NATIONAL BUREAU OF STANDARDS REPORT

9651

ALLOCATING SERVICE PERIODS TO MINIMIZE DELAY TIME

by

W.A. Horn

Technical Report

to

Northeast Corridor Transportation Project

NBS

U.S. DEPARTMENT OF COMMERCE
NATIONAL BUREAU OF STANDARDS

THE NATIONAL BUREAU OF STANDARDS

The National Bureau of Standards[1] provides measurement and technical information services essential to the efficiency and effectiveness of the work of the Nation's scientists and engineers. The Bureau serves also as a focal point in the Federal Government for assuring maximum application of the physical and engineering sciences to the advancement of technology in industry and commerce. To accomplish this mission, the Bureau is organized into three institutes covering broad program areas of research and services:

THE INSTITUTE FOR BASIC STANDARDS . . . provides the central basis within the United States for a complete and consistent system of physical measurements, coordinates that system with the measurement systems of other nations, and furnishes essential services leading to accurate and uniform physical measurements throughout the Nation's scientific community, industry, and commerce. This Institute comprises a series of divisions, each serving a classical subject matter area:

—Applied Mathematics—Electricity—Metrology—Mechanics—Heat—Atomic Physics—Physical Chemistry—Radiation Physics—Laboratory Astrophysics[2]—Radio Standards Laboratory,[2] which includes Radio Standards Physics and Radio Standards Engineering—Office of Standard Reference Data.

THE INSTITUTE FOR MATERIALS RESEARCH . . . conducts materials research and provides associated materials services including mainly reference materials and data on the properties of materials. Beyond its direct interest to the Nation's scientists and engineers, this Institute yields services which are essential to the advancement of technology in industry and commerce. This Institute is organized primarily by technical fields:

—Analytical Chemistry—Metallurgy—Reactor Radiations—Polymers—Inorganic Materials—Cryogenics[2]—Office of Standard Reference Materials.

THE INSTITUTE FOR APPLIED TECHNOLOGY . . . provides technical services to promote the use of available technology and to facilitate technological innovation in industry and government. The principal elements of this Institute are:

—Building Research—Electronic Instrumentation—Technical Analysis—Center for Computer Sciences and Technology—Textile and Apparel Technology Center—Office of Weights and Measures—Office of Engineering Standards Services—Office of Invention and Innovation—Office of Vehicle Systems Research—Clearinghouse for Federal Scientific and Technical Information[3]—Materials Evaluation Laboratory—NBS/GSA Testing Laboratory.

---

[1] Headquarters and Laboratories at Gaithersburg, Maryland, unless otherwise noted; mailing address Washington, D. C., 20234.

[2] Located at Boulder, Colorado, 80302.

[3] Located at 5285 Port Royal Road, Springfield, Virginia 22151.

NATIONAL BUREAU OF STANDARDS REPORT

NBS PROJECT	NBS REPORT
2058456	9651

ALLOCATING SERVICE PERIODS TO MINIMIZE DELAY TIME

by

W.A. Horn

Technical Report

to

Northeast Corridor Transportation Project

U.S. DEPARTMENT OF COMMERCE
NATIONAL BUREAU OF STANDARDS

ALLOCATING SERVICE PERIODS TO MINIMIZE DELAY TIME

W.A. Horn

National Bureau of Standards

ABSTRACT

Consider a facility which must divide its services, during the time interval $[0,T]$, among N streams of arrivals. The problem treated is that of finding a pattern of service which minimizes total delay to the members of the streams, taking into account the "dead time" which begins each service period. For each stream, it is required that final queue size equal initial size, and that the queue be empty sometime in $[0,T]$. Conditions for feasibility of solutions are given in the case where the instantaneous service rates are bounded above by known constants. In the event that all streams have constant arrival rates and are to be served the same number of times, an optimal service pattern is derived using a recent result of R. Rangarajan and R.M. Oliver.

Key words

Transportation theory, queueing theory, traffic flow, switching theory, scheduling, allocation.

ALLOCATING SERVICE PERIODS TO MINIMIZE DELAY TIME

W.A. Horn

National Bureau of Standards

1. PROBLEM STATEMENT

This paper presents a further investigation into the type of problem studied in [1]. The problem pertains to the allocation of servicing times among several incoming streams which require "processing" of some kind by a single "server" capable of handling only one stream at a time. The server might for example be a switching point or a congestion point (e.g., a tunnel entrance) in a transport network, in which case "serving" a stream simply means permitting passage to its flow. Or, the server might be a computer handling reservations being arranged at several points, or exercising control on a real-time basis over operations along several links.

The time period during which servicing occurs is assumed to be $[0,T]$, where $T > 0$. The rate at which "customers" arrive at stream i is assumed to be the known continuous function $a_i(t)$, while the outflow or service-rate function $s_i(t)$, which will be defined below, is bounded above by the constant $C_i > 0$, the capacity when servicing stream i . It is further assumed that $a_i(t)$ crosses the level C_i at most a finite number of times in $[0,T]$. That is, $a_i^{-1}(C_i)$ is a set with a finite number of connected components.

The size of the waiting queue in stream i at time t is designated by $Q_i(t)$, and we let $q_i = Q_i(0)$, where it is assumed that each q_i is non-negative. Then clearly

$$(1) \qquad Q_i(t) = q_i + \int_0^t (a_i(\tau) - s_i(\tau))d\tau .$$

A cumulative waiting time function $W_i(t)$ is defined for each i by

$$W_i(t) = \int_0^t Q_i(\tau)d\tau ,$$

and a total waiting time function $W(t)$ is defined by

$$W(t) = \sum_{i=1}^{N} W_i(t) ,$$

where N is the number of streams.

Since $s_i(t)$ will depend on the times at which servicing actually occurs, it is clear that W_i and W also depend on these variables. Furthermore, W_i and W depend on the quantities q_i. These functional dependencies may be expressed explicitly when convenient.

Next we define $s_i(t)$. As a preliminary, we further limit the generality of the problem by assuming that a given stream i is serviced, within $[0,T]$, only during a finite set of closed intervals $\{[x_{ij},y_{ij}]\}_{j=1}^{m(i)}$ where $m(i)$ is the total number of such intervals of service to stream i. No two intervals of servicing, for any streams, may overlap except at their endpoints, and the set of all service intervals for all streams covers $[0,T]$. (The endpoints of these intervals comprise the <u>switching pattern</u>.) Furthermore, for some set of positive constants $\{d_i\}$ called the "dead" times, it is assumed that $x_{ij} + d_i \le y_{ij}$, for $i = 1,2,\ldots,N$, and $j = 1,2,\ldots,m(i)$.

If stream i is not being serviced at time t, then $s_i(t)$ is defined to be 0. If t lies in the service interval $[x_{ij},y_{ij}]$, then $s_i(t)$ is defined by

$$s_i(t) = \begin{cases} 0 : t \in [x_{ij}, x_{ij}+d_i], \\ C_i : t \in (x_{ij}+d_i, y_{ij}] \text{ and } Q_i(t) > 0, \\ \min(C_i, a_i(t)) : t \in (x_{ij}+d_i, y_{ij}] \text{ and } \\ \quad Q_i(t) \le 0 . \end{cases}$$

[It is clear from (1) that Q_i depends on s_i, whereas the above definition states that s_i depends on Q_i, at least on the intervals $(x_{ij}+d_i, y_{ij}]$. Thus it is necessary to show that there exist unique functions s_i and Q_i which satisfy (1) and the equation for s_i. This will not be detailed here, although it is quite simple due to the particular form of $a_i(t)$. Briefly, the proof consists of constructing the unique s_i and Q_i by breaking up each interval $(x_{ij}+d_i, y_{ij}]$ into subintervals where either $a_i(t) \leq C_i$ or $a_i(t) > C_i$. The functions s_i and C_i are then defined in the natural way and easily shown to be unique.]

The above definition of s_i, together with the fact that $q_i \geq 0$, assures that $Q_i(t) \geq 0$ for all $t \in [0,T]$, since if $Q_i(t) < 0$ for some t then there would exist $t' < t$ such that $Q_i(t') = 0$, by the continuity of Q_i. Let t_0' be the greatest such t'. Then $Q_i(r) \leq 0$ for all r such that $t_0' \leq r \leq t$, and so $a_i(r) - s_i(r) \geq 0$ in this interval, by the definition of s_i. This contradicts the fact that

$$0 > Q_i(t) = Q_i(t_0') + \int_{t_0'}^{t} (a_i(r) - s_i(r))dr .$$

The general problem to be considered is that of finding a finite set of intervals $[x_{ij}, y_{ij}]$ of service, as defined and restricted above, such that the total waiting time $W(T)$ is a minimum, given the initial values q_i.

We will consider a variant of the general problem in this paper by introducing the restrictions (for $i=1,2,\ldots,N$)

$$(2) \qquad q_i = Q_i(0) = Q_i(T) ,$$

$$(3) \qquad Q_i(t) = 0 \quad \text{for some} \quad t \in [0,T] ,$$

and by allowing the minimization not only over the intervals of servicing but also over the parameters q_i as well. That is, the q_i will not be considered fixed but variable, subject to (2) and (3). Constraints (2) and (3) are called the feasibility constraints, and a solution is feasible if it satisfies them.

Imposing constraint (2) has the property of making the problem periodic, in the following sense. If we suppose that a_i is defined for all $t \geq 0$, rather than for $t \in [0,T]$, with $a_i(t+T) = a_i(t)$, and that the switching pattern is extended in such a way that stream i is being serviced at time $t+T$ if and only if it is also being serviced at time t, then $Q_i(0) = Q_i(T)$ implies

$$Q_i(t+T) = Q_i(t) .$$

It is also seen from this that

$$W_i(T+t) - W_i(t) = W_i(T) ,$$

or,

$$W_i(nT+\theta) = nW_i(T) + W_i(\theta) .$$

This in turn implies that

$$W(t)/t \to W(T)/T \quad \text{as} \quad t \to \infty ,$$

so that, in a sense, minimizing $W(T)$ and thus $W(T)/T$ is equivalent to minimizing the "long-term average delay". The motivation for this restriction is that the applications we have in mind refer to ongoing systems rather than isolated occurrences. It seems of little practical use (except for emergency evacuation operations and the like) to formulate the problem as if what happened after time T were of no concern. Condition (2), at least for periodic arrival patterns, implies stability in the sense of "repeatability" for the situation, in particular ruling out unbounded growth of queues over the long run. It might prove worthwhile to investigate the problem variant in which (2) is replaced by

$$(2') \qquad Q_i(T) \leq Q_i(0) ,$$

but this version will not be studied here.

2. FEASIBILITY AND CONSISTENCY CONDITIONS

So far it is not known whether there exist feasible solutions of the problem of section 1. This section will be devoted to finding necessary and sufficient conditions for the existence of such solutions.

Since we are considering the problem where not only the points of switching are allowed to vary, but also the q_i, the first question which arises is how much freedom the q_i have, for a given switching pattern, so that conditions (2) and (3) may still be satisfied. The following two lemmas answer this.

LEMMA 1. Let Q_i and Q_i' be the queue-size functions associated with the initial values q_i and q_i', respectively, and having the same switching pattern. Then $|Q_i(t) - Q_i'(t)|$ is a non-increasing function of t.

PROOF. First, $Q_i - Q_i'$ does not change sign. For if $Q_i(a) - Q_i'(a) > 0$ and $Q_i(b) - Q_i'(b) < 0$, where $a < b$, then by continuity $Q_i(t) = Q_i'(t)$ for some $t \epsilon (a,b)$. But then $Q_i(\tau) = Q_i'(\tau)$ for all $\tau \geq t$, contradicting $Q_i(b) < Q_i'(b)$.

Now suppose (say) $Q_i(0) > Q_i'(0)$. Then $Q_i(t) - Q_i'(t) = |Q_i(t) - Q_i'(t)|$, by the above. But

$$Q_i(t) - Q_i'(t) = q_i - q_i' + \int_0^t (s_i'(\tau) - s_i(\tau))d\tau ,$$

and since $s_i'(\tau) \leq s_i(\tau)$

for all $\tau \in [0,T]$ because $Q_i \geq Q_i'$, we have that $|Q_i(t) - Q_i'(t)|$ is non-increasing. Similarly if $Q_i(0) < Q_i'(0)$ or $Q_i(0) = Q_i'(0)$.

LEMMA 2. Let Q_i and Q_i' be two queue-size functions associated with the same switching pattern during $[0,T]$ and both satisfying (2). Then Q_i and Q_i' differ by a constant. Hence there exists at most one feasible solution for each switching pattern.

PROOF. From lemma 1,

$|Q_i(0) - Q_i'(0)| \geq |Q_i(t) - Q_i'(t)| \geq |Q_i(T) - Q_i'(T)|$ for all $t \in [0,T]$. But by (2), $|Q_i(0) - Q_i'(0)| = |Q_i(T) - Q_i'(T)|$. Hence, by the continuity of the queue-size functions, $Q_i(t) - Q_i'(t) = q_i - q_i'$, a constant.

Now if two solutions having the same switching pattern are both feasible, then they satisfy (3) in addition to (2). But if $q_i > q_i'$, for example, then $Q_i(t) > Q_i'(t) \geq 0$, by the above, and Q_i is not feasible. Therefore $q_i = q_i'$ and so $Q_i = Q_i'$.

The next question to be treated is which patterns of switching times admit feasible solutions.

THEOREM 3. Suppose that stream i is serviced during successive intervals of length $L_{i1}, L_{i2}, \ldots, L_{im(i)}$, and let d_i be the dead time for stream i. Then a necessary and sufficient condition that a set of q_i exist for which the given switching pattern produces a feasible solution is

$$C_i \sum_{j=1}^{m(i)} L_{ij} \geq \int_0^T a_i(t)dt + m(i)C_i d_i$$

for all i.

PROOF. Let the above inequality be satisfied. It is clear that, when q_i is sufficiently large, $Q_i(t) > 0$ for all $t \in [0,T]$. Then, by definition, $s_i(t) = 0$ during non-service intervals and during a part d_i of each service interval, and $s_i(t) = C_i$ otherwise. Thus for such q_i,

$$\int_0 s_i(t)dt = C_i \sum_{j=1}^{m(i)} (L_{ij} - d_i)$$

$$= C_i \sum_{j=1}^{m(i)} L_{ij} - C_i m(i) d_i$$

$$\geq \int_0^T a_i(t)dt,$$

so that

$$Q_i(T) = q_i + \int_0^T (a_i(t) - s_i(t))dt$$

$$\leq q_i .$$

On the other hand, for $q_i = 0$ it is clear that $Q_i(T) \geq q_i$. But for a fixed switching pattern, $Q_i(T)$ is a continuous function of q_i. (In fact, by lemma 1, $|Q_i(T) - Q_i'(T)| \leq |q_i - q_i'|$.) Thus there exists a q_i for which $Q_i(T) = q_i$.

Let q_i^0 be the infimum of all such q_i satisfying $q_i = Q_i(T)$ for the given pattern of switching. Let

$$b = \inf \{Q_i(t): t \in [0,T] , Q_i(0) = q_i^0\} .$$

If $b=0$, then q_i^0 gives a feasible solution. But if $b > 0$, then using the value $q_i = q_i^0 - b/2$ as the initial queue size, we find that $Q_i(t) \geq b - b/2 = b/2$ for all $t \in [0,T]$, by lemma 1. Thus by the definition of s_i we have the same value of $s_i(t)$, for each $t \in [0,T]$, for the two initial values q_i^0 and $q_i^0 - b/2$. But this implies that the solution with initial value $q_i^0 - b/2$ also satisfies (2), contradicting the definition of q_i^0.

This proves sufficiency.

Necessity follows from (2), via the inequality

$$\int_0^T a_i(t)dt = \int_0^T s_i(t)dt$$

$$\leq C_i \sum_{j=1}^{m(i)} (L_{ij} - d_i) .$$

LEMMA 4. Suppose there exist numbers $T_i > 0$ such that

$$\sum_{i=1}^{N} T_i = T$$

and

$$C_i T_i \geq \int_0^T a_i(t)dt + m(i)C_i d_i .$$

Then there exists a feasible solution for which stream i is serviced exactly $m(i)$ times (some of which may be consecutive), and there exists an optimal such solution relative to the particular $m(i)$'s .

PROOF. By Theorem 3, if $\{L_{ij}\}$ is a set satisfying

$$\sum_{j=1}^{m(i)} L_{ij} = T_i ,$$

where each $L_{ij} \geq d_i$, then L_{ij} represents a feasible solution. By lemma 2, q_i is uniquely determined by the switching pattern. But the vector of switching times is a point of

$$[0,T]^{\Sigma m(i)}$$

a compact set. The conditions(given in Theorem 2) defining those switching patterns which correspond to feasible solutions are linear inequalities in the L_{ij}'s , and so in the switching times themselves;

thus the minimization is to take place over a closed subset of the compact set.

Now it may be shown that $W(T)$ depends continuiously on the switching times. In fact, $W(T)$ is uniformly continuous in the vector of switching times. (The proof, which is lengthy but not difficult, will not be given here. However, briefly, the following argument is used. Let $\{(x_{ij},y_{ij})\}$ and $\{(x'_{ij},y_{ij})\}$ be two feasible switching patterns which differ, for any stream i, in at most one term. If the first pattern, for example, results in less service to stream i, let q_i be the initial value for $Q_i(t)$ associated with this pattern. Apply the second pattern to this starting value, obtaining a new $Q_i(T) = q'_i$. It can be shown that this q'_i is the initial value of the i-th queue associated with the second pattern.)

Since $W(T)$ depends continuously on the switching times, which in turn range over a compact set, an optimal solution must exist.

The last result permits us to state explicit conditions for the existence of an optimal solution in the simplest case, namely for constant a_i.

COROLLARY 5. Suppose $a_i(t) = a_i$, a constant. If $\{L_{ij}\}$ are the service interval lengths of a switching pattern such that

$$(4) \qquad \sum_{j=1}^{m(i)} L_{ij} \geq m(i)d_i + a_i T/C_i \ ,$$

then the pattern is feasible. Furthermore, if numbers $T_i > 0$ exist such that

$$\sum_{i=1}^{N} T_i = T$$

and

$$T_i \geq m(i)d_i + a_i T/C_i \ ,$$

then an optimal feasible solution exists.

PROOF. From Theorem 3 and Lemma 4.

Now let $\rho_i = a_i/C_i$ and $d = \sum_{i=1}^{N} m(i)d_i$. Then we have the following.

THEOREM 6. For constant a_i , the following conditions are necessary and sufficient for the existence of a feasible solution with $m(i)$ servicings of stream i which is optimal with respect to the given set $\{m(i)\}$:

$$(5) \qquad \sum_{i=1}^{N} \rho_i < 1 \ .$$

(6) $$T \geq \frac{d}{1 - \sum_{i=1}^{N} \rho_i}$$

PROOF. If $\{L_{ij}\}$ is feasible, by Lemma 5 it must satisfy (4). Summing (4) over all i, we get

$$T \geq \sum_{i=1}^{N} m(i)d_i + T\sum_{i=1}^{N} (a_i/C_i)$$

(7) $$= d + T \sum_{i=1}^{N} \rho_i ,$$

or equivalently

$$T(1 - \sum_{i=1}^{N} \rho_i) \geq d > 0 .$$

Since $T > 0$, we have $1 - \sum_{i=1}^{N} \rho_i > 0$, proving (5), and also

$$T \geq \frac{d}{1 - \sum_{i=1}^{N} \rho_i} ,$$

proving (6).

Now suppose (5) and (6) are satisfied. Then (7) is also satisfied. Thus it is possible to find $T_i > 0$ such that

$$T_i \geq m(i)d_i + a_i T/C_i \ .$$

The result follows from Corollary 5.

3. OPTIMAL SOLUTIONS FOR CONSTANT ARRIVAL RATES AND EQUAL m(i)'s

This section will deal with the special case in which each $a_i(t)$ is a constant function, and the $m(i)$ are all equal to some common value m. This case was treated in [1], with the further provision that $m=1$, that is, that each stream was serviced exactly once.

A method of finding the optimum service period length for each stream was found in [1], and this is the result to be generalized here. It will be shown that an optimum (not necessarily the optimum) switching pattern is obtained for the case $m(i) = m$, a_i constant, when the interval $[0,T]$ is broken up into m subintervals of equal length, the switching pattern is optimized over the first interval $[0,T/m]$ by the methods of [1] with one servicing per stream, and this pattern is repeated cyclically for each of the other $m-1$ subintervals. In addition to this, an auxiliary result of [1] will be used to find the optimum value of m.

LEMMA 7. Suppose that in the interval $[s,t]$ the function $f(x)$ is either increasing at a rate p or decreasing at a rate $-q$ almost everywhere. Suppose that $f(s) = f(t) = 0$ and that $f(x) \geq 0$ for $x \in (s,t)$. Suppose further that f increases during $[s,s+A_1)$, decreases during $(s+A_1,s+A_1+B_1)$, etc., where each interval of increase of length A_i is followed by an interval of decrease of length B_i , and each interval of decrease of length B_i is followed by an interval of increase of length A_{i+1} , the last intervals being of length A_n and B_n respectively. Then

$$\int_s^t f(x)dx \geq \frac{p(p+q)}{2q} \sum_{i=1}^{n} A_i^2 .$$

PROOF. The proof is geometric in nature and is by induction on n . First, for $n=1$ the inequality reduces to

$$\int_s^t f(x)dx \geq \frac{p(p+q)}{2q} A_1^2 .$$

Referring to figure 1, we see that $\int_s^t f(x)dx$ is the area of the triangle whose altitude is $A_1 p$ and whose base is A_1+B_1 , where

$$B_1 = A_1 p/q .$$

Thus $\int_s^t f(x)dx = \frac{p(p+q)}{2q} A_1^2$, proving the case $n=1$.

Assume that the lemma has been proved for 1,2,...,n-1. Looking at fingure 2, we see that

$$\int_s^t f(x)dx \geq \frac{p(p+q)}{2q} \sum_{i=1}^{n-1} A_i^2 + \text{Area } YWVZX$$

$$\geq \frac{p(p+q)}{2q} \sum_{i=1}^{n-1} A_i^2 + \Delta XYZ$$

$$= \frac{p(p+q)}{2q} \sum_{i=1}^{n-1} A_i^2 + \frac{p(p+q)}{2q} A_n^2$$

$$= \frac{p(p+q)}{2q} \sum_{i=1}^{n} A_i^2 .$$

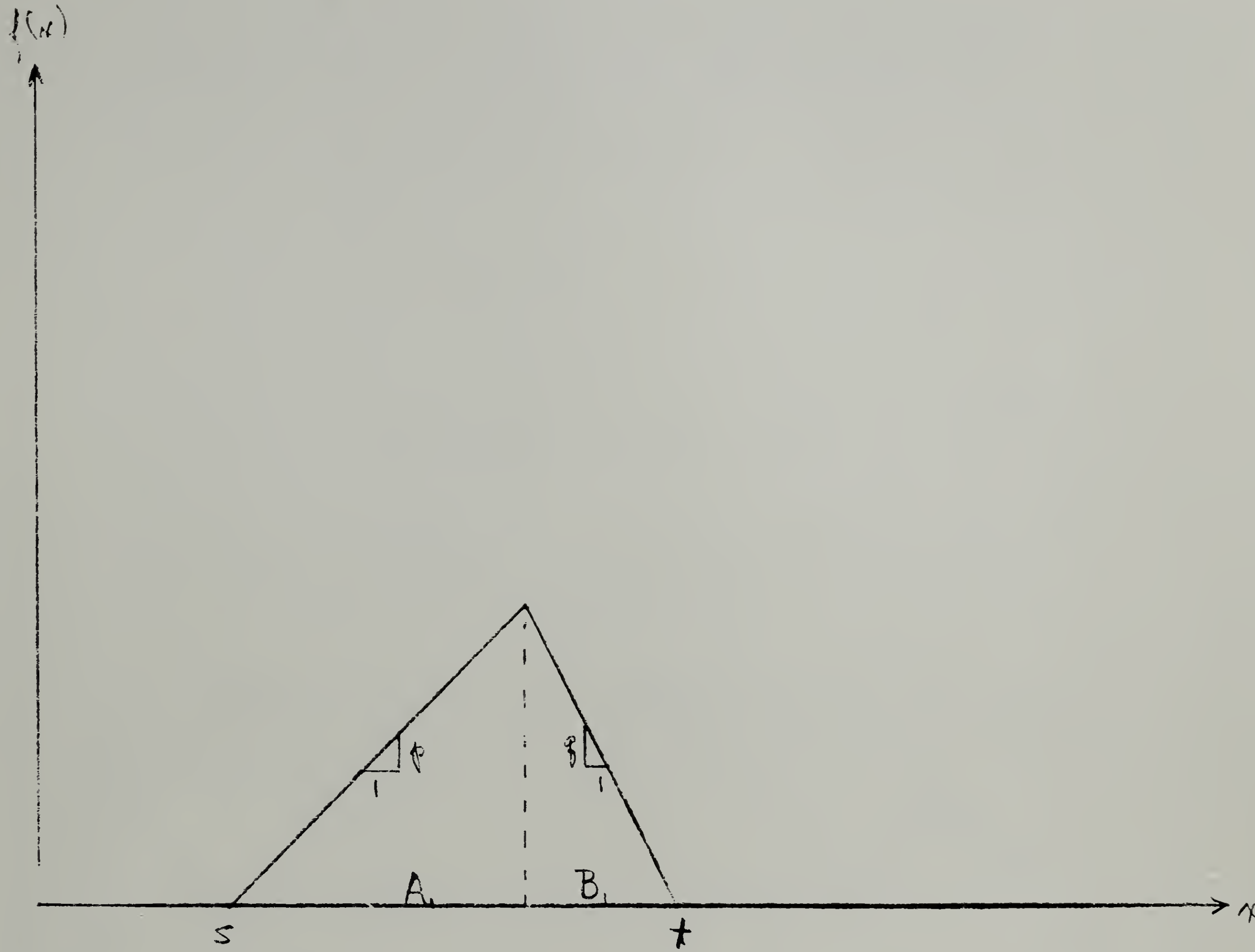

Figure 1

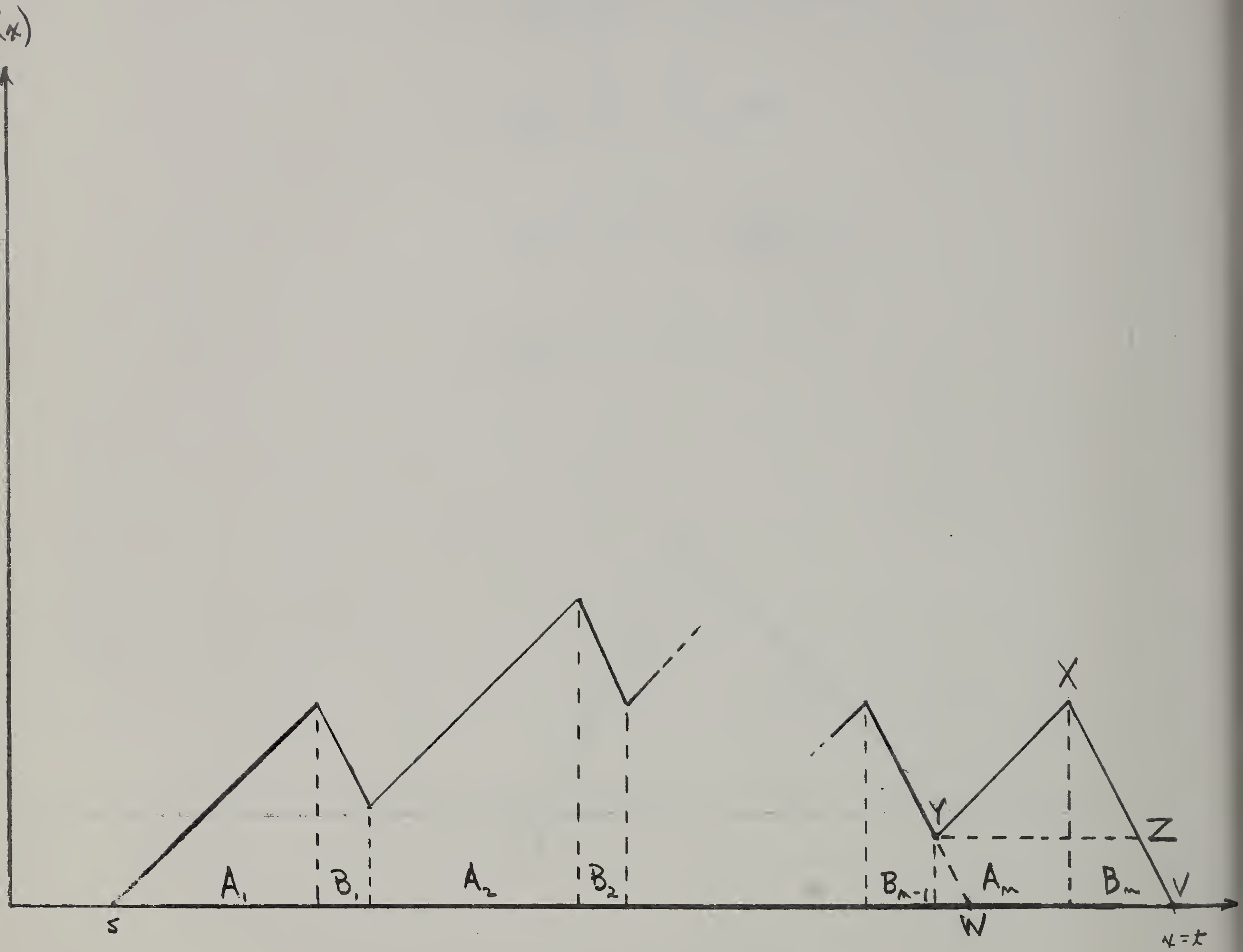

Figure 2

We now state the following notations to be used during the rest of this section. I_{ij} will denote the j-th interval of service for lane i during $[0,T]$. Because of the cyclic character of the problem we also define

$$I_{ik} \equiv I_{ij}$$

where $1 \leq j \leq m$ and $k \equiv j \pmod{m}$. We denote the length of I_{ij} by L_{ij}. The interval contained between I_{ij} and $I_{i,j+1}$ (or the two intervals, if $j = m$) will be denoted by J_{ij} and its length (respectively, the sum of their lengths) by M_{ij}.

LEMMA 8. Let S be a feasible switching pattern for a problem with constant a_i. Then

$$W_i(T) \geq \frac{a_i C_i}{2(C_i - a_i)} \sum_{j=1}^{m(i)} (M_{ij} + d_i)^2 .$$

PROOF. By definition, W_i is the integral of Q_i, which is always non-negative. Thus W_i may be thought of as the area under the curve representing Q_i. Since $C_i > a_i$ in order to satisfy feasibility, Q_i is either rising at a rate a_i (when $s_i=0$), falling at a rate $-(C_i - a_i)$ (when $s_i = C_i$), or constant at 0 (when $s_i = a_i$). We may represent this as in figure 3. It is also clear that the curve is rising only during the time that lane i is not being serviced, plus the dead time, that is, during intervals of length $M_{ij} + d_i$.

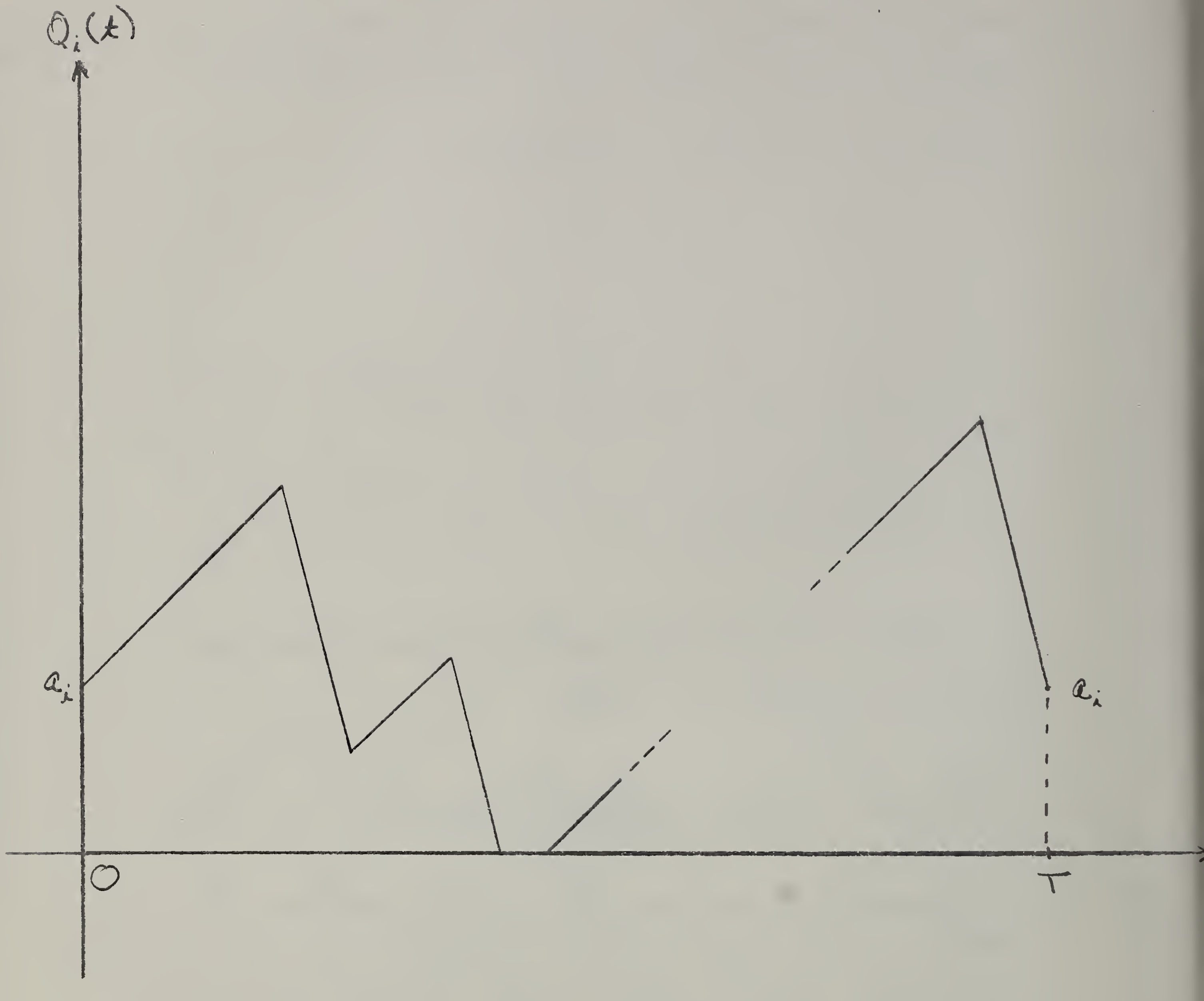

Figure 3

Now let us group the I_{ij} as follows. By condition (3), there exists $t_0 \epsilon [0,T]$ such that $Q_i(t_0) = 0$. Let I_{ij} be the interval of service to lane i to which t_0 belongs. (Obviously t_0 must lie in some interval of service.) Let I_{ik} be the first interval after I_{ij} during which $Q_i(t)$ is again 0. Then the lengths of the intervals of non-service to i between I_{ij} and I_{ik}, namely $J_{ij}, J_{i,j+1}, \ldots, J_{i,k-1}$, will be $M_{ij}, M_{i,j+1}, \ldots, M_{i,k-1}$. If we let z_{iq} represent the right endpoint of any interval I_{iq}, then from lemma 7 we have

$$\int_{z_{ij}}^{z_{ik}} Q_i(t)dt \geq \frac{a_i C_i}{2(C_i - a_i)} \sum_{\nu=j}^{k-1} (M_{i\nu} + d_i)^2 .$$

Summing this inequality over all intervals in which Q_i becomes 0, we get

$$W_i(T) = \int_0^T Q_i(t)dt \geq \frac{a_i C_i}{2(C_i - a_i)} \sum_{j=1}^{m(i)} (M_{ij} + d_i)^2 .$$

This proves the lemma.

THEOREM 9. Suppose that, for constant a_i, $\{I_{ij}\}$ represents an optimal feasible solution, where $m(i) = m$. Then there exists an optimal feasible solution $\{\bar{I}_{ij}\}$ with

$$\bar{L}_{ij} = (1/m) \sum_{k=1}^{m} L_{ik} , \; 1 \leq i \leq n ,$$

where $\bar{L}_{ij}$ is the length of interval $\bar{I}_{ij}$, L_{ij} the length of I_{ij}.

PROOF. Consider the solution defined by $\bar{L}_{ij} = (1/m) \sum_{k=1}^{m} L_{ik}$, where $\bar{I}_{1,1}$ is the first interval and where interval $\bar{I}_{ij}$ is followed by $\bar{I}_{i+1,j}$ for $i < n$, and interval $\bar{I}_{nj}$ is followed by $I_{1,j+1}$. Then we have

$$\bar{M}_{ij} = \bar{M}_i , \quad 1 \leq j \leq n .$$

By lemma 8, the waiting time for the original solution satisfies

$$W_i(T) \geq \frac{a_i C_i}{2(C_i - a_i)} \sum_{j=1}^{m} (M_{ij} + d_i)^2 ,$$

and the Cauchy-Schwartz inequality yields

$$\sum_{j=1}^{m} (M_{ij} + d_i)^2 \geq m(\bar{M}_{ij} + d_i)^2 = \sum_{j=1}^{m} (\bar{M}_{ij} + d_i)^2 .$$

Thus it suffices to show that the new solution (which by corollary 5 is feasible) obeys Lemma 8 with <u>equality</u>, i.e. that

$$\bar{W}_i(T) = \int_0^T \bar{Q}_i(t)dt = \frac{a_i C_i}{2(C_i - a_i)} \sum_{j=1}^{m} (\bar{M}_{ij} + d_i)^2 .$$

From the proofs of lemmas 7 and 8, it should be apparent that this is equivalent to proving that $\bar{Q}_i$ is reduced to 0 during <u>each</u> service interval $\bar{I}_{ij}$. The condition for this is

(8) $$(C_i - a_i)(\bar{L}_{ij} - d_i) \geq a_i(\bar{M}_{ij} + d_i) ,$$

which will now be demonstrated.

By corollary 5,

$$m\bar{L}_{ij} = \sum_{j=1}^{m} L_{ij} \geq md_i + a_i T/C_i ,$$

and so

$$C_i(\bar{L}_{ij} - d_i) \geq a_i(T/m) = a_i(\bar{M}_{ij} + d_i + (\bar{L}_{ij} - d_i)) ,$$

implying (8) as desired.

Note that theorem 9 merely establishes the form of one optimal solution. It is clear that we could get a set of optimal solutions by permuting indices. Furthermore, it could probably be shown that these solutions are the only ones which are optimal, but this would involve solving a set of simultaneous linear equations relating the M_{ij} and L_{ij} and this is not considered here to be worth the trouble.

Also, nothing is said about the case where $m(i) \neq m(k)$. This seems to be a much harder case to analyze generally. However, if $N=2$ the above proof may be modified so that this problem is completely solved. For in that case it is obvious that $m(1) = m(2)$, since otherwise two servicings of some lane would be consecutive, giving a higher value for W than that obtained by merging the two consecutive servicings into one. Also, $L_{ij} = M_{2,j-1}$ and $L_{2j} = M_{ij}$. Thus the condition $M_{ij} = M_{ik}$ is equivalent to $L_{ij} = L_{ik}$.

Finally, we shall discuss the optimum value of m to choose in the above problem, given T . Note that if $\{I_{ij}\}$ denotes an optimal solution on the interval $[0,T]$ as described above, where each lane is serviced once during the interval $[0,T/m]$ and this procedure is repeated m times, then we have

$$mW(T/m) = W(T) .$$

This recalls the work of [1] in which the function to be minimized is $W(T)/T$ instead of $W(T)$. (Of course, for constant T , minimizing $W(T)$ is equivalent to minimizing $W(T)/T$, and the first part of [1] considers the problem in this way.) We show how the results of [1] can be applied to the above problem.

Suppose that T' is a variable which is allowed to take on only the discrete set of values T/m , m an integer, in the problem where each lane is serviced exactly once during the interval $[0,T']$. If $W(T')/T'$ attains its minimum value on this discrete set at the point $T' = T'_0 = T/m_0$, then $W(T)$ in the original problem attains its minimum value over all m for $m = m_0$. For we have

$$\frac{W(T/m_0)}{T/m_0} \leq \frac{W(T/m)}{T/m} ,$$

for $m \neq m_0$; or ,

$$m_0 W(T/m_0) \leq mW(T/m) .$$

But by the above remarks, $mW(T/m) = W(T)$, where $W(T)$ is calculated for m cycles of service in the interval $[0,T]$, and similarly for $m_0 W(T/m_0)$.

Thus it is clear that finding m_0 such that $W(T)$, optimized over all switching patterns, attains its least value when each stream is serviced m_0 times, is equivalent to finding m_0 such that $W(T/m_0)/(T/m_0)$ is a minimum for the optimized solution of the one-cycle case. Now a method is given in [1] to optimize $W(T')/T'$ when T' is a continuous variable, and it is shown that $W(T')/T'$ is monotone decreasing to the left of the optimum, $T' = T'_0$, and monotone increasing to the right. Thus if n is an integer such that

$$T/(n+1) \leq T'_0 \leq T/n ,$$

it is clear that the optimum value m_o sought above must be either n or $n+1$. In order to determine which of these values is optimum, it is only necessary to substitute in the expression for the optimized value of $W(T')/T'$, also given in [1]. Thus m_o is easily determined.

REFERENCE

[1] R. Rangarajan and R.M. Oliver, "Allocations of Servicing Periods that Minimize Average Delay for N Time-Shared Traffic Streams", Transportation Science, Vol. 1 (1967), pp. 74-80.

DEPARTMENT OF COMMERCE ★ NATIONAL BUREAU OF STANDARDS ★