The Output of a Queuing System
Author(s): Paul J. Burke
Source: *Operations Research*, Vol. 4, No. 6 (Dec., 1956), pp. 699-704
Published by: INFORMS
Stable URL: http://www.jstor.org/stable/166919
Accessed: 26-02-2016 15:53 UTC

THE OUTPUT OF A QUEUING SYSTEM

PAUL J. BURKE
Bell Telephone Laboratories, New York, New York
(Received June 25, 1956)

For a queuing system with Poisson input, a single waiting line without defections, and identically distributed independent (negative) exponential service times, the equilibrium distribution of the number of service completions in an arbitrary time interval is shown to be the same as the input distribution, for any number of servers. This result has applications in problems of tandem queuing. The essence of the proof is the demonstration of the independence of an interdeparture interval and the state of the system at the end of the interval.

THE PROBLEM OF the output or 'efflux,' as termed by MORSE,[1] of a queuing system apparently has been previously considered, but not investigated, in the literature. Thus Morse in the reference cited states, "A little thought will convince one that the efflux from a single-channel, exponential service channel, fed by Poisson arrivals, must be Poisson with the same rate as the arrivals . . . ."

In similar vein, for the case of two gates in series with arrivals and service times random at the first gate, O'BRIEN has stated, "The arrival of customers at gate 2 will be random and the average arrival rate will be exactly the same as that for gate 1 . . . ." [2]

Neither statement is supported by analysis. Since the truth of these statements is far from obvious, a proof seems necessary. It is the main purpose of this paper to supply that proof and, although the quoted statements apply only to single-server queues, to generalize the result to multiserver queues.*

The motivation for studying the output distribution arises chiefly in problems of tandem queuing, which occur in a variety of applications. One example is that of customers in a store who must first be waited on by sales clerks and then, after being served by these clerks, must then be served by wrappers or cashiers. Another, more complicated example, is the setting up of a telephone call through a switching system. One type of the latter problem has been studied by V. E. BENES in an unpublished memorandum. This problem of Benes' is concerned with a two-stage queue, the first or register stage having random input, exponential service times, and an arbitrary number of servers, while the second or marker stage has one server and general service time. The principal result is that the equilibrium distribution of queue length and elapsed service time in the second

* A different method of proof which includes the present result as a special case was developed, after the completion of this paper, by DR. EDGAR REICH of the Rand Corporation. Dr. Reich's method will be published in the *Annals of Mathematical Statistics*.

stage is the same as though the input to the second stage were random. In Benes' words, "If we had considered the marker system by itself with Poisson input at rate μ calls per second, we would have obtained precisely the equation . . . for the generating function of the number of people present and the age of the marker." Thus if it had been known that the output from the first stage was in fact random, the result of this paper would have followed from well-known theorems with a considerable saving in analysis.

Still another example of tandem queuing is analyzed in a paper by R. R. P. JACKSON.[3] In one case of Jackson's problem, random input to the first stage is assumed; and each stage has a single server with exponential service time. Again, certain of Jackson's results follow immediately, if it is known that the output of each stage has the same Poisson distribution as the original input, and, furthermore, these results can be generalized to any number of stages.

It is intuitively clear that, in tandem queuing processes of the type mentioned above, if the output distribution of each stage was of such character that the queuing system formed by the second stage was amenable to analysis, then the tandem queue could be analyzed stage-by-stage insofar as the separate delay and queue-length distributions are concerned. Such a stage-by-stage analysis can be expected to be considerably simpler than the simultaneous analysis heretofore necessary. Fortunately, under the conditions stated below, it is true that the output has the required simplicity for treating each stage individually.

THE MODEL

A SUMMARY statement of the theorem proved in this paper is that the output of a queuing system with Poisson input and exponential holding times is again Poisson. The details of the hypothesis are as follows. A single-stage queue with random input is assumed. The average interarrival interval has length $1/\lambda$. That is, the probability of the arrival of a call (customer) during an interval of length dt is taken equal to $\lambda\, dt$, within infinitesimals of higher order, independent of the state of the system, arrival times of previous calls, or any other conditions whatever.

There are s servers (channels) each having an exponential holding time distribution with average $1/\mu$. The holding times are completely independent of all conditions. Hence the probability that a call which is receiving service at the beginning of an interval dt will terminate during the interval is $\mu\, dt$ within infinitesimals of higher order.

Under these assumptions and the further condition $s\mu>\lambda$, it is well known that there is an equilibrium distribution of the states (number of calls in the system). Furthermore this distribution is the same as that of the states encountered by calls entering the system.

All calls remain in the system until they have received service. Otherwise, the queue discipline, or order of service, is irrelevant, since the output and not the delay distribution is of interest.

OUTLINE OF PROOF

IN ORDER TO SHOW that the equilibrium distribution of the number of calls completing service during an arbitrary interval of length T is Poisson with parameter λT under the conditions of the model, an equivalent result, that the time intervals between successive call completions are independently distributed with the same exponential distribution as the time intervals between arrivals, will be obtained.

Owing to the randomness of the input and to the exponential holding-time distribution, the output process is Markoffian with respect to the state of the system, i.e., given the state of the system at any time t, no further knowledge concerning the output distribution subsequent to t is gained from the previous history of the system. It will be shown that an interdeparture interval and the state of the system at the end of the interval are independent at equilibrium. Together with the Markoffian property, this independence implies that all interdeparture intervals are independent. It will be shown simultaneously that the equilibrium distribution of an interdeparture interval is exponential, and hence it follows that the output distribution, or distribution of call completions, is Poisson.

PROOF

IT SHOULD BE noted first that the probability of the system being in some state k immediately after a call departs is the same as the probability that an arriving call will find the system in state k. A departing call that leaves the system in state k represents a transition from state $k+1$ to state k, while an arriving call's finding the system in state k is a transition from state k to state $k+1$. The number of transitions from state k to state $k+1$ cannot differ by more than one from those from state $k+1$ to state k in any time interval. Hence the proportion of calls leaving the system in state k approaches the same limit as the proportion of calls finding the system in state k. The latter limit is known to be equal to the equilibrium probability of the system being in state k at an arbitrary instant. These equilibrium probabilities comprise the distribution,[4] which in the present notation may be written

$$p_k = p_0\,(\lambda/\mu)^k/k!, \qquad (0 \leqq k < s)$$

$$p_k = p_0\,(\lambda/\mu)^k/(s!\,s^{k-s}), \qquad (k \geqq s)$$

where p_0 is determined by the requirement

$$\textstyle\sum_{k=0}^{k=\infty} p_k = 1.$$

Let L denote the length of an arbitrary interdeparture interval and

$n(t)$ the state at a time t after the last previous departure. Let $F_k(t)$ be the probability that $n(t)=k$ and jointly that $L>t$. It may be helpful to note that

$$\textstyle\sum_{k=0}^{k=\infty} F_k(t)=F(t)$$

is the falling distribution of the length of an interdeparture interval and

$$F_k(0)=p_k$$

at equilibrium.

For an infinitesimal interval of length dt,

$$F_0(t+dt)=F_0(t)\ (1-\lambda\ dt),$$

within infinitesimals of higher order, since $L>t+dt$ if and only if $L>t$ and no calls arrive during dt. Similarly,

$$F_k(t+dt)=F_k(t)\ (1-\lambda\ dt-j\mu\ dt)+F_{k-1}(t)\ \lambda\ dt,$$

where $j=k$ for $k<s$ and $j=s$ for $k\geqq s$. These equations reduce to the differential equations

$$\begin{aligned} F_0'(t)&=-\lambda\ F_0(t),\\ F_k'(t)&=\lambda\ F_{k-1}(t)-(\lambda+j\mu)\ F_k(t), \end{aligned} \tag{1}$$

subject to the initial conditions (which imply the existence of equilibrium)

$$F_k(0)=p_k.$$

Equations (1) can be solved by induction to yield

$$F_k(t)=p_k\ e^{-\lambda t} \tag{2}$$

as the unique solutions subject to the initial conditions. This result implies that the marginal distribution of the interdeparture intervals is exponential with parameter λ, i.e., the same as interarrival distribution. Also, the independence of L and $n(L)$ can be readily established from it. The probability that

$$t+dt>L>t \quad\text{and}\quad n(L+0)=k$$

is

$$F_{k+1}(t)\ (k+1)\ \mu\ dt \qquad \text{for} \quad k+1\leqq s$$

and

$$F_{k+1}(t)\ s\ \mu\ dt \qquad \text{for} \quad k+1>s.$$

But these expressions reduce to

$$(1/k!)\ (\lambda/\mu)^k\ p_0\ e^{-\lambda t}\ \lambda\ dt$$

and

$$[1/(s!\ s^{k-s})]\ (\lambda/\mu)^k\ p_0\ e^{-\lambda t}\ \lambda\ dt$$

respectively, which are factored into the marginal probability functions of $n(L)$ and L, thus proving the independence of L and $n(L)$. The independence of the length of an arbitrary interval and all subsequent intervals follows from the last result together with the Markoff property, as is shown by the formal argument following.

Let Λ represent the set of lengths of an arbitrary number of interdeparture intervals subsequent to the arbitrary interval of length L, and let $P(—)$ represent the probability function of the chance variable(s) represented within the brackets. The Markoff property implies

$$P(\Lambda|n(L))=P(\Lambda|n(L),\, L) \tag{3}$$

where to avoid ambiguity $n(L)$ may be taken to mean $n(L+0)$. The independence of $n(L)$ and L is equivalent to

$$P(n(L),L)=P(n(L))\ P(L). \tag{4}$$

The joint probability function of the initial interval-length, the state at the end of the interval, and the set of subsequent interval-lengths may be written as

$$P(L,n(L),\Lambda)=P(\Lambda|L,n(L))\ P(L,n(L)). \tag{5}$$

Substituting (3) and (4) into (5), one has

$$\begin{aligned} P(L,n(L),\Lambda)&=P(\Lambda|n(L))\ P(n(L))\ P(L)\\ &=P(\Lambda,n(L))\ P(L). \end{aligned} \tag{6}$$

Whence from (6),

$$P(L,\Lambda)=\sum_{n(L)=0}^{n(L)=\infty} P(\Lambda,n(L))\ P(L)=P(L)\ P(\Lambda).$$

From this result follows the mutual independence of all intervals, which concludes the proof.

It may be remarked that for the equilibrium output to be Poisson uniformly for all values of the parameter μ, the input (assumed to have a fixed average) must be Poisson. This follows from the fact that the output distribution may be made to approximate the input by allowing μ to become infinite.

AN EXAMPLE

As AN ILLUSTRATION of the application of the result of this paper, an idealization of the situation involving sales clerks and cashiers mentioned above will be considered. It will be assumed that customers have access to any of the clerks who may be free and that the clerks have equal access to all the merchandise. After service by a clerk, a customer proceeds to cashiers and has access to all of them regardless of the type of purchase. Also, there is a single queue in front of the clerks and another single queue in front of the cashiers. Service by the clerks and by the cashiers is order-of-arrival.

It will be further assumed that evidence exists that the service times of clerks may be satisfactorily approximated by an exponential distribution with average 1.5 minutes, while the service times of cashiers are almost constant at 1 minute. The problem is to determine the numbers of clerks and cashiers necessary so that the probabilities of a customer's being delayed more than three minutes in front of the clerks or more than two minutes in front of the cashiers are each less than 0.05 for a period of several hours during which customer arrivals are random (Poisson) with an average of two per minute.

Since the service times of the clerks are exponential, it may be inferred from the theorem of this paper that the input to the cashiers is Poisson with an average of two per minute regardless of the number of clerks, provided only that these exceed three, and hence the number of cashiers can be determined independently of the number of clerks.

The required number of clerks can be found with the aid of plotted delay distributions for exponential holding times.[5] For five clerks, the occupancy is 0.60, since arrivals average three per service time; and the probability of delay greater than two service times (three minutes) is found graphically to be 0.0047, which more than meets the criterion. For four clerks, however, the occupancy is 0.75 and the probability of delay greater than two service times is 0.070, which does not meet the criterion. Hence five clerks are necessary.

Similarly, to determine the number of cashiers, plotted delay distributions for constant holding times may be used.[6] If there are three cashiers, the occupancy will be 0.67, and the probability of delay greater than two service times will be less than 0.01. Hence five clerks and three cashiers will be required to meet the service criteria.

ACKNOWLEDGMENTS

THANKS are due to W. S. HAYWARD, JR. for suggesting the method of differential equations, to E. REICH for criticism and correction of part of the proof, and to R. H. DAVIS for helpful discussions.

REFERENCES

1. PHILIP M. MORSE, "Stochastic Properties of Waiting Lines," *Opns. Res.* **3**, 256 (1955).
2. GEORGE G. O'BRIEN, "Some Queuing Problems," *J. Soc. Indust. Appl. Math.* **2**, 134 (1954).
3. R. R. P. JACKSON, "Queueing Systems with Phase Type Service," *Opnl. Res. Quart.* **5**, 109–120 (1954).
4. T. C. FRY, *Probability and Its Engineering Uses*, 382, Van Nostrand, New York, 1928.
5. E. C. MOLINA, "Application of the Theory of Probability to Telephone Trunking Problems," *Bell System Tech. J.* **6**, 461–494 (1927).
6. C. D. CROMMELIN, "Delay Probability Formulae," *Post Office Elec. Eng. J.* **26**, 266–274 (1934).