THE QUEUE AS A STOPPED RANDOM WALK

Elena M. Johnson
Bell Telephone Laboratories, Inc.
Holmdel, New Jersey USA

ABSTRACT

A single-server queuing system is analyzed on the basis of its underlying stochastic processes, which we here define in terms of the stopping variables of certain random walks. By applying results from the theory of stopped random walks the first moments of these stochastic processes may be determined as well as lower bounds for the second moments.

We begin with a description of the general, single-server queue as a stopped random walk with emphasis on the stopping variables which arise therein. Results are then derived for the idle and busy periods in the general system, that is, GI/G/1. The first moments are derived exactly in terms of the moments of one particular stopping variable: namely, the number of customers served during a busy period. For the second moments, lower bounds are derived. Also in the general case, an expression for the expected waiting time in similar terms is obtained. These results are then applied to specific cases of the single-server queue, i.e., M/G/1 and GI/M/1.

The queue length is not amenable to similar analysis in GI/G/1. Expressions for the moments of this stochastic process are derived separately for the systems M/G/1 and GI/M/1.

The approach employed in this paper provides a simplification over previous methods for obtaining at least lower bounds for the moments of various distributions in queuing systems. In addition, in dealing with systems where no steady state is achieved, approximations may be obtained for the moments of queue-length, waiting time, and idle periods.

1. GI/G/1 AS A STOPPED RANDOM WALK

In this paper we propose to derive expressions for the first and second moments of the idle period, busy period, waiting time, and queue length for the single-server queuing system. The expressions are derived in terms of stopping variables of a stopped random walk which we define within the context of a queuing system.

In order to prepare the way for our analysis a description of the system GI/G/1 is given. Consider customers arriving at a single server. Let \( u_1, u_2, \ldots \) be the intervals of time between successive arrivals and \( v_1, v_2, \ldots \) be the length of successive service times. We assume that \( \{u_n\} \) and \( \{v_n\} \) form independent sequences of independent, identically distributed random variables. Thus, the sequences form two independent renewal processes.

Now define another sequence of random variables \( \{X_n\} \) such that \( X_n = v_n - u_n \) for \( n \geq 1 \). Then \( a = E[X_n] = b - a \), and \( \sigma^2 = \text{var}[X_n] = \sigma^2 + \sigma_0^2 \). Consider the sequence of partial sums \( \{S_n\} \) where \( S_0 = 0 \) and \( S_n = X_1 + X_2 + \ldots + X_n \) \((n \geq 1)\). This sequence \( \{S_n\} \) constitutes a random walk.

Let \( W_n \) be the waiting time (before going into service) of the \( n \)th arrival to the system. We have the recurrence relation,

\[
W_{n+1} = \begin{cases} 
W_n + v_{n+1} - u_{n+1} & \text{if } W_n + v_{n+1} - u_{n+1} < 0 \\
0 & \text{if } W_n + v_{n+1} - u_{n+1} \geq 0
\end{cases}
\]

or

\[
W_{n+1} = \max(W_n + X_{n+1}, 0).
\]

With \( W_0 = 0 \), it can be shown that

\[
W_n = \max(0, S_1, S_2, \ldots, S_n) = M_n \quad (\text{say}).
\]

The first index \( N > 0 \) for which \( S_N < 0 \); that is, the first time the random walk \( \{S_n\} \) enters the interval \((-\infty, 0]\), is called the first descending ladder epoch of this random walk, and the sum \( S_N \) is the descending ladder height corresponding to \( N \). Thus, \( N = \min(n : n \geq 1, W_n = 0 \implies W_n = 0) \) (most of our results are independent of the initial condition. Thus it will be dropped). It can be shown that the duration of the first idle period is given by \( I = -S_N \geq 0 \). In addition, the \( N \) has the

* Here for two random variables \( X, Y \) which have the same distribution we write \( X \sim Y \).
Further significance in our analysis in that the busy period B is clearly
\[ B = v_1 + v_2 + \ldots + v_N, \]
and the busy cycle R is
\[ R = u_1 + u_2 + \ldots + u_N. \]
We will also be interested in ladder epochs other than the first. The ith descending ladder epoch is defined by \( N_i = \min(n: S_n - S_{n-1} \leq 0) \), and the corresponding ladder height is given by \( Z_i = S_{N_i} - S_{N_i-1} \). Now let \( \bar{N}(n) \) be the number of idle periods up to the time of the nth arrival. In terms of the descending ladder epochs, \( \bar{N}(n) = \max(k: N_k \leq n) \). Then if \( I_n \) is the total idle time up to the time of the nth arrival
\[ I_n = I_1 + I_2 + \ldots + I_{\bar{N}(n)}, \]
where \( I_k \) is the duration of the kth idle period. Also,
\[ W_n = S_n + I_1 + I_2 + \ldots + I_{\bar{N}(n)}. \]
Figure 1 gives a graphical representation of this random walk and shows the relationship of all the parameters.

In a random walk the ascending ladder epochs and corresponding ladder heights are also defined. Thus, \( N = \min(n: S_n > 0) \) and \( Z = S_n \) define the ascending ladder and epoch and corresponding ladder height respectively. As before, succeeding ascending ladder epochs and heights are defined: \( N_j = \min(n: S_n - S_{n-1} > 0) \) and \( Z_j = S_{N_j} - S_{N_j-1} \). The significance of these random variables in the queuing context is not immediately obvious. However, for example, recall that
\[ W_n = \max(0, S_1, S_2, \ldots, S_n). \]
The distribution of this maximum (for finite \( n \) as well as \( n = \infty \)) can be obtained in terms of the distribution of the random variable \( Z = S_n > 0 \) as follows:
\[ W_n \sim Z + Z_2 + \ldots + Z_{\bar{N}(n)} \]
where \( \bar{N}(n) = \max(k: N_k \leq n) \). In addition, it is known that there exists a duality relation between a given queuing system and one obtained by interchanging the interarrival and service times such that the ascending ladder epochs in one system become the descending ladder epochs in the dual system, with care being exercised to consider the sign of the ladder heights and the strength of the inequalities.

2. THE IDLE PERIOD, BUSY PERIOD AND WAITING TIME

We retain all the notation of the previous section. In addition let
\[ A = \frac{1}{n} \sum_{i=1}^{n} \Pr(S_n \leq 0), \]
\[ F = \frac{1}{n} \sum_{i=1}^{n} \Pr(S_n > 0) \]
and
\[ C = \frac{1}{n} \sum_{i=1}^{n} \Pr(S_{n+1} > 0) - \frac{1}{n} \sum_{i=1}^{n} \Pr(S_n > 0) \]
From a discussion in Prabhu[5] of a result by Feller the probability generating function of \( \bar{N} \) is given by
\[ A(z) = E(z) = 1 - e^{-C} \]
We now prove that if the series \( F \) converges, then
\[ E[R] = e^F \]
and, if furthermore \( F = \sum_{i=1}^{\infty} \Pr(S_n > 0) \) also converges then
\[ E[N^2] = 2e^F + e^F. \]
Since \( A + E = \infty \), it follows that \( A = \infty \), and \( \bar{N} \) is a proper variable, that is \( \Pr(\bar{N} = 0) = 1 \). From (1)
\[ 1 - A(z) = \frac{1 - z}{1 - z} \sum_{i=1}^{n} \Pr(S_n > 0) \]
so that
\[ E[\bar{N}] = 1 / \frac{1 - A(z)}{1 - z} = e^F. \]
Further
\[ \frac{1}{2} \Pr(E[\bar{N}(n-1)] = 1 / \frac{1 - A(z)}{1 - z} \]
This leads to (3) (Let us note that the conditions \( F = \infty \) and \( A = \infty \) correspond to the case \( n = 0 \)).
We said that \( \bar{N}(n) = \max(k: N_k \leq n) \) is number of idle periods up to the time of the nth arrival. If \( p < 1 \), we will show that as \( n \to \infty \)
\[ \bar{N}(n) = \text{var}(N) E[N] - E[N^2] - 2E[N]^2 + E[N] \]
and
\[ \text{var}(\bar{N}(n)) = n \text{var}(N) \]
Also, if \( p = 1 \)
\[ \bar{E}[\bar{N}(n)] = \frac{2Cn^{1/2}}{\sqrt{n}} \]
where \( C \) is as defined in (1). Clearly, the event \( E = \{ \text{an arriving customer finds the queue empty} \} \) is a recurrent event. The number of arrivals to the succeeding occurrence has the same distribution as \( N \). Then \( \bar{N}(n) \) is the number of times \( E \) has occurred up to the time of the
n + n arrival. Let $v_n^* = \Pr(E \text{ occurs at the time of the } n\text{th arrival})$, which finds the 
queue empty. Then $E[N_n] = v_1^* + v_2^* + \ldots + v_n^*$
and $v_1^* = 1$. When $p < 1$ the event $E$ is persistent,
aperiodic, so as $n \to \infty$, $v_n^* \to 1$. We refer to
Feller[2].

Theorem 9 from Feller[2], which derives expressions
for the mean and variance of finite recurrent
events, to obtain equations (4) and (5).
For a proof of (6) the reader is referred to 
Feller[4].

For the idle period in GI/G/l with $p < 1$, we obtain the following results:

$$E[I] = (-\alpha)E[N]$$

and

$$E[I^2] \geq \sigma^2 E[N] - \alpha^2 E[N^2]$$

except in the case D/D/l where $X_n$ is constant
with probability one. To show this we consider
the variables $X_n = v_n - u_n (n > 1)$, which are independent,
and identically distributed with mean $\alpha$ and variance $\sigma^2$. The descending ladder
epoch $N$ is clearly a stopping variable for the
sequence $(S_n)$. Recall that $I = S_N$. In order to
apply Theorem 2 developed by Chow, Robbins,
and Teicher[1] we must have $N < \infty$ with probabil­
ity one and $E[N] < \infty$. Both conditions are sati­s­
ﬁﬁed provided $\alpha < (p - 1)$. The expression for $E[I]$ follows directly from the theorem. For $E[I^2]$ we have:

$$E[I^2] = \sigma^2 E[N] - 2\alpha E[N] - \alpha^2 E[N^2]$$

The term $E[N]$ is obviously positive, and since
$p < 1$, the second term on the right side of (9)
is positive and (8) follows. In the case D/D/l
the inequality is trivial.

For $p = 1$,

$$E[I] = \frac{\alpha^2}{C_2}$$

is a result which is based on a theorem given in 

If the term $E[N]$ can be explicitly evaluated,
(9) gives the exact value for $E[I^2]$. M/G/l is
such a case as will be shown.

In GI/G/l, if $p < 1$, then the busy period will
terminate with probability one. If $p < 1$ then
$E[B] = \alpha E[N]$

and

$$E[B^2] \geq \sigma^2 E[N] - b^2 E[N^2]$$

except in GI/D/l where the above is trivial.
The busy period is given by $B = v_1 + v_2 + \ldots + v_p$.
Thus the derivations of (11) and (12) take the
same form as the derivations of the results for $I$.

Recall from the description of the general queue
that the waiting time may be given as:

$$W = S_1 + S_2 + \ldots + S_N$$

Then

$$E[W] = \alpha n + E[I]E[N]$$

We have $E[N] = u_1 + u_2 + \ldots + u_n$ where $u_n$
= Printh arrival finds the queue empty. Then $u_n$
= $\frac{1}{1-p}$ as $n \to \infty$, and applying Theorem 9 of
Feller[2] as before, if $p < 1$

$$E[W] = \alpha E[N^2] - 2\alpha E[N] + E[N]$$

If $p = 1$ ($\alpha = 0$),

$$E[W] = E[I]E[N]$$

and from (6) and (10)

$$E[W]^2 \approx \frac{\sigma^2}{\pi} n^{1/2}$$

Recall from our earlier discussion that

$$W = Z_1 + Z_2 + \ldots + Z_N$$

where $Z_k$ are successive ascending ladder heights
and $N(n) = \max(k: N_k < n)$. We have, for $p > 1$,

$$E[Z] = \alpha E[N]$$

and

$$\text{EN}(n) = \frac{E[N]}{E[N] - n(1-p)}$$

so

$$E[W]^2 \approx \frac{\sigma^2}{\pi} n^{1/2}$$

3. APPLICATIONS TO M/G/l

Consider the system M/G/l in which customers ar­
rive in a Poisson process at rate $\lambda$, and the
distribution function of the service times is $B(t)$
($0 \leq t \leq \infty$). Let $\phi(\theta)$ be the Laplace-Stieltjes
transform of the service time distribution. We
assume a ﬁnite mean service time equal to $b$ and
$p = \lambda b$. In this case:

$$\alpha = E[X_n] = \frac{\rho - 1}{\lambda}$$

and

$$\sigma^2 = \text{var}[X_n] = E[v^2] + \frac{(1 - p^2)}{b^2}.$$
and

\[ \text{var}(N(n)) = n[\lambda^2 E(Y^2) + \rho - \rho^2]. \]  

(29)

For the idle period we get in M/G/l:

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

(30)

as expected. For \( E[I^2] \), recall that in M/G/l the distribution of the idle period is independent of the number of customers served in a busy period. Therefore the product moment term in (9) may be evaluated exactly to get

\[ E[I^2] = \frac{2(1-\rho)}{(1-\rho)^2}. \]

(31)

Turning to the total idle time up to the \( n \)-th arrival, recall that

\[ I_n = I_1 + \ldots + I_N(n). \]

Applying our results we get

\[ E[I_n] = n \alpha \]

and

\[ \text{var}(I_n) = n \alpha^2. \]

(32)

For the busy period in M/G/l with \( \rho < 1 \)

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

(34)

and

\[ E[B^2] \geq \frac{(1-2\rho)E[I^2]}{(1-\rho)^2} = \frac{2b}{(1-\rho)^2}. \]

(35)

In M/D/l

\[ E[B^2] = \frac{\rho^2 \sigma^2}{(1-\rho)^3}. \]

(36)

The expected waiting time in M/G/l is obtained by substituting the expressions for \( E[N] \) in (14). Thus,

\[ E[W_n] = \frac{\lambda E[I^2]}{2(1-\rho)}. \]

(37)

4. APPLICATIONS TO GI/M/l

In the system GI/M/l the interarrival times have an arbitrary distribution function \( B(t) \) \( (0 < t < \infty) \) with a finite mean \( b \) and second moment \( \sigma^2 \).

The traffic intensity is \( \rho = (\lambda b)^{-1} \), and

\[ \alpha = E[X_1] = -b(1-\rho) \]

\[ \sigma^2 = \text{var}[X_1] = E[u^2] - b^2(1-\rho^2). \]

(38)

(39)

In this system, if the interarrival and service time distributions are interchanged, the resulting system is M/G/1. We make use of this duality between the two systems to derive expressions for \( E[N] \) and \( E[N^2] \) in GI/M/l. Recall that we derived the probability generating functions of \( N \) and \( N_i \), the descending and ascending ladder epochs in M/G/l. The duality relation implies that the ascending ladder epochs in M/G/l become the descending ladder epochs in GI/M/l. Therefore, from (25) the probability generating function of \( N \) in GI/M/l is

\[ Y(t) = \frac{t^\xi}{1-t^\xi}. \]

Thus,

\[ E[N] = \frac{1}{1-\xi}. \]

(40)

and

\[ E[N^2] = \frac{2\xi}{(1-\xi)^2} \cdot \frac{1}{(1-K'(\xi))} + \frac{1}{(1-\xi)}. \]

(41)

where \( \xi \) is the smallest positive root of the equation \( \xi = K(\xi) = \psi(\lambda - \xi) \) and \( \psi(b) \) is the Laplace-Stieltjes transform of the interarrival time distribution.

Thus, we have the following results for \( \rho < 1 \)

\[ E[N(n)] = n(1-\xi) + \frac{1}{1-\xi} \cdot \frac{1}{1-K'(\xi)} - \xi \]

(42)

\[ \text{var}(N(n)) = n(1-\xi) \left[ \frac{2\xi}{(1-\xi)^2} \cdot \frac{1}{(1-K'(\xi))} - \xi \right] \]

(43)

from (4) and (5).

\[ E[I] = \frac{b(1-\rho)}{1-\xi}. \]

(44)

Further,

\[ E[I^2] > \frac{E[u^2] - 2(1-\rho)b^2}{(1-\xi)^2} \cdot \frac{1}{(1-K'(\xi))} + \frac{b^2}{1-\xi}. \]

(45)

and since the busy cycle \( R = U_1 + U_2 + \ldots + U_N \),

\[ E[R] = \frac{b}{1-\xi}. \]

(46)

and

\[ E[R^2] > \frac{2\xi b^2}{(1-\xi)^2(1-K'(\xi))} + \frac{b^2}{1-\xi}. \]

(47)

The expected waiting time is:

\[ E[W_n] = \frac{-\alpha K'(\xi)}{(1-\xi)(1-K'(\xi))}. \]

(49)

5. QUEUE LENGTH

Another stochastic process of practical interest in a queuing system is the queue length; that is the total number of customers in the system. In order to analyze this process we consider the particular systems M/G/l and GI/M/l. The general queue does not seem amenable to the treatment given here.

Consider a sequence of independent random variables \( \{F_n: n \geq 1\} \) with a common distribution concentrated on \( (-1,0,1,2,...) \). Let us write the probability generating function as

\[ K(t) = \sum_{n=0}^\infty K_n t^n \]

where \( K(t) \) is itself a probability generating function. Let \( S_0 = 0; S_n = Y_1 + Y_2 + \ldots + Y_n \) \( (n > 1) \). For the random walk \( \{S_n\} \) we define the descending ladder epochs \( N_i \) by setting \( N_1 = \text{min}(n: S_n < 0) \) and \( N_i = \text{min} \{n > N_{i-1}: S_n - S_{N_{i-1}} < 0\} \) \( (i > 2) \). The ladder heights are defined as before.

Using Wald's identity, we may write

\[ E[Y_S] = 1 \]

(51)

where

\[ 1 = t K(z). \]

However, it is clear that \( S_N = -l \) with probability 1. So we get

\[ E[Y_S(t)] = z = \xi(t) \]

(52)
and
\[ E[t^N] = \frac{l-\xi}{1-\xi}. \]  

5.1 Queue Length in M/G/1

In this system let \( t_n \) (\( n \geq 0 \)) denote the epoch of the departure of the \( n \)th arrival to the system. \( Q_n \) is defined as the number of customers in the system at time \( t_n + 0 \). We assume \( Q_0 = 0 \). Let \( U_n \) denote the number of customers arriving during the service time of the \( n \)th customer. The \( U_n \) are independent, identically distributed random variables with probability generating function
\[ U \left( e^t \right) = K(t) = \psi(\lambda - \lambda t) \]
where \( \psi(0) \) is the transform of the service time distribution. Then \( E[U_n] = K'(1) = -\lambda \psi'(0) = \rho \).

Clearly, \( Q_n \) may be defined as
\[ Q_{n+1} = \max(U_n, Q_n + U_n - 1) \quad n \geq 0 \]

Let \( Y_n = U_n - 1 \), and \( S_n = Y_1 + Y_2 + \ldots + Y_n \) \((n \geq 1)\). Since \( E[t^S_n] = t^{-1}K(t) \) it follows from the previous discussion that the \( K \) in (50) is the same as \( K \) derived in Section 2, and \( N \) and \( N' \) of the random walk \( (S_n) \) here have the same distribution and physical interpretation as the \( N \) and \( N' \) derived in that previous section. Thus,
\[ Q_n = S_n + N(n) \]

where \( N(n) = \max(i : \ N_i > n) = 1 + \) number of busy cycles in \((0,n)\). With \( p < 1 \) \( E[Q_n] \) may be written as
\[ E[Q_n] = np - 1 + v^* + \ldots + v^{n-1} \]

where \( v^* = P_n(\text{nth customer finds the system empty})(n > 1) \) and \( v_0 = 1 \). Once again, applying Feller[2] Theorem 9, we get
\[ E(Q_n) = \rho + E[N^2] - 2E[N^2] + E[N] \quad n \geq 0 \]
and for \( M/G/1 \):
\[ E[Q_n] = \rho + \frac{\lambda^2 E[v^2]}{2(1-\rho)}. \]

5.2 Queue Length in GI/M/1

In GI/M/1, we let \( t_n \) (\( n \geq 0 \)) be the epoch of the \( n \)th arrival and \( Q_n \) denote the number of customers in the system at time \( t_n + 0 \). We assume \( Q_0 = 0 \), and define \( V_n \) as the number of departures from the system in the interval \((t_n,t_{n+1})-0\) for \( n \geq 0 \). The \( V_n \) are independent, identically distributed random variables with probability generating function
\[ E[V_n] = K(t) = \psi(\lambda - \lambda t) \]
where \( \psi(0) \) is the transform of the interarrival distribution
\[ E(V_n) = K'(1) = \frac{1}{\rho}. \]

If we let \( Y_n = V_n - 1 \) and \( S_n = Y_0 + Y_1 + \ldots + Y_n \) then
\[ Q_n = \max(0,Q_n - Y_n + 1) \]
or
\[ Q_n \sim \min(S_n). \]

Then
\[ Q_n \sim [Z_1 + Z_2 + \ldots + Z_N(n)] \]
where \( Z_i = -1 \) with probability one, and \( N(n) = \max(i : N_i \leq n) \).

Thus,
\[ Q_n \sim N(n) \quad \text{and} \]
\[ E(Q_n) = EN(n). \]

For \( \rho > 1 \) and since \( E(N) = 1 - \frac{1}{\rho} \)
\[ E[Q_n] = n\left(1 - \frac{1}{\rho}\right) \]
and
\[ \text{var}(Q_n) \approx n\left[\frac{\lambda^2 E[v^2]}{\rho} - \frac{1}{\rho^2} + \frac{1}{\rho}\right]. \]

If \( \rho < 1 \) we note that
\[ E[Q_n] = v_1 + \ldots + v^n - \frac{1}{1 - \rho(N - 1)} + \frac{1}{\rho}. \]

6. SUMMARY

We have given a description of the single-server queue as a stopped random walk. Once in this context, the results obtained are derived in terms of the moments of the number of customers served during a busy period which is a particular stopping variable.

Expressions for the first moments of the idle period, busy period and waiting times are derived in GI/G/1. In addition lower bounds for the second moments of the idle and busy periods are obtained. The general results are then applied to the systems M/G/1 and GI/M/1.

The analysis of queue-length is made in the two systems M/G/1 and GI/M/1 separately, and expressions for the mean queue-length are derived.

Queueing systems with \( \rho \geq 1 \) are of practical interest. We have included some results concerning such systems. In particular expressions for the mean idle period, mean waiting time in GI/G/1 and the mean and variance queue length in GI/M/1 are obtained. It is hoped that an extension of the method presented here may provide more information concerning such systems.

BIBLIOGRAPHY