Regular Paper

Performance Limits of Parallel Server Systems Based on Deterministic Optimal Routing

Kazumasa Oida† and Kazumasa Shinjo†

This paper describes the performance limits of parallel server (PS) systems in which every server has its own queue. The average packet delays of PS systems depend on the routing policy, which assigns each arriving packet to one of the parallel servers. The optimal routings are numerically calculated based on the condition that the input traffic is completely deterministic. These optimal routings show that under a heavy traffic load, PS systems outperform a single server (SS) system. When an infinite number of packets arrive simultaneously, the expected average delay of a PS system that includes 10 servers is 20% smaller than that of an SS system but is 60% larger than that of an SS system that has a "shortest remaining processing time" discipline.

1. Introduction

How far can we improve the performance of parallel server systems? In this paper, we present optimal routings for parallel server (PS) systems in which every server has its own queue. We show that under a heavy traffic load, a PS system can outperform a single server (SS) system whose transmission (service) rate is the same as the total transmission rate of the PS system.

The average packet delay of the PS system is uniquely determined by a routing policy that assigns each arriving packet to one of the parallel servers. For example, assuming that packets arrive according to a Poisson process, the sizes of the arriving packets have a negative exponential distribution, and all p parallel servers in the PS system have the same transmission rate. If each packet is uniformly assigned to one of the p parallel servers according to a Bernoulli process, then by using the M/M/1 model, the average packet delay of the PS system is p times larger than that of the SS system (e.g., Ref. 1)). In contrast, Kingman[2)] showed that under a heavy traffic load, if the packets are assigned according to a "join the shortest queue" (JSQ) policy, the average waiting time of the two parallel servers is smaller than that of the SS system.

An optimal routing policy that minimizes the average packet delay of the PS system absolutely depends on the information available for routing decisions. Most previous studies have dealt with this optimal routing policy based on "stochastic input," in which all packets arrive according to a certain stochastic process and the service time (or the size) of each packet has a certain probability distribution[3)–8)]. For example, Winston[3)] considered the case of Poisson arrivals and exponential service times and showed that the JSQ policy is optimal. Other heuristic routing policies based on the stochastic input can be found in Refs. 9) and 10).

In this paper, we consider an optimal routing based on "deterministic input," in which all of the input packet arrival times as well as all of the packet sizes are fully known in advance. Aicardi et al.[11)] considered the case in which the inter-arrival times and sizes of all packets are deterministic and constant. Our model does not assume that they should be constant. The reason we are studying an optimal routing based on deterministic input is that the obtained optimal routing, called a deterministic optimal routing, achieves the performance limits of the PS systems. This is because the deterministic optimal routing makes the best use of the entire body of information on the input traffic. Therefore, the value of PS systems can be determined according to the performance limits. We believe that efforts spent on this old routing problem should depend on the potential performance.

In Section 2, we first formulate an optimal routing problem for assigning a deterministic input traffic load to parallel servers. The problem was numerically solved with an optimization algorithm, so that an optimal numerical so-

† ATR Adaptive Communications Research Laboratories

lution of the problem represents a deterministic optimal routing. Next, based on the numerical results, we show that PS systems can outperform SS systems only if there is heavy traffic, while under light or moderate traffic the "join the shortest delay" (JSD) policy is almost always optimal. These results indicate that solving this routing problem is significant only if the traffic is heavy. We also show two characteristics of the optimal numerical solutions, which were first introduced by Oida et al.[13] for $p = 2$. We extend the results to the case where $p > 2$.

In Section 3, in order to verify the two characteristics of the optimal solutions, a novel routing policy was created based on these characteristics: if the average delay of the created optimal routing policy is nearly equal to the optimal values of the problem, then these characteristics are proven to be true. Accordingly, our work can be regarded as research into finding the natural laws that may exist in an analytically intractable combinatorial optimization problem.

In Section 4, we consider the performance limits of PS systems based on the assumption that all packets simultaneously arrive. This simultaneous arrival model presents the maximum advantage of PS systems. We show that the average delay of a PS system is at most 20% smaller than that of an SS system but is at least 60% larger than that of an SS system with the shortest remaining processing time discipline.

In Section 5, we discuss the application of our results. Most traffic on the Internet today is bursty, so many SS systems can be replaced by PS systems for better performance. From the results of the simultaneous arrival model, if the batch size of arriving packets is *a priori* known, then by using the optimal "fix queue based on size" (FS) policy, the optimal number of servers that maximizes the performance of the PS system can be determined. This approach is quite practical because the model does not require input traffic to be deterministic.

2. Deterministic Optimal Routing

2.1 Formulation of the Optimization Problem

We have formulated the routing problem for the PS system discussed in the former section.

Consider a parallel server (PS) system (Fig. 1) in which there are p (≥ 2) homogeneous parallel servers ($S_k, k = 1, \ldots, p$). In a

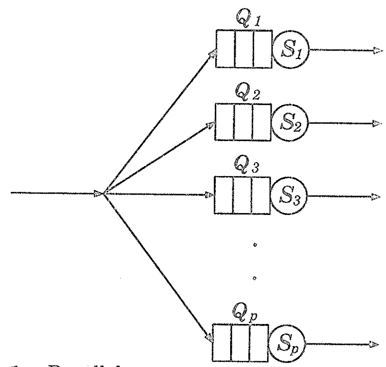


Fig. 1 Parallel server system having p servers.

single server (SS) system, $p = 1$. Each server (S_k) has its own infinite-capacity queue (Q_k), and the transmission rates (C/p) of all servers are identical. Each arriving packet (e_i) joins one of the queues according to a routing policy and is transmitted on a first-come first-served (FCFS) basis. Once a packet joins a queue, it cannot change its queue (that is, no jockeying). The total number (n) of arriving packets ($e_i, i = 1, \ldots, n$) is finite. Let x_i and t_i be the size of packet e_i and its arrival time at the PS system, respectively. We assume the values of the sizes and the arrival times ($x_i, t_i, i = 1, \ldots, n$) of all packets are given (the deterministic input assumption), and $t_1 \leq t_2 \leq \cdots \leq t_n$. Let $W_k(t)$ denote the emptying time of server S_k at time t. Strictly speaking, $W_k(t)$ is equal to the sum of the remaining transmission time of the packet being transmitted by S_k at time t and the total transmission time of all packets waiting in Q_k at time t. If the assignment (u_i) of packet e_i to queue Q_k is given by $u_i = k$, then we have

$$W_k(t_{i+1}) = \max\left(W_k(t_i) + \frac{\theta_k^i x_i}{C/p} - (t_{i+1} - t_i), 0\right), \quad (1)$$
$$k = 1, \ldots, p,$$

where

$$\theta_k^i = \begin{cases} 1, & \text{if } u_i = k, \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

Then, the average packet delay becomes

$$D_{p,n} = \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{p}\theta_k^i W_k(t_i) + \frac{1}{n}\sum_{i=1}^{n}\frac{x_i}{C/p}, \quad (3)$$

where $W_1(t_1) = W_2(t_1) = \cdots = W_p(t_1) = 0$. The first and the second terms on the right-hand side of Eq. (3) represent the average waiting time and the average transmission time, respectively. Accordingly, the optimization prob-

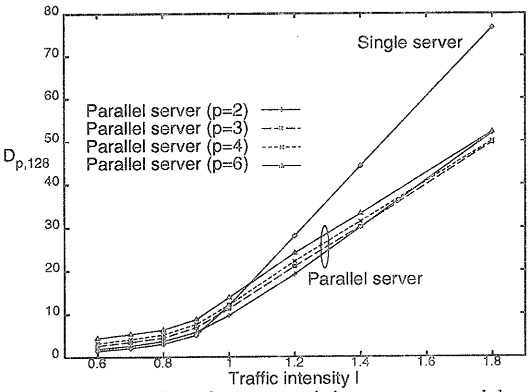


Fig. 2 Comparison between minimum average delays of four PS systems ($p = 2, 3, 4, 6$) and average delays of SS system.

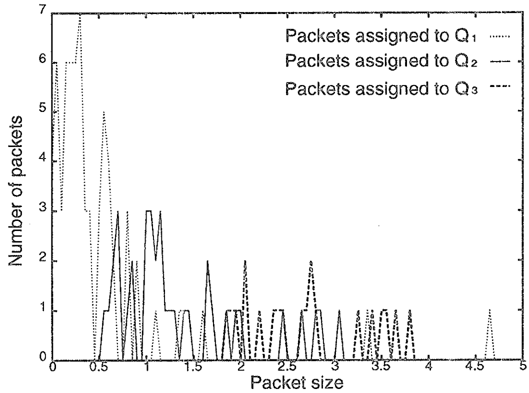


Fig. 3 Size distribution of three packet groups corresponding to three queues in PS system where $p = 3$ and $I = 1.8$.

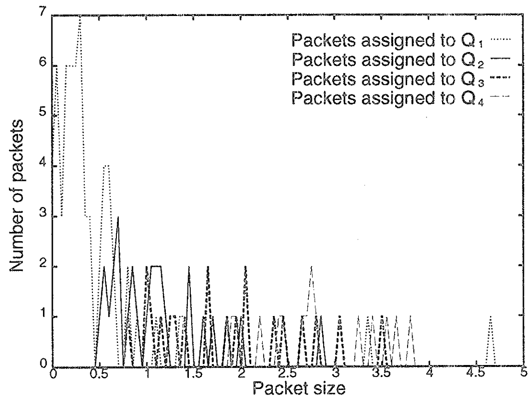


Fig. 4 Size distribution of four packet groups corresponding to four queues in PS system where $p = 4$ and $I = 1.8$.

lem P to be solved in this section can be described as

$$\mathrm{P}\begin{cases} \text{Minimize} & D_{p,n}(u_1, u_2, \ldots, u_n) \\ \text{Subject to} & u_i \in S, \\ & t_i \le t_{i+1},\ i = 1, \ldots, n-1, \\ & x_i > 0,\ i = 1, \ldots, n, \\ & C > 0, \end{cases}$$

where $S = \{1, 2, \ldots, p\}$. Note that problem P can also be regarded as an optimization problem that minimizes the "average waiting time" since the second term on the right-hand side of Eq. (3) does not depend on the routing decisions $\{u_i\}$.

2.2 Calculation Conditions

We numerically solved the optimization problem P with the Hamiltonian Algorithm (HA)[12], which is one of the iterative algorithms used to search for the global optimum solution $\{u_i^*\}$. Suppose that the HA generates a routing sequence $\{u_i^0\}, \{u_i^1\}, \cdots, \{u_i^R\}$. Let $\{u_i^*(R)\}$ be one of the routings in the sequence satisfying $D_{p,n}(\{u_i^*(R)\}) = \min_{0 \le k \le R} D_{p,n}(\{u_i^k\})$. In this paper, routing $\{u_i^*(R)\}$ represents the optimal numerical solution of problem P.

Calculations in this paper were made on parallel computers, and the number of iterations R was more than 10^4. The values for packet sizes $\{x_i\}$ and inter-arrival times $\{t_{i+1} - t_i\}$ were randomly generated based on a negative exponential distribution. In addition, identical values for $\{x_i\}$ and $\{t_{i+1} - t_i\}$ were used for all numerical calculations in this paper.

2.3 Numerical Results

We numerically calculated the optimal values ($D_{p,n}(\{u_i^*(R)\})$ of problem P when $p = 2, \ldots, 6$ and $n = 128$ and obtained the following results. **Under light or moderate traffic, all PS systems are inferior to the SS system; in contrast, under heavy traffic, all PS systems are superior to the SS system.**

The input traffic intensity (I) assigned to the SS system and the PS systems can be described as

$$I = \frac{\lambda}{C\mu}, \tag{4}$$

where $1/\mu$ and $1/\lambda$ are mean values of distributions that generate $\{x_i\}$ and $\{t_{i+1} - t_i\}$, respectively. **Figure 2** compares the average packet delays ($D_{p,128}$) of the PS systems ($p = 2, 3, 4, 6$) with those of the SS system when $1/\mu = 1/\lambda = 1$. (In order to use identical values of $\{x_i\}$ and $\{t_{i+1} - t_i\}$ for all calculations, we varied I by changing the total transmission rate C in (4)). From Fig. 2, if $I \le 0.9$, the SS system scores the best performance. In contrast, if $I > 1.1$, all PS systems outperform the SS system. This result demonstrates that optimal routing policies should be considered under the heavy traffic assumption. Note that if $I \leq 0.9$, the average delay increases with an increase in p, while if $I \rightarrow \infty$, from the lines corresponding to the PS systems, the average delay seems to decrease with an increase in p. We will show in Section 4 that this is true when $n = \infty$.

We also observed the following two characteristics which are common to $p = 2, \ldots, 6$. They were first reported in Ref. 13) when $p = 2$.

Characteristic 1: Under heavy traffic, the optimal routing assigns a queue for each arriving packet based on its size.

Figure 3 and 4 show the size distributions of the packets for p packet groups, where each distribution is assigned to one of the p queues $(Q_1, \ldots, Q_p)$ by the optimal solution $\{u_i^*(R)\}$ when p is 3 and 4, respectively, and $I = 1.8$. As these figures show, the optimal solution determines a queue for each arriving packet based on its size. For example, Fig. 3 shows that most of the small packets are assigned to Q_1, most of the large packets are assigned to Q_3, and the rest are assigned to Q_2. We call this routing a "fix queue based on size" (FS) policy. The assignment $(\tilde{u}_i)$ of packet e_i to queue Q_j according to this FS policy is defined as

$$\tilde{u}_i(\alpha_1^p, \ldots, \alpha_{p-1}^p) = j, \quad \text{if } \alpha_{j-1}^p \leq x_i < \alpha_j^p, \tag{5}$$

where $0 = \alpha_0^p \leq \alpha_1^p \leq \cdots \leq \alpha_{p-1}^p \leq \alpha_p^p = \infty$. The optimal values of tuples $(\alpha_1^p, \ldots, \alpha_{p-1}^p)$ for $p \geq 2$ are discussed in Section 4.

Characteristic 2: Under light or moderate traffic, the average delay of the optimal routing is almost equal to that of the JSD policy.

The assignment $(\hat{u}_i)$ of packet e_i to queue Q_j according to the "join the shortest delay" (JSD) policy can be expressed by

$$\hat{u}_i = j, \quad \text{if } W_j(t_i) = \min_{k \in S} W_k(t_i). \tag{6}$$

Figure 5 compares the average delays $(D_{p,128})$ of four different routings when $p = 2, 4$. The figure shows that the average delays of the JSD policy $\{\hat{u}_i\}$ are almost equal to those of the optimal numerical solution $\{u_i^*(R)\}$ when $I \leq 0.9$.

3. Mimic Optimal Routing

In this section, we present a mimic optimal routing derived from "Characteristic 1" and

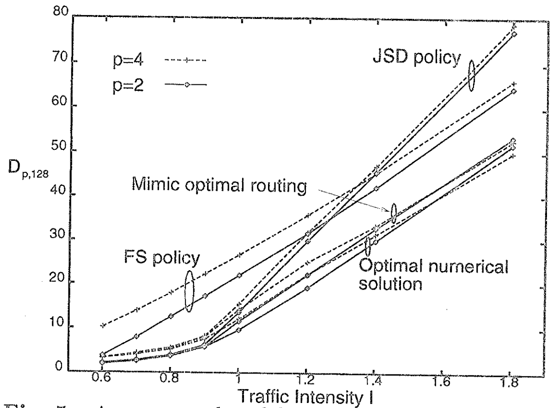


Fig. 5 Average packet delays for four routings: optimal numerical solution $\{u_i^*(R)\}$, mimic optimal routing $\{u_i^{mimic}\}$, FS policy $\{\tilde{u}_i\}$, and JSD policy $\{\hat{u}_i\}$. $1/\lambda = 1/\mu = 1$ and $p = 2, 4$.

"Characteristic 2" of Section 2. The mimic optimal routing is used to verify the characteristics of the optimal numerical solutions. If the mimic optimal routing attains a performance near the optimal value of problem P, then we can conclude that the two characteristics are true.

From the two characteristics, the optimal routing may be expressed as the weighted sum of two routing policies (the JSD policy and the FS policy), and the weights of these two policies would depend on the current input traffic intensity. In order to make a mimic optimal routing based on this conjecture, we first introduce two functions: $d(i, j)$ and $w(I)$. The function $d(i, j)$ represents the distance between packet e_i and queue Q_j and is defined as

$$d(i,j) = \begin{cases} 0, & \text{if } \alpha_{j-1}^p \leq x_i \leq \alpha_j^p, \\ \alpha_{j-1}^p - x_i, & \text{if } x_i < \alpha_{j-1}^p, \\ x_i - \alpha_j^p, & \text{if } \alpha_j^p < x_i, \end{cases} \tag{7}$$

where $\alpha_0^p, \ldots, \alpha_p^p$ correspond to those used in (5). As the distance $d(i, j)$ decreases, the probability that packet e_i is assigned to queue Q_j increases. Next, the function $w(I)$ is defined as

$$w(I) = \frac{1}{2}\{1 + \mathrm{erf}(\frac{I - \gamma}{\sqrt{2}\sigma})\}, \tag{8}$$

where erf denotes the error function, and γ and σ are parameters. The value of $w(I)$ increases from 0 to 1 for I, with the increase occurring around $I = \gamma$, and the increase rate is determined by $\sigma(> 0)$. The weights of the FS and the JSD policies correspond to $w(I_i)$ and $1 - w(I_i)$, respectively, when the current input traffic intensity is I_i. I_i is measured in the time interval $[t_i - \delta_1, t_i + \delta_2]$, where $\delta_1(> 0)$ and $\delta_2(> 0)$ are parameters; *i.e.*, suppose that

$$\begin{cases} t_{i-r-1} < t_i - \delta_1 \le t_{i-r}, \\ t_{i-s} \le t_i + \delta_2 < t_{i-s+1}, \end{cases} \quad (9)$$

where r and s are non-negative integers, then $I_i = \lambda_i / C\mu_i$, where

$$\begin{cases} 1/\mu_i = (\sum_{k=0}^{r+s} x_{i-r+k})/(r+s+1), \\ 1/\lambda_i = (\delta_1 + \delta_2)/(r+s+1). \end{cases} \quad (10)$$

Now, we describe the assignment (u_i^{mimic}) of packet e_i to queue Q_j according to the mimic optimal routing as:

$$u_i^{mimic} = j, \quad \text{if } M_{i,j} = \min_{k \in S} M_{i,k}, \quad (11)$$

where

$$M_{i,j} = w(I_i)d(i,j) + \phi_i(1 - w(I_i))(W_j(t_i) - \min_{k \in S} W_k(t_i)) \quad (12)$$

$$\phi_i = \begin{cases} 0, \text{ if } W_1(t_i) = \cdots = W_p(t_i) = 0, \\ \dfrac{1/\mu_i}{\frac{1}{p}\sum_{k=1}^{p} W_k(t_i)}, \text{ otherwise.} \end{cases} \quad (13)$$

Here, ϕ_i is a scaling factor for the waiting time relative to the packet size at time t_i. Note that if $w(I_i) = 0$, the mimic optimal routing described in (11) is equal to the JSD policy; if $w(I_i) = 1$, the mimic optimal routing is the FS policy.

Figure 5 plots the average packet delays of two mimic optimal routings for $p = 2, 4$, where the parameters $(\gamma, \sigma, \delta_1, \delta_2)$ used in (8)-(10) are $(0.89, 0.55, 60, 58)$ and $(1.0, 0.52, 50, 65)$, respectively, and the values of tuples $(\alpha_1^p, \alpha_2^p, \ldots, \alpha_{p-1}^p)$ for $p = 2, 4$ are listed in Table 1 of the next section. For all I, $D_{2,128}(\{u_i^{mimic}\})$ and $D_{4,128}(\{u_i^{mimic}\})$ closely approximate $D_{2,128}(\{u_i^*(R)\})$ and $D_{4,128}(\{u_i^*(R)\})$, respectively. This means that "Characteristic 1" and "Characteristic 2" of Section 2 are confirmed.

The mimic optimal routing indicates two important possibilities. First, the optimization problem P can be reduced. This is because problem P has n variables: $u_1, u_2, \ldots, u_n$, but the mimic optimal routing has only $p + 3$ undetermined parameters: $\alpha_1^p, \alpha_2^p, \ldots, \alpha_{p-1}^p, \gamma, \sigma, \delta_1, \delta_2$. Consequently, if $n \gg p$, the calculation time for obtaining a minimum average delay will decrease considerably with the mimic optimal routing. Second, a practical routing policy that achieves a performance limit may be created based on (11), which does not require all the information on the input packets but only requires $p + 2$ data $(I_i, 1/\mu_i, W_1(t_i), \ldots, W_p(t_i))$ as the current state of the PS system when the shape of the weight curve $w(I; \gamma, \sigma)$ and the value of a tuple $(\alpha_1^p, \alpha_2^p, \ldots, \alpha_{p-1}^p)$ are given. Note that suitable values for the parameters $(\gamma, \sigma, \alpha_1^p, \alpha_2^p, \ldots, \alpha_{p-1}^p)$ can be determined according to $\{x_i\}$ and $\{t_{i+1} - t_i\}$ and that I_i and $1/\mu_i$ require forecasts of the average intensity and average packet size of arriving packets, respectively. Therefore, a practical routing policy will be created based on (11) if the packet arrival process and the packet size distribution are known in advance.

4. Performance Advantage of Parallel Server System

In this section, we consider a simultaneous arrival model, in which all packets simultaneously arrive, for two reasons. First, to obtain the optimal values of tuples in (5). Second, to show how much better the performance of a PS system is than that of an SS system. In Section 2, we showed that PS systems are superior to the SS system when the input traffic intensity is high. Then, in Section 3, we confirmed that the optimal routing for the PS systems uses an FS policy in this traffic environment. Consequently, the simultaneous arrival model and the FS policy maximize the performance advantage of PS systems.

4.1 Simultaneous Arrival Model

Consider a case in which all n packets arrive simultaneously at a PS system. For simplification, the total transmission rate of all servers is one; *i.e.*,

$$t_1 = t_2 = \cdots = t_n, \quad C = 1. \quad (14)$$

In this case, we suppose that the decisions of routing $\{u_i\}$ are made in the following order: $u_1, u_2, \ldots, u_n$. By using constraints (14), the average packet delay $D_{p,n}$ in problem P can be rewritten in the following simple form.

$$J_{p,n} = \frac{p}{n} \sum_{k=1}^{p} u_k^t Q u_k, \quad (15)$$

where

$$u_k = \begin{pmatrix} \theta_k^1 \\ \theta_k^2 \\ \vdots \\ \theta_k^n \end{pmatrix}, Q = \begin{pmatrix} x_1 & 0 & \ldots & 0 \\ x_1 & x_2 & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ x_1 & x_2 & \ldots & x_n \end{pmatrix}.$$

Consequently, we have the optimization problem P′:

$$\mathrm{P}' \begin{cases} \text{Minimize} & J_{p,n}(u_1, u_2, \ldots, u_n) \\ \text{Subject to} & u_i \in S, \\ & x_i > 0,\ i = 1, \ldots, n. \end{cases}$$

4.2 Expected Average Packet Delay for FS policy

Here, we formulate the expected value of J_n according to the FS policy. Let X_i denote a continuous random variable representing the size of packet e_i. Assume that $X_1, X_2, \ldots, X_n$ are all independent and have an identical probability distribution with a density function $p(x)$. Let $\tilde{J}_{p,n}(\alpha_1^p, \ldots, \alpha_{p-1}^p)$ be the expected value of $J_{p,n}$ according to the FS policy $\{\tilde{u}_i(\alpha_1^p, \ldots, \alpha_{p-1}^p)\}$.

Lemma 1 For all $n \geq 1$ and $p \geq 2$,

$$\tilde{J}_{p,n} = pm + \frac{p(n-1)}{2} F_p(\alpha_1^p, \ldots, \alpha_{p-1}^p), \quad (17)$$

where $m = \int_0^\infty x dP(x)$, $dP(x) = p(x)dx$, and

$$F_p(\alpha_1^p, \ldots, \alpha_{p-1}^p) = \sum_{k=1}^p \int_{\alpha_{k-1}^p}^{\alpha_k^p} dP(x) \int_{\alpha_{k-1}^p}^{\alpha_k^p} x dP(x), \quad (18)$$

$$0 = \alpha_0^p \leq \alpha_1^p \leq \cdots \leq \alpha_p^p = \infty. \quad (19)$$

The proof of Lemma 1 is omitted here since it closely resembles the proof of Eq. (17) in Ref. 13).

On the other hand, from Eq. (15), the average packet delay of the SS system is $J_{1,n} = \frac{1}{n}\{nx_1 + (n-1)x_2 + \cdots + 2x_{n-1} + x_n\}$, so the expected average packet delay becomes

$$\tilde{J}_{1,n} = \frac{(n+1)}{2} m. \quad (20)$$

4.3 Single Server versus Parallel Server

We now compare the expected average delays of the PS systems and the SS system.

Let α_p^* be an optimal vector $(\alpha_1^p, \ldots, \alpha_{p-1}^p)$ that minimizes $F_p(\alpha_1^p, \ldots, \alpha_{p-1}^p)$ subject to (19) when the packet size distribution is given. From (17), α_p^* also minimizes $\tilde{J}_{p,n}(\alpha_1^p, \ldots, \alpha_{p-1}^p)$ for any $n (\geq 2)$. Let $\tilde{J}_{p,n}^*$ and F_p^* be the minimum values of $\tilde{J}_{p,n}$ and F_p, respectively. From (17) and (20),

$$\frac{\tilde{J}_{p,n}^*}{\tilde{J}_{1,n}} = p\left(\frac{2}{n+1} + \frac{n-1}{n+1}\frac{F_p^*}{m}\right). \quad (21)$$

Lemma 2 If the packet sizes have a negative exponential distribution, then for any $m > 0$, F_p^*/m is constant.

Lemma 2, which is proven in the Appendix, indicates that if packet sizes have a negative exponential distribution, then the ratio $\tilde{J}_{p,n}^*/\tilde{J}_{1,n}$ does not depend on the mean value m of the distribution. Table 1 shows $\alpha_p^*, p = 2, \ldots, 6$ when $m = 1$. If $m = m'$, the values of the vectors α_p^*, $p = 2, \ldots, 6$ in Table 1 have to be modified to $m'\alpha_p^*$, $p = 2, \ldots, 6$.

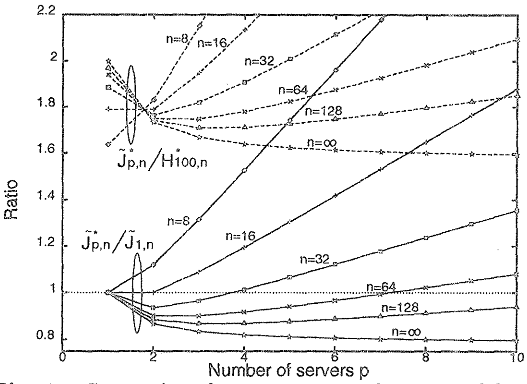


Fig. 6 Comparison between expected average delays in PS systems and SS system when packet sizes have a negative exponential distribution.

Table 1 Values of optimal vectors $\alpha_p^* = (\alpha_1^p, \ldots, \alpha_{p-1}^p)$, $p = 2, \ldots, 6$, when packet sizes have a negative exponential distribution with a mean of one.

p	$\alpha_p^* = (\alpha_1^p, \ldots, \alpha_{p-1}^p)$
2	(1.117)
3	(0.73, 1.666)
4	(0.561, 1.147, 2.032)
5	(0.464, 0.902, 1.448, 2.306)
6	(0.4, 0.755, 1.158, 1.683, 2.526)

Figure 6 plots $\tilde{J}_{p,n}^*/\tilde{J}_{1,n}$, when the packet sizes have a negative exponential distribution. The figure shows that when n (the number of arriving packets) is small, the SS system is better than the PS systems ($\tilde{J}_{p,n}^*/\tilde{J}_{1,n} > 1$). Conversely, when n is large, the PS systems surpass the SS system. When $n = \infty$, the minimum average delay of the PS system that includes 10 servers is approximately 20% smaller than the average delay of the SS system. Note that if $n = \infty$, the ratio $\tilde{J}_{p,n}^*/\tilde{J}_{1,n}$ decreases as p increases.

4.4 Single Server with SRPT

Finally, we compare the performance of the PS systems with that of an SS system that has a "shortest remaining processing time" (SRPT) discipline. In this paper, we have so far only considered the FCFS discipline for packet scheduling. Here, we consider the performance of an SS system that has an SRPT discipline. The SRPT discipline is an optimal packet scheduler that minimizes the number of packets in the SS system at all times independent of any assumptions about the distribution of either the inter-arrival times or the service times[14]. The SRPT discipline always serves the packet with the shortest remaining transmission times. Note that this discipline is preemptive; however, in the simultaneous arrival model, preemption by this discipline never occurs.

We will approximate the SRPT discipline with a prioritized FCFS (p-FCFS) discipline. In the p-FCFS discipline, all arrival packets are divided into g groups such that packet e_i belongs to group j, if $\beta^g_{j-1} \leq x_i < \beta^g_j$, when $0 = \beta^g_0 \leq \beta^g_1 \leq \cdots \leq \beta^g_{g-1} \leq \beta^g_g = \infty$, and packets in the shorter packet size group are given higher priority. Within each group, the FCFS discipline is then applied.

Consider an SS system that has the p-FCFS discipline. Assume that the constraints (14) hold (although all packets simultaneously arrive, the arrival order is assumed to be as follows: $e_1, \ldots, e_n$) and that $X_1, X_2, \ldots, X_n$ are all independent and have identical probability distributions with the density function $p(x)$. Let H_n and $H_{g,n}$ denote the expected average delays of SS systems that have the SRPT and p-FCFS disciplines, respectively, and g and n represent the numbers of groups and packets, respectively.

Lemma 3 For all $n \geq 2$ and $g \geq 2$,

$$H_{g,n} = m + \frac{(n-1)}{2} G_g(\beta^g_1, \ldots, \beta^g_{g-1}), \tag{22}$$

where $m = \int_0^\infty x dP(x)$, $dP(x) = p(x)dx$, and

$$G_g(\beta^g_1, \ldots, \beta^g_{g-1}) = \sum_{k=1}^g \{ \int_{\beta^g_{k-1}}^{\beta^g_k} dP(x) \int_0^{\beta^g_k} x dP(x) + \int_{\beta^g_{k-1}}^{\beta^g_k} x dP(x) \int_{\beta^g_k}^\infty dP(x) \}, \tag{23}$$

$$0 = \beta^g_0 \leq \beta^g_1 \leq \cdots \leq \beta^g_g = \infty. \tag{24}$$

Lemma 3 is proven in the Appendix.

Let β^*_g be an optimal vector $(\beta^g_1, \ldots, \beta^g_{g-1})$ that minimizes $H_{g,n}$ and G_g subject to (24), and let $H^*_{g,n}$ and G^*_g be $H_{g,n}(\beta^*_g)$ and $G_g(\beta^*_g)$, respectively. Since $H_n = \lim_{g\to\infty} H^*_{g,n}$, from (17) and (22),

$$\frac{\tilde{J}^*_{p,n}}{H_n} = \lim_{g\to\infty} \frac{\tilde{J}^*_{p,n}}{H^*_{g,n}} = p\frac{2 + (n-1)F^*_p/m}{2 + (n-1)\lim_{g\to\infty} G^*_g/m}. \tag{25}$$

If the packet sizes have a negative exponential distribution, then G^*_g/m is independent of the mean value $m(> 0)$ of the distribution. (This proof is exactly the same as the proof of Lemma 2.) Thus the ratio $\tilde{J}^*_{p,n}/H^*_{g,n}$ is independent of m. From now on, we only consider the case in which the packet sizes have a negative exponential distribution with mean one. Function G^*_g monotonically decreases with g (the proof is shown in the Appendix) and converges to a constant (G); therefore, if g is large enough, $G^*_g \approx G$. Figure 6 shows the plot of $\tilde{J}^*_{p,n}/H^*_{100,n}$. (We selected G^*_{100} as an approximate value of G. Since $G^*_{90} - G^*_{100} < 10^{-5}$, G^*_{100} may be sufficiently close to G.) This figure shows that for all p and n, $\tilde{J}^*_{p,n} > H^*_{100,n}$. If $n = \infty$, the minimum average delay of a PS system that has 10 servers is approximately 60% larger than the average delay of an SS system with the SRPT discipline.

5. Discussion

We have considered PS systems in which the transmission rates of all p servers are C/p, so for any p the total transmission rate of all servers is identical. We then assume that the transmission rates of all p servers are C instead of C/p. Then the minimum average delay of the PS systems becomes $\tilde{J}^*_{p,n}/p$, which is at least p times smaller than $\tilde{J}_{1,n}$ for all (p, n) pairs that satisfy $\tilde{J}^*_{p,n}/\tilde{J}_{1,n} < 1$ in Fig. 6. In other words, when n is large, in the simultaneous arrival model, the performance of the PS systems becomes more than p times higher than that of the SS system.

Figure 6 indicates that the optimal number of servers that minimizes $\tilde{J}^*_{p,n}/\tilde{J}_{1,n}$ can be determined if the batch size, which corresponds to n in Fig. 6, is *a priori* known. For example, if $n = 32$, then the optimal number of servers is 2. Note that this figure is not based on the deterministic input assumption but simply requires that the optimal FS policy be used for all PS systems; therefore, this approach is quite applicable.

We have come to the conclusion that the PS systems are suitable for a bursty traffic environment. The question is whether such a bursty traffic environment is common or not.

On the Internet today, the window-based flow control[15] is used between two hosts. A receiver reports a window size to a sender for limiting the amount of data the sender can transmit. The window size depends on the buffer size of the receiver. Such window-based flow control makes traffic bursty. For example, if the window size is 1 Mbyte and the size of all packets is 1024 bytes, then the sender can continuously transmit 1024 packets.

6. Conclusions

We have studied a deterministic optimal routing for homogeneous PS systems both numerically and analytically and learned that finding the optimal routing policy is significant only if the traffic is heavy. The performance of PS systems based on deterministic optimal routing can be summarized as follows:

(1) The performance merit of PS systems increases with the input traffic intensity and with the number of simultaneous arriving packets.

(2) When the packet sizes have a negative exponential distribution, the minimum average delay of a PS system is at most 20% smaller than the average delay of an SS system but at least 60% larger than the average delay of an SS system that has an SRPT discipline.

We also presented the following additional results:

(3) The characteristics of the optimal routing for parallel servers ($p = 2, \ldots, 6$) were verified.

(4) An extended mimic optimal routing was described for $p \geq 2$. This routing can be used not only to reduce the size of the optimization problem but also as a basis for creating a practical routing policy that achieves a performance limit.

In the future, we will create a practical optimal routing policy based on this mimic optimal routing. We also plan to investigate the impact of the weight curve's shape on the average packet delay.

Acknowledgments We are grateful to Dr. Bokuji Komiyama, Dr. Shigeru Saito, Dr. Shinsuke Shimogawa, and Mr. Yoshihiro Kitagawa for their assistance.

References

1) Bertsekas, D. P. and Gallager, R. G.: *Data Networks*, Prentice Hall International (1991).

2) Kingman, J. F. C.: Two similar queues in parallel, *Ann. Math. Stat.*, Vol.32, pp.1314-1323 (1961).

3) Winston, W.: Optimality of the shortest line discipline, *Journal of Applied Probability*, Vol.14, pp.181-189 (1977).

4) Ephremides, A., Varaiya, P. and Walrand, J.: A simple dynamic routing problem, *IEEE Transactions on Automatic Control*, Vol.25, No.4, pp.690-693 (1980).

5) Weber, R. W.: On the optimal assignment of customers to parallel servers, *Journal of Applied Probability*, Vol.15, pp.406-413 (1978).

6) Whitt, W.: Deciding which queue to join: some counterexamples, *Operations Research*, Vol.34, No.1, pp.55-62 (1986).

7) Houck, D. J.: Comparison of policies for routing customers to parallel queueing systems, *Operations Research*, Vol.35, No.2, pp.306-310 (1987).

8) Banawan, S. A. and Zahorjan, J.: Load sharing in heterogeneous queueing systems, *Proc. of IEEE INFOCOM'89*, pp.731-739 (1989).

9) Yum, T. P. and Schwartz, M.: The join-biased-queue rule and its application to routing in computer communication networks, *IEEE Transactions on Communications*, Vol.29, No.4, pp.505-511 (1981).

10) Shenker, S. and Weinrib, A.: The optimal control of heterogeneous queueing systems: a paradigm for load-sharing and routing, *IEEE Transactions on Computers*, Vol.38, No.12, pp.1724-1735 (1989).

11) Aicardi, M., Minciardi, R. and Pesenti, R.: Optimal routing in a simple network with deterministic arrival and service times, *Proc. of Annual Allerton Conf. on Communication, Control, and Computing*, pp.640-649 (1990).

12) Shinjo, K. and Sasada, T.: Hamiltonian systems with many degrees of freedom: asymmetric motion and intensity of motion in phase space, *Physical Review*, E54, 4685 (1996).

13) Oida, K. and Shinjo, K.: Characteristics of deterministic optimal routing for a simple traffic control problem, *Proc. of IEEE Int'l Performance, Computing, and Communications Conf.*, pp.386-392 (1999).

14) Schrage, L.: A proof of the optimality of the shortest remaining processing time discipline, *Operations Research*, Vol.16, pp.687-690 (1968).

15) Jacobson, V., Braden, R. and Borman, D.: TCP extensions for high performance, RFC 1323 (1993).

Appendix

A.1 Proof of Lemma 2

Let $p_m(x)$ be a density function of a negative exponential distribution with mean m; *i.e.*, $p_m(x) = \frac{1}{m}e^{-x/m}, x \geq 0, m > 0$. Let $F_{p,m}$ denote F_p defined in (18) when $p(x) = p_m(x)$; *i.e.*,

$$
\begin{aligned}
&F_{p,m}(\alpha_1^p, \ldots, \alpha_{p-1}^p) \\
&= \sum_{k=1}^{p} \int_{\alpha_{k-1}^p}^{\alpha_k^p} dP_m(x) \int_{\alpha_{k-1}^p}^{\alpha_k^p} x dP_m(x),
\end{aligned}
$$

where $dP_m(x) = p_m(x)dx$. If $t = mx$ $(m > 0)$, then

$$
\begin{aligned}
&F_{p,1}(\alpha_1^p, \ldots, \alpha_{p-1}^p) \\
&= \sum_{k=1}^{p} \int_{\alpha_{k-1}^p}^{\alpha_k^p} dP_1(x) \int_{\alpha_{k-1}^p}^{\alpha_k^p} x dP_1(x) \\
&= \sum_{k=1}^{p} \int_{m\alpha_{k-1}^p}^{m\alpha_k^p} dP_m(t) \int_{m\alpha_{k-1}^p}^{m\alpha_k^p} \frac{t}{m} dP_m(t) \\
&= \frac{1}{m} F_{p,m}(m\alpha_1^p, \ldots, m\alpha_{p-1}^p). \qquad (26)
\end{aligned}
$$

Eq. (26) indicates that if $a^* = (a_1^*, a_2^*, \ldots, a_{p-1}^*)$ minimizes $F_{p,1}$, then $ma^* = (ma_1^*, ma_2^*, \ldots, ma_{p-1}^*)$ minimizes $F_{p,m}$, and vice versa, and that for all $m > 0$ $F_{p,1}(a^*) = \frac{1}{m}F_{p,m}(ma^*)$.

A.2 Proof of Lemma 3

Let T_n^g denote the total packet delay of an SS system with p-FCFS discipline, where $g(\geq 2)$ and $n(\geq 2)$ represent the numbers of groups and packets, respectively, and let

$$
\delta_k^n = \begin{cases} 1, & \text{if } \beta_{k-1}^g \leq x_n < \beta_k^g, \\ 0, & \text{otherwise.} \end{cases}
$$

$T_n^g - T_{n-1}^g$ is equal to the sum of the delay of packet e_n and the total waiting time increase of other packets due to packet e_n's arrival; accordingly, it can be expressed as

$$
\begin{aligned}
T_n^g - T_{n-1}^g = &\sum_{k=1}^{g} [\delta_k^n \{\sum_{i=1}^{n-1} x_i(\delta_1^i + \cdots + \delta_k^i) \\
&+ x_n + \sum_{i=1}^{n-1} x_n(\delta_{k+1}^i + \cdots + \delta_g^i)\}].
\end{aligned}
$$

Then, the expected value of $T_n^g - T_{n-1}^g$ can be written as

$$
\begin{aligned}
&E[T_n^g - T_{n-1}^g] \\
&= \int_0^\infty \cdots \int_0^\infty (T_n^g - T_{n-1}^g) dP(x_1) \cdots dP(x_n) \\
&= \sum_{k=1}^{g} \{\int_{\beta_{k-1}^g}^{\beta_k^g} dP(x_n) \sum_{i=1}^{n-1} \int_0^{\beta_k^g} x_i dP(x_i) \\
&\quad + \int_{\beta_{k-1}^g}^{\beta_k^g} x_n dP(x_n) \\
&\quad + \int_{\beta_{k-1}^g}^{\beta_k^g} x_n dP(x_n) \sum_{i=1}^{n-1} \int_{\beta_k^g}^{\infty} dP(x_i)\} \\
&= m + (n-1) G_g(\beta_1^g, \ldots, \beta_{g-1}^g), \qquad (27)
\end{aligned}
$$

where

$$
\begin{aligned}
&G_g(\beta_1^g, \ldots, \beta_{g-1}^g) \\
&= \sum_{k=1}^{g} \{\int_{\beta_{k-1}^g}^{\beta_k^g} dP(x) \int_0^{\beta_k^g} x dP(x) \\
&\quad + \int_{\beta_{k-1}^g}^{\beta_k^g} x dP(x) \int_{\beta_k^g}^{\infty} dP(x)\}.
\end{aligned}
$$

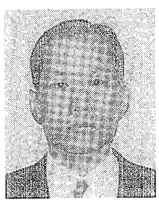
By using (27), we have

$$
\begin{aligned}
H_{g,n} &= E[T_n^g/n] \\
&= \frac{1}{n} E[(T_n^g - T_{n-1}^g) + (T_{n-1}^g - T_{n-2}^g) \\
&\quad + \cdots + (T_2^g - T_1^g) + T_1^g] \\
&= m + \frac{(n-1)}{2} G_g(\beta_1^g, \ldots, \beta_{g-1}^g).
\end{aligned}
$$

A.3 G_g is a monotonically decreasing function

Let $\beta_g^* = (b_1, \ldots, b_{g-1})$ and $\beta_{g+1} = (0, b_1, \ldots, b_{g-1})$. For any $g \geq 2$, we have $G_{g+1}(\beta_{g+1}^*) \leq G_{g+1}(\beta_{g+1}) = G_g(\beta_g^*)$.

(Received April 15, 1999)
(Revised June 4, 1999)
(Accepted July 3, 1999)



Kazumasa Oida received the M.E. degree in information engineering from Hokkaido University in 1985. In 1996, he joined the ATR Adaptive Communications Research Laboratories, where he has been engaged in research on agent-based adaptive routing algorithms in communication networks and performance analysis of parallel server systems. He is a member of IPSJ and JSIAM.

Kazumasa Shinjo has been with the ATR Adaptive Communications Research Laboratories since 1996.