SOME RESULTS FOR INFINITE SERVER POISSON QUEUES

by

SHELDON M. ROSS AND MARK BROWN

OPERATIONS RESEARCH CENTER

UNIVERSITY OF CALIFORNIA - BERKELEY
SOME RESULTS FOR INFINITE SERVER POISSON QUEUES

by

Sheldon M. Ross
Department of Industrial Engineering
and Operations Research
University of California, Berkeley

and

Mark Brown
Department of Operations Research
Cornell University
Ithaca, New York

September 1968

ORC 68-23

This research has been partially supported by the U.S. Army Research Office-Durham under Contract DA-31-124-ARO-D-331 and the National Science Foundation under Grant GP-8695 with the University of California. Reproduction in whole or in part is permitted for any purpose of the United States Government.
ABSTRACT

A generalization of the M/G/« queueing system with batch arrivals to one with time dependent arrival rates, service times, and batch size distributions is considered. It is shown that both \( W(t) \), the number of people being served at \( t \), and \( S(t) \), the number of people who have completed service by \( t \), are distributed as compound Poisson laws. The distributions of the traffic time average \( T^{-1} \int_0^T W(t) dt \) and the occupation time \( O(t) \) (the amount of time past \( t \) until the system becomes empty, under the assumption that no new customers are served after \( t \)) are also derived.

The limiting proportion of busy time and the asymptotic behavior of the traffic time average are also discussed in the time homogeneous case.
SOME RESULTS FOR INFINITE SERVER POISSON QUEUES

by

Sheldon M. Ross
Mark Brown

0. Introduction and Summary

We consider a queueing model in which arrivals occur according to a nonhomogeneous Poisson Process in batches of varying size, and in which a customer is served immediately upon arrival by one of an infinite number of servers.

We allow for the possibility that both the batch size and service time distributions might depend on the arrival time and thus denote by \( P_t(r) \), the probability that a batch arriving at time \( t \) will contain \( r \) customers \((r > 1)\), and by \( G_t() \), the service time distribution of a customer arriving at time \( t \). We also let \( m(t) \) denote the mean value function of the Poisson Process of arrivals. This system is thus the generalization of \( M/G/\infty \) with batch arrivals to time dependent arrival rates, service times and batch size distributions.

In the first section we show that both \( W(t) \), the number of customers being served at time \( t \), and \( S(t) \), the number of customers who have completed service by time \( t \), are distributed as compound Poisson laws. In the second section we derive, in the time homogeneous case \((G_t() = G(), P_t() = P(), m(t) = \lambda t)\), the limiting proportion of time that the system is nonempty.

In the third section we derive the distribution of the occupation time \( 0(t) \); where \( 0(t) \) is defined as the amount of time past \( t \) until the system becomes empty, under the assumption that no new customers are served after time \( t \).

In the fourth section we derive the distribution of the traffic time average \( T^{-1} \int_0^T W(t)dt \), and its asymptotic behavior is discussed in the time homogeneous case.
In [7] Shanbag considered a special case of the above model, $G_t = G$, $P_t() = P()$, and by solving a differential equation, derived the joint generating function of $W(t)$ and $S(t)$. His method, however, does not seem applicable to the present model unless some conditions (such as t-continuity) are placed on $P_t()$ and $G_t()$.

Benes (rf. [6], p. 123) has previously obtained the distribution of the traffic time average for the case $M/M/\infty$, and in a more recent paper [5] Rao has generalized this to the case $CG.I/G/\infty$ where $CG.I$ stands for any stationary stream of random jumps (batch arrivals) for which the times between successive jumps are independent and identically distributed. His results thus include $M/G/\infty$ with batch arrivals as a special case. The method employed in the present paper differs from those used in the above papers.
1. Distribution of $W(t)$ and $S(t)$

Throughout this paper we shall assume $P_x(r)$ is a measurable function of $x$ for all $r$ and that $G_x(t - x)$ is a measurable function of $x$ for all $t$.

We shall suppose that the process first begins at $t = 0$, and we let $B(t)$ be the number of batches which have arrived by time $t$. Then $W(t) = \sum_{i=1}^{B(t)} y_i$,

where $y_i$ denotes the number of arrivals from the $i^{th}$ batch that are being served at time $t$. Thus

$$P(W(t) = k) = \sum_{n=0}^{\infty} \frac{e^{-m(t)} (m(t))^n}{n!} p\left\{ \sum_{i=1}^{n} y_i = k \mid B(t) = n \right\}. \quad (1)$$

Now conditional on $B(t) = n$, the (unordered) arrival times of the batches are distributed as a sample of independent and identically distributed (i.i.d.) random variables with a distribution given by $F(x) = \begin{cases} \frac{m(x)/m(t)}{\int_{0}^{t} x \leq t} & \frac{1}{t} \int_{t}^{\infty} \frac{dx}{x} \geq t \end{cases}$.

Thus conditional on $B(t) = n$, $\sum_{i=1}^{B(t)} y_i$ is distributed as the sum of $n$ i.i.d. random variables $Z_1, ..., Z_n$ each having probability distribution

$$P(Z_i = j) = \int_{0}^{t} \frac{1}{m(t)} \sum_{r=j}^{\infty} P_x(1 - G_x(t - x))^j (G_x(t - x))^{r-j} dm(x) \quad (2)$$

Thus,

$$P(W(t) = k) = \sum_{n=0}^{\infty} \frac{e^{-m(t)} (m(t))^n}{n!} P(Z_1 + ... + Z_n = k) \quad (3)$$

and so $W(t)$ has a compound Poisson distribution with Poisson parameter $m(t)$ and with jumps distributed according to (2) — i.e., $W(t) = \sum_{i=1}^{B(t)} Z_i$ where $Z_i$ are i.i.d. according to (2) and are independent of $B(t)$. The probability
generating function $\psi_W(t)(s) = E(s^W(t))$ is given by

$$\psi_W(t)(s) = \exp\left\{ \sum_{j=0}^{\infty} \frac{(s^j - 1)}{j!} \int_0^t \sum_{r=j}^{\infty} P_x(r) \frac{r^j}{j!} (1 - G_x(t - x))^j (G_x(t - x))^r - \frac{rj}{j!} dm(x) \right\}.$$ (4)

A similar analysis may be done to show that $S(t)$ has a compound Poisson distribution with Poisson parameter $m(t)$ and jump distribution $V$ where

$$P[V = j] = \int_0^t \frac{1}{m(t)} \sum_{r=j}^{\infty} P_x(r) \frac{r^j}{j!} (1 - G_x(t - x))^j (G_x(t - x))^r - \frac{rj}{j!} dm(x), \quad j \geq 0.$$ (5)
2. Limiting Proportion of Busy Time (Homogeneous Case)

In this section we suppose that \( m(t) = \lambda t \), \( G_x = G \), and \( P_t = P \).

From (2) and (3) it follows that

\[
P(W(t) = 0) = \exp \left\{ -\lambda \int_0^t (1 - G^r(x)) P(r) dx \right\} = e^{-\lambda M} \quad \text{as} \quad t \to \infty ,
\]

where

\[
M = \int_0^\infty (1 - G^r(x)) P(r) dx .
\]

Now as time passes, there will be periods of time at which the queue is empty which will alternate with periods at which the queue is busy. Let \( A_i \) be the length of the \( i \)th empty period and \( D_i \) the length of the \( i \)th busy periods.

The sequences \( \{A_i\}_1^\infty \) and \( \{D_i\}_1^\infty \) are independent renewal sequences, and thus the sequence \( \{A_1, D_1, A_2, D_2, \ldots\} \) is an alternating renewal sequence. Then it is known (see [4]) that \( P\{\text{Queue is empty at } t\} \to \frac{E(A)}{E(A) + E(D)} = \frac{1}{1 + \lambda} \quad \text{as} \quad t \to \infty ,\)

and thus from (6) we have that

\[
ED = \frac{\lambda M - 1}{\lambda} .
\]

Let \( C_i = A_i + D_i \), then \( \{C_i\} \) is a renewal sequence. Let \( N(t) \) denote the number of \( C \)-renewals up to time \( t \). Since

\[
\frac{N(t)}{t} \xrightarrow{a.s.} \frac{1}{EC} = \lambda e^{-\lambda M}
\]

If \( r \) denotes a random batch size, and \( Y_1, \ldots, Y_r \) the service times then

\[
M = E[\text{Max}(Y_1, \ldots, Y_r)] \quad \text{and this is finite if} \quad \int_0^\infty (1 - G(y) dy \quad \text{and} \quad \int_0^\infty r P(r) \quad \text{are finite.}
\]

\(^{\dagger}\)Equation (7) was derived by a different method in [7].
and

\[ \frac{1}{t} \sum_{i=1}^{N(t)} A_i \xrightarrow{a.s.} e^{-\lambda M} \text{ as } t \to \infty, \]  
(8)

and from Wald's Fundamental Equation of Sequential Analysis that

\[ E\left[ \frac{1}{t} \sum_{i=1}^{N(t)} A_i \right] \to e^{-\lambda M} \text{ as } t \to \infty. \]  
(9)

Now let \( A(t) = \) amount of time the queue is empty up to time \( t \)
\( D(t) = \) amount of time the queue is busy up to time \( t \).

Now \( |A(t) - \sum_{i=1}^{N(t)} A_i| \leq A_{N(t)+1} \), and since \( \frac{1}{t} E[A_{N(t)+1}] \to 0 \) and \( \frac{1}{t} A_{N(t)+1} \to 0 \) with problem 1 as \( t \to \infty \) we have by (8) and (9) that

\[ \frac{A(t)}{t} \xrightarrow{a.s.} e^{-\lambda M} \text{ as } t \to \infty. \]  
(10)

and

\[ E\left[ \frac{A(t)}{t} \right] \to e^{-\lambda M} \text{ as } t \to \infty. \]  
(10)

Also since \( D(t) = t - L(t) \) we have that

\[ \frac{D(t)}{t} \xrightarrow{a.s.} 1 - e^{-\lambda M} \]  
and

\[ E\left[ \frac{D(t)}{t} \right] \to 1 - e^{-\lambda M}. \]  
(11)
It can also be shown that \( A(t) \) and \( D(t) \) (suitably normalized) both having limiting normal distribution (see [8]).
3. Occupation Time

The occupation time $O(t)$ is defined as the amount of time past $t$ until the system becomes empty when no new customers are served after time $t$.

We say that a batch is served when all members of that batch have been served. Now the time points at which batches being served at time $t$ arrived may be shown (see [3], p. 497 or [1], p. 4) to form a nonhomogeneous Poisson Process with mean value function $\bar{\mu}(y) = \int_0^y (1 - \bar{G}_x(t - x)dm(x)) , y < t$, where

$$\bar{G}_x(a) = \sum_r P_r(r)G_x(r)$$; and thus given that there are $n$ batches being served at $t$ their (unordered) arrival times have the same distribution as an i.i.d. sample from

$$F(y) = \int_0^y (1 - \bar{G}_x(t - x)dm(x))/\int_0^t (1 - \bar{G}_x(t - x)dm(x)) \quad y < t$$

and so

$$O_x(t) = P(O(t) \leq x)$$

$$= \sum_n P(W(t) = n)\left(\int_0^t \frac{\bar{G}_y(t + x - y) - \bar{G}_y(t - y)}{1 - \bar{G}_y(t - y)} dF(y)\right)^n \quad (12)$$

$$= \psi_W(t)(\int_0^t \bar{G}_y(t + x - y) - \bar{G}_y(t - y)) dm(y)/\int_0^t (1 - \bar{G}_y(t - y)dm(y)$$.
4. Traffic Time Average

In order to obtain the distribution of \( \bar{W}_T = \frac{1}{T} \int_0^T W(t) dt \) we first note

\[
\int_0^T W(t) dt = \sum_{i=1}^T \sum_{j=1}^{t_i} \min (x_{ij}, T - \tau_i)
\]  

(13)

where

\( \tau_i = \text{arrival time of } i^{\text{th}} \text{ batch} \)

\( t_i = \text{number in } i^{\text{th}} \text{ batch} \)

\( x_{ij} = \text{service time of } j^{\text{th}} \text{ member of } i^{\text{th}} \text{ batch} \)

and thus

\[
\int_0^T W(t) dt = \frac{B(T)}{T} \sum_{i=1}^T L_i
\]  

(14)

where \( L_i \) is the sum at \( T \) of all the service times of members of the \( i^{\text{th}} \) batch. It thus follows as in Section 1 that

\[
\int_0^T W(t) dt = \frac{B(T)}{T} \sum_{i=1}^T R_i, \text{ where } R_i \text{ are i.i.d. independent of } B(T)
\]  

(15)

and where

\[
P(R_i < a) = \int_0^T \frac{1}{m(t)} \sum_r p_x(r) C(r) x_i(t) (a) dm(x)
\]  

(16)

where
\[ G_{x,T}(a) = \begin{cases} G_X(a) & a < T - x, \\ 1 & \end{cases} \]

and

\[ G^{(r)}_{x,T}(a) \] is the \( r \)-fold convolution.

Letting \( \phi_{G_{x,T}}(u) = \int_0^\infty e^{iua} dG_{x,T}(a) \), we have that

\[ \phi_U(u) = E(\exp(\frac{i u W_T}{T})) \]

\[ = \exp \left\{ \int_0^T \int_0^1 \mathbb{P}(x) \left( \left( \phi_{G_{x,T}}(u/T) \right)^r - 1 \right) dm(x) \right\}. \]
5. Homogeneous Case

We suppose that \( G = G \), \( P = P \), and \( M(t) = \lambda t \); also \( \mu_B = \sum r P(r) \), \( \mu_B^2 = \sum r^2 P(r) \), \( \mu_G = \int x dG(x) \), and \( \mu_G^2 = \int x^2 dG(x) \) are all assumed finite.

Let \( L_1 \) be the sum of the service times of members of the \( i \)th batch, and let \( \mu_L = \mu_B \mu_G \), and \( \mu_L^2 = \mu_B \sigma_G^2 + \mu_G^2 \). \( \sigma_G^2 = \int (x - \mu_G)^2 dG(x) \). Let

\[
S_T = \sqrt{T} \left( \frac{B(T)}{1} L_1 - \lambda \mu_L \right)
\]

and

\[
S_T^* = \sqrt{T} \left( \frac{B(T)}{1} L_1 - \lambda \mu_L \right).
\]

Now, \( \text{Var} (S_T - S_T^*) = \lambda \left( \mu_L - \text{EL}_1 \right)^2 + \lambda \left( \sigma_L^2 - \text{Var} L_1 \right) \to 0 \) as \( T \to \infty \). Also \( \text{E} (S_T^* - S_T) \to 0 \). Thus, \( S_T^* \) converging in distribution implies that \( S_T \) also converges in distribution to the same limit law. Now

\[
\phi_{s_T}(t) = \exp \left\{ \lambda T (\phi_{L_1}(t/\sqrt{T}) - 1) - i t \sqrt{T} \mu_L \right\},
\]

and

\[
\phi_{L_1}(t/\sqrt{T}) = 1 + i t \mu_L/\sqrt{T} - t^2 \mu_L^2/2T + o(T^{-1}),
\]

implying that

\[
\phi_{s_T}(t) = \exp \left\{ -\lambda \mu_L^2 t^2/2 \right\} \text{ as } T \to \infty.
\]

Thus,
\sqrt{T} (\bar{W}_T - \lambda \mu L) \overset{1}{\to} \text{Normal} \ (0, \ \lambda \mu^2).

(20)

Now, let \( N(a,b) \) be the number of customers arriving in \((a,b)\), and let \( N(t) = N(0,t) \).

Lemma 1:

\[ \frac{W(t)}{N(t)} \overset{a.s.}{\to} 0 \text{ as } t \to \infty. \]

Proof:

\[ \frac{W(t)}{N(t)} \leq \frac{N(t-n,t)}{N(t)} + \frac{W(t,n)}{N(t-n)}, \]

where \( W(t,n) \) denotes the number of customers arriving in \((0,t-n)\) whose service time is greater than \(n\). Thus,

\[ \lim_{t \to \infty} \frac{W(t)}{N(t)} \leq 1 - G(n) \text{ a.s. for all } n. \]

Q.E.D.

Theorem 1:

\[ \frac{1}{T} \int_0^T W(t) dt \overset{a.s.}{\to} \lambda \mu L \text{ as } T \to \infty. \]

Proof:

Suppose first that service times are bounded, i.e., \( G(M) = 1 \) for some \( M < \infty \), and let \( \{X_i, i = 1, \ldots, W(T)\} \) be the service times of customers being served at \( T \). Then

\[ \frac{B(T)}{L_i} \leq \frac{1}{N(T)} \sum \frac{W(T)}{L_i} \overset{1}{\to} 0 \text{ by} \]
Lemma 1. Thus,

\[ \frac{1}{T} \int_0^T W(t) dt = \frac{N(T)}{T} \frac{1}{N(T)} \sum \frac{B(T)}{1} L_1 = \lambda \mu L, \]

and so the result follows in the bounded case.

Now suppose only that \( \mu_B \) and \( \mu_C \) are finite. Let \( W^M(t) \) denote the number of customers being served at \( t \) whose service time at \( t \) is less than \( M \). Also, let \( \bar{N}(t) \) denote the number of batches arriving in \((0,t)\) having a member whose service time is greater than or equal to \( M \), and let \( \bar{L}_i \) be the sum of the service times of the \( i \)th such batch. Then

\[ \frac{1}{T} \int_0^T W(t) dt - \frac{1}{T} \int_0^T W^M(t) dt \leq \frac{1}{T} \sum \bar{L}_i \]

\[ \text{a.s.} \]

\[ \lambda E L_1 \left( 1 - \frac{1}{r} \sum P(r) G^r(M) \right), \]

where the convergence follows from the fact that \( \bar{N}(t) \) is a Poisson Process with mean value function \( \lambda t \left( 1 - \frac{1}{r} \sum P(r) G^r(M) \right) \). Now

\[ E L_1 = \sum_{r} P(r) \int_{y_1 + \ldots + y_r} dG(y_1) \ldots dG(y_r) / \lambda - \sum_{r} P(r) G^r(M) \]

(23)

where \( M^* = \{ \text{Max}(y_1, \ldots, y_r) \geq M \} \).

Thus, combining (22) and (23) we arrive at

\[ \frac{1}{T} \int_0^T (W(t) - W^M(t)) dt \leq \lambda \sum_{r} P(r) \int_{y_1 + \ldots + y_r} dG(y_1) \ldots dG(y_r) \text{ a.s.} \]

(24)
as \( T \to \infty \). However, the right-hand side of (24) goes to zero as \( M \to \infty \) (since \( \nu_B \) and \( \nu_G \) are finite), and so the result follows from the bounded case. If either \( \nu_B \) and \( \nu_G \) is infinite, the result follows from truncation.

Q.E.D.
REFERENCES


**Title:** SOME RESULTS FOR INFINITE SERVER POISSON QUEUES

**Authors:** ROSS, Sheldon M. and BROWN, Mark

**Date:** September 1968

**Contract or Grant No.:** DA-31-124-ARO-D-331

**Project No.:** 20014501B14C

**Other Report No.:** ORC 68-23

**Abstract:** SEE ABSTRACT.
Infinite Server
Nonhomogeneous Poisson Process
Occupation Time
Traffic Time Average