of the sojourn time at \(Q_2\):

\[
k(1,s) = \frac{\beta(2)(s)}{1 - \beta(1)(-r/a) - s} \frac{\beta(2)(s)}{1 - \beta(1)(-r/a) - s}.
\]

Remark 3.1

Substitution of \(\beta(s) = e^{-as}\) and \(\beta(2)(s) = e^{-as}\) with \(y(s) = E\{e^{-as}\}\) of the model with \(T(1) \equiv s\), and \(T(2) \equiv s \geq \beta\) leads to the expressions for \(k(r,s)\) and \(h(s)\) that were obtained for that model in [1].

Remark 3.2

One can prove that the expression \((1 - a) / (1 - a)\) has the following interpretation: it is the LST of the limiting distribution of the s.v. \(\eta_j\), with \(\eta_j\) the amount of work in \(Q_2\) immediately before the arrival in \(Q_2\) of the first customer of the \(j\)th busy period of \(Q_1\).

We now turn to the consideration of the joint limiting distribution of \(\eta_1^{(1)} + \eta_2^{(2)}\), the actual waiting times of \(C_n\) in the system.

**Theorem 3.1**

For \(a_2 < 1\),

\[
E\{e^{-as} \beta(1)(-s) \beta(2)(s)\} = \frac{1}{1 - \beta(1)(-a/s) \beta(2)(s)} \frac{as(1 - a_2)}{1 - as},
\]

| \(\mid r \mid \leq 1, Re s > 0\). |

**Proof (sketch)**

In pg. 69 of [1] we have shown that for a general tandem queueing model (the model described in Section 1 of the present study, but without any assumption on the supports of the service time distributions):

\[
E\{\exp\{\frac{-r\eta_1^{(1)}}{a} - \eta_1^{(1)}\} \Pr\{\eta_1^{(1)} \leq s\}\}
\]

which leads to (cf. [1] pg. 71):

\[
E\{\exp\{-r\eta_1^{(1)}\} \Pr\{\eta_2^{(2)} < \eta + y\}\}
\]

Taking the LST in (3.9) yields (3.8) after lengthy calculations, which are somewhat similar to those in (2.7) and (2.8) and will be omitted.

Substitution of \(s = 0\) and \(\beta(1)(-r/a) = \rho\) in (3.8) leads to the Pollaczek-Khinchine formula for an M/G/1 queue (cf. Cohen [4]):

\[
E\{e^{-as} \beta(1)(-s) \beta(2)(s)\} = \frac{1}{\alpha a - 1 + \beta(1)(s)}
\]

Similarly (3.8) leads - after appropriate substitutions for \(r\) and with the help of (3.5) and (3.6) - to the following results (\(\eta_n\) is a s.v. with the limiting distribution of \(\eta_1^{(1)} + \eta_2^{(2)}\), the total waiting time of \(C_n\) in the system).

**Theorem 3.2**

For \(a_2 < 1\),

\[
E\{e^{-as} \beta(1)(-as) \beta(2)(s)\} = \frac{1}{1 - \beta(1)(-a/s) \beta(2)(s)} \frac{as(1 - a_2)}{1 - as},
\]

\[
\mid \beta(1)(-a/s) \beta(2)(s) \mid, \mid r \mid \leq 1, Re s > 0\). |

**Proof (sketch)**

In the case \(\eta_1^{(1)} \equiv s, \eta_2^{(2)} \equiv s + a_2 \geq s\), the term between square brackets in the right-hand side of (3.12) is equal to one and the right-hand side of (3.12) is equal to the LST of the limiting distribution of the waiting times.
in an M/G/1 system with arrival intensity $\lambda^{-1}$ and service time distribution $B(\cdot)$. This result is implicitly contained in Friedman [5].

COROLLARY 3.1

For $\lambda < 1$, (with $'$ denoting a derivative)

$$E\{w(1)^2\} = \frac{\beta_2(2)}{2\alpha(1 - \lambda)} - \frac{\beta_3(1)}{2\alpha(1 - \lambda^2)},$$

$$E\{w(2)^2\} = \frac{\beta_2(2)}{2\alpha(1 - \lambda^2)} - \frac{\beta_3(1)}{2\alpha(1 - \lambda)},$$

$$E\{w(1)w(2)\} = \frac{\beta_2(2)}{2\alpha(1 - \lambda^2)} - \frac{\beta_3(1)}{2\alpha(1 - \lambda)},$$

$$\beta_2(2) = \frac{\beta_1(1)}{\alpha} - \frac{\beta_3(1)}{\alpha^2} - \frac{\beta_4(1)}{\alpha^3}.$$

These results follow after straightforward (but sometimes very tedious) calculations from (3.10), (3.11) and (3.12).

We have omitted the lengthy expression for $E\{w(3)^2\}$, but we have used it to calculate

$$E\{w(1)^2\} = \frac{\beta_2(2)}{2\alpha(1 - \lambda)} - \frac{\beta_3(1)}{2\alpha(1 - \lambda^2)}.$$

The expression for the covariance might also have been derived from (3.8).

Remark 3.5

For the case $\lambda(1) \geq \beta_1$, $\lambda(2) = \beta + \varepsilon_n \geq \beta$ (3.17) reduces to the following expression, which has also been discussed in [1] and [2]:

$$\text{cov}(\varepsilon_n(1), \varepsilon_n(2)) = (\beta(1))^2 \frac{\beta(2)}{\alpha} - \frac{\beta(1)}{\alpha^2} - \frac{\beta(2)}{\alpha^3}.$$

In the present more general model $\text{cov}(\varepsilon_n(1), \varepsilon_n(2))$ is again linearly dependent on $\beta(2)$, although it is not directly proportional to $(\beta(2) - \beta(1))$.

Remark 3.6

In Section 1 we have mentioned a result of Tembe and Wolff [8] concerning the optimal order of queues in a tandem system; this result implies that in the present model the total time spent in the system is stochastically larger than the total time spent in the system with reversed order of queues. Obviously a similar result holds for the total waiting time (= total time in system minus service times). The total waiting time in the "reversed" system is obviously equal to the waiting time in its first queue, which is an M/G/1 queue ("Q3") with arrival intensity $\lambda^{-1}$ and service time distribution $B(\cdot)$. For the queue $Q_3$ we have, with an obvious notation:

$$E\{w(3)^2\} = \frac{\beta_3(2)}{2\alpha(1 - \lambda)}.$$

Remark 3.7

The approach of the present study can be carried over to allow Erlangian interarrival time distribution. It is also possible to determine the joint distribution of the sojourn times in $Q_1$ and $Q_2$, and of the amounts of work in $Q_1$ and $Q_2$, following the approach outlined in [1]; but these results will be omitted.

Appendix

Define

$$Y_n \overset{\text{df}}{=} \varepsilon_n(1),$$

$$\bar{Y}_n = \sum_{k=1}^{n} Y_k.$$

In the present more general model $\text{cov}(\varepsilon_n(1), \varepsilon_n(2))$ is again linearly dependent on $\beta(2)$, although it is not directly proportional to $(\beta(2) - \beta(1))$. 

Remark 3.6

In Section 1 we have mentioned a result of Tembe and Wolff [8] concerning the optimal order of queues in a tandem system; this result implies that in the present model the total time spent in the system is stochastically larger than the total time spent in the system with reversed order of queues. Obviously a similar result holds for the total waiting time (= total time in system minus service times). The total waiting time in the "reversed" system is obviously equal to the waiting time in its first queue, which is an M/G/1 queue ("Q3") with arrival intensity $\lambda^{-1}$ and service time distribution $B(\cdot)$. For the queue $Q_3$ we have, with an obvious notation:

$$E\{w(3)^2\} = \frac{\beta_3(2)}{2\alpha(1 - \lambda)}.$$

Remark 3.7

The approach of the present study can be carried over to allow Erlangian interarrival time distribution. It is also possible to determine the joint distribution of the sojourn times in $Q_1$ and $Q_2$, and of the amounts of work in $Q_1$ and $Q_2$, following the approach outlined in [1]; but these results will be omitted.
with \( n^{(1)} \) the number of customers served during a busy period of the M/G/1 queue \( Q_{1} \).

**THEOREM A.1**

For \( |p| \leq 1, \Re s \geq 0, \)
\[
E(p^n e^{-sx}) = p \beta^{(2)}(s) \beta^{(1)}((1-E(p e^{-sx}))/\alpha - s). \tag{A.1}
\]

**Proof**


Let
\[
a(r;p,s)d!f r-p \beta^{(1)}((1-r)/\alpha - s)\beta^{(2)}(s), \quad |p| \leq 1, \Re s \geq 0,
\]
be a function of \( r \). Let \( \zeta(p,s) \) be the zero of \( a(r;p,s) \) which has the smallest absolute value. Such a zero exists since \( E(p^n e^{-sx}) \) is according to Theorem A.1 a zero of \( a(r;p,s) \), and \( |E(p^n e^{-sx})| \leq 1 \) for \( |p| \leq 1, \Re s > 0; \)
for \( p = 1, s = 0, r = 1 \) is obviously a zero.

**THEOREM A.2**

a. The function \( a(r;p,s) \) has only one zero in \( |r| < 1 \) if
   (i) \( |p| < 1, \Re s \leq 0; \)
   (ii) \( |p| < 1, \Re s > 0, \)
   (iii) \( |p| \leq 1, \Re s > 0, a_1 > 1. \)

b. If \( p = 1, s = 0, a_1 \leq 1, \) then \( \zeta(1,0) = 1. \)

**Proof**

(a) and (ii) follow easily by using Rouché's Theorem (cf. Titchmarsh [9]). Now consider (iii). For \( \epsilon > 0 \) sufficiently small and \( a_1 > 1 \) we have with \( |r| = 1 - \epsilon: \)
\[
|p \beta^{(1)}((1-r)/\alpha - s) \beta^{(2)}(s)| \leq |\beta^{(1)}((1-r)/\alpha - s)|
\]
\[
\leq |\beta^{(1)}(\epsilon/\alpha)| = 1 - \epsilon a_1 + o(\epsilon) \approx 1 - \epsilon = |r|,
\]
and application of Rouché's Theorem to (A.2) leads to the desired result. Part b is a special case of a statement proved in Takács [7], pg. 48.

**THEOREM A.3**

\[
\zeta(p,s) = E(p^{n^{(1)}} e^{-sx^{(1)}})
\]
in all cases mentioned in Theorem A.2.

**Proof**

Trivial, except for the case \( p = 1, s = 0, a_1 > 1, \) because \( r = 1 \) is then a second zero of \( a(r;p,s) \) in the region \( |r| \leq 1 \). But according to M/G/1 theory \( E(p^{n^{(1)}} e^{-sx^{(1)}}) < 1 \) if \( a_1 > 1, \) so again \( E(p^{n^{(1)}} e^{-sx^{(1)}}) \) is the zero of \( a(r;p,s) \) in \( |r| \leq 1 \) which has the smallest absolute value.

**REFERENCES**