A PRIORITY PROBLEM IN QUEUING THEORY

by

N. K. Jaiswal and H. C. Jain

Defence Science Laboratory, Delhi

ABSTRACT

A new type of priority is defined in which the priority unit has a right of interrupting the service of a nonpriority unit but only after the completion of the particular phase in progress, it being assumed that a unit demands a particular number of phases according to some probability distribution. The effect of such priority has been investigated. A particular case of this sort of priority corresponds to the head-of-line priority assignment discussed by Cobham.

Introduction

In many queuing problems, a customer demands a certain number of service-phases to be performed upon him. A problem of this type was discussed by Gaver[1] assuming the queue discipline to be "first come, first served". But in many practical situations, priority is imposed on certain customers either due to the higher payment or otherwise (for the same type of service) and therefore the investigation of the same problem with priorities appears to be more reasonable and practical.

For the sake of simplicity, only two priority classes have been considered. The priority and non-priority units arrive according to a Poisson distribution with mean rate λ_1 and λ_2 respectively. Let c_j and c_j' be the probabilities that a priority and a non-priority customer demands for j phases respectively. It is further assumed that a priority or a non-priority phase has a negative exponential service-time distribution with mean $1/\mu$. Obviously, the priority units are serviced prior to the non-priority ones but if there are no priority units, the service facility takes up a non-priority unit, if any and starts servicing on its phases. If a priority unit arrives at this stage, he has a right to interrupt the service operation on a non-priority unit only after the completion of the particular phase in progress. Since, in such a situation, neither the service on a non-priority is totally finished, nor it is at once interrupted, this constitutes a new type of assignment and is obviously more practicable. This sort of priority assignment is in between the head-of-the-line priority introduced by Cobham[2] and the preëmptive priority discussed by White & Christie[3]. However, by putting $c_j' = 1$, this problem reduces to that of Cobham for a wide class of service time distribution of a priority unit and an exponential distribution for the non-priority units.

Detailed Balance Equations

Let $P_{1,m,n,}$ and $P_{2,m,n,}$ denote the probability that a priority phase and a non-priority phase is in service, m priority and n non-priority phases are present in the system, including the phase in service. Assuming $P_{1,o,n,} = o$, $n \geqslant 1$, $P_{2,m,o,} = o, m \geqslant 1$ and $P_{1,o,o,} = P_{2,o,o,} = P_o$, the equations of detailed balance are

$$\lambda_1 \sum_{j=1}^{m} c_j P_{1,m-j,n} + \lambda_2 \sum_{j=1}^{n} c'_j P_{1,m,n-j} + \mu P_{2,m,n+1}$$

$$+ \mu P_{1,m+1,n} - (\lambda_1 + \lambda_2 + \mu) P_{1,m,n} = 0 \quad \overline{m \geqslant 1, n \geqslant 1} \qquad (1)$$

$$\lambda_1 \sum_{j=1}^{m} c_j P_{2,m-j,n} + \lambda_2 \sum_{j=1}^{n} c'_j P_{2,m,n-j}$$

$$-(\lambda_1 + \lambda_2 + \mu) P_{2,m,n} = 0 \quad \overline{m \geqslant 1, n \geqslant 1} \quad \ldots \quad \ldots \quad (2)$$

$$\mu P_{1,m+1,0} + \lambda_1 \sum_{j=1}^{m} c_j P_{1,m-j,o} + \mu P_{2,m,1}$$

$$-(\lambda_1 + \lambda_2 + \mu) P_{1,m,n} = 0 \quad \underline{m \geqslant 1} \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (3)$$

$$\mu P_{2,o,n+1} + \lambda_2 \sum_{j=1}^{n} c_j P_{2,o,n-j} + \mu P_{1,1,n}$$

$$-(\lambda_1 + \lambda_2 + \mu) P_{2,o,n} = o \quad n \geqslant 1 \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (4)$$

$$\mu P_{1,1,o} + \mu P_{2,o,1} - (\lambda_1 + \lambda_2) P_o = 0 \quad \ldots \quad \ldots \quad \ldots \quad (5)$$

Solution

Let us define the following generating functions.

$$F_{1,m}(y) = \sum_{n=0}^{\infty} y^n P_{1,m,n} \qquad F_{2,m}(y) = \sum_{n=1}^{\infty} y^n P_{2,m,n}$$

$$H_1(x, y) = \sum_{m=1}^{\infty} x^m F_{1,m}(y) \qquad H_2(x, y) = \sum_{m=0}^{\infty} x^m F_{2,m}(y)$$

$$H(x, y) = H_1(x, y) + H_2(x, y) + P_o$$

And let,

$$A(x) = \sum_{j=1}^{\infty} c_j x^j \text{ and } B(y) = \sum_{j=1}^{\infty} c'_j y^j$$

Multiplying the equations, (1) to (4) by the appropriate powers of x and y and adding, we finally obtain,

$$H_1(x, y)\left\{\lambda_1+\lambda_2+\mu-\lambda_1 A(x)-\lambda_2 B(y)-\frac{\mu}{x}\right\}$$

$$=\frac{\mu}{y} H_2(x, y)+\lambda_1 P_o A(x)-\mu\left[F_{11}(y)+\frac{1}{y} F_{2,o}(y)\right] \quad \ldots(6)$$

$$H_2(x, y)\left[\lambda_1+\lambda_2+\mu-\lambda_1 A(x)-\lambda_2 B(y)\right]$$

$$=\frac{\mu}{y} F_{2,o}(y)+\mu F_{11}(y)-P_o\left[\lambda_1+\lambda_2-\lambda_2 B(y)\right] \quad \ldots\ldots\ldots(7)$$

In order to find a relation between $F_{11}(y)$ and $F_{2,o}(y)$, we multiply equation (4) by y^n and sum over n from one to infinity. We thus, get a relation,

$$\mu F_{11}(y)=\left[\lambda_1+\lambda_2+\mu-\frac{\mu}{y}-\lambda_2 B(y)\right] F_{2,o}(y)$$

$$+P_o\left[\lambda_1+\lambda_2-\lambda_2 B(y)\right] \quad \ldots \quad (8)$$

Eliminating $F_{11}(y)$ from (7) and (8) we get

$$H_2(x, y)\left[\lambda_1+\lambda_2+\mu-\lambda_1 A(x)-\lambda_2 B(y)\right]$$

$$=F_{2,o}(y)\left[\lambda_1+\lambda_2+\mu-\lambda_2 B(y)\right] \quad \ldots \quad (9)$$

and then from (6) and (9) we get

$$H_1(x,y)=$$

$$\frac{-F_{2,o}(y)\left[\lambda_1+\lambda_2+\mu-\lambda_2 B(y)\right]\left[-\frac{\mu}{y}+\lambda_1+\lambda_2+\mu-\lambda_1 A(x)-\lambda_2 B(y)\right]-P_o Z}{Z\left[Z-\frac{\mu}{x}\right]} \quad \ldots \quad (10)$$

and

$$H_2(x, y)=\frac{F_{2,o}(y)\left[\lambda_1+\lambda_2+\mu-\lambda_2 B(y)\right]}{Z} \quad \ldots \quad (11)$$

where $Z=\lambda_1+\lambda_2+\mu-\lambda_1 A(x)-\lambda_2 B(y)$

therefore

$$H(x, y)=\frac{P_o \mu(1-x)}{\mu-x Z}$$

$$+\frac{F_{2,o}(y)\left[\lambda_1+\lambda_2+\mu-\lambda_2 B(y)\right] \mu(y-x)}{y Z(\mu-x Z)}$$

The mean number of priority phases in the system

$$\frac{\partial}{\partial x} H(x, y) \Big/_{\substack{x=1 \\ y=1}}$$

This expression involves $F_{2,0}(1)$. To evaluate it, we find from (11)

$$H_2(1, 1) = F_{2,0}(1) \frac{(\lambda_1 + \mu)}{\mu}$$

$$\therefore F_{2,0}(1) = \frac{\mu}{\mu + {}_1} H_2(1,1)$$

Also, since the mean service rate is the same for both type of units, the mean fraction of the time, the service channel is working on non-priority phases is proportional to the fraction of the time, the channel is busy and to the fraction of the non-priority phases which arrive and therefore

$$H_2(1, 1) = \sum_{m=0}^{\infty} \sum_{n=1}^{\infty} P_{2, m, n} =$$

$$\left[\frac{\lambda_1 \sum_{j=1}^{\infty} jc_j + \lambda_2 \sum_{j=1}^{\infty} jc_j'}{\mu} \right] \left[\frac{\lambda_2 \sum_{j=1}^{\infty} j c_j'}{\lambda_1 \sum_{j=1}^{\infty} jc_j + \lambda_2 \sum_{j=1}^{\infty} jc_j'} \right]$$

$$H_2(1,1) = \frac{\lambda_2}{\mu} \sum_{j=1}^{\infty} jc'_j \qquad \therefore F_{2,0}(1) = \frac{\lambda_2}{\lambda_1 + \mu} \sum_{j=1}^{\infty} jc_j'$$

Putting this value in (12), we get

L_1 = The mean number of priority phases in the system

$$= \frac{\mu \lambda_1 \frac{(P + Q)}{2} + \lambda_1 \lambda_2 PR}{\mu (\mu - \lambda_1 P)}$$

$$\text{where } P = \sum_{j=1}^{\infty} jc_j, \quad Q = \sum_{j=1}^{\infty} j^2 c_j \text{ and } R = \sum_{j=1}^{\infty} jc_j'$$

If no priority is imposed *i.e.* $\lambda_2 = 0$

$$L_1 = \frac{\lambda_1 \frac{P + Q}{2}}{\mu - \lambda_1 P}$$

If no priority is assumed in our model, *i.e.*, the units are serviced according to the "first come first served" discipline the mean queue length

$$L=\frac{\left[\lambda_1\sum_{j=1}^{\infty} jc_j + \lambda_2\sum_{j=1}^{\infty} jc_j'\right]+\frac{1}{2}\left[\lambda_1\sum_{j=1}^{\infty} j(j-1)c_j + \lambda_2\sum_{j=1}^{\infty} j(j-1)c_j'\right]}{\mu-\lambda_1\sum_{j=1}^{\infty} jc_j - \lambda_2\sum_{j=1}^{\infty} jc_j'}$$

$$L=\frac{\lambda_1\frac{(P+Q)}{2}+\lambda_2\frac{(R+S)}{2}}{\mu-\lambda_1 P-\lambda_2 R} \qquad \text{where } S=\sum_{j=1}^{\infty} j^2 c_j'$$

Hence mean number of non-priority phases in the system

$$=L_2=L-L_1=\frac{\lambda_1\frac{P+Q}{2}+\lambda_2\frac{R+S}{2}}{\mu-\lambda_1 P-\lambda_2 R}-\frac{\mu\lambda_1\frac{P+Q}{2}+\lambda_1\lambda_2 PR}{\mu(\mu-\lambda_1 P)}$$

Waiting Time Distribution

If $W(\xi, t)\,d\xi$, is the probability that a priority unit wait for a time between ξ and $\xi + d\xi$, then

$$W(\xi,t)\,d\xi=\sum_{n=o}^{\infty}\sum_{m=1}^{\infty} P_{1,m,n}(t)\frac{(\mu\xi)^{m-1}}{\underline{m-1}}\exp(-\mu\xi)\mu\, d\xi$$

$$+\sum_{n=1}^{\infty}\sum_{m=o}^{\infty} P_{,2m,n}\frac{(\mu\xi)^{m}}{\underline{|m}}\exp(-\mu\xi)\mu d\xi + P_o\,\delta_o,\xi$$

The Laplace transform of the waiting time distribution is given by

$$W(s,t)=\sum_{n=o}^{\infty}\sum_{m=1}^{\infty} P_{1,m,n}(t)\left(\frac{\mu}{\mu+s}\right)^{m}+\sum_{n=1}^{\infty}\sum_{m=o}^{\infty} P_{2,m,n}(t)\left(\frac{\mu}{\mu+s}\right)^{m+1}+P_o$$

$$=H_1(x,1)+x\,H_2(x,1)+P_o$$

$$=\frac{s}{s-\lambda_1\left[1-\sum_{j=1}^{\infty} c_j\left(\frac{\mu}{\mu+s}\right)^j\right]}\left[\frac{\mu-\lambda_1 P}{\mu}-\frac{s\lambda_2 R}{\mu(\mu+s)}\right]$$

Mean waiting time of a unit in the queue $=\left(-\frac{\partial w}{\partial s}\right)_{s=0}$

$$=\frac{\lambda_2 R+\lambda_1\frac{P+Q}{2}}{\mu(\mu-\lambda_1 P)}$$

If there is no priority unit *i.e.*, $\lambda_2 = o$

$$W(s,t)=\frac{s\left[1-\frac{\lambda_1}{\mu}\sum_{j=1}^{\infty} jc_j\right]}{s-\lambda_1\left[1-\sum_{j=1}^{\infty} c_j\left(\frac{\mu}{\mu+s}\right)^j\right]}$$

which is the same as that of Gaver [1]. Since the Laplace transform of the service-time distribution of a priority unit is obviously given by

$$H(s) = \text{Laplace transform of} \sum_{j=1}^{\infty} c_j \frac{(\mu\xi_o)^{j-1}}{\underline{|j-1}} exp\,(-\mu\xi)\,\mu d\xi$$

$$= \sum_{j=1}^{\infty} c_j \left(\frac{\mu}{\mu + s}\right)^j$$

We get

$$W\,(s,t) = \frac{s}{s - \lambda_1\,[\,1 - H\,(s)\,]} \left[\frac{\mu - \lambda_1 \sum\limits_{j=1}^{\infty} jc_j}{\mu} - \frac{s\lambda_2 \sum\limits_{j=1}^{\infty} jc_j'}{\mu\,(\mu + s)}\right]$$

This is the required relation between Laplace transform of the waiting time distribution of the priority unit and its servicing time distribution.

If $c_j = 1, j = 1$ and $c'_j = o, j \neq 1$, the non-priority unit brings only one phase and then our priority assignment becomes similar to that of Cobham[2]. In particular,

(1) If $c_j = 1, j = k$ and $c_j = o$, the service-time distribution of a priority unit is k — Erlang type and the Laplace transform of the waiting-time distribution becomes

$$W\,(s,\,t) = \frac{s}{s - \lambda_1 \left[\,1 - \left(\frac{\mu}{\mu + s}\right)^k\,\right]} \left[\frac{\mu - \lambda_1 k}{\mu} - \frac{s\lambda_2}{\mu\,(\mu + s)}\right]$$

Since

$$W\,(s,\,t) = \int_o^{\infty} e^{-s\xi}\, w(\xi,t)\, d\xi$$

$$= \int_o^{\infty} \left(1 - s\xi + \frac{s^2\xi^2}{\underline{|2}} \;\ldots\ldots\ldots\right) w\,(\xi,\,t)\;d\xi$$

$$= 1 - sM_1 + \frac{s^2}{\underline{|2}}\, M_2 \qquad \ldots\ldots\ldots$$

Where M_1, is the mean waiting time, M_2, M_3, etc. are the various moments of the distribution. These can be thus obtained, either by differentiating partially with respect to s and putting $s = o$, or by expanding $W\,(s,\,t)$ in ascending powers of s. Thus for the k-Erlang service time distribution,

$$M_1 = \frac{\lambda_2 + 1/2\ \{k\,(k + 1)\,\lambda_1\}}{\mu\,(\mu - \lambda_1 k)}$$

Which can be verified with the Cobham's formula for mean waiting time.

(2)

$k = 1,$

$$M_1 = \frac{\lambda_1 + \lambda_2}{\mu\,(\mu - \lambda_1)}$$

which is the same as due to Morse[4]. It is obvious, that the variance $\sigma^2 = M_2 - M_1{}^2$ can be easily calculated in any of the cases and other higher moments can be calculated.

Acknowledgements

We wish to express our thanks to Maj. H. S. Subba Rao for guiding us throughout the preparation of this paper. Thanks are also due to Shri T. R. Thiagarajan.

References

(1) Gaver, D. P., *J. Ops. Res. Soc. Am.* **2**, 139, 1954.

(2) Cobham, *J. Ops. Res. Soc. Am.* **2**, 70, 1954.

(3) White H. and Christie Lee G., *J. Ops. Res. Soc. Am.* **6**, 79, 1958.

(4) Morse, P. M., "*Queues, Inventories and Maintenance*", Wiley, New York, 1958.