A SIMPLE DYNAMIC ROUTING PROBLEM

by

A. Ephremides, P. Varaiya and J. Walrand

Memorandum No. UCB/ERL M79/37

7 May 1979

A SIMPLE DYNAMIC ROUTING PROBLEM

by

A. Ephremides, P. Varaiya and J. Walrand

Memorandum No. UCB/ERL M79/37

7 May 1979

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

A SIMPLE DYNAMIC ROUTING PROBLEM

A. Ephremides,[†] P. Varaiya and J. Walrand

Department of Electrical Engineering and Computer Sciences
and the Electronics Research Laboratory
University of California, Berkeley, California 94720

ABSTRACT

As jobs arrive they have to be routed to one of two similar exponential servers. It is shown that if the queue lengths at both servers are observed then the optimal decision is to route jobs to the shorter queue, whereas if the queue lengths are not observed then it is best to alternate between queues, provided the initial distribution of the two queue sizes is the same. The optimality of these routing strategies is independent of the statistics of the job arrivals.

Research sponsored in part by the National Science Foundation under Grants ENG76-16816 and ENG77-22752 and the Miller Institute. The authors are grateful to Han-Shing Liu for helpful comments.

[†]On leave from Department of Electrical Engineering, University of Maryland.

1. Introduction

While the analysis of static (stationary) routing strategies in queueing networks has received moderate attention [1,2], the study of dynamic strategies has barely begun. By dynamic strategy one means a policy which, at each time, bases the choice of a route for a job upon the information then available. The difficulties involved in the study of dynamic strategies seem to stem from two sources. First, the information available may be "non-classical" [3] in nature; for example, the information available at different nodes may be different. Second, even the most simple dynamic strategy can lead to queue behavior whose statistical characteristics are not yet adequately understood.

In this paper we study the most elementary problem in which the first kind of difficulty mentioned above is absent. Consider the network depicted in Figure 1. Jobs arrive at times $0 < t_1 < t_2 < \ldots < T$ during the interval $[0,T]$. These arrival times are known in advance. (It will be seen later that this knowledge is useless.) At each time t_i it has to be decided whether to send the just-arrived job to queue 1, i.e., choose $r_t = 1$, or to send it to queue 2, $r_t = 0$. Each queue is serviced by an independent exponential server with the same parameter μ. The decision is to be based on the information available at t_i. We consider two different information patterns. In the first case the number of jobs x_t^i in each queue (the job in service included) is known at each t. In the second case it is known that at time 0 both queues are equal $x_0^1 = x_0^2$, but nothing further is known about the queue behavior (except for the arrival times). The problem is to find for each case the optimal decision rule which minimizes

$$E\{\int_0^T (x_t^1+x_t^2)dt+\frac{1}{2\mu}x_T^1(x_T^1+1)+\frac{1}{2\mu}x_T^2(x_T^2+1)\}, \tag{1.1}$$

which is the total expected time for the completion of service on all jobs which arrive before T.

It is shown in the next section that in the first case the optimum rule is to send each job to the shorter queue. In Section 3 it is shown that in the second case the best decision is to follow the "round-robin" rule: send the jobs arriving at $t_1,t_3,t_5,\ldots$ to queue 1 and those arriving at $t_2,t_4,t_6,\ldots$ to queue 2. Observe that neither rule depends upon the arrival times, hence they are a fortiori optimal when these times are unknown or random. In Section 4 the second case is studied further to show that the policy which assigns a customer to the queue with shorter expected queue length need not be optimum if $x_0^1 \neq x_0^2$, whereas if $x_0^1 = x_0^2$ this policy reduces to the round-robin rule. Some concluding remarks concerning Bernoulli splitting are collected in Section 5.

2. The Send-to-Shorter-Queue (SS) Policy

In this section it is assumed that $x_t = (x_t^1,x_t^2)$ is observed at each t. Let $V(t,x^1,x^2)$ be the (expected) cost incurred over [t,T] when $x_t = (x^1,x^2)$. (In case both queue lengths are equal when a job arrives, the SS policy assigns the job to queue 1.)

Lemma 2.1. (i) $V(T,x^1,x^2) = \frac{1}{2\mu}[x^1(x^1+1)+x^2(x^2+1)]$.

(ii) If $t = t_i$ is an arrival epoch then

$$V(t-,x^1,x^2) = \begin{cases} V(t+,x^1+1,x^2) & \text{if } x^1 \leq x^2 \\ V(t+,x^1,x^2+1) & \text{if } x^1 > x^2. \end{cases}$$

(iii) If there is no arrival during [t-dt,t], then

$$V(t-dt,x^1,x^2) = V(t,x^1,x^2)(1-2\mu dt)+\{V(t,(x^1-1)^+,x^2)+V(t,x^1,(x^2-1)^+)\}\mu dt + (x^1+x^2)dt.$$

(Here for any number y, $y^+ = \max(y,0)$.)

Proof. (i) follows from (1.1) and (ii) from the definition of the SS policy. Suppose there is no arrival during [t-dt,t] and $x_{t-dt} = (x^1,x^2)$. Then the only change in x_t is due to service completions and so by properties of the exponential distribution of the server

$$x_t = \begin{cases} ((x^1-1)^+,x^2) & \text{with probability } \mu dt \\ (x^1,(x^2-1)^+) & \text{with probability } \mu dt \\ (x^1,x^2) & \text{with probability } 1-2\mu dt. \end{cases}$$

Hence (iii) follows by evaluating

$$V(t-dt,x^1,x^2) = E\{V(t,x_t^1,x_t^2)\,|\,x_{t-dt}=(x^1,x^2)\}$$ ¤

Remark. In the above and later on we continue to neglect terms of magnitude o(dt).

Lemma 2.2. $V(t,x^1,x^2) \le V(t,y^1,y^2)$ whenever $x^1 \le y^1$, $x^2 \le y^2$.

Proof. From (1.1) it is clear that

$$V(T,x^1,x^2) \le V(T,y^1,y^2) \text{ when } x^i \le y^i.$$

Suppose now that the assertion is true for t+ where t is an arrival epoch. We shall prove the assertion for t-. Since $x^i \le y^i$, therefore

either (i) $x^1 \geq x^2$, $y^1 \geq y^2$, or (ii) $x^1 \leq x^2 \leq y^2 < y^1$, or (iii) $x^1 > x^2$, $y^1 > y^2$, or (iv) $x^2 < x^1 \leq y^1 \leq y^2$. In case (i)

$$V(t-,x^1,x^2) = V(t+,x^1+1,x^2) \leq V(t+,y^1+1,y^2) = V(t-,y^1,y^2)$$

where the inequality follows from the hypothesis. Under (ii)

$$V(t-,x^1,x^2) = V(t+,x^1+1,x^2) \leq V(t+,y^1,y^2) \leq V(t+,y^1,y^2+1) = V(t-,y^1,y^2)$$

where again both inequalities follow from the hypothesis. The remaining cases are treated similarly.

Finally suppose the assertion is true for t and that no arrivals occur during [t-dt,t]. Suppose $x^i \leq y^i$. Then $(x^i-1)^+ \leq (x^i-1)^+ \leq (y^i-1)^+$ and so the result follows from Lemma 2.1 (iii). ¤

Lemma 2.3. $V(t,x,y) = V(t,y,x)$.

Proof. By (1.1) the assertion holds for T. Suppose it is true for t+, t being an arrival epoch. Let $x < y$. Then

$$V(t-,x,y) = V(t+,x+1,y) = V(t+,y,x+1), \text{ by hypothesis,}$$

$$V(t-,y,x) = V(t+,y,x+1).$$

On the other hand if the hypothesis is true for t and [t-dt,t] contains no arrival epoch, then

$$V(t-dt,x,y) = V(t,x,y)(1-2\mu dt) + [V(t,(x-1)^+,y)+V(t,x,(y-1)^+]\mu dt + (x+y)dt$$

$$= V(t,y,x)(1-2\mu dt) + [V(t,y,(x-1)^+) + V(t,(y-1)^+,x)]\mu dt$$

$$+ (x+y)dt, \text{ by hypothesis}$$

$= V(t-dt,y,x)$. ¤

Lemma 2.4. Suppose $x^1 + x^2 = y^1 + y^2$ and $|x^1-x^2| \leq |y^1-y^2|$. Then $V(t,x^1,x^2) \leq V(t,y^1,y^2)$.

Proof. Because of Lemma 2.3 we may assume $0 \leq x^2 - x^1 < y^2 - y^1$. The assertion is clearly true for T. Suppose it is true for t+, t an arrival epoch. Then

$$V(t-,x^1,x^2) = V(t+,x^1+1,x^2)$$

$$V(t-,y^1,y^2) = V(t+,y^1+1,y^2)$$

Now $|x^1+1-x^2| \leq |y^1+1-y^2|$ and so, by hypothesis, $V(t-,x^1,x^2) \leq V(t-,y^1,y^2)$. Similarly, if the assertion is true for t and [t-dt,t] contains no arrivals then one can verify that

$$V(t-dt,x^1,x^2) \leq V(t-dt,y^1,y^2).$$ ¤

The next lemma states an appealing property of the function V. Its proof can be easily constructed as well.

Lemma 2.5. For each t and y $V(t,\cdot,y)$ is convex i.e., $V(t,x+1,y) - V(t,x,y)$ increases with x.

Theorem 2.1. The SS policy is optimal.

Proof. It is enough to show that V satisfies the dynamic programming equations at each arrival epoch t, i.e.,

$$V(t-,x^1,x^2) = \underset{0\leq r\leq 1}{\text{Min}} \{rV(t+,x^1+1,x^2) + (1-r)(V(t+,x^1,x^2+1)\} \qquad (2.1)$$

where r is the probability of routing the just-arrived job to queue 1. Because of Lemma 2.3 we may assume $x^1 \leq x^2$. Then $|x^1 + 1 - x^2| \leq |x^1 - (x^2+1)|$ and so, by Lemma 2.4, $V(t+,x^1+1,x^2) \leq V(t+,x^1,x^2+1)$ and so the right-hand side is minimized by r = 1 which is also the routing decision of the SS policy. □

3. The Round-Robin (RR) Policy

It is assumed that the initial queue lengths are known and equal, $x_0^1 = x_0^2$, and no further observations are made. The RR policy assigns the odd-numbered arrivals (at $t_1,t_3,t_5\ldots$) to queue 1 and the remainder (at $t_2,t_4,t_6,\ldots$) to queue 2. Let x_t^1, x_t^2, $t \geq 0$, denote the resulting random process of queue lengths. Consider any other assignment policy and denote the resulting random queue lengths by $y_t^1,y_t^2,t \geq 0$. We wish to compare the behavior of (x_t^1,x_t^2) and (y_t^1,y_t^2).

Definition 3.1. Let x^i, y^i i = 1,2 be non-negative integer-valued random variables. Then $(x^1,x^2) \prec (y^1,y^2)$ if there exist random variables $\tilde{x}^i,\tilde{y}^i$, possibly defined on a different probability space, such that

$$x^i \text{ and } \tilde{x}^i,\ y^i \text{ and } \tilde{y}^i \text{ have the same distribution,} \tag{3.2}$$

$$\tilde{x}^1 \leq \tilde{x}^2 \leq \tilde{x}^1 + 1 \quad \text{a.s.} \quad \text{or} \quad \tilde{x}^2 \leq \tilde{x}^1 \leq \tilde{x}^2 + 1 \quad \text{a.s.} \tag{3.2}$$

$$\tilde{x}^1 + \tilde{x}^2 \leq \tilde{y}^1 + \tilde{y}^2 \quad \text{a.s.} \tag{3.3}$$

Lemma 3.1. Suppose $(x^1,x^2) \prec (y^1,y^2)$. Let f(n), n = 0,1,... be any convex increasing function. Then

$$E[f(x^1)+f(x^2)] \leq E[f(y^1)+f(y^2)]$$

Proof. Let $\tilde{x}^i$, $\tilde{y}^i$ be as in the definition. Observe that $Ef(x^i) = Ef(\tilde{x}^i)$ and $Ef(y^i) = Ef(\tilde{y}^i)$. We may suppose

$$\tilde{x}^1 \leq \tilde{x}^2 \leq \tilde{x}^1 + 1 \quad \text{a.s.} \tag{3.4}$$

Let ω be a sample point and suppose $\tilde{y}^1(\omega) \leq \tilde{y}^2(\omega)$. It is easy to see from (3.3) and (3.4) that either

$$0 \leq \tilde{x}^2(\omega) - \tilde{x}^1(\omega) \leq \tilde{y}^2(\omega) - \tilde{y}^1(\omega), \tag{3.5}$$

or

$$\tilde{x}^i(\omega) \leq \tilde{y}^i(\omega), \; i = 1,2. \tag{3.6}$$

Since f is convex and increasing it follows from (3.5) and (3.3) or from (3.6) that

$$f(\tilde{x}^1(\omega)) + f(\tilde{x}^2(\omega)) \leq f(\tilde{y}^1(\omega)) + f(\tilde{y}^2(\omega))$$

The assertion follows by taking expectations. ¤

We can now state the main result .

Theorem 3.1. For each $t \geq 0$, $(x_t^1, x_t^2) \prec (y_t^1, y_t^2)$.

Corollary 3.1. For each $t \geq 0$, $Ex_t^1 + Ex_t^2 \leq Ey_t^1 + Ey_t^2$. In particular the RR policy is optimal.

Proof. The first assertion follows from Theorem 3.1 and Lemma 3.1 by taking $f(n) = n$. To prove the optimality of the RR rule observe that, if no arrivals are permitted after T, then the cost (1.1) is equal to $E\int_0^\infty (x_t^1+x_t^2)dt$ for the RR rule and $E\int_0^\infty (y_t^1+y_t^2)dt$ for the alternative policy. The optimality is immediate. ¤

We now prove the Theorem with the aid of two lemmas.

Lemma 3.2. Let $\tilde{x}^i$, $\tilde{y}^i$ $i = 1,2$ be random variables with $\tilde{x}^1 \leq \tilde{x}^2 \leq \tilde{x}^1 + 1$ a.s. and $\tilde{x}^1 + \tilde{x}^2 \leq \tilde{y}^1 + \tilde{y}^2$ a.s. Let $S^1, S^2, \ldots$ be a sequence of random variables mutually independent and independent of the $\tilde{x}^i$, $\tilde{y}^i$, and each exponentially distributed with parameter μ. Let

$$S_t = \max\{s \mid \sum_{i=1}^{s} S^i \leq t\} \tag{3.7}$$

Let

$$\tilde{x}_t^i = (\tilde{x}^i - S_t)^+, \quad \tilde{y}_t^i = (\tilde{y}^i - S_t)^+ \tag{3.8}$$

Then

$$\tilde{x}_t^1 \leq \tilde{x}_t^2 \leq \tilde{x}_t^1 + 1 \text{ a.s.} \quad \text{and} \quad \tilde{x}_t^1 + \tilde{x}_t^2 \leq \tilde{y}_t^1 + \tilde{y}_t^2 \text{ a.s.}$$

Proof. Let ω be a sample point, let $\tilde{x}^i(\omega) = \hat{x}^i$, $\tilde{y}^i(\omega) = \hat{y}^i$, $S_t(\omega) = \hat{s}$. Suppose $\hat{x}^1 \leq \hat{x}^2 \leq \hat{x}^1 + 1$, $\hat{x}^1 + \hat{x}^2 \leq \hat{y}^1 + \hat{y}^2$. It is trivial to check that then $(\hat{x}^1 - \hat{s})^+ \leq (\hat{x}^2 - \hat{s})^+ \leq (\hat{x}^1 - \hat{s})^+ + 1$ and $(\hat{x}^1 - \hat{s})^+ + (\hat{x}^2 - \hat{s})^+ \leq (\hat{y}^1 - \hat{s})^+ + (\hat{y}^2 - \hat{s})^+$. ¤

Lemma 3.3. For each t there exist random variables $\tilde{x}_t^i$ (defined on some probability space) such that $\tilde{x}_t^i$ and x_t^i have the same distribution and either $\tilde{x}_t^1 \leq \tilde{x}_t^2 \leq \tilde{x}_t^1 + 1$ a.s. or $\tilde{x}_t^2 \leq \tilde{x}_t^1 \leq \tilde{x}_t^2 + 1$ a.s.

Proof. Let $0 < t_1 < t_2 \ldots < t_N < T$ be the arrival times. Recall that $x_0^1 = x_0^2$ by assumption. Hence if $0 \leq t < t_1$, x_t^1 and x_t^2 have the same distribution and so the assertion is true for $0 \leq t < t_1$. Let $\tilde{x}^1 = \tilde{x}^2$ be a random variable with distribution as $x_{t_1-}^1$ and $x_{t_1-}^2$. Now at t_1+, $x_{t_1+}^1 = x_{t_1-}^1 + 1$ and $x_{t_1+}^2 = x_{t_1-}^2$. Let $\tilde{x}_{t_1+}^1 = \tilde{x}^i + 1$ and $\tilde{x}_{t_1+}^2 = \tilde{x}^2$. Then

$\tilde{x}^i_{t_1+}$ and $x^i_{t_1+}$ have the same distribution and $\tilde{x}^2_{t_1+} \leq \tilde{x}^1_{t_1+} \leq \tilde{x}^2_{t_1+} + 1$. Let $t_1 < t_1+t < t_2$. Let S_t be a random variable defined as in (3.7) and independent of the $\tilde{x}^i_{t_1+}$. It is easy to see that $x^i_{t_1+t}$ and $\tilde{x}^i_{t_1+t} = (\tilde{x}^i_{t_1} - S_t)^+$ have the same distribution and that

$$\tilde{x}^2_{t_1+t} \leq \tilde{x}^1_{t_1+t} \leq \tilde{x}^2_{t_1+t} + 1 \text{ a.s.} \tag{3.9}$$

Thus the assertion is true for $t_1 \leq t < t_2$ also. Now at t_2, according to the RR rule $x^1_{t_2+} = x^1_{t_2-}$ and $x^2_{t_2+} = x^2_{t_2-} + 1$. Let $\tilde{x}^1_{t_2+} = \tilde{x}^1_{t_2-}$ and $\tilde{x}^2_{t_2+} = x^2_{t_2-} + 1$. Then $\tilde{x}^i_{t_2+}$ and $x^i_{t_2+}$ have the same distribution and, because of (3.9).

$$\tilde{x}^1_{t_2+} \leq \tilde{x}^2_{t_2+} \leq \tilde{x}^1_{t_2+} + 1 \text{ a.s.}$$

We can now proceed in the same way and prove the assertion for $t_2 \leq t < t_3$. The assertion follows by repeating the argument. □

Proof of Theorem 3.1. We prove the result by induction on N, the number of arrivals. Since $x^1_0 = x^2_0 = y^1_0 = y^2_0$, then, if N = 0, x^1_t, x^2_t, y^1_t, y^2_t all have the same distribution and so the assertion is immediate. Suppose the result is true for N-1, and let t_N be the time of Nth arrival. If $t < t_N$ the result is immediate by the induction hypothesis. So suppose $t \geq t_N$. By the induction hypothesis there exist random variables $\tilde{x}^i_{t_N-}$, $\tilde{y}^i_{t_N-}$ such that $\tilde{x}^i_{t_N-}$ and $x^i_{t_N-}$ and $\tilde{y}^i_{t_N-}$ and $y^i_{t_N-}$ have the same distribution and such that (3.2), (3.3) hold. Without losing generality we may suppose that $\tilde{x}^2_{t_N-} \leq \tilde{x}^1_{t_N-} \leq \tilde{x}^2_{t_N-} + 1$ a.s. Then $x^2_{t_N+} = x^2_{t_N} + 1$,

$x^1_{t_N+} = x^t_{t_N-}$. Suppose, without losing generality, that the alternative policy assigns the arrival at t_N to the second queue so that $y^2_{t_N+} = y^2_{t_N-} + 1$ and $y^1_{t_N+} = y^1_{t_N-}$. Let $\tilde{x}^1 = \tilde{x}^1_{t_N-}$, $\tilde{x}^2 = \tilde{x}^2_{t_N-} + 1$, $\tilde{y}^1 = \tilde{y}^1_{t_N-}$, $\tilde{y}^2 = \tilde{y}^2_{t_N-} + 1$. It is clear that $\tilde{x}^1 \le \tilde{x}^2 \le \tilde{x}^1 + 1$, and $\tilde{x}^1 + \tilde{x}^2 \le \tilde{y}^1 + \tilde{y}^2$. Let $\tau = (t-t_N) \ge 0$, and define $\tilde{x}^i_\tau$, $\tilde{y}^i_\tau$ by (3.9). Then, x^i_t and $\tilde{x}^i_\tau$, and y^i and $\tilde{y}^i_\tau$ have the same distribution and so it follows from Lemma 3.1 that $(x^1_t, x^2_t) \prec (y^1_t, y^2_t)$. ¤

4. The Send-to-Expected Shorter Queue (SES) Policy

We suppose again that the initial queue length distributions are known and that no further observations are made. If the initial distributions of the two queues are the same, then the RR policy coincides with the policy which assigns an arrival to the queue with the shorter expected queue length. From Theorems 2.1, 3.1 it might be conjectured that this SES will be optimal even when the initial distributions are not the same. We give here an example to show that this conjecture is false.

Suppose the initial distribution is as follows. $x^1_0 = 1$ a.s.,

$$x^2_0 = \begin{cases} n \text{ with probability } (1+\varepsilon)n^{-1} \\ 0 \text{ with probability } 1 - (1+\varepsilon)n^{-1} \end{cases}$$

where $\varepsilon > 0$. There are only two arrivals, the first at $t_1 = 0$ and the second at $t_2 = T$ to be specified later. Since $Ex^1_0 = 1 < Ex^2_0 = 1 + \varepsilon$, the SES policy assigns the first arrival to the first queue so

$$x^1_{0+} = 2 \text{ a.s.}, \qquad x^2_{0+} = x^2_0 \text{ a.s.} \tag{4.1}$$

Let x_t^i, $t \geq 0$ be the resulting queue lengths assuming no further arrivals. Then at time T the second arrival will be sent to the queue with the shorter expected length and hence this customer will face an expected waiting time of $\mu^{-1}(Ex_T^1 \wedge Ex_T^2)$ and so the total cost incurred by the SES policy is

$$J_1 = E\int_0^\infty (x_t^1+x_t^2)dt + \mu^{-1}(Ex_T^1 \wedge Ex_T^2) + \mu^{-1} \tag{4.2}$$

where the last term is simply the expected service time for the second arrival. (Here $\wedge$ denotes minimum.)

Consider the alternative policy which assigns the first arrival to the second queue giving queue lengths

$$y_{0+}^1 = x_0^1 = 1 \text{ a.s.}, \quad y_{0+}^2 = x_0^2 + 1 = \begin{cases} n+1 & \text{with prob } (1+\varepsilon)n^{-1} \\ 1 & \text{with prob } 1-(1+\varepsilon)n^{-1} \end{cases} \tag{4.3}$$

Let y_t^i be the resulting queue lengths assuming no further arrivals and suppose that the arrival at T is sent to the first queue. The resulting cost is

$$J_2 = E\int_0^\infty (y_t^1+y_t^2)dt + \mu^{-1}Ey_T^1 + \mu^{-1} \tag{4.4}$$

We will show that for certain values of n,T $J_1 - J_2 > 0$ and so the SES policy cannot be optimal.

To evaluate J_1, J_2 observe that

$$E\int_0^\infty x_t^i dt = \frac{1}{2\mu} Ex_{0+}^i(x_{0+}^i+1)$$

and so, from (4.1),

$$E\int_0^\infty x_t^1 dt = 3\mu^{-1}, \quad E\int_0^\infty x_t^2 dt = \frac{\mu^{-1}}{2}(1+\varepsilon)(n+1).$$

Similarly, from (4.3),

$$E\int_0^\infty (y_t^1+y_t^2)dt = (3+\varepsilon)\mu^{-1} + \frac{\mu^{-1}}{2}(1+\varepsilon)(n+1),$$

and so, substituting into (4.2), (4.4),

$$\mu(J_1-J_2) = -\varepsilon + (Ex_T^1 \wedge Ex_T^2) - Ey_T^1 \tag{4.5}$$

Let S^i be a sequence of independent random variables each exponentially distributed with parameter μ. Let

$$S_T = \max\{s \mid \sum_{i=1}^{s} S^i \leq T\}.$$

Then from (4.1) and (4.3) it follows that

$$Ex_T^1 = E(2-S_T)^+, \quad Ex_T^2 = (\frac{1+\varepsilon}{n})E(n-S_T)^+, \quad Ey_T^1 = E(1-S_T)^+ \tag{4.6}$$

From properties of the exponential distribution

$$\text{Prob}\{S_T = 0\} = \text{Prob}\{S^1 > T\} = e^{-\mu T},$$

$$\text{Prob}\{S_T = 1\} = \text{Prob}\{S^1 < T \leq S^2\} = \mu T e^{-\mu T}$$

and so, from (4.6)

$$Ex_T^1 = 2e^{-\mu T} + \mu T e^{-\mu T}, \quad Ey_T^1 = e^{-\mu T}. \tag{4.7}$$

Now

$$Ex_T^2 \geq \frac{1+\varepsilon}{n}(n-ES_T) = 1 + \varepsilon - \frac{\mu T}{n} .$$

Select T so that $Ex_T^1 < 1$ and then select n so that $Ex_T^1 < 1 - \frac{\mu T}{n}$. Then $Ex_T^1 < Ex_T^2$ and so, from (4.5),

$$\mu(J_1-J_2) = - \varepsilon + Ex_T^1 - Ey_T^1$$

$$= - \varepsilon + e^{-\mu T} + \mu T e^{-\mu T}, \quad \text{from (4.7)}$$

$$> 0$$

for ε sufficiently small.

5. Concluding remarks

Suppose the arrivals during [0,T] form a Poisson stream of rate $\lambda < 2\mu$. According to the result of Foschini-Salz [4], as $T \to \infty$ and under heavy traffic ($\lambda \to 2\mu$), the average system time due to the SS policy approaches the same value as that given by the M/M/2 system. Hence it is significantly lower than the average time incurred by Bernoulli splitting i.e., by randomly assigning an arrival with probability one-half to either queue. On the other hand if the RR policy is adopted the distribution of the interarrival times at each node is the sum of two independent exponential random variables each with parameter λ i.e., the two-stage Erlang distribution Er(2). Thus the RR policy applied to Poisson arrivals results in two parallel queuing systems each of which is Er(2)/M/1. It is easy to show that the resulting average system time is also less than that obtained by Bernoulli splitting. Thus

Bernoulli splitting, while analytically attractive since it preserves exponential interarrivål distributions, appears to be a very poor assignment policy.

References

[1] L. Fratta, M. Gerla and L. Kleinrock, "The flow deviation method: An approach to store-and-forward communication networks," Networks, 3, 97-133, 1973.

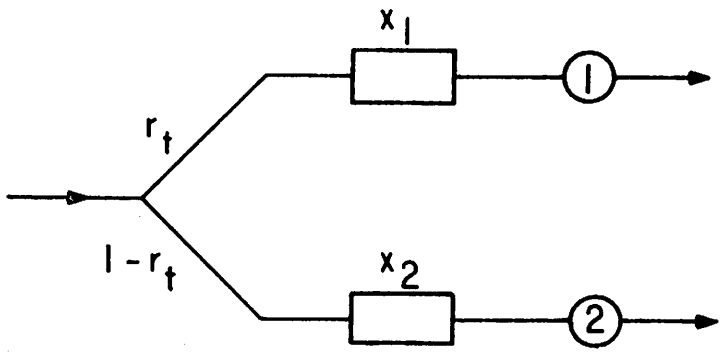
[2] R. Gallager, "A minimum delay routing algorithm using distributed computation," IEEE Trans. Comm., 25 (1), 1977.

[3] H. Witsenhausen, "Separation of estimation and control for discrete time systems," Proceedings IEEE, 59, 1557-1566, 1971.

[4] G. T. Foschini and J. Salz, "A basic dynamic routing problem and diffusion," IEEE Trans. Comm., 26, 320-328, 1978.

Figure Caption

Fig. 1. Network with the queues.


x_1
r_t
1
$1 - r_t$
x_2
2