OPTIMAL NUMBER OF SERVICE CHANNELS FOR A SERVICE LOSS SYSTEM

JOHN WALKER

Nanyang Business School

Nanyang Technological University

Singapore, 639798

Abstract

A simple graphical means of support in determining the optimal number (and associated sensitivity analysis) of service channels for a service loss system is presented. A service loss system is a multi-server system in which customers unable to receive service when the service channels are fully occupied are turned away or "lost". The graphical analysis indicates the optimal number of service channels as a function of the mean arrival rate per period x the mean customer service time and the ratio of the marginal cost of an extra service channel per period to the mean contribution margin to profit per customer served per period.

(SERVICE LOSS SYSTEMS, OPTIMAL SERVICE CHANNELS, GRAPHICAL ANALYSIS)

Introduction

Many service systems allow arriving customers to queue if the service channels are fully occupied when the customer enters the system. In contrast, service loss systems allow no queueing and so customers unable to receive service immediately are turned away or "lost". The diverse nature of service loss systems provides the motivation for this paper. Examples of such systems range from telephone switching systems and the stock control of expensive spare-parts to car parks and restaurants, etc. In this paper an application in determining the optimal number of items of equipment for a "rental" company to make available for hire is described.

The most widely used model of service loss systems is the M/G/s loss queueing model. The M/G/s loss queueing model considered is defined by the following assumptions.

- Customers arrive according to a Poisson probability distribution with a mean arrival rate of λ customers per period.
- The marginal cost of providing an extra service channel is C per channel per period.
- There are available s service channels.
- Customers arriving when all s service channels are occupied are denied service and are lost.
- The service times of successful customers are independent and identically distributed random variables with a mean of τ periods.
- The contribution margin to profit per customer served (revenue minus costs associated with the act of service) is P per period.

Let B(s,λτ) denote the probability that all s service channels are occupied (equivalently the proportion of customers per period denied service).

$$B(s,\lambda\tau) = (\lambda\tau)^s/s! \;/\; \sum_{k=0}^{s} (\lambda\tau)^k/k!, \qquad (1)$$

Direct numerical calculation of B(s,λτ) using formula (1) is difficult for large values of λτ and s. However, a fast and accurate computational scheme is given by Cooper (1981).

$$B(k,\lambda\tau) = \lambda\tau B(k-1,\lambda\tau) \,/\, [k + \lambda\tau B(k-1,\lambda\tau)],\ k \geq 1 \qquad (2)$$

$$B(0,\lambda\tau) = 1$$

Tijms (1986) states that the M/G/s loss model is one of the most useful models of queueing theory and discusses several applications, including telephone switching systems, stock control and reliability problems. The usefulness of the model lies in the applicability of the Poisson arrival process to many real world processes, and the insensitivity of formula (1) to changes in the distribution of service time (only the mean service time is required).

Minimising Mean Total Cost per Period

To determine the optimal number of service channels, the objective function of mean total cost per period, k(s), can be used. The mean total cost per period is considered to be the sum of the service channel provision cost per period plus the cost of the lost contribution margin to profit per period.

$$k(s) = Cs + \lambda B(s,\lambda\tau)P\tau$$

Dividing through by P results in a scaled mean total cost per period K(s),

$$K(s) = [C/P]s + \lambda\tau B(s,\lambda\tau)$$

Since $B(s,\lambda\tau)$ is convex, Messerli (1972), K is convex and the condition that s minimises K is that $\nabla K(s) \geq 0$ and $\nabla K(s+1) \leq 0$ where

$$\begin{aligned} \nabla K(s) &= K(s-1) - K(s) && \text{if } s > 0 \\ &= \infty && \text{if } s = 0 \end{aligned}$$

Defining $\nabla B(s,\lambda\tau)$ in a similar fashion,

$$\begin{aligned} \nabla K(s) &= \lambda\tau\nabla B(s,\lambda\tau) - C/P \\ \nabla K(s+1) &= \lambda\tau\nabla B(s+1,\lambda\tau) - C/P \end{aligned}$$

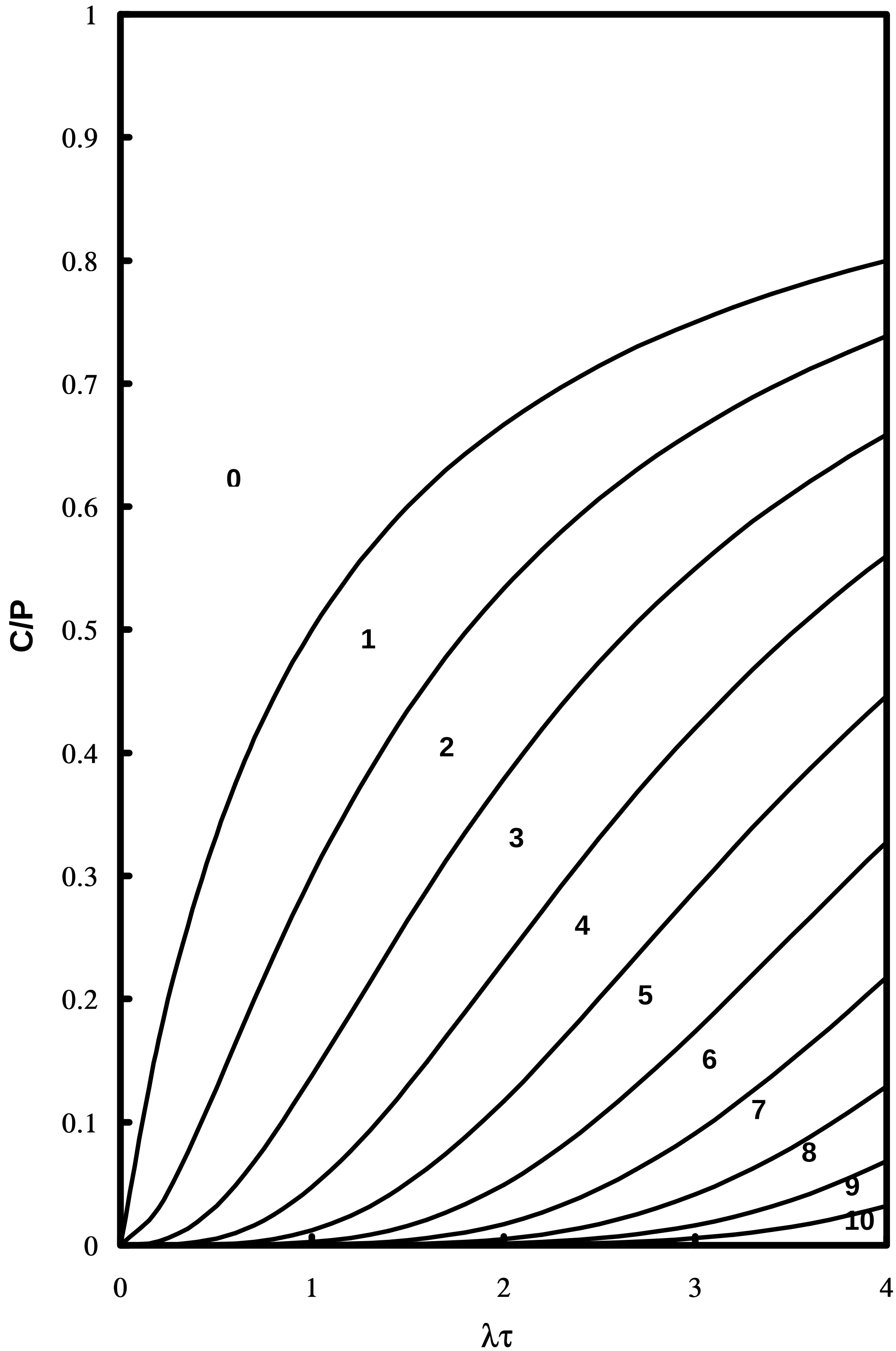
Then s minimises K when

$$\lambda\tau\nabla B(s+1,\lambda\tau) \leq C/P \leq \lambda\tau\nabla B(s,\lambda\tau) \qquad (3)$$

Graphical Analysis for Optimality Regions

For a given $\lambda\tau$ and s inequality (3) provides an *interval of* optimality for the parameter [C/P]; within this interval the given value of s is optimal. Plotting such intervals over a range of interest for $\lambda\tau$, C/P and s allows *optimality regions* to be determined. In considering increasing values of s formula (2) is used to determine $\nabla B(s,\lambda\tau)$. For given values of $\lambda\tau$ and C/P the optimality regions show "at a glance" the optimal number of service channels, s, and also the sensitivity of s to the values of $\lambda\tau$ and C/P. Figure 1 provides an illustrative plot.

Figure 1 Optimal Number of Service Channels



Applications in "Rental" Companies

For many companies the option of hiring items for which there is only an intermittent requirement is less expensive and more convenient than carrying the cost of owning and maintaining the items. As a result many "rental" companies have come into existence and the diverse nature of items available for hire illustrates the demand for this service. Such items include building contractor's construction equipment, office equipment, specialist tools, DIY tools, vehicles, personnel, etc. Figure 1 providing a graphical means of support in determining the optimal number of units of a particular item to make available for hire. The hiring process is defined by the following assumptions.

- Requests for hire of the item arrive according to a Poisson probability distribution with mean arrival rate λ units per day.
- The marginal cost of providing an extra unit is C per unit per day.
- The company has available s units of the item.
- Requests arriving when all s units are on hire are denied service and are lost.
- The hire times of successful requests are independent and identically distributed random variables with a mean of τ days.
- The contribution margin to profit from the hire of the item (revenue minus costs incurred in preparing item upon return for next hire, etc.) is P per unit per day.

It is necessary to establish the correspondence between the hiring process and the M/G/s loss queueing model By identifying item units on hire with busy servers and hire times with service times, it is readily seen the number of item units on hire is distributed as the number of busy servers in the M/G/s loss queueing model.

Example 1 Consider an item which has to be periodically replaced. In determining C the rental company has first to determine an appropriate replacement period, estimate the maintenance costs over the period, the resale value at the end of the period and the appropriate cost of capital for discounting purposes. Some of the difficulties likely to be encountered are enumerated in Walker (1993) and it is clear that the resulting value of C can only be regarded as an estimate. The graphical analysis provides, for given values of $\lambda\tau$ and P, the ranges within which C may vary for optimal values of s.

Example 2 Consider an item which has a seasonal pattern of demand and for which it is possible to purchase/dispose of units at the start/end of each season. Then, in addition to the sensitivity analysis on C as in example 1, for given estimates of λ for each season, the graphical analysis provides a simple, quick and visual means of determining an appropriate optimal dynamic purchase/disposal policy.

Conclusions

A simple graphical means of support in determining the optimal number of service channels for a service loss system is presented. The graphical analysis indicates the optimal number of service channels as a function of $\lambda\tau$ and C/P. The graphical analysis also has the distinct advantage of allowing for simultaneous errors of estimation and/or changes in one or more of the parameters λ, τ, C and P.

References

Cooper, R.B. (1981), *Introduction to Queueing Theory*, North-Holland (Elsevier), Amsterdam.

Messerli, E.J. (1972), "Proof of a Convexity Property of the Erlang B Formula", *Bell Systems Technical Journal*, Vol. 51, pp. 951 - 953.

Tijms, H.C. (1986), *Stochastic Modelling and Analysis: A Computational Approach*, Wiley, Chichester.

Walker, J. (1993), "A Decision Support Tool for Machine Replacement Problems", *Asia-Pacific Journal of Operational Research*, Vol. 10, pp. 171 - 184.