SECOND MOMENT OF QUEUE SIZE WITH STATIONARY ARRIVAL PROCESSES AND ARBITRARY QUEUE DISCIPLINE

R. Haji

Department of Industrial Engineering
Sharif University of Technology
Tehran, Iran

Abstract In this paper we consider a queuing system in which the service times of customers are independent and identically distributed random variables, the arrival process is stationary and has the property of orderliness, and the queue discipline is arbitrary. For this queuing system we obtain the steady state second moment of the queue size in terms of the stationary waiting time distribution of a similar queuing system in which the queue discipline is first-in-first-out.

Key Words Queuing Theory, Stochastic Processes, Applied Probability, Moments of Queue Size, Stochastic Service Systems

چکیده در این مقاله سیستم صفی مورد نظر است که در آن زمان خدمت مشتریان متغیرهای تصادفی مستقل با توزیع یکسان، فرآیند ورودی ساکن با خاصیت ترتیبی بوده، و نظم صف اختیاری است. برای این سیستم صف گشتاور دوم طول صف درحالت پایدار برحسب توزیع ساکن زمان انتظار دریک سیستم صف مشابه که نظم صف آن خروج به ترتیب ورود است بدست آمده است.

INTRODUCTION

The purpose of this paper is to investigate the relationship of the second moment of the queue size with the stationary waiting time distribution. Under great generality, the relation between the first moment of the queue size and the first moment of waiting time distribution can be obtained from the well-known Little's formula [4,7,11] which states that the expected number of customers in the queue is equal to the product of the arrival rate and expected waiting time distribution. In other words, the steady state first moment of the queue size is equal to the first moment of the number of customers who arrive during a time interval D, a random variable distributed as the stationary waiting time. Making the following assumptions:

1) arrivals join a single queue in front of one or more service stations,
2) service times of customers are independent and identically distributed (i.i.d.) random variables,
3) the arrival process is stationary and has the property of orderliness; that is, the probability that two or more customers arrive in sufficiently small interval of time is negligible when compared with the probability of one or no arrivals,

we will show in this paper that for any given queue discipline, the steady second moment of the queue size is equal to the second moment of the number of customers who arrive during a time interval D, which is distributed as the stationary waiting time in a queuing system which is equivalent to the above queuing system in every respect except its discipline which is first-in-first-out (FIFO).

We will use the following notations:

$A(t)$= the cumulative number of arrivals to time t,

$E[A^2(t)]$= second moment of $A(t)$,

$d(t)$= the cumulative number of departures from the queue to time t,

λ= a constant, the arival rate of customers,

$N(t)$= number of customers in the queue at time t.

m(t)= expected number of customers to arrive in the interval (0,t) given that a customer arrived at time zero,

D= waiting time of an arbitrary customer in the queue when the queue disciopline is FIFO,

H(.)= the stationary distribution of D, that is, $H(x)=Pr\{D\leq x\}$.

EFFECT OF QUEUE DISCIPLINE

Suppose we begin our observation of the system at time t=0, chosen as a time when the queue is empty. Let $0<T_1<T_2<...$ be the ordered arrival times of customer and $0<u_1<u_2<...$ be the ordered departure times of the customers from the queue. Clearly $u_1=T_1$. The ith customer may be delayed a time period D_i, The period during which he is waiting for other customers. Let S_i be his service time, i.e., the period during which he actually receives service. We will say he is in queue during his delay time, in service during his service time, and, in either case, in the system. Thus, the term system will refer abstractly to the facility and all customers being served or delayed. Now, let w_i be the waiting time of the ith customer in the system which is simply the sum of his delay (his waiting time in the queue) and his service time:

$$w_i = D_i + S_i.$$

With this notation, the ith customer leaves the queue at time $T_i + D_i$ and leaves the system at time $T_i + w_i$.

It is convenient to represent the evolution of the system by drawing graphs of the cumulative number of arrivals to time, t, A(t), and the cumulative number of departures from the queue to time t, d(t). These are both step functions which increase by one at each time of T_k and u_k, respectively, as shown in Figure 1. At any time t the number of customers in the queue is the vertical distance between the two curves, N(t)= A(t) - d(t).

If there is a queue of more than one customer, the order in which customers enter service need not be the same as the order in which arrive. If the departure time of the kth arriving customer from the queue is u_{n_k}, then this customer waits a time

$$D_k = u_{n_k} - T_k$$

in the queue. For FIFO discipline, $n_k = k$ and D_k is the horizontal distance from the curve A(t) to d(t) at height between k-1 and k. Clearly the waiting time distribution depends on the order of service. Now if all customers are identical (have exactly the same constant service time) it is reasonable to assume that the departure curve d(t) does not depend upon the order in which the customers are served. If the service times are i.i.d. random variables and d(t) is a random function of t, we can interpret this to mean that the stochastic properties of d(t) do not depend upon the order of service. This is turn implies that the distribution of N(t) does not depend upon the queue discipline [12]. Therefore to find the second moment of queue size for any queue discipline, we need only find it for the case of FIFO discipline.

Let

$E[N^2(t)]$ = second moment of N(t) the number of customers in queue at time, t, and

$E[A^2(D)]$=second moment of A(D) the number of customers who arrive in a time interval D, a random variable distributed as the stationary waiting time distribution when the queue discipline is FIFO.

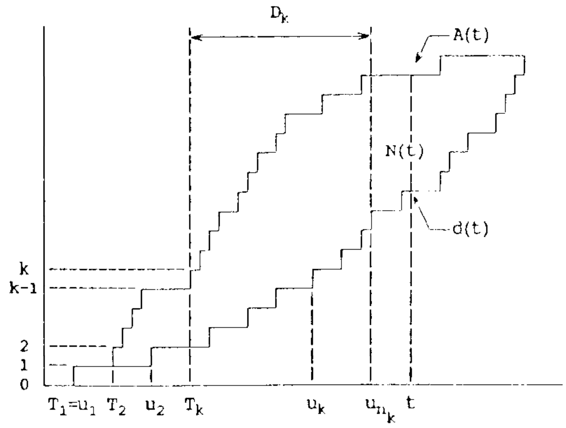

Figure 1. Cumulative number of customers

In what follows, we will prove the following theorem:

Theorem: For any queuing system, if

(i) the arrival process is stationary and has the property of orderliness

(ii) service times of customers are independent and identically distributed random variables,

(iii) for the case of FIFO discipline the waiting time of a customer is independent of the future arrival process,

then, the steady state second moment of the queue size for any queue discipline is equal to the second moment of the number of customers who arrive in a time interval D, a random variable which is distributed as the stationary waiting time for FIFO discipline. That is,

$$E[A^2(D)] = E[N^2(t)]. \qquad (1)$$

$E[A^2(t)]$

Let $P_i(h)$ be the probability of i arrivals in an interval of length h. Then the assumption that the probability of two or more arrivals in a sufficiently small interval of time is negligible when compared with probability of one (or no) arrival, may be defined by the condition [5,12]

$$1-P_0(h) - P_1(h)= o(h) \qquad \text{as } h \to 0 \qquad (2)$$

For a stationary arrival process [5,10]

$$\lim_{t \to 0} \frac{1-P_0(t)}{t} = \theta,$$

a constant, or equivalently

$$1- P_0(t)= \theta t + o(t) \qquad \text{as } t \to 0,$$

and for the orderliness of any stationary arrival process, (2) implies that $\theta= \lambda$, where λ is the arrival rate [5,10,12]. Divide the time axis arbitrary small intervals,

$$h_j = t_{j+1} - t_j,$$

as in Figure 2. Let

$$t_k > t_j, \qquad \text{for all } k > j.$$

and

$$C_t = \{ j: 0 < t_{j+1} \le t \}.$$

If

$$A_j = \text{number of customers who arrive in } h_j,$$

then,

$$E[A(t)] = E[\sum_{j\in C_t} A_j]$$

$$= \sum_{j\in C_t} E[A_j].$$

From (2),

$$A_j= 1 \text{ or } 0 \qquad \text{as } h_j \to 0. \qquad (3)$$

Thus,

$$E[A(t)]= \sum_{j\in C_t} P[A_j= 1]= \sum_{j\in C_t} \lambda h_j = \lambda t, \qquad (4)$$

which is true for any stationary arrival process [1,10,12].

To find the second moment of A(t) we can write

$$E[A^2(t)]= E[(\sum_{j\in C_t} A_j)^2]$$

$$= E [\sum_{j\in C_t} A_j^2] + 2\sum_j \sum_{k>j} E[A_j A_k]. \qquad j,k\in C_t$$

$$= \sum_{j\in C_t} E[A_j^2] + 2\sum_j \sum_{k>j} E[A_j A_k]. \qquad j,k\in C_t \qquad (5)$$

From (3), as $h_j \to 0$

$$E[A_j^2]= E[A_j]= P\{A_j=1\}= \lambda h_j,$$

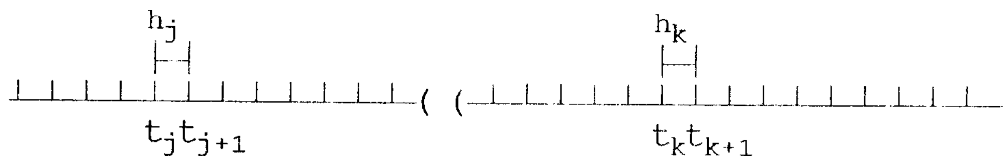

Figure. 2. Time axis divisions.

and from (4)

$$\sum_{j\in C_1} E[A_j^2]= \lambda t, \qquad (6)$$

Conditioning on A_j we can also write, as $h_j \to 0$

$$E[A_jA_k]= E[A_k|A_j= 1] P\{A_j= 1\}$$
$$= E[A_k|A_j= 1] \lambda h_j. \qquad (7)$$

From (5), (6), and (7) we can write

$$E[A^2(t)] = \lambda t + 2\sum_j \sum_{k>j} E[A_k| A_j= 1] \lambda h_j, \quad j,k\in C_1 \qquad (8)$$

By virtue of the additivity of mathematical expectation we can write

$$\sum_{k>j} E[A_k| A_j= 1]= m(t-t_j) \qquad k, j\in C_1 \qquad (9)$$

where

$m(t-t_j)$= expected number of arrivals in the interval $(t-t_j)$, given that one arrived at t_j.

In view of (9) we can write (8) as

$$E[A^2(t)]= \lambda t + 2 \sum_{j\in C_1} m(t-t_j) \lambda h_j$$

or equivalently

$$E[A^2(t)]= \lambda t + 2\lambda \int_0^{\infty} m(t-x)dx$$
$$= \lambda t + 2\lambda \int_0^{\infty} m(x)\, dx \qquad (10)$$

Now, let

A(D)= number of arrivals in a time interval D, a random variable with probability distribution function H(.).

Then, by conditioning on D, we have

$$E[A^2(D)]= E[E[A^2(D)|D]]$$
$$= \int_0^{\infty} E[A^2(D) |D= u]\, dH(u),$$

and from (10)

$$E[A^2(D)]= \int_0^{\infty} [\lambda u + 2\lambda \int_0^{\infty} m(x)dx]\, dH(u).$$

Integrating by part, we have

$$E[A^2(D)]= \lambda E(D) + 2\lambda \int_0^{\infty} m(u) [1- H(u)]\, du. \qquad (11)$$

$E[N^2(D)]$

Let

N_j= number of customers who arrived in h_j and are still in the queue at time t,

and $\qquad C_2= \{j: -\infty < t_{j+1} \le t\}$.

Then, we can write

$$N(t)= \sum_{j\in C_2} N_j,$$

and

$$E[N(t)]= E[\sum_{j\in C_2} N_j]$$
$$= \sum_{j\in C_2} E[N_j]. \qquad (12)$$

Clearly,

$$N_j= 1 \text{ or } 0 \quad \text{as } h_j \to 0. \qquad (13)$$

Thus we can write

$$E[N_j]= Pr \{N_j= 1\}$$
$$= Pr \{N_j= 1 | A_j= 1\} Pr \{A_j= 1\}$$
$$= Pr \{N_j= 1 | A_j= 1\} \lambda h_j.$$

But, given a customer arrived in h_j then he will be in the queue at time t if his waiting time D is greater than $t-t_j$. That is,

$$Pr\{N_j= 1 | A_j= 1\}= Pr\{D>t-t_j\}$$
$$= H^c(t-t_j),$$

where

$$H^c(t-t_j)= 1-H(t-t_j).$$

Thus,

$$E[N_j]= \lambda H^c(t-t_j) h_j \qquad (14)$$

and from (12)

$$E[N(t)]= \lambda \sum_{j\in C_2} H^c(t-t_j)h_j$$

or equivalently

$$E[N(t)]= \lambda \int_{-\infty}^{t} H^c(t-s)ds= \lambda \int_{0}^{t} H^c(u)du$$
$$= \lambda E[D]. \qquad (15)$$

This is the well-known Little's Formula [4,7,11] which holds for more general queueing systems than the one considered here.

To find $E[N^2(t)]$ we can write

$$E[N^2(t)]= E[(\sum_{j\in C_2} N_j)^2]= E[\sum_{j\in C_2} N_j^2 + 2\sum_{j}\sum_{k>j} N_jN_k] \quad j,k\in C_2$$
$$= \sum_{j\in C_2} E[N_j^2] + 2\sum_{j}\sum_{k>j} E[N_jN_k] \quad j,k\in C_2 \qquad (16)$$

From (13)

$$E[N_j^2]= \Pr\{N_j^2=1\}$$
$$= \Pr\{N_j= 1\}$$
$$= \lambda H^c(t-t_j)\, h_j, \qquad (17)$$

and

$$E[N_jN_k]= E[N_k| N_j=1] \Pr\{N_j= 1\}$$
$$= E[N_k| N_j=1]\, \lambda H^c(t-t_j)\, h_j. \qquad (18)$$

But,

$$E[N_k | N_j= 1]= E[A_k| N_j= 1].$$

The reason is that $N_j=1$; that is a customer arrived in h_j and is still in the queue, implies that the customer who arrived in h_k must be in the queue at time t, since $t_k>t_j$ and the queue discipline is FIFO.

Furthermore, we can write

$$E[A_k| N_j=1] = E[A_k|A_j= 1]. \qquad (18a)$$

To see this, we note that $N_j=1$ implies a customer arrived in h_j and his waiting time is greater than $t-t_j$. But, for the above queuing system with FIFO discipline the waiting time of a customer in the queue is independent of the number of arrivals A_k, who arrive after him. Thus we can write

$$E[N_kN_j]= E[A_k|A_j= 1]\, \lambda H^c(t-t_j)\, h_j. \qquad (19)$$

Substituting (17) and (19) in (16) we have

$$E(N^2(t))= \sum_{j\in C_2} \lambda H^c(t-t_j)h_j + 2\sum_{j}\sum_{k>j} E[A_k|A_j=1]\, \lambda H^c(t-t_j)\, h_j \quad j,k\in C_2$$
$$= \lambda \sum_{j\in C_2} H^c(t-t_j)+2\lambda \sum_{j\in C_2} H^c(t-t_j)\, h_j \sum_{k>j} E[A_k|A_j= 1]. \quad j,k\in C_2$$

From (9) we can write

$$E[N^2(t)]= \lambda \sum_{j\in C_2} H^c(t-t_j)\, h_j + 2\lambda \sum_{j\in C_2} H^c(t-t_j)\, m(t-t_j)h_j,$$

or equivalently

$$E[N^2(t)]= \lambda \int_{-\infty}^{t} H^c(t-u)du + 2\lambda \int_{-\infty}^{t} m(t-u)\, H^c(t-u)\, du,$$

or

$$E[N^2(t)]= \lambda \int_{0}^{\infty} H^c(u)du + 2\lambda \int_{0}^{\infty} m(u)\, H^c(u)\, du$$
$$= \lambda E[D] + 2\lambda \int_{0}^{\infty} m(u)\, H^c(u)\, du. \qquad (20)$$

In view of (11) we can write (20) as

$$E[N^2(t)]= E[A^2(D)],$$

which proves the stated theorem.

Remarks

(i) Since service times of customers are i.i.d. random variables, the distribution of queue size does not depend on the queue discipline. Thus, if we denote the queue size for FIFO and non-FIFO queue discipline, respectively, by N_F and N_{nF} then Equation 20 states that

$$E[N_{nF}^2]= E[N_F^2]= E[N^2(t)]= \lambda E[D] + \lambda \int_0^\infty m(u)\ H^c(u)\ du. \quad (21)$$

(ii) Furthermore, let

$H_{nf}(.)$= distribution of waiting time in queue for non-FIFO discipline.

Since the distribution of waiting depends on the queue discipline [12], Equation 20 does not hold if we replace H(.) by $H_{nf}(.)$ in Equation 20. This point can also be seen from the fact that Equation 18a does not hold for any queue discipline other than FIFO.

For Poisson arrival process at rate λ we have $m(u)=\lambda u$ Thus for Poisson arrival we can write (20) as:

$$E[N^2(t)]= \lambda E[D] + 2\lambda \int_0^\infty \lambda u\ H^c(u)\ du$$

$$= \lambda E[D] + \lambda^2 \int_0^\infty 2uH(u)\ du$$

$$= \lambda E[D] + \lambda^2 E(D^2) \quad (22)$$

Let σ_q= standard deviation of N(t), and σ_w= standard deviation of D, then from the fact that $E[N^2(t)]= \sigma_q^2 + (E[N(t)])^2$, $E[D^2]= \sigma_w^2 + (E[D])^2$, and $E[N(t)]= \lambda E[D]$ we can write (22) as

$$\sigma_q^2 = \lambda E[D] + \lambda^2\sigma_w^2. \quad (23)$$

An obvious application of 20 (or equally of Equation 1) is that the knowledge about the distribution of waiting time determines the variance of the queue size, a quantity that is of interest in the analysis and designing of queuing systems. Consider the application of "L=λW" (L= E[N], W= E[D]) where the knowledge about either L or W leads to the knowledge about the other quantity. As in the application of "L= λW", in the case of queues with Poisson arrivals, the knowledge about either σ_q or σ_w in (23) also leads to the knowledge about the other quantity.

Examples

In the following examples let

P_k = the steady state probability that n customers are in the system.

(a) (M/H2/1) model

Consider a single-channel queuing system in which the arrival process is Poisson with rate λ and the service times have the following hyper-exponential distribution function:

$$g(x)= dG(x)/dx= (1/4)\lambda e^{-\lambda x} + (3/4)(2\lambda)e^{-2\lambda x}, \quad x\geq 0. \quad (24)$$

The moments of this distribution are:

$$E[S]= 1/\mu= 5/(8\lambda),\ E[S^2]= 7/(8\lambda^2), \text{and}\ E[S^3]= 33/(16\lambda^3). \quad (25)$$

From (25) and Pollaczek-Khintchine formula [2] we have

$$E[D]= 7/(6\lambda) \quad \text{and} \quad E[N]= 7/6 \quad (26)$$

In general, for M/G/1 model [2] we have

$$E[D^2]= [\lambda/(1-\rho)]\ \{E[S^2]E[D] + (E[S^3]/3)\},\ (\rho= \lambda/\mu). \quad (27)$$

From (25) and (26), and the fact that ρ= 5/8, we can write (27) as

$$E[D^2]= 8\lambda/3\ \{(49/48\lambda^3) + (11/16\lambda^3)\}= 41/9\lambda^2 \quad (28)$$

Applying (20) to this queuing system we have (from (22) for Poisson arrivals)

$$E[N^2]= \lambda\ E[D] + \lambda^2\ E[D^2]$$

and from (26) and (28) we can write

$$E[N^2] = 103/18. \tag{29}$$

To see that this is in fact the value of the second moment of the queue size for this queuing system, we first note that for this queuing system (See [6] p. 196)

$$P_k = (3/32)(2/5)^k + (9/32)(2/3)^k.$$

Thus

$$E[N^2] = \sum_{k=1}^{\infty} (K-1)^2 P_k = 103/18$$

which is shown to be the same as (29), as was the intention.

(b) (G/M/1) model

Consider a single-channel queuing system in which the times between successive arrivals have an arbitrary distribution G and the service times are exponentially distributed with rate μ. For this queuing system [9]

$$Pr\{D>u\} = H^c(u) = \beta\, e^{-\mu(1-\beta)u}$$

where

$$\beta = \int_0^{\infty} e^{-\mu(1-\beta)u}\, dG(u). \tag{30}$$

Thus,

$$E[D] = \int_0^{\infty} H^c(u)\, du = \beta/u(1-\beta), \text{ and } E[N] = \lambda E[D] = \rho\beta/(1-\beta). \tag{31}$$

Let

$$s = \mu(1-\beta). \tag{32}$$

Then, we can write

$$H^c(u) = \beta\, e^{-su},\ u \geq 0, \tag{33}$$

and

$$\beta = G^*(s) \tag{34}$$

where $G^*(s)$ is the Laplace-Stieltjes transform (LST) of $G(t)$ [2]. That is,

$$G^*(s) = \int_0^{\infty} e^{-st}\, dG(t)$$

Now, applying Equation 20 to this queuing system we have

$$\begin{aligned} E[N^2] &= \lambda E[D] + 2\lambda \int_0^{\infty} m(u)\, H^c(u)\, du \\ &= \lambda E[D] + 2\lambda \int_0^{\infty} m(u)\, \beta\, e^{-su}\, du \\ &= \lambda E[D] + 2\lambda\, \beta\, m(s), \end{aligned} \tag{35}$$

where $m(s)$ is the Laplace transform (LT) of the renewal function $m(t)$, that is,

$$m(s) = \int_0^{\infty} e^{-st}\, m(t)\, dt$$

But, the relation between $m(s)$ and the LST of $m(t)$, $m^*(s)$, is [12]

$$m(s) = m^*(s)/s$$

Furthermore, the relation between $m^*(s)$ and $G^*(s)$ is [8]

$$m^*(s) = G^*(s)/[1-G^*(s)]$$

Thus, we can write (35) as

$$E(N^2) = \lambda E[D] + 2\lambda\beta\, \{G^*(s)/[1-G^*(s)]\}\, (1/s),\ s = \mu(1-\beta). \tag{36}$$

From (31), (32), and (34) we can write (36) as

$$E(N^2) = \frac{\rho\beta}{1-\beta} + \frac{2\lambda\beta^2}{(1-\beta)}\left[\frac{1}{\mu(1-\beta)}\right] = \frac{\rho(1+\beta)\beta}{(1-\beta)^2}. \tag{37}$$

To see that this is in fact true, we note that for this queuing system [9]

$$P_k = \rho\,(1-\beta)\,\beta^{k-1}, \qquad k \geq 1,\ (\rho = \lambda/u)$$

$$P_0 = 1-\rho.$$

Thus we can write

$$E[N^2] = \sum_{k=1}^{\infty} (K-1)^2 P_k = \frac{\rho(1+\beta)\beta}{(1-\beta)^2}$$

which is shown to be the same as (37), as was the intention.

REFERENCES

1. F.G. Beutler and Q.A.Z. Leneman: The Theory of Stationary Point Processes, *Acta Mathematica*, Vol. 116, pp. 159-197 (1966).
2. D. Gross and C.M. Harris: "*Fundamentals of Queuing Theory*," 2nd ed., Wiley, New York (1985).
3. R. Haji,: Derivation of Second Moment of the Queue Size, presented in the International Conference on Recent Developments in Statistical Data Analysis and Inference, Neuchatel, Switzerland, August (1989).
4. A.Y. Khintchine: "*Mathematical Methods in the Theory of Queuing*," Griffin, London (1960).
5. W.S. Jewell: A Simple Proof of L= λW, *Operations Research* 15, pp. 1109-1116 (1967).
6. L. Kleinrock: "*Queueing Systems*." Vol. 1, Wiley, New York (1975).
7. J.D.C. Little: A Proof for the Queuing Formula: L= λW, *Operations Research*, Vol. 9, pp. 383-387 (1961).
8. E. Parzen: "*Stochastic Processes*," Holden-Day, San Francisco, Calif. (1962).
9. S.M. Ross: "*Introduction to Probability Models*," Academic Press, New York (1989).
10. S.M. Ross "*Stochastic Processes*," John Wiley, New York (1983).
11. S. Stidham Jr: A Last Word on L= λW, *Operations Research*, 22, pp. 417-421 (1974).