COMMUN. STATIST.-STOCHASTIC MODELS, 1(2), 257-262 (1985)

WAITING TIMES IN THE NON-PREEMPTIVE PRIORITY M/M/c QUEUE

Offer Kella and Uri Yechiali

Department of Statistics
Tel-Aviv University, Tel-Aviv 69978, Israel

Key Words and Phrases: waiting times; M/M/c queue; non-preemptive priority.

ABSTRACT

By using a probabilistic equivalence between the M/G/1 queue with multiple server's vacations and the M/M/c system, we derive the Laplace-Stieltjes transform of the waiting time W_k of a class-k customer in the non-preemptive priority M/M/c queue where all customers have the same mean service time. We also calculate the first two moments of W_k.

1. INTRODUCTION

Consider a non-preemptive M/M/c queueing system with n priority classes. Customers of class i arrive according to a Poisson process with rate λ_i, $1 \leq i \leq n$. Service times for <u>all</u> customers are exponentially distributed with parameter μ.

0822-0287/85/0102-0257$3.50/0

Customers of class i have non-preemptive priority over customers of class j whenever $i < j$, and service within each class follows the FCFS rule. Our analysis is based on the observation that an arriving customer waits for service if and only if all servers are busy, and on a probabilistic equivalence between the waiting time of an arbitrary customer in an M/G/1 queue with multiple server's vacations (Model 2 of Levy and Yechiali (1975)) and that of a class-k customer in the non-preemptive M/M/c system <u>given that all servers are busy.</u>

2. THE LST OF W_k

From the point of view of class-k customers, classes $1,2,\ldots,k-1$ and classes $k+1,\ldots,n-1,n$ may be grouped into two distinct classes denoted by class-a and class-b, respectively. Hence, let

$$\lambda_a = \sum_{i<k} \lambda_i, \quad \lambda = \sum_{i=1}^{n} \lambda_i, \quad \rho_i = \lambda_i(c\mu)^{-1}, \quad \rho_a = \lambda_a(c\mu)^{-1},$$

$$\sigma_j = \sum_{i=1}^{j} \rho_i, \quad \rho = \sigma_n = \lambda(c\mu)^{-1}.$$

Now, let γ be the length of time from an instant when all servers are busy, an arbitrary customer enters service and there are no class-a customers in queue, until the first moment thereafter that the number of busy servers decreases to $c-1$ or a customer from one of the classes $k,k+1,\ldots,n$ enters service. As all service times are exponential with parameter μ, the interdeparture times within the time interval γ are exponentially distributed with parameter $c\mu$. Hence, γ may be viewed as the busy period of an M/M/1 queue with arrival rate λ_a and service rate $c\mu$.

It is well known (Kleinrock, p. 215) that the Laplace-Stieltjes transform (LST) of γ is given by

$$\tilde{\gamma}(s) \equiv E[e^{-s\gamma}] = \left\{ s+\lambda_a+c\mu-[(s+\lambda_a+c\mu)^2 - 4\lambda_a c\mu]^{\frac{1}{2}} \right\} (2\lambda_a)^{-1}. \qquad (1)$$

In addition, a realization of γ which starts with a service of a class-k customer may be viewed as a (generalized) service time in an M/G/1 queue serving class-k customers only. Alternatively, a realization of γ which starts either with service of a class-b customer or by any other customer causing the number of occupied servers to rise from c-1 to c, may be interpreted as a <u>vacation time</u> in an M/G/1 queue with multiple server's vacations where only customers of class-k type are considered as actual customers.

As a consequence of these observations, <u>given that all servers are busy,</u> the waiting time of a class-k customer is probabilistically equivalent to that of an arbitrary customer in an M/G/1 queue with multiple server's vacations in which the arrival rate is λ_k and <u>both</u> the service and vacation times are distributed like γ. Using these facts we state

<u>Theorem</u>. The LST of the waiting time W_k of a class-k customer

in a non-preemptive M/M/c queue with the same service rate for all classes is given by

$$\tilde{W}_k(s) = (1-\pi) + \pi \frac{c\mu(1-\sigma_k)\ (1-\tilde{\gamma}(s))}{s-\lambda_k+\lambda_k\tilde{\gamma}(s)}, \tag{2}$$

where

$$\pi = \frac{(\lambda/\mu)^c}{c!(1-\rho)}\left[\sum_{i=0}^{c-1}\frac{(\lambda/\mu)^i}{i!} + \frac{(\lambda/\mu)^c}{c!(1-\rho)}\right]^{-1},$$

is the probability that all servers are busy.

Proof: Levy and Yechiali (1975) showed that for an M/G/1 queue with multiple server's vacations, arrival rate λ, service times V and vacation lengths U, the LST of the waiting time W of an arbitrary customer is given by

$$\tilde{W}(s) = \frac{(1-\lambda EV)(1-\tilde{U}(s))}{EU[s-\lambda(1-\tilde{V}(s))]}, \tag{3}$$

where $\tilde{U}(S)$ and $\tilde{V}(S)$ denote the LST of U and V, respectively. Applying this result to the present model it readily follows that

$$E[e^{-sW_k}|\text{all servers are busy}] = \frac{(1-\lambda_k E\gamma)(1-\tilde{\gamma}(s))}{E\gamma[s-\lambda_k+\lambda_k\tilde{\gamma}(s)]}. \tag{4}$$

From (1) we obtain

$$E\gamma = [c\mu(1-\rho_a)]^{-1} = [c\mu(1-\sigma_{k-1})]^{-1}. \tag{5}$$

By substituting (5) into (4) the proof is complete.

Since $\tilde{\gamma}(s)$ is the solution of the equation $\tilde{\gamma}(s) = c\mu[s+\lambda_a - \lambda_a\tilde{\gamma}(s)+c\mu]^{-1}$ we redily have that

$$\tilde{\gamma}(s)[s-\lambda_k(1-\tilde{\gamma}(s))] = (1-\tilde{\gamma}(s))\ c\mu[1-\sigma_k\tilde{\gamma}(s)]\ . \qquad (6)$$

Substituting (6) in (2) yields.

$$\tilde{W}_k(s) = (1-\pi) + \pi\ \frac{(1-\sigma_k)\tilde{\gamma}(s)}{1-\sigma_k\tilde{\gamma}(s)}\ . \qquad (7)$$

Equation (7) has been obtained by Davis (1966) in a more elaborate way.

3. MOMENTS OF W_k

From equation (1) we derive

$$E\gamma^2 = 2(c\mu)^{-2}(1-\sigma_{k-1})^{-3}\ . \qquad (8)$$

By differentiating (2) or (7) and applying (5) and (8) we get, after some calculations, that

$$EW_k = \pi[c\mu(1-\sigma_k)(1-\sigma_{k-1})]^{-1}\ , \qquad (9)$$

and

$$EW_k^2 = 2\pi(1-\sigma_k\sigma_{k-1})[(c\mu)^2(1-\sigma_k)^2(1-\sigma_{k-1})^3]^{-1}. \qquad (10)$$

Result (9) has been derived by Cobham (1954) using expected value arguments.

BIBLIOGRAPHY

Cobham, A. (1954). Priority Assignment in Waiting Line Problems, Opns. Res., 2, 70-76.

Davis, R.H. (1966). Waiting-Time Distribution of a Multi-Server, Priority Queueing System, Opns. Res., 14, 133-136.

Kleinrock, L. (1975). Queueing Systems, Vol. I: Theory, Wiley.

Levy, Y. and Yechiali, U. (1975). Utilization of Idle Time in an M/G/1 Queueing System, Mgmt. Sci., 22, 202-211.

Received 6/28/84, Revised 11/13/84

Recommended by Y. Takahashi, Editor