

HETEROGENEOUS-SERVER LOSS SYSTEMS WITH ORDERED ENTRY: AN ANOMALY

Robert B. COOPER and Sharmila PALAKURTHI

Florida Atlantic University, Boca Raton, FL 33431, USA

Received February 1989

Revised May 1989

We consider a loss system with Poisson arrivals and ordered entry, and we suppose that server i has service-time distribution function F_i . Consistent with the results of others, we verify that, in contrast with the classical Erlang loss system ($F_i = F$ for all i), when $F_i \neq F_j$ then the loss probability is not insensitive to the form of F_i or F_j (even when F_i and F_j have the same first moment). Our main point is to note a remarkable counterintuitive fact that is an easy consequence of this result: the policy of assigning arrivals to the fastest idle server does not necessarily produce the lowest loss probability.

heterogeneous servers * loss systems * ordered entry * individual versus social optimization * insensitivity

1. Introduction

We consider a loss system with Poisson arrivals and servers numbered $1, 2, \dots, n$. The server-assignment policy is *ordered entry* (or *ordered hunt*), according to which each arriving customer is assigned to the lowest-numbered idle server. (Arrivals that find all servers busy are cleared from the system.) A customer assigned to server i holds it for a service time whose duration has distribution function F_i . (That is, the service-time distribution 'belongs' to the server, not to the customer.)

It has been known since the time of Erlang that when $F_i = F$ for all i (*homogeneous* servers), the equilibrium probability that an arrival finds all n servers busy is given by the Erlang loss formula

$$E_{1,n}(a) = \frac{a^n/n!}{\sum_{k=0}^n a^k/k!}, \quad (1)$$

where $a = \lambda/\mu$, λ is the Poisson arrival rate, and $\mu^{-1} = \int_0^\infty x dF(x)$ is the mean service time.

The insensitivity of the formula (1) to the form of F is quite remarkable (see, e.g., Cohen [1]), and it implies that (1) is insensitive also to the form of the server-assignment policy. This might lead one

to conjecture that (1) remains valid even when the servers are *heterogeneous* ($F_i \neq F_j$ when $i \neq j$) but with the same mean service time (i.e.,

$$\int_0^\infty x dF_i(x) = \int_0^\infty x dF_j(x) = \mu^{-1});$$

the truth of this conjecture has been demonstrated (by Fakinos [3,4]) for the policy according to which arriving customers are assigned to the idle servers uniformly at random (i.e., an arriving customer has an equal chance of being assigned to each of the idle servers).

Wolff and Wrightson [6] considered a 2-server Poisson-input loss system with multiple classes of customers, with each class characterized by its own service-time distribution and its own preference for the individual servers (i.e., an arrival from the class i that finds both servers idle will select a particular server with probability α_i). They showed that (1) applies for this model also; in passing, they noted that numerical results showed that for our model (in which the service times belong to the servers, not to the customer classes), formula (1) does not apply. In a different context, van Dijk [2] and others (see the references in [2]) also have noted this violation of insensitivity.

In this note we verify algebraically (by a simple example) that the insensitivity property does not remain valid for the common-mean heterogeneous-server loss system when the server-assignment policy is ordered entry. Our main point is to note a remarkable and counterintuitive fact that is an easy consequence of this result: the policy of assigning arrivals to the fastest idle server (in loss systems with heterogeneous servers with different mean service times) does not necessarily produce the lowest loss probability among all ordered-entry server-assignment policies. (More specifically, we consider a particular heterogeneous-server loss system with two servers, and we show that the server-assignment policy that always assigns the slower server to an arrival that finds both servers idle produces a smaller loss probability than the policy that always assigns the faster idle server.)

Why is it surprising that the policy of assigning every arrival to the slowest idle server can produce the best (i.e., lowest probability of loss) system performance? Because it seems apparent that a policy according to which each customer is assigned the server that will, on average, serve him in the least time and thereby be available for service again soonest, will hence result in the least chance that succeeding arrivals will find all servers busy. In other words, it seems apparent here that a policy of *individual optimization* (the arriving customer's service time is minimized) should always produce *social optimization* (the number of customers who get served is maximized); our counterexample shows this presumption to be false.

Indeed, in a study of ordered-entry loss systems with heterogeneous servers with exponential service times (server i has exponential service times with mean μ_i^{-1} , $\mu_i \neq \mu_j$), Yao [7] proved that "... arranging the servers in decreasing (increasing) order of their service rates will yield the smallest (largest) overflow and blocking. While this statement may not sound too surprising ...". (Our results do not contradict Yao's, because we do not retain his assumption of heterogeneous exponential service times.)

Our proof of sensitivity (by exhibiting a simple counterexample that disagrees with (1)) and its consequence that it is better to assign the slow server than the fast is so simple as to be trivial; but its counterintuitive nature makes this note, we feel, worthy of publication.

2. A 2-server heterogeneous loss system with ordered entry

We consider a 2-server loss system in which one of the servers (server G) draws its service times from a population with distribution function G , with mean service time γ ; and the service times of the other server (server H) have distribution function H , with mean η . The arrival process is Poisson, with rate λ , and the server-assignment policy is ordered entry.

In particular, we assume that $G = E_1$ and $H = E_2$ are the exponential distribution function and the 2-phase Erlangian distribution function, respectively, with the same mean $\gamma = \eta = \mu^{-1}$; that is, $G(t) = E_1(t)$ and $H(t) = E_2(t)$, where

$$E_m(t) = 1 - \sum_{\nu=0}^{m-1} \frac{(m\mu t)^\nu}{\nu!} e^{-m\mu t} \quad (t \geq 0).$$

Let $P(i, j)$ be the equilibrium joint probability that server E_1 is in state i ($i = 0$ when server E_1 is idle, and $i = 1$ when E_1 is busy) and server E_2 is in state j ($j = 0$ when E_2 is idle, and $j = k$ when E_2 is in phase k).

We assume first that the order of selection is $E_1 \rightarrow E_2$; that is, server E_1 is always assigned to an arrival that finds both servers idle. Then the probabilities $P(i, j)$ are uniquely determined by the conservation-of-flow equations

$$\lambda P(0, 0) = \mu P(1, 0) + 2\mu P(0, 2),$$

$$(\lambda + 2\mu)P(0, 1) = \mu P(1, 1),$$

$$(\lambda + 2\mu)P(0, 2) = \mu P(1, 2) + 2\mu P(0, 1),$$

$$(\lambda + \mu)P(1, 0) = \lambda P(0, 0) + 2\mu P(1, 2),$$

$$(\mu + 2\mu)P(1, 1) = \lambda P(0, 1) + \lambda P(1, 0),$$

$$(\mu + 2\mu)P(1, 2) = \lambda P(0, 2) + 2\mu P(1, 1),$$

and the normalization equation

$$P(0, 0) + P(0, 1) + P(0, 2) + P(1, 0) + P(1, 1) + P(1, 2) = 1.$$

The probability $B(E_1, E_2)$ of loss (the probability that both servers are busy) when the entry order is $E_1 \rightarrow E_2$ is given by

$$B(E_1, E_2) = P(1, 1) + P(1, 2),$$

which (as determined by MACSYMA) can be written as

$$B(E_1, E_2) = \frac{a^2(5 + 5a + a^2)}{(1 + a)(9 + 11a + 6a^2 + a^3)}, \quad (2)$$

where $a = \lambda/\mu$.

We now assume that the order of selection is reversed, $E_2 \rightarrow E_1$. Then, a similar analysis shows that the loss probability $B(E_2, E_1)$ is given by

$$B(E_2, E_1) = \frac{a^2(4 + 6a + a^2)}{(1 + a)(9 + 11a + 7a^2 + a^3)}. \quad (3)$$

Comparison of equations (1), (2), and (3) shows that, for $a > 0$,

$$B(E_2, E_1) < B(F, F) < B(E_1, E_2), \quad (4)$$

where $B(F, F)$ is given by (1) with $n = 2$, and F is any distribution function with mean μ^{-1} . (In particular, (4) shows that

$$B(E_2, E_1) < B(E_2, E_2) = B(E_1, E_1) < B(E_1, E_2).$$

The last of these inequalities follows also from results of Niu [5], who shows that a server receiving input according to an interrupted Poisson process, which describes the overflow from E_1 , has a loss probability that increases as its service-time distribution becomes more regular; he calls this "perhaps a surprising phenomenon". Statement (4) verifies that (in contrast to the results of Fakinos for the Poisson-input common-mean heterogeneous-server loss system with random server-assignment policy) the loss probability for the ordered-entry server-assignment policy is not insensitive to the form of the service-time distribution functions.

3. The fastest idle server is not necessarily the best choice

Now suppose that the mean service time for server E_2 is increased slightly, to $\mu^{-1} + \delta$. Then,

since the loss probabilities are (clearly) continuous functions of the mean service time associated with any particular server, the values of $B(E_2, E_1)$ and $B(E_1, E_2)$ will change only slightly; that is, one can choose $\delta > 0$ such that (4) remains valid. Hence, even though the mean service time for server E_2 is greater (by δ) than that for server E_1 , nevertheless $B(E_2, E_1) < B(E_1, E_2)$, and our assertion is proved.

References

- [1] J.W. Cohen, "Sensitivity and insensitivity", *Delft Progr. Rep.*, 159-173 (1980).
- [2] N. van Dijk, "A proof of simple insensitive bounds for a pure overflow system", *J. Appl. Probab.* **26**, 113-120 (1989).
- [3] D. Fakinos, "The $M/G/k$ blocking system with heterogeneous servers", *J. Opl. Res. Soc.* **31**, 919-927 (1980).
- [4] D. Fakinos, "The generalized $M/G/k$ blocking system with heterogeneous servers", *J. Opl. Res. Soc.* **33**, 801-809 (1982).
- [5] S.-C. Niu, "A single server queueing loss model with heterogeneous arrival and service", *Oper. Res.* **28**, 584-593 (1980).
- [6] R.W. Wolff and C.W. Wrightson, "An extension of Erlang's loss formula", *J. Appl. Probab.* **13**, 628-632 (1976).
- [7] D.D. Yao, "The arrangement of servers in an ordered-entry system", *Oper. Res.* **35**, 759-763 (1987).