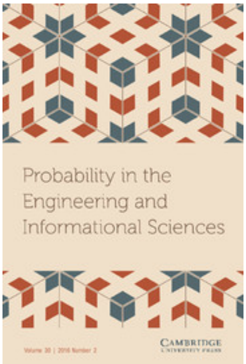

Optimal Sequencing in Multiserver Systems

Dimitrios G. Pandelis and Demosthenis Teneketzis

Probability in the Engineering and Informational Sciences / Volume 10 / Issue 03 / July 1996, pp 377 - 396
DOI: 10.1017/S0269964800004423, Published online: 27 July 2009

Probability in the Engineering and Informational Sciences, **10**, 1996, 377–396. Printed in the U.S.A.

OPTIMAL SEQUENCING IN MULTISERVER SYSTEMS

DIMITRIOS G. PANDELIS
Environmental Research Institute of Michigan
P.O. Box 134001
Ann Arbor, Michigan 48113-4001

DEMOSTHENIS TENEKETZIS
Department of Electrical Engineering and Computer Science
University of Michigan
Ann Arbor, Michigan 48109-2122

We analyze service systems where N servers ($N > 2$) move one at a time along an array of stations to satisfy a known number of requests for service. Processing a request consists of determining the server to satisfy the request. The cost of processing a request is determined by the distance the server that performs the request has to move. We determine qualitative properties of sequencing strategies that minimize the expected cost incurred by the service of all requests when the array of stations is on a circle or an interval.

1. INTRODUCTION—PROBLEM FORMULATION

The problems investigated in this paper are motivated by disk storage systems. Requests for information on the disk are received; to process a request a read-write head is moved to the address of the required information and the read-write operation is performed. There are N arms each with a read-write head. It is assumed that the cost of moving a head to the position of the requested information is a major fraction of the total cost of processing the request. Furthermore, it is assumed that all the arms are controlled by one controller. The

This research was supported in part by NSF grant NCR-9204419.

 0269-9648/96 $11.00 + .10

objective is to minimize the cost of processing a known number of requests that occur according to a given distribution. Similar problems arise in (i) multiple access mechanisms for high-performance computers [1] and (ii) storage systems where loading/unloading mechanisms are moved individually along a common track.

The mathematical model used to formulate and analyze the preceding problems is the following. We consider a service facility consisting of a linear array of K stations or a continuum of stations located along a circle or an interval and N servers moving one at a time from station to station to satisfy requests for service. We assume that the number of requests that have to be served is fixed and known. Requests for service, each designating one station on the circle or the interval, arrive singly. The positions for requests for service are distributed according to an arbitrary probability distribution on the circle or the interval. Processing of the kth request does not begin until processing of the $(k-1)$th request is completed. The location of a request becomes known at the time it is requested. Processing a request consists of determining the server to satisfy the request based on the current server positions, the position where service is requested, and the number of the remaining requests. The server selected is then moved to the position of the request. Processing of the next request is done based on the new server positions. We assume that the cost of processing a request is determined by the distance the server that performs the request has to move. The objective is to determine a server-selection policy that minimizes the expected cost incurred by the service of all the requests.

Two-server models similar to the preceding have been studied by Calderbank, Coffman, and Flatto [2,3]. Under the assumption that the positions of the requests are uniformly distributed, they obtained the following results: (i) the optimal server-selection policy on the interval is characterized by a threshold rule, and (ii) the optimal policy on a K-dimensional sphere ($K \geq 2$) is the one that uses the nearest server. Variations of the two-server models of Calderbank et al. [2,3] have been investigated by Hofri [5], where both servers are allowed to move at the same time, and by Calderbank et al. [4], where the servers are separated by a fixed distance.

The contributions of this paper are the following. We determine properties of optimal server-selection strategies for systems with an arbitrary number of identical servers and requests that are arbitrarily distributed on a circle or a line interval. More specifically, we prove that requests are served by the server immediately to their right or to their left; furthermore, the optimal selection is based on a threshold rule. We prove that optimal thresholds possess a monotonicity property. We extend the results to situations where some parts of the circle are more expensive to traverse than others, and we investigate systems with nonidentical servers. We also provide an alternative proof of the fact that the nearest-server policy is optimal for a system with two servers and requests uniformly distributed on a circle or a sphere and show by counterexample that for multiserver systems and requests uniformly distributed on a circle the nearest-server policy is not optimal in general.

The paper is organized as follows. In Section 2, we first investigate the two-server problem on the circle and present a proof of the optimality of the nearest-server policy; this proof is different from the one presented in Calderbank et al. [2]. Then, we analyze the multiserver sequencing problem. Finally, we show that the sequencing problem on the interval is a special case of the sequencing problem on the circle. In Section 3, we analyze systems with nonidentical servers as well as systems where parts of the circle are more expensive to traverse than others; finally, we present a proof of the optimality of the nearest-server policy for two-server sequencing problems on the sphere. We conclude in Section 4 with a summary of the main results.

2. OPTIMIZATION ON THE CIRCLE

We now analyze the problem where the service facility consists of stations along a circle and each of the N servers can move either in the clockwise or the counterclockwise direction. We proceed by first providing a proof of the optimality of the nearest-server policy for the two-server problem; this proof is different from the one appearing in Calderbank et al. [2]. Then, we study the N-server problem ($N > 2$). We first show by counterexample that for multiserver systems and requests uniformly distributed on the circle the nearest-server policy is not optimal in general. Then, we determine qualitative properties of optimal server selection strategies. The main result of our analysis (Theorem 2) states that any request is served by the server immediately to its right or its left and that the optimal selection is based on a threshold rule. We prove a monotonicity property of the optimal thresholds and discuss a recursive procedure for computing them.

2.1. The Two-Server Problem

For a system with two servers, it has been shown by Calderbank et al. [2] that, when the requests are uniformly distributed on the circle, it is optimal to serve them with the nearest server. In this section, we provide an alternative proof of that result. Without loss of generality, we assume that the circle is of unit length and we denote by $V_n(\alpha)$, $0 \le \alpha \le \frac{1}{2}$, the minimum expected cost of serving n requests when the servers are initially at a distance α on the circle (α is the length of the shortest of the two arcs defined by the positions of the two servers). The following theorem holds.

THEOREM 1:

(i) *For any n, $V_n(\alpha)$, $0 \le \alpha \le \frac{1}{2}$, is nonincreasing in α.*

(ii) *It is optimal to serve any request with the nearest server.*

PROOF: The proof is by induction on n, the number of requests. Consider Figure 1, where the servers are located at points A and B (the length of arc AB being equal to α), C is the midpoint of arc AB, and D and E are the diametrically opposite points of A and C, respectively. Part (ii) is trivial for $n = 1$. Because of symmetry, we can derive an expression for $V_1(\alpha)$ by averaging over the requests closer to B and multiplying by 2. With x being the distance of the request from B, we get

$$V_1(\alpha) = 2\left[\int_0^{\alpha/2} x\,dx + \int_0^{(1-\alpha)/2} x\,dx\right] = \frac{\alpha^2 + (1-\alpha)^2}{4},$$

where the two integrals correspond to arcs CB and BE. It is straightforward to show that the first derivative of the right-hand side (RHS) of the preceding expression is nonpositive for $0 \le \alpha \le \frac{1}{2}$, so part (i) is true for $n = 1$. We now assume that parts (i) and (ii) are true for $n = k$ and prove them for $n = k + 1$.

We prove part (ii) first. It suffices to show that it is optimal to serve the requests that are closer to B than to A with the server at B. Consider a request in arc CB at a distance x from B. The expected costs incurred by serving the request with the servers at B and A are $x + V_k(\alpha - x)$ and $\alpha - x + V_k(x)$, respectively. The result follows from $\alpha - x \ge x$ and the induction hypothesis for part (i). For a request in arc BD at a distance x from B, the corresponding costs are $x + V_k(\alpha + x)$ and $\alpha + x + V_k(x)$, and the result follows from $\alpha + x \ge x$ and the induction hypothesis for part (i). Finally, for a request in arc DE at a distance x from B, the corresponding costs are $x + V_k(1 - \alpha - x)$ and $1 - \alpha - x + V_k(x)$. Note that $(1-\alpha)/2 \ge x$, because $(1-\alpha)/2$ is the length of arc BE. Then, the result follows from $1 - \alpha - x \ge x$ and the induction hypothesis for part (i).

We proceed now to prove part (i). To get an expression for $V_{k+1}(\alpha)$ we use the fact that it is optimal to serve a request with the nearest server and then take

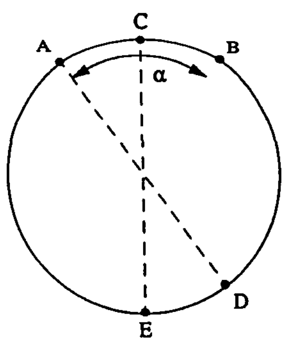

FIGURE 1.

the average over the requests closer to B than to A twice. With x being the distance of the request from B, we get

$$\frac{V_{k+1}(\alpha)}{2} = \int_0^{\alpha/2} [x + V_k(\alpha - x)]\,dx + \int_0^{(1/2)-\alpha} [x + V_k(\alpha + x)]\,dx + \int_{(1/2)-\alpha}^{(1-\alpha)/2} [x + V_k(1 - \alpha - x)]\,dx$$
$$= \frac{\alpha^2 + (1-\alpha)^2}{8} + \int_0^{\alpha/2} V_k(\alpha - x)\,dx + \int_0^{(1/2)-\alpha} V_k(\alpha + x)\,dx + \int_{(1/2)-\alpha}^{(1-\alpha)/2} V_k(1 - \alpha - x)\,dx,$$

where the three integrals correspond to arcs CB, BD, and DE, respectively. Using Leibniz's rule for differentiating integrals, we get

$$\frac{1}{2}\,\frac{dV_{k+1}(\alpha)}{d\alpha} = \frac{\alpha}{2} - \frac{1}{4} + \int_0^{\alpha/2} \frac{dV_k(\alpha - x)}{d\alpha}\,dx + \frac{1}{2}\,V_k\left(\alpha - \frac{\alpha}{2}\right) + \int_0^{(1/2)-\alpha} \frac{dV_k(\alpha + x)}{d\alpha}\,dx + (-1)V_k\left(\alpha + \frac{1}{2} - \alpha\right)$$
$$+ \int_{(1/2)-\alpha}^{(1-\alpha)/2} \frac{dV_k(1 - \alpha - x)}{d\alpha}\,dx + \left(-\frac{1}{2}\right) V_k\left(1 - \alpha - \frac{1-\alpha}{2}\right) - (-1)V_k\left(1 - \alpha - \frac{1}{2} + \alpha\right).$$

After some straightforward algebra, we get

$$\frac{dV_{k+1}(\alpha)}{d\alpha} = \alpha - \frac{1}{2} + V_k\left(\frac{1-\alpha}{2}\right) - V_k\left(\frac{\alpha}{2}\right). \tag{1}$$

Because $0 \le \alpha \le \frac{1}{2}$, we have $0 \le (\alpha/2) \le (1-\alpha)/2 \le \frac{1}{2}$, which by the induction hypothesis for $n = k$ implies that $V_k\big((1-\alpha)/2\big) \le V_k(\alpha/2)$. Therefore, the RHS of Eq. (1) is nonpositive and the proof is complete. ■

2.2. The N-Server Problem ($N > 2$)

In this section, we investigate optimal strategies for systems with more than two servers. We start with some notation. We denote by $V_n(S_1, S_2, \ldots, S_N)$ the minimum expected cost of serving n requests when the servers are initially located at points $S_1, S_2, \ldots, S_N$ on the circle, by $V_n^\pi(S_1, S_2, \ldots, S_N)$ the expected cost of serving n requests incurred under policy π when the servers are initially located at points $S_1, S_2, \ldots, S_N$ and by $V_n^\pi(S_1, S_2, \ldots, S_N; R)$ the conditional expected value of $V_n^\pi(S_1, S_2, \ldots, S_N)$ given that the first request is located at point R. We also denote by ℓ_{AB}^{-}, ℓ_{AB}^{+} the distances on the circle from point A to point B in the counterclockwise and clockwise directions, respectively, and by ℓ_{AB} the length of the shorter arc AB, that is, $\ell_{AB} = \min\{\ell_{AB}^{-}, \ell_{AB}^{+}\}$.

In contrast with two-server systems, even when the requests are uniformly distributed on the circle, it is not, in general, optimal to serve a request with the nearest server. We illustrate this point by the following example. Consider the system of Figure 2. We have three servers, one located at point A and two at point B and two requests with the first one located at point R, where $\ell_{AB} = \ell_{AB}^{+} = \alpha$, $\ell_{AR} = r$, and $r < (\alpha/2)$. We will show that for certain values of r it is optimal to serve the request with one of the servers at B, although the server at A is closer to the request. Let π and π' be the policies that serve the request with the servers at A and B, respectively, and proceed optimally afterward. We have

$$V_2^{\pi}(A,B,B;R) - V_2^{\pi'}(A,B,B;R) = r + V_1(R,B,B) - \{\alpha - r + V_1(A,R,B)\}. \tag{2}$$

Using the fact that for one request it is optimal to use the nearest server, we get

$$V_1(R,B,B) = 2\left[\int_0^{(\alpha-r)/2} x\,dx + \int_0^{(1-\alpha+r)/2} x\,dx\right] = \frac{(\alpha-r)^2}{4} + \frac{(1-\alpha-r)^2}{4}$$

and

$$V_1(A,R,B) = 2\left[\int_0^{r/2} x\,dx + \int_0^{(\alpha-r)/2} x\,dx + \int_0^{(1-\alpha)/2} x\,dx\right]$$
$$= \frac{r^2}{4} + \frac{(\alpha-r)^2}{4} + \frac{(1-\alpha)^2}{4}. \tag{3}$$

From Eqs. (2) and (3), we get

$$V_2^{\pi}(A,B,B;R) - V_2^{\pi'}(A,B,B;R) = 2r - \alpha + \frac{r(1-\alpha)}{2}.$$

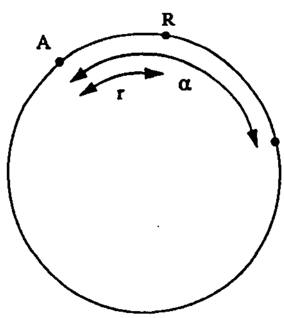

FIGURE 2.

Note that $\lim_{r \to (\alpha/2)} V_2^{\pi}(A,B,B;R) - V_2^{\pi'}(A,B,B;R) = [\alpha(1-\alpha)]/4$, which is a positive expression for any $0 < \alpha < \frac{1}{2}$. This implies that for r sufficiently close to $\alpha/2$ it is optimal to serve the request with one of the servers at point B.

We now turn to the general N-server problem. The following lemma is needed for the establishment of properties of optimal strategies for systems with N servers $(N > 2)$.

LEMMA 1: *For any n, we have*

$$|V_n(S_1,S_2,\ldots,S_N) - V_n(S_1',S_2,\ldots,S_N)| \le \ell_{S_1S_1'}. \tag{4}$$

PROOF: Let S and S' be the systems in which the servers are initially located at points $S_1,S_2,\ldots,S_N$ and $S_1',S_2,\ldots,S_N$, respectively. Let π be the optimal policy for system S and $\tilde{\pi}$ be a policy that serves each request in system S' using the same server as π. For any sequence of requests, the costs due to policies π and $\tilde{\pi}$ differ only the first time the server at S_1 and S_1', respectively, serves a request. In this case, the distance moved by that server may be up to $\ell_{S_1S_1'}$ greater in system S' than in system S. Therefore,

$$V_n(S_1',S_2,\ldots,S_N) \le V_n^{\tilde{\pi}}(S_1',S_2,\ldots,S_N) \le V_n(S_1,S_2,\ldots,S_N) + \ell_{S_1S_1'}, \tag{5}$$

where the first inequality follows from the fact that $\tilde{\pi}$ is not necessarily optimal for system S'. By interchanging S_1 and S_1' in Eq. (5), we get Lemma 1. ■

Properties of an optimal policy are given in the following theorem.

THEOREM 2: *Consider a request located at point D, and let S^-, S^+ be the positions of the servers closest to D in the counterclockwise and clockwise directions, respectively (see Figs. 3 and 4). Then, we have the following:*

(i) *It is optimal to serve the request either with the server at S^- or the server at S^+.*

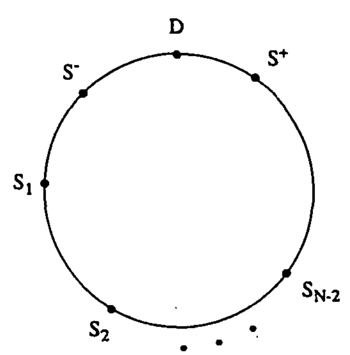

FIGURE 3.

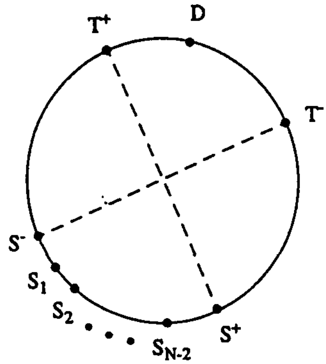

FIGURE 4.

(ii) *There exists a threshold point T such that it is optimal to serve the request with the server at S^- if $\ell^+_{S^-D} \leq \ell^+_{S^-T}$ and with the server at S^+ otherwise.*

PROOF:

(i) Let $S_1, S_2, \ldots, S_{N-2}$ be the positions of the rest $N-2$ servers and n be the number of requests including the one at D. Let π and π' be the policies that serve the request by moving the servers at S^- and S_k, $1 \leq k \leq N-2$, respectively, in the clockwise direction, and proceed optimally afterward. We have

$$V_n^{\pi'}(S_1,\ldots,S_{N-2},S^-,S^+;D) - V_n^{\pi}(S_1,\ldots,S_{N-2},S^-,S^+;D)$$
$$= \ell^+_{S_kD} + V_{n-1}(S_1,\ldots,S_{k-1},S_{k+1},\ldots,S_{N-2},S^-,S^+,D)$$
$$- \{\ell^+_{S^-D} + V_{n-1}(S_1,\ldots,S_{k-1},S_k,S_{k+1},\ldots,S_{N-2},S^+,D)\}. \quad (6)$$

The RHS of Eq. (6) is nonnegative because

$$|V_{n-1}(S_1,\ldots,S_{k-1},S_{k+1},\ldots,S_{N-2},S^-,S^+,D)$$
$$- V_{n-1}(S_1,\ldots,S_{k-1},S_k,S_{k+1},\ldots,S_{N-2},S^+,D)|$$
$$\leq \ell_{S_kS^-} \leq \ell^+_{S_kS^-} = \ell^+_{S_kD} - \ell^+_{S^-D},$$

where the first inequality follows from Lemma 1. By the same argument, we can show that when we move clockwise, using the server at S^- results in less cost than using the server at S^+. Thus, by serving the request with the server at S^- moving in the clockwise direction, we incur less cost than by serving the request with any other server moving in the clockwise direction. We can similarly show that by serving the request with the server at S^+ moving in the counterclockwise direction we incur less cost than by serving the request with any other server moving in the counterclockwise direction. Therefore, it is optimal to serve the request either with the server at S^- or the server at S^+.

(ii) From the proof of part (i), we know that the servers at S^+ and S^- are optimal for requests at points D such that $\ell_{S-D} = \ell^-_{S-D}$ and $\ell_{S+D} = \ell^+_{S+D}$,[1] respectively. Thus, to prove part (ii), we only need to consider requests at points D such that $\ell_{S-D} = \ell^+_{S-D}$ and $\ell_{S+D} = \ell^-_{S+D}$.[2] For such points, we first show that the incentive to use the server at S^+ instead of the one at S^- increases as the request moves closer to S^+. Let π^+ and π^- be the policies that use the servers at S^+ and S^-, respectively, and proceed optimally afterward. The incentive to use the server at S^+ for a request at D is

$$\begin{aligned} I(D) &= V_n^{\pi^-}(S_1,\ldots,S_{N-2},S^-,S^+;D) - V_n^{\pi^+}(S_1,\ldots,S_{N-2},S^-,S^+;D) \\ &= \ell^+_{S-D} + V_{n-1}(S_1,\ldots,S_{N-2},S^+,D) \\ &\quad - \{\ell^-_{S+D} + V_{n-1}(S_1,\ldots,S_{N-2},S^-,D)\}. \end{aligned} \tag{7}$$

For points D_1, D_2, such that D_2 is closer to S^+ than D_1, we will show that $I(D_2) \geq I(D_1)$. We have

$$\begin{aligned} &V_n^{\pi^-}(S_1,\ldots,S_{N-2},S^-,S^+;D_2) + V_n^{\pi^+}(S_1,\ldots,S_{N-2},S^-,S^+;D_1) \\ &= \ell^+_{S-D_2} + V_{n-1}(S_1,\ldots,S_{N-2},S^+,D_2) \\ &\quad + \ell^-_{S+D_1} + V_{n-1}(S_1,\ldots,S_{N-2},S^-,D_1) \\ &\geq \ell^+_{S-D_1} + V_{n-1}(S_1,\ldots,S_{N-2},S^+,D_1) \\ &\quad + \ell^-_{S+D_2} + V_{n-1}(S_1,\ldots,S_{N-2},S^-,D_2) \\ &= V_n^{\pi^-}(S_1,\ldots,S_{N-2},S^-,S^+;D_1) \\ &\quad + V_n^{\pi^+}(S_1,\ldots,S_{N-2},S^-,S^+,D_2), \end{aligned} \tag{8}$$

where the inequality follows from Lemma 1 because $\ell^+_{S-D_2} - \ell^+_{S-D_1} = \ell^-_{S+D_1} - \ell^-_{S+D_2} = \ell_{D_1D_2}$. From Eqs. (7) and (8), we get $I(D_2) \geq I(D_1)$. The proof of part (ii) is now complete when the points S^-, S^+ are as in Figure 4, that is, $\ell_{S-S+} = \ell^-_{S-S+}$. To complete the proof of part (ii) for the case where points S^-, S^+ are as in Figure 3 ($\ell_{S-S+} = \ell^+_{S-S+}$), it suffices to show that $I(S^+) \geq 0$ and $I(S^-) \leq 0$. From Eq. (7), we have

$$\begin{aligned} I(S^+) = \ell^+_{S-S+} &+ V_{n-1}(S_1,\ldots,S_{N-2},S^+,S^+) \\ &- V_{n-1}(S_1,\ldots,S_{N-2},S^-,S^+) \end{aligned}$$

and

$$\begin{aligned} I(S^-) = V_{n-1}&(S_1,\ldots,S_{N-2},S^+,S^-) \\ &- \ell^-_{S+S-} - V_{n-1}(S_1,\ldots,S_{N-2},S^-,S^-). \end{aligned}$$

Then, $I(S^+) \geq 0$ and $I(S^-) \leq 0$ follow directly from Lemma 1. ■

Let points $S_1, S_2, \ldots, S_N$ be located on the circle as shown in Figure 5: S_{k+1} is the closest point to S_k in the clockwise direction, $k = 1, 2, \ldots, N$ (we define $S_{N+1} = S_1$). For such a configuration, we denote by $T_n^k(S_1, S_2, \ldots, S_N)$, $k =$

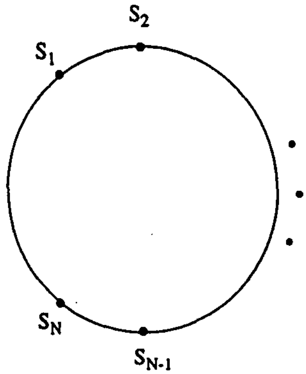

FIGURE 5.

1,2, . . . ,N, the threshold point between S_k and S_{k+1} when we have n requests and the servers are located at $S_1, S_2, \ldots, S_N$. Then, $T_n^k(S_1, S_2, \ldots, S_N)$ is given by the point T that satisfies

$$\ell_{TS_K} + V_{n-1}(S_1, \ldots, S_{k-1}, T, S_{k+1}, \ldots, S_N)$$
$$= \ell_{TS_{k+1}} + V_{n-1}(S_1, \ldots, S_k, T, S_{k+2}, \ldots, S_N), \qquad k = 1,2, \ldots, N-1,$$

and

$$\ell_{TS_N} + V_{n-1}(S_1, S_2, \ldots, T) = \ell_{TS_1} + V_{n-1}(T, S_2, \ldots, S_{N-1}), \qquad k = N. \quad (9)$$

Equation (9) defines a recursive procedure for the computation of thresholds. The computation of these thresholds is tedious, and we will not pursue it any further in this paper. Instead, we prove a monotonicity property of the optimal thresholds.

THEOREM 3:

(i) *Suppose that the servers are located at positions $S_1, S_2, \ldots, S_N$, not necessarily in that order. Let S^-, S^+ be the positions of the servers closest to S_2 in the counterclockwise and clockwise directions, respectively. Then, for all S_2' between S_2 and S^+, we have for any n*

$$\frac{d}{dS_1}[V_n(S_1, S_2', \ldots, S_N) - V_n(S_1, S_2, \ldots, S_N)] \le 0,$$

where the preceding derivative is defined by

$$\frac{d}{dS_1} V_n(S_1, S_2, \ldots, S_N)$$
$$= \lim_{\ell^+_{S_1 S_1'} \to 0} \frac{V_n(S_1', S_2, \ldots, S_N) - V_n(S_1, S_2, \ldots, S_N)}{\ell_{S_1 S_1'}}.$$

(ii) *Consider a request at position D. Let S_ℓ, S_r be the positions of the servers closest to D in the counterclockwise and clockwise directions, respectively. If it is optimal to serve the request at D with the server at S_ℓ when the servers are located at positions $S_1, S_2, \ldots, S_N$, then it is optimal to serve it with the same server when any server is moved in the clockwise direction at a distance small enough that the order of the servers around the circle is not changed.*

PROOF: The proof is by induction on n. Part (i) holds trivially for $n = 0$. We assume that it holds for $n = k - 1$ and prove that both parts (i) and (ii) hold for $n = k$.

By successive application of the induction hypothesis for part (i), we can show that

$$\frac{d}{dS_1}\,[V_{k-1}(S_1, S_2', \ldots, S_N') - V_{k-1}(S_1, S_2, \ldots, S_N)] \le 0, \tag{10}$$

where $S_2', \ldots, S_N'$ are such that $\ell^+_{S_1 S_m'} \ge \ell^+_{S_1 S_m}$, $m = 2, \ldots, N$.

First, we prove part (ii) for $n = k$. Let π_ℓ and π_r be the policies that serve the request at D with the servers at S_ℓ and S_r, respectively, and proceed optimally afterward. We have three cases.

Case 1: A server located at some location $S_m \ne S_\ell, S_r$ moves to location S_m'. Then,

$$\begin{aligned}
&V_k^{\pi_\ell}(S_1, \ldots, S_m', S_\ell, S_r, \ldots, S_N; D) - V_k^{\pi_r}(S_1, \ldots, S_m', S_\ell, S_r, \ldots, S_N; D)\\
&\quad = \ell_{S_\ell D} + V_{k-1}(S_1, \ldots, S_m', D, S_r, \ldots, S_N)\\
&\qquad - \{\ell_{S_r D} + V_{k-1}(S_1, \ldots, S_m', S_\ell, D, \ldots, S_N)\}\\
&\quad \le \ell_{S_\ell D} + V_{k-1}(S_1, \ldots, S_m, D, S_r, \ldots, S_N)\\
&\qquad - \{\ell_{S_r D} + V_{k-1}(S_1, \ldots, S_m, S_\ell, D, \ldots, S_N)\}\\
&\quad = V_k^{\pi_\ell}(S_1, \ldots, S_m, S_\ell, S_r, \ldots, S_N; D)\\
&\qquad - V_k^{\pi_r}(S_1, \ldots, S_m, S_\ell, S_r, \ldots, S_N; D) \le 0.
\end{aligned} \tag{11}$$

The first inequality in Eq. (11) is true because Eq. (10) implies

$$\frac{d}{dS_m}\,[V_{k-1}(S_1, \ldots, S_m, D, S_r, \ldots, S_N) - V_{k-1}(S_1, \ldots, S_m, S_\ell, D, \ldots, S_N)] \le 0.$$

The second inequality in Eq. (11) follows from the fact that it is optimal to use the server at S_ℓ when a server is located at S_m. Therefore, when the server moves from S_m to S_m', it is still optimal to use the server at S_ℓ.

Case 2: The server at S_{ℓ} moves to a position S'_{ℓ} between S_{ℓ} and D. Then,

$$
\begin{aligned}
&V_k^{\pi_r}(S_1,\ldots,S'_\ell,S_r,\ldots,S_N;D) - V_k^{\pi_\ell}(S_1,\ldots,S'_\ell,S_r,\ldots,S_N;D)\\
&\quad = \ell_{S'_\ell D} + V_{k-1}(S_1,\ldots,D,S_r,\ldots,S_N)\\
&\qquad - \{\ell_{S_r D} + V_{k-1}(S_1,\ldots,S'_\ell,D,\ldots,S_N)\}\\
&\quad \leq \ell_{S_\ell D} + V_{k-1}(S_1,\ldots,D,S_r,\ldots,S_N)\\
&\qquad - \{\ell_{S_r D} + V_{k-1}(S_1,\ldots,S_\ell,D,\ldots,S_N)\}\\
&\quad = V_k^{\pi_r}(S_1,\ldots,S_\ell,S_r,\ldots,S_N;D) - V_k^{\pi_\ell}(S_1,\ldots,S_\ell,S_r,\ldots,S_N;D) \leq 0.
\end{aligned}
\tag{12}
$$

The first inequality in Eq. (12) follows from $\ell_{S_\ell D} - \ell_{S'_\ell D} = \ell_{S_\ell S'_\ell}$ and Lemma 1. The second inequality in Eq. (12) is true because it is optimal to use the server at S_ℓ. Therefore, when the server moves from S_ℓ to S'_ℓ, it is still optimal to use that server.

Case 3: The server at S_r moves to position S'_r. The proof is identical to that of Case 2.

We now prove part (i) for $n = k$. By conditioning on the location of the first request, we get

$$
\begin{aligned}
\frac{d}{dS_1} V_k(S_1,S_2,\ldots,S_N) &= \frac{d}{dS_1} E\{V_k(S_1,S_2,\ldots,S_N;D)\}\\
&= \lim_{\ell^+_{S_1 S'_1}\to 0} \frac{E\{V_k(S'_1,S_2,\ldots,S_N;D)\} - E\{V_k(S_1,S_2,\ldots,S_N;D)\}}{\ell_{S_1 S'_1}}.
\end{aligned}
$$

We can show by the coupling argument used in the proof of Lemma 1 that for any D

$$
|V_k(S'_1,S_2,\ldots,S_N;D) - V_k(S_1,S_2,\ldots,S_n;D)| \leq \ell_{S_1 S'_1}.
$$

Then, by the bounded convergence theorem, it is valid to interchange the order of limit and expectation to get

$$
\begin{aligned}
&\frac{d}{dS_1} V_k(S_1,S_2,\ldots,S_N)\\
&\quad = E\left\{ \lim_{\ell^+_{S_1 S'_1}\to 0} \frac{V_k(S'_1,S_2,\ldots,S_N;D) - V_k(S_1,S_2,\ldots,S_N;D)}{\ell_{S_1 S'_1}} \right\}\\
&\quad = E\left\{ \frac{d}{dS_1} V_k(S_1,S_2,\ldots,S_N;D) \right\}.
\end{aligned}
$$

Therefore, it suffices to show that for any D

$$\frac{d}{dS_1}\,[V_k(S_1,S_2',\ldots,S_N;D) - V_k(S_1,S_2,\ldots,S_N;D)] \le 0.$$

We adopt the convention that if the optimal policy is indifferent between the request's two adjacent servers, then it chooses the server counterclockwise of the request. This convention implies that the optimal policy remains the same when the position of the server at S_1 is changed clockwise to a position S_1' sufficiently close to S_1. This is because when the optimal policy for S_1 is indifferent between the two adjacent servers, the server counterclockwise of the request is optimal by convention, and it is also optimal for S_1' by part (ii). We denote by S, S' the position of the server that the optimal policy uses to serve the request at D for $(S_1,S_2,\ldots,S_N)$ and $(S_1,S_2',\ldots,S_N)$, respectively. There are several cases to consider.

Case 1: $S = S' = S_1$.

$$\frac{d}{dS_1}\,[V_k(S_1,S_2',\ldots,S_N;D) - V_k(S_1,S_2,\ldots,S_N;D)]$$

$$= \frac{d}{dS_1}\,[\ell_{S_1D} + V_{k-1}(D,S_2',\ldots,S_N) - \ell_{S_1D} - V_{k-1}(D,S_2,\ldots,S_N)] = 0.$$

Case 2: $S = S_1$, $S' = S_2'$.

If D is such that $\ell_{S_1D} = \ell^+_{S_1D}$ (see Fig. 6), then by part (ii) the server at S_1 would also be optimal for $(S_1,S_2',\ldots,S_N)$. Therefore, $\ell_{S_1D} = \ell^-_{S_1D}$. We have

$$\frac{d}{dS_1}\,[V_k(S_1,S_2',\ldots,S_N;D) - V_k(S_1,S_2,\ldots,S_N;D)]$$

$$= \frac{d}{dS_1}\,[\ell_{S_2'D} + V_{k-1}(S_1,D,\ldots,S_N) - \ell_{S_1D} - V_{k-1}(D,S_2,\ldots,S_N)]$$

$$= \lim_{\ell^+_{S_1S_1'}\to 0} \frac{V_{k-1}(S_1',D,\ldots,S_N) - \ell_{S_1'D} - V_{k-1}(S_1,D,\ldots,S_N) + \ell_{S_1D}}{\ell_{S_1S_1'}}$$

$$= \lim_{\ell^+_{S_1S_1'}\to 0} \frac{V_{k-1}(S_1',D,\ldots,S_N) - V_{k-1}(S_1,D,\ldots,S_N) - \ell_{S_1S_1'}}{\ell_{S_1S_1'}} \le 0,$$

where the inequality follows from Lemma 1.

Case 3: $S = S_1$, $S' \ne S_1, S_2'$.

The proof is similar to that of Case 2.

Case 4: $S = S_2$, $S' = S_1$.

If D is such that $\ell_{S_1D} = \ell^-_{S_1D}$ (see Fig. 7), then by part (ii) the server at S_1 would also be optimal for $(S_1,S_2,\ldots,S_N)$. Therefore, $\ell_{S_1D} = \ell^+_{S_1D}$. We have

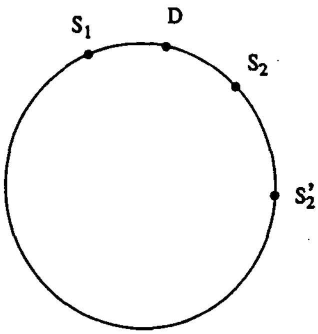

FIGURE 6.

$$\frac{d}{dS_1}\left[V_k(S_1,S_2',\ldots,S_N;D)-V_k(S_1,S_2,\ldots,S_N;D)\right]$$

$$=\frac{d}{dS_1}\left[\ell_{S_1D}+V_{k-1}(D,S_2',\ldots,S_N)-\ell_{S_2D}-V_{k-1}(S_1,D,\ldots,S_N)\right]$$

$$=\lim_{\ell^{+}_{S_1S_1'}\to 0}\frac{\ell_{S_1'D}-V_{k-1}(S_1',D,\ldots,S_N)-\ell_{S_1D}+V_{k-1}(S_1,D,\ldots,S_N)}{\ell_{S_1S_1'}}$$

$$=\lim_{\ell^{+}_{S_1S_1'}\to 0}\frac{V_{k-1}(S_1,D,\ldots,S_N)-V_{k-1}(S_1',D,\ldots,S_N)-\ell_{S_1S_1'}}{\ell_{S_1S_1'}}\le 0,$$

where the inequality follows from Lemma 1.

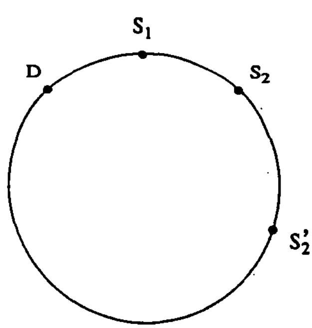

FIGURE 7.

Case 5: $S = S_2$, $S' = S_2'$.

$$\frac{d}{dS_1}[V_k(S_1,S_2',\dots,S_N;D) - V_k(S_1,S_2,\dots,S_N;D)]$$

$$= \frac{d}{dS_1}[\ell_{S_2'D} + V_{k-1}(S_1,D,\dots,S_N) - \ell_{S_2D} - V_{k-1}(S_1,D,\dots,S_N)] = 0.$$

Case 6: $S = S_2$, $S' \neq S_1, S_2'$.

D is between S^- and S^+ because the server at S_2 is optimal for $(S_1,S_2,\dots,S_N)$. If D is between S_2' and S^+ (see Fig. 8), then by part (ii) the server at S_2' would be optimal for $(S_1,S_2',\dots,S_N)$. Therefore, D is between S^- and S_2' and the server at S^- is optimal for $(S_1,S_2',\dots,S_N)$. We have

$$\frac{d}{dS_1}[V_k(S_1,S_2',\dots,S_N;D) - V_k(S_1,S_2,\dots,S_N;D)]$$

$$= \frac{d}{dS_1}[\ell_{S^-D} + V_{k-1}(S_1,D,S_2',\dots,S_N) - \ell_{S_2D} - V_{k-1}(S_1,S^-,D,\dots,S_N)]$$

$$\leq 0,$$

where the inequality follows from Eq. (10).

Case 7: $S \neq S_1, S_2$, $S' = S_1$.

The proof is similar to that of Case 4.

Case 8: $S \neq S_1, S_2$, $S' = S_2'$.

D is between S^- and S^+ because the server at S_2' is optimal for $(S_1,S_2',\dots,S_N)$. If D is between S^- and S_2 (see Fig. 8), then by part (ii) the server at S_2 would be optimal for $(S_1,S_2,\dots,S_N)$. Therefore, D is between S_2 and S^+ and the server at S^+ is optimal for $(S_1,S_2,\dots,S_N)$. We have

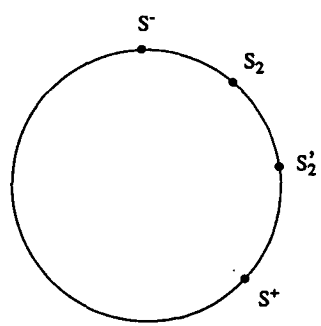

FIGURE 8.

$$\frac{d}{dS_1}[V_k(S_1,S_2',\ldots,S_N;D) - V_k(S_1,S_2,\ldots,S_N;D)]$$

$$= \frac{d}{dS_1}[\ell_{S_2'D} + V_{k-1}(S_1,D,S^+,\ldots,S_N) - \ell_{S^+D} - V_{k-1}(S_1,S_2,D,\ldots,S_N)]$$

$$\leq 0,$$

where the inequality follows from Eq. (10).

Case 9: $S \neq S_1, S_2$, $S' \neq S_1, S_2'$.

If D is between S^-, S_2 or S_2', S^+, the servers at S^- and S^+ are, respectively, optimal for both $(S_1, S_2, \ldots, S_N)$ and $(S_1, S_2', \ldots, S_N)$. Then the result follows directly from the induction hypothesis. If D is between S_2, S_2', the server at S^+ is optimal for $(S_1, S_2, \ldots, S_N)$ and the server at S^- is optimal for $(S_1, S_2', \ldots, S_N)$. We have

$$\frac{d}{dS_1}[V_k(S_1,S_2',\ldots,S_N;D) - V_k(S_1,S_2,\ldots,S_N;D)]$$

$$= \frac{d}{dS_1}[\ell_{S^-D} + V_{k-1}(S_1,D,S_2',S^+,\ldots,S_N)$$

$$- \ell_{S^+D} - V_{k-1}(S_1,S^-,S_2,D,\ldots,S_N)] \leq 0,$$

where the inequality follows from Eq. (10).

Consider now the case when D is not between S^- and S^+. If either the server at S_ℓ or the server at S_r is optimal for both $(S_1, S_2, \ldots, S_N)$ and $(S_1, S_2', \ldots, S_N)$, the result follows directly from the induction hypothesis. The case when the server at S_ℓ is optimal for $(S_1, S_2, \ldots, S_N)$ and the server at S_r is optimal for $(S_1, S_2', \ldots, S_N)$ contradicts part (ii). Therefore, the only remaining possibility is the server at S_r to be optimal for $(S_1, S_2, \ldots, S_N)$ and the server at S_ℓ to be optimal for $(S_1, S_2', \ldots, S_N)$. We have

$$\frac{d}{dS_1}[V_k(S_1,S_2',\ldots,S_N;D) - V_k(S_1,S_2,\ldots,S_N;D)]$$

$$= \frac{d}{dS_1}[\ell_{S_\ell D} + V_{k-1}(S_1,D,S_r,S_2',\ldots,S_N) - \ell_{S_r D}$$

$$- V_{k-1}(S_1,S_\ell,D,S_2,\ldots,S_N)] \leq 0,$$

where the inequality follows from Eq. (10). ■

Remark: Consider the problem where the stations are located along an interval. The problem for the interval can be formulated as a special case of that for the circle as follows. Assume that the N servers are initially located within a certain semicircle and the probability distribution of the requests is such that they always occur on that semicircle. It is then clear that it is not optimal for any server to move outside the semicircle. Therefore, the results obtained in this section for the circle apply to the interval problem as well.

3. GENERALIZATIONS

3.1. Sequencing on a Nonuniform Circle

We consider the problem where it is more expensive to traverse some portions of the circumference of the circle than others. With each location R on the circle, we associate a cost function $p(R)$, so that the cost of moving a server from point A to point B in the clockwise and counterclockwise directions is

$$c_{AB}^{+} = \oint_{A}^{B} p(R)\, dR, \quad \text{and} \quad c_{AB}^{-} = \oint_{B}^{A} p(R)\, dR,$$

respectively, where $\oint_A^B$ denotes the line integral from A to B in the clockwise direction.

A cost defined by the preceding expressions can be interpreted as a generalized distance between two points on the circle. The results of Section 2 apply in this case as well with the notion of the actual distance replaced with that of the generalized distance. Properties of the optimal policy are given in the following theorem.

THEOREM 4: *Consider a request located at point D, and let S^-, S^+ be the positions of the servers closest to D in the counterclockwise and clockwise directions, respectively. Then, we have the following:*

(i) *It is optimal to serve the request either with the server at S^- or the server at S^+.*

(ii) *There exists a threshold point T such that it is optimal to serve the request with the server at S^- if $c_{S^-D}^{+} \leq c_{S^-T}^{+}$ and with the server at S^+ otherwise.*

(iii) *If it is optimal to serve the request with the server at S^- when the servers are located at positions $S_1, S_2, \ldots, S_N$, then it is optimal to serve it with the same server when any server is moved in the clockwise direction at a distance small enough that the order around the circle is not changed.*

PROOF: The proof follows identical steps to the proofs of Theorems 2 and 3. ∎

3.2. Nonuniform Server Sequencing

We consider the problem where some servers are more expensive to move than others. We assume that we have N servers $S_1, S_2, \ldots, S_N$ and a cost c_i, $i = 1,2,\ldots,N$, associated with server S_i. The cost of serving a request with server S_i, $i = 1,2,\ldots,N$, is equal to $c_i \ell$, where ℓ is the distance traveled by S_i.

It is easy to show by example that it is not necessarily optimal to serve a request with one of its two adjacent servers. This fact considerably complicates the problem of determining the optimal server for each request. One approach to solving the problem is by pairwise comparisons of servers. Using this approach, we get the following property of the optimal policy.

THEOREM 5: *Consider two servers at positions S_1 and S_2 and a request at point D. Let S_1' and S_2' be the diametrically opposite points of S_1 and S_2, respectively. Then, we have the following:*

(i) *There exists a threshold point T between S_1 and S_2 (see Fig. 9) such that it is optimal to serve the request with the server at S_1 instead of the one at S_2 if D is between S_1 and T and vice versa if it is between T and S_2.*

(ii) *There exists a threshold point T' between S_1' and S_2' (see Fig. 9) such that it is optimal to serve the request with the server at S_1 instead of the one at S_2 if D is between S_2' and T' and vice versa if it is between T' and S_1'.*

PROOF: The proof is along the lines of that of part (ii) of Theorem 2. ■

We have been unable to obtain any results for requests located between S_1', S_2 or S_2, S_1'. For this case, we have the following conjecture.

Conjecture: Assume that the server at S_1 is more expensive to move than the server at S_2. Then, we have the following:

(i) It is optimal to serve a request between S_2 and S_1' with the server at S_2 instead of the server at S_1.

(ii) There exists a threshold T'' in the semicircle $S_1 S_2' S_1'$ such that it is optimal to serve the request with the server at S_1 instead of the one at S_2 if D is between S_1 and T'' and vice versa if D is between T'' and S_1'.

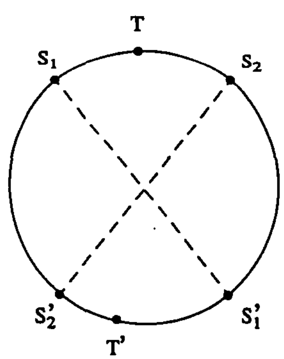

FIGURE 9.

Remark: Part (ii) of the conjecture does not contradict part (ii) of Theorem 5 for the following reason. If T'' is between S_1 and S_2', then T' coincides with S_2'. On the other hand, if T'' is between S_2' and S_1', then T'' coincides with T'.

3.3. Sequencing on a Sphere

We consider the problem where the service stations are located on a sphere. Calderbank et al. [2] showed for two-server systems that when the requests are uniformly distributed on the sphere it is optimal to serve them with the nearest server. In this section, we provide a different proof of that result. We begin by noting that the shortest path between two points on a sphere lies on the great circle defined by those two points. We denote by ℓ_{AB} the length of the shortest path on the sphere from point A to point B. We also define $V_n(\cdot)$ and $V_n^\pi(\cdot)$ as in the beginning of Section 2.2. The following lemma holds.

LEMMA 2: *For any n, we have*

$$|V_n(S_1,S_2) - V_n(S_1',S_2)| \leq \ell_{S_1 S_1'}.$$

PROOF: Let π be the optimal policy for (S_1,S_2) when we have n requests to process. We construct policy $\tilde{\pi}$ for (S_1',S_2) as follows: $\tilde{\pi}$ is identical to π except the first time π serves a request with the server at S_1. Then, $\tilde{\pi}$ serves the request by moving the server at S_1' to S_1 and then to the location of the request. Therefore,

$$V_n(S_1',S_2) \leq V_n^{\tilde{\pi}}(S_1',S_2) = V_n(S_1,S_2) + \ell_{S_1 S_1'}, \tag{13}$$

where the inequality follows from the fact that $\tilde{\pi}$ is not necessarily optimal for (S_1',S_2). By interchanging S_1 and S_1' in Eq. (13), we get Lemma 2. ∎

Lemma 2 can be used to prove the basic result of Section 3.3 that is given by the following theorem.

THEOREM 6: *For two-server systems and requests uniformly distributed on the sphere, it is optimal to serve any request with the nearest server.*

PROOF: Let S_1,S_2 be the locations of the two servers and D the location of the request. Assume that $\ell_{S_2 D} < \ell_{S_1 D}$. Then, there exists a point on arc $S_1 D$ such that $\ell_{S_2' D} = \ell_{S_2 D}$. Let π_1 and π_2 be the policies that serve the request at D with the server at S_1 and S_2, respectively, and proceed optimally afterward. For any n, we have

$$V_n^{\pi_1}(S_1,S_2;D) = \ell_{S_1 D} + V_{n-1}(S_2,D) = \ell_{S_1 D} + V_{n-1}(S_2',D), \tag{14}$$

$$V_n^{\pi_2}(S_1,S_2;D) = \ell_{S_2 D} + V_{n-1}(S_1,D) = \ell_{S_2' D} + V_{n-1}(S_1,D), \tag{15}$$

where the second equality in Eq. (14) follows from the fact that, because the requests are uniformly distributed on the sphere, the expected cost depends only on the distance between the two servers and not their actual positions. From Eqs. (14) and (15), we get

$$V_n^{\pi_1}(S_1,S_2;D) - V_n^{\pi_2}(S_1,S_2;D) = \ell_{S_1S_2'} + V_{n-1}(S_2',D) - V_{n-1}(S_1,D) \geq 0,$$

where the inequality follows from Lemma 2. Therefore, it is optimal to serve the request with the nearest server. ■

4. CONCLUSIONS

The main results of the paper present qualitative properties of optimal multi-server sequencing strategies on the circle and the interval. For systems with identical servers, these strategies utilize the server immediately to the right or immediately to the left of a request and are described by thresholds that depend on the location of all servers, the number of the remaining requests, and their distribution. The optimal thresholds possess a monotonicity property; their computation remains a challenging and formidable task. We believe that our analysis does provide a useful insight into the problem of designing sequencing algorithms for the classes of problems mentioned in the introduction.

Acknowledgment

The authors are grateful to the anonymous reviewer whose excellent comments helped to substantially improve the results and the presentation of this paper.

Notes

1. In Figure 4, the set of such points are the arc S^-T^+ traversed in the clockwise direction, and the arc S^+T^- traversed in the counterclockwise direction, where T^- and T^+ are the diametrically opposite points of S^- and S^+, respectively. Such a situation does not arise in Figure 3.
2. In Figure 4, the set of such points is the arc T^+T^- traversed in the clockwise direction. In Figure 3, the set of such points is the arc S^-S^+ traversed in the clockwise direction.

References

1. Birnbaum, J.S. (1982). Computers: A survey of trends and limitations. *Science* 215: 760–765.
2. Calderbank, A.R., Coffman, E.G., Jr., & Flatto, L. (1985). Sequencing problems in two-server systems. *Mathematics of Operations Research* 10: 585–598.
3. Calderbank, A.R., Coffman, E.G., Jr., & Flatto, L. (1985). Sequencing two servers on a sphere. *Stochastic Models* 1: 17–28.
4. Calderbank, A.R., Coffman, E.G., Jr., & Flatto, L. (1988). Optimal directory placement on disk storage devices. *Journal of the Association for Computing Machinery* 35: 433–446.
5. Hofri, M. (1983). Should the two-headed disk be greedy? — Yes, it should. *Information Processing Letters* 16: 83–85.