Report Number--- NPS62-81-039

NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

DYNAMIC MULTICOMMODITY FLOW SCHEDULES

by

Adam Feit

December 1981

Thesis Advisor: J. M.Wozencraft

Prepared for:
Defense Advanced Research Projects Agency
1400 Wilson Blvd
Arlington, Virginia 22209

NAVAL POSTGRADUATE SCHOOL
Monterey, California

This thesis prepared in conjunction with research supported in part by Defense Advanced Research Projects Agency under ARPA Order No. 3924, Program Code No. OD20.

Released as a
Technical Report by:

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER NPS 62-81-039	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Dynamic Multicommodity Flow Schedules		5. TYPE OF REPORT & PERIOD COVERED Ph.D. Thesis; December 1981
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Adam Feit		8. CONTRACT OR GRANT NUMBER(s) ARPA Order No. 3924, Program Code No. OD20.
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Postgraduate School Monterey, California 93940		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Defense Advanced Research Projects Agency 1400 Wilson Blvd Arlington, Virginia 22209		12. REPORT DATE December 1981
		13. NUMBER OF PAGES 258
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Dynamic Routing, Data Communication Network, Minimum Time Problem, Min-Max Criterion, Transportation Problem, Linear Programming

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Some new results in the scheduling of dynamic multicommodity flows in data communication networks are presented.

A new performance measure for effective delivery of backlogged data to their destinations is defined and the solution to the resulting delivery problem is obtained through a sequential linear optimization methodology. Properties of an optimal dynamic multicommodity flow schedule are studied in detail, taking advantage

#20 - ABSTRACT - (CONTINUED)

where possible of the linear programming formulation. The special case of the delivery problem in a single destination network also is analyzed.

Application of the results to stochastic delivery problems in which the data inputs to the network are modelled as Poisson processes is addressed, and a new dynamic data communication network analysis is presented.

Finally, the delivery problem on networks with capacitated links and with traversal delays is considered and some new results obtained.

Approved for public release; distribution unlimited.

Dynamic Multicommodity Flow Schedules

by

Adam Feit
Lieutenant Colonel, Israeli Defense Forces
M.Sc., Technion-Israeli Institute of Technology, 1974

Submitted in partial fulfillment of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

from the

NAVAL POSTGRADUATE SCHOOL
December 1981

ABSTRACT

Some new results in the scheduling of dynamic multicommodity flows in data communication networks are presented.

A new performance measure for effective delivery of backlogged data to their destinations is defined and the solution to the resulting delivery problem is obtained through a sequential linear optimization methodology. Properties of an optimal dynamic multicommodity flow schedule are studied in detail, taking advantage where possible of the linear programming formulation. The special case of the delivery problem in a single destination network also is analyzed.

Application of the results to stochastic delivery problems in which the data inputs to the network are modelled as Poisson processes is addressed, and a new dynamic data communication network analysis is presented.

Finally, the delivery problem on networks with capacitated links and with traversal delays is considered and some new results obtained.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

I wish to express special gratitude to my advisor, Prof. J.M. Wozencraft, who proposed the research topic and generously contributed of his time through numerous discussions and careful reading of the manuscript. His invaluable suggestions and comments are reflected throughout the dissertation. It has been an honor and a pleasure to be associated with him.

I am thankful to the members of my doctoral committee: Prof. D.P. Gaver, Jr., Prof. R.W. Hamming, Prof. P.H. Moose and Prof. M.A. Morgan, for their counsel which led to improvements in the context and the form of the dissertation.

I cannot adequately thank here my wife Carmela and our children--Galia and Asaf for the atmosphere of love and enjoyment they have created. To them I dedicate this dissertation.

I. INTRODUCTION

A. MOTIVATION

The design of data communication networks has received much attention during the past decade and the interest in this field is constantly growing. This fact is explained by the rapidly expanding role being played by data processing in today's society and the apparent advantage in sharing powerful computational resources.

The overall subject of data communication networks can be looked upon from many different points of view, each representing an intellectually challenging problem. Just to name two: there is the problem of cost effective topological network design and there is the problem of determining, in a given network, the routes which data should follow from their source to their destination. The problem of assigning routes has been one of the most intensively studied areas in the field of data communication networks.

It is possible to view the complete routing problem as composed of two parts: the first, calculation of routing assignments and second, their implementation in an actual data routing control strategy in a real, basically stochastic, network environment. It is the first part that we are mostly concerned with in this research, although we touch briefly on the other.

A network model for calculating routing assignments can be roughly classified as static or dynamic. These reflect the

long and the short term stochastic behavior of a network, respectively. A finer taxonomy is possible, regarding the frequency with which a routing assignment has to be recomputed to support efficient communication. A desirable complete routing methodology must provide fast adaptation to a changing network environment and be stable at the same time. It should be dead-lock free as well as distributed to minimize computational complexity and control information flows. Clearly, there are substantial difficulties in achieving all these attributes simultaneously or even in qualifying trade-off relations among them.

In spite of the progress made in some areas, and in particular in the computation of static routing assignments, many basic questions remain unanswered. For example, definition of a meaningful performance measure for dynamic routing assignments which will give rise to mathematically tractable problem formulation, seems to have been missing. Thus, it appears that new insight into the relationships that make up a dynamic network model should be helpful in coping with these questions. The new conceptual ideas and analytic tools for studying them introduced in this research are a step in this direction.

B. SURVEY OF PREVIOUS WORK

In [15] Kleinrock introduced an analytical model of a data communication network. Many algorithms use this model to compute routing assignments, for example, the algorithms due to Cantor and Gerla [17], Fratta, Gerla and Kleinrock [16] and

Defenderfer [18]. In all those algorithms the steady-state delay of messages in each link is calculated explicitly as

$$t_{ij} = \frac{1}{c_{ij} - f_{ij}}$$

where:

f_{ij} is the data flow rate in link [i,j] (messages/second);

c_{ij} is the capacity of link [i,j] (messages/second);

t_{ij} is the expected delay/message experienced by all messages using that link.

A routing assignment is selected to minimize the expected weighted delay T,

$$T = \sum_{[i,j]} f_{ij} t_{ij}$$

of messages traversing the network. This analysis is based on the assumption that message arrivals to each link can be considered as Poisson process with independent, exponentially distributed, message length, which requires Kleinrock's famous "independence assumption" that messages "lose their identity" at each node and are assigned new independent lengths.

All the algorithms referenced thus far exploit the convexity of the objective function. A different approach, leading to a linear problem formulation, was suggested by Wozencraft in [29]. Instead of minimizing the average expected delay, the objective is to minimize the maximal saturation ratio f_{ij}/c_{ij}

in the network. Subsequently the next maximal saturation ratio is minimized, etc. This Min-Max policy was extensively studied by Ros [8] and as a result considerable insight into the problem of static routing was gained. One of the objectives of this research is to study possible generalization of the Min-Max approach and in particular its sequential character to deal with dynamic network models.

Kleinrock's steady-state and independence assumptions make his model appropriate to describe long term stochastic behavior of a data communication model, but less applicable for computation of dynamic routing assignments.

One of the most important efforts in the direction of finding a more appropriate model for describing network dynamics was proposed by Segall [1]. His approach is to view the contents of data queues at the network nodes in a continuous state space setting rather than as an integer number of messages or bits. The continuous nature of the state is justified by recognizing that any individual message contributes very little to the overall behavior of a network. Having established the model, Segall expresses the dynamic routing assignment problem as a linear-optimal control problem for which a close-loop solution is sought. This formulation has been used to investigate how to minimize

$$D = \int_{t_0}^{t_f} \left(\sum_{(i,k)} q_i^k(t) \right) dt$$

where:

t_f is some time at which all queues are empty;

$q_i^k(t)$ is the queue size at node i of data destined to node k, at time t.

Here D is the total delay experienced by the queued messages during the period $[t_0, t_f]$, during which all the queued data are to be delivered to their destinations.

The solution to this problem is approached by means of the minimum principle of Pontryagin [30], since this principle provides not only necessary but also sufficient condition for optimality in that case. A comprehensive set of principles describing the closed loop solution has been obtained by Moss [2] for the case in which all backlogged data have the same destination and the input to the network is continuous and constant. Additional study of the single destination case is provided in [11] and [31]. Unfortunately the optimal control approach runs into difficulties when the general, multicommodity case is considered and no general solution has yet been obtained.

C. SYNOPSIS OF THESIS

To cope with the difficulties posed by the multicommodity dynamic routing assignment problem, a different approach is taken in this research. It involves a change in the performance measure. Rather than trying to minimize the total delay experienced by the backlogged data, the concept is to deliver all of that data to their destinations in the shortest time possible. Since eventually there are many dynamic routing assignments that can do so, one which maximizes the total amount

of delivered data over time, is selected. We refer to this problem formulation as the "delivery problem" and to the corresponding dynamic routing assignment as the "flow schedule." There are several advantages associated with the above statement of the delivery problem. Most important is the ability to find the desired flow schedule by solving a sequence of hierarchically related linear programs. Also, it turns out that the new performance measure reveals some structural properties that bring new insight to the problem of dynamic flow scheduling. The main purpose of this research is to study those properties in detail.

The new performance measure (we call it the "delivery function") and related optimality criterion are explained in Ch. II. Also, by introducing a concept of "global optimality" we are able to relate the optimal delivery function to the "total delay" criterion. In Ch. II we also introduce the basic network model and the notational convention to be used throughout the thesis.

In Ch. III a solution algorithm to the delivery problem is presented. It consists of solving a sequence of hierarchically related linear programs. In principle, each linear program that is solved contracts the space of remaining feasible flow schedules, until finally an optimal piecewise constant flow schedule is obtained. The optimality of the resulting flow schedule is formally derived in Ch. III using some basic results regarding the properties of feasible flow schedules.

In Ch. IV we study in detail the structural properties of an optimal flow schedule. By exploiting various properties of

linear programming we are able to derive several results that characterize an optimal flow schedule. One key result is the description of critical sets of commodities and the capacity resources (links) they must saturate for various periods of time. Also, the properties of those saturating flows, and in particular their total rates, are determined. In this chapter the important idea of optimal solution "stability" is introduced, which allows one to express most of the above properties in terms of the optimal dual variables associated with the linear programs of the solution algorithm.

We devote Ch. V to discussion of the single destination network. The solution algorithm of Ch. III is specialized to handle this case. Considerable additional simplifications are obtained by observing that the single destination delivery problem may be interpreted as a single commodity flow problem so that advantage may be taken of many well established results. The concept of stability is revised and exploited as part of the solution algorithm.

Continuing the discussion from Ch.II we show that the optimal single destination flow schedule is also globally optimal and thus also solves the single destination "minimal total delay problem" [11]. The computational advantages of the new algorithm are addressed briefly.

In Ch. VI we analyze the stochastic delivery problem. Here, in addition to the backlogged data considered so far, we are concerned with Poisson arrivals of messages to the network. Following Yee [12] the expected minimal time to empty an M/M/1

queueing system is taken to be the new performance measure. It is shown that the theory of dynamic flow scheduling, derived earlier for the deterministic case, and in particular the sequential linear optimization methodology, can be applied (at least, in principle) to solve the stochastic delivery problem.

In Ch. VII we consider a more general setting for a delivery problem. Here we associate with each link a traversal delay, in addition to a capacity constraint. Although it would be possible to continue our discussion within the context of data communication networks, it is more natural to choose the transportation problem as the framework for our investigation. This allows a generalization of link capacity constraints to include loading and unloading constraints. The addition of traversal delays greatly complicates the delivery problem. It is possible, however, to exhibit a (conceptually) simple solution procedure for the case of bi-partite transportation networks. A discrete time approximation for general network models also is discussed and a particular example of military application is presented.

In Ch. VIII we summarize the most important results of this research and indicate areas for future study.

Finally, we defer to the Appendix a number of proofs, examples and short discussions which would tend to blur the main ideas if left within the body of the thesis.

II. PROBLEM FORMULATION

A. COMMUNICATION NETWORK MODEL

1. Topological Representations

A data communication network may be modelled as a set of nodes interconnected by a number of links. The nodes represent physical locations at which data may enter or exit the network and the links represent unidirectional channels over which data is transmitted from node to node. A typical data communication model is shown in Fig. II.1.

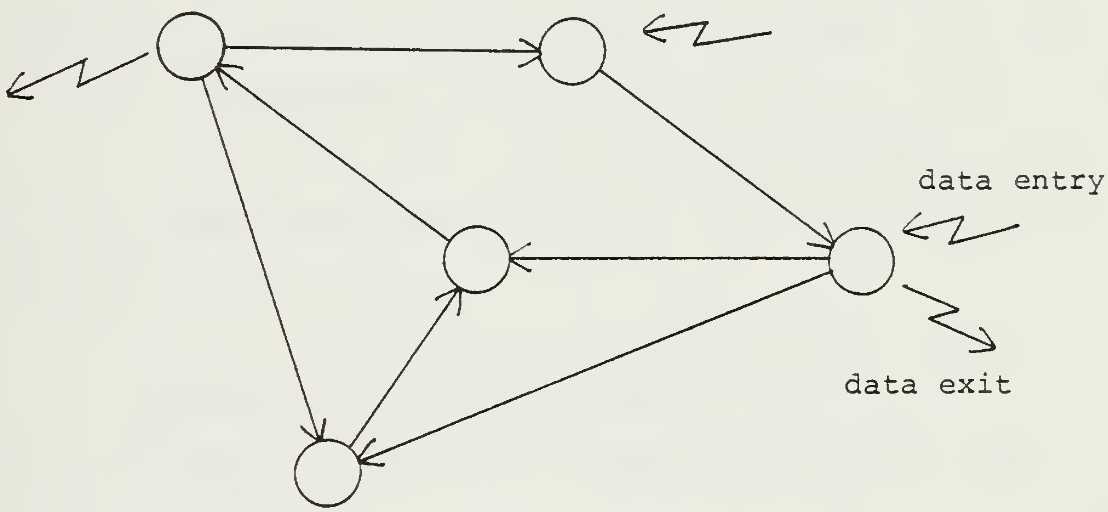


Fig. II.1. Communication Network

With each link we associate a channel capacity which indicates the upper bound on the data flow rate for that channel. With each node we associate, at every instant of time, the amounts of data awaiting transmission to each destination at the corresponding location. The collection of node descriptors for all

the nodes in the network constitutes the state of the system, or equivalently, system congestion at any given time t.

We will say that any data in the network is commodity (i,k) if its origin (entry node) was node i and its final destination (exit node) is node k. We shall also say that any data in the network is commodity k if its final destination is node k.

Consider a data communication network $G(V,L_0)$, where $V = \{1,2,...,n\}$ is a set of n nodes and $L_0 = \{[i,j]\}$ is a set of links. By the notation [i,j] we mean the link that connects node i to node j, in that direction. We will also use the notation $L_0 = \{1,2,...,\ell\}$ and denote an element of L_0 by e. If link e corresponds to link [i,j] then h(e) = i and t(e) = j.† We say that node i can communicate to node k iff there is at least one directed chain of links going form node i to node k. We also define:

N_0	=	set $\{(i,k)\}$ of node pairs such that node i can communicate to node k, $i \neq k$.
$q_i^k(t)$	=	amount of commodity k, stored in node i at time t, $\forall (i,k) \in N_0$.
$f_{ij}^k(t)$	=	flow rate of commodity k on link [i,j] at time t, $\forall [i,j] \in L_0$ and $\forall (i,k) \in N_0$.
c_{ij}	=	capacity of link [i,j], $\forall [i,j] \in L_0$.
$a_i^k(t)$	=	flow rate of commodity k arriving at node i from outside the network, at time t, $\forall (i,k) \in N_0$.

---

†h--head, t--tail $h(e) \rightarrow t(e)$

We reserve the use of respective capital letters for sets and vector notation, interchangeably. For example, $F(t) \triangleq (f_{12}^{2}(t), f_{13}^{2}(t), \ldots, f_{ij}^{k}(t) \ldots)$ denotes a vector (set) of flows.

2. Dynamic System Equations and Constraints

The flows and the queues just defined must satisfy three basic constraints: non-negativity, conservation and capacity. The non-negativity constraint states that

$$f_{ij}^{k}(t) \geq 0, \quad \forall [i,j] \in L_0, \ \forall (i,k) \in N_0 \text{ and } \forall t. \qquad \text{(II.1)}$$

The conservation constraint may be written as

$$r_i^k(t) \triangleq \sum_{j(\neq i)} f_{ij}^{k}(t) - \sum_{j(\neq i)} f_{ji}^{k}(t) \begin{cases} \geq a_i^k(t), & \text{if } q_i^k(t) > 0, \\ = a_i^k(t), & \text{otherwise,} \end{cases}$$

$$\forall (i,k) \in N_0 \text{ and } \forall t. \qquad \text{(II.2)}$$

Constraint (II.2) accounts for the fact that at all times the amount of any commodity stored at any node is a non-negative quantity. Moreover, the fact that the net delivery rate of commodity k from node i is non-negative, $r_i^k(t) \geq 0$, $\forall (i,k) \in N_0$ and $\forall t$, implies that data is not stored at intermediate nodes en route from its entry node to its exit node in the network.†

---

†Intermediate storage of commodities is discussed in Appendix C.

Finally, the capacity constraint is

$$f_{ij}(t) \triangleq \sum_{k(\neq i)} f_{ij}^{k}(t) \leq c_{ij}, \quad \forall [i,j] \in L_0 \text{ and } \forall t \qquad \text{(II.3)}$$

where $f_{ij}(t)$ denotes the <u>aggregate flow rate</u> on link $[i,j]$ at time t.

<u>Definition (II.1).</u>

A set of flows $F(t)$, $t_0 \leq t \leq t_1$ is a <u>feasible multicommodity flow schedule</u> if it satisfies constraints (II.1)-(II.3) for all $t \in [t_0, t_1]$.

□

We will assume, for mathematical convenience, that data input flow rates are identically zero during the time interval under consideration, i.e.

$$a_i^k(t) \equiv 0, \quad \forall (i,k) \in N_0, \quad t_0 \leq t \leq t_1 \qquad \text{(II.4)}$$

In Ch. VI the behavior of a communication network in a stochastic environment is considered and $a_i^k(t) = a_i^k$, $\forall (i,k) \in N$ will be interpreted as a rate of a Poisson process.

We follow the model proposed by Segall [1], where the contents of the queues at the nodes are viewed as continuous quantities, rather than as integer number of messages (in Ch. VI we recognize the existence of separate messages but model their size as a continuous quantity). This macroscopic point of view not only provides a model that is analytically simpler than others, but also is justified by recognizing that any

individual message (bit) contributes very little to the over-all behavior of the network.

Let $Q(t_0)$ and $Q(t_1)$ be the system states at times t_0 and t_1 respectively. We say that the state $Q(t_1)$ is reachable from the state $Q(t_0)$ if there exists a feasible multicommodity flow schedule $F(t)$, $t_0 \leq t \leq t_1$ for which

$$q_i^k(t_1) = q_i^k(t_0) - \int_{t_0}^{t_1} r_i^k(t)dt, \quad \forall (i,k) \in N_0 \tag{II.5}$$

or in shorthand notation, $F(t): Q(t_0) \rightarrow Q(t_1)$. To every feasible multicommodity flow schedule we adjoin a delivery function $D(t)$, $t_0 \leq t \leq t_1$ which represents the total amount of data delivered to their destination by time t.

$$D(t) = \sum_{(i,k) \in N_0} [q_i^k(t_0) - q_i^k(t)], \quad t_0 \leq t \leq t_1 \tag{II.6a}$$

or equivalently

$$D(t) = \sum_{(i,k) \in N_0} \int_{t_0}^{t} r_i^k(\alpha)d\alpha, \quad t_0 \leq t \leq t_1 \tag{II.6b}$$

From the nature of the model that we constructed and in particular from the fact that $q_i^k(t_0) > 0 \rightarrow (i,k) \in N_0$,[†] we conclude that the zero state is reachable from any other finite state within some finite time t_1, i.e. there exist both t_1 and

---

[†]To avoid excessive notation complexity we will assume in the sequel that $q_i^k(t_0) > 0 \leftrightarrow (i,k) \in N_0$. The case of $q_i^k(t_0) > 0 \rightarrow (i,k) \in N_0$ is included in the examples.

$F(t)$, $t_0 \le t \le t_1$ such that $F(t): Q(t_0) \to Q(t_1) = \underline{0}$. Consequently, for every initial state there is some minimal value of t_1. We define the minimal total delivery time t_1^0 as

$$t_1^0 = \min_{\{F(t)\}} \{t_1 | F(t): Q(t_0) \to Q(t_1) = \underline{0}\} \qquad \text{(II.7)}$$

We conclude this section with a basic result regarding the nature of feasible multicommodity flow schedules.

Theorem II.1[†]

Let $Q(t_0)$ and $Q(t_1)$ be any two states such that $Q(t_1)$ is reachable from $Q(t_0)$. Then there exists a feasible multicommodity constant flow schedule $F_1(t) = F$, $t_0 \le t \le t_1$ such that $F_1(t): Q(t_0) \to Q(t_1)$.

□

This result implies that in order to transfer a system into a reachable state it is sufficient to look for an appropriate flow schedule within the subset of constant flow schedules. The benefits of this property will become evident in following chapters.

3. Piecewise Constant Flow Schedules

From this point on we will narrow our interest to the subset of feasible multicommodity flow schedules which are piecewise constant. By an M-part constant-flow schedule $F_M(t)$, $t_0 \le t \le t_1$ we mean

---

[†]See Appendix A for proof of this theorem.

$$F_M(t) = \begin{cases} F_M(1), & t_2 < t \leq t_1 \\ \vdots & \\ F_M(m), & t_{m+1} < t \leq t_m \\ \vdots & \\ F_M(M), & t_0 \leq t \leq t_M \end{cases} \quad \text{(II.8)}$$

Consulting (II.6) we immediately conclude that the corresponding delivery function $D_M(t)$, $t_0 \leq t \leq t_1$ is piecewise linear. We write

$$D_M(t) = \begin{cases} D_M(t_1)-\rho_M(1)\cdot(t_1-t), & t_2 < t \leq t_1 \\ \vdots & \\ D_M(t_m)-\rho_M(m)\cdot(t_m-t), & t_{m+1} < t \leq t_m \\ \vdots & \\ D_M(t_M)-\rho_M(M)\cdot(t_M-t), & t_0 \leq t \leq t_M \end{cases} \quad \text{(II.9a)}$$

where $\rho_M(m)$ is the total delivery rate in the m-th interval,

$$\rho_M(m) \triangleq \sum_{(i,k)\in N_0} r_i^k(m) \quad \text{(II.9b)}$$

The $r_i^k(m)$, $(i,k) \in N_0$ is the net delivery rate of the k-th commodity from node i (see (II.2)) in the m-th interval, corresponding to the flow schedule segment $F_M(m)$ in (II.8). Clearly, $\rho_M(m)$ is the slope of the delivery function in that interval. An example of a piecewise linear delivery function is shown in Fig. II.2

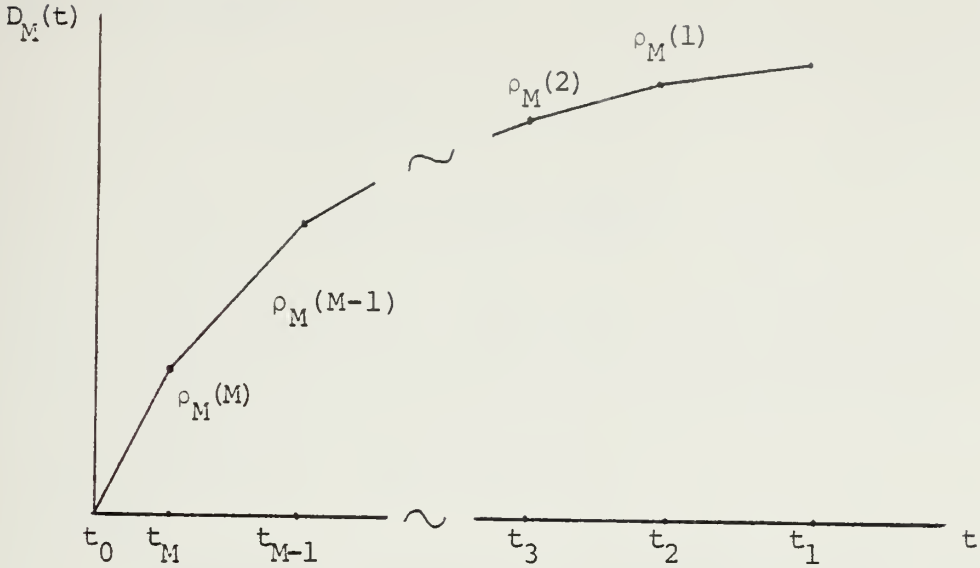


Fig. II.2. Piecewise Linear Delivery Function

4. <u>Example</u>

To fix ideas, in this subsection we provide a simple example of a communication network, and use it to illustrate the various notation and definitions introduced previously.

Consider the network $G(V,L_0)$ shown in Fig. II.3.

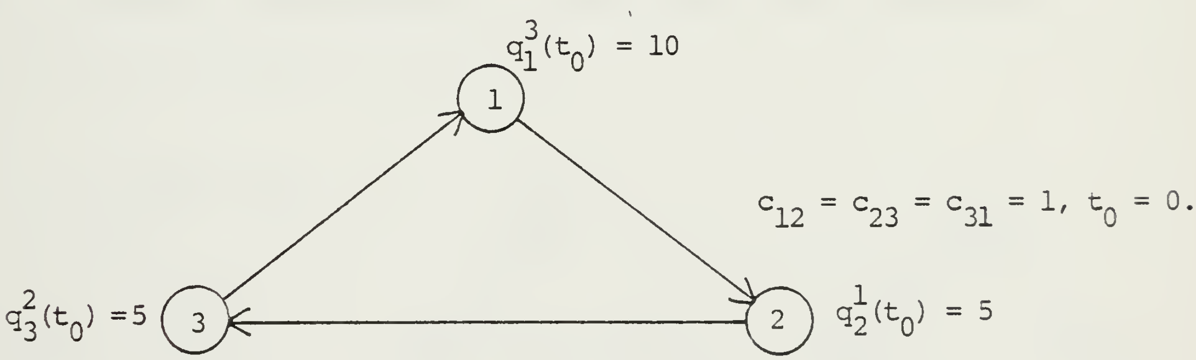


Fig. II.3. Communication Network with Queued Data

For this netowrk

$$V = \{1,2,3\}$$

$$L_0 = \{[1,2],[2,3],[3,1]\}$$

$$N_0 = \{(1,2),(1,3),(2,1),(2,3),(3,1),(3,2)\}$$

$$Q(0) = \{0,10,5,0,0,5\}$$

Consider a constant flow schedule $F_1(t)$, $0 \leq t \leq 15$ given by its components:

$$f_{12}^2 = \frac{1}{3}, \quad f_{12}^3 = \frac{2}{3}, \quad f_{23}^1 = \frac{1}{3}, \quad f_{23}^3 = \frac{2}{3}, \quad f_{31}^1 = \frac{1}{3},$$

$$f_{31}^2 = \frac{1}{3}.$$

It is easy to check that $F_1(t)$ satisfies constraints (II.1)-(II.3) for all t, $t \in [0,15]$ and that $F_1(t): Q(0) \rightarrow Q(15) = \underline{0}$. We find it useful to decompose the flow pattern into commodity chain flows as shown in Fig. II.4.

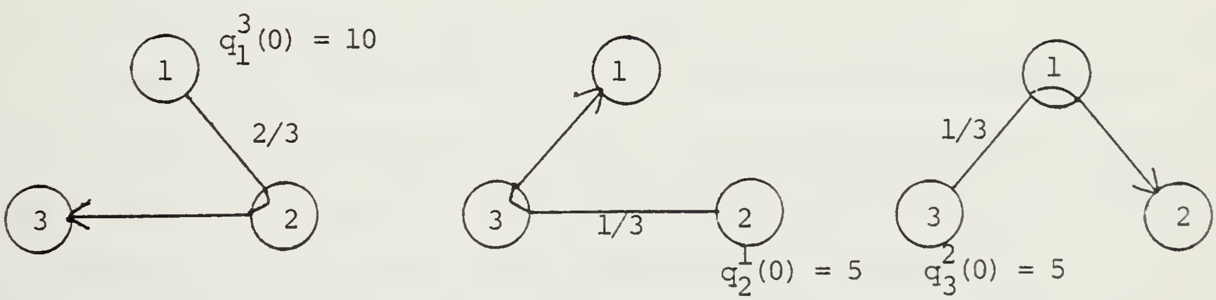


Fig. II.4. Chain Flow Decomposition

Using (II.2) we find the net delivery rates:

$$r_1^2 = 0, \quad r_1^3 = \frac{2}{3}, \quad r_2^1 = \frac{1}{3}, \quad r_2^3 = 0, \quad r_3^1 = 0,$$

$$r_3^2 = \frac{1}{3}.$$

The total delivery rate is thus (their sum) $\rho_1(1) = \frac{4}{3}$ and the resulting delivery function is given in Fig. II.5.

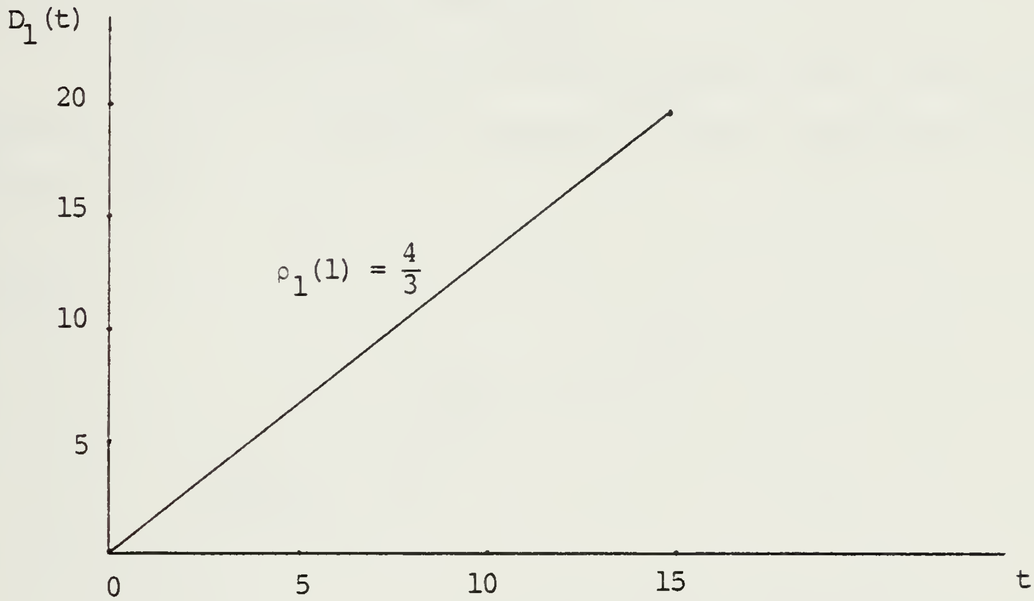


Fig. II.5. Delivery Function

B. OPTIMAL DELIVERY FUNCTION CONCEPT

Recall that our objective is to find a flow schedule that will efficiently deliver a given set of data backlogs to their destinations. Thus, we need to establish what are the desired properties of an optimal delivery function and find its generating flow schedule. In defining the optimality criteria we were led by the following goals:

(i) physically meaningful criterion;

(ii) computational tractability of the solution procedure;

(iii) gaining new insight into the problem.

We will show that the criterion we have chosen actually attains these goals. In this section we discuss property (i) above. Properties (ii) and (iii) will be considered in the next chapters.

1. Convex Delivery Functions

Consider some non-convex piecewise linear delivery function $D_M(t)$. A typical example of such a function is shown in Fig. II.6.

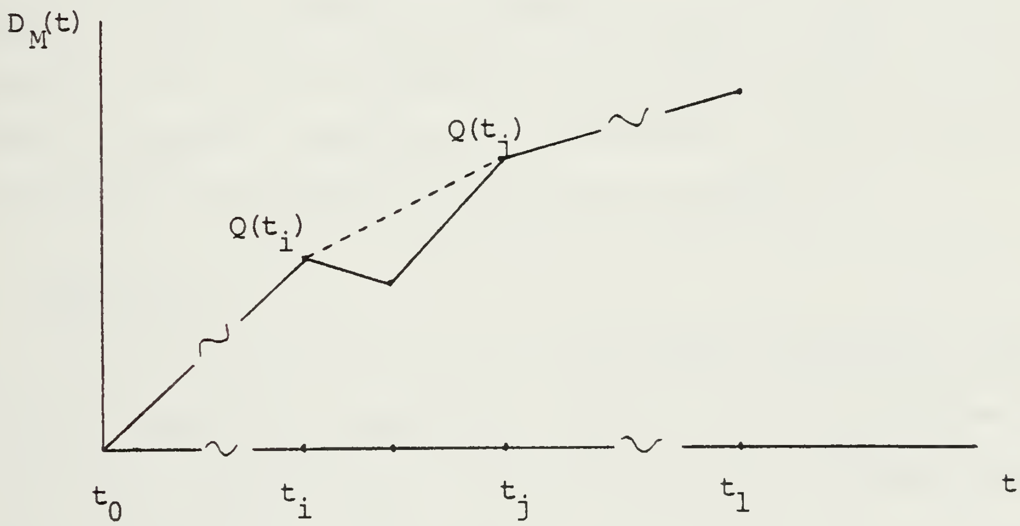


Fig. II.6. Non-Convex Delivery Function

Let $Q(t_i)$ and $Q(t_j)$ be the system states at times t_i and t_j, respectively. Since $Q(t_j)$ is reachable from $Q(t_i)$ (by assumption), then according to Thm. II.1, it is possible to construct a constant flow schedule $F_1(t)$, $t_i < t \leq t_j$ such that

$F_1(t): Q(t_i) \to Q(t_j)$. To this flow schedule corresponds a linear delivery function, indicated by the broken line in Fig. II.6.

We define a new flow schedule $F_{M-1}(t)$, $t_0 \leq t \leq t_1$ as

$$F_{M-1}(t) = \begin{cases} F_M(t), & t_0 \leq t \leq t_i \text{ and } t_j < t \leq t_1 \\ F_1(t), & t_i < t \leq t_j \end{cases}$$

where $F_M(t)$ is the original flow schedule corresponding to $D_M(t)$. The new delivery function[†] $D_{M-1}(t)$, $t_0 \leq t \leq t_1$ is convex. It is not difficult to see that a similar procedure may be applied, repeatedly if necessary, to any non-convex flow schedule. We summarize this fact in Lemma II.1.

<u>Lemma II.1</u>

Let $D_M(t)$, $t_0 \leq t \leq t_1$ be a non-convex piecewise linear delivery function. Then there exists a convex piecewise linear delivery function $D_K(t)$, $t_0 \leq t \leq t_1$ and $K < M$ such that $D_K(t) \geq D_M(t)$, $\forall t \in [t_0, t_1]$.

□

There should be no doubt that in the context of our problem the delivery function $D_K(t)$ is preferable to $D_M(t)$. As a result

---

[†]By "delivery function" we mean, unless otherwise indicated, a feasible, convex, piecewise linear delivery function. Similarly, we use loosely "flow schedule" to mean a feasible, multicommodity piecewise constant flow schedule.

we conclude that the search for "optimal" delivery functions may be confined to the subset of convex delivery functions.

2. Preference Relation

Here we answer the following question: If $D_M^A(t)$, $t_0 \leq t \leq t_1^A$ and $D_K^B(t)$, $t_0 \leq t_1 \leq t_1^B$, are two delivery functions, which is preferable? From our previous discussion of piecewise linear delivery functions (see (II.9)) we know that, for a given initial system state, a delivery function is completely specified by its corner points and delivery rates. We define T to be a descriptor vector for a delivery function $D_M(t)$.

$$T \triangleq (t_1, \rho_1, t_2, \rho_2, \ldots, t_M, \rho_M) \qquad \text{(II.10)}$$

Definition II.2

Given two convex, piecewise linear delivery functions $D_M^A(t)$, $t_0 \leq t \leq t_1^A$ and $D_K^B(t)$, $t_0 \leq t \leq t_1^B$, we say that $D_M^A(t)$ dominates $D_K^B(t)$ iff $T_j^A < T_j^B$ and $T_i^A = T_i^B$, $i = 1,2,\ldots,j-1$ for some j, $j \leq \min(M,K)$, where T_j denotes the j-th component of T. □

The implication of Def. II.2 will become evident in the following examples. In Fig. II.7, $D_3^A(t)$ dominates D_2^B since $\rho_3^A(1) < \rho_2^B(1)$. In Fig. II.8, $D_2^B(t)$ dominates $D_3^A(t)$ since $t_1^B < t_1^A$. In Fig. II.9, $D_2^A(t)$ dominates $D_3^B(t)$ since $t_2^A < t_2^B$.

3. Optimality Criterion

The definition of optimal flow schedule now follows directly from the preference relation that we established.

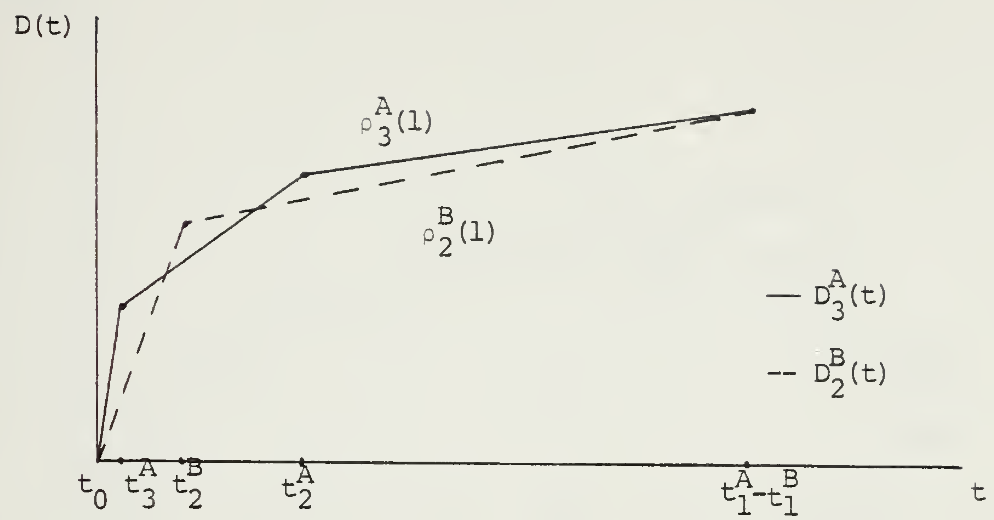


Fig. II.7. Comparison of Delivery Functions

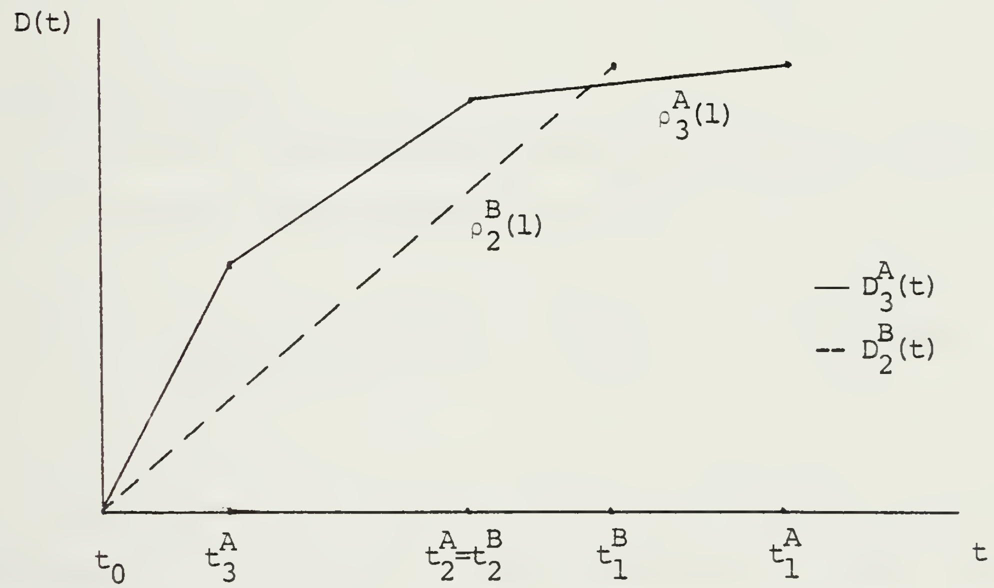


Fig. II.8. Comparison of Delivery Functions

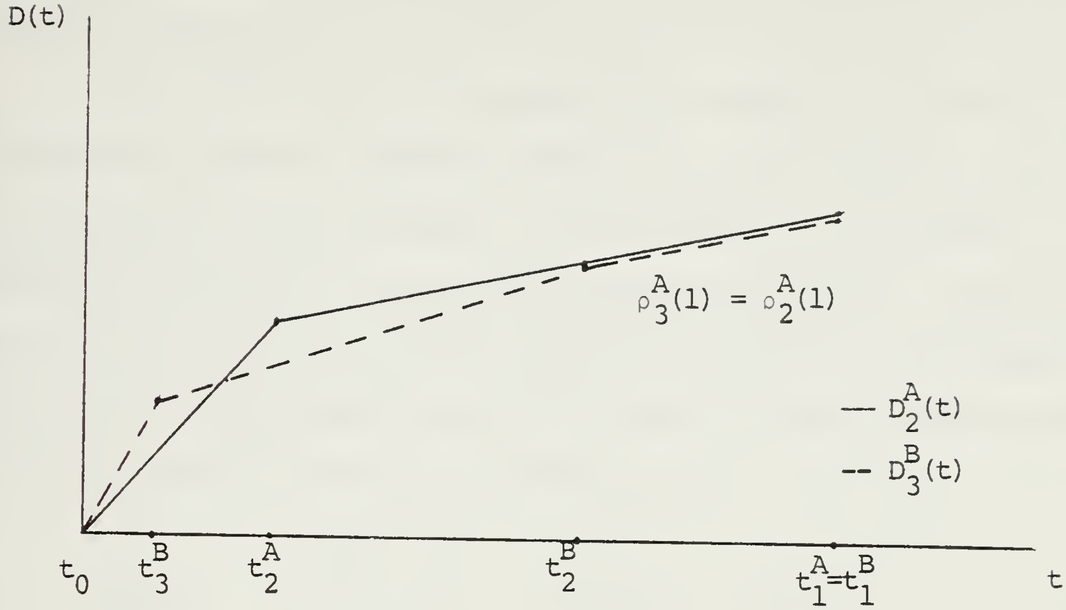


Fig. II.9. Comparison of Delivery Functions.

Definition II.3

We say that a delivery function $D_M^0(t)$, $t_0 \leq t \leq t_1^0$ is optimal for an initial system state $Q(t_0)$ iff

$$T_j^0 = \min_{\{D_K(t)\}} \{T_j \mid T_1^0, T_2^0, \ldots, T_{j-1}^0\}, \quad j = 2, \ldots, 2M \qquad \text{(II.11)}$$

□

In words, $D_M^0(t)$ is an optimal delivery function if it is not dominated by any other delivery function. We shall call a flow schedule $F_M^0(t)$, $t_0 \leq t \leq t_1^0$ which generates $D_M^0(t)$ an <u>optimal flow schedule</u>.

There is a technical question concerning the existence of $D_M^0(t)$: Can we be sure that M, the number of corner points is finite, because if not the delivery function will not be piecewise

linear. We defer the discussion of this point to Appendix E, since we need additional results before we can cope with it.

Our definition of optimality is based on a particular preference relation, which we believe to be a natural one within the framework of the optimal delivery problem. Obviously, other preference relations (and thus optimality criteria), based on the delivery function concept, can be designed. For example, the minimal total delay objective (see [1], [2]) may be specified, in terms of a delivery function, as

$$\min_{\{D(t)\}} \int_{t_0}^{t_1} \Big[\sum_{(i,j)\in N_0} q_i^k(t_0) - D(t) \Big] dt \tag{II.12}$$

where $Q(t_1) = \underline{0}$, and the optimal delivery function is that one which satisfies (II.12).

A legitimate question to be raised is: Since the optimal delivery function approach as well as the minimal total delay objective seem to have merits in the same workframe, are they related? To answer this question we need to introduce the concept of <u>global optimality</u>.

<u>Definition II.4</u>

We say that a delivery function $D_M^*(t)$, $t_0 \le t \le t_1^*$ is globally optimal iff

$$D_M^*(t) \ge D_K(t), \ \forall t \in [t_0, t_1^*] \text{ and } \forall D_K(t). \tag{II.13}$$

□

Theorem II.2

Suppose there exists a globally optimal delivery function $D_M^*(t)$. Then $D_M^*(t)$ and its generating flow schedule $F_M^*(t)$ solve the optimal delivery as well as the minimal total delay problems.

Proof:

If we denote the total amount of data stored in the network at time t_0 by q_0 then (II.12) may be written

$$D_K^0(t) = \arg\{\min_{\{D(t)\}} [(t_1-t_0)q_0 - \int_{t_0}^{t_1} D(t)dt]\} \quad \text{(II.14)}$$

The graphic representation of (II.14) is shown in Fig. II.10

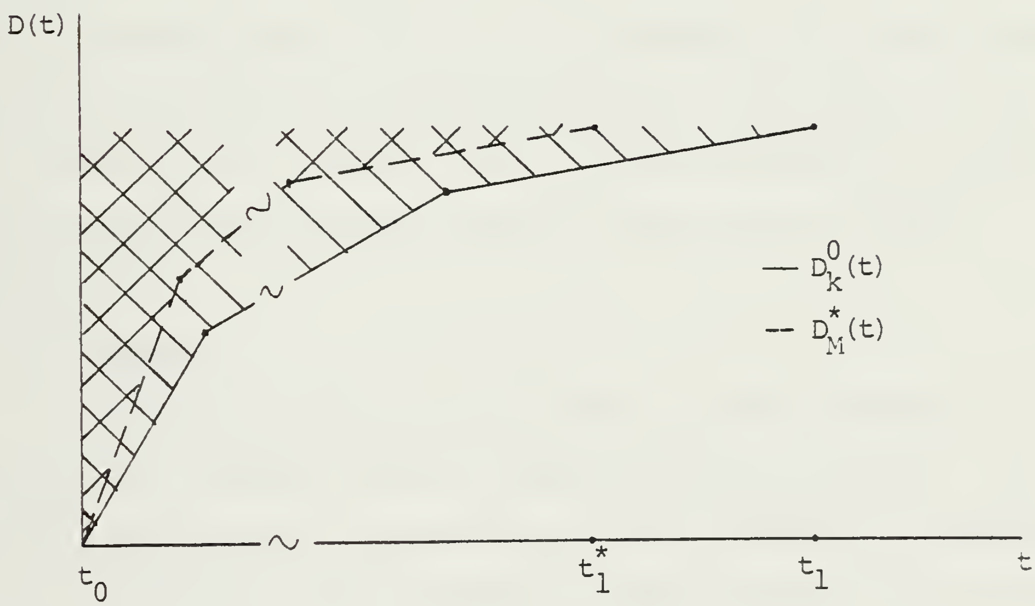


Fig. II.10. Graphic Representation of Eq. (II.14).

The objective is to minimize the shaded area. Assume that $D_K^0(t)$ solves the problem and $D_K^0(t) \neq D_M^*(t)$. Then, since by definition $D_M^*(t) \geq D_K^0(t)$, $\forall t \in [t_0, t_1^*]$ the double shaded area is smaller than the shaded area. This contradicts the assumption that $D_K^0(t) \neq D_M^*(t)$ and completes the proof with regard to the minimal total delay problem.

Next, assume that $D_K^0(t)$ solves the optimal delivery problem. Looking at $D_K^0(t)$ from t_1 backwards, and arguing that by definition it satisfies (II.11), we find that it is identical to $D_M^*(t)$. This completes the proof with regard to the optimal delivery problem.

□

We conclude from Thm. II.2 that if either criterion produces a globally optimal D(t), then they are equivalent. Unfortunately, the conjecture that a globally optimal delivery function always exists for any multicommodity delivery problem is disproved by counterexample as shown in Appendix B. We shall prove later, however, that a globally optimal delivery function <u>does</u> always exist when all flows have a single destination.

4. <u>Example</u>

We conclude this section with a simple example of optimal delivery function and its generating flow schedule. We use the same delivery problem as in Sec. A.4 of this chapter.

An optimal flow schedule is shown in Fig. II.11, were we use the chain flow representation.

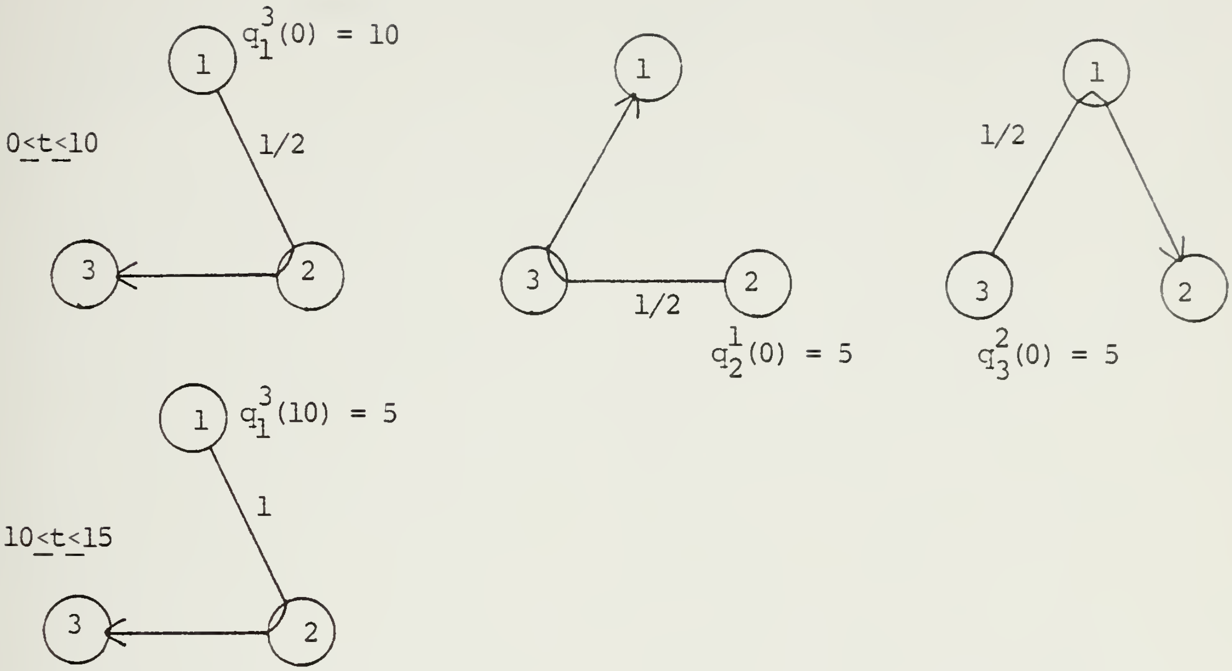


Fig. II.11. Optimal Flow Schedule

The optimal delivery function is plotted in Fig. II.12. The broken line depicts the delivery function that we obtained

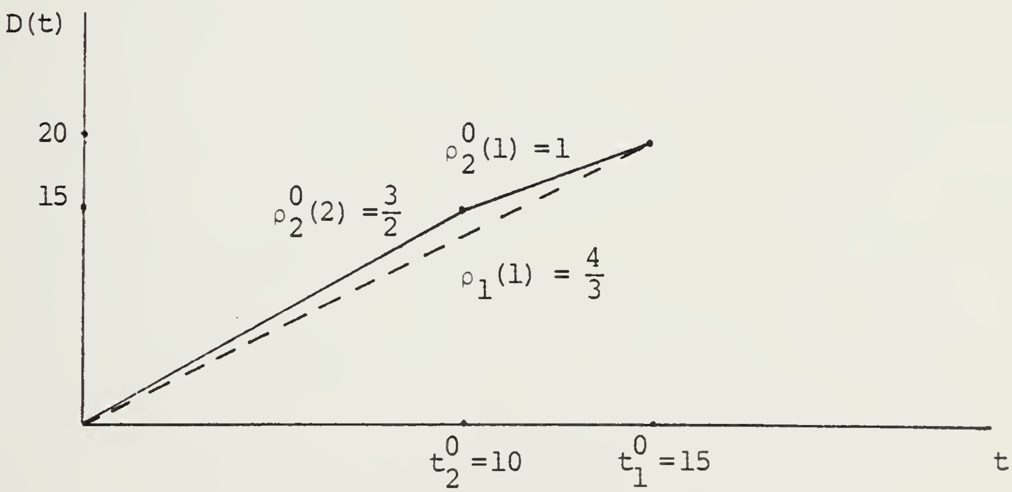


Fig. II.12. Optimal Delivery Function

for this problem in the previous section. A methodology for obtaining an optimal delivery function is treated in the next chapter.

III. SOLUTION ALGORITHM

The definition of optimal delivery function (see Def. II.3) suggests in itself a sequential structure of the solution algorithm. We start by solving for t_1^0, the minimal time in which all the queues can be delivered to their destinations. Then, keeping t_1^0 fixed we search for a flow schedule which generates a delivery function with a minimal possible delivery rate in its first interval. We denote this rate by ρ_1^0. The next step is to hold t_1^0 and ρ_1^0 fixed and search for t_2^0, the next corner point, etc. We will refer to the problems in which the minimal time (or corner location) are being found as <u>Minimal Time Problems</u> (MTP, for short). We will call the other problems, <u>Minimal Rate Problems</u> (or MRP). Since there may be more than one corner point to consider, we will usually add an index to indicate which corner we are dealing with. We will show that both types of problems can be formulated as <u>linear programming problems</u> (LP), and as such enjoy tremendous computational benefits. The mathematical properties of the optimal solutions as well as their meaning will be the subject of Ch. IV.

A. THE FIRST CORNER POINT

1. The First Minimal Time Problem

Let $F_M^0(t)$ be an optimal flow schedule and $D_M^0(t)$ its delivery function. Thm. II.1 assures us of the existence of an $F_1^0(t)$, $0 \le t \le t_1^0$ such that $F_1^0(t): Q(0) \to Q(t_1^0) = \underline{0}$. The relation between $D_1^0(t)$ and $D_M^0(t)$ is pictured in Fig. III.1.

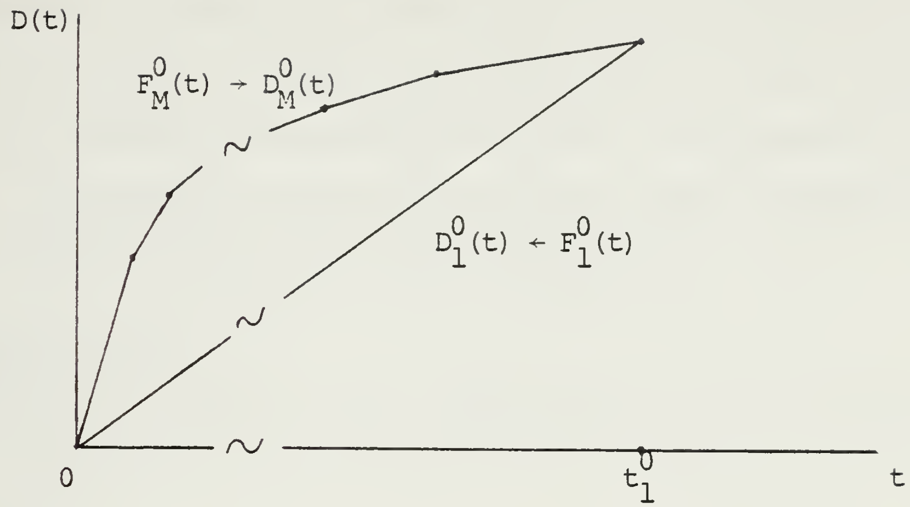


Fig. III.1. $D_1^0(t)$ versus $D_M^0(t)$

An important consequence of the above is that our search for t_1^0 can be confined to the subset of constant flow schedules (with one corner only) $\{F_1(t)\}$.

Definition III.1

The First Minimal Time Problem is given by

MTP(1): $$\min t_1$$

s.t.

$$t_1\{\sum_{j(\neq i)} f_{ij}^k(1) - \sum_{j(\neq i)} f_{ji}^k(1)\} = q_i^k(0), \ \forall (i,k) \in N_0$$

$$\sum_{k(\neq i)} f_{ij}^k(1) \leq c_{ij}, \ \forall [i,j] \in L_0$$

$$t_1,\ f_{ij}^k(1) \geq 0, \ \forall [i,j] \in L_0, \ \forall (i,k) \in N_0 \quad \text{(III.1a)}$$

□

The minimal value of the objective function of MTP(1) represents the minimal time by which the set of commodity queues Q(0) can be completely delivered to their destination by some constant flow schedule $F_1(t)$. We summarize this fact with the previous observation that we need to look only within the class of one corner flow schedules by

Theorem III.1

The minimal value of the objective function of MTP(1) equals t_1^0.

□

Problem (III.1a) has some quadratic constraints. To overcome this difficulty we use a simple transformation of variables. Define

$$u_{ij}^k(1) \triangleq t_1 f_{ij}^k(1), \quad \forall (i,k) \in N_0, \ \forall [i,j] \in L_0 \qquad \text{(III.2)}$$

The transformation relies on the assumption that $t_1^0 > 0$. Obviously this is the case in any problem of interest. Introducing (III.2) into (III.1a) results in

MTP(1): $\min t_1$

s.t.

$$\sum_{(j\neq i)} u_{ij}^k(1) - \sum_{j(\neq i)} u_{ji}^k(1) = q_i^k(0), \ \forall (i,k) \in N_0$$

$$-t_1 c_{ij} + \sum_{k(\neq i)} u_{ij}^k(1) \leq 0, \ \forall [i,j] \in L_0 \qquad \text{(III.1b)}$$

$$t_1, u_{ij}^k(1) \geq 0, \ \forall [i,j] \in L_0, \ \forall (i,k) \in N_0$$

□

which is an LP. The new variable $u_{ij}^{k}(1)$ represents the total amount of data destined to node k that traverses link [i,j] during a period of length t_1.

2. <u>The First Minimal Rate Problem</u>

Now that we have found t_1^0 we want to find a flow schedule which satisfies both t_1^0 and ρ_1^0.

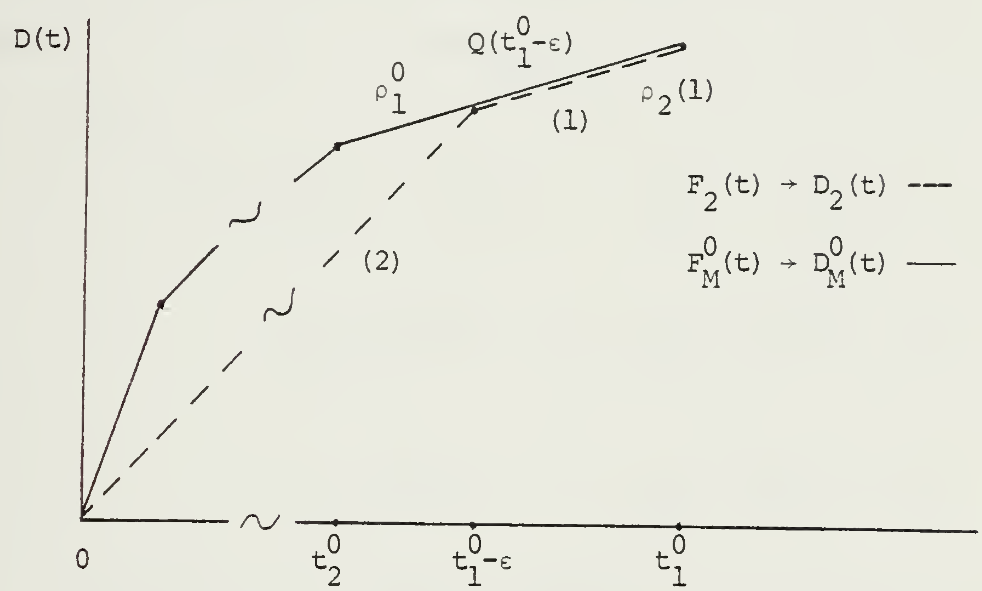


Fig. III.2. $D_M^0(t)$ versus $D_2(t)$

Let $D_M^0(t)$ be an optimal delivery function and let $Q(t_1^0-\varepsilon)$ be the state of the system at some time $t_1^0-\varepsilon$, where $0 < \varepsilon < t_1^0-t_2^0$. From Thm. II.1 we know that there exists a constant flow schedule which can transfer the system from its initial state Q(0) to $Q(t_1^0-\varepsilon)$. If we combine this flow schedule with part of the optimal flow schedule in the interval $(t_1^0-\varepsilon,t_1^0]$, we obtain a two part flow schedule $F_2(t)$, $0 \leq t \leq t_1^0$. The

resulting delivery function $D_2(t)$, $0 \le t \le t_1^0$ is shown in Fig. III.2. We observe that the delivery function $D_2(t)$ has a delivery rate, in the interval $(t_1^0-\varepsilon, t_1]$, $\rho_2(1)$ which is identical to ρ_1^0.

Definition III.2

The First Minimal Rate Problem is given by

$$\text{MRP}(1): \quad \min \sum_{(i,k)\in N_0} r_i^k(1)$$

s.t.

$$(t_1^0-\varepsilon) r_i^k(2) + \varepsilon r_i^k(1) = q_i^k(0), \ \forall (i,k) \in N_0$$

$$\sum_{k(\neq i)} f_{ij}^k(1) \le c_{ij}, \ \forall [i,j] \in L_0 \qquad \text{(III.3)}$$

$$\sum_{k(\neq i)} f_{ij}^k(2) \le c_{ij}, \ \forall [i,j] \in L_0$$

$$r_{ij}^k(1),\ r_{ij}^k(2),\ f_{ij}^k(1),\ f_{ij}^k(2) \ge 0, \ \forall [i,j] \in L_0, \ \forall (i,k) \in N_0$$

for any ε such that $0 < \varepsilon < t_1^0 - t_2^0$,

where

$$r_i^k(p) \triangleq \sum_{j(\neq i)} f_{ij}^k(p) - \sum_{j(\neq i)} f_{ji}^k(p), \ \forall (i,k) \in N_0,$$

$p = 1,2.$ □

As in the case of MTP(1) also here we can state that

Theorem III.2

The minimal value of the objective function of MRP(1) equals ρ_1^0. □

The condition on ε in (III.3) requires knowledge of t_2^0 which has not yet been determined. This difficulty is only of a theoretical nature, however, since in practice we can choose ε as small as we wish. If the very small ε we picked is too large, i.e., $\varepsilon > t_1^0 - t_2^0$, then we will miss one corner point of the optimal delivery function and the solution we obtain will be suboptimal in this sense. On the other hand, there is not much to lose by overlooking corners in the optimal delivery function which are an infinitesimal distance apart.

B. SUBSEQUENT CORNER POINTS

By now it should be clear that the procedure of Section A in principle can be applied M-times in sequence to obtain the optimal delivery function. Thus, for example, the optimal delivery function solution to MTP(m), which we denote by $D_m^0(t)$ coincides with the optimal delivery function $D_M^0(t)$ in the first 2m-1 elements of the descriptor vectors (see (II.10) and Def. II.2). These elements are $t_1^0, \rho_1^0, \ldots, \rho_{m-1}^0, t_m^0$. The optimal delivery function solution to MRP(m), which we denote by $D_m(t)$ coincides with the optimal delivery function $D_M^0(t)$ in 2m descriptor elements, namely $t_1^0, \rho_1^0, \ldots, t_m^0, \rho_m^0$. Since the term "optimal delivery function" may be confused with partial solutions we will usually use "optimal delivery function of order m" whenever we mean a solution to MTP(m) or MRP(m) for some m, $1 \leq m < M$.

Using Fig. III.3 and Fig. III.4 to help identify the variables we can write linear programs for solving MTP(m) and MRP(m) as follows:

1. The m-th Minimal Time Problem

Definition III.3

The m-th Minimal Time Problem is given by

MTP(m): $\min t_m$

s.t.

$$\sum_{p=1}^{m}\Big(\sum_{j(\neq i)} u_{ij}^{k}(p) - \sum_{j(\neq i)} u_{ji}^{k}(p)\Big) = q_i^k(0), \ \forall (i,k) \in N_0$$

$$\sum_{j(\neq i)} u_{ij}^{k}(p) - \sum_{j(\neq i)} u_{ji}^{k}(p) \geq 0, \ \forall (i,k) \in N_0, \ p = 1,2,\ldots,m$$

$$-t_m c_{ij} + \sum_{k(\neq i)} u_{ij}^{k}(m) \leq 0, \ \forall [i,j] \in L_0$$

$$-t_{m-1} c_{ij} + \sum_{k(\neq i)} u_{ij}^{k}(m-1) \leq 0, \ \forall [i,j] \in L_0 \qquad \text{(III.4)}$$

$$\sum_{k(\neq i)} u_{ij}^{k}(p) \leq \Delta t_p^0 c_{ij}, \ \forall [i,j] \in L_0,$$

$$p = 1,2,\ldots,m-2$$

$$-t_{m-1}\rho_{m-1}^{0} + \sum_{(i,k)\in N_0}\Big(\sum_{j(\neq i)} u_{ij}^{k}(m-1) - \sum_{j(\neq i)} u_{ji}(m-1)\Big) = 0$$

$$\sum_{i,k\in N_0}\Big(\sum_{j(\neq i)} u_{ij}^{k}(p) - \sum_{j(\neq i)} u_{ji}(p)\Big) = \Delta t_p^0 \rho_p^0, p = 1,2,\ldots,m-2$$

$$t_m + t_{m-1} = t_{m-1}^{0}$$

$$u_{ij}^{k}(p), t_m, t_{m-1} \geq 0, \quad \forall (i,k) \in N_0, \ \forall [i,j] \in L_0, \ p = 1,2,\ldots,m$$

for given $t_1^0, \rho_1^0, \ldots, t_{m-1}^0, \rho_{m-1}^0$. □

We used the following change of variables:

$$u_{ij}^{k}(p) \triangleq \begin{cases} \Delta t_p^0 f_{ij}^{k}(p), & p = 1,2,\ldots,m-2 \\ & \qquad \forall (i,k) \in N_0, \ \forall [i,j] \in L_0 \\ t_p f_{ij}^{k}(p), & p = m-1,m \end{cases} \tag{III.5}$$

and

$$\Delta t_p^0 \triangleq t_p^0 - t_{p+1}^0, \quad p = 1,2,\ldots,m-2.$$

Fig. III.3 shows the relations between the various parameters of MTP(m).

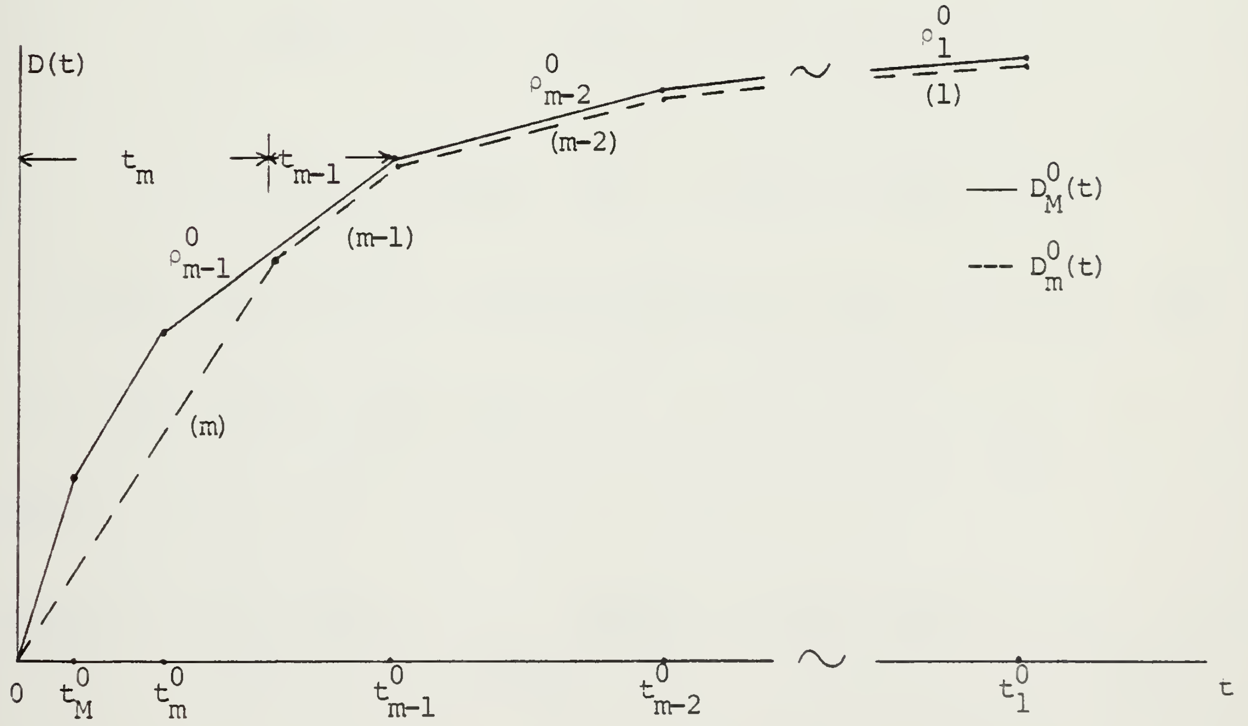


Fig. III.3. The m-th Minimal Time Problem

Using Thm. II.1 and the same reasoning as in Section A we see that

Theorem III.3

The minimal value of the objective function of MTP(m) equals t_m^0.

□

2. The m-th Minimal Rate Problem

Definition III.4

The m-th minimal rate problem is given by

$$\text{MRP}(m): \qquad \min \sum_{(i,k)\in N_0} r_i^k(m)$$

s.t.

$$(t_m^0-\varepsilon)\, r_i^k(m+1)+\varepsilon r_i^k(m) + \sum_{p=1}^{m-1} \Delta t_p^0 r_i^k(p) = q_i^k(0), \quad \forall (i,k) \in N_0$$

$$\sum_{k(\neq i)} f_{ij}^k(p) \leq c_{ij}, \quad \forall [i,j] \in L_0, \quad p = 1,2,\ldots,m+1$$

(III.6)

$$\sum_{(i,k)\in N_0} r_i^k(p) = \rho_p^0, \quad p = 1,2,\ldots,m-1$$

$$r_i^k(p),\ f_{ij}^k(p) \geq 0, \quad \forall (i,k) \in N_0,\ \forall [i,j] \in L_0,\ p = 1,2,\ldots,m+1$$

for given $t_1^0, \rho_1^0, \ldots, \rho_{m-1}^0, t_m^0, \varepsilon$,

□

where

$$r_i^k(p) \triangleq \sum_{j(\neq i)} f_{ij}^k(p) - \sum_{j(\neq i)} f_{ji}^k(p), \quad \forall (i,k) \in N_0,\ p = 1,2,\ldots,m+1$$

(III.6a)

$$\Delta t_p^0 \triangleq t_p^0 - t_{p+1}^0, \quad p = 1,2,\ldots,m-1$$

and ε is any real number such that $0 < \varepsilon < t_m^0 - t_{m+1}^0$. Fig. III.4 shows the relations between the various parameters of MRP(m).

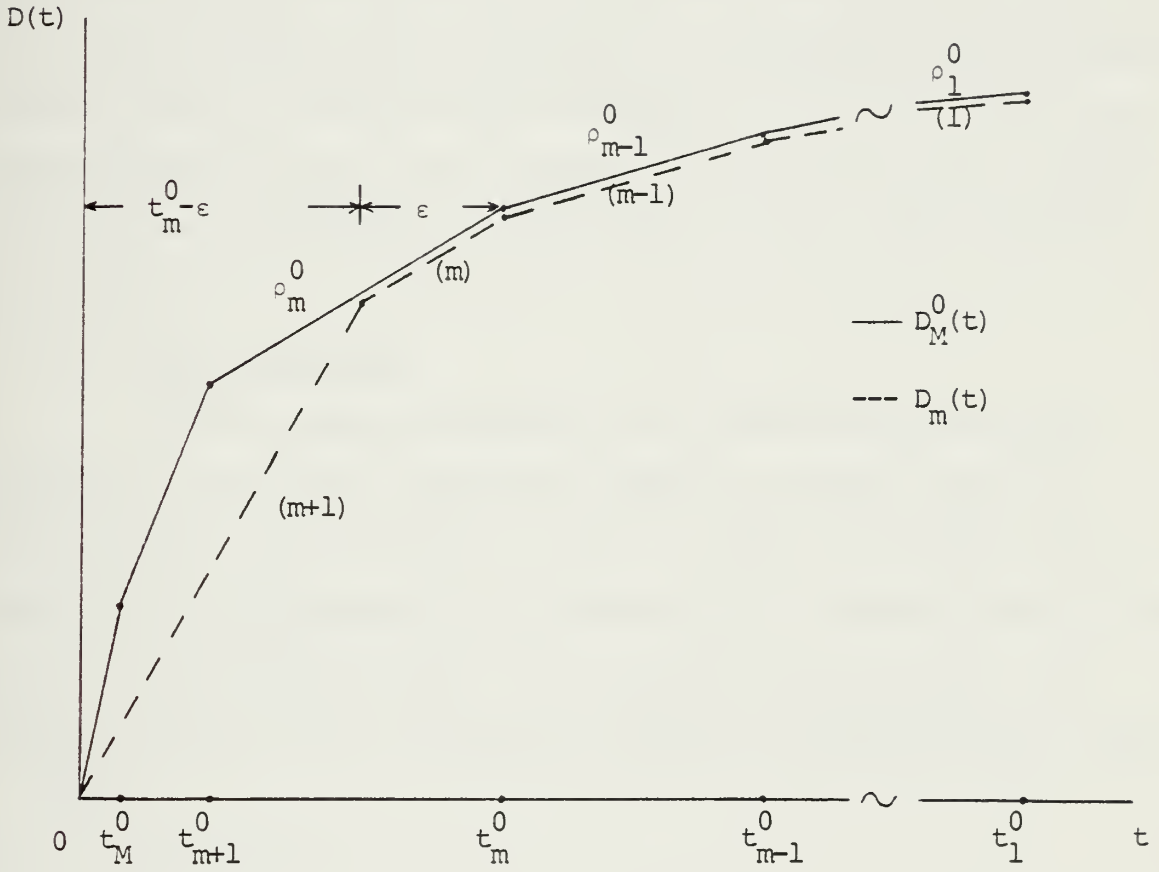


Fig. III.4. The m-th Minimal Rate Problem

Again, using Thm. II.1 to prove the existence of $D_m(t)$ as shown in Fig. III.4 we conclude that

Theorem III.4

The minimal value of the objective function of MRP(m) equals ρ_m^0. □

C. SUMMARY OF THE SOLUTION PROCEDURE

In Sections A and B we showed how to formulate the MTP(m) and MRP(m) in LP form. The solution algorithm consists of an iterative procedure in which we solve that pair of problems for each new corner point of the optimal delivery function. We refer to this kind of procedure as Sequential Linear Optimization (SLO). We will see in Chapters V, VI and VII that the SLO methodology is a very powerful tool for solving a variety of complex multicommodity network and certain other problems as well.

1. The Algorithm

We now define conditions for algorithm termination. Suppose we have just solved the MTP(M), and thus found t_M^0. The optimal flow schedule of order M, $D_M^0(t)$ is actually the optimal flow schedule we are looking for. Suppose that we solve now the MRP to obtain ρ_M^0. It is obvious that the new flow schedule $F_{M+1}(t)$ will generate a delivery function $D_{M+1}(t)$ which

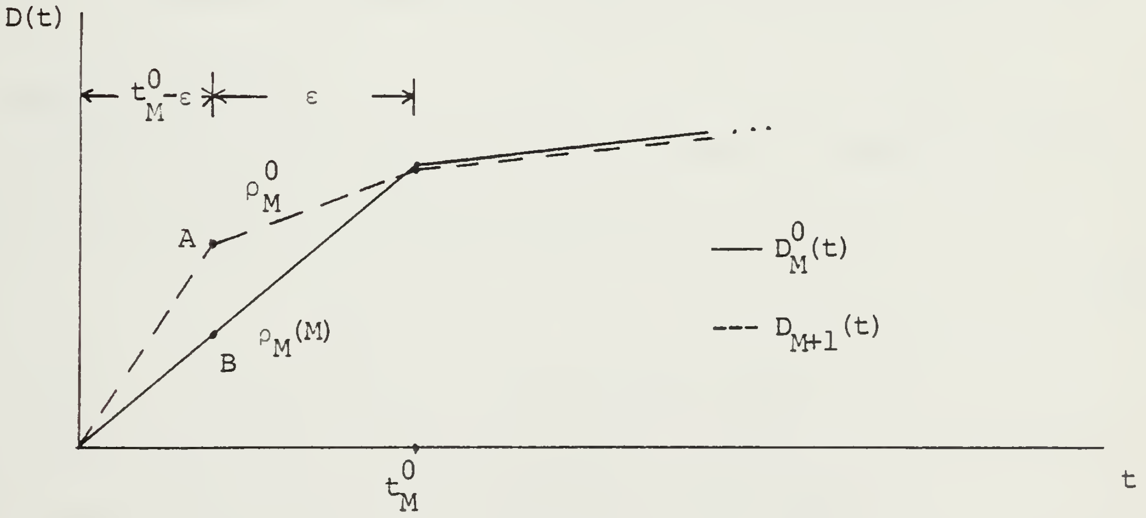


Fig. III.5. Stopping Rule

is identical to $D_M^0(t)$, and the points A and B in Fig. III.5 will coincide. Thus the stopping rule can be written as

$$\rho_M(M) = \rho_M^0 \rightarrow \text{stop} \qquad \text{(III.7)}$$

We can summarize the solution procedure "formally" in the following statement.

Algorithm:

$m = 0$

Loop

$m \leftarrow m+1$

solve MTP(m) and MRP(m)

if $\rho_M(M) = \rho_M^0$ then stop

Repeat

□

2. Computational Complexity

We now propose a conjecture which we believe to be correct although we have not been able to prove it rigorously.

Conjecture III.1

Let $D_M^0(t)$ be an optimal delivery function. Then the number of corner points M is bounded by $|N_0|$.

$$M \leq |N_0| \qquad \text{(III.9)}$$

where $|N_0|$ denotes cardinality of the set N_0.

□

We briefly discuss this conjecture in Appendix F.

We do not intend to comment further on the issue of computational complexity. There is a vast literature which

deals with efficient solutions of medium to large linear programs, and much of it is dedicated to multicommodity structures (see [3],[4]). The objective of this thesis is to show how to apply these LP methods to the solution of certain dynamic network problems.

We conclude this chapter with a simple example of a computer solution to optimal delivery problem. An additional example is provided in Appendix G.

3. Computer-Solution Example

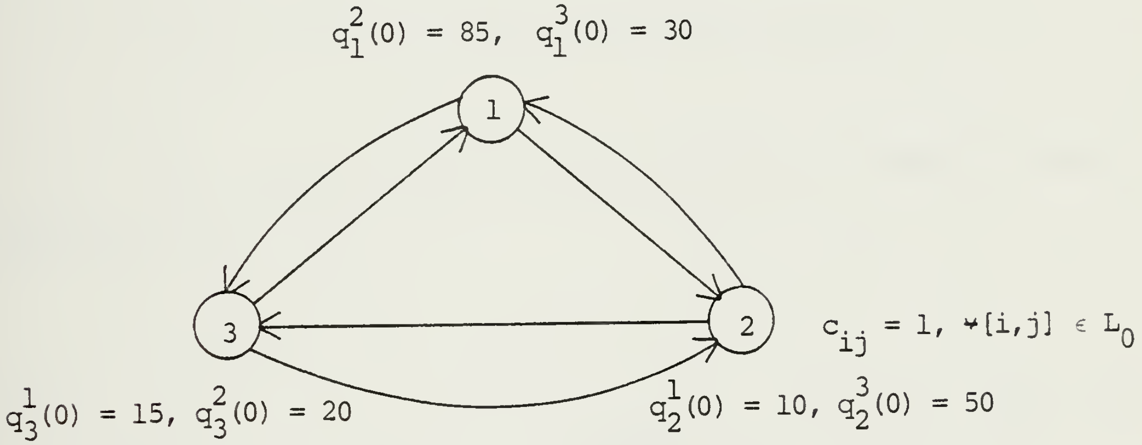


Fig. III.6. Delivery Problem

The delivery problem in Fig. III.6 was solved using the algorithm presented in this chapter. As a result the optimal delivery function and its generating optimal flow schedule were found.

The optimal delivery function is shown in Fig. III.7.

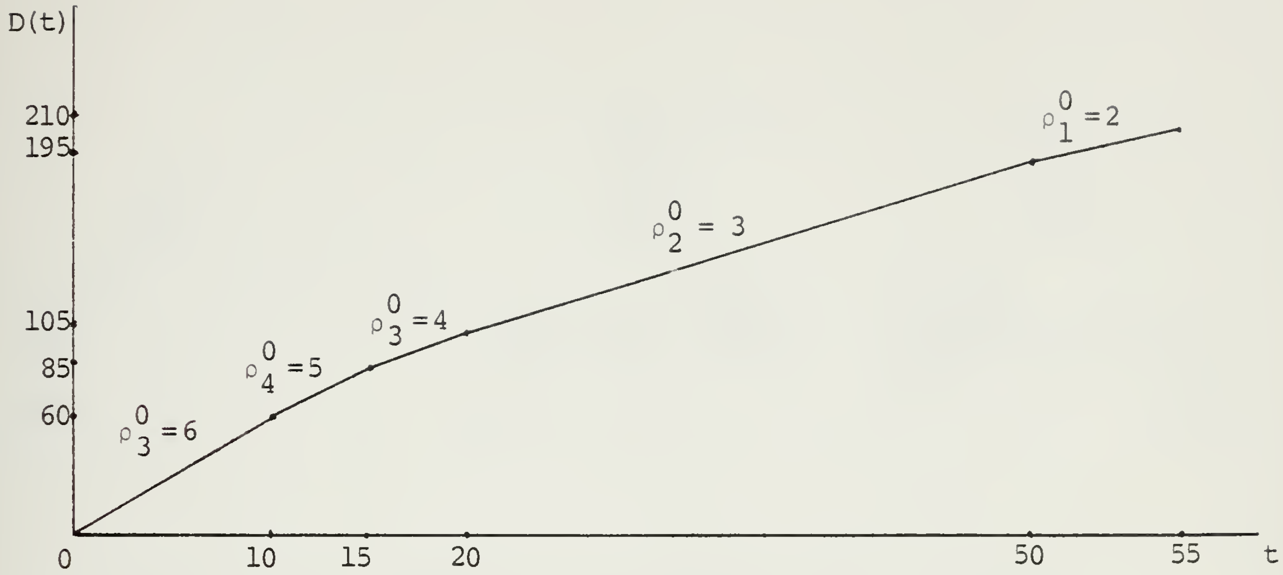


Fig. III.7. Optimal Delivery Function

The optimal flow schedule solution is shown in Figs. III.7a to e.

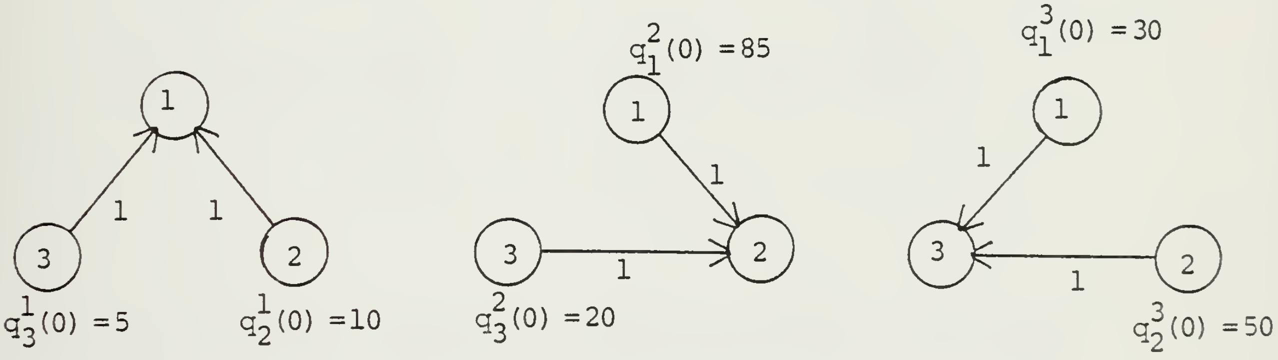


Fig. III.7a. Optimal Flow Schedule for $t \in [0,10]$

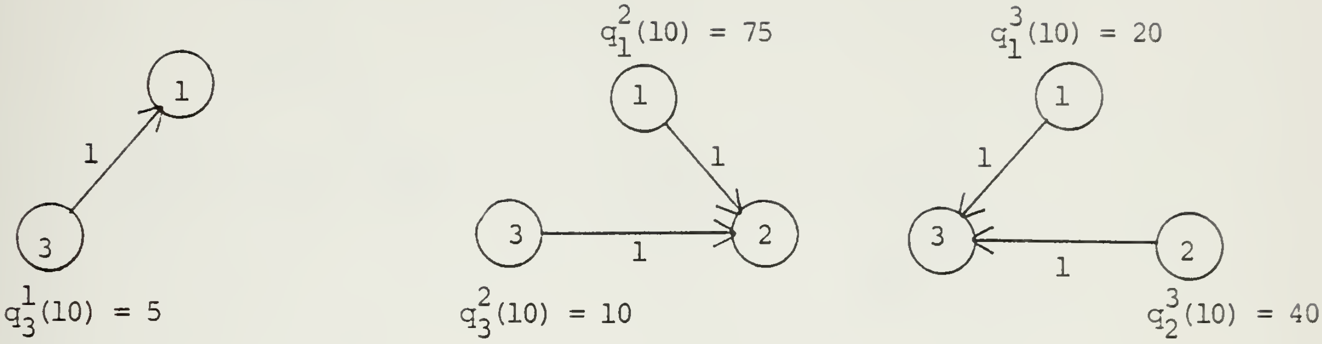


Fig. III.7b. Optimal Flow Schedule for $t \in (10,15]$

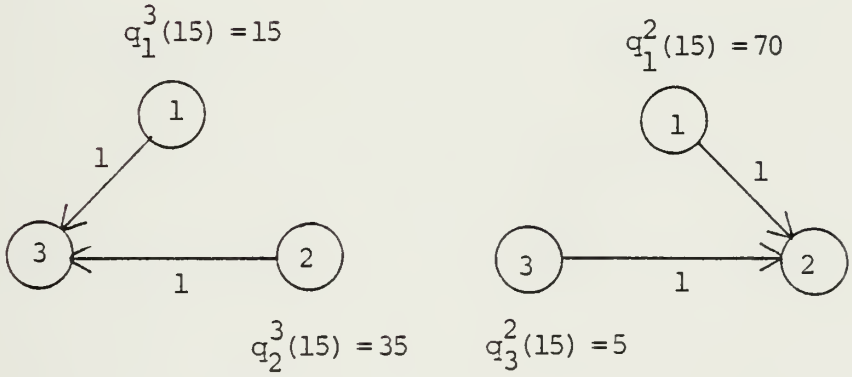


Fig. III.7c. Optimal Flow Schedule for $t \in (15,20]$

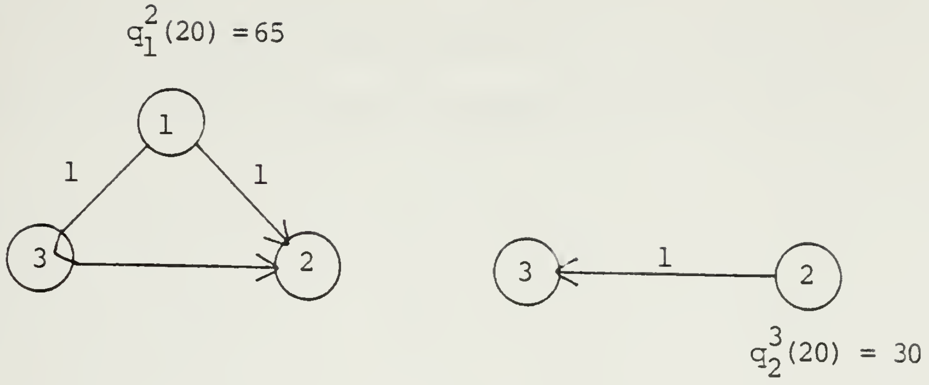


Fig. III.7d. Optimal Flow Schedule for t $\in$ (20,50]

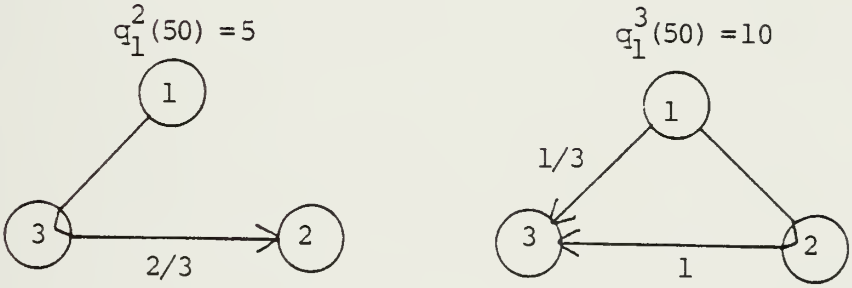


Fig. III.7e. Optimal Flow Schedule for t $\in$ (50,57.5]

It is somewhat surprising to find that even a simple delivery problem, such as the one considered here, should give rise to what might seem a complex optimal flow schedule. It is one of our aims in the following chapters, to show that what

appears to be a "random" looking optimal flow schedule has a lot of well defined structure to it. In particular, we shall return to discuss the above example at the end of Ch. V.

IV. PROPERTIES OF THE OPTIMAL SOLUTION

In Ch. III we have presented a sequential linear optimization methodology for solving the optimal delivery problem. The availability of potent computational tools, like the simplex method (see [5]), makes the attractiveness of linear programming obvious. But it is not only the computational advantage that LP enjoys. For example, the linear dependence of optimal solution to changes, within certain bounds, in the right hand side (RHS) makes the LP formulation specially appealing from the sensitivity analysis point of view. In this chapter, we use this and other known results of linear programming to determine some of the structural properties enclosed within a multicommodity delivery network problem.

A. THE FIRST CORNER POINT

1. On the Minimum Time t_1^0

a. The Primal Problem

For convenience we restate the First Minimal Time Problem (see (III.1b)) in standard LP form.

MTP(1): $\min t_1$

s.t.

$$\sum_{j(\neq i)} u_{ij}^k(1) - \sum_{j(\neq i)} u_{ji}^k(1) = q_i^k(0), \quad \forall (i,k) \in N_0$$

$$-t_1 c_{ij} + \sum_{k(\neq i)} u_{ij}^k(1) + s_{ij}(1) = 0, \quad \forall [i,j] \in L_0 \qquad \text{(IV.1)}$$

$$t_1,\ s_{ij}(1),\ u_{ij}^k(1) \geq 0, \quad \forall [i,j] \in L_0, \quad \forall (i,k) \in N_0$$

□

The slack variable $s_{ij}(1)$ gives the amount of unused capacity of link $[i,j]$ over the period $[0,t_1^0]$, $\forall [i,j] \in L_0$.

<u>Example</u>:

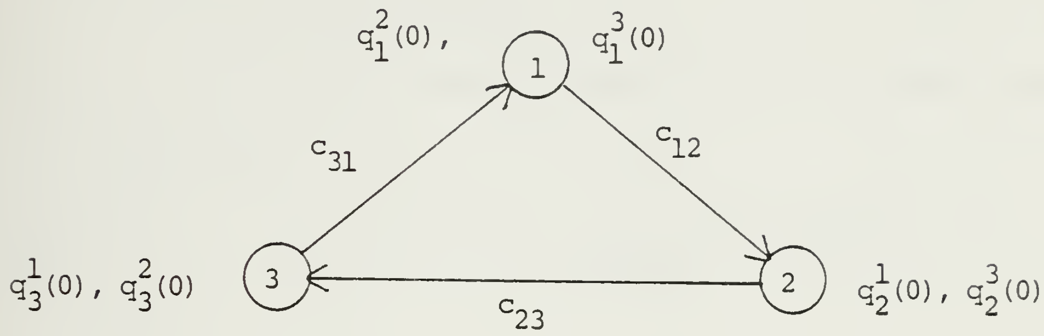


Fig. IV.1. Three Node Network Delivery Problem

The matrix formulation of (IV.1) for the problem in Fig. IV.1 is given below.

$$\min t_1$$

s.t.

$$\begin{vmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ -c_{12} & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ -c_{23} & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ -c_{31} & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \end{vmatrix} \begin{vmatrix} t_1 \\ s_{12}(1) \\ s_{23}(1) \\ s_{31}(1) \\ u_{12}^2(1) \\ u_{12}^3(1) \\ u_{23}^1(1) \\ u_{23}^3(1) \\ u_{31}^1(1) \\ u_{31}^2(1) \end{vmatrix} = \begin{vmatrix} q_1^2(0) \\ q_1^3(0) \\ q_2^1(0) \\ q_2^3(0) \\ q_3^1(0) \\ q_3^2(0) \\ 0 \\ 0 \\ 0 \end{vmatrix} \qquad \text{(IV.2)}$$

$$X^T(1) \triangleq (t_1, s_{12}(1), s_{23}(1), s_{31}(1), u_{12}^2(1), u_{12}^3(1), u_{23}^1(1)$$

$$u_{23}^3(1), u_{31}^1(1), u_{31}^2(1)) \geq \underline{0}$$

□

b. The Dual Problem

The dual linear program to (IV.1) can be written [6] as follows:

DMTP(1):
$$\max \sum_{(i,k)\in N_0} \sigma_i^k(1) q_i^k(0)$$

s.t.

$$- \sum_{[i,j]\in L_0} \pi_{ij}(1) c_{ij} \leq 1$$

$$\pi_{ij} \leq 0, \ \forall [i,j] \in L_0 \qquad \text{(IV.3)}$$

$$\sigma_i^k(1) - \sigma_j^k(1) + \pi_{ij}(1) \leq 0, \ \forall (i,k) \in N_0, \ \forall [i,j] \in L_0,$$

□

where

$$\sigma_k^k \triangleq 0$$

and

$$\Lambda(1) \triangleq (\Sigma(1), \Pi(1))$$

is a vector of dual variables. The vector $\Sigma(1)$ has n(n-1) components corresponding to the n(n-1) conservation constraints

in (IV.1). Vector $\Pi(1)$ has ℓ components corresponding to the ℓ capacity constraints in (IV.1).

Applying the complementary slackness theorem of linear programming (see [6], p. 77) to (IV.1) and (IV.3) we have

Lemma IV.1

Let $X(1)$ and $\Lambda(1)$ be optimal solutions for the primal and dual problems, respectively. Then $\forall [i,j] \in L_0$ and $\forall (i,k) \in N_0$

$$\text{(i)} \quad u_{ij}^k(1) > 0 \rightarrow -\pi_{ij}(1) = \sigma_i^k(1) - \sigma_j^k(1)$$

$$-\pi_{ij}(1) > \sigma_i^k(1) - \sigma_j^k(1) \rightarrow u_{ij}^k(1) = 0$$

$$\text{(ii)} \quad s_{ij}(1) > 0 \rightarrow \pi_{ij}(1) = 0 \qquad \text{(IV.4)}$$

$$\pi_{ij}(1) < 0 \rightarrow s_{ij}(1) = 0$$

where

$$\sigma_k^k(1) \triangleq 0$$

□

In Ch. II.A.2 we concluded that in our network model, the zero state is reachable from any initial state of the system, provided that $q_i^k(0) > 0 \rightarrow (i,k) \in N_0$. This is equivalent to saying that problem (IV.1) has a finite optimal solution. Using now the duality theorem of linear programming (see [6], p. 72) we conclude that the dual problem (IV.3) has

also a finite optimal solution, and the corresponding values of the objective functions are equal. An immediate consequence of this argument is the following lemma.

<u>Lemma IV.2</u>

Let t_1^0 be the first minimal time. Then

$$t_1^0 = \sum_{(i,k)\in N_0} \sigma_i^k(1) q_i^k(0) \qquad \text{(IV.5)}$$

□

c. Stability

In this paragraph we briefly discuss the question of uniqueness of the optimal dual solution and its relation to sensitivity analysis.

Consider a linear program in standard (matrix) form

$$\text{LP:} \qquad \min z = \Gamma \cdot X$$

s.t.

$$A X = b \qquad \text{(IV.6)}$$

$$X \geq \underline{0}$$

□

where Γ is the cost vector, A is the matrix of coefficients, b is the Right Hand Side (RHS) vector of requirements and X is the vector of primal variables.† And its dual:

---

†The interpretation of these quantities in the Minimal Time Problem context can be deduced by comparing (IV.6) to (IV.2).

DLP: $$\max z = \Lambda \cdot b$$

s.t.

$$\Lambda A \leq \Gamma \qquad \text{(IV.7)}$$

where Λ is the vector of dual variables.

Definition IV.1

For a given requirement vector b, an LP is said to be stable or to be at a stable point if the optimal dual solution Λ is unique. □

Let the LP (IV.6) be written

LP: $$\min z = \Gamma_B \cdot X_B + \Gamma_D \cdot X_D$$

s.t.

$$BX_B + DX_D = b \qquad \text{(IV.8)}$$

$$X_B, X_D \geq \underline{0}$$

□

where X_B is a vector of basic variables. We can transform (IV.8) into an equivalent linear program

LP: $$\min z = \Gamma_B B^{-1} b + (\Gamma_D - \Gamma_B B^{-1} D) X_D$$

s.t.

$$X_B + B^{-1} D X_D = B^{-1} b \qquad \text{(IV.9)}$$

$$X_B, X_D \geq \underline{0}$$

□

If X_B is a basic optimal solution ($X_D = \underline{0}$) it implies that:

(i) $X_B = B^{-1}b \geq \underline{0}$, i.e., it is primal feasible

$$\text{(IV.10)}$$

(ii) $\Gamma_D - \Gamma_B B^{-1} D \geq 0$, i.e., it is dual feasible

and

$$\Lambda = \Gamma_B B^{-1} \qquad \text{(IV.11)}$$

is the related optimal dual solution to DLP. It should be noted that the primal feasibility does not depend on the cost vector Γ, and that the dual feasibility does not depend on the requirement vector b.

Let us now consider a new requirement vector $\hat{b}$ such that

$$\hat{b} = b - \Delta b \qquad \text{(IV.12)}$$

where

$$\Delta b \triangleq \varepsilon \hat{1}_{\Delta b}$$

Thus, ε denotes the magnitude of the perturbation vector Δb and $\hat{1}_{\Delta b}$ is a unit vector in its direction. In what follows we will be concerned only with perturbations resulting in a requirement vector $\hat{b}$ for which the LP has a feasible solution. We call such perturbations, feasible. If Δb is a feasible perturbation then, since the dual feasibility (IV.10(ii)) does not depend on the requirement vector, the new solution $\hat{X}_B$ to the perturbed problem will be optimal if it is primal feasible, i.e.

$$\hat{X}_B = B^{-1}(b - \Delta b) = X_B - B^{-1}\Delta b \geq \underline{0} \qquad \text{(IV.13)}$$

An immediate consequence of condition (IV.12) is that if $X_B > \underline{0}$, i.e. if the optimal primal solution is not degenerate, then there exists some real and positive ε_1 such that condition (IV.13) is satisfied for all feasible directions $\hat{1}_{\Delta b}$ and all $0 < \varepsilon < \varepsilon_1$. On the other hand, if X_B is degenerate it is always possible to find some direction of perturbation such that (IV.13) is not satisfied unless the magnitude, ε, of this perturbation is zero.

Now, if primal feasibility is satisfied for the new (and therefore optimal) solution $\hat{X}_B$, then from (IV.9) we have

$$\min \hat{z} = \Gamma_B B^{-1}(b - \Delta b) = \min z - \Gamma_B B^{-1}\Delta b \qquad \text{(IV.14)}$$

The change

$$\Delta z \triangleq \min z - \min \hat{z} \qquad \text{(IV.15)}$$

in the optimal value of the cost function is

$$\Delta z = \Gamma_B B^{-1}\Delta b \qquad \text{(IV.16a)}$$

or in terms of the related optimal dual solution (IV.11)

$$\Delta z = \Lambda \cdot \Delta b \qquad \text{(IV.16b)}$$

Definition IV.2

Consider an optimal primal solution to LP and let Λ be the related optimal dual solution. We say that a feasible perturbation Δb is acceptable if the change Δz in the optimal value of the cost function due to that perturbation is

$$\Delta z = \Lambda \cdot \Delta b$$

□

It is clear from our discussion that if an optimal primal solution happens not to be degenerate then all feasible directions give rise to a nonzero acceptable perturbation. We now show that the same property holds for a weaker condition than primal solution non-degeneracy, namely stability.

Let Δb be a feasible perturbation and assume that (IV.13) is not satisfied unless $\varepsilon = 0$. This also implies that X_B is degenerate, i.e. there is at least one basic variable at zero level. If we force the magnitude of the perturbation to be somewhat larger than zero this causes one or more of the zero level basic variables to become negative. To satisfy primal feasibility, a new optimal basis $\hat{B}$ has to be found by exchanging the basic negative variable(s) with appropriate[†] non basic variable(s). The optimal value of the cost function for the perturbed problem can be written as

$$\min \hat{z} = \hat{\Gamma}_B \hat{B}^{-1}(b - \varepsilon \hat{1}_{\Delta b}) \qquad \text{(IV.17)}$$

[†]This is usually performed by a Dual Simplex pivot step(s) (see [7]).

If we let ε go to zero, then in the limit we have

$$\lim_{\varepsilon \to 0} \{\min \hat{z}\} = \min z \qquad \text{(IV.18)}$$

from which we conclude that the optimal perturbed basis $\hat{B}$ is also an optimal basis for the original (unperturbed) LP.

If we assume that the unperturbed optimal solution to LP was stable (unique optimal dual), then we have

$$\hat{\Gamma}_B \hat{B}^{-1} = \Lambda \qquad \text{(IV.19)}$$

where Λ is the related optimal dual solution to the original DLP. As a consequence of (IV.19) we state

Theorem IV.1

At a stable point, there exists a real and positive ε_1 such that all the feasible perturbations Δb, where

$$\Delta b = \varepsilon \hat{1}_{\Delta b}$$

and

$$0 < \varepsilon < \varepsilon_1$$

are acceptable.

□

The application of Thm. IV.1 to the optimal delivery problem, and in particular to MTP(1) results in

Corollary IV.1

At a stable point, let t_1^0 be the first minimal time and let $\Delta Q = \{\Delta q_i^k\}$ be a small change in the data queues sizes. Then

the change Δt_1^0 in the first minimal time which is caused by that perturbation of queues is

$$\Delta t_1^0 = \sum_{(i,k)\in N_0} \sigma_i^k(1)\Delta q_i^k \tag{IV.20}$$

□

Necessary and sufficient condition for optimal dual solution being unique is not obvious but partial results exist. The next two results (see [5], p. 144) indicate conditions for the existence of optimal stable solutions.

<u>Result IV.1</u>

If LP has at least one not degenerate optimal basic solution then the optimal dual solution is unique.

□

<u>Result IV.2</u>

If LP has a degenerate optimal solution then the optimal dual solution is never unique if it is not degenerate.

□

To make the digression from our main exposition as short as possible, we defer an illustration of stability in optimal delivery problems to Appendix D. For the rest of our study we will assume, without loss of generality, the existence of stability whenever we consider perturbation problems. This allows us to evaluate the effect of a perturbation on the optimal value of a cost function without considering the perturbed optimal basis, which is very convenient from the mathematical point of view.

Our interest in the stability concept is strongly motivated by yet another property that we find very useful for

analysis purposes. Consider the LP in (IV.6), and in particular let

$$X = (Y,S) \tag{IV.21}$$

where S is a vector of slack variables. We say that a constraint is <u>critical</u> if its slack variable is zero in <u>all</u> optimal primal solutions. In the next section we show that if an optimal solution to LP is stable, then all the critical constraints can be uniquely identified with the help of the related optimal dual solution. It is again worth noting, however, that there is no loss in generality associated with the stability assumption, since in Ch. V we present an algorithm to identify critical constraints even when the optimal primal solution is unstable.

d. Critical Sets

In [8] Ros showed that the optimal solution to the static flow routing problem imposes a certain partition of links and commodities into hierarchically related sets. Our study of the optimal delivery problem (= dynamic flow routing) reveals the existence of a similar structure. We obtain a somewhat more general result which involves critical sets of links and commodities as well as critical flow rates. In hindsight, the existence of a more general structure is not surprising in view of the higher complexity of dynamic versus static flow routing problems. Here we present results related to the critical sets of links and commodities. The critical flow structure is considered in Section A.2. Extension of these results to

subsequent corners of the optimal flow schedule will be studied in Section B.

We start with definition of the critical set L_1.

Definition IV.3

At a stable point, a link $[i,j] \in L_0$ belongs to the set L_1 iff $\pi_{ij}(1) < 0$.

□

The links in L_1 have the property that they must be saturated for the whole period $[0,t_1^0]$ in any flow schedule $F_K(t)$, $0 \le t \le t_1^0$ such that $F_K(t): Q(0) \to Q(t_1^0) = \underline{0}$. To see that this is true we need to show that the links in L_1 are saturated in all optimal primal solutions to MTP(1), and that no other links are saturated in at least one optimal solution. The first fact follows from (IV.4(ii)). Now suppose there is an optimal solution for which there is some link $[i,j] \in L_0$ such that $s_{ij}(1) = 0$ and $\pi_{ij}(1) = 0$. Applying the theorem of strong complementary slackness (see [9], p. 54) to (IV.1) and (IV.3) we have

Lemma IV.3

Given a pair of primal and dual programs with feasible solutions, there exists at least one pair of optimal solutions $X(1)$ and $\Lambda(1)$ for which:

$$\text{(i)} \quad u_{ij}^k(1) > 0 \leftrightarrow -\pi_{ij}(1) = \sigma_i^k(1) - \sigma_j^k(1)$$

$$-\pi_{ij}(1) > \sigma_i^k(1) - \sigma_j^k(1) \leftrightarrow u_{ij}^k(1) = 0$$

$$\text{(ii)} \quad s_{ij}(l) > 0 \leftrightarrow \pi_{ij}(l) = 0 \qquad \text{(IV.23)}$$

$$\pi_{ij}(l) < 0 \leftrightarrow s_{ij}(l) = 0$$

where

$$\sigma_k^k(l) \triangleq 0$$

□

Since by the assumption of stability the optimal dual solution is unique, Lemma IV.3(ii) implies an existence of a different optimal primal solution for which the link [i,j] is not saturated any more.

In passing we note that the optimal solutions referred to in Lemma IV.3 are not necessarily basic solutions. Also, Lemma IV.3 differs from Lemma IV.1 in that the properties (i) and (ii) are implied here in both directions.

To see that the links in L_1 are not only saturated in all optimal primal solutions but in any flow schedule that empties all data queues by time t_1^0, it suffices to refer to the proof technique of Thm. II.1.[†] There it is shown how a flow schedule $F_K(t)$, $0 \leq t \leq t_1^0$ such that $F_K(t)\colon\ Q(0) \to Q(t_1^0) = \underline{0}$ can be replaced by a constant flow schedule, $F_1(t) = F$, $0 \leq t \leq t_1^0$ such that $F_1(t)\colon Q(0) \to Q(t_1^0) = \underline{0}$. The construction used there assures us that a link $[i,j] \in L_0$ is saturated in the constant flow schedule iff it is saturated in $F_K(t)$ for the whole period $[0,t_1^0]$. This proves our point.

[†]See Appendix A.

Consider some commodity $(i,k) \in N_0$, which uses links in the set L_1. Let $\underline{P}(i,k)$ be a collection of all directed link chains connecting node i to node k. Suppose there is some chain $p_i^k \in \underline{P}(i,k)$ such that $p_i^k \cap L_1 = \emptyset$, i.e. the set L_1 and the chain p_i^k have no links in common. In this case it would be possible to divert more of commodity (i,k) flow into this chain. Recall from Lemma IV.3 that there exists at least one optimal primal solution for which all the links $[i,j] \not\in L_1$ are not saturated. This makes that flow diversion possible. But if a flow is diverted from the chains cutting through L_1, some links in L_1 will become unsaturated which contradicts the definition of the set L_1. Thus, any commodity (i,k) which is using links of L_1 enjoys the property

$$p_i^k \cap L_i \neq \emptyset, \quad \forall p_i^k \in \underline{P}(i,k) \qquad \text{(IV.24)}$$

If we denote by N_1 the set of all commodities that use links of the set L_1 then (IV.24) leads to the following theorem.

<u>Theorem IV.2</u>

The set L_1 is a disconnecting† set for commodities in N_1.

□

We say that a chain $p_i^k \in \underline{P}(i,k)$ is active if the flow of commodity k is non-zero on each of its links. Let us pick for any commodity (i,k) which is using links of the set

---

†We follow here the terminology used by Ford & Fulkerson in [10].

L_1, one of its active chains. From Lemma IV.1(i) we have that for each link of that chain

$$\sigma_\alpha^k(l) = -\pi_{\alpha\beta}(l) + \sigma_\beta^k(l), \ \forall [\alpha,\beta] \in p_i^k \tag{IV.25}$$

Using "chain substitution" we find that

$$\sigma_i^k(l) = -\pi_{i\alpha}(l) + \sigma_\alpha^k(l) = -\pi_{i\alpha}(l) - \pi_{\alpha\beta}(l) + \sigma_\beta^k(l)$$

$$= \ldots = - \sum_{[\alpha,\beta] \in p_i^k} \pi_{\alpha\beta}(l) \tag{IV.26}$$

Since at a stable point,

$$\pi_{\alpha\beta}(l) \begin{cases} < 0, & \text{iff } [\alpha,\beta] \in L_1 \\ = 0, & \text{otherwise} \end{cases}$$

We conclude that commodities $(i,k) \in N_0$ which use the set L_1 have their optimal dual variable $\sigma_i^k(l)$ positive. We summarize this observation in a formal definition of the set N_1.

<u>Definition IV.4</u>

At a stable point, a commodity $(i,k) \in N_0$ belongs to the set N_1 iff $\sigma_i^k(l) > 0$.

□

Again, as in the case of the set L_1, it can be shown that the set N_1 is unique, i.e. the same for all flow schedules $F_K(t)$, $0 \leq t \leq t_1^0$ such that $F_K(t): Q(0) \to Q(t_1^0) = \underline{0}$.

Let us consider again the optimal delivery function $D_M^0(t)$ and its generating optimal flow schedule $F_M^0(t)$. More precisely, let us look at the first part of $F_M^0(t)$, namely $F_M(1)$, $t_2^0 < t \leq t_1^0$. The delivery rate in this interval ρ_1^0, is the minimal delivery rate over all flow schedules $F_K(t)$, $0 \leq t \leq t_1^0$ such that $F_K(t): Q(0) \rightarrow Q(t_1^0) = \underline{0}$. From the results in this section we know that the critical set of links L_1 has to be saturated at all times $t \in [0,t_1^0]$ and consequently for all $t \in (t_2^0,t_1^0]$. The only commodities that can saturate L_1 are those in the critical set N_1. This being the case, we may expect that ρ_1^0 is the minimal value of flow rate with which the commodities in N_1 can saturate the link set L_1. We address this proposition in the next subsection.

2. On the Minimal Rate ρ_1^0

From the discussion of stability property in Paragraph (c) of the last subsection, we conclude that there exists a real ε, $\varepsilon > 0$ such that all feasible perturbations $\Delta Q = \{\Delta q_i^k\}$ that satisfy the equation

$$\sum_{(i,k)\in N_0} \sigma_i^k(1)\Delta q_i^k = \varepsilon \tag{IV.27}$$

where $\Delta q_i^k \geq 0$, $\forall(i,k) \in N_0$ are acceptable. This means that ε describes the change in the optimal value of the minimal time t_1^0 which is caused by that perturbation.

The plot in Fig. IV.2 depicts the optimal delivery function $D_1^0(t)$, $0 \leq t \leq t_1^0$ (the solution to MTP(1)) and its perturbed version, say $D_1^p(t)$, $0 \leq t \leq t_1^0 - \varepsilon$.

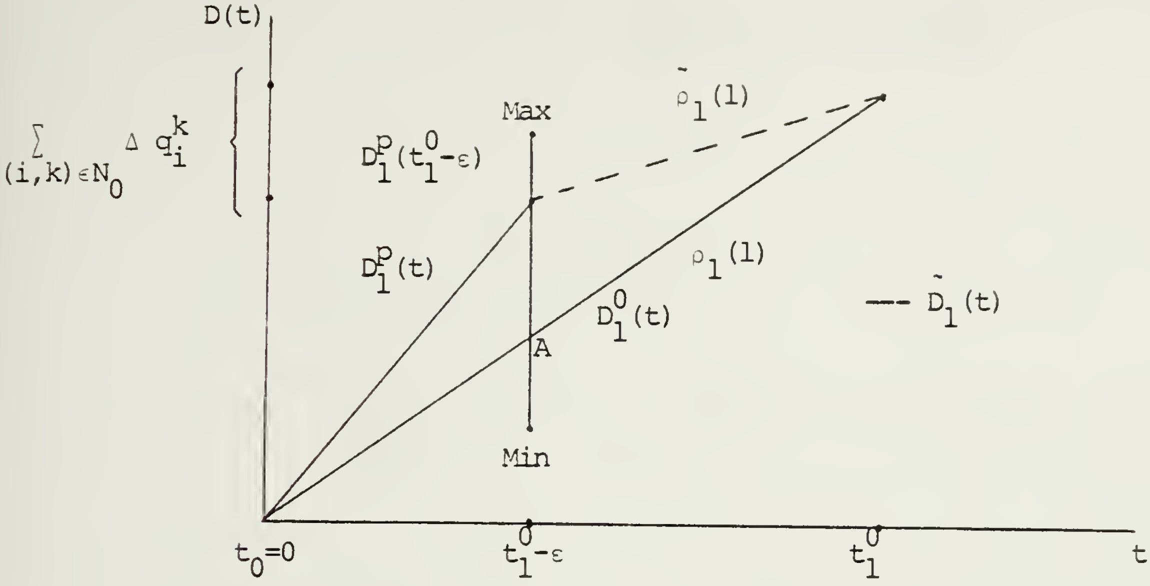


Fig. IV.2. The First Perturbation Problem

The perturbed flow schedule $F_1^p(t)$, $0 \leq t \leq t_1^0-\varepsilon$ delivers all but Δq_i^k, $\forall(i,k) \in N_0$ of each of the backlogged data queues $q_i^k(0)$ to its destination.

Suppose we can find now another flow schedule $\tilde{F}_1(t)$, $t_1^0-\varepsilon < t \leq t_1^0$ such that $\tilde{F}_1(t): \Delta Q \to \underline{0}$. This flow schedule will generate a delivery function $\tilde{D}_1(t)$, $t_1^0-\varepsilon < t \leq t_1^0$ which is shown by the broken line in Fig. IV.2. The value of the delivery function $D_1^p(t)$ at time $t_1^0-\varepsilon$ depends on the perturbation ΔQ that was chosen, and the locus of its possible values is shown by the vertical line in Fig. IV.2. Our objective is to find a perturbation ΔQ, for which the value $D_1^p(t_1^0-\varepsilon)$ will attain its maximum. This obviously is equivalent to finding a delivery function $\tilde{D}_1(t)$, $t_1^0-\varepsilon < t \leq t_1^0$ with minimal value

of the delivery rate $\tilde{\rho}_1(1)$. We can formalize now the above construction in the following <u>Perturbation Problem</u>:

PP(1):

$$\min \sum_{(i,k)\in N_0} \Delta q_i^k$$

s.t.

$$\sum_{(i,k)\in N_1} \sigma_i^k(1)\Delta q_i^k = \varepsilon$$

$$\Delta q_i^k - \{\sum_{j(\neq i)} \tilde{f}_{ij}^k - \sum_{j(\neq i)} \tilde{f}_{ji}^k\}\varepsilon = 0, \ \forall (i,k) \in N_0$$

$$\sum_{k(\neq i)} \tilde{f}_{ij}^k \leq c_{ij}, \ \forall [i,j] \in L_0 \qquad \text{(IV.28a)}$$

$$\Delta q_i^k, \tilde{f}_{ij}^k \geq 0, \ \forall (i,k) \in N_0, \ \forall [i,j] \in L_0$$

where the positive value ε is chosen small enough so that <u>all</u> ΔQ that satisfy (IV.27) are acceptable.

□

Due to stability assumption such an ε exists and we can rewrite (IV.28a) as

PP(1):

$$\min \sum_{(i,k)\in N_0} \tilde{r}_i^k$$

s.t.

$$\sum_{(i,k)\in N_1} \sigma_i^k(1)\tilde{r}_i^k = 1$$

$$\sum_{j(\neq i)} \tilde{f}_{ij}^k - \sum_{j(\neq i)} \tilde{f}_{ji}^k - \tilde{r}_i^k = 0, \ \forall (i,k) \in N_0 \qquad \text{(IV.28b)}$$

$$\sum_{k(\neq i)} \tilde{f}_{ij}^{k} \leq c_{ij}, \quad \forall [i,j] \in L_0$$

$$\tilde{r}_i^k, \tilde{f}_{ij}^k \geq 0, \quad \forall (i,k) \in N_0, \forall [i,j] \in L_0,$$

where

$$\tilde{r}_i^k \triangleq \frac{\Delta q_i^k}{\varepsilon}, \quad \forall (i,k) \in N_0$$

□

Moreover, for each acceptable perturbation and in particular for the optimal solution to PP(1) there <u>exists</u> a feasible flow schedule $F_1^p(t)$, $0 \leq t \leq t_1^0-\varepsilon$ such that $D_1^p(t_1^0-\varepsilon) = \sum_{(i,k)\in N_0} (q_i^k(0)-\Delta q_i^k)$. We stress this observation here, since as a result we can state the next theorem.

<u>Theorem IV.3</u>

At a stable point, the minimal value of the objective function of PP(1) is equal to ρ_1^0.

□

An immediate implication of Thm. IV.3 is that the minimal flow rate ρ_1^0 can be found in a much simpler way than by solving MRP(1). The number of variables and constraints in PP(1) is considerably smaller (we are not concerned with flow variables corresponding to the period $[0,t_1^0-\varepsilon]$) than in MRP(1). The only limitation is that the formulation of PP(1) is valid for stable points.

A more important consequence of Thm. IV.3 is that it enables us to obtain a new insight into the problem by a

careful study of (IV.28b). First, we want to indicate that problem (IV.28b) is feasible. As an example we can pick any ε, $0 < \varepsilon < t_1^0$ and a perturbation

$$\Delta q_i^k \triangleq \varepsilon r_i^k(1), \quad \forall (i,k) \in N_0, \tag{IV.29}$$

where $r_i^k(1)$ is the net delivery rate for the flow schedule $F_1^0(t)$, $\forall (i,k) \in N_0$. Then

$$\tilde{f}_{ij}^k(1) = f_{ij}^k(1), \quad \forall k, \ \forall [i,j] \in L_0 \tag{IV.30}$$

is a solution to PP(1). The corner point of the resulting delivery function $D_1^p(t)$ is schematically indicated as point A in Fig. IV.2. We say that the flow pattern[†] of the optimal solution to MTP(1) is a feasible solution to PP(1).

Another simple observation results from the structure of PP(1).

Lemma IV.4

In the optimal solution to PP(1)

$$\tilde{r}_i^k = 0 \quad \text{if} \quad \sigma_i^k(1) \leq 0 \tag{IV.31}$$

□

This, together with the fact that $\sigma_i^k(1) = 0$, $\forall (i,k) \notin N_1$ supports our previous observation that only commodities in N_1 play a role in the first segment of the optimal delivery function.

[†]By flow pattern we mean the flow composition, without any particular reference to time, e.g. an m-piece flow schedule is constructed of m flow patterns.

The perturbation equation

$$\sum_{(i,k)\in N_1} \sigma_i^k(1)\tilde{r}_i^k = 1$$

has a very important interpretation: it expresses the condition that the flows of commodities in N_1 need to satisfy in order to saturate the set L_1. This interpretation can be deduced by the following argument. The flow shcedule $\tilde{F}_1(t)$, $t_1^0-\varepsilon < t \leq t_1^0$ whose flow pattern is obtained from the solution of PP(1) must saturate L_1 with commodities from N_1, since it is a part of two segment flow schedule which delivers all the queues by t_1^0. Now, if we look carefully at the constraints of (IV.28b) there is nothing there, beside the perturbation equation, that can account for this saturation property.

<u>Theorem IV.4</u>

At a stable point, let $\sum(1)$ be the optimal dual solution to MPT(1). Then a feasible flow pattern F, of commodities in N_1 saturates the set L_1 iff

$$\sum_{(i,k)\in N_1} \sigma_i^k(1) r_i^k = 1.$$

□

The next topic we want to consider is the lower bound on ρ_1^0. Suppose that we remove the feasibility constraints from (IV.28b) and end up with

$$\hat{\rho}_1^0 = \min \sum_{(i,k)\in N_1} \tilde{r}_i^k$$

s.t.

$$\sum_{(i,k)\in N_1} \sigma_i^k(1)\tilde{r}_i^k = 1 \qquad \text{(IV.32)}$$

$$\tilde{r}_i^k \geq 0, \quad \forall (i,k) \in N_1.$$

□

It is obvious that $\hat{\rho}_1^0 \leq \rho_1^0$, and thus can serve as a lower bound. It also is easy to see that for this simple LP

$$\hat{\rho}_1^0 = 1/\sigma_{max}(1), \qquad \text{(IV.33)}$$

where

$$\sigma_{max}(1) = \max_{(i,k)\in N_1} \{\sigma_i^k(1)\}.$$

Theorem IV.5

At a stable point, the minimal rate ρ_1^0 satisfies

$$\rho_1^0 \geq 1/\sigma_{max}(1), \qquad \text{(IV.34)}$$

where

$$\sigma_{max}(1) = \max_{(i,k)\in N_1} \{\sigma_i^k(1)\}.$$

□

A related result states

Lemma IV.5

At a stable point, if the optimal delivery function of order one, $D_1^0(t)$, $0 \leq t \leq t_1^0$ satisfies

$$\sigma_i^k(1) = \sigma_{max}(1), \ \forall (i,k) \in N_0 \qquad \text{(IV.35)}$$

then it is the optimal solution to the delivery problem.

Proof:

Suppose that condition (IV.35) is satisfied. By Thm. IV.4 we must have that (in this case $N_1 = N_0$)

$$\sum_{(i,k) \in N_0} \sigma_i^k(1) r_i^k(1) = 1 \qquad \text{(IV.36)}$$

Substituting (IV.35) into (IV.36) results in

$$\sum_{(i,k) \in N_0} r_i^k(1) = 1/\sigma_{max}(1) \qquad \text{(IV.37)}$$

But the left hand side of (IV.37) is equal, by definition, to the total flow rate $\rho_1(1)$. Comparing with (Iv.34) we must conclude that $D_1^0(t)$ is already the optimal delivery function and $\rho_1(1) = \rho_1^0$.

□

An interesting case occurs when

$$\sigma_i^k(1) = \sigma_{max}(1), \ \forall (i,k) \in N_1 \qquad \text{(IV.38)}$$

and

$$N_1 \subset N_0$$

The delivery rate of the flow schedule solution to MTP(1), $F_1^0(t)$, may be viewed as composed of two constituents.

$$\rho_1(1) = \rho(N_1) + \rho(N_0 - N_1), \tag{IV.39}$$

where $\rho(A)$ denotes the delivery rate due to commodities in the set A. Using Thm. IV.4 and substituting (IV.38) into the perturbation equation we obtain

$$\rho(N_1) = 1/\sigma_{max}(1) \tag{IV.40}$$

Since all the backlogs are delivered by the time t_1^0, we conclude that

$$\rho(N_0-N_1) = \frac{1}{t_1^0} \sum_{(i,k)\in N_0-N_1} q_i^k(0) \tag{IV.41}$$

We now show that in this particular case there exists a solution to PP(1) in which ρ_1^0 attains its lower bound, namely $1/\sigma_{max}(1)$. Fig. IV.3 will help the reader to follow the argument.

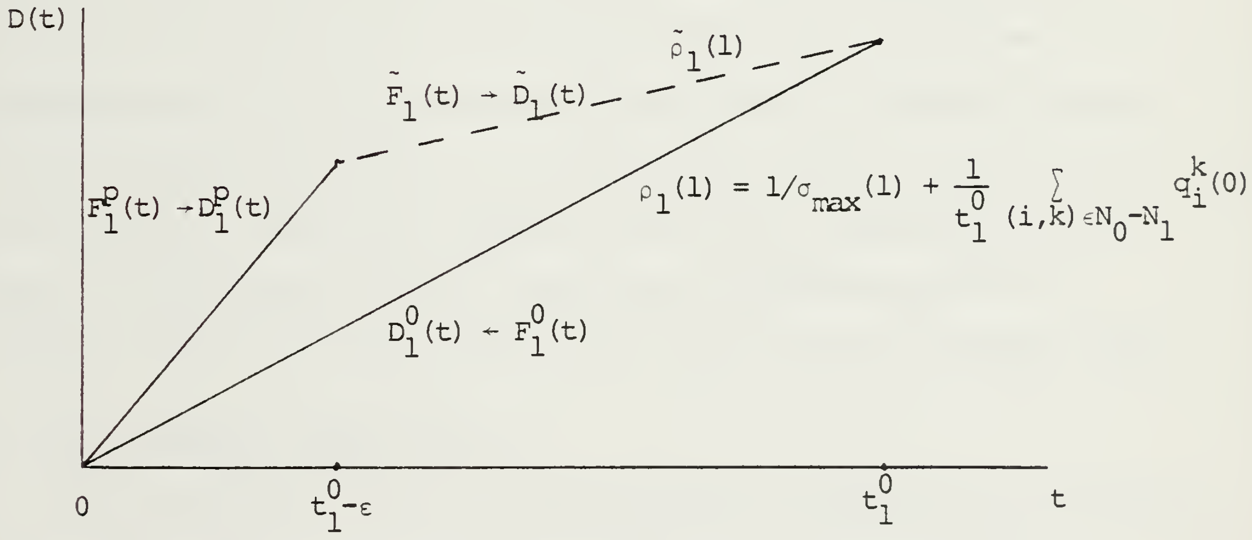


Fig. IV.3. Special Case of the First Perturbation Problem

Let us take the flow schedule solution to MTP(1), $F_1^0(t)$, $0 \leq t \leq t_1^0$ and increase the flows of all the commodities in the set $N_0 - N_1$. This is possible since none of these commodities passes through L_1, and thus all use only unsaturated links. As a result all the queues of commodities not in N_1 will be fully delivered prior to t_1^0, say by some time $t_1^0 - \varepsilon$. Let us denote this new segment of a flow schedule by $F_1^p(t)$, $0 \leq t \leq t_1^0 - \varepsilon$ as shown in Fig. IV.3. In the remaining time interval $(t_1^0 - \varepsilon, t_1^0]$, only the commodities in N_1 will continue to flow with a total rate $\tilde{\rho}_1(1) = 1/\sigma_{max}(1)$. This is denoted by $\tilde{F}_1(t)$, $t_1^0 - \varepsilon < t \leq t_1^0$ in Fig. IV.3. Comparing this value of $\tilde{\rho}_1(1)$ with the lower bound on ρ_1^0 (IV.34) makes our point.

Two observations are appropriate here. First, in this particular case it is not necessary at all to solve the PP(1) since the minimal rate value is known to be $\rho_0^1 = 1/\sigma_{max}(1)$ ahead of time. Second, there is a slight change in the notion of the next time problem. In MTP(2) we will be looking explicitly for the minimal time t_2^0 by which all the queues of commodities not in N_1 can be fully delivered to their destinations, given that queues of commodities in N_1 are delivered by the time t_1^0. These two observations have a considerable impact in the case of single destination networks where the special case we consider here turns out to be the general case.

B. THE SECOND CORNER POINT

We now begin a study parallel to that carried out in Section A, and define and describe the same type of concepts and

results that were derived for the first corner of the optimal delivery function. Since most results will be totally analogous we will not dwell on their proofs, unless the reasoning behind a proof is very much different from that in Section A.

1. On the Minimal Time t_2^0

The Second Minimal Time Problem can be written (see Def. III.3) in standard LP form as

MTP(2): $\min t_2$

s.t.

$$\sum_{p=1}^{2} \Big(\sum_{j(\neq i)} u_{ij}^k(p) - \sum_{j(\neq i)} u_{ji}^k(p) \Big) = q_i^k(0), \ \forall (i,k) \in N_1$$

$$\sum_{j(\neq i)} u_{ij}^k(2) - \sum_{j(\neq i)} u_{ji}^k(2) = q_i^k(0), \ \forall (i,k) \notin N_1$$

$$\sum_{j(\neq i)} u_{ij}^k(p) - \sum_{j(\neq i)} u_{ji}^k(p) - d_i^k(p) = 0, \ \forall (i,k) \in N_1, \quad p = 1,2$$

$$-t_1 c_{ij} + \sum_{k:(i,k)\in N_1} u_{ij}^k(1) = 0, \ \forall [i,j] \in L_1 \qquad \text{(IV.42)}$$

$$-t_1 c_{ij} + \sum_{k:(i,k)\in N_1} u_{ij}^k(1) + s_{ij}(1) = 0, \ \forall [i,j] \notin L_1$$

$$-t_2 c_{ij} + \sum_{k:(i,k)\in N_1} u_{ij}^k(2) = 0, \ \forall [i,j] \in L_1$$

$$-t_2 c_{ij} + \sum_{k(\neq i)} u_{ij}^k(2) + s_{ij}(2) = 0, \ \forall [i,j] \notin L_1$$

$$-t_1\rho_1^0 + \sum_{(i,k)\in N_1} \Big(\sum_{j(\neq i)} u_{ij}^k(1) - \sum_{j(\neq i)} u_{ji}^k(1)\Big) = 0$$

$$t_1 + t_2 = t_1^0$$

$$d_i^k(p),\ u_{ij}^k(p),\ s_{ij}(1),\ s_{ij}(2),\ t_1,\ t_2 \geq 0, \quad \forall [i,j] \in L_0$$

$$\forall (i,k) \in N_0$$

for given t_1^0 and ρ_1^0.

□

In the formulation of (IV.42) we have incorporated some of the results we derived earlier in this section. All the flows in the first time interval $(t_2,t_1^0]$ are due to commodities in the set N_1. This is so because any commodity $(i,k) \notin N_1$ ceases to flow (its queue is completely delivered), in an optimal solution, prior to time t_1^0. Also, there is no need for slack variables in capacity constraints for links in L_1, since we know that they must be saturated for all $t \in [0,t_1^0]$. We note that the third constraint in (IV.42) accounts for no intermediate data storage ($d_i^k(p)$ denotes the surplus variable for this constraint, $\forall (i,k) \in N_1$, $p = 1,2$). In Fig. IV.4 we indicate the basic relationship between the various parameters of (IV.42).

Let $\Lambda(2) \triangleq (\Sigma(2),\Omega(1),\Omega(2),\Pi(1),\Pi(2),\sigma_\rho(1),\sigma_t(2))$ be the vector of dual variables for the dual problem to (IV.42). The vector $\Sigma(2)$ has $|N_0|$ components corresponding to the first two delivery constraints of (IV.42). The vectors $\Omega(1)$ and $\Omega(2)$, have $|N_1|$ components each and they correspond to the

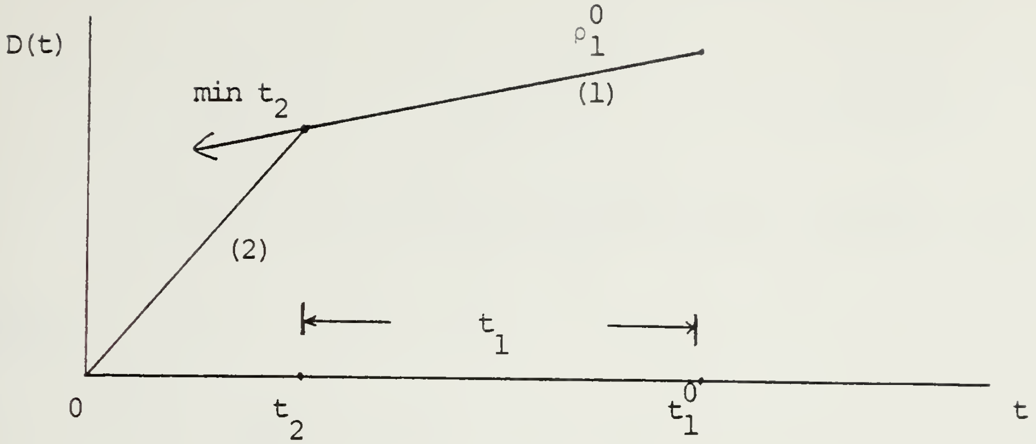


Fig. IV.4. The Second Minimal Time Problem

"no intermediate storage" constraints for both epochs (p = 1,2 in the third constraint of (IV.42)). Vectors $\Pi(1)$ and $\Pi(2)$ have $|L_0|$ components each and correspond to the capacity constraints in both time periods. Finally, $\sigma_\rho(1)$ and $\sigma_t(2)$ represent the dual variables corresponding to the rate and the time constraints, respectively.

Lemma IV.6 (IV.1)†

Let X(2) and $\Lambda(2)$ be optimal solutions for the primal and dual problems, respectively. Then $\forall [i,j] \in L_0$ and $\forall (i,k) \in N_0$

$$\text{(i)} \quad u_{ij}^k(2) > 0 \rightarrow -\pi_{ij}(2) = \sigma_i^k(2) - \sigma_j^k(2)$$

$$-\pi_{ij}(2) > \sigma_i^k(2) - \sigma_j^k(2) \rightarrow u_{ij}(2) = 0$$

---

†We will indicate in the brackets the analgous result in Section A.

(ii) $s_{ij}(2) > 0 \rightarrow \pi_{ij}(2) = 0$ if $[i,j] \not\in L_1$ (IV.43)

$\pi_{ij}(2) < 0 \rightarrow s_{ij}(2) = 0$ if $[i,j] \not\in L_1$

$\pi_{ij}(2)$ is unconstrained in sign if $[i,j] \in L_1$

where

$$\sigma_k^k(2) \triangleq 0$$

□

Application of the duality theorm of linear programming results in

Lemma IV.7 (IV.2)

Let t_2^0 be the second minimal time. Then

$$t_2^0 = \sum_{(i,k)\in N_0} \sigma_i^k(2) q_i^k() + \sigma_t(2) t_1^0 \qquad \text{(IV.44)}$$

□

It is worth noting that t_2^0 depends on the queue sizes and on t_1^0.

It is not difficult to see that at a stable point $\sigma_t(2) < 0$: Suppose that we resolve MTP(2) while we let $t_1^0 \rightarrow t_1^0 + \Delta t_1^0$. Now it is possible to deliver some small amount of each commodity in this additional interval and we expect t_2^0 to decrease as a result by some Δt_2^0. Recalling that at a stable point, $\sigma_t(2)$ relates the perturbation to the change in an objective function, we conclude that $\sigma_t(2) < 0$.

The discussion of stability in Section A.l.c applies without change to MTP(2) and we obtain:

Lemma IV.8 (Corollary IV.1)

At a stable point let t_2^0 be the second minimal time and let $(\Delta Q, \Delta t_1^0)$ be a small change in the data backlogs sizes and in the minimal total delivery time t_1^0. The corresponding change in the second minimal time Δt_2^0 is given by

$$\Delta t_2^0 = \sum_{(i,k)\in N_0} \sigma_i^k(2)\Delta q_i^k + \sigma_t(2)\Delta t_1^0 \qquad \text{(IV.45)}$$

□

Following the discussion in Section A.1.d we define the critical set L_2.

Definition IV.5 (IV.3)

At a stable point, link $[i,j] \in L_0$ belongs to the set L_2 either if $[i,j] \in L_1$ or if $[i,j] \notin L_1$ and $\pi_{ij}(2) < 0$.

□

The links in L_2 have the property that they are saturated for the whole period $[0,t_2^0]$ in any flow schedule $F_K(t)$, $0 \leq t \leq t_1^0$ such that

$$\begin{aligned} &\text{(i)} \quad F_K(t): \; Q(0) \to \emptyset \\ &\text{(ii)} \quad \rho_K(1) = \rho_1^0, \; \forall t \in (t_2^0, t_1^0] \end{aligned} \qquad \text{(IV.46)}$$

We note that the necessity to break the definition of L_2 into two exclusive cases results from the fact that $\pi_{ij}(2)$ is in general not restricted in sign (see Lemma IV.6(ii)), for links $[i,j] \in L_1$. Thus, a simple statement like

$[i,j] \in L_2 \leftrightarrow \pi_{ij}(2) < 0$, will not be correct. With the same precaution we define the set N_2 as follows.

Definition IV.6 (IV.4)

At a stable point, a commodity $(i,k) \in N_0$ belongs to the set N_2 either if $(i,k) \in N_1$ or if $(i,k) \notin N_1$ and $\sigma_i^k(2) > 0$.

□

The commodities in N_2 have the property that they must flow through the set L_2 and saturate it during the interval $[0,t_2^0]$ for all $F_K(t)$ such that (IV.46) is satisfied. Using the same arguments as those leading to (IV.24) we have

Theorem IV.6 (IV.2)

The set L_2 is a disconnecting set for commodities in L_2.

□

A false impression may result from our discussion, namely that each corner point in the optimal delivery function implies a new pair of critical sets. This is true only for the first corner point. It is possible that, for example, $L_2 = L_1$ and correspondingly $N_2 = N_1$. This would be the case when the corner point occurs because the commodities in N_1 can not maintain the minimal flow rate ρ_1^0 any longer (backwards in time). At this point a new rate ρ_2^0 will be computed without change in either L_1 or N_1. This brings us into the discussion of ρ_2^0.

2. On Minimal Rate ρ_2^0

Our objective here is to study the properties of ρ_2^0 and its interpretation. As before, we will rely on the stability assumption to simplify analysis.

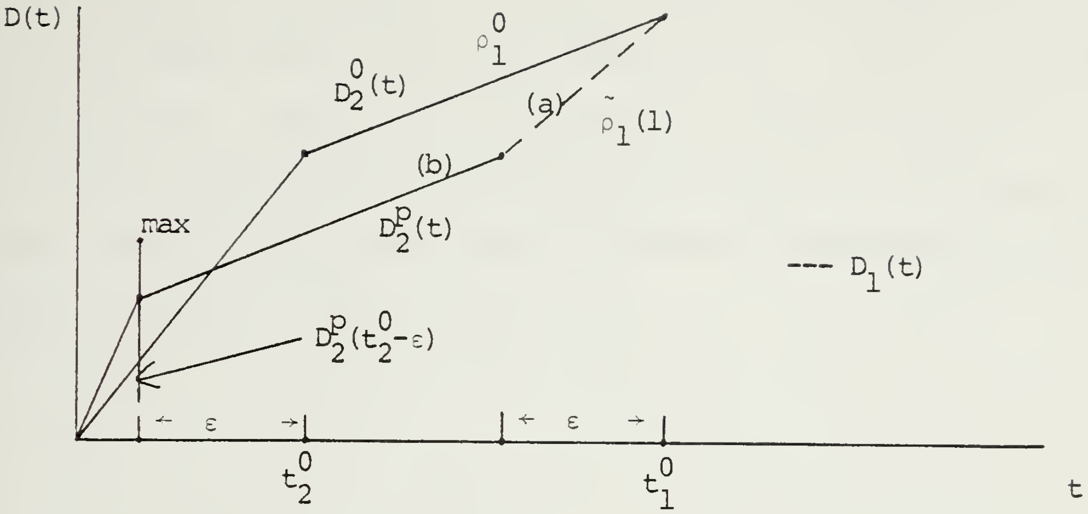


Fig. IV. 5a. The Second Perturbation Problem

Fig. IV.5a depicts the optimal delivery function $D_2^0(t)$, $0 \le t \le t_1^0$ (solution to MTP(2)) and its perturbed version $D_2^p(t)$, $0 \le t \le t_1^0-\varepsilon$. The perturbation equation derived in Section A (IV.27) is not applicable here. We want to find a perturbation ΔQ such that $D_2^p(t_2^0-\varepsilon)$ will be maximal but at the same time we must preserve previous results, i.e. the minimal rate ρ_1^0 and its duration $t_1^0-t_2^0$. The perturbed delivery function $D_2^p(t)$ in Fig. IV.5a has the required properties. It is described by a generalized perturbation equation

$$\varepsilon = \sum_{(i,k)\in N_0} \sigma_i^k(2)\Delta q_i^k + \varepsilon\sigma_t(2) \qquad \text{(IV.47a)}$$

which may be also written as

$$\frac{1}{1-\sigma_t(2)} \sum_{(i,k)\in N_0} \sigma_i^k(2) r_i^k = 1 \qquad \text{(IV.47b)}$$

As in PP(1), we require from the flow schedule $\tilde{F}_1(t)$, $t_1^0-\varepsilon \leq t \leq t_1^0$ to satisfy $\tilde{F}_1(t)$: $\Delta Q \to \underline{0}$. The broken line in Fig. IV.4 denotes the delivery function $\tilde{D}_1(t)$ which is generated by $\tilde{F}_1(t)$. Now, if we permute the segments denoted by (b) and (a), by permuting the corresponding flow schedules in time (this is always possible), we obtain the desired structure as shown in Fig. IV.5b.

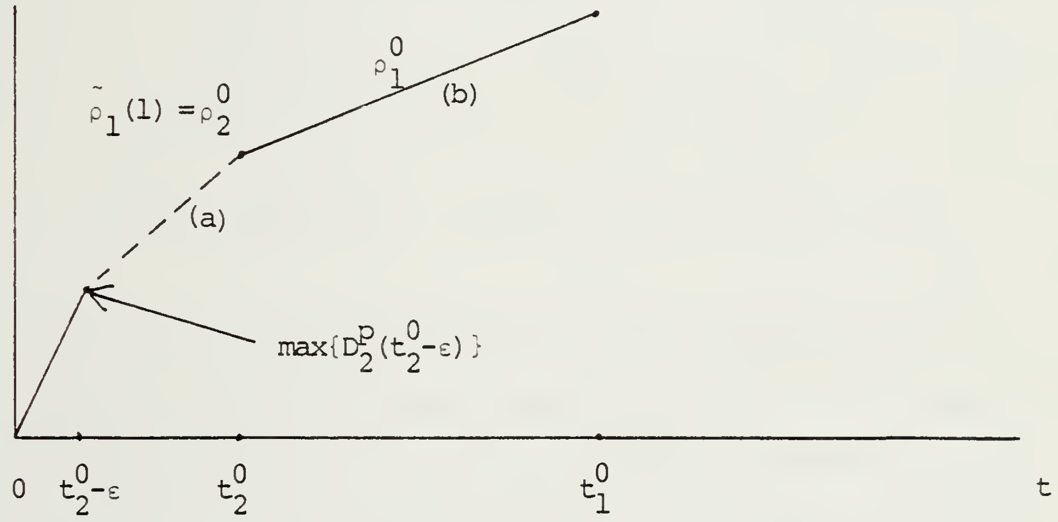


Fig. IV.5b. The Second Perturbation Problem (Permuted)

Since maximizing $D_2^p(t_2^0-\varepsilon)$ is equivalent to minimizing the delivery rate $\tilde{\rho}_1(1)$ we can finally state the PP(2).

PP(2):

$$\min \sum_{(i,k)\in N_0} \tilde{r}_i^k$$

s.t.

$$\frac{1}{1-\sigma_t(2)} \sum_{(i,k)\in N_0} \sigma_i^k(2)\tilde{r}_i^k = 1 \tag{IV.48}$$

$$\sum_{j(\neq i)} \tilde{f}_{ij}^k - \sum_{j(\neq i)} \tilde{f}_{ji}^k - \tilde{r}_i^k = 0, \quad \forall (i,k) \in N_0$$

$$\sum_{k(\neq i)} \tilde{f}_{ij}^k \leq c_{ij}, \quad \forall [i,j] \in L_0$$

$$\tilde{f}_{ij}^k, \tilde{r}_i^k \geq 0, \quad \forall [i,j] \in L_0, \forall (i,k) \in N_0$$

□

By the stability assumption, <u>all</u> perturbations ΔQ that satisfy the perturbation equation (IV.47b) are acceptable, and for each perturbation there exists a perturbed delivery function $D_2^p(t)$ of the form discussed. Due to this fact we can state the following.

<u>Theorem IV.7</u> (IV.3)

At a stable point, the minimal value of the objective function for PP(2) is equal to ρ_2^0.

□

The analysis of PP(2) leads to exactly the same results as for PP(1), aside from the slight modifications introduced by the factor $1/1-\sigma_t(2)$ in the generalized perturbation equation. We state those results here for completeness:

From the structure of PP(1) we have

Lemma IV.9 (IV.4)

In the optimal solution to PP(2)

$$\tilde{r}_i^k = 0, \text{ if } \sigma_i^k(2) \leq 0 \tag{IV.49}$$

□

This, together with the fact that $\sigma_i^k(2) = 0, \forall (i,k) \notin N_2$ supports our previous observation that only commodities in N_2 play a role in the second segment of the optimal delivery function.

The interpretation of the generalized perturbation equation is analogous to that in PP(1). Observing that this is the only condition in the statement of PP(2) that can account for the fact that commodities in N_2 must saturate L_2, we have

Theorem IV.9 (IV.4)

At a stable point, let $\Sigma(2)$ be the optimal dual solution to MTP(2). Then a feasible flow pattern F, of commodities in N_2, saturates the set L_2 iff

$$\frac{1}{1-\sigma_t(2)} \sum_{(i,k)\in N_2} \sigma_i^k(2) r_i^k = 1$$

□

The lower bound on ρ_2^0 has basically the same form as the lower bound on ρ_1^0.

Theorem IV.10 (IV.5)

At a stable point, the minimal rate ρ_2^0 satisfies

$$\rho_2^0 \geq \frac{1-\sigma_t(2)}{\sigma_{max}(2)}, \qquad \text{(IV.50)}$$

where

$$\sigma_{max}(2) = \max_{(i,k)\in N_2} \{\sigma_i^k(2)\}$$

□

We can be certain that the flow schedule solution $D_2^0(t)$, $0 \leq t \leq t_1^0$ generates the optimal delivery function if (cf. Lemma IV.5)

$$\sigma_i^k(2) = \sigma_{max}(2), \quad \forall (i,k) \in N_2 \qquad \text{(IV.51)}$$

The special case that we discussed at the end of the last section applies here as well. As we indicated there we will discuss it in detail in the single destination networks case in Chapter V.

C. SAMPLE PROBLEM

From previous sections it is apparent how everything studied so far generalizes and applies to any corner point. From a conceptual point of view only two corner points have to be studied, since corner three, or any subsequent corner presents no significant difference with respect to the second corner point. We believe that it will be more enlightening to present

a detailed study of a sample problem than repeat the same ideas again.

Example:

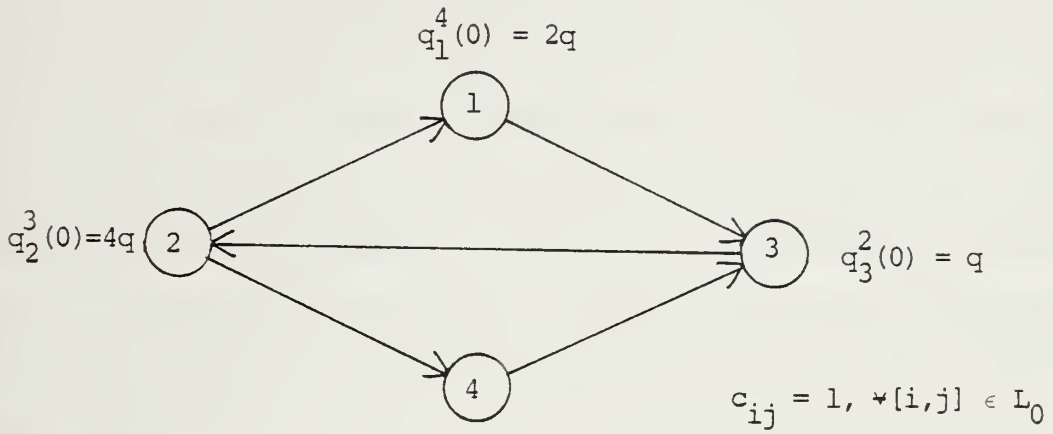


Fig. IV.6. Sample Delivery Problem

As the first step in solving the delivery problem in Fig. IV.6 we wish to find the composition of the first critical set N_1. Obviously, N_1 can be only one of the following:

(i) $\{(1,4)\}$, (ii) $\{(2,3)\}$, (iii) $\{(3,2)\}$, (iv) $\{(1,4),(3,2)\}$,

(v) $\{(3,2),(2,3)\}$, (vi) $\{(1,4),(2,3)\}$, (vii) $\{(1,4)(2,3),(3,2)\}$.

In what follows we consider explicitly each one of the above possibilities.

(i) -- For this case commodity (1,4) must have all its chains (here, there is only one) going through L_1, which implies that at least one of the links [1,3], [3,2] or [2,4] must be in L_1.

But this would imply that at least one of the commodities (2,3) or (3,2) uses links in L_1 and thus belongs to N_1. This contradicts (i).

(ii) -- For this case commodity (2,3) must have both of its available chains going through L_1, which implies that at least one link in each of the following pairs, $\{[2,1],[1,3]\}$ and $\{[2,4],[4,3]\}$ must be in L_1. It is not difficult to see that if link [2,1] is saturated so is [1,3]. Similarly, if link [4,3] is saturated so is [2,4]. This implies that commodity (1,4) uses links in L_1 and thus belongs to N_1. This contradicts (ii).

(iii) -- In principle the same type of argument applies here; we write in shortened notation:

$$N_1 = \{(3,2)\} \rightarrow [3,2] \in L_1 \rightarrow (1,4) \in N_1 \rightarrow N_1 \neq \{(3,2)\}.$$

(iv) - In this case $L_1 = \{[3,2]\}$. Suppose that commodity (1,4) is using link [3,2] with some flow rate α, and consequently commodity (3,2) is using that link (remaining capacity) with flow rate $1-\alpha$. Then it is true (remember that at this stage we are solving for a constant flow schedule) that

$$t_1^0 = \frac{q_1^4(0)}{\alpha} = \frac{q_3^2(0)}{1-\alpha}$$

or equivalently

$$t_1^0 = \frac{2q}{\alpha} = \frac{q}{1-\alpha}$$

We can eliminate the flow variable α by using the law of proportions, namely $\frac{a}{b} = \frac{c}{d} = \frac{a+c}{b+d}$, such that

$$t_1^0 = \frac{2q + q}{\alpha + 1-\alpha} = 3q$$

Now, we can also solve for α, to obtain $\alpha = 2/3$. The remaining capacity on links [1,3] and [2,4] is thus 1/3 (actually, less than 1/3, say $\frac{1}{3}-\varepsilon$, $\varepsilon > 0$, since commodity (2,3) $\notin N_1$ and in the solution to MTP(1) it must flow only through unsaturated links). We find that the time required to deliver the queue of commodity (2,3) is

$$t_1 = \frac{4q}{2(\frac{1}{3} - \varepsilon)} > 6q > t_1^0,$$

which contradcits our assumption in (iv).

(v) -- We have already seen in (iii) that if commodity (3,2) $\in N_1$ then also commodity (1,4) $\in N_1$, which contradicts (v).

Exercising our advantage over the reader in knowing the solution, let us study possibility (vii) prior to (vi).

(vii) -- This case is equivalent to (iv) if we let $\varepsilon \equiv 0$ (now commodity (2,3) $\in N_1$) in our discussion there. Then it must be true (for an optimal solution to MPP(1)) that

$$t_1^0 = \frac{q_1^4(0)}{\alpha} = \frac{q_3^2(0)}{1-\alpha} = \frac{q_2^3(0)}{2(1-\alpha)}$$

or equivalently,

$$t_1^0 = \frac{2q}{\alpha} = \frac{q}{1-\alpha} = \frac{4q}{2(1-\alpha)}$$

where the last equality can never be satisfied for non-zero α. We conclude that (vii) cannot be accepted as correct.

(vi) -- For this case $L_1 = \{[1,3],[2,4]\}$ (the remaining commodity (3,2) uses the unsaturated link [3,2] only, as required). The chain flow decomposition of the constant flow schedule solution to MTP(1) is shown in Fig. IV.7.

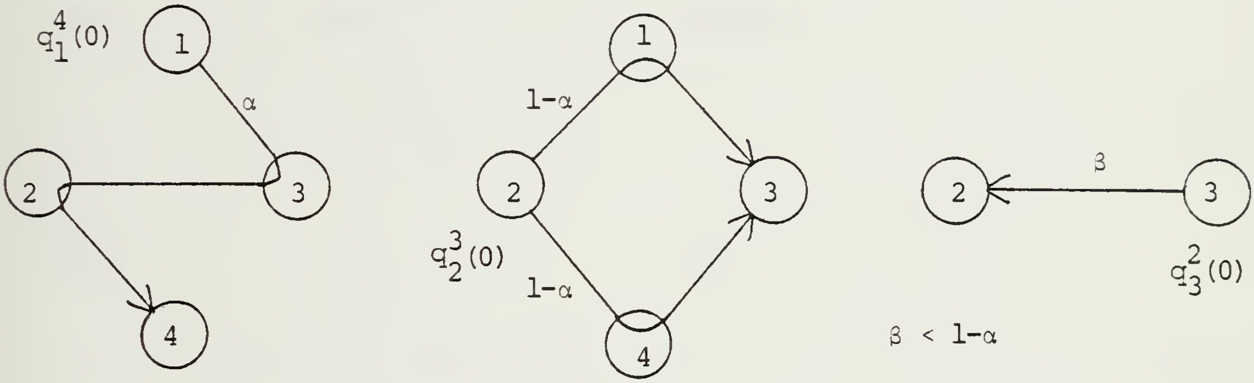


Fig. IV.7. Chain Flow Decomposition

It must be true then that

$$t_1^0 = \frac{q_1^4(0)}{\alpha} = \frac{q_2^3(0)}{2(1-\alpha)} = \frac{q_3^2(0)}{\beta}, \qquad \text{(IV.52)}$$

where

$$\beta < 1-\alpha.$$

If we consider in (IV.52) only those elements that have the variable α associated with them, we have (using the law of proportions)

$$t_1^0 = \frac{2q_1^4(0) + q_2^3(0)}{2\alpha + 2 - 2\alpha} = q_1^4(0) + \frac{1}{2} q_2^3(0) \qquad \text{(IV.53)}$$

For the queues sizes we have selected,

$$t_1^0 = 4q \qquad \text{(IV.54)}$$

Using (IV.52) and (IV.54) we can evaluate α,

$$\alpha = \frac{2q}{4q} = \frac{1}{2}, \qquad \text{(IV.55)}$$

and β

$$\beta = \frac{q}{4q} = \frac{1}{4}, \qquad \text{(IV.56)}$$

which is less than $1-\alpha = \frac{1}{2}$ (as desired).

Now that we have established t_1^0, N_1 and L_1 we turn to calculate ρ_1^0, the minimal delivery rate in the first interval of the optimal delivery function. Recalling that (see (IV.5))

$$t_1^0 = \sum_{(i,k)\in N_1} \sigma_i^k(1) q_i^k(0),$$

and comparing to (IV.53) results in

$$\sigma_1^4(1) = 1, \quad \sigma_2^3(1) = \frac{1}{2}. \qquad \text{(IV.57)}$$

To obtain ρ_1^0 we have to solve PP(1) (IV.28b),

PP(1): $$\min \sum_{(i,k)\in N_1} \tilde{r}_i^k$$

s.t.

$$1 = \sum_{(i,k)\in N_1} \sigma_i^k(1)\tilde{r}_i^k, \qquad \text{(IV.58)}$$

and

$\{\tilde{r}_i^k\}$, $\forall (i,k) \in N_1$ is a feasible set of delivery rates.

□

Consulting (IV.57), problem (IV.58) can be solved by inspection yielding

$$(\tilde{r}_1^4, \tilde{r}_2^3) = (1,0), \qquad \text{(IV.59)}$$

which is shown in Fig. IV.8.

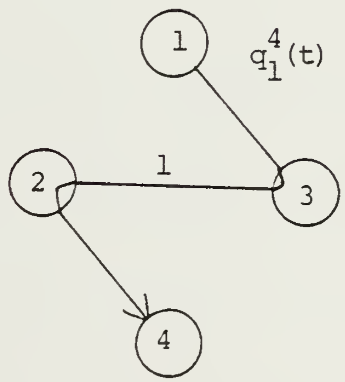


Fig. IV.8. Solution to the First Minimal Time Problem

Thus, we conclude that

$$\rho_1^0 = 1.$$

The fact that the solution to PP(1) is unique, considerably simplifies the formulation of MTP(2). Let $Q(t_2)$ denote the system state at time t_2, $0 < t_2 \leq t_1^0$. From our study we know that there exists a constant flow schedule $F(t)$, $0 \leq t \leq t_2$ such that $F(t)\colon Q(0) \to Q(t_2)$. This fact can be expressed by the chain flow decomposition in Fig. IV.7 if we substitute $Q(0)-Q(t_2)$ for $Q(0)$. Also, we cannot be sure any more (nor is it required) that $\beta < 1-\alpha$, since it is possible that $N_2 = N_1 \cup (3,2)$. Because of the uniqueness property of the optimal solution to PP(1), we have that the components of $Q(t_2)$ are

$$q_i^k(t_2) = \begin{cases} (t_1^0 - t_2)\rho_1^0, & (i,k) = (1,4) \\ 0, & \text{otherwise.} \end{cases} \qquad \text{(IV.60)}$$

Since $q_i^k(t_2) \leq q_i^k(0)$, $\forall (i,k) \in N_0$ then in particular

$$q_1^4(t_2) = (t_1^0 - t_2)\rho_1^0 \leq q_1^4(0), \qquad \text{(IV.61a)}$$

or equivalently

$$t_2 \geq t_1^0 - q_1^4(0) = 2q \qquad \text{(IV.61b)}$$

Equation (IV.61b) presents a lower bound on the minimal value of t_2, i.e. $t_2^0 \geq 2q$.

Let us check whether it is possible that $t_2^0 = 2q$. This is equivalent to asking whether the queues of commodities (2,3) and (3,2) can be delivered within the time interval [0,2q]. Since commodity (2,3) and (3,2) have no common links in their respective chain flow decomposition we can assign a flow rate of 2 to commodity (2,3) and a flow rate of 1 to commodity (3,2). It is easy to see that it will take $\frac{4q}{2} = 2q$ and $\frac{q}{1} = q$ units of time to deliver the respective queues to their destinations.

The complete optimal flow schedule is now shown in Fig. IV.9.

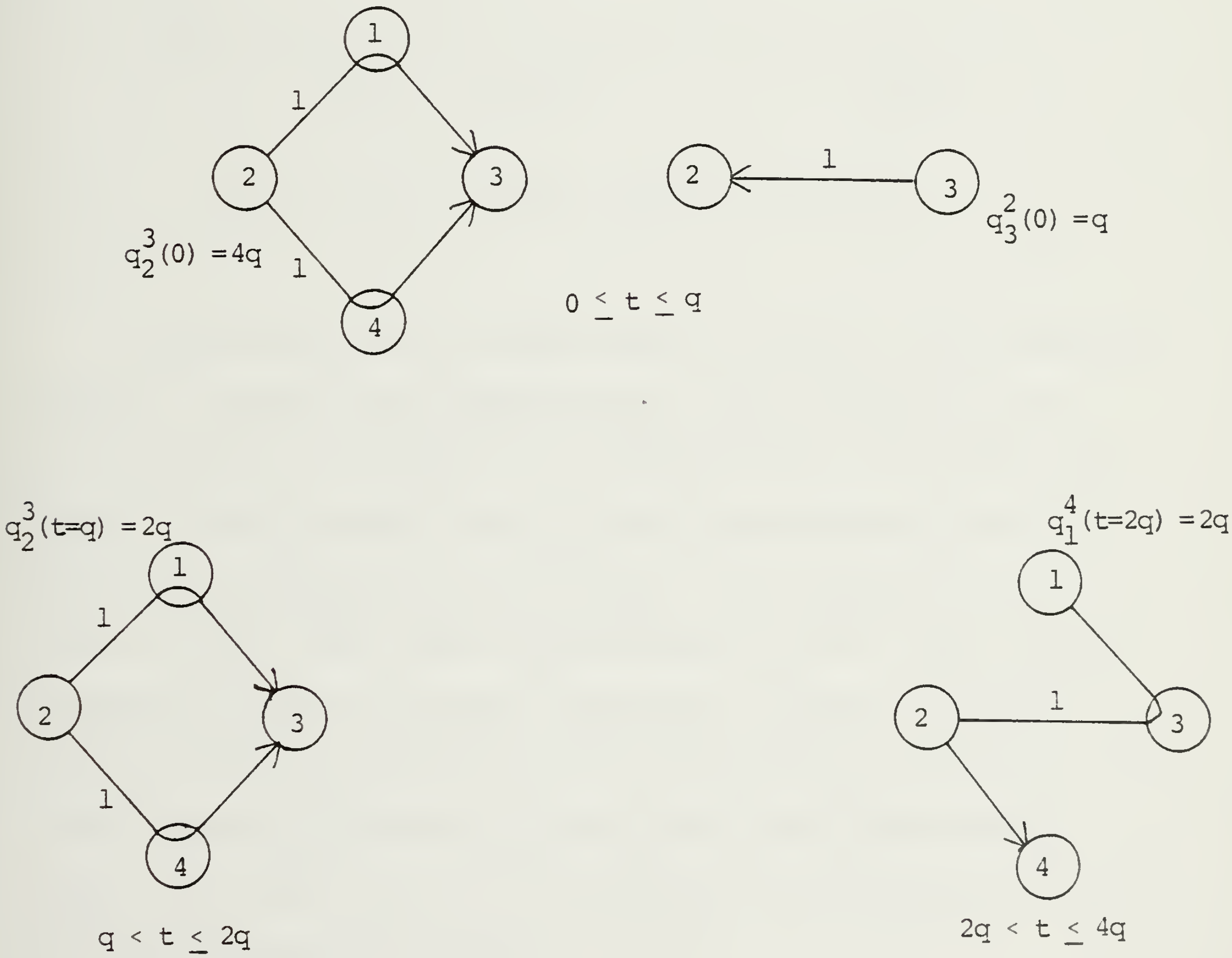


Fig. IV.9. Optimal Flow Schedule

The optimal delivery function is shown in Fig. IV.10.

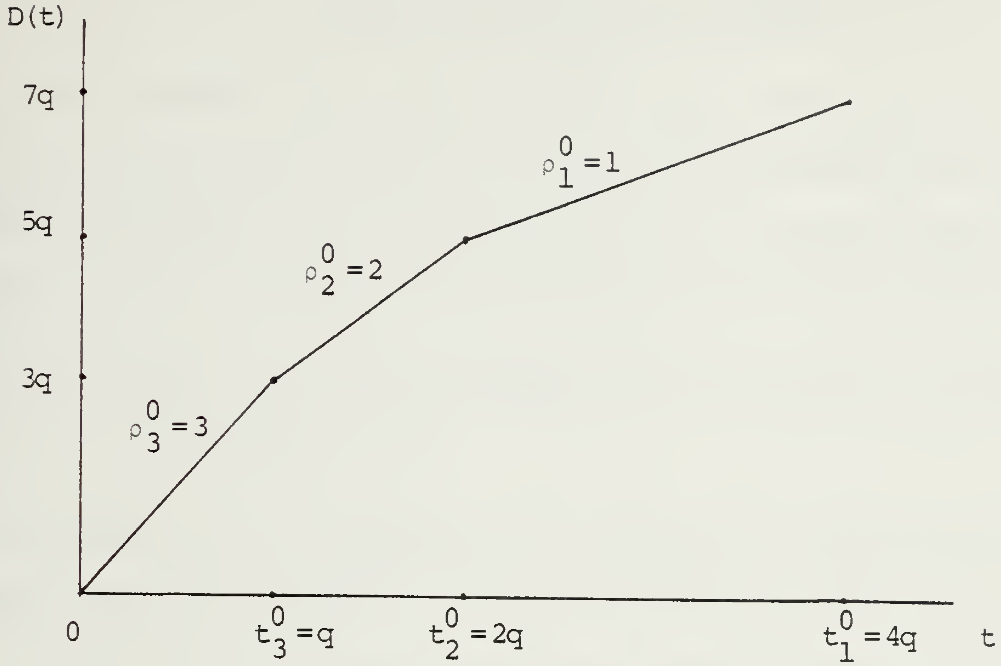


Fig. IV.10. Optimal Delivery Function.

□

A remaining problem, which does not arise in the foregoing example, concerns the possibility of loops existing in the flow solution. Evidently, such loops cannot affect the optimality of the delivery function, but nonetheless their existence is not aesthetically pleasing. Two comments are in order: First, such loops cannot appear if the input traffic between all pairs of nodes is non-zero. Second, given the time and rate parameters of the optimum delivery function, all loops can be eliminated by solving the last flow problem again, but this time with the objective of maximizing the sum of the link slack variables.

V. SINGLE DESTINATION NETWORKS

By a Single Destination Network (SDN, for short) we mean that all the data traffic in a network is destined to a single node. Without loss of generality we will assume that node to be n, $n \in V$. Accordingly, for present purposes we redefine

N_0 = set $\{i\}$ of all nodes such that node i can communicate to node n, $i \neq n$.

In general, we will simplify the notation by dropping out the destination indication, which is implicitly understood to be n.

Naturally, all the results that were derived for multi-commodity case apply to SDN. Their generality though, tends to hide some of the unique properties of SDN which we study here.

A. THE FIRST CORNER POINT

The single destination variant of MTP(1) is given by

MTP(1): $\min t_1$

s.t.

$$\sum_{j(\neq i)} u_{ij} - \sum_{j(\neq i)} u_{ji} = q_i(0), \quad \forall i \in N_0$$

$$-t_1 c_{ij} + u_{ij} + s_{ij} = 0, \quad \forall [i,j] \in L_0 \qquad \text{(V.1)}$$

$$t_1, u_{ij}, s_{ij} \geq 0, \quad \forall [i,j] \in L_0$$

□

where we have used the previously defined change of variables

$$u_{ij} \triangleq t_1 f_{ij}, \quad \forall [i,j] \in L_0 .$$

The dual linear programming form to MTP(1) can be formulated exactly as in the multicommodity case. The same statement applies to the discussion of stability and to the definitions and properties of the sets L_1 and N_1. Let us assume, at this point, that the optimal solution to MTP(1) is stable. Hence, the sets L_1 and N_1 are uniquely determined by MTP(1). Later on in this chapter, when we discuss the solution algorithm for SDN, we will relax this assumption.

By Thm. IV.2, the set L_1 is a disconnecting set for nodes in N_1, i.e. every chain that connects any of the nodes $i \in N_1$ to the destination n has at least one of its links in the set L_1. We shall see in the next few paragraphs that the maximal flow rate $\rho_{max}(N_1)$ with which data can be delivered from the set N_1 to the destination node n is given by the Max-flow Min-cut theorem (see [10], p. 11]

$$\rho_{max}(N_1) = \max_{\{F\}} \sum_{i \in N_1} r_i = CS(N_1), \qquad (V.2)$$

where $CS(N_1)$ is the value of a minimal cut-set, separating the set N_1 from node n.

It is important to realize that in the SDN case the multicommodity delivery problem turns into single commodity† problem

---

†Discussion of the differences between single commodity and multicommodity can be found in [9].

which makes the notion of minimal cut-set meaningful. To prove (V.2), consider an optimal flow solution to MTP(1). A typical source node i, $i \in N_1$ is shown in Fig. V.1(a). The initial queue $q_i(0)$ is diminished with a rate $r_i(1)$ (net delivery rate of data from node i) such that $q_i(t_1^0) = 0$. Part (b) in Fig. V.1 describes an equivalent setup made up of a virtual node v connected to node i by an infinite capacity link [v,i]. There is a constant data input with rate $r_i(1)$ to node v.

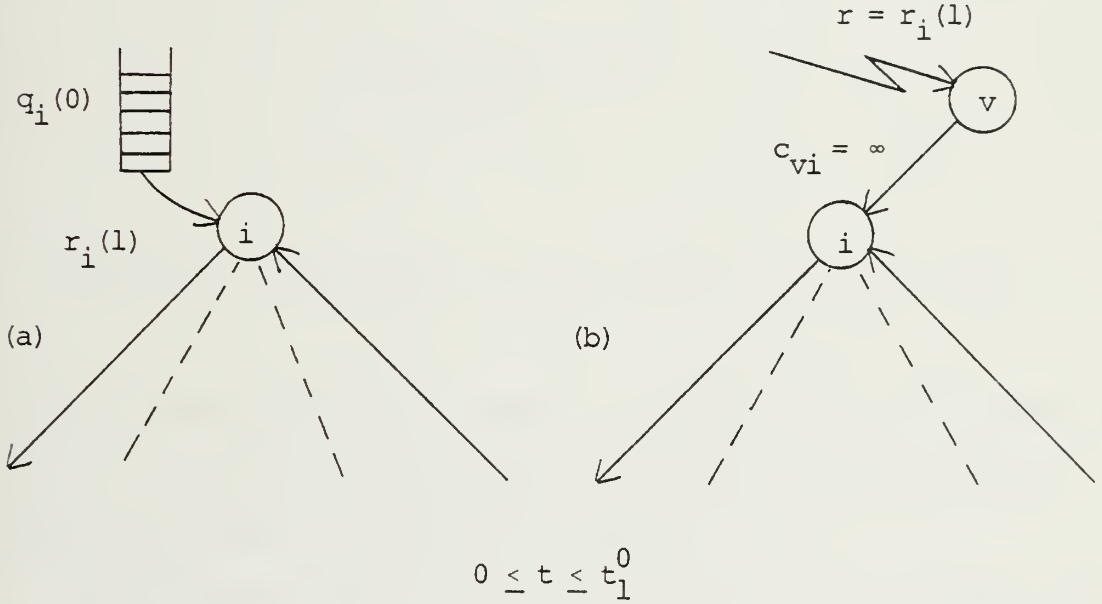


Fig. V.1. A Source Node in an Optimal Solution to MTP(1)

If we extend the model to all nodes in the set N_1, the result is as shown in Fig. V.2.

Recall now that the critical set N_1 consists of all nodes i, $i \in N_0$ which determine the minimal time t_1^0. An implication of this characterization is that it is impossible to increase any of the initial queues (while not changing the others)

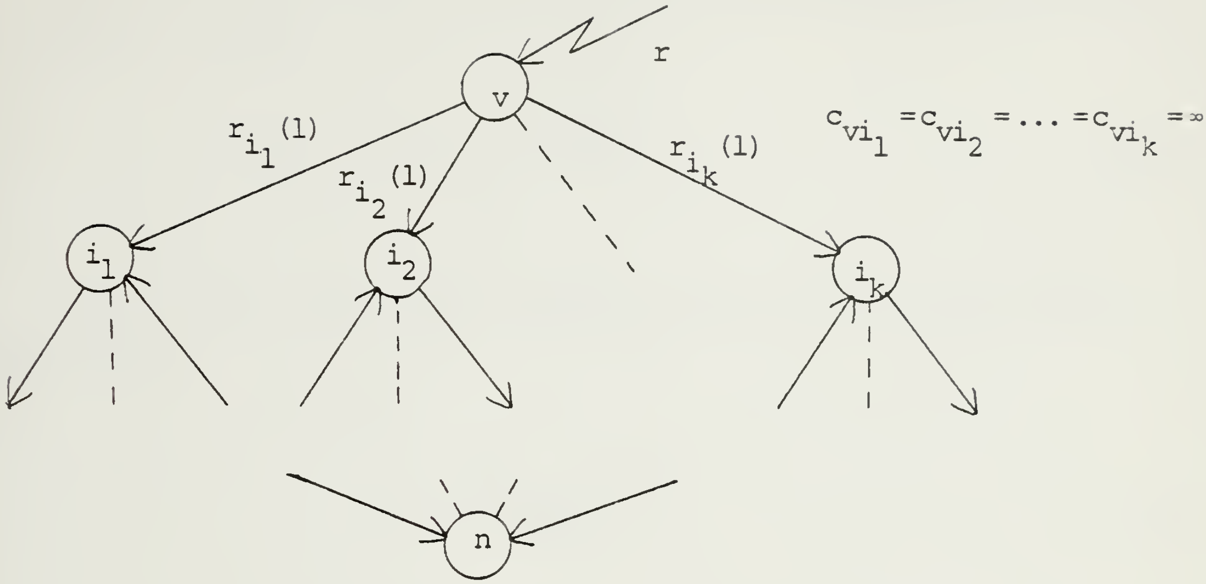


Fig. V.2. The Flow Pattern of $F_1^0(t)$, $0 \leq t \leq t_1^0$

without causing t_1^0 to increase. Equivalently no delivery rate $r_i(1), i \in N_1$ can be increased, even momentarily, without decreasing some of the other delivery rates.

To show that the flow pattern originating in node v and terminating in node n, as shown in Fig. V. 2, is maximal it suffices to demonstrate that there is no flow augmenting path (see [10], p. 12] from node v to node n. The existence of such a flow augmenting path would manifest itself in an allowable increase of flow rate on exactly one of the links [v,i], for some $i \in N_1$. But this would be equivalent to an increase in exactly one delivery rate $r_i(1)$, for some $i \in N_1$ (without changing any of the others) which is just what we have shown to be impossible. This brings us to the conclusion:

Theorem V.1

Let $F_1^0(t)$, $0 \leq t \leq t_1^0$ be an optimal flow schedule solution to MTP(1). Then

$$\sum_{i \in N_1} r_i(1) = CS(N_1) \qquad (V.3)$$

□

As all the queues are reduced to zero by the time t_1^0 we can write

$$t_1^0 = \frac{\sum_{i \in N_1} q_i(0)}{CS(N_1)} \qquad (V.4)$$

But it is also true that (see Lemma IV.2)

$$t_1^0 = \sum_{i \in N_1} \sigma_i(1) q_i(0) \qquad (V.5)$$

Comparison of (V.4) and (V.5) raises a question about the functional relation between the set of optimal dual variables and the value of the minimal cut-set $CS(N_1)$. The next lemma answers this question.

Lemma V.1

At a stable point, let $\sum(1)$ be an optimal dual solution to MTP(1). Then

$$\sigma_i(1) = 1/CS(N_1), \quad \forall i \in N_1 \qquad (V.6)$$

Proof:

Suppose that not all $\sigma_i(1)$, $i \in N_1$ are equal, and let a and b be a pair of sources in N_1 such that $\sigma_a(1) > \sigma_b(1)$. Define a new delivery problem for which

$$\hat{q}_i(0) = \begin{cases} q_a(0)-\Delta a, & \text{if } i = a \\ q_b(0)+\Delta b, & \text{if } i = b \\ q_i(0), & \text{otherwise} \end{cases} \tag{V.7}$$

and

$$-\sigma_a(1)\Delta a + \sigma_b(1)\Delta b = 0 \tag{V.8}$$

where

$$\Delta b > \Delta a > 0$$

At a stable point, we can always find a perturbation $(\Delta a, \Delta b)$ for which (V.8) is satisfied and the perturbation is acceptable.

It is not difficult to see that condition (V.8) implies the fact (see Corollary IV.1) that

$$\hat{t}_1^0 = t_1^0 \tag{V.9}$$

where $\hat{t}_1^0$ is the first minimal time for the new problem. The total delivery rate of sources in N_1 is, for the new flow schedule,

$$\hat{\rho}(N_1) = \frac{\sum_{i \in N_1} q_i(0)}{\hat{t}_1^0} \qquad \text{(V.10a)}$$

which can be also written as

$$\hat{\rho}(N_1) = \rho(N_1) + \varepsilon \qquad \text{(V.10b)}$$

where

$$\varepsilon \triangleq \frac{\Delta b - \Delta a}{t_1^0}$$

We already know that $\rho(N_1) = \rho_{max}(N_1)$, and from (V.10b) and the assumption that $\sigma_a(1) > \sigma_b(1)$ we have that $\varepsilon > 0$. Thus

$$\hat{\rho}(N_1) > \rho_{max}(N_1) = CS(N_1) \qquad \text{(V.11)}$$

which of course is impossible. This contradicts our initial assumption about an existence of unequal dual variables. As a consequence we may rewrite (V.5) as

$$t_1^0 = \sigma_{max}(1) \sum_{i \in N_1} q_i(0) \qquad \text{(V.12)}$$

and comparison to (V.4) completes the proof.

□

It is interesting to observe in consequence that in SDN, the first minimal time t_1^0 is always equally sensitive to changes in any of the queues in the set N_1.

One of the significant properties of the set L_1 is that it must be saturated by flows originating in the set N_1, throughout the interval $[0,t_1^0]$. The total rate of the saturating flows was found to be maximal and thus equal to $CS(N_1)$. It is quite obvious that this flow rate cannot drop from its maximal value, even momentarily, since this would cause some of the data coming from sources in N_1 to be delivered later than t_1^0. This observation holds for all flow schedules, as long as they terminate by time t_1^0. We conclude that no segment of an optimal delivery function may have a slope less than $CS(N_1)$.

Lemma V.2

Let $D_M^0(t)$, $0 \leq t \leq t_1^0$ be an optimal delivery function in SDN. Then

$$\rho_m^0 \geq CS(N_1), \quad m = 1,2,\ldots,M. \qquad \text{(V.13)}$$

□

We will show now that (V.13) is satisfied with the equality for the first segment of an optimal delivery function. Suppose that in the optimal solution to MTP(1) we find that $N_1 = N_0$. The total delivery rate of $F_1^0(t)$, $0 \leq t \leq t_1^0$ is maximal and equal to $CS(N_0)$. It can be easily seen with the help of the sketch in Fig. V.3 that the existence of a delivery function (shown by the broken line) that dominates $D_1^0(t)$ would imply a delivery rate $\rho'(N_0) > CS(N_0)$ which is of course impossible. We must conclude that $D_1^0(t)$, $0 \leq t \leq t_1^0$ is the optimal delivery function. This proves our claim for the case of $N_1 = N_0$.

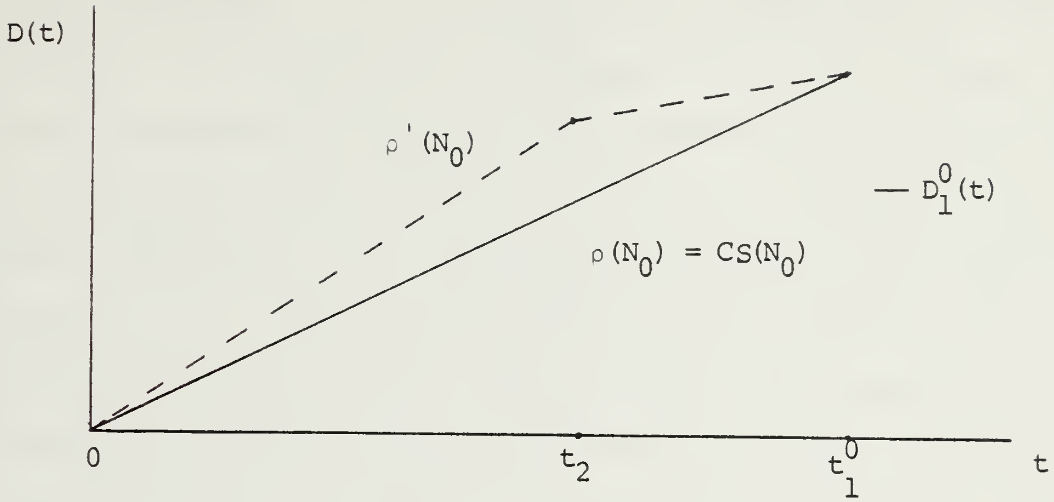


Fig. V.3. Delivery Function with Maximal Flow Rate

In general we may expect in a solution to MTP(1) that $N_1 \subset N_0$. In this case it is possible to increase all the delivery rates of sources in $N_0 - N_1$ since they use (by definition) only unsaturated links. Hence, all the data backlogged in these nodes will be completely delivered prior to time t_1^0, say at some time $t_1^0 - \varepsilon$. In the remaining interval only data from the set N_1 will continue to flow in the network, with the rate of $CS(N_1)$. This simple construction proves the existence of a two part ($\varepsilon > 0$) flow schedule with a delivery rate of $CS(N_1)$ in the interval $(t_1^0 - \varepsilon, t_1^0]$.

<u>Corollary V.2</u>

Let $D_M^0(t)$ be the optimal delivery function in SDN. Then

$$\rho_1^0 = CS(N_1) \tag{V.14}$$

□

The proof leading to Corollary V.2 deserves some additional discussion. Let us consider an optimization problem in which the objective is to minimize the total delivery time of data backlogged in N_0-N_1, while keeping the delivery time of data queued in N_1 at time t_1^0. This is a slightly more formal statement of the construction method we used to prove Corollary V.2. Since the solution to this problem will automatically satisfy (V.14) we may use this new formulation as a substitute for our previous formulation, which required both MRP(1) and MTP(2) in order to obtain the value of t_2^0. The substitute optimization problem can be written as follows.

$$\min t_2$$

s.t.

$$\sum_{p=1}^{2}\Big(\sum_{j(\neq i)} u_{ij}(p) - \sum_{j(\neq i)} u_{ji}(p)\Big) = q_i(0), \ \forall i \in N_1, \ p = 1,2$$

$$\sum_{j(\neq i)} u_{ij}(p) - \sum_{j(\neq i)} u_{ji}(p) \geq 0, \ \forall i \in N_1, \ p = 1,2$$

$$\sum_{j(\neq i)} u_{ij}(1) - \sum_{j(\neq i)} u_{ji}(1) = q_i(0), \ \forall i \in N_0-N_1 \qquad \text{(V.15)}$$

$$t_2 c_{ij} + u_{ij}(1) \leq t_1^0 c_{ij}$$

$$-t_2 c_{ij} + u_{ij}(2) \leq 0, \ \forall [i,j] \in L_0$$

$$u_{ij}(p),\ t_2 \geq 0, \ \forall [i,j] \in L_0, \ p = 1,2,$$

□

where we have used the notation

(i) $u_{ij}(1) \triangleq (t_1^0 - t_2) f_{ij}$

$\forall [i,j] \quad L_0$

(ii) $u_{ij}(2) \triangleq t_2 f_{ij}$

The formulation (V.15) depends on the knowledge of the set N_1 (and the time t_1^0), and hence on stability of MTP(1). As we indicated before, we will later show that the stability requirement is not necessary.

We now show not only that MRP(1) is not needed in SDN, but that the optimization problem in (V.15) can be formulated in a much more efficient way (with regard to the number of variables and the number of constraints). The basic idea is that the flow pattern of data delivered from the set N_1 in the period $(t_2^0 - t_1^0]$ can be made identical to its pattern in the interval $[0, t_2^0]$. In other words, let $p_i^n \in \underline{P}(i,n)$ be an active link chain used by some source i, $i \in N_1$ with a rate $r_i[p_i^n]$, $\forall t \in [0, t_2^0]$. Then the same chain can be used to forward data from source i with the same rate $r_i[p_i^n]$, $\forall t \in (t_2^0, t_1^0]$. In order to see it we need the following result.

Theorem V.2

There exists an optimal flow schedule $F_M^0(t)$, $0 \le t \le t_1^0$ for which

$$r_i(t) = \frac{q_i(0)}{t_1^0}, \ \forall t \in [0, t_1^0] \text{ and } \forall i \in N_1 \qquad \text{(V.16)}$$

Proof:

Before we start with the proof, we remind the reader that $r_i(t)$ is the net delivery rate of data from node i at time t (see (II.2)).

Let i and j be any two nodes such that $i \in N_1$ and $j \in N_0 - N_1$. Let $\underline{P}(i,n)$ and $\underline{P}(j,n)$ be the sets of all directed chains connecting node i and node j, respectively, to destination node n. Select any active chain $p_i^n \in \underline{P}(i,n)$ and let x be the head node of the first link in that chain that belongs to L_1 (there is at least one such link since $i \in N_1$). Let $p_i^x \subset p_i^n$ denote the partial chain connecting node i to node x. Our claim is that

$$\text{if } p_j^n \text{ is active then } p_i^x \cap p_j^n = \emptyset, \quad \forall p_j^n \in \underline{P}(j,n). \qquad (V.17)$$

This claim can be easily verified by referring to Fig. V.4.

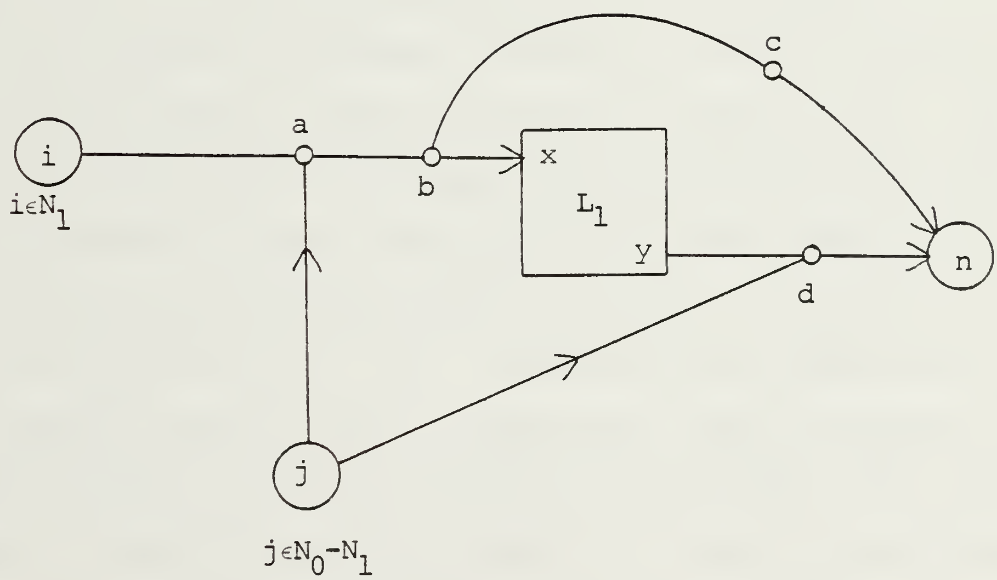


Fig. V.4. Illustration for Theorem V.2

Suppose that there is some active chain $\hat{p}_j^n \in \underline{P}(j,n)$ such that

$$\hat{p}_j^n \cap p_i^n = [a,b]. \tag{V.18}$$

Since $j \notin N_1$, the flow on this chain $r_j[\hat{p}_j^n]$, cannot go through L_1, so that the chain must have the form (see Fig. V.4)

$$p_j^n = (j,\ldots,a,b,\ldots,c,\ldots,n). \tag{V.19}$$

But this would imply the existence of a chain (not necessarily active)

$$\hat{p}_i^n = (i,\ldots,a,b,\ldots,c,\ldots,n), \tag{V.20}$$

which violates the fact that L_1 is a disconnecting set for all nodes in N_1. We must conclude that (V.17) is true.

Now, consider the optimal solution to MTP(1) and let p_i^n be an active chain in that solution, for some $i \in N_1$. Our proof of claim (V.17) indicates that no data flow from any of the sources in N_0-N_1 may <u>ever</u> (in any flow schedule) use the partial chain p_i^x, and this is true for all partial chains of this type for all $i \in N_1$. In view of this observation we may require, without loss of generality, that an optimal flow schedule will have the same chain flow structure $(p_i^x, r_i[p_i^x], \forall i \in N_1)$ as the flow solution to MTP(1). This in turn implies (V.16) since in MTP(1) the delivery rates of all the sources,

and in particular those in N_1, have the form

$$r_i(t) = \frac{q_i(0)}{t_1^0}, \; \forall i \in N_0, \; \forall t \in [0,t_1^0]. \qquad \text{(V.21)}$$

This concludes our proof.

□

Thm. V.2 does not mean that the solution to MTP(2) has no effect on the chain flow structure of sources in N_1, but only that any effect must occur beyond the critical set L_1. This is schematically indicated by point d in Fig. V.4.

Suppose next that we have solved MTP(2), so that the chain flows of sources in N_1 incorporate interactions with the chain flows originating in N_0-N_1. Since in the period $(t_2^0,t_1^0]$ only the sources in N_1 are active we may require, without loss of generality, that their chain structure in that interval, will remain the same as in the interval $[0,t_2^0]$. This result leads us to a new formulation of MTP(2), which is described next.

B. SUBSEQUENT CORNER POINTS

Given N_1, t_1^0 and armed with the results of the last section we can formulate the Second Minimal Time Problem as

MPT(2): $$\min t_2$$

s.t.

$$-t_2 \frac{q_i(0)}{t_1^0} + \sum_{j(\neq i)} u_{ij} - \sum_{j(\neq i)} u_{ji} = 0, \quad \forall i \in N_1$$

$$\sum_{j(\neq i)} u_{ij} - \sum_{j(\neq i)} u_{ji} = q_i(0), \; \forall i \in N_0-N_1 \qquad \text{(V.22)}$$

$$-t_2 c_{ij} + u_{ij} \leq 0, \quad \forall [i,j] \in L_0$$

$$t_2, u_{ij} \geq 0, \quad \forall [i,j] \in L_0$$

□

where we have used the transformation of variables

$$u_{ij} \triangleq t_2 f_{ij}, \quad \forall [i,j] \in L_0.$$

It should be noted that the number of variables as well as constraints is exactly the same here as in MTP(1) (cf. V.1).

At a stable point, the optimal dual solution to MTP(2) may be used to identify the sets L_2 and N_2 (see Def. IV.4). Moreover, using similar arguments to those used in the proof of Thm. V.1, an analogous result can be obtained for sources in N_2. We state without proof:

<u>Theorem V.3</u> (Thm. V.1)

Let $F_2^0(t)$, $0 \leq t \leq t_1^0$ be an optimal flow schedule solution to MTP(2). Then

(i) $$\sum_{i \in N_2} r_i(2) = CS(N_2)$$

(ii) $$\sum_{i \in N_1} r_i(1) = \sum_{i \in N_1} r_i(2) = CS(N_1)$$

□

Proposition (ii) above was already proved in Thm. V.2, but we include it here for completeness.

Similarly, it can be shown (by analogy to Lemma V.1) that at a stable point, the optimal dual variables corresponding

to sources in N_2-N_1 are positive and equal, i.e.

$$\sigma_i(2) = \sigma_{max}(2), \forall i \in N_2-N_1 \tag{V.23}$$

The functional relation of $\sigma_{max}(2)$ to the value of the minimal cutset $CS(N_2)$ can be determined as follows. Using Thm. V.3, we have

$$t_2^0 = \frac{t_2^0 CS(N_1) + \sum_{i \in N_2-N_1} q_i(0)}{CS(N_2)} \tag{V.24a}$$

or equivalently

$$t_2^0 = \frac{1}{CS(N_2) - CS(N_1)} \sum_{i \in N_2-N_1} q_i(0) \tag{V.24b}$$

But we also have (from Corollary IV.1 and the form of RHS of (V.22)) that

$$t_2^0 = \sigma_{max}(2) \sum_{i \in N_2-N_1} q_i(0). \tag{V.25}$$

Comparing (V.25) and (V.24b) we conclude that

Lemma V.3

At a stable point, let $\sum(2)$ be an optimal dual solution to MTP(2). Then

$$\sigma_i(2) = \frac{1}{CS(N_2)-CS(N_1)}, \forall i \in N_2-N_1 \tag{V.26}$$

□

Next, using reasoning completely parallel to that in the proof of Lemma V.2 and its corollary we arrive at a similar result, which we state formally as:

<u>Lemma V.4</u>

Let $D_M^0(t)$ be an optimal delivery function in SDN. Then

$$\rho_2^0 = CS(N_2) \tag{V.27}$$

□

The implication of this result and its constructive proof is that we can proceed directly to solve a new version of MTP(3), without bothering to solve MRP(2). This new version of MTP(3) may be phrased as follows:-- "Minimize the total delivery time of queues in the set N_0-N_2, such that all queues in N_2-N_1 are delivered by the time t_2^0 and all queues in N_1 are delivered by the time t_1^0." In order to show that this MTP(3), like MTP(2), can be formally stated in an efficient way, we need to prove a result parallelto Thm. V.2.

<u>Theorem V.4</u>

There exists an optimal flow schedule $F_M^0(t)$, $0 \le t \le t_1^0$ for which

$$\text{(i)} \quad r_i(t) = \frac{q_i(0)}{t_1^0}, \ \forall t \in [0,t_1^0] \text{ and } \forall i \in N_1$$

$$\text{(ii)} \quad r_i(t) = \begin{cases} \dfrac{q_i(0)}{t_2^0}, \ \forall t \in [0,t_2^0] \\ 0 \quad , \text{ otherwise} \end{cases} \quad \forall i \in N_2-N_1 \tag{V.28}$$

Proof:

Proposition (i) is a restatement of Thm. V.2 and is included for completness.

The proof is in essence identical to that of Thm. V.2.

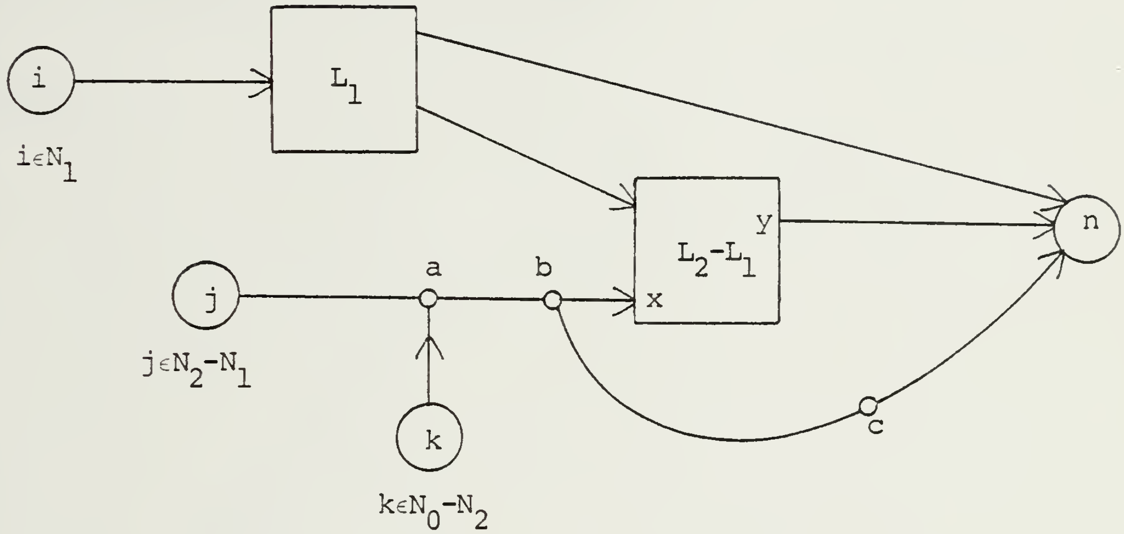


Fig. V.5. Illustration for Thm. V.4

Our basic claim is that for all sources j, $j \in N_2-N_1$, any active subchain p_j^x remains free from any interaction with flows originating in the set N_0-N_2. Suppose that a flow coming from some node k, $k \in N_0-N_2$ interferes with a flow coming from some node j, $j \in N_2-N_1$ in link [a,b]. Since the flow from k cannot use the set L_2 (and in particular L_2-L_1) by definition, it must use the bypassing chain which goes through node c. But this would imply that the node j has a chain outside the set L_2 which is impossible since L_2 is a disconnecting set for nodes in N_2. This basically completes the major arguments of the proof. The remaining details are as in the proof of Thm. V.2. □

By now the reader has no doubt surmised (correctly) that all the results derived up to here can be extended to subsequent corners of the optimal delivery function. We therefore conclude that the general character of the optimal delivery function is as illusted in Fig. V.6, where

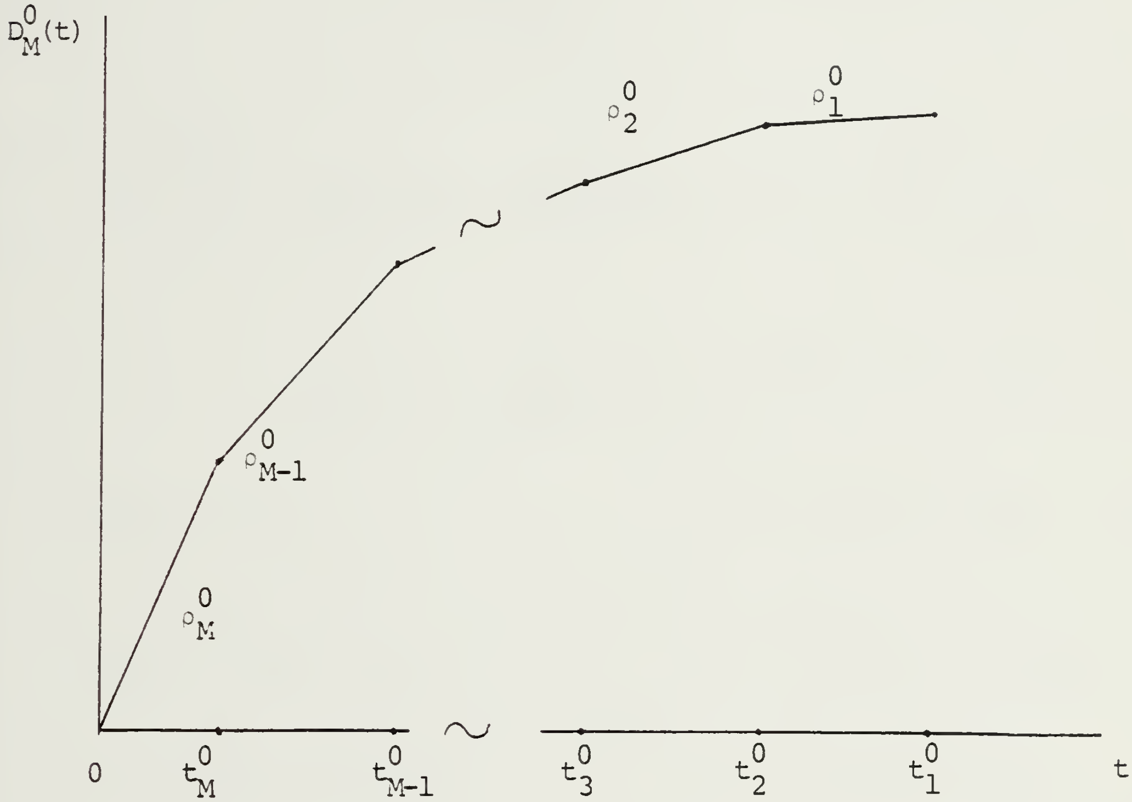


Fig. V.6. Optimal Delivery Function in Single Destination Networks

$$\rho_m^0 = CS(N_m) \; , \qquad m = 1,2,\ldots,M$$

$$t_m^0 = \begin{cases} \dfrac{1}{CS(N_1)} \displaystyle\sum_{i \in N_1} q_i(0), & m = 1, \\ \dfrac{1}{CS(N_m) - CS(N_{m-1})} \displaystyle\sum_{i \in N_m - N_{m-1}} q_i(0), & m = 2,3,\ldots,M. \end{cases} \tag{V.29a}$$

and

$$N_M \equiv N_o.$$

The general form of MTP(m), m = 1,2,...,M is given by

MTP(m): $\min t_m$

s.t.

$$-t_m \frac{q_i(0)}{t_k^0} + \sum_{j(\neq i)} u_{ij} - \sum_{j(\neq i)} u_{ji} = 0, \ \forall i \in \begin{cases} N_1, & k = 1 \\ N_k - N_{k-1}, & k = 2,3,\ldots,M \end{cases}$$

$$\sum_{j(\neq i)} u_{ij} - \sum_{j(\neq i)} u_{ji} = q_i(0), \ \forall i \in N_0 - N_{m-1}$$

$$-t_m c_{ij} + u_{ij} = 0, \ \forall [i,j] \in L_{m-1}{}^{\dagger} \qquad \text{(V.29b)}$$

$$-t_m c_{ij} + u_{ij} + s_{ij} = 0, \ \forall [i,j] \in L_0 - L_{m-1}$$

$$t_m, u_{ij}, s_{ij} \geq 0, \ \forall [i,j] \in L_0, \qquad \square$$

where $t_1^0, t_2^0, \ldots, t_{m-1}^0$ are given, and

$$N_M \equiv N_o.$$

---

†There is no need for slack variables since the links $[i,j] \in L_{m-1}$ are saturated for all $t \in [0, t_m^0]$.

C. GLOBAL OPTIMALITY

We stated previously that one of the distinctions between a multicommodity delivery problem and a delivery problem in SDN is exhibited in the fact that every optimal delivery function in SDN is also globally optimal. This is not true for multicommodity case, as the counter example in Appendix B indicates.

Theorem V.5

Let $D_M^0(t)$ be an optimal delivery function in SDN. Then it is also globally optimal.

Proof:

Assume to the contrary that there exists some other delivery function $D_K(t)$, $0 \le t \le t_1^0$ for which

$$\hat{D}_K(t') > D_M^0(t'), \text{ for some } t' \in [0,t_1^0]. \qquad (V.30)$$

With the help of the optimal delivery function let us find m, $m \in \{1,2,\ldots,M\}$ such that $t' \in (t_{m+1}^0, t_m^0]$. Let us mark now all the data stored in nodes of the set N_0-N_m so it will be distinguishable from the data stored in the nodes of the set N_m. Before applying the flow schedule $\hat{F}_K(t)$, $0 \le t \le t_1^0$ which generates the delivery function $\hat{D}_K(t)$, let us place an observer at node n. His duty is to count how much marked Q_m and unmarked Q_u data is delivered to node n, up to time t'.

$$\hat{D}_K(t') = Q_m + Q_u \qquad (V.31)$$

It is clear that

$$Q_m \leq \sum_{i \in N_o - N_m} q_i(0). \qquad \text{(V.32)}$$

The observer cannot count more marked data than there was initially in the network. Also,

$$Q_u \leq CS(N_m) \cdot t' \qquad \text{(V.33)}$$

since no more than $CS(N_m)$ of unmarked data per unit time can reach the node n at any given time. Thus

$$\hat{D}_K(t') \leq \sum_{i \in N_0 - N_m} q_i(0) + CS(N_m) \cdot t'. \qquad \text{(V.34)}$$

But the right hand side of (V.34) is exactly the value of $D_M^0(t)$ at time t', which contradicts (V.30) and completes the proof.

□

The proof of Thm. V.5 is not dependent anywhere on the fact that the flow schedule $F_K(t)$ is feasible in the narrow sense (see Appendix A), i.e. does not allow for intermediate storage of data in the network. Combining this observation with the result of Thm. II.2 we know that the optimal delivery function and its generating flow schedule also solve the minimal total delay problem over the class of flow schedules that allow intermediate data storage. As such, we obtain a much simpler solution algorithm to that problem than the one described in [11].†

---

†The work of Shats and Segall seems to be the only known open loop solution to the minimal total delay problem in SDN.

D. SOLUTION ALGORITHM FOR SINGLE DESTINATION NETWORKS

Up to now we have established that the optimal delivery function and its generating flow schedule can be obtained by solving a sequence of specialized MTP's. In addition, the size of each of the problems is limited to $2\ell+1$ variables and $n+\ell-1$ constraints independently of which corner point we are solving for. The only condition that we assumed, for the above to be true, is that a solution to MTP(m), $m = 1,2,\ldots,M$ identifies uniquely the critical set N_m, or equivalently, is stable. When we say the set N_m, we really have in mind the nodes in $N_m - N_{m-1}$, since the nodes in N_{m-1} were supposedly identified at previous corners.

Consider an optimal dual solution $\sum(1)$ for MTP(1). Define

$$N_1^1 \;=\; \text{set } \{i\} \text{ of all nodes } i \in N_0 \text{ such that } \sigma_i(1) > 0,$$

and

$$L_1^1 \;=\; \text{set } \{[i,j]\} \text{ of all links } [i,j] \in L_0 \text{ such that } \pi_{ij} < 0.$$

A useful interpretation of the sets N_1^1 and L_1^1 can be obtained with the help of the following lemma.

<u>Lemma V.4</u>

Let $\sum(1)$ be an optimal dual solution to MTP(1). Then

$$\sigma_i(1) > 0 \;\rightarrow\; i \in N_1 \qquad \text{(V.35)}$$

Proof:

Let i, $i \in N_0$ be some node for which $\sigma_i(1) > 0$. From (IV.10) we have (for SDN, $k = n$)

$$\sigma_i(1) = - \sum_{[\alpha,\beta] \in p_i^n} \pi_{\alpha\beta}(1), \tag{V.36}$$

where p_i^n is any active chain connecting source i to the destination n. Since $\pi_{\alpha\beta} \leq 0$, $\forall [\alpha,\beta] \in L_0$ we conclude that there exists at least one link in the chain p_i^n, for which $\pi_{\alpha\beta} < 0$. From the slackness theorem (see Lemma IV.1(ii)) we know that this particular link will be saturated (zero slack variable) in all optimal primal solutions, and hence belongs to the critical set L_1. Any source node using this link must belong to the set N_1. We conclude that $i \in N_1$.

□

It is clear now that the set N_1^1 consists of all the members of N_1 that were uniquely identified by the optimal solution to MTP(1).

$$N_1^1 \subseteq N_1 \tag{V.37}$$

and similarly

$$L_1^1 \subseteq L_1$$

where the equality holds if the solution to MTP(1) is stable. It should be observed also that the set N_1^1 cannot be empty. This can be easily deduced from application of (IV.5) to MTP(1):

$$t_1^0 = \sum_{i \in N_0} \sigma_i(1) q_i(0) \tag{V.38}$$

Since $t_1^0 > 0$ and $q_i(0) \geq 0$, $\forall i \in N_0$ there must be at least one node i, $i \in N_0$ for which $\sigma_i(1) > 0$. This also implies that there is at least one link $(\alpha,\beta] \in L_0$ such that $\pi_{\alpha\beta}(1) < 0$.

Up to this point we have no way of identifying the remaining members of N_1. But let us try to proceed with the solution algorithm in spite of this fact. We will use the set N_1^1 instead of N_1; as a result we obtain a slightly different formulation for MTP(2).

MTP(2): $\min t_2$

s.t.

$$-t_2 \frac{q_i(0)}{t_1^0} + \sum_{j(\neq i)} u_{ij} - \sum_{j(\neq i)} u_{ji} = 0, \ \forall i \in N_1^1$$

$$\sum_{j(\neq i)} u_{ij} - \sum_{j(\neq i)} u_{ji} = q_i(0), \ \forall i \notin N_1^1 \tag{V.39}$$

$$-t_2 c_{ij} + u_{ij} = 0, \ \forall [i,j] \in L_1^{1\dagger}$$

$$-t_2 c_{ij} + u_{ij} + s_{ij} = 0, \ \forall [i,j] \notin L_1^1$$

$$s_{ij}, t_2, u_{ij} \geq 0, \ \forall [i,j] \in L_0$$

□

---

† In this equation we don't need slack variables since we know that all the links in L_1 have always to be saturated.

Suppose now that actually $N_1^1 \subset N_1$. Then it is impossible to deliver all the queues in the set $N_1 - N_1^1$ prior to time t_1^0 while keeping at the same time the delivery rates of all the nodes in N^1 at the value $q_i(0)/t_1^0$. We therefore conclude that

<u>Lemma V.5</u>

Let t_2^0 be the optimal value of the cost function for problem (V.39). Then

$$N_1^1 \subset N_1 \rightarrow t_2^0 = t_1^0 \tag{V.40}$$

□

This does not necessarily seem like progress until we consider the optimal dual solution to (V.39). Using again (V.5) for our problem here, we obtain

$$t_2^0 = t_1^0 = \sum_{i \notin N_1^1} \sigma_i(2) q_i(0) \tag{V.41}$$

which implies (similarly to (V.38)) that there must be at least one node i, $i \notin N_1^1$ such that $\sigma_i(2) > 0$. We want to show now that this node belongs to N_1. Let us pick some active chain for that node. If the chain passes through L_1^1, we are done, since this implies that $i \in N_1$. If the chain does not go through L_1^1, then we can repeat the argument used to prove Lemma V.4.

Defining

$$N_1^2 = N_1^1 \cup \{i | \sigma_i(2) > 0 \text{ and } i \notin N_1^1\} \tag{V.42}$$

and

$$L_1^2 = L_1^1 \cup \{[i,j] \mid \pi_{ij} < 0 \text{ and } [i,j] \not\in L_1^1\} \qquad (V.43)$$

we can solve (V.39) again, this time using N_1^2 and L_1^2 instead of N_1^1 and L_1^1, respectively. In view of the precedding discussion we are assured that after a number of iterations k_1, $k_1 \leq |N_1|$ the whole set N_1 (and L_1) will be identified and we may proceed to solve the original MTP(3).

The general idea behind the solution procedure and how it applies to subsequent corner points should now be clear. The only consequence of instability in any of the corner points is to increase the number of iterations needed to reach the next corner. If we let K denote the total number of iterations, where

$$K = \sum_{i=1}^{M} k_i \qquad (IV.44)$$

then we obviously have

$$M \leq K \leq |N_0| \qquad (V.45)$$

where M is the number of corners in the optimal delivery function in SDN.

E. REMARK ON MULTICOMMODITY FLOW SCHEDULES

In this subsection we briefly discuss a class of multicommodity problems for which the optimal dual variables satisfy, at every corner, the following relation:

$$\sigma_i^k(m) = \sigma_{max}(m), \forall (i,k) \in N_m, m = 1,2,\ldots,M. \qquad (V.46)$$

In particular, it can be shown that the data flow rate with respect to the set N_m is maximal in the period $[0,t_m^0]$, i.e.

$$\sum_{(i,k)\in N_m} r_i^k(t) = \rho_{max}(N_m), \forall t \in [0,t_m^0] \qquad (V.47)$$

The delivery of all data queues corresponding to the set N_0-N_m can be accomplished prior to t_m^0, say by $t_0-\varepsilon$, since by definition the flows of commodities in N_0-N_m only use unsaturated links.[†] Thus the generalized perturbation equation must be satisfied by flows of commodities in the set N_m (see Thm. IV.9)

$$\frac{\sigma_{max}(m)}{1-\sigma_t(m)} \sum_{(i,k)\in N_m} r_i^k(t) = 1, \forall t \in (t_m^0-\varepsilon,t_m^0] \qquad (V.48)$$

Using (V.47) in (V.48) we have

$$\rho_{max}(N_m) = \frac{1-\sigma_t(m)}{\sigma_{max}(m)} \;. \qquad (V.49)$$

As a consequence, the optimal delivery function for this class of multicommodity problems is characterized by:

$$\text{(i)} \quad \rho_m^0 = \rho_{max}(N_m) = \frac{1-\sigma_t(m)}{\sigma_{max}(m)} \qquad (V.50)$$

[†]A more detailed argument can be found in Chapter IV.A.2.

$$\text{(ii)} \quad t_m^0 = \frac{\sum_{(i,k)\in N_m - N_{m-1}} q_i^k(0)}{\rho_{max}(N_m) - \rho_{max}(N_{m-1})}, \quad m = 2,3,\ldots,M$$

and

$$t_1^0 = \frac{\sum_{(i,k)\in N_1} q_i^k(0)}{\rho_{max}(N_1)}$$

Moreover, it is easy to see that the proof of global optimality for SDN applies without change to the multicommodity case considered here if we use $\rho_{max}(N_m)$ instead of $CS(N_m)$, since the two are not equal in general for the multicommodity case.

It turns out that computer solution example which we considered in Chapter III.3 falls into this category of "SDN like" multicommodity problems, i.e. multicommodity problems for which the "optimal" solution is also globally optimal. For convenience we restate that delivery problem.

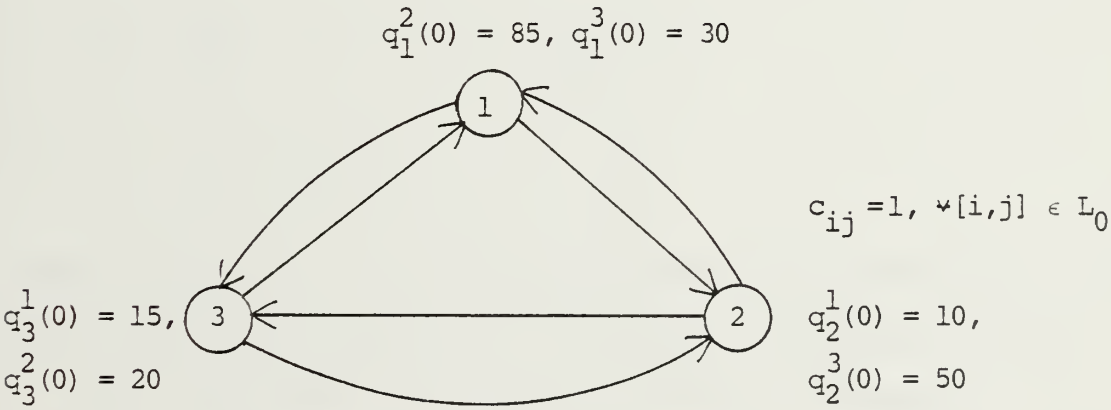


Fig. V.7. Delivery Problem of Chapter III.C.3.

The critical sets for this problem are (from the computer solution):

$$N_1 = \{(1,2),(1,3)\}, \quad N_2 = \{(2,3)\} \cup N_1,$$

$$N_3 = \{(3,2)\} \cup N_2, \quad N_4 = \{(3,1)\} \cup N_3, \qquad \text{(V.51)}$$

and $N_5 = \{(2,1)\} \cup N_4$.

It is not difficult to see from Fig. V.7 that the maximal flow rates with respect to the critical sets are:

$$\rho_{max}(N_1) = 2, \quad \rho_{max}(N_2) = 3, \quad \rho_{max}(N_3) = 4,$$

$$\text{(V.52a)}$$

$$\rho_{max}(N_4) = 5, \quad \text{and} \quad \rho_{max}(N_5) = 6.$$

Using (V.50) we obtain the optimal corner times

$$t_1^0 = \frac{115}{2} = 57.5, \quad t_2^0 = \frac{50}{3-2} = 50$$

$$\text{(V.52b)}$$

$$t_3^0 = \frac{20}{4-3} = 20, \quad t_4^0 = \frac{15}{5-4} = 15$$

and $\quad t_5^0 = \frac{10}{6-5} = 10$

Comparison with Fig. III.7 will immediately reveal that (V.52) is a correct description of the optimal delivery function.

This completes our promise at the end of Chapter III, to show that what looks like a random flow schedule has indeed a lot of structure, and thus simplicity, to it.

F. SAMPLE PROBLEM

We conclude this chapter on single destination networks with a short study of a sample delivery problem. This gives us an opportunity to illustrate some of the special properties that are characteristic of SDN.

In [11][†] Shats and Segall presented an interesting algorithm for solving minimal total delay problems in SDN. We have shown (see Thm. V.5) that the optimal delivery function is also globally optimal in SDN and thus (cf. Thm. II.2) serves as an optimal solution to the minimal total delay problem. Unlike in [11], the sizes of linear programs that are required in our solution procedure are independent of the number of corners of the optimal delivery function. This makes our algorithm adequate to solve large network problems, a task which could not be handled by the authors there. We believe that there is also another advantage to our methodology, namely the additional insight it provides.

We have adopted one of the computer solution examples from [11, p. 73] as our sample problem. We intend to show that by using concepts introduced here it can be solved, with very little effort, actually by inspection only.

As the first step in solving the delivery problem in Fig. V.8, we wish to find the composition of the first critical set N_1. Obviously, N_1 can be only one of the following:

[†]The study in this reference aroused our interest in SDN, and inspired many of the results obtained in this chapter.

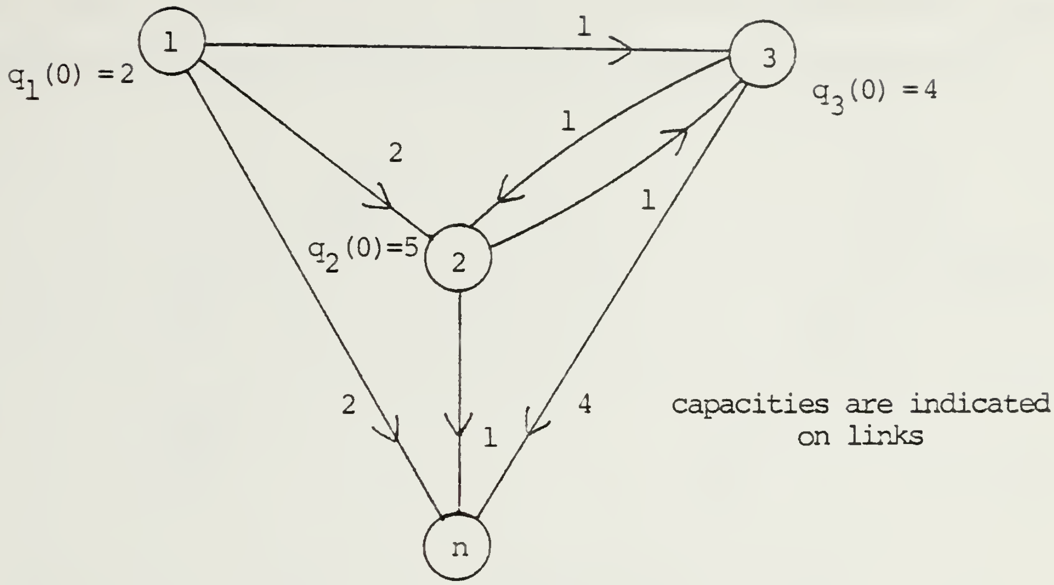


Fig. V.8. Single Destination Delivery Problem

(i) {1}, (ii) {2}, (iii) {3}, (iv) {1,2},

(v) {2,3}, (vi) {1,3}, (vii) {1,2,3}.

Let us now consider all the possibilities for which node $1 \in N_1$ (i.e.: (i), (iv), (vi) and (vii)). As a result of our study we know that the chain flows originating in N_1 must saturate $CS(N_1)$ in the optimal solution for all $t \in [0,t_1^0]$. An immediate consequence of this statement is that node 2 has no available chains, during that period, to send its data to the destination n. This leaves us with possibilities (ii), (iii) and (v). By using exactly the same reasoning we can discard possiblity (iii).

Now, suppose that $N_1 = \{2,3\}$. Since the flow originating in N_1 must saturate $CS(N_1)$, one of the following chain flow decompositions must be part of the optimal flow schedule:

Fig. V.8a. Chain Flow Decomposition for $N_1 = \{2,3\}$

or

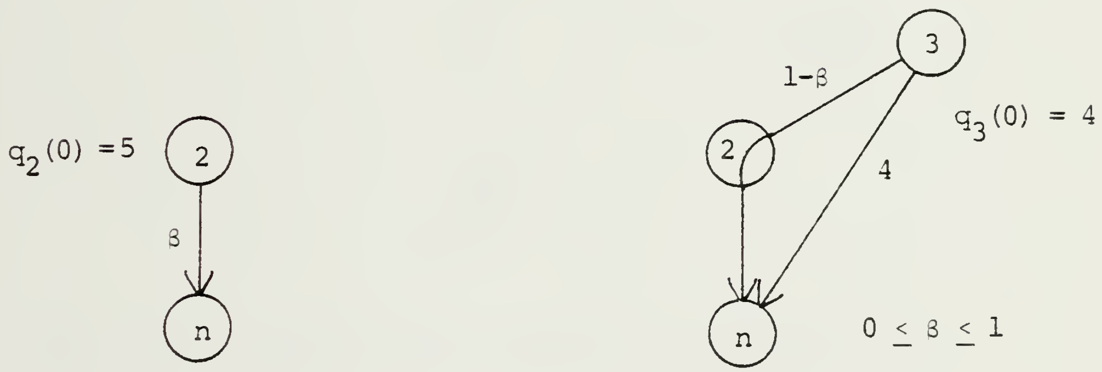


Fig. V.8b. Alternate Chain Flow Decomposition for $N_1 = \{2,3\}$

Then, it also must be true that

$$\text{(a)} \quad t_1^0 = \frac{5}{1+\alpha} = \frac{4}{4-\alpha}$$

or

(b) $t_1^0 = \frac{5}{\beta} = \frac{4}{3+\beta}$

For the first case we have a solution for α, $\alpha = 1\,\frac{7}{9}$ and for the second case, $\beta = -15$. Both of course are infeasible and we conclude that $N_1 = \{2\}$.

We are ready to construct the optimal flow schedule. The flows from node 2 must saturate CS(2), which is equal to 2. The chain flow decomposition which achieves this is unique and is shown in Fig. V.9.

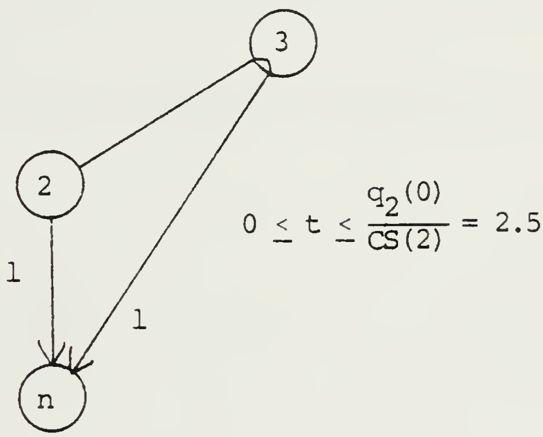


Fig. V.9. Optimal Flow Schedule for Source (2).

We are left essentially with a delivery problem shown in Fig. V.10.

The final solution should be obvious to the eye at this point. For completeness though, we continue with our formal exposition.

Suppose now that

$$\{1,3\} \in N_2 - N_1$$

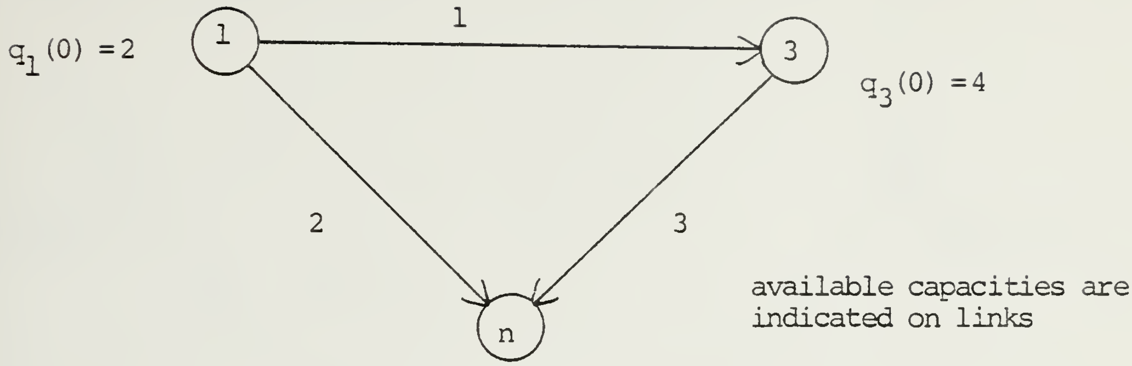


Fig. V.10. Delivery Problem for Sources in $N_0 - N_1$

this would imply the chain flow decomposition which is shown in Fig. V.11.

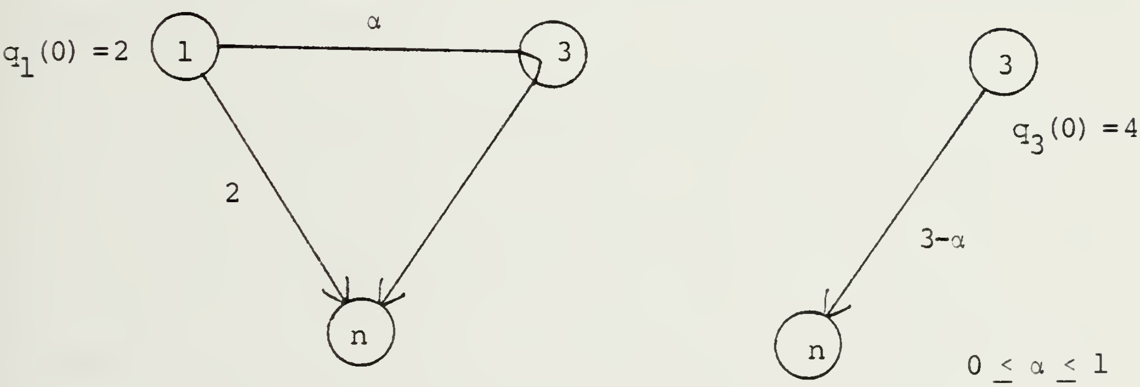


Fig. V.11. Chain Flow Decomposition for $N_2 - N_1 = \{1,3\}$

Also,

$$t_2^0 = \frac{2}{2+\alpha} = \frac{4}{3-\alpha}$$

This leads to $\alpha = -1/5$, which is unacceptable. The only remaining possibility is shown in Fig. V.12.

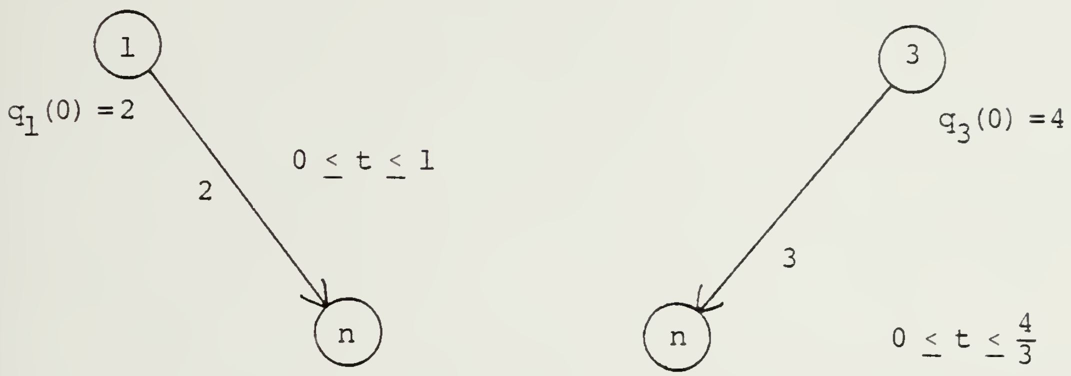


Fig. V.12. Optimal Flow Schedule for Sources (1) and (3)

The solution presented in Fig. V.12 is equivalent to the following statements:

(i) $N_2 = \{3\} \cup N_1$

(ii) $N_3 = \{1\} \cup N_2$

where

$$N_1 = \{1\}$$

The resulting optimal delivery function is shown in Fig. V. 13, where

$$\rho_1^0 = CS(2) = 2, \quad \rho_2^0 = CS(2,3) = 5$$

$$\rho_3^0 = CS(2,3,1) = 7$$

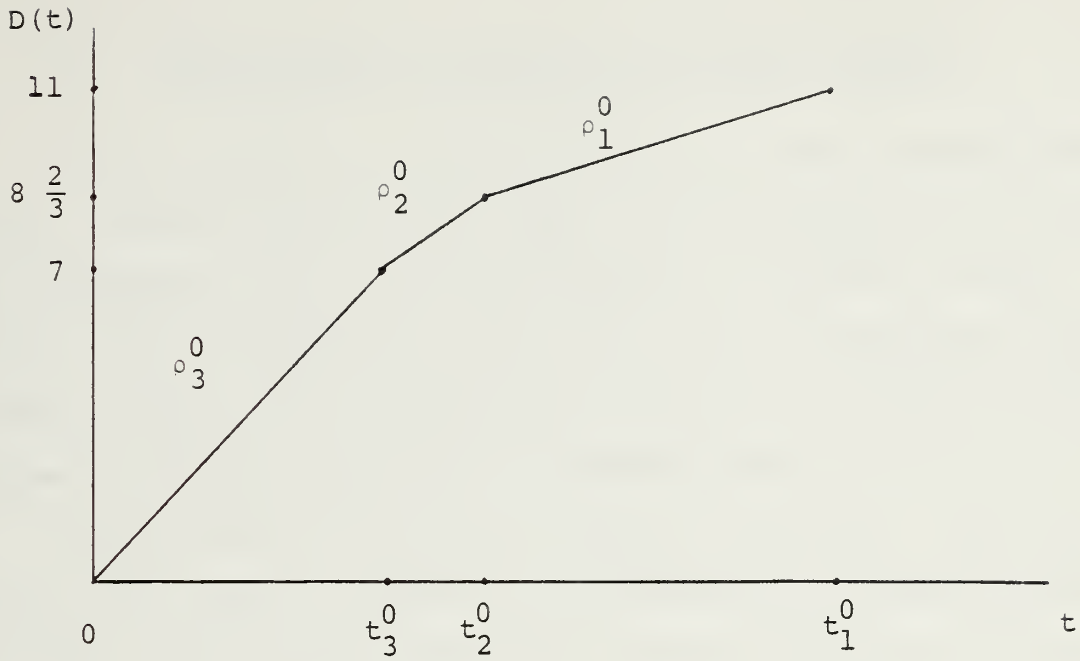


Fig. V.13. Optimal Delivery Function

and

$$t_1^0 = \frac{q_2(0)}{CS(2)} = 2\,\frac{1}{2}$$

$$t_2^0 = \frac{q_3(0)}{CS(2,3)-CS(2)} = 1\,\frac{1}{3}$$

$$t_3^0 = \frac{q_1(0)}{CS(2,3,1)-CS(2,3)} = 1$$

□

VI. APPLICATION TO STOCHASTIC DELIVERY PROBLEMS

We have studied in depth so far the optimal delivery problem. The mode that we have used in our discussion is based on the assumption that during the period of interest the data input rate is identically zero. We find it convenient, especially in view of the forthcoming discussion, to refer to that class of delivery problems as "deterministic."

We now focus on stochastic delivery problems. Here the data input rates are assumed to be governed by some stochastic process. In this new framework, the time necessary to empty a data queue (deliver its contents to their destinations) is no longer a deterministic value but a random variable. In [12] Yee suggested use of the expected delivery time as a performance measure for dynamic routing. We demonstrate that the Sequential Linear Optimization (SLO) methodology, which we used to solve the deterministic case, can be applied to the stochastic case with Yee's performance measure. Before we do so though, we briefly summarize the most common stochastic routing model [13] and indicate why the new performance measure seems to be advantageous.

A. BACKGROUND†

Our point of departure is the original (cf. Chapter II) ℓ-link, n-node model for a communication network. Data entering

---

†In this section we closely follow the discussion in [13].

the network from external sources forms a Poisson process with a rate of a_i^k (messages per second) for those messages entering the network at node i and destined for node k. All messages are assumed to have lengths that are drawn independently from an exponential distribution with mean $1/\mu$ (bits). The combined effect of finite link capacities and random fluctuations in the actual arrival rate of messages to the network causes queueing delays. In order to accommodate these queues we assume that all nodes in the network have unlimited storage capacity. At any given time, the state of the network (or its congestion) is described by the set of data queues.

With each link $[i,j] \in L_0$ we associate, in addition to its capacity, c_{ij}, a queue q_{ij} of all messages waiting to be transmitted over that link. The routing of messages (flow pattern) in the network is accomplished by determining for each node what fraction of the incoming traffic, for each commodity (i.e. destination), is to be directed to which link queue.

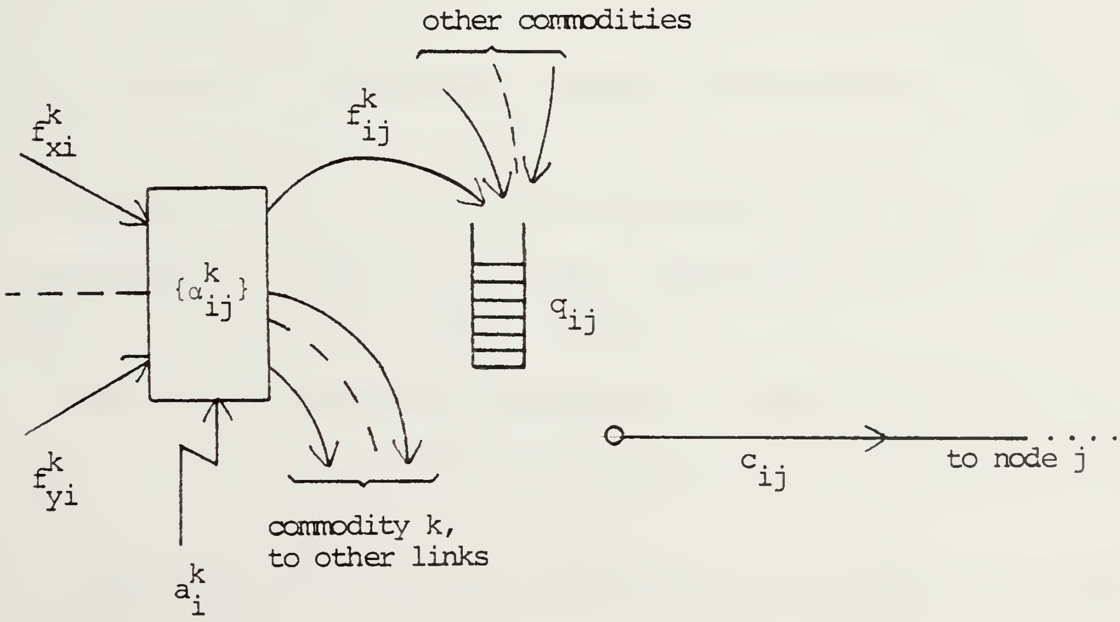


Fig. VI.1. Schematic Representation of Node-Link Queueing Model

The data rate conservation equations for any node i, $i \in V$ can be written (with the help of Fig. VI.1) as

$$\sum_{j(\neq i)} f_{ij}^{k} - \sum_{j(\neq i)} f_{ji}^{k} = a_i^k, \quad \forall (i,k) \in N_0 \tag{VI.1}$$

and the routing variables are defined by

$$\alpha_{ij}^{k} \triangleq \frac{f_{ij}^{k}}{\sum_{j(\neq i)} f_{ji}^{k} + a_i^k} \quad \forall [i,j] \in L_0, \quad \forall k \tag{VI.2}$$

We are now faced with analysis of a network of queues. A similar problem was studied by Jackson [14], and he was able to establish that an imbedded queueing and serving facility offered a solution identical to the same facility acting independently from the network, but with Poisson arrivals at a rate offered by the network. In order to apply this remarkable result here it is necessary to assume that every message, once it arrives at its intended queue q_{ij}, has its length randomly selected anew from an exponential distribution with mean $1/\mu$. This destroys the dependence between interarrival and service (transmission) times. This assumption was studied extensively by Kleinrock in [15], with the conclusion that the so-called "independence assumption", albeit rigorously unjustified, leads in practice to useful results.

With the independence assumption, we see that any link $[i,j] \in L_0$ is now representable as an M|M|1[†] queue with Poisson

---

[†]Detailed study of M|M|1 queues can be found in any basic text on queueing theory.

arrivals of rate $f_{ij} = \sum_k f_{ij}^k$ and exponential service rate with mean $1/\mu c_{ij}$. For notational simplicity we will assume that all capacities in the network include the factor μ, and thus the mean service (transmission) time of a message over link $[i,j]$ is $1/c_{ij}$. In what follows we assume steady-state operation of the M|M|1 system, and thus require that

$$f_{ij} < c_{ij} \qquad \text{(VI.3)}$$

For the queueing model described above the average delay T of a message passing through the network is given by [13]

$$T = \frac{1}{a} \sum_{[i,j]\in L_0} \frac{f_{ij}}{c_{ij} - f_{ij}} \qquad \text{(VI.4)}$$

where

$$a \triangleq \sum_{(i,k)\in N_0} a_i^k$$

and

$$f_{ij} = \sum_{k(\neq i)} f_{ij}^k$$

The most common statement of the routing problem involves minimization of the average delay T as the objective function, subject to constraint (VI.1). Various solution methods for this non-linear problem have been presented (e.g. [16] through [20]) in past years. In [8] a different approach leading to a linear programming formulation was taken, namely a saturation ratio f_{ij}/c_{ij} is defined for each link $[i,j] \in L_0$, and

the worst of them is minimized. This procedure is then iterated until all saturation ratios have been minimized.

When implementing any of the above methods in actual routing control it is necessary to update the estimates of input data rates more or less often (depending on the nature of external data sources), and to recalculate the routing variables in order to adapt to possible changes. It should be noted that in (VI.4) the flow rates $\{f_{ij}^k\}$, and respectively the routing variables $\{\alpha_{ij}^k\}$ depend only on the input rates $\{a_i^k\}$ and not on the actual congestion $\{q_{ij}\}$ in the network at the update time. It is reasonable that inclusion of global congestion information in determination of the routing variables should improve the adaptivity properties of any routing methodology. In particular we follow [12] in suggesting the expected time needed to empty a queueing system as a practical performance measure which uses congestion information.

B. ON THE EXPECTED TIME TO EMPTY A QUEUEING SYSTEM

In this section we will derive (following Yee) the formula for t_1, the expected time needed to empty, for the first time an $M|M|1$ system.

The first question in this respect that we wish to answer is: If a message arrives to an empty system, how long will it take, on average, before the system becomes empty again? Some thought will show that this is exactly the expected length of a "busy period" t_b, which is known to be for $M|M|1$

$$t_b = \frac{1}{c-a} \qquad \text{(VI.5)}$$

where a and 1/c are the arrival rate and the expected service time, respectively.

Now, suppose we look at a queueing system and find (q-1) messages awaiting transmission, and one message being transmitted. Let us agree to put any message that arrives from now on in buffer (b) (see Fig. VI.2). Also, we will always empty the (b) buffer prior to servicing the (a) buffer.

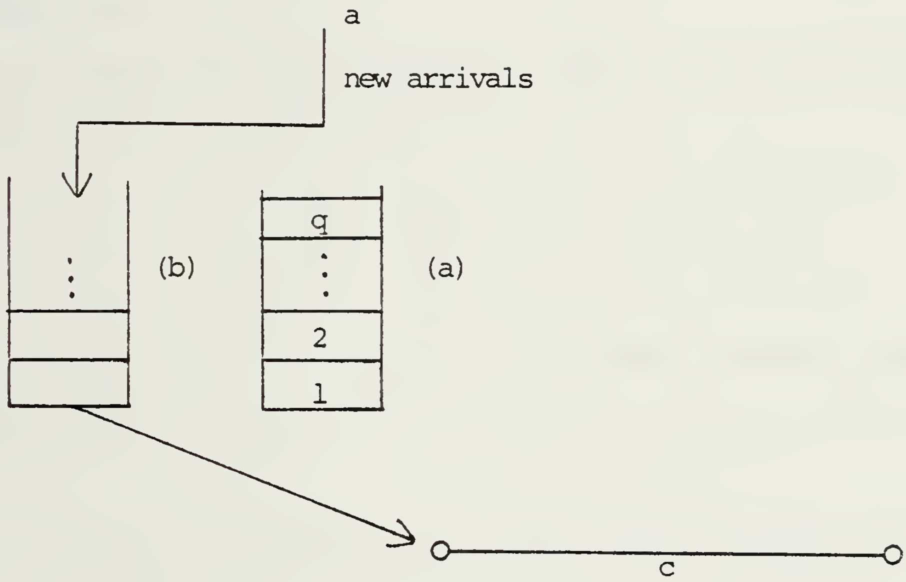


Fig. VI.2. Queueing System

It is clear that the next message in buffer (a) must wait, before beginning transmission, the length of a busy period (VI.5). (We note that the change in queueing discipline that we introduced does not affect the distribution of busy/idle periods for $M|M|1$.) Following the same argument it is clear

that the system becomes empty for the first time after waiting, on average,

$$t_1 = qt_b = \frac{q}{c-a} \qquad \text{(VI.6)}$$

We summarize this part of our discussion in the following lemma.

Lemma VI.1

The expected time t_1 needed to empty, for the first time, an $M|M|1$ system with q messages is

$$t_1 = \frac{q}{c-a}$$

where a and 1/c are the arrival rate and expected service time, respectively.

□

Consider a commodity $(i,k) \in N_0$ which is characterized at a given moment, say t = 0, by its queue $q_i^k(0)$ and its arrival rate a_i^k. Suppose we dedicate to this commodity a part of the capacity resources of the network in such a way that they can support a constant flow rate f_i^k of commodity (i,k) from node i to node k. From our previous discussions we know that it is impossible to assign to commodity (i,k) more capacity than CS(i,k) the value of a minimal cutset separating node i from node k, i.e.

$$f_i^k \leq CS(i,k) \qquad \text{(VI.7)}$$

This situation is depicted schematically in Fig. VI.3.

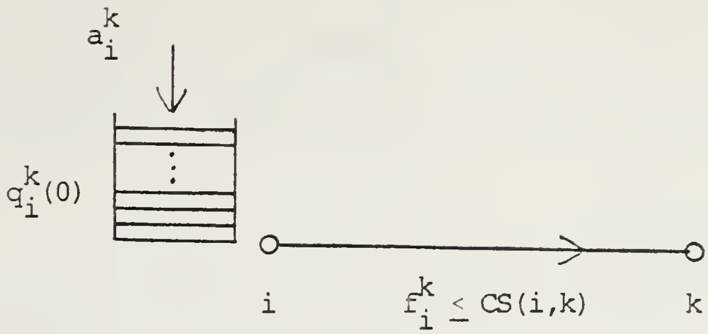


Fig. VI.3. Queueing System for Commodity (i,k)

At this point we can write that the expected time t_i^k needed to empty, for the first time, a queue of commodity (i,k) is

$$t_i^k = \frac{q_i^k(0)}{f_i^k - a_i^k} \tag{VI.8}$$

We would like to consider the same construction simultaneously for all commodities in the network. Since the capacity resources are limited the following constraints must be satisfied:

(i) link capacity constraint

$$\sum_{k(\neq i)} f_{ij}^k \leq c_{ij}, \ \forall [i,j] \in L_0 \tag{VI.9}$$

where f_{ij}^k denotes the part of link capacity c_{ij} that is dedicated for commodity (i,k) use.

(ii) capacity assignment continuity

$$\sum_{j(\neq i)} f_{ij}^k - \sum_{j(\neq i)} f_{ji}^k = f_i^k, \ \forall (i,k) \in N_0 \tag{VI.10}$$

The constraint preserves a constant total assignment, say f_i^k, of network capacity to commodity (i,k) along all paths from node i to node k. Substituting (VI.10) into (VI.8) we obtain

$$t_i^k = \frac{q_i^k(0)}{\sum_{j(\neq i)} f_{ij}^k - \sum_{j(\neq i)} f_{ji} - a_i^k}, \quad \forall (i,k) \in N_0 \qquad \text{(VI.11)}$$

From all feasible capacity assignments that satisfy (VI.9) and (VI.10) we ask for one which has the minimal largest expected delivery time. This is the same Min-Max criterion we used in the formulation of the First Minimal (deterministic) Time Problem. The formulation of the corresponding First Minimal Expected Time Problem (METP(1)) and subsequent optimization problems are the subject of the next section.

C. SEQUENTIAL LINEAR OPTIMIZATION FORMULATION

With every feasible capacity assignment, $\{f_{ij}^k\}$ we associate a descriptor vector $T = (t_1, t_2, \ldots, t_M)$, $M \leq |N_0|$ of distinct expected times to empty the queues. We assume, without loss of generality that the components of T are ordered such that $t_i > t_j$, if $i > j$. Now, given two feasible capacity assignments F^A and F^B, we say that F^A dominates F^B iff $t_j^A < t_j^B$ and $t_i^A = t_i^B$, $i = 1,2,\ldots,j-1$ for some j, $j \leq \min(M_A, M_B)$ (cf. Def. II.2). The definition of optimal capacity assignment now follows directly.

Definition VI.1

We say that a capacity assignment is optimal for a given network state $\{q_i^k(0)\}$ and arrival rates $\{a_i^k\}$, $\forall (i,k) \in N_0$ iff

$$t_m^0 = \min_{\{f_{ij}^k\}} \{t_m | t_1^0, t_2^0, \ldots, t_{m-1}^0\}, \; m = 1,2,\ldots,M \qquad \text{(VI.12)}$$

□

Our objective is to show that the optimal capacity assignment can be obtained by solving an appropriate deterministic optimal delivery problem with <u>constant</u> rate data inputs.

We now present the statement of the First Minimal Expected Time Problem (METP(1), for short), in which the largest expected queue delivery time is minimized.

METP(1): $\min t_1$

$$\frac{q_i^k(0)}{\sum_{j(\neq i)} f_{ij}^k - \sum_{j(\neq i)} f_{ji}^k - a_i^k} = t_1, \quad \forall (i,k) \in N_0$$

$$\sum_{k(\neq i)} f_{ij}^k \leq c_{ij}, \; \forall [i,j] \in L_0 \qquad \text{(VI.13)}$$

$$t_1, f_{ij}^k \geq 0, \; \forall [i,j] \in L_0, \; \forall (i,k) \in N_0$$

□

The optimal solution to problem (VI.13) is a function of the initial network congestion $\{q_i^k(0)\}$ and the expected arrival rates $\{a_i^k\}$. This <u>open-loop</u> solution can be implemented, at least in principle, as <u>closed-loop</u> routing control by continuously

recalculating it in time, with the current network state (congestion) as initial condition for each problem. It should be pointed out that we do neither imply nor believe that this kind of implementation is possible unless the frequency with which a new solution is recomputed can be adjusted to match the actual transmission and computation capabilities of a network.

In spite of this fact, we will assume in the sequel that it is possible to recompute the capacity assignment with every new message arrival to the network. As a result we obtain a theoretical model which provides some new insight into what is the effect of including congestion information, in addition to that of arrival rates, on dynamic routing strategies.

Following the assumption above we specify the set N_0, in this chapter, to include those commodities (i,k) for which $q_i^k(0) > 0$. By doing so we do not allocate in (VI.13) capacity resources to empty queues (as long as they remain empty). If one "deletes" from problem (VI.13) those commodities for which $q_i^k(0) = 0$ (the expected time needed to empty those queues is not well defined), then its interpretation as a "minimal expected time to empty" is strictly valid, and the network could operate for some finite time with links saturated while also obeying the capacity assignments.

Using the transformation

$$u_{ij}^k \triangleq t_1 f_{ij}^k, \forall [i,k] \in L_0, \forall (i,k) \in N_0$$

and introducing slack variables we obtain the LP formulation of METP(1) in standard form.

METP(1): $\min t_1$

s.t.

$$-t_1 a_i^k + \sum_{j(\neq i)} u_{ij}^k - \sum_{j(\neq i)} u_{ji}^k = q_i^k(0), \forall (i,k) \in N_0$$

$$-t_1 c_{ij} + \sum_{k(\neq i)} u_{ij}^k + s_{ij} = 0 \qquad \text{(VI.14)}$$

$$t_1, u_{ij}^k \geq 0, \forall [i,j] \in L_0, \forall (i,k) \in N_0$$

□

The reader will recognize problem (VI.14) as a statement of the First Minimal (deterministic) Time Problem (cf. III.1b) with constant rate data inputs $\{a_i^k\}$ to the network. Although both formulations are mathematically identical, here we interpret t_1 to be the expected value of a delivery time.

It is easy to see (cf. VI.14) that as t_1 reduces to its minimal value t_1^0 a subset of links, say L_1, is bound to become critical. For this set of links the second constraint in (VI.13) holds with equality or, equivalently, the respective slack variables in (VI.14) must be identically zero.

It can be expected that many feasible capacity assignments will solve problem (VI.14). Because of this circumstance, we may ask for the "best" among many solutions. One way to approach this question will be to apply the same min-max criterion as in METP(1) to those links which are not critical, i.e. to all [i,j] such that $[i,j] \notin L_1$. We therefore seek next to minimize expected delivery time t_2, while retaining t_1^0 for all commodities that were assigned capacity in the set L_1.

We denote the set of these commodities by N_1. The Second Minimal Expected Time Problem can then be stated as

METP(2): $\min t_2$

$$\frac{q_i^k(0)}{\sum_{j(\neq i)} f_{ij}^k - \sum_{j(\neq i)} f_{ji}^k - a_i^k} = t_1^0, \quad \forall (i,k) \in N_1$$

$$\frac{q_i^k(0)}{\sum_{j(\neq i)} f_{ij} - \sum_{j(\neq i)} f_{ji} - a_i^k} \leq t_2, \quad \forall (i,k) \in N_0 - N_1 \qquad \text{(VI.15)}$$

$$\sum_{k(\neq i)} f_{ij}^k = c_{ij}, \quad \forall [i,j] \in L_1$$

$$\sum_{k(\neq i)} f_{ij}^k \leq c_{ij}, \quad \forall [i,j] \in L_0 - L_1$$

$$t_2,\ f_{ij}^k \geq 0, \quad \forall [i,j] \in L_0,\ \forall (i,k) \in N_0$$

□

Again, using the transformation

$$u_{ij}^k \triangleq t_2 f_{ij}^k \qquad \text{(VI.16)}$$

and rearranging results in an LP formulation

METP(2): $\min t_2$

s.t.

$$-t_2\left(\frac{q_i^k(0)}{t_1^0} + a_i^k\right) + \sum_{j(\neq i)} u_{ij}^k - \sum_{j(\neq i)} u_{ji}^k = 0, \quad \forall (i,k) \in N_1$$

$$-t_2 a_i^k + \sum_{j(\neq i)} u_{ij}^k - \sum_{j(\neq i)} u_{ji}^k = q_i^k(0), \quad \forall (i,k) \in N_0 - N_1$$

$$-t_2 c_{ij} + \sum_{k(\neq i)} u_{ij}^k = 0, \quad \forall [i,j] \in L_1 \qquad \text{(VI.17)}$$

$$-t_2 c_{ij} + \sum_{k(\neq i)} u_{ij}^k + s_{ij} = 0, \quad \forall [i,j] \in L_0 - L_1$$

$$t_2,\ u_{ij}^k,\ s_{ij} \geq 0, \quad \forall [i,j] \in L_0, \quad \forall (i,k) \in N_0$$

for given t_1^0.

□

It may also happen that the solution to METP(2) is not unique. The procedure could then be repeated by identifying additional critical links [i,j] and defining L_2 to be a set of links including the new critical links as well as those in L_1. Similarly we define N_2 to be the set of all commodities that have capacity assignments in the set L_2. Continuing to iterate in this way until we have exhausted all commodities, we ultimately generate a capacity assignment for which (VI.12) is satisfied and thus is <u>optimal</u>.

For completeness, we present the linear programming formulation of the m-th Minimal Expected Time Problem.

METP(m): $\min t_m$

s.t.

$$-t_m\left(\frac{q_i^k(0)}{t_p^0}+a_i^k\right) + \sum_{j(\neq i)} u_{ij}^k - \sum_{j(\neq i)} u_{ji}^k = 0, \quad \forall (i,k) \in N_p - N_{p-1}, \quad p = 1,2,\ldots,m-1$$

$$-t_m a_i^k + \sum_{j(\neq i)} u_{ij}^k - \sum_{j(\neq i)} u_{ji}^k = q_i^k(0), \quad \forall (i,k) \in N_0 - N_{m-1}$$

$$-t_m c_{ij} + \sum_{k(\neq i)} u_{ij}^k = 0, \quad \forall [i,j] \in L_{m-1} \qquad \text{(VI.18)}$$

$$-t_m c_{ij} + \sum_{k(\neq i)} u_{ij}^k + s_{ij} = 0, \quad \forall [i,j] \in L_0 - L_{m-1}$$

$$t_m,\ u_{ij}^k \geq 0, \quad \forall [i,j] \in L_0, \quad \forall (i,k) \in N_0$$

for given $t_1^0, t_2^0, \ldots, t_{m-1}^0$

$$N_{-1} \triangleq \emptyset$$

□

The SLO methodology establishes a hierarchical order among the critical sets, as indicated in Fig. VI.4. The arrows there indicate, for a given set of links, which are the commodities that may (must--for a horizontal arrow) use it. The question of identification of critical sets (or their partial composition) was considered in detail in Chapter V.5 and we need not repeat it here.

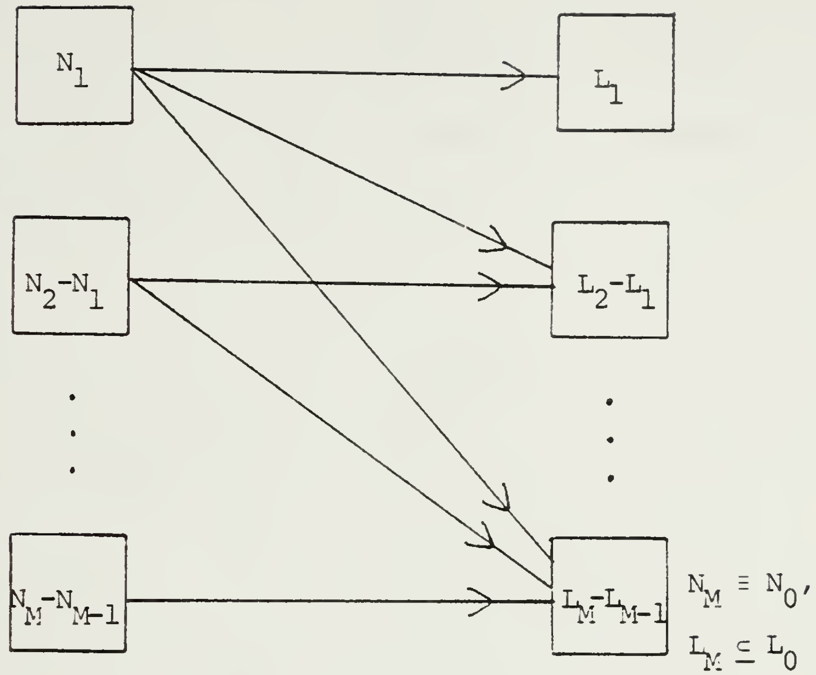


Fig. VI.4. Hierarchical Structure of Critical Sets

The following example of a stochastic delivery problem illustrates some of the concepts that we have introduced in this section.

Example:

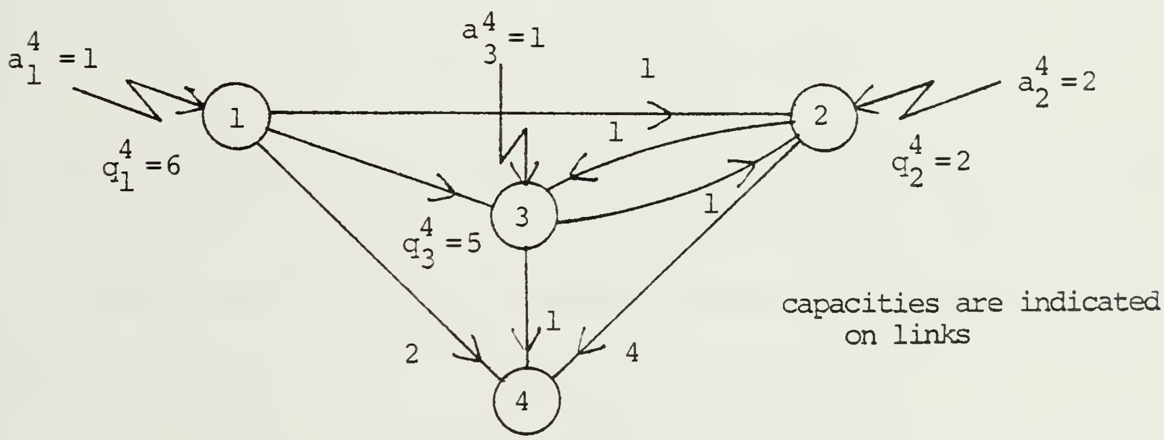


Fig. VI.5. Stochastic Delivery Problem

Following the argumentation we used in discussion of the network shown in Fig. VI.5 (cf. Chapter V.F) it can be shown that the first critical set N_1 is composed of commodity (3,4) only. In this case the assignment of capacity to commodity (3,4) is unique and shown in Fig. VI.5a.

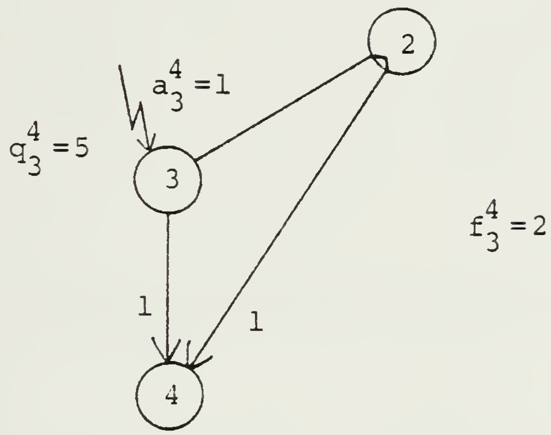


Fig. VI.5a. Optimal Capacity Assignment for Commodity (3,4)

The minimal expected time to empty the queue q_3^4 is given by

$$t_1^0 = \frac{q_3^4}{f_3^4 - a_3^4} = \frac{5}{2-1} = 5$$

Commodities (1,4) and (2,4) belong to N_2-N_1 and the corresponding capacity assignment is shown in Fig. VI.5b. The second minimal expected time is computed (using the values in Fig. VI.5b) to be

$$t_2^0 = \frac{q_1^4}{f_1^4 - a_1^4} = \frac{6}{2.5 - 1} = 4 \ (\frac{2}{2.5 - 2} = 4)$$

□

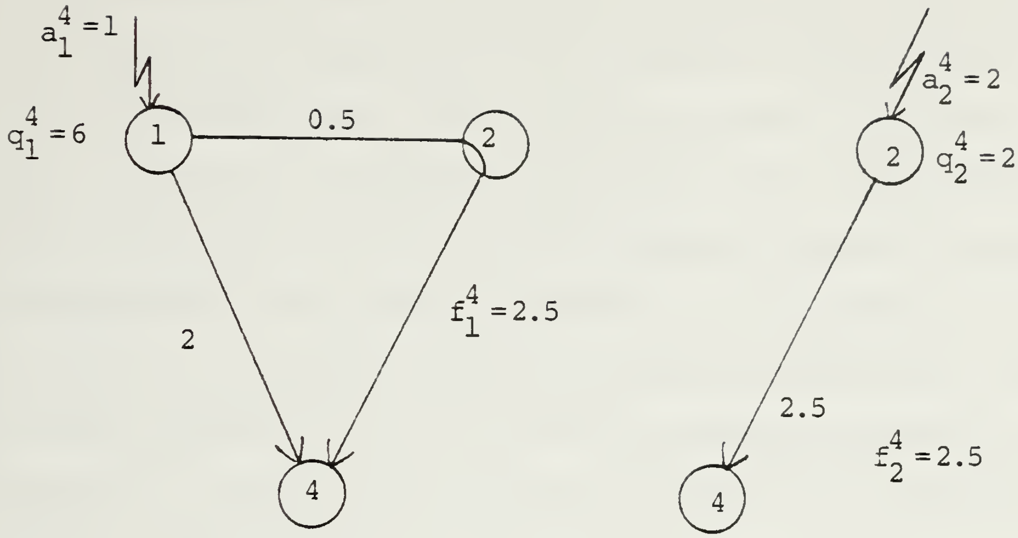


Fig. VI.5b. Optimal Capacity Assignment for Commodities in N_2-N_1

We conclude this section with an observation regarding the optimality criterion in Definition VI.1. Since the mathematical formulation of MTP(1) and METP(1) are identical we can interpret the stochastic delivery problem as a deterministic delivery problem with constant rate inputs. We could ask then for an optimal capacity assignment (that may change with time) which consists of two consistent parts, one that accommodates the constant input rates in such a way that the other enables optimal delivery of the backlogged data (in the optimal delivery function sense). Since the modifications necessary to make the time (MTP(m)) and the rate (MRP(m)) problems handle constant inputs are trivial, this approach would result in an optimal capacity assignment schedule (utilizing a sequence of corresponding METP's and MERP's).

Although our previous results indicate that this sequential optimization approach of interleaved time/rate problems is more powerful than that of time problems only, we do not use it because of conceptual difficulty that arises between firm scheduling of events in time and the potential necessity to recompute a new capacity schedule. For example, it is possible that in an optimal capacity assignment schedule, a queue of some commodity $(i,k) \in N_0$ will be assigned capacity only following a point in time which is beyond the recomputation instant. This can lead, at least in theory, to tremendous delays in the delivery of that queue. For this reason we find it conceptually more satisfactory to consider constant capacity assignment (as derived in this section) in our model.

D. DISCUSSION

In the last section we have shown that an optimal capacity assignment can be obtained as a result of solving a sequence of linear programming programs (METP's). We also have indicated that the mathematical formulation of a stochastic delivery problem is identical to that of a corresponding deterministic problem with constant rate data inputs. Due to this similarity we can apply the theory developed in previous chapters to derive and understand the structural properties of an optimal solution.

The derivation of routing variables has been studied mainly in two extreme situations. In the first case only the steady-state arrival rates of messages are taken into consideration. The performance measure objective usually is to minimize expected

delay experienced by a message traversing the network or, as in [8], to minimize a sequence of saturation ratios. In any case, the underlying network model is that of Kleinrock [15], as described in Section A. In the other case, only the congestion state of a network is utilized to compute a routing flow schedule that attains optimal delivery function. This approach is studied in detail in this thesis.[†] The dynamic delivery problem introduced in this chapter, provides a theoretical model which will combine both types of information, i.e. the expected arrival rates of messages and the existing congestion in a network. We combine the objective of emptying the set of initial queues with the probabilistic information about future expected arrival rates of messages. We believe that the resulting optimal capacity assignment provides a fairly accurate analytic model for a desired dynamic (short term) routing strategy (routing variables). How to use such routing variables (whatever their origin) in the implementation of an actual network control system is a complex matter and only initial[††] results are available. We will not consider this issue here.

We have indicated before that it is unrealistic to expect that the new dynamic model can actually be used to compute routing variables in real network environments. We do suggest, however, that it may prove useful in simulation studies as

---

[†]For relation to other results, see Chapter I.

[††]See for example [21], [22] and [23].

a reference against which actual routing strategies can be compared.

In this connection, it is worth noting that it is also possible to solve the dynamic routing problem which results when all of the a_i^k (not just those for which $q_i^k(0)$ is non-zero) are involved in allocation of the $\{f_{ij}^k\}$. In practice this requires interpretation of the set N_0 in (VI.14) and in subsequent time problems as the set of all commodities. This does not change anything in the mathematical procedure used to solve those problems. The conceptual difference, however, arises due to the assignment of capacity resources to commodities for which $q_i^k(0) = 0$. Since the message arrivals to the network are stochastic in nature, it now may happen that for a certain period of time the assigned capacity f_i^k (for the commodities in question) will not be fully utilized. This model variant, although in theory inferior to the one discussed earlier, is probably an acceptable compromise between the static and the dynamic network models. Its obvious advantage results from the fact that it is not necessary to recompute the capacity assignment with every new message arrival to the network but rather as frequently as the actual computational facilities allow it. Finally, we mention that the new stochastic delivery model seems to be free from the "independence assumption." Moreover, observe that if all the q_i^k are set equal to one, the resulting (static) solution optimizes the objective function which (sequentially) seeks to minimize the maximum expected length of the queueing system busy periods.

VII. APPLICATION TO NETWORKS WITH TRAVERSAL DELAYS

A. INTRODUCTION

Up to this point we have considered multicommodity delivery problems for which the delay in delivery was caused by finite capacity of network links. A natural extension of this model is to consider networks with traversal delays. Association of traversal delay with each one of the links complicates the mathematics of the optimal delivery problem, but provides a considerably improved model for transportation applications.

The classical transportation problem (see, for example, [7]) refers to the shipment of assets† from a set of sources to a set of destinations, to satisfy given demand at minimal cost. An important class of extensions of this problem recognizes the existence of queueing and traversal delays, and consequently looks into the question of minimal time demand satisfiability. We prefer to view this issue in a more general framework of Minimal Time Redistribution Problem (MTRP); given initial and desired distributions of assets, in some geographical locations, the objective is to redistribute the assets accordingly in minimal time. Thus from our point of view there is no inherent distinction anymore between "source" and "destination" nodes.

---

†We use "asset" instead of the more common term "commodity" to distinguish it from our definition of commodity in Chapter II. There commodity was identified with destination node, where here one type of asset may be demanded in many locations, and a location may have demand for many types of assets.

An obvious and important application of this class of problems is to military logistics planning. In particular, assume that an outbreak of hostilities in a number of locations requires redistribution of assets (troops, tanks, supplies, etc.) in minimal time. The same model may be used for supply of aid to disaster-struck areas or for transportation of perishable supplies, and many others.

The usefulness and applicability of minimal time problems has attracted great attention (see [24] for survey and exhaustive list of references). Most of the research, however, has been done in the area of bi-partite transportation networks. Hammer [25]† provided an algorithm to solve a single asset, uncapacitated minimal time problem. Tapiero and Soliman [27] have treated the multi-asset version of the capacitated minimal time problem as an optimal control problem, using a maximum principle and a continuous state-space framework. Their paper does not contain proof of their algorithm. Bookbinder and Sethi [24] use basically the same approach but elaborate more on the mathematical programming aspect of their algorithm, for which only convergence to a local minimum is assured. In both cases it is unclear whether or not the algorithms are computationally manageable for problems of practical size.

In [24] an important observation is made, namely that at least for bi-partite transportation networks, <u>capacity linking constraints</u> (to be discussed in the next section) cause most

---

†For similar results see [26].

of the complexity in MTRP. The authors predict there that it may not be easy or even possible to include a capacity linking constraint and still provide a linear programming formulation of the problem. In this chapter we study this question and come to a conclusion that Sequential Linear Optimization methodology can be used to solve MTRP. Also, we analyze the special nature of capacity linking constraints and their influence on solution procedure complexity. Then we consider possible extension (by discrete time modelling) of the results to general networks. Application to military decision problems is provided in the formulation of the Maximally Delayed Decision Problem (MDDP).

B. TRANSPORTATION NETWORK MODEL

1. Topological Representation

A useful way to view the redistribution of assets over a set of locations is in terms of a network model composed of nodes and links. The links represent unidirectional means of asset transportation and the nodes represent physical locations. A typical transportation model is shown in Fig. VII.1.

With each link we associate a traversal time, i.e. the time required by the corresponding transportation mode to traverse the distance between the locations represented by the head and tail nodes of that link, respectively.

In general, we allow for a variety of capacity constraints, the most common of which concern loading, unloading

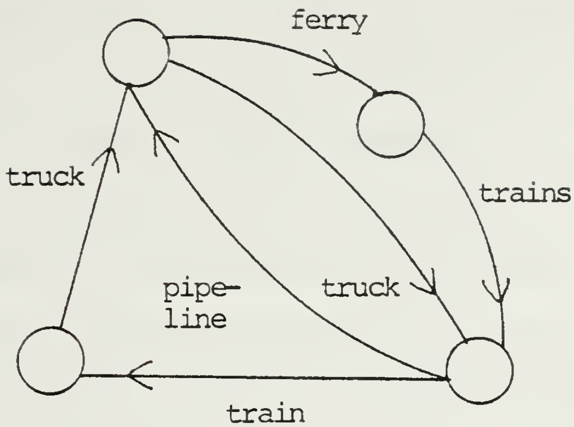


Fig. VII.1. Transportation Network

and link capacities.† The loading and unloading constraints (sometimes referred to as "linking constraints") provide an upper bound on the volume of assets that can be loaded or unloaded (in general, with respect to particular transport mode) at a given location per unit time. The link capacity constraint represents the upper bound on the volume of transportation mode and thus on the total amount of assets per unit time that can be sent over that link.

With each node we associate, at every interval of time, the amount of assets stored at the corresponding location. The collection of these descriptors for all the nodes in the network and for assets in transit constitutes the state of the system.

---

†One also may consider an upper bound on the amount of assets allowed at any given time at a particular node.

Since none of our results is dependent on the number of asset types nor on the number of different transportation modes we will, for the sake of simplicity, limit the notation to represent a single type of asset and a single mode of transportation. Generalization to more than one type of each is straightforward. Furthermore, since most of the notation and the meaning of network parameters are the same as in Chapter II, we will keep our discussion brief whenever possible.

Consider a transportation network $G(V,L_0)$, where $V = \{1,2,\ldots,n\}$ is a set of n nodes and $L_0 = \{[i,j]\}$ is a set of ℓ links. We also define:

$q_i(t)$ = amount of asset stored at node i at time t, $\forall i \in V$;

$f_{ij}(t)$ = rate of asset flow <u>leaving</u> node i on link $[i,j]$ at time t, $\forall [i,j] \in L_0$

c_{ij} = capacity of link $[i,j]$ (or of the transportation mode represented by that link), $\forall [i,j] \in L_0$

a_i = loading capacity at node i, $\forall i \in V$

b_j = unloading capacity at node j, $\forall j \in V$

τ_{ij} = traversal delay from node i to node j along link $[i,j]$, $\forall [i,j] \in L_0$.

We reserve the use of respective capital letters for set and vector notation, interchangeably. For example, the quantities of assets stored in network nodes at time t are given by $Q(t) = (q_1(t),q_2(t),\ldots,q_n(t))$.

2. Dynamic System Equations and Constraints

The quantities just defined must satisfy three basic constraints: non-negativity, conservation and capacity. The non-negativity constraint states that

$$f_{ij}(t) \geq 0 \qquad \forall [i,j] \in L_0, \ \forall t,$$
$$q_i(t) \geq 0, \qquad \forall i \in V, \ \forall t. \qquad \text{(VII.1)}$$

The conservation constraint may be written as

$$q_i(t_2) = q_i(t_1) - \int_{t_1}^{t_2} \Big(\sum_{j(\neq i)} f_{ij}(\alpha) \Big) d\alpha$$
$$+ \int_{t_1}^{t_2} \Big(\sum_{j(\neq i)} f_{ji}(\alpha - \tau_{ji}) \Big) d\alpha, \ \forall i \in V \qquad \text{(VII.2)}$$

and

$$t_2 > t_1.$$

Constraint (VII.2) accounts for the fact that flows arriving at node i, $i \in V$ over link [j,i] at time t, left node j at time $t-\tau_{ji}$, where τ_{ji} is the traversal delay associated with link [j,i]. Finally, the capacity constraints are

$$f_{ij}(t) \leq c_{ij}, \quad \forall [i,j] \in L_0, \ \forall t$$
$$\sum_{j(\neq i)} f_{ij}(t) \leq a_i, \quad \forall i \in V \qquad \text{(VII.3)}$$

$$\sum_{j(\neq i)} f_{ji}(t) \leq b_j, \quad \forall j \in V$$

Definition VII.1

A set of flows F(t) is a feasible flow schedule if it satisfies (VII.1)-(VII.3) for all t.

□

Let Q_0 and Q_1 denote the initial and the desired (terminal) distributions of assets (network states), respectively. We say that Q_1 is reachable from Q_0 if there exists a feasible flow schedule F(t), $t_0 \leq t \leq t_1$ such that F(t): $Q_0(t_0) \rightarrow Q_1(t_1)$ and $0 \leq t_0 < t_1 < \infty$.† Consequently, for any such pair (Q_0, Q_1) there is some minimal value of t_1. We define the minimal redistribution time t_1^0 as

$$t_1^0 = \min_{\{F(t)\}} \{t_1 | F(t): Q_0(0) \rightarrow Q_1(t_1)\} \qquad \text{(VII.4)}$$

Definition VII.2

We say that a feasible flow schedule F(t): $Q_0(0) \rightarrow Q_1(t_1)$ is an optimal solution to the minimal time redistribution problem if $t_1 = t_1^0$.

□

C. BI-PARTITE NETWORKS

1. Problem Statement

In this section we study the minimal time redistribution problem on bi-partite networks. A bi-partite network is one

---

†We of course exclude the trivial case where $Q_0 \equiv Q_1$. Also, we assume without loss of generality that the initial time t_0 is identically zero.

whose node set can be partitioned into two subsets S and D, so that each link has its head node in S and its tail node in D. A typical network of this nature is shown in Fig. VII.2.

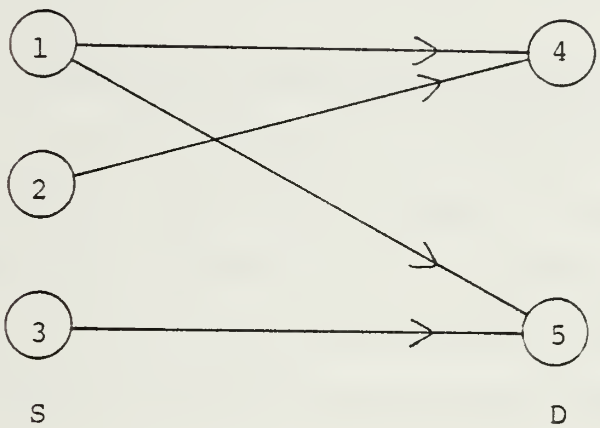


Fig. VII.2. Bi-partite Network

It is customary to associate the set S with "supply" nodes, and the set D with "demand" nodes. Various amounts of assets are stored initially at each of the $I \triangleq |S|$ supply nodes and there is a specified requirement for assets at each one of the $J \triangleq |D|$ demand nodes. Following our notation we write:

$$
\begin{aligned}
Q_0 &\triangleq (s_1, s_2, \ldots, s_I, 0, 0, \ldots, 0) \\
Q_1 &\triangleq (-, -, \ldots, -, -, d_1, \ldots, d_J)
\end{aligned}
\qquad \text{(VII.5)}
$$

The notation (-) indicates "don't care" situation. It is not important how many assets remain in the supply set S (as long as these are non-negative quantities).

We can restate now the Minimal Time Redistribution Problem as follows:

For any given pair of system states (Q_0, Q_1) and network parameters $T = \{\tau_{ij}\}$, $C = \{c_{ij}\}$, $A = \{a_i\}$, $B = \{b_j\}$, where $i = 1,2,\ldots,I$ and $j = 1,2,\ldots,J$, find a minimal time flow schedule.

2. Structure of the Minimal Time Flow Schedule

In this subsection we analyze the structural properties of the minimal time flow schedule. We derive a result which parallels that of Thm. II.1 and enables us later on to formulate the minimal time redistribution problem in LP form. We start with the trivial but important observation that

Lemma VII.1

Let $F(t)$, $0 \leq t \leq t_1$ be a feasible flow schedule such that $F(t)\colon Q_0(0) \rightarrow Q_1(t_1)$. Then we may always take

$$f_{ij}(t) = 0 \begin{cases} t_1 - \tau_{ij} < t < t_1, & \text{if } \tau_{ij} < t_1 \\ & \quad \forall [i,j] \in L_0 \\ 0 \leq t \leq t_1, & \text{if } \tau_{ij} > t_1 \end{cases} \tag{VII.6}$$

Proof:

Consider a bi-partite network (like that shown in Fig. VII.2) and assume that there is some feasible flow solution $F(t)$, $0 \leq t \leq t_1$ such that $F(t)\colon Q_0(0) \rightarrow Q_1(t_1)$. It is quite obvious that this flow schedule cannot use any link $[i,j]$ for which $\tau_{ij} > t_1$, since all the flows on this link will arrive at their

destination j later than t_1, and thus be of no use in the context of the minimal time problem. Also, there is no point in sending flows over a useful link $[i,j]$ ($\tau_{ij} < t_1$) beyond time $t_1-\tau_{ij}$ because they will not reach their destination in time. $\square$

Combining (VII.2), (VII.5) and (VII.6) we can express the desired system state with respect to the demand nodes as

$$d_j = \sum_i \int_0^{t_1-\tau_{ij}} f_{ij}(t)dt, \quad \forall j \in D \qquad \text{(VII.7)}$$

Similarly, the non-negativity constraint (VII.1) with respect to the supply nodes can be written as

$$s_i \geq \sum_j \int_0^{t_1-\tau_{ij}} f_{ij}(t)dt, \quad \forall i \in S \qquad \text{(VII.8)}$$

The minimal time redistribution problem reduces to finding a flow schedule which satisfies (VII.7) and (VII.8), subject to capacity constraints, in minimal time.

We now show that the search for a minimal time flow schedule may be confined similarly to the optimal delivery problem, to the class of piecewise constant flow schedules. Furthermore, it is possible even to narrow this class to flow schedules which we call linking flow schedules. In order to present this subclass we need to introduce some additional notation.

For a given bi-partite transportation network we define for each supply node i, $i \in S$ a vector $T_i \triangleq (\tau_i(1),\tau_i(2),\ldots,\tau_i(n_i))$

of all distinct traversal times associated with outgoing links of that node. The components of T_i are assumed to be ordered so that $\tau_i(k) < \tau_i(r)$ if $k < r \le n_i$, and n_i is the number of vector components. By n_{ij} we denote the ordinal position of an element of T_i such that $\tau_i(n_{ij}) = \tau_{ij}$. For the network in Fig. VII.3 we have

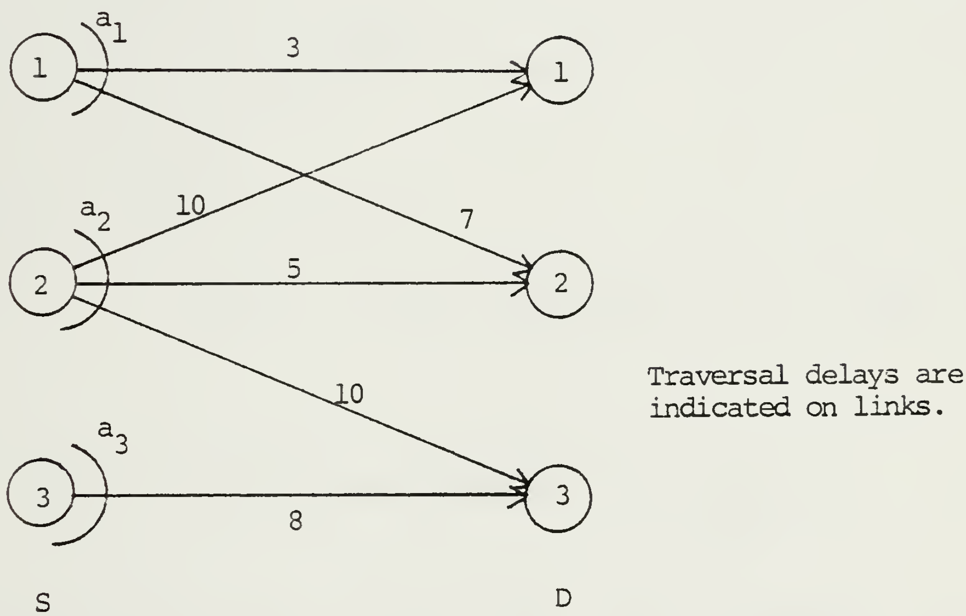


Fig. VII.3. Bi-partite Transportation Network

$T_1 = (3,7)$, $T_2 = (5,10)$, $T_3 = (8)$

$n_1 = 2$, $n_2 = 2$, $n_3 = 1$

$n_{11} = 1$, $n_{12} = 2$, $n_{21} = 2$, $n_{22} = 1$, $n_{23} = 2$,

$n_{33} = 1.$

We now use this example† to demonstrate the structure of a linking flow schedule.

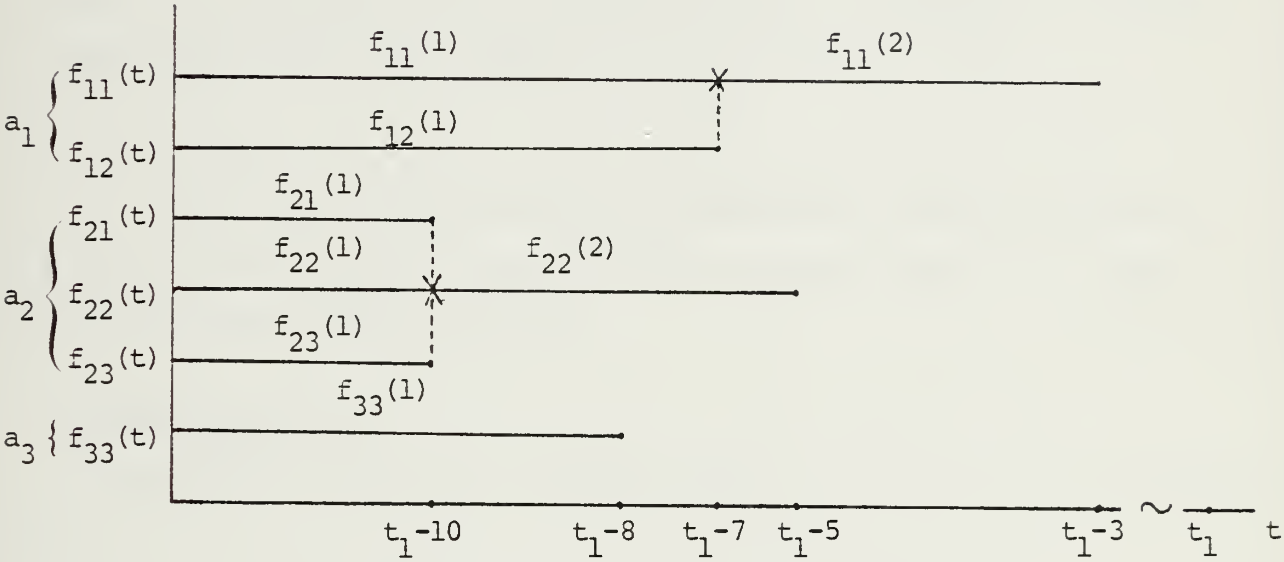


Fig. VII.4. Linking Flow Schedule ($t_1 > \max_{L_0} \tau_{ij}$)

The length of a horizontal line in Fig. VII.4 expresses the time duration of effective flows on that link (Lemma VII.1). The labels above a line identify the constant segments of a flow. The crosses (or corner points) indicate the time instances at which the flow may change its value. We do not consider flow termination points as corner points in this context.

---

†We consider loading and link capacities. Unloading capacity constraints are studied later on in this chapter.

The flow schedule in Fig. VII.4 was obtained by using the following construction rule.

Construction Rule:

A flow[†] on a link [a,b] has a corner point at time t *iff* link [a,b] is linked by a loading capacity constraint to some other link [c,d] such that $t = t_1 - \tau_{cd}$, and $\tau_{cd} > \tau_{ab}$.

□

Now we are in position to formulate a general description of a linking flow schedule for the case when $t_1 > \max_{[i,j]} \tau_{ij}$, i.e. all the links are usable.

Definition VII.3

A feasible linking flow schedule F(t), $0 \leq t \leq t_1$ is described by

$$f_{ij}(t) = \begin{cases} f_{ij}(1), & 0 \leq t \leq t_1 - \tau_{ij}(1) \\ \vdots & \vdots \\ f_{ij}(m), & t_1 - \tau_{ij}(m-1) < t \leq t_1 - \tau_{ij}(m) \\ \vdots & \vdots \\ f_{ij}(n_{ij}), & t_1 - \tau_{ij}(n_{ij}-1) < t \leq t_1 - \tau_{ij}(n_{ij}) \end{cases} \qquad \text{(VII.9)}$$

where $T_i \triangleq (\tau_{ij}(1), \ldots, \tau_{ij}(n_i))$ is a vector all distinct traversal delays of links linked to [i,j] by a loading constraint, $\forall [i,j] \in L_0$. Also, we define $\tau_{ij}(0) \triangleq t_1$.

□

[†]Of each of the asset-types, in multiasset case.

For notational convenience, we define $\delta_i(m)$ to denote the duration of the m-th flow segment $f_{ij}(m)$,

$$\delta_i(m) \triangleq \tau_{ij}(m-1) - \tau_{ij}(m), \quad m = 1,2,\ldots,n_{ij} \qquad \text{(VII.10)}$$

Since our interest lies, as indicated before, within the class of linking flow schedules, it is useful to rewrite the constraints (VII.1)-(VII.3) (we also use (VII.7) and (VII.8)) as they apply to linking flow schedules.

(i)--non-negativity

$$f_{ij}(m) \geq 0, \quad \forall [i,j] \in L_0, \quad m = 1,2,\ldots,n_{ij}$$

$$s_i \geq \sum_j \sum_{m=1}^{n_{ij}} f_{ij}(m)\delta_i(m), \quad \forall i \in S \qquad \text{(VII.11)}$$

(ii)--conservation

$$d_j = \sum_i \sum_{m=1}^{n_{ij}} f_{ij}(m)\delta_i(m), \quad \forall j \in D \qquad \text{(VII.12)}$$

(iii)--capacity

$$f_{ij}(m) \leq c_{ij}, \quad \forall [i,j] \in L_0, \; m = 1,2,\ldots,n_{ij}$$

$$\sum_j f_{ij}(m) \leq a_i, \quad \forall i \in S, \; m = 1,2,\ldots,n_{ij} \qquad \text{(VII.13)}$$

Finally we state and prove the main result of this section.

Theorem VII.1

Let Q_0 and Q_1 be any two states of a given bi-partite transportation network such that $Q_1(t_1)$ is reachable from $Q_0(0)$ by some flow schedule F(t), $0 \leq t \leq t_1$. Then there exists a feasible linking flow schedule $\hat{F}(t)$, $0 \leq t \leq t_1$ such that $\hat{F}(t): Q_0(0) \to Q_1(t_1)$.

Proof:

Define (for useful links)

$$\hat{f}_{ij}(m) \triangleq \frac{1}{\delta_i(m)} \int_{t_1-\tau_{ij}(m-1)}^{t_1-\tau_{ij}(m)} f_{ij}(t)dt, \qquad \text{(VII.14)}$$

$$\forall [i,j] \in L_0, \quad m = 1,2,\ldots,n_{ij}$$

If we start now in state $Q_0(0)$ and apply the new flow schedule $\hat{F}(t)$ as defined in (VII.14), the system will be transfered by time t_1 into a new state, say $\hat{Q}_1(t_1)$ (cf. VII.12), such that

$$\hat{d}_j(t_1) = \sum_i \sum_{m=1}^{n_{ij}} \hat{f}_{ij}(m)\delta_i(m), \quad \forall j \in D \qquad \text{(VII.15)}$$

Substituting (VII.14) into (VII.15) results in

$$\hat{d}_j(t_1) = \sum_i \int_0^{t_1-\tau_{ij}} f_{ij}(t)dt, \quad \forall j \in D \qquad \text{(VII.16)}$$

and from (VII.7) we conclude that

$$\hat{d}_j(t_1) = d_j(t_1), \quad \forall j \in D \qquad \text{(VII.17)}$$

Equation (VII.17) is essentially the desired result but we still must show that $\hat{F}(t)$, as defined in (VII.14), satisfies constraints (VII.11)-(VII.13). The first part of (VII.11) is trivial due to the requirement $f_{ij}(t) \geq 0$, $\forall [i,j] \in L_0$, $\forall t$. The second part of this constraint is shown to be satisfied by substituting (VII.14) into it and comparing the result with (VII.8).

To show that $\hat{f}_{ij}(m) \leq c_{ij}$, $\forall [i,j] \in L_0$, $m = 1,2,\ldots,n_{ij}$ we conclude from (VII.14) that

$$\hat{f}_{ij}(m) \leq \max_{[t_1-\tau_{ij}(m-1),t_1-\tau_{ij}(m)]} \{f_{ij}(t)\}, \qquad m = 1,2,\ldots,n_{ij} \tag{VII.18}$$

But since by assumption the right hand side of (VII.18) must satisfy the link capacity constraint, so must $\hat{f}_{ij}(m)$.

To show that the loading constraint holds for the new flow schedule, we write (by assumption)

$$\sum_j f_{ij}(t) \leq a_i, \quad \forall i \in S, \ \forall t \tag{VII.19}$$

Integrating both sides of (VII.19) over the interval $[t_1-\tau_{ij}(m-1), t_1-\tau_{ij}(m)]$, for any m such that $1 \leq m \leq n_{ij}$ results in

$$\frac{1}{\delta_i(m)} \int_{t_1-\tau_{ij}(m-1)}^{t_1-\tau_{ij}(m)} \left(\sum f_{ij}(t)\right)dt \leq a_i, \tag{VII.20}$$

$$\forall i \in S, \quad m = 1,2,\ldots,n_{ij}$$

But (VII.20) can be written as

$$\sum_j \frac{1}{\delta_i(m)} \int_{t_1-\tau_{ij}(m-1)}^{t_1-\tau_{ij}(m)} f_{ij}(t)dt = \sum_j \hat{f}_{ij}(m) \leq a_i \quad \text{(VII.21)}$$

$$\forall i \in S, \quad m = 1,2,\ldots,n_{ij}$$

which completes the proof.

□

An obvious consequence of Thm. VII.1, with respect to minimal time flow schedules, is given in the following corollary.

Corollary VII.1

If there exists a feasible minimal time flow schedule then there also exists a feasible minimal flow schedule of the linking type.

□

It is worth noting the difference between the optimal solution to MTP(1) for a network without traversal delays, and the minimal time flow schedule here. In the first case it was sufficient to consider a constant flow schedule, whereas here we need a piecewise constant solution, which serves as a first indication of higher complexity of delivery problems on networks with traversal delays.

3. Solution Algorithm

Armed with the results of the last subsection we may state the Minimal Time Redistribution Problem as follows.

MTRP: $\min t_1$

s.t.

$$\sum_i \sum_{m=1}^{n_{ij}} f_{ij}(m)\delta_i(m) = d_j, \quad \forall j \in D$$

$$\sum_j \sum_{m=1}^{n_{ij}} f_{ij}(m)\delta_i(m) \le s_i, \quad \forall i \in S \qquad \text{(VII.22)}$$

$$f_{ij}(m) \le c_{ij} \quad \forall [i,j] \in L_0, \; m = 1,2,\ldots,n_{ij}$$

$$\sum_j f_{ij}(m) \le a_i, \quad \forall i \in S, \; m = 1,2,\ldots,n_{ij}$$

$$t_1, f_{ij}(m) \ge 0, \quad \forall [i,j] \in L_0, \; m = 1,2,\ldots,n_{ij}$$

□

We have assumed that $t_1^0 > \max_{L_0}\{\tau_{ij}\}$, i.e. that all links are usable.

Inspection of the first two constraints of (VII.22) shows that in both the coefficient of $f_{ij}(1)$, $\forall [i,j] \in L_0$ has the form (VII.10):

$$\delta_i(1) = t_1 - \tau_i(1), \quad \forall i \in S$$

The value of this coefficient is unknown since it is a function of the variable t_1. Define

$$u_{ij}(1) \triangleq f_{ij}(1)\delta_i(1), \quad \forall [i,j] \in L_0 \qquad \text{(VII.23)}$$

Using (VII.23) and rearranging (VII.22) results in

MTRP: $\min t_1$

s.t.

$$\sum_i u_{ij}(1) + \sum_i \sum_{m=2}^{n_{ij}} f_{ij}(m)\delta_i(m) = d_j, \ \forall j \in D$$

$$\sum_j u_{ij}(1) + \sum_j \sum_{m=2}^{n_{ij}} f_{ij}(m)\delta_i(m) \leq s_i, \ \forall i \in S \qquad \text{(VII.24)}$$

$$-t_1 c_j + u_{ij}(1) \leq -\tau_i(1) c_{ij}, \ \forall [i,j] \in L_0$$

$$f_{ij}(m) \leq c_{ij}, \ \forall [i,j] \in L_0, \ m = 2,\ldots,n_{ij}$$

$$-t_1 a_i + \sum_j u_{ij}(1) \leq -\tau_i(1) a_i, \ \forall i \in S$$

$$\sum_j f_{ij}(m) \leq a_i, \ \forall i \in S, \ m = 2,3,\ldots,n_{ij}$$

$$t_i, \ u_{ij}(1), \ f_{ij}(m) \geq 0, \ \forall [i,j] \in L_0, \ m = 2,3,\ldots,n_{ij}$$

□

which is a linear programming problem.

It should be noted that the formulation in (VII.24) is valid only when $t_1 - \tau_i(1) > 0$, $\forall i \in S$ which makes the inverse transformation of (VII.23) meaningful. But this is taken care of by the assumption that $t_1^0 > \max_{L_0} \{\tau_{ij}\}$. Let $T \triangleq (\tau(1),\ldots,\tau(n))$ be the vector of all distinct traversal delays of the bi-partite network. The components of T are

assumed to be in descending order of their size, and n is their number. Thus for example the statement: $t_1^0 > \max_{L_0} \{\tau_{ij}\}$, is equivalent to

$$t_1^0 > \tau(1) \tag{VII.25}$$

Suppose now, as may be the case, that $t_1^0 < \tau(1)$, and in particular that

$$\tau(k+1) < t_1^0 < \tau(k) \tag{VII.26}$$

for some k, $1 \leq k \leq n-1$. It should be clear that the solution to MTRP will yield (due to the nonnegativity of $u_{ij}(1)$, $\forall [i,j] \in L_0$)

$$t_1^0 = \tau(1) \tag{VII.27}$$

Equation (VII.27) suggests itself as an indicator of the fact that $t_1^0 < \tau(1)$. We may think of solving MTRP again, but this time we "remove" all the links for which $\tau_{ij} = \tau(1)$. It may happen now that $t_1^0 = \tau(2)$, in which case we repeat the procedure with respect to links for which $\tau_{ij} = \tau(2)$. If $\tau(2) < t_1^0 \leq \tau(1)$ then we have found an optimal solution.

Definition VII.4

The m-th, $m \geq 1$, Minimal Time Redistribution Problem (MTRP(m)) is equivalent to MTPR for which all the links $[i,j] \in L_0$, such that $\tau_{ij} > \tau(m)$ are considered to be not usable.

□

Actually, we are free to search for t_1^0 in any order, and in particular we may implement a "binary search" technique [28]. The resulting algorithm will have the following form:

Algorithm:

$$m \leftarrow \left\lceil \frac{n}{2} \right\rceil$$

Loop

Solve MTRP(m) (VII.28)

if $t_1^0 = \tau(m+1)$ then $m \leftarrow m + \left\lceil \frac{n-m}{2} \right\rceil$,
repeat loop.

if $t_1^0 > \tau(m)$ then $m \leftarrow \left\lceil \frac{m}{2} \right\rceil$, repeat loop.

else stop,

where n is the number of distinct traversal delays in the network and $\lceil X \rceil$ denotes the ceiling (i.e. the smallest integer $\geq$ X) of the number X.

□

Since the complexity of a "binary search" is logarithmic with respect to the number of elements being searched we conclude that

Lemma VII.1

The MTRP can be solved in $0(\lg_2 n)$ number of steps, each being a solution of an LP, where n is the number of distinct traversal delays in the network.

□

In the next subsection we give a simple example of MTRP.

4. Example

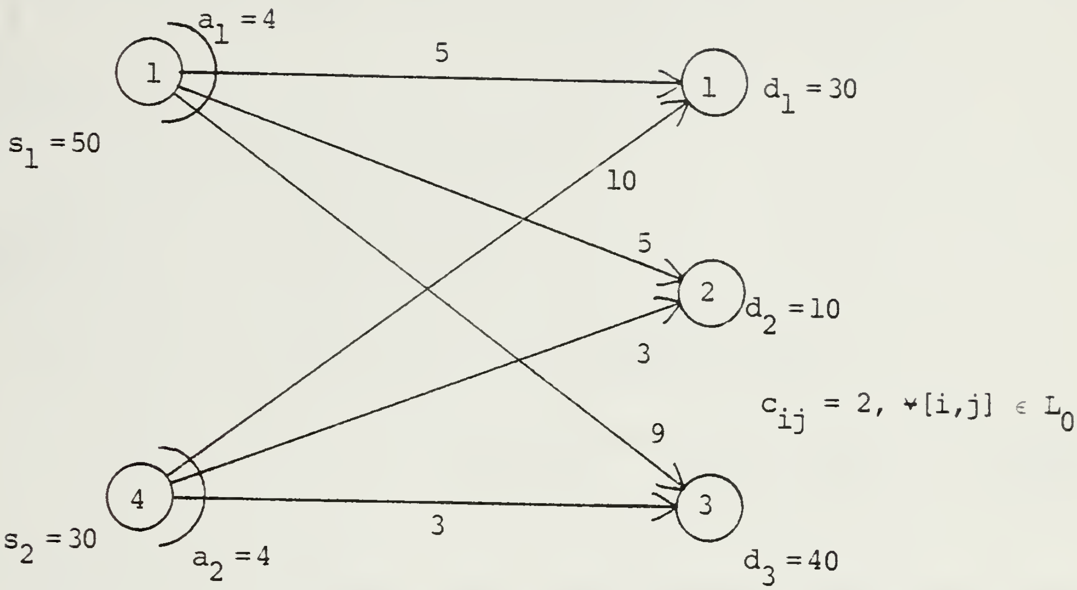


Fig. VII.5. Minimal Time Redistribution Problem on Bi-partite Network

The MTRP shown in Fig. VII.5 was solved using the methodology of the last subsection. The resulting minimal time flow schedule is shown in Fig. VII.6a. We find it useful to present the same flow schedule from a demand node point of view, i.e. with respect to the arrival time. This is done in Fig. VII.6b. We conclude the example with the observation that the existence of unsaturated links and/or loading constraints (cf. $f_{21}(1)$, $f_{22}(1)$, $f_{23}(2)$) in the optimal solution suggests the optimal delivery function as the final goal of optimization. We leave this point open for a future study. We will touch again upon this issue when we consider discrete time networks.

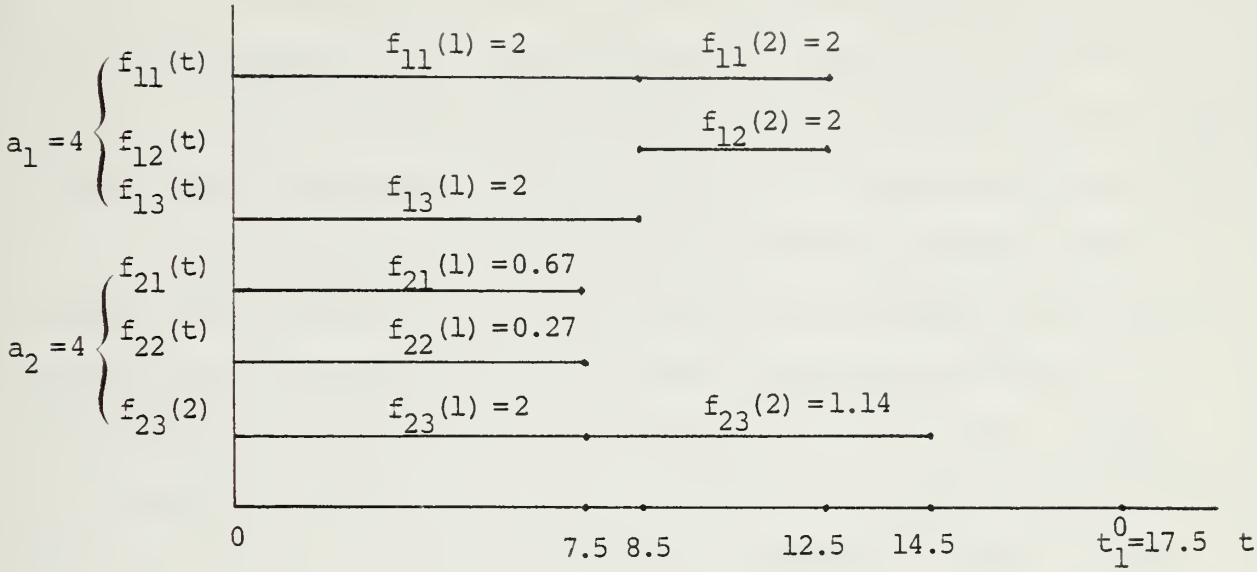


Fig. VII.6a. Minimal Time Flow Schedule (flow departure)

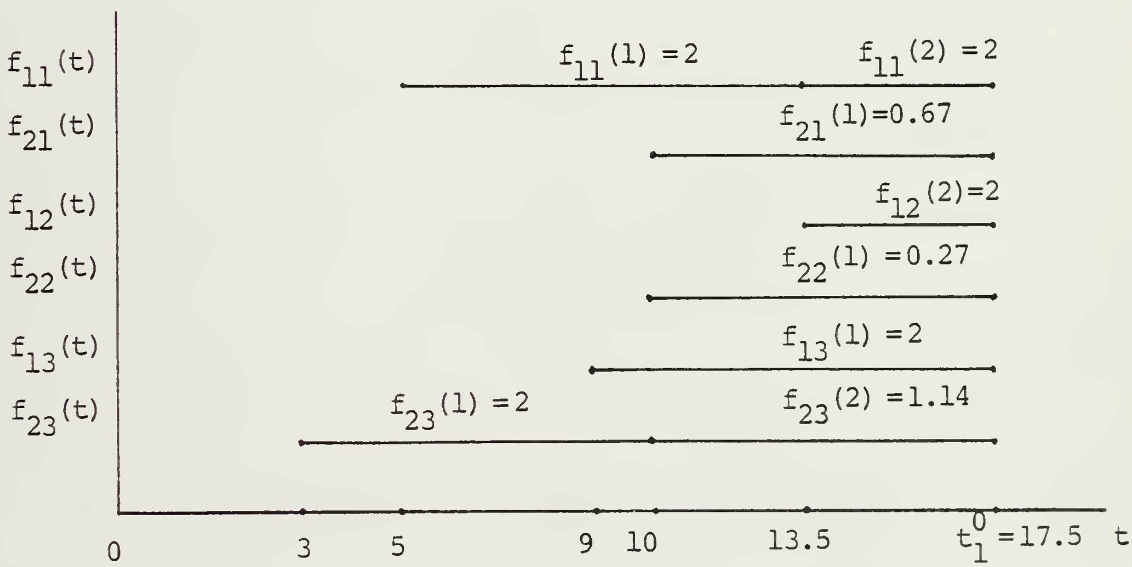


Fig. VII.6b. Minimal Time Flow Schedule (flow arrival)

5. Introducing Unloading Constraints

Up to this point we have considered the MTRP (on bipartite networks) with link and loading capacity constraints only. In this subsection we introduce the unloading constraints. We will use the network in Fig. VII.3 for illustration purposes.

The unloading constraint introduces linking of the flows with respect to their arrival time at demand node set D. The loading constraints had the same effect with respect to the departure time from supply node set S. Fig. VII.7 shows the unloading linkage of flows. (Note that from a demand node point of view, a flow on link [i,j] commences at time τ_{ij} and terminates at time t_1.)

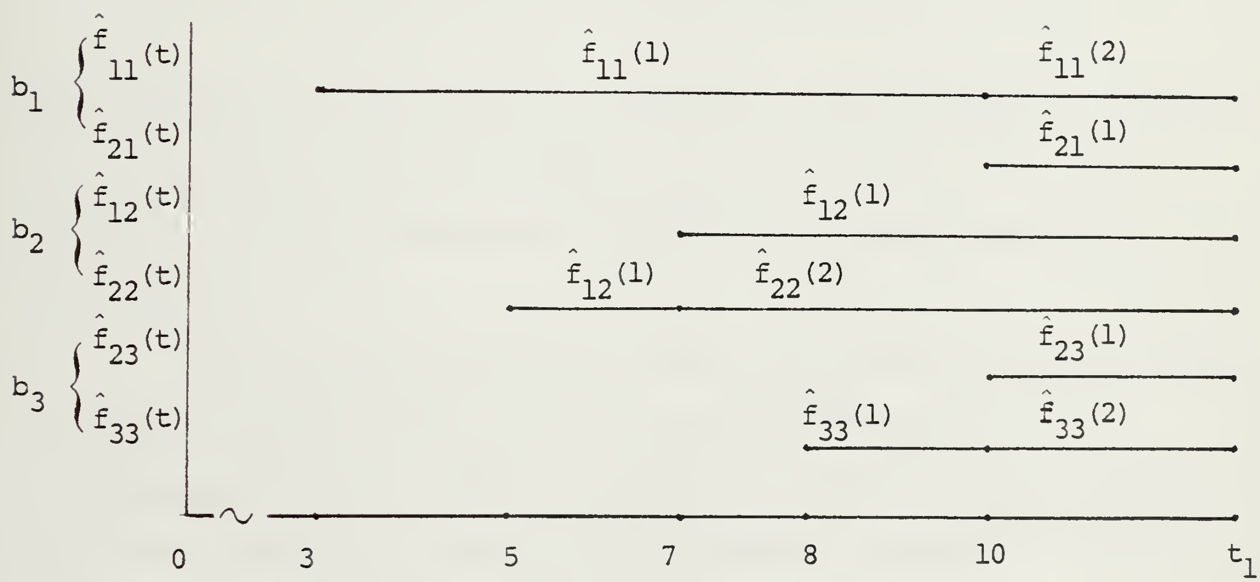
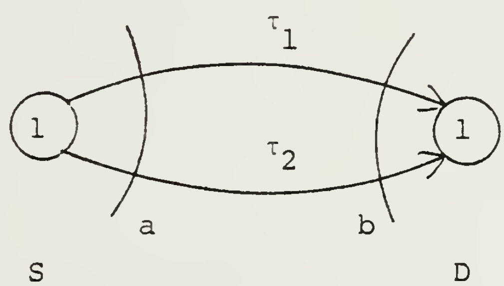


Fig. VII.7. Linking by Unloading Constraints

When both loading and unloading constraints are in effect, problem formulation is complicated by the requirement that the two grids of corner points, namely that in Fig. VII.4 and that in Fig. VII.7 must be consistent; if any flow variable changes value at a particular point of time (corner point), then all other flows linked to it (now by both constraints) must also be allowed to change. Consider for example the artificial network in Fig. VII.8.



$c_{ij} = c, \forall [i,j] \in L_0$

Fig VII.8. Network with Linking Constraints

The sequence of plots below illustrates the procedure of finding a consistent flow schedule. In each figure, the crosses identify already established corners, while the circles denote new ones.

After some thought the reader can convince himself that if the traversal times (and/or the solution time t_1^0) are incommensurate, the number of corner points would be infinite, in which case the solution is not piecewise constant. This difficulty may be overcome by approximating the continuous time axis by a finite sequence of discrete time values with some

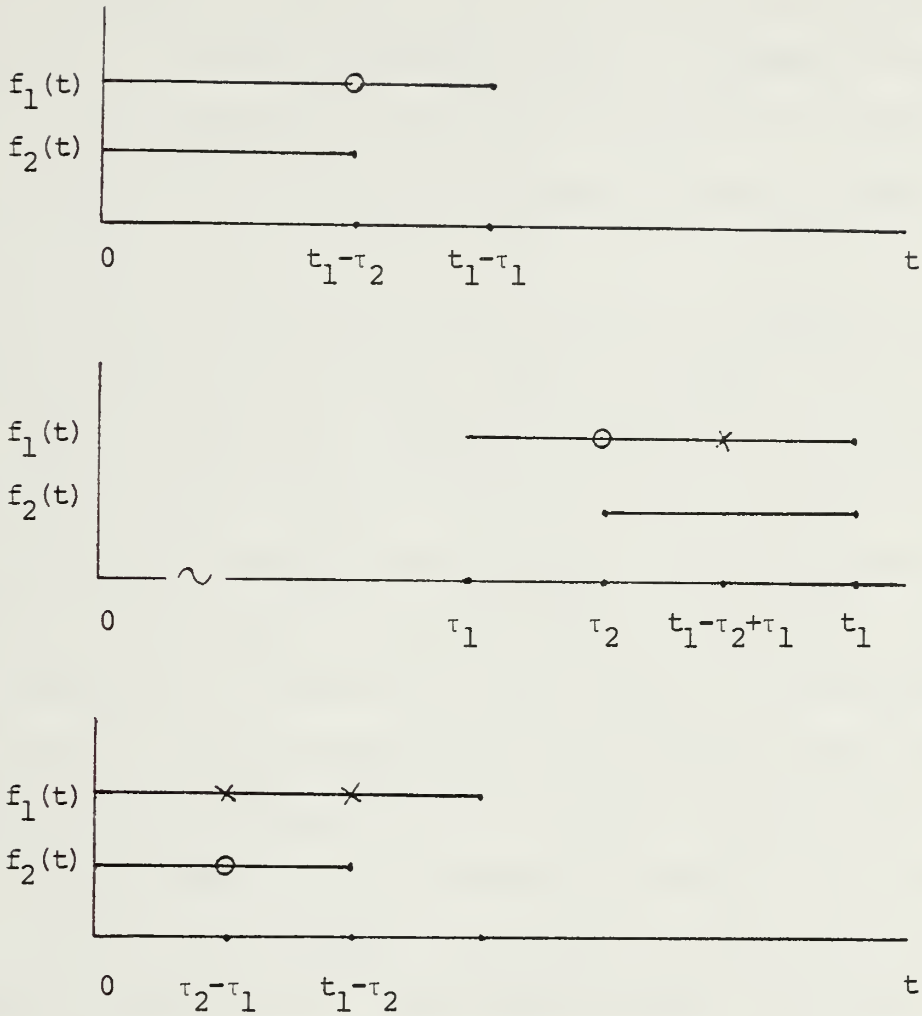


Fig. VII.9. Linking Flow Schedule

resolution Δt, $\Delta t > 0$. We note that the same reasoning applies to general networks with traversal delays and hence defer further discussion of discrete time networks to the next section, where we no longer restrict consideration to the bi-partite network.

D. DISCRETE TIME APPROXIMATION

In this section we briefly discuss a discrete time approximation of a minimal time redistribution problem. We have shown

already that this approximation arises naturally in the context of bi-partite networks with both loading and unloading constraints. The same reasoning applies to general networks. Hence we must assume that the corner points of a minimal time flow schedule may occur at any integer multiple of some basic unit of time Δt, $\Delta t > 0$.

It is worth mentioning that for general networks the number of corner points in a minimal time flow schedule can be very large, even if only link capacity constraints are present. This is so since a flow over any link in the network has to reflect all the possible time patterns of flow arrivals (intended for that link) to the head node of that link. Although this phenomenon is unrelated to the notion of commensurability its effect for large networks may be in practice the same.

We start by breaking a given interval of time $[0,T]$, where $T-\Delta t < t_1 \leq T$, into $k \triangleq T/\Delta t$ parts. A feasible flow schedule $F(t)$, $0 \leq t \leq t_1$ is assumed to be constant throughout each segment of duration Δt. Also, the traversal delays are approximated to the next nearest integer multiple of Δt. Thus from here on we will interpret τ_{ij} as the integer specifying the number of basic time units Δt which make up the traversal time from node i to node j. We use the following notation for the initial and desired states (i.e. distribution of assets) of a network:

$$Q_0 \triangleq (s_1, s_2, \ldots, s_n)$$

$$Q_1 \triangleq (d_1, d_2, \ldots, d_n)$$

and also the convention

$f_{ij}(r) \triangleq$ the amount of assets shipped out of node i, over a link [i,j] at time instant $r\Delta t$, and arriving at node j at time instant $(r+\tau_{ij})\Delta t$, $r = 0,1,\ldots,k$

$q_i(r) \triangleq$ the amount of assets stored at node i throughout the r-th interval, $r = 0,1,\ldots,k$.

We also observe that a link [i,j] can carry useful flows that were initiated up to and including the $(k-\tau_{ij})$-th interval. Any flow launched beyond this time will not reach node j by time $k\cdot\Delta t$.

We formulate now the minimal time redistribution problem as

MTRP: Find a minimal value of a time segment index k, say k^0, for which the optimal value of the cost function in the following linear program is zero.

LP: $\min \alpha$

s.t.

$$\alpha + q_i(k) = d_i, \quad \forall i \in V$$

$$q_i(0) + \sum_{j(\neq i)} f_{ij}(0) = s_i, \quad \forall i \in V$$

$$q_i(r) - q_i(r-1) + \sum_{j(\neq i)} f_{ij}(r) - \sum_{j(\neq i)} f_{ji}(r-\tau_{ji}) = 0, \quad \forall i \in V, \; r = 1,2,\ldots,k^{\dagger}$$

---

†For notational simplicity, here and in the sequel, we use $r = 0,1,\ldots,k$ although $f_{ji}(r-\tau_{ji}) \equiv 0$, unless $0 \leq r-\tau_{ji} \leq k-\tau_{ji}$, $\forall [i,j] \in L_0$.

$$\sum_{j(\neq i)} f_{ij}(r) \leq a_i, \; \forall i \in V, \; r = 0,1,\ldots,k$$

$$\sum_{j(\neq i)} f_{ji}(r-\tau_{ji}) \leq b_i, \forall i \in V, \; r = 0,1,\ldots,k \qquad \text{(VII.29)}$$

$$f_{ij}(r) \leq c_{ij}, \; \forall [i,j] \in L_0, \; r = 0,1,\ldots,k$$

$$\alpha, \; q_i(r), \; f_{ij}(r) \geq 0, \; \forall i \in V, \; \forall [i,j] \in L_0, \; r = 0,1,\ldots,k$$

where Q_0 and Q_1 are given.

□

It is not difficult to see that $k^0 \Delta t$ approximates the optimal value of the minimal time t_1^0, within one basic time interval. More precisely

$$(k^0-1)\Delta t < t_1^0 \leq k^0 \Delta t \, . \qquad \text{(VII.30)}$$

This is achieved of course at the expense of solving the LP in (VII.29) a number of times, for different values of the time segment index k, until the smallest value of k, say k^0, is discovered for which min $\alpha = 0$. Using a binary search technique (similar to (VII.28)) results in needing $0 \; (\lg_2 \frac{\Delta T}{\Delta t})$ steps, where ΔT is the size of the potential range of values of t_1^0.

The additional computational complexity that is introduced by this approximation scheme is counterbalanced, at least partially, by our ability to implement the notion of an optimal delivery function. We remind the reader that in the case of

exact solution (of a bi-partite network without unloading constraints) it was in general impossible to implement any requirements with respect to delivery time (the optimal delivery function and in particular the minimal rate problem fall within this category) while preserving a piecewise constant flow schedule.

The discrete time version of a delivery function (compare with (II.6)) is given by

$$D(r) \triangleq \sum_{i \in D} q_i(r), \quad r = 0,1,\ldots,k^0 \qquad \text{(VII.31)}$$

where the summation is over all nodes for which the demand is defined (demand set). The delivery function in (VII.31) was defined for flow schedules which do not allow intermediate storage or, in general, the amount of assets at any node i and any time t cannot exceed the demand d_i at that node. In the redistribution problem this assumption is not true anymore and a generalized formulation of delivery function is necessary. We define

$$D(r) \triangleq \sum_{i \in D} w_i(r), \quad r = 0,1,\ldots,k^0 \qquad \text{(VII.32)}$$

where

$$w_i(r) = \begin{cases} q_i(r), & q_i(r) < d_i \\ d_i, & \text{otherwise} \end{cases}$$

The formulation in (VII.32) accounts for the fact that a surplus delivery of assets (beyond the demand d_i) is not considered helpful and thus must be disregarded. The reader will notice that if $q_i(r) \leq d_i$, $\forall r$, as the case is in the delivery problem of Chapter II, Definition (VII.32) reduces to (VII.31).

Since the corner points of the minimal time schedule are fixed by the approximation method and the minimal time index solution k^0, the optimal (generalized) delivery function is defined as follows.

Definition VII.5

A delivery function $D^0(r)$, $r = 0,1,\ldots,k^0$ is said to be optimal if it satisfies

$$D^0(r) = \max_{F(k)} \{D(r) \mid D^0(k^0), D^0(k^0-1), \ldots, D^0(r+1)\}, \qquad \text{(VII.33)}$$

$$r = k^0-1,\ldots,0$$

where

$$D^0(k^0) \triangleq \sum_{i \in D} d_i$$

□

The sequential linear optimization procedure now consists of solving the minimal time problem (VII.29) followed by a sequence of k^0-1 LP problems. The formulation of the m-th, $m = k^0-1, k^0-2, \ldots, 1$ problem in that sequence is derived next.

We start by observing that

$$\max D(r) = \min\{\sum_{i \in D} d_i - D(r)\}$$

or equivalently

$$\max D(r) = \{\min \sum_{i \in D} h_i(r)\} \tag{VII.34a}$$

where

$$h_i(r) \triangleq d_i - w_i(r) = \begin{cases} d_i - q_i(r), & d_i > q_i(r) \\ 0, & \text{otherwise} \end{cases}$$

We can state now the m-th optimization problem as follows:

$$\min \sum_{i \in D} h_i(m)$$

s.t.

$$q_i(k^0) = d_i, \ \forall i \in D$$

$$q_i(0) + \sum_{j(\neq i)} f_{ij}(0) = s_i, \ \forall i \in V$$

$$q_i(r) - q_i(r-1) + \sum_{j(\neq i)} f_{ij}(r) - \sum_{j(\neq i)} f_{ji}(r-\tau_{ji}) = 0, \ \forall i \in V, \quad r = 1,2,\ldots,k^0 \dagger$$

$$\underline{h_i(r) + q_i(r)} \geq d_i, \ \forall i \in D \quad r = m, m+1, \ldots, k^0-1$$

---

†See (VII.29).

$$\sum_{i\in D} h_i(r) \quad = \quad \sum_{i\in D} d_i - D^0(r), \; r = m+1, m+2, \ldots, k^0$$

$$\sum_{j(\neq i)} f_{ij}(r) \quad \leq \quad a_i, \; \forall i \in V, \; r = 0,1,\ldots,k^0$$

$$\sum_{j(\neq i)} f_{ji}(r-\tau_{ji}) \quad \leq \quad b_i, \; \forall i \in V, \; r = 0,1,\ldots,k^0 \qquad \text{(VII.34b)}$$

$$f_{ij}(r) \quad \leq \quad c_{ij}, \; \forall [i,j] \in L_0, \; r = 0,1,\ldots,k^0$$

$$h_i(r), \; q_i(r), \; f_{ij}(r) \quad \geq \quad 0, \; \forall i \in V, \; \forall [i,j] \in L_0, \; r = 1,\ldots,k^0$$

where Q_0, Q_1, k^0 and $D^0(r)$, $r = k^0, k^0-1, \ldots, m+1$ are given.

□

It is important to notice that the definition of $h_i(r)$ in (VII.34a) is satisfied in the linear problem formulation (VII.34b). The underlined constraint and the form of the cost function in (VII.34b) provide for this fact. If $q_i(r) \geq d_i$, then the cost function will force the corresponding $h_i(r)$ to be equal to zero and when $q_i(r) < d_i$ then $h_i(r)$ is exactly the difference between the two values.

We use the following example to illustrate some of the ideas presented in this section.

Example:

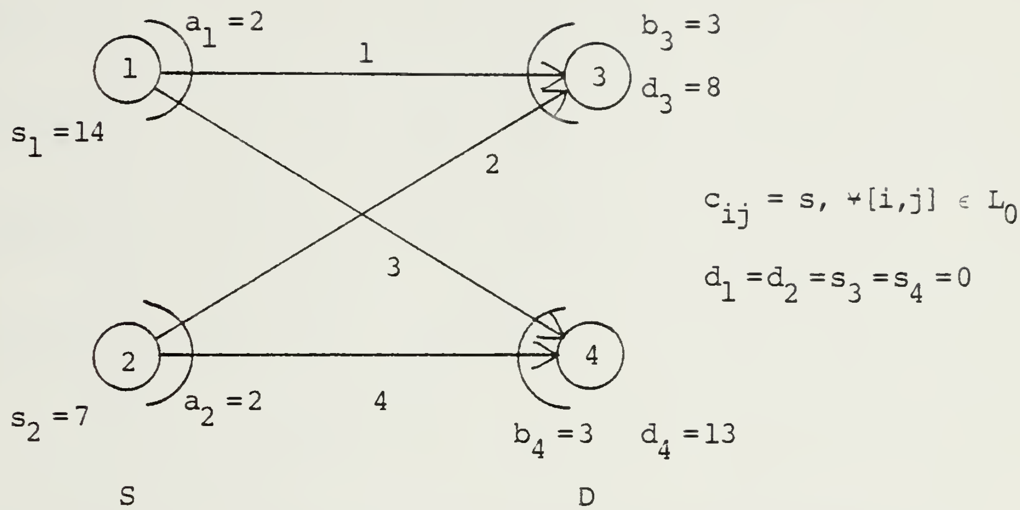


Fig. VII.10. Discrete Time Redistribution Problem

For the problem in Fig. VII.10 the traversal times (in units of Δt) are indicated on the corresponding links. The relevant loading and unloading constraints are shown per unit of time Δt. This problem was solved using the methodology presented earlier in this section. The resulting minimal time flow schedule is schematically illustrated in Fig. VII.11.

	Flow \ Shipment instance	0	1	2	3	4	5	6	7
$a_1=2$	$f_{13}(t)$		1	0	0	0	2	2	
	$f_{14}(t)$	2	1	2	2	2			
$a_2=2$	$f_{23}(t)$		1	1	1				
	$f_{24}(t)$	2	1	1					

Fig. VII.11. Minimal Time Flow Schedule (loading)

The same flow schedule is shown with respect to arrival time in Fig. VII.12.

	Flow \ Arrival instance	0	1	2	3	4	5	6	7 (k^0)	8	9
$b_3=3$	$f_{13}(t)$			1	0	0	0	2	2		
	$f_{23}(t)$				1	1	1				
$b_4=4$	$f_{14}(t)$				2	1	2	2	2		
	$f_{24}(t)$					2	1	1			
	ρ_m^0			1	3	4	4	5	4		

Fig. VII.12. Minimal Time Flow Schedule (unloading)

The optimal delivery function is shown in Fig. VII.13.

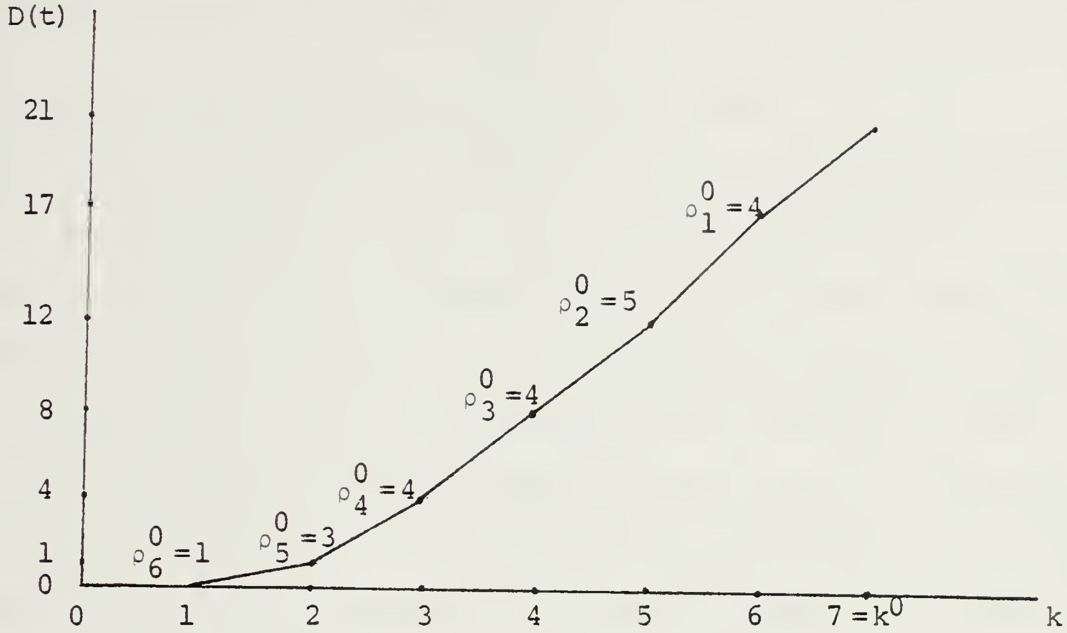


Fig. VII.13. Optimal Delivery Function

An observation with respect to Fig. VII.13 is appropriate here. In contradiction to the optimal delivery function studied in the main body of this thesis, here the objective function is no longer convex. This is still one more distinction between the delivery problem on networks without and with traversal delays.

We have mentioned in the introduction to this chapter that the minimal time redistribution problem has important application in military logistics planning. The SLO methodology and the results obtained thus far allow us to extend this

statement beyond the classical transportation problem, into a domain of complex decision making processes. We study one such application in the next section.

E. MAXIMALLY DELAYED DECISION PROBLEM (MDDP)

In this section we are concerned with a particular aspect of the decision making process in a military environment. Assume initially that a possible strategy to follow has been determined and that the problem of interest is the redistribution of available assets to implement this strategy. Since the various possible distributions of assets over a given set of locations constitute a state space of our model, any strategy implies one or more corresponding state trajectories (from initial to desired distribution) in that space. If there is a set of possible strategies under consideration, the ability to analyze their corresponding state trajectories can help identify which of these strategies is most attractive.

Consider a military environment characterized by a network $G = (V,L_0)$ and an initial distribution of assets $Q_0 = (s_1,...,s_n)$, and suppose there is a set $P = (p_1,p_2,...,p_M)$ of M strategies. Each strategy is described in turn by its corresponding terminal distribution of assets Q_m, $m = 1,2,...,M$. We assume, without loss of generality, that all the states Q_m, $m = 1,2,...,M$ are reachable from Q_0, i.e. there exists $F_m(k)$: $Q_0(0) \rightarrow Q_m(K)$ for some $K < \infty$, which is called the horizon time. The decision maker is faced with the problem of selecting the most desirable (in a military context) strategy from the set P. An inherent

property of most military situations is the necessity to make decisions subject to a high degree of uncertainty (as to the enemy's state, for example). It seems natural that the decision maker should seek to delay the decision process as much as possible, while still ensuring that all the terminal distributions $\{Q_m\}$ are attainable within the prescribed time limit K. By delaying the decision instant, a number of advantages are achieved, for example:

(i) minimize unnecessary, sometimes irreversible, commitments of assets.

(ii) maximize enemy's uncertainty as to which strategy is selected.

(iii) gain time to acquire additional information which may influence the decision process.

As a first step in delaying the decision instant by, say, k_1 units of time, we consider finding a common flow schedule $F^1(k)$, $0 \leq k \leq k_1$ which will transfer the system from its initial state $Q_0(0)$ to some intermediate state $Q^1(k_1)$. Obviously, the flow schedule $F^1(k)$, the state $Q^1(k_1)$ and the resulting decision delay of k_1 units make up an acceptable solution to the delayed decision problem iff all the terminal states Q_m, $m = 1,2,\ldots,M$ are reachable from $Q^1(k_1)$ within the remaining time $K-k_1$. This leads us to the statement of the First Maximally Delayed Decision Problem.

MMDP(1): Find the largest time index k_1, say k_1^0, for which the optimal value of the cost function in the following LP is zero.

LP: $$\min \alpha_1$$

s.t.

$$F^1(k): \quad Q_0(0) \rightarrow Q^1(k_1)$$

$$F_m(k): \quad Q^1(k_1) \rightarrow Q_m(K) - \underline{\alpha}_1, \; m = 1,2,\ldots,M \qquad \text{(VII.34)}$$

where $\underline{\alpha}_1 = (\alpha_1, \alpha_1, \ldots, \alpha_1)$ is an n component vector and $\alpha_1 \geq 0$.

□

To simplify the notation we have used vector formulation, where each of the constraints in (VII.34) represents a shorthand for all the constraints in LP(VII.29) with the distinction that $Q^1(k_1)$ here (which corresponds to $Q_1 = (d_1, d_2, \ldots, d_n)$ there) is unknown.

It is true that as k_1 increases to its maximal value k_1^0, some subset of trajectories connecting $Q^1(k_1^0)$ to their respective terminal states is bound to become critical, so that the corresponding terminal states become unreachable from $Q^1(k)$ for any $k > k_1^0$. We denote by P_1 the subset of strategies that become critical at k_1^0. The decision to select any one of these strategies has to be made prior to or at time $k_1^0 \Delta t$.

If $P_1 \subset P$, then we may follow the same line of reasoning as before to formulate a second optimization problem, which will let us identify the second critical set of strategies and their corresponding critical time.

MDDP(2): Find the largest time index k_2, say k_2^0, for which the optimal value of the cost function in the following LP is zero.

LP: $$\min \alpha_2$$

s.t.

$$F^1(k):\ Q_0(0) \rightarrow Q^1(k_1^0)$$

$$F_m(k):\ Q^1(k_1^0) \rightarrow Q_m(K),\ \forall m \in P_1$$

$$F^2(k):\ Q^1(k_1^0) \rightarrow Q^2(k_2) \qquad \text{(VII.35)}$$

$$F_m(k):\ Q^2(k_2) \rightarrow Q_m(K) - \underline{\alpha}_2,\ \forall m \notin P_1$$

for given k_1^0, where $\underline{\alpha}_2 = (\alpha_2,\alpha_2,\ldots,\alpha_2)$ and $\alpha_2 \geq 0$. □

Continuing to iterate in this way until we have exhausted all strategies, we will generate a finite number M_0, $M_0 \leq M-1$ of pairs (k_m^0, P_m), $m = 1,2,\ldots,M_0$ with the following properties

(i) $0 \leq k_1^0 < k_2^0 < \ldots < k_{M_0}^0 \leq K.$

(ii) $P_i \cap P_j = \emptyset,\ \ i,j = 1,2,\ldots,M_0,\ i \neq j.$ (VII.36)

(iii) $\bigcup_{i=1}^{M_0} P_i = P$

We see, as before, that the SLO methodology gives rise to a hierarchical structure (Fig. VII.14) in which the solution of

the individual optimization problems is carried out in a sequential manner, and that each such solution constrains the solution space of problems lower in the hierarchy.

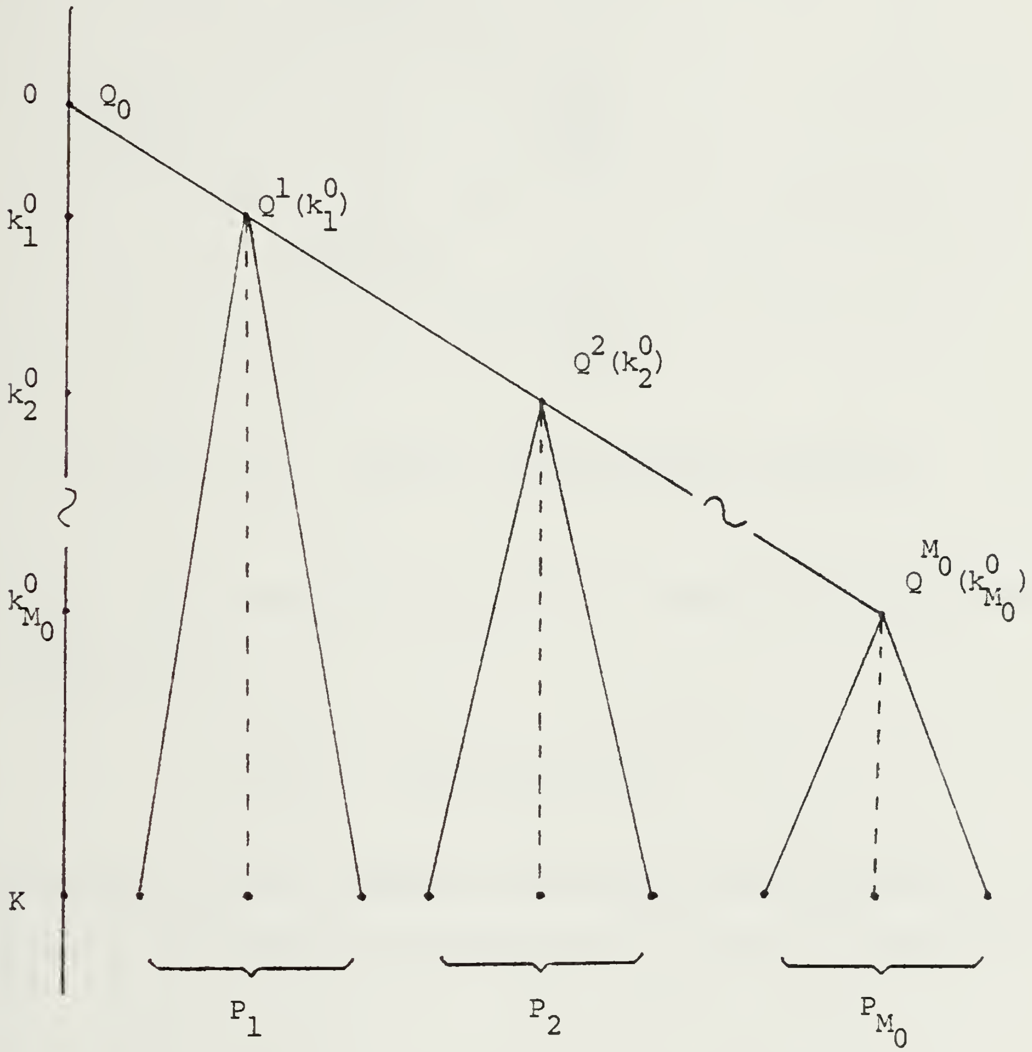
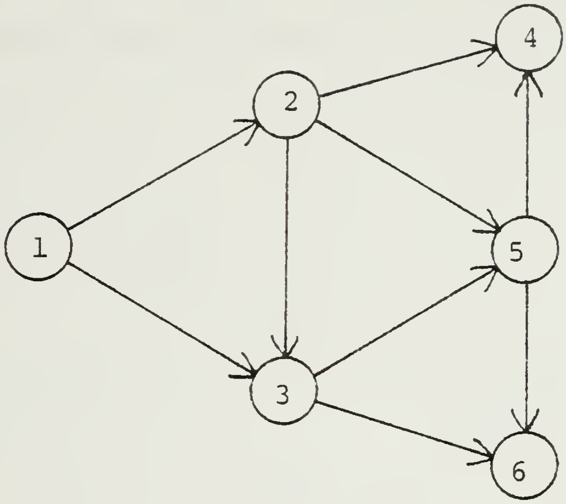


Fig. VII.14. Hierarchical Structure of the Strategy Sets w/r to Decision Instance

We conclude this section with an example of MDDP.

Example:



$c_{ij} = 2, \forall [i,j] \in L_0$

$\tau_{ij} = 1, \forall [i,j] \in L_0$

$a_i = b_i = \infty, \forall i \in V$

Fig. VII. 15. General Transportation Network

For the network in Fig. VII.15 the initial distribution of assets is

$$Q_0 = (20,0,0,0,0,0)$$

and there are two possible strategies characterized by their respective terminal distributions Q_1 and Q_2 at time $K\Delta t$, where $K = 10$.

$$Q_1 = (0,0,0,0,0,20), \quad Q_2 = (0,0,0,15,5,0)$$

The decision maker is faced with the problem of identifying a flow schedule which will enable him to delay as much as possible the final decision as to which of the strategies he is to follow.

The problem was solved using the methodology of this section. It was found that the latest instant of time by which a decision which strategy to follow has to be made is $k_1^0 \Delta t$, where $k_1^0 = 4$.

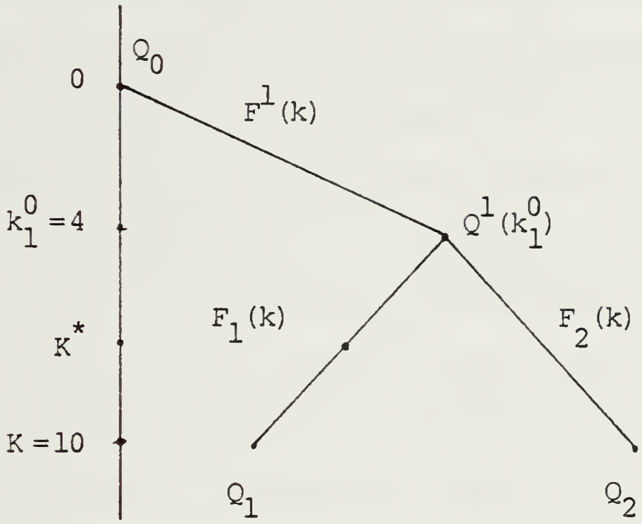


Fig. VII.16. Hierarchical Structure of the Decision Process

The first critical strategy set P_1 consists of strategy p_2. The fact that strategy p_1 is not critical (and the only one left) indicates that once it is decided to select p_1 at time $k_1^0 = 4$, it is possible to meet the desired distribution Q_1 prior to time K = 10, say at time K* (see Fig. VII.16). The appropriate minimal time flow schedule can be found by solving a minimal time redistribution problem with $Q_0 = Q^1(k_1^0)$ and Q_1. The complete flow schedule solution to the MDDP in Fig. VII.15 is given in Appendix H.

□

VIII. CONCLUSIONS

A. SUMMARY OF IMPORTANT RESULTS

In Chapter I we expressed the desire to obtain new insight into the problem of dynamic routing assignments. As an initial step toward achieving this goal a new performance measure for efficient delivery of backlogged data to their destinations is presented and called the optimal delivery function. The first important result is that the corresponding optimal flow schedule can be obtained by solving a hierarchical sequence of linear programming problems. Due to this fact the optimal delivery problem is computationally tractable even for moderately large networks. Furthermore, the well established results of linear programming are exploited to derive and understand the properties of an optimal flow schedule. Among the important results, at each hierachical level, the optimal flow schedule shows:

(i) A critical set of commodities that share a common delivery time.

(ii) A critical set of network links that must be saturated by those commodities throughout their delivery period.

(iii) The minimal rate with which saturation can be achieved.

Another important result relates, at each hierarchy level, the properties of an optimal flow schedule to the optimal dual variables associated with the corresponding linear program. In particular, the composition of the critical sets can be uniquely identified and the minimal rate of the saturation

flow can be calculated, whenever the optimal solution is stable.

The optimal delivery problem is then studied in the case of single destination networks. Two important results are obtained. The first states that the optimal delivery function is globally optimal, i.e. no flow schedule can deliver more commodity to the destination at any time, than the optimal flow schedule. The second result is the algorithm for solution of the single destination delivery problem. It is composed of a sequence of linear programs whose size is independent of the total number of hierarchy levels, which makes it computationally efficient. In this context a method for identifying critical sets, when the optimal solution is not stable, is presented.

An important result of this thesis is presented in the form of a new dynamic network analysis in which the optimal capacity assignment for routing purposes takes into account not only the backlogged messages but also the expected arrival rates of messages at the network. It is shown that the "minimal expected time to empty a queueing system" objective leads to a mathematical formulation which is identical to that of the deterministic delivery problem with constant flow inputs, and thus most of the previously derived results apply. It is also shown that a slight modification of the dynamic network analysis leads to a reasonable problem formulation which can be expected to provide a bound on the performance of more computationally

tractable routing procedures which might be implemented in a real network environment.

In the last chapter a number of results concerning the delivery problem in capacitated networks with traversal delays are obtained. In particular we mention the solution algorithm for the bi-partite network case.

B. AREAS OF FUTURE WORK

Many additional areas of research appear to be ready for further investigation.

One important point for future study concerns the efficiency of the proposed algorithm for solving the multicommodity delivery problem. In particular, decentralized computation capability should be investigated for its efficient implementation in dynamic routing schemes.

A further study of the delivery problem on general networks with traversal delays is needed. Here the question of computational tractability seems to be crucial. Our results for the bi-partite and general networks should provide an appropriate starting point for this effort.

Throughout the thesis we presented a number of conjectures. The questions of the number of corner points in the optimal multicommodity delivery function and its possible global optimality, in the case of intermediate storage flow schedules, remain open.

APPENDIX

A. PROOF OF THEOREM II.1

Thm. II.1 was stated in terms of flow schedules which satisfy constraints (II.1)-(II.3). We prove this theorem here for a wider class of flow schedules, namely those that allow intermediate data queuing. More precisely, we say that a multicommodity flow schedule is <u>feasible in a wide sense</u> (w/s) if it satisfies the following conditions:

$$\text{(i)} \quad f^k_{ij}(t) \geq 0, \ \forall [i,j] \in L_0 \ \text{ and } \ \forall t.$$

$$\text{(ii)} \quad q^k_i(t) \geq 0, \ \forall (i,k) \in N_0 \ \text{ and } \ \forall t. \tag{A.1}$$

$$\text{(iii)} \quad \sum_{k(\neq i)} f^k_{ij}(t) \leq c_{ij}, \ \forall [i,j] \in L_0 \ \text{ and } \ \forall t$$

where

$$q^k_i(t_1) = q^k_i(t_0) - \int_{t_0}^{t_1} \{ \sum_{j(\neq i)} f^k_{ij}(t) - \sum_{j(\neq i)} f^k_{ji}(t) \} dt.$$

The basic difference from Chapter II is that here the delivery rate of commodity k from node i at time t,

$$r^k_i(t) \triangleq \sum_{j(\neq i)} f^k_{ij}(t) - \sum_{j(\neq i)} f^k_{ji}(t) \tag{A.2}$$

is not restricted any more to be a non-negative quantity.

Theorem II.1 (Extended)

Let $Q(t_0)$ and $Q(t_1)$ be any two states such that $Q(t_1)$ is reachable from $Q(t_0)$ by some feasible (w/s) multicommodity flow schedule $F(t)$, $t_0 \leq t \leq t_1$. Then there exists a feasible (w/s) multicommodity constant flow schedule $F_1(t) = \hat{F}$, $t_0 \leq t \leq t_1$ such that $F_1(t)$: $Q(t_0) \rightarrow Q(t_1)$.

Proof:

Define

$$\hat{f}_{ij}^k \triangleq \frac{1}{t_1 - t_0} \int_{t_0}^{t_1} f_{ij}^k(t)dt, \quad \forall [i,j] \in L_0, \ \forall k \qquad \text{(A.3)}$$

If we start now in the state $Q(t_1)$ and apply the new flow schedule $\hat{F}$ for a duration $(t_1 - t_0)$ we will transfer the system into some new state, say $\hat{Q}(t_1)$.

$$\hat{q}_i^k(t_1) = q_i^k(t_0) - (t_1 - t_0)\{\sum_{j(\neq i)} \hat{f}_{ij}^k - \sum_{j(\neq i)} \hat{f}_{ji}^k\}, \qquad \text{(A.4a)}$$

$$\forall (i,k) \in N_0.$$

Substituting (A.3) into (A.4a) gives:

$$\hat{q}_i^k(t_1) = q_i^k(t_0) - \int_{t_0}^{t_1} \{\sum_{j(\neq i)} f_{ij}^k(t) - \sum_{j(\neq i)} f_{ji}^k(t)\}dt, \qquad \text{(A.4b)}$$

$$\forall (i,k) \in N_0$$

and from (A.1) we conclude that

$$\hat{q}_i^k(t_1) \equiv q_i^k(t_1), \quad \forall (i,k) \in N_0 \qquad \text{(A.5)}$$

Equation (A.5) is the desired result but we must still show that $\hat{F}$, as defined in (A.3), satisfies the constraints in (A.1):

(i) obvious, integration of a non-negative (by assumption) function results in a non-negative value.

(ii) From (A.4a) we have that for any $t \in [t_0, t_1]$

$$\hat{q}_i^k(t) = q_i^k(t_0) - (t-t_0)\hat{r}_i^k, \quad \forall (i,k) \in N_0 \tag{A.6}$$

where

$$\hat{r}_i^k \triangleq \sum_{j(\neq i)} \hat{f}_{ij}^k - \sum_{j(\neq i)} \hat{f}_{ji}^k, \quad \forall (i,k) \in N_0. \tag{A.7}$$

Since $\hat{r}_i^k$ is not a function of time we see (using (A.5) and (A.6)) that $q_i^k(t)$ falls on the line segment joining $q_i^k(t_0)$ and $q_i^k(t_1)$. But $q_i^k(t_0)$ and $q_i^k(t_1)$ are non-negative (by assumption) which leads to

$$\hat{q}_i^k(t) \geq 0, \quad \forall (i,k) \in N_0, \ \forall t \in [t_0, t_1]. \tag{A.8}$$

(iii) Define

$$\hat{f}_{ij} \triangleq \sum_{k(\neq i)} \hat{f}_{ij}^k \tag{A.9}$$

We need to show that $\hat{f}_{ij} \leq c_{ij}$, $\forall [i,j] \in L_0$. Using (A.3) we can write

$$\hat{f}_{ij} = \frac{1}{t_1 - t_0} \int_{t_0}^{t_1} \{ \sum_{k(\neq i)} f_{ij}^k(t) \} dt, \quad \forall [i,j] \in L_0 \tag{A.10}$$

By assumption

$$\max_{t\in[t_0,t_1]} \{ \sum_{k(\neq i)} f_{ij}^k(t)\} \leq c_{ij}, \quad \forall [i,j] \in L_0 \tag{A.11}$$

which results in

$$\hat{f}_{ij} \leq c_{ij}, \quad \forall [i,j] \in L_0 \tag{A.12}$$

and completes the proof.

□

We derive two corollaries of Thm. II.1.

Corollary II.1.1

In Thm. II.1 let $Q(t_1) = 0$. Then the constant flow schedule $F_1(t) = \hat{F}$, $t_0 \leq t \leq t_1$ is feasible in the narrow sense (n/s).

Proof:

Using equation (A.6) together with the requirement that $Q(t_1) = 0$, we have

$$0 = q_i^k(t_0) - (t_1 - t_0)\hat{r}_i^k, \quad \forall (i,k) \in N_0 \tag{A.13}$$

But $q_i^k(t_0) \geq 0$, $\forall (i,k) \in N_0$ and thus $\hat{r}_i^k \geq 0$, $\forall (i,k) \in N_0$ and there is no buildup of queues.

□

Corollary II.1.2.

In Thm. II.1 let $F(t)$, $t_0 \leq t \leq t_1$ be a feasible (n/s) multicommodity flow schedule. Then $F_1(t) = \hat{F}$, $t_0 \leq t \leq t_1$ is also feasible in a narrow sense.

Proof:

From equation(A.3) and the definition of a net delivery rate, we have that

$$\hat{r}_i^k = \frac{1}{t_1 - t_0} \int_{t_0}^{t_1} r_i^k(t)\,dt, \quad \forall (i,k) \in N_0 \qquad \text{(A.14)}$$

Obviously if $r_i^k(t) \geq 0$, $\forall t \in [t_0, t_1]$ and $\forall (i,k) \in N_0$, so is $\hat{r}_i^k$.

□

We prove now a result which is frequently used in our study.

Lemma II.1.1

Let $Q(t_0)$ and $Q(t_1)$ be any two states such that $Q(t_1)$ is reachable from $Q(t_0)$. Let $\{F(t)\}$, $t_0 \leq t \leq t_1$ be the set of all flow schedules such that $F(t)$: $Q(t_0) \rightarrow Q(t_1)$. Also, let $\{F_1(t)\}$, $t_0 \leq t \leq t_1$ be the set of all constant flow schedules such that $F_1(t)$: $Q(t_0) \rightarrow Q(t_1)$. Any link $[i,j] \in L_0$ that is saturated in all flow schedules in the set $\{F_1(t)\}$ is saturated for the whole period $[t_0, t_1]$ in all flow schedules in the set $\{F(t)\}$.

Proof:

Suppose that a link $[i,j] \in L_0$ is saturated in all the flow schedules in the set $\{F_1(t)\}$. Let there be some schedule $\hat{F}(t)$, $t_0 \leq t \leq t_1$ for which this link is not saturated for the whole period $[t_0, t_1]$. Then by Thm. II.1 and in particular (A.10) it is possible to construct a constant flow schedule $\hat{F}_1(t)$, $t_0 \leq t \leq t_1$ such that $F_1(t)$: $Q(t_0) \rightarrow Q(t_1)$, but for which

the aggregate flow $f_{ij} < c_{ij}$, so that the link is not saturated. This contradicts the definition of the set $\{F_1(t)\}$.

□

B. GLOBAL OPTIMALITY AND MULTICOMMODITY FLOW SCHEDULES

In this section we show by counterexample that the conjecture that every optimal delivery function is also globally optimal is false for the multicommodity case. The counterexample applies only to flow schedules which are feasible in the narrow sense,† i.e. do not allow intermediate data storage. It is still an open question whether the same conjecture holds for the more general class of flow schedules, namely those that allow for queues buildup.

Consider the delivery problem in Fig. B.1.

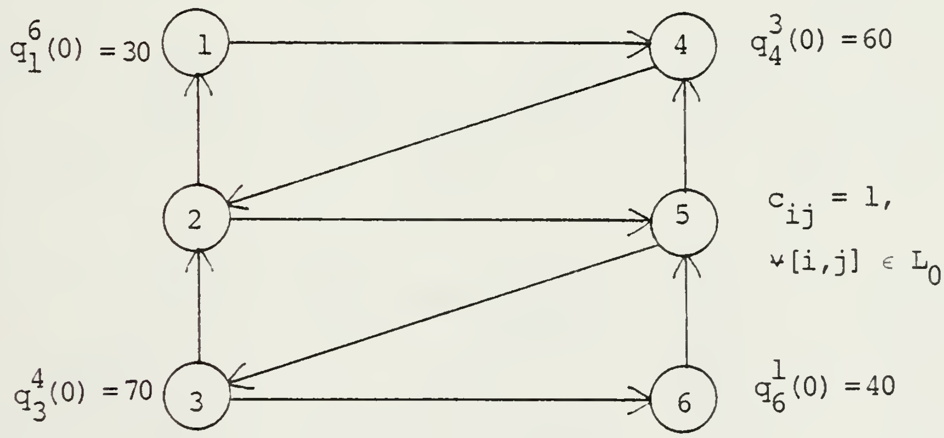


Fig. B.1. Delivery Problem

---

†For definition and related issues see Section A and Section C of the appendix.

Using the algorithm of Chapter III an optimal flow schedule solution is obtained. It is described in Fig. B.2.

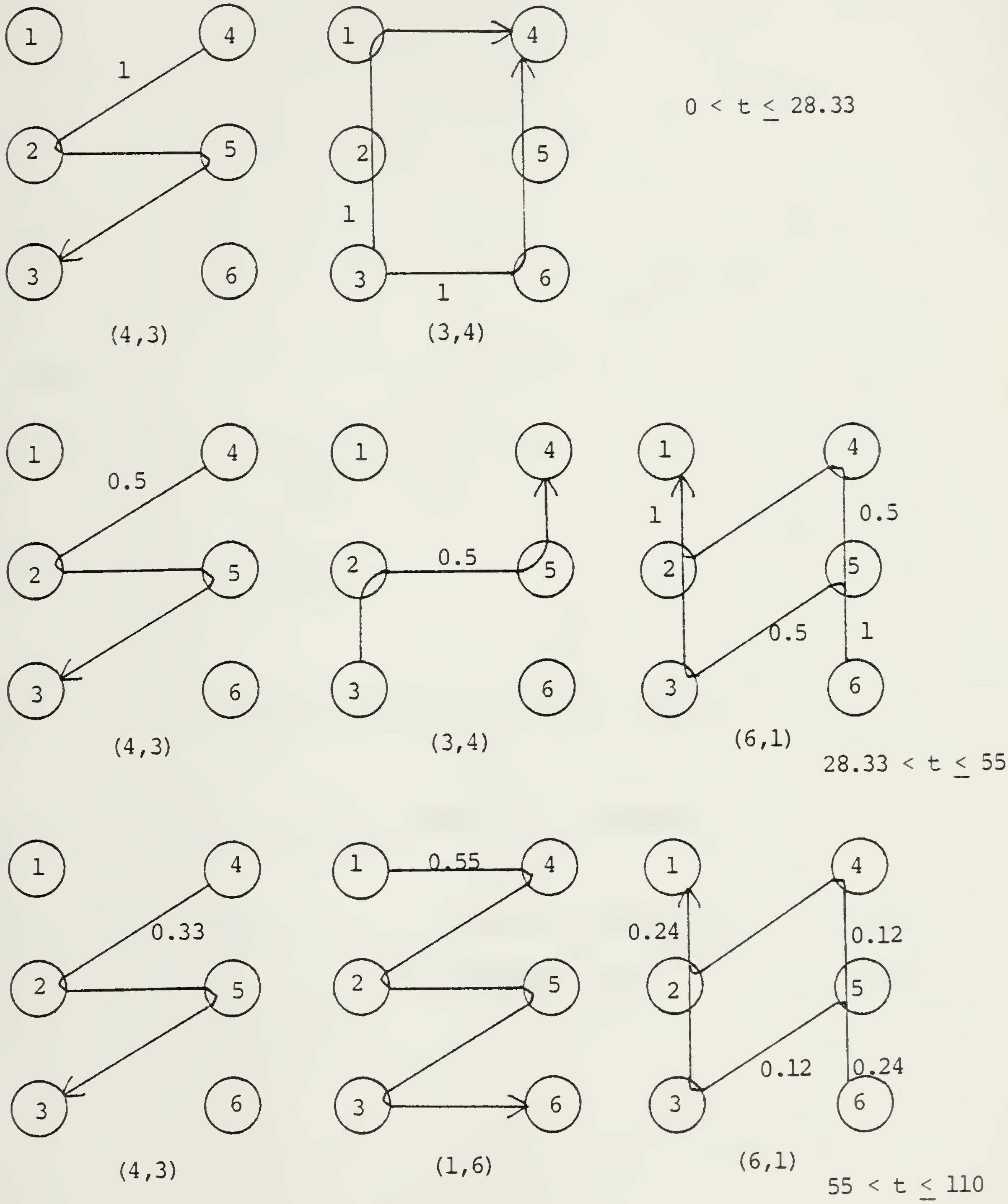


Fig. B.2. Chain Flow Decomposition of $F_3^0(t)$.

The optimal delivery function which is generated by this flow schedule is shown by the solid line in Fig. B.3.

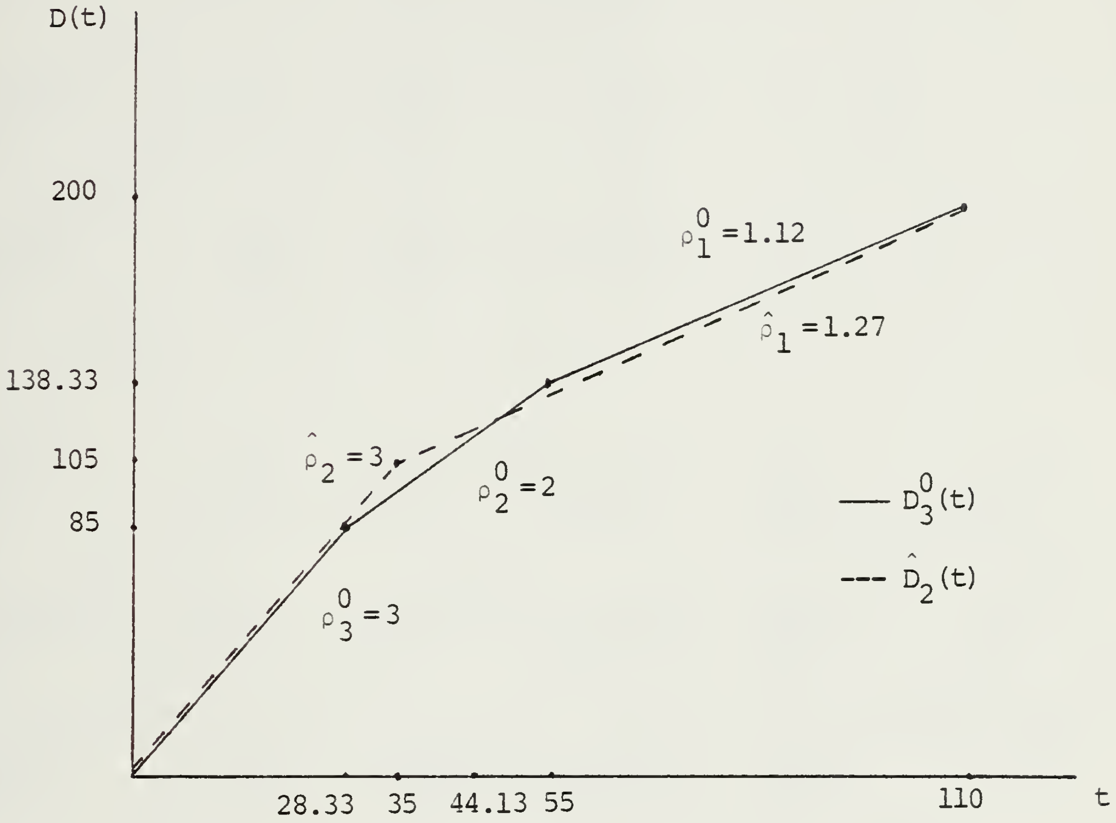


Fig. B.3. Optimal Delivery Function

The delivery function $\hat{D}_2(t)$ shown by the broken line in Fig. B.3 corresponds to the flow schedule $\hat{F}_2(t)$, $0 < t \leq t_1^0$ which is depicted in Fig. B.4.

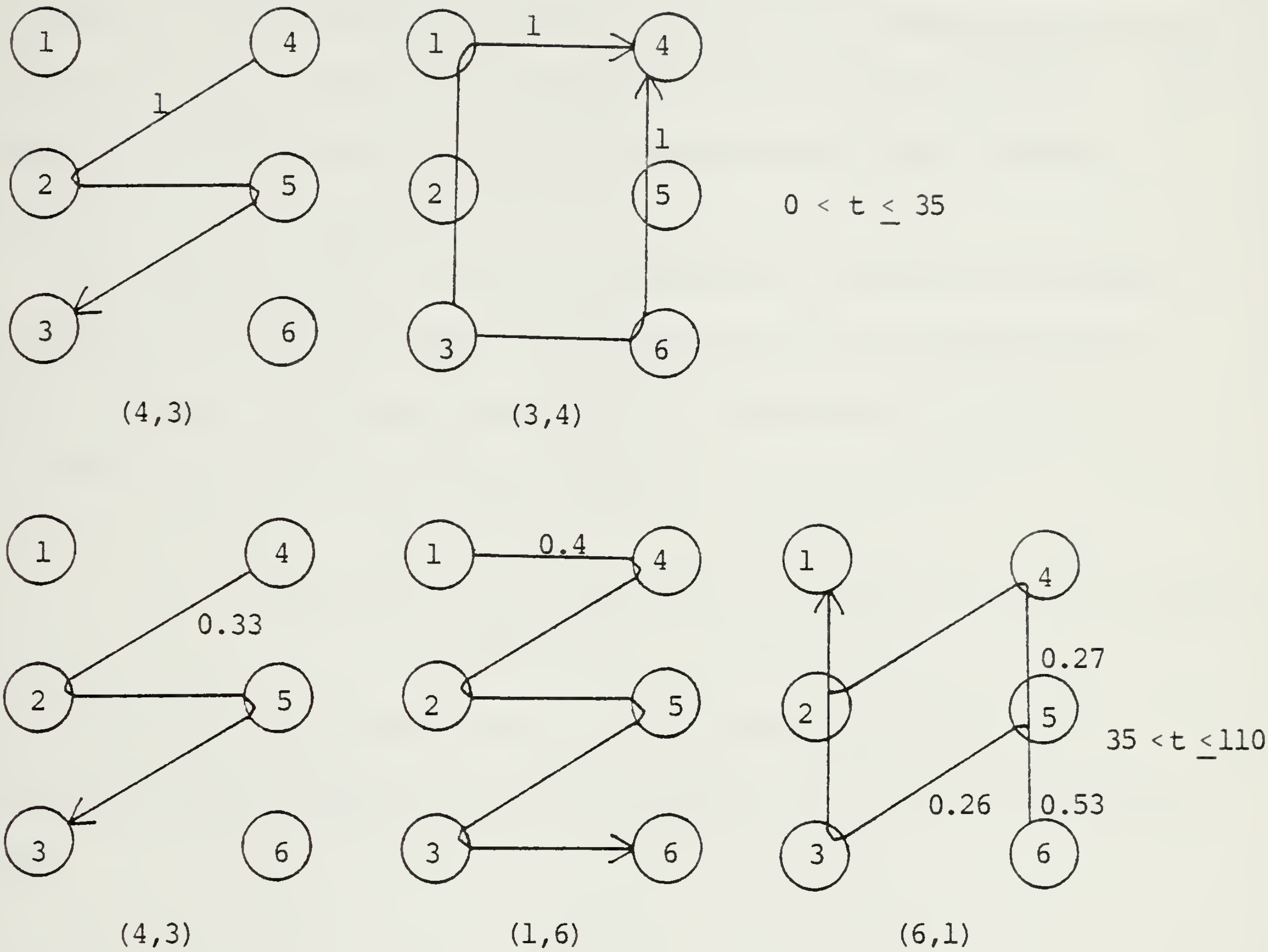


Fig. B.4. Chain Flow Decomposition of $\hat{F}_2(t)$

We have that

$$\hat{D}_2(t) \quad > \quad D_3^0(t), \ \forall t \in (28.33, 44.13) \qquad \text{(B.1)}$$

from which we conclude that not every optimal multicommodity flow schedule is also globally optimal.

C. INTERMEDIATE QUEUEING OF DATA

In this thesis most of the results concerning the deterministic delivery problem in communications networks are derived in terms of flow schedules that do not allow intermediate queueing of data (i.e. that are feasible in the narrow sense[†]). This approach avoids unnecessary complexity which could obscure insight into the problem.

For completeness, we now show that the solution algorithm in Chapter III can be easily modified to allow intermediate data storage. The implication of w/s feasibility (see (A.1)) is that the net delivery rate $r_i^k(t)$,

$$r_i^k(t) \triangleq \sum_{j(\neq i)} f_{ij}^k(t) - \sum_{j(\neq i)} f_{ji}^k(t) \not\geq 0 \qquad \text{(C.1)}$$

is not limited any more to be a non-negative quantity. As a consequence the conservation constraint (II.2) must be changed into

$$q_i^k(t) \triangleq q_i^k(0) - \int_0^t \{ \sum_{j(\neq i)} f_{ij}^k(\alpha) - \sum_{j(\neq i)} f_{ji}^k(\alpha)) d\alpha \geq 0, \qquad \text{(C.2)}$$

$$\forall (i,k) \in N_0, \ \forall t$$

in order to account for the non-negativity of all queues at all times.

Now, it is not difficult to see that the formulation of MTP(m) (see (III.4)) in terms of a w/s feasible flow has the following form.

[†]This section should be read after Section A of this appendix.

MTP(m)(w/s): $\min t_m$

s.t.

$$\sum_{p=1}^{m} \Big(\sum_{j(\neq i)} u_{ij}^{k}(p) - \sum_{j(\neq i)} u_{ji}^{k}(p) \Big) \leq q_i^k(0), \ \forall (i,k) \in N_0$$

$$\sum_{p=n}^{m} \Big(\sum_{j(\neq i)} u_{ij}^{k}(p) - \sum_{j(\neq i)} u_{ij}^{k}(p) \Big) \leq q_i^k(0), \ \forall (i,k) \in N_0, \quad n = m, m-1, \ldots, 2$$

$$-t_m c_{ij} + \sum_k u_{ij}^k(p) \leq 0, \ \forall [i,j] \in L_0$$

$$-t_{m-1} c_{ij} + \sum_k u_{ij}^k(m-1) \leq 0, \ \forall [i,j] \in L_0 \quad \text{(C.3a)}$$

$$\sum_k u^k{}_{ij}(p) \leq \Delta t_p^0 c_{ij}, \quad \forall [i,j] \in L_0, \quad p = 1,2,\ldots,m-2$$

$$-t_{m-1} \rho_{m-1}^0 + \sum_{(i,k)\in N_0} \Big(\sum_{j(\neq i)} u_{ij}^k(m-1) - \sum_{j(\neq i)} u_{ji}(m-1) \Big) = 0$$

$$\sum_{(i,k)\in N_0} \Big(\sum_{j(\neq i)} u_{ij}^k(p) - \sum_{j(\neq i)} u_{ji}(p) \Big) = \Delta t_p^0 \rho_p^0, \quad p = 1,2,\ldots,m-2$$

$$t_m + t_{m-1} = t_{m-1}^0$$

$$u_{ij}^k(p),\ t_m,\ t_{m-1} \geq 0, \ \forall (i,k) \in N_0, \ \forall [i,j] \in L_0, \ p = 1,2,\ldots,m$$

for given $t_1^0, \rho_1^0, \ldots, t_{m-1}^0, \rho_{m-1}^0$.

□

Inspection of (C.3) reveals that it differs from the original statement of MTP(m) in the form of one constraint only (underlined). A similar modification can be applied to MRP(m) (III.6) to allow for intermediate data storage.

MRP(m)(w/s): $\min \sum_{(i,k)} f_{ij}^k(m)$

s.t.

$$(t_m^0-\varepsilon)r_i^k(m+1) + \varepsilon r_i^k(m) + \sum_{p=1}^{m-1} \Delta t_p^0 r_i^k(p) = q_i^k(0), \forall (i,k) \in N_0$$

$$(t_m^0-\varepsilon)r_i^k(m+1) \leq q_i^k(0), \forall (i,k) \in N_0$$

$$(t_m^0-\varepsilon)r_i^k(m+1) + \varepsilon r_i^k(m) \leq q_i^k(0), \forall (i,k) \in N_0 \qquad \text{(C.3b)}$$

$$(t_m^0-\varepsilon)r_i^k(m+1) + \varepsilon r_i^k(m) + \sum_{p=n}^{m-1} \Delta t_p^0 r_i^k(p) \leq q_i^k(0), \forall (i,k) \in N_0, \quad n = m-1, m-2, \ldots, 2$$

$$\sum_k f_{ij}^k(p) \leq c_{ij}, \forall [i,j] \in L_0, \quad p = 1,2,\ldots,m+1$$

$$\sum_{(i,k)\in N_0} r_i^k(p) = \rho_p^0, \; p = 1,2,\ldots,m-1$$

$$f_{ij}^k(p), r_i^k(p) \geq 0, \forall (i,k) \in N_0, \forall [i,j] \in L_0, \quad p = 1,2,\ldots,m+1.$$

for given $t_1^0, \rho_1^0, \ldots, \rho_{m-1}^0, t_m^0, \varepsilon$

$\square$

where

$$r_i^k(p) = \sum_{j(\neq i)} f_{ij}^k(p) - \sum_{j(\neq i)} f_{ji}^k(p), \; \forall (i,k) \in N_0, \; p = 1,2,\ldots,m+1$$

$$t_p^0 = t_p^0 - t_{p+1}^0, \quad p = 1,2,\ldots,m-1$$

and

ε is any real number such that $0 < \varepsilon < t_m^0 - t_{m+1}^0$.

The change in the form of the cost function is necessary to ensure that we minimize delivery rate and not the total flow rate in the network.

We conclude this section with an example of a simple delivery problem for which we compare the delivery functions resulting from both types of flow schedules.

Example:

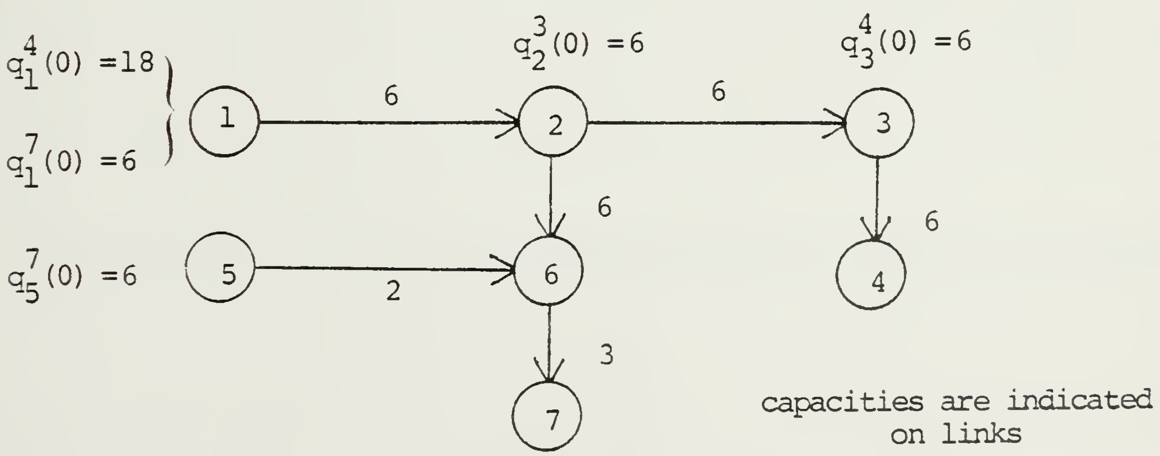


Fig. C.1. Delivery Problem

An optimal flow schedule solution for the problem in Fig. C.1 is shown in Figs. C.2a and C.2b.

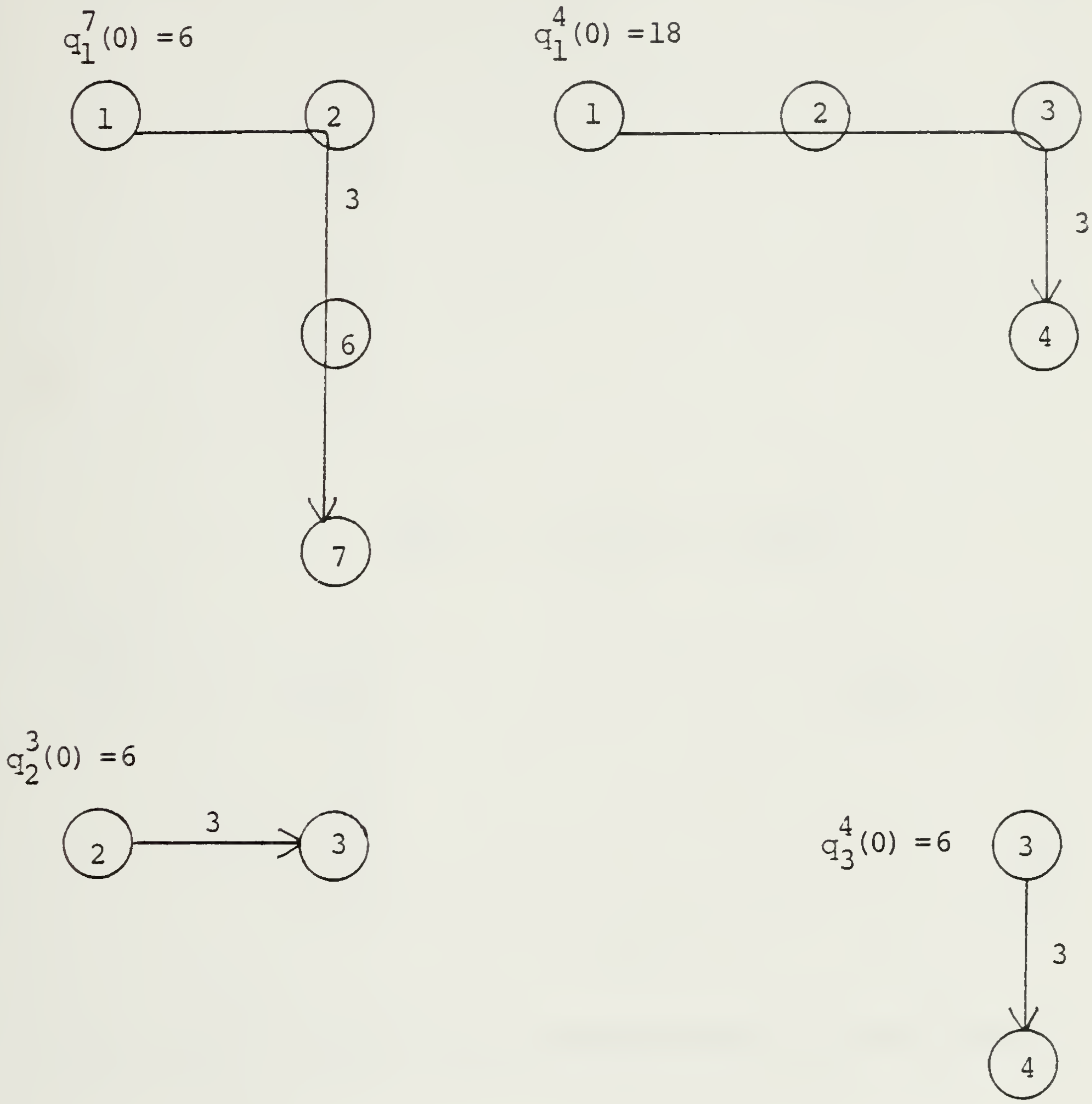


Fig. C.2a. Chain Flow Decomposition for the Period [0,1]

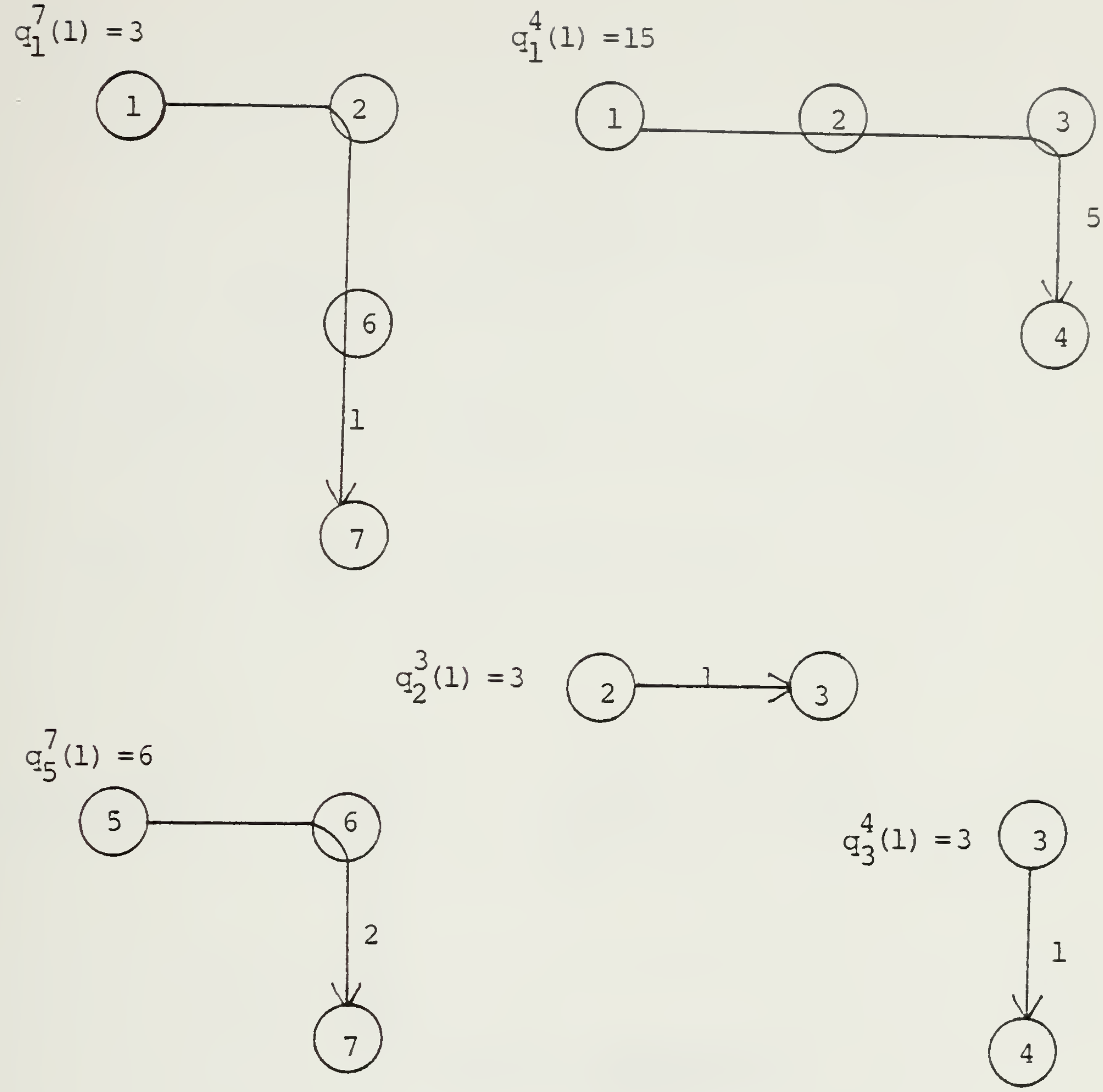


Fig. C.2b. Chain Flow Decomposition for the Period (1,4]

The resulting optimal delivery function is shown in Fig. C.3.

Suppose now that we allow for intermediate data queueing and in particular consider the flow schedule in Figs. C.4a and C.4b. Observe that in the first period, [0,1], commodity (1,7) flows with rate six from node 1 to node 6 and continues

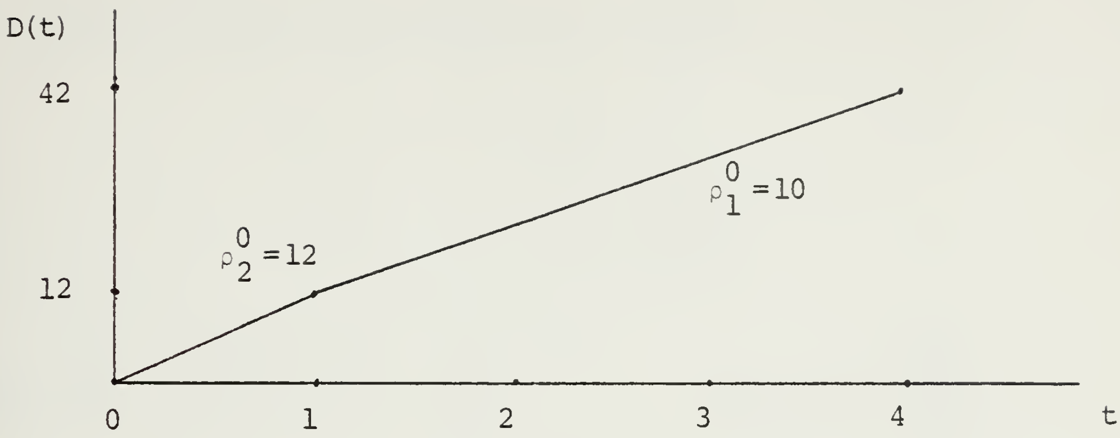


Fig. C.3. Optimal Delivery Function

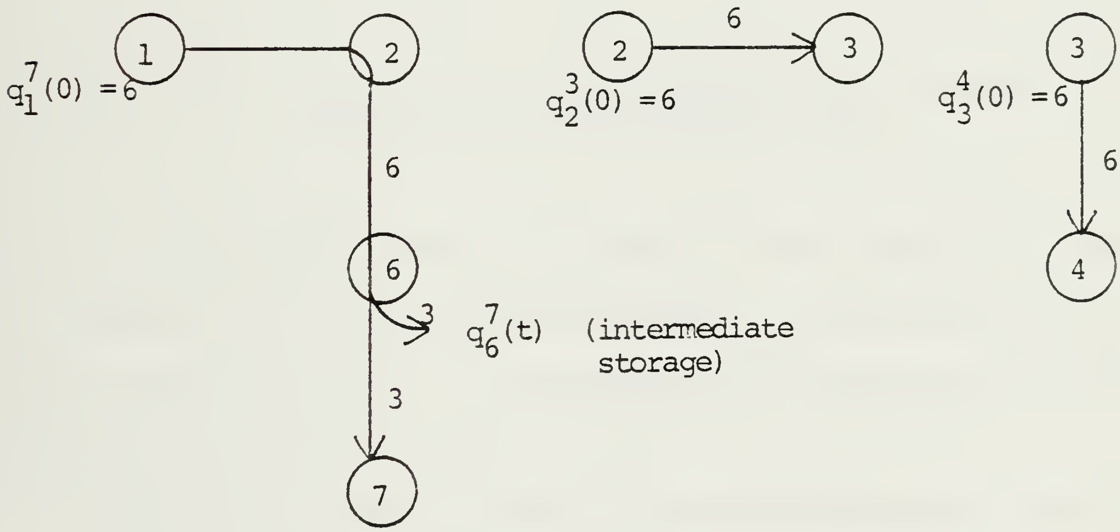


Fig. C.4a. Chain Flow Decomposition with Intermediate Queueing for the Period [0,1]

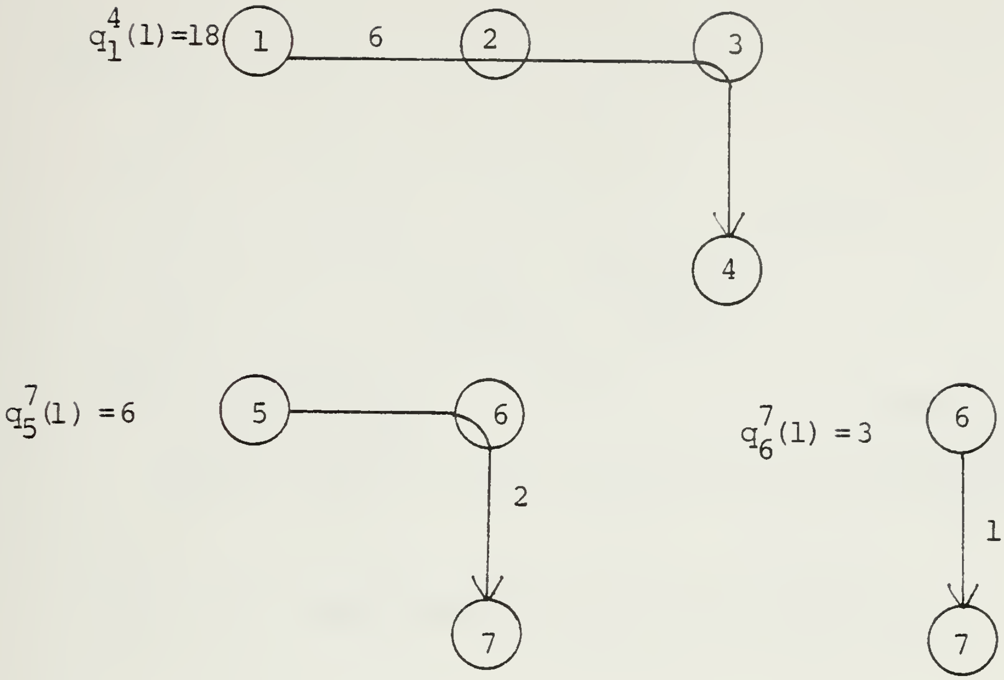


Fig. C.4b. Chain Flow Decomposition with Intermediate Queueing for the Period (1,4]

with rate three from node 6 to node 7. As a result, three units of commodity (1,7) are stored in node 6 in an intermediate queue denoted by $q_6^7(t)$. The contents of this queue are delivered in the second period (1,4].

The delivery function which is generated by the new flow schedule is superior (dominates) the optimal delivery function. Both delivery functions are displayed in Fig. C.5.

We conclude that flow schedules which allow intermediate data queueing may have advantages in certain instances over flow schedules which are feasible in the narrow sense. Nevertheless, it seems clear that wide-sense (w/s) feasible optimal

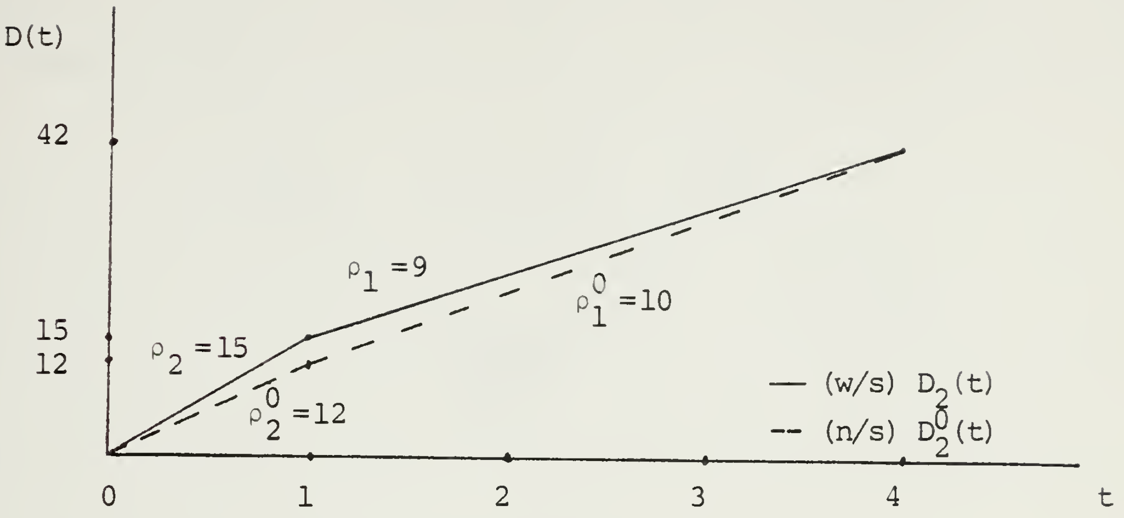


Fig. C.5. Comparison of Delivery Functions

flow schedules (which allow for intermediate data queueing) do not produce delivery functions that differ substantially in their basic characteristic (piecewise linear, convex, etc.) from those produced by narrow sense (n/s) optimal flow schedules.

D. MORE ON STABILITY

In this section we complete the discussion of stability (from Chapter IV.A.c) with examples of delivery problems that illustrate this concept.

1. Unstable Delivery Problem

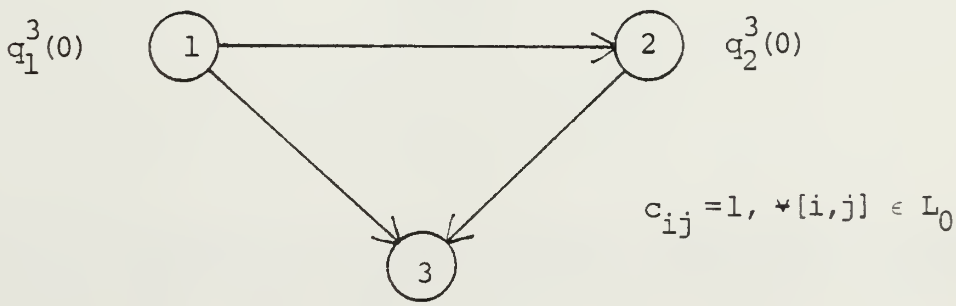


Fig. D.1. Delivery Problem

For the case when $q_1^3(0) = q_2^3(0) = q$, an optimal constant flow schedule solution can be found by inspection. It is shown in Fig. D.2.

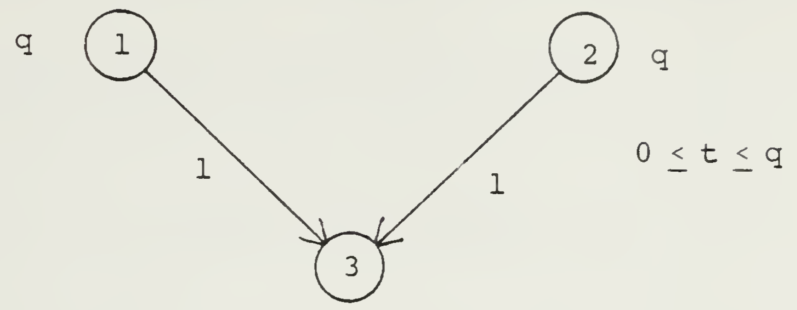


Fig. D.2. Optimal Constant Flow Schedule

Formally, the MTP(1) for this problem can be written as:

MTP(1)[†]: $\min t_1$

s.t.

$$\begin{vmatrix} 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 & 0 & 0 & 1 \end{vmatrix} \begin{vmatrix} t_1 \\ u_{12} \\ u_{23} \\ u_{13} \\ s_{12} \\ s_{23} \\ s_{13} \end{vmatrix} = \begin{vmatrix} q \\ q \\ 0 \\ 0 \\ 0 \end{vmatrix} \qquad (D.1)$$

$$(t_1, u_{12}, u_{23}, u_{13}, s_{12}, s_{23}, s_{13}) \geq \underline{0}$$

[†]Since both commodities have a common destination (node 3), we will not use the upper index notation to indicate the destination, i.e. $u_{ij}^k \to u_{ij}$, $\forall [i,j] \in L_0$.

An optimal basic solution X_B[†] is found to be

$$X_B^T = (t_1, u_{23}, u_{13}, s_{12}, s_{13}) = (q,q,q,q,0), \qquad (D.2)$$

and the related optimal dual solution is

$$\Lambda = \Gamma_B \cdot B^{-1} = (0,1,0,0,-1) \qquad (D.3)$$

where

$$\Gamma_B = (1,0,0,0,0,0,0).$$

Now, suppose that $q_2^3(0) = q-\varepsilon$, $\varepsilon > 0$. It is not difficult to see that in the new optimal flow schedule, commodity (1,3) must use the link chain ([1,2],[2,3]) in addition to link [1,3]. Therefore it has the form shown in Fig. D.3. Also,

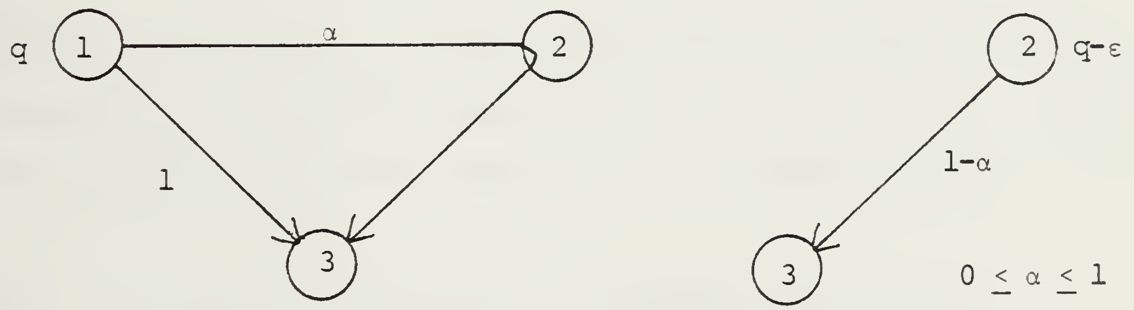


Fig. D.3. Structure of the Optimal Perturbed Flow Schedule

[†]Recall that the flow rate variables f_{ij} are related to the u_{ij}-variables by the transformation

$$f_{ij} = u_{ij}/t_1^0, \quad \forall [i,j] \in L_0.$$

it must be true that

$$\hat{t}_1^0 = \frac{q}{1+\alpha} = \frac{q-\varepsilon}{1-\alpha} =^{\dagger} \frac{2q-\varepsilon}{2} = q - \frac{\varepsilon}{2} \tag{D.4}$$

We can now calculate α from (D.4):

$$\alpha = \frac{q}{q-\frac{\varepsilon}{2}} - 1 = \frac{\frac{\varepsilon}{2}}{q-\frac{\varepsilon}{2}} \tag{D.5}$$

The change Δz in the optimal value of the cost function that was caused by the perturbation is

$$\Delta z = t_1^0 - \hat{t}_1^0 = q - (q-\frac{\varepsilon}{2}) = \frac{\varepsilon}{2} \tag{D.6}$$

If we use the optimal dual solution (D.3) to evaluate Δz, we obtain

$$\Delta z = \Lambda\cdot\Delta b = (0,1,0,0,-1)\cdot(0, \ ,0,0,0)^T = \varepsilon, \tag{D.7}$$

which is incorrect. We conclude that the original optimal solution is not stable (optimal dual is not unique). Actually, the new optimal basic solution is (cf. Figs. D.3 and D.5)

$$\hat{X}_B^T = (t_1,u_{12},u_{23},u_{13},s_{12}) = (q-\frac{\varepsilon}{2},\frac{\varepsilon}{2},q-\frac{\varepsilon}{2},q-\frac{\varepsilon}{2},q-\varepsilon) \tag{D.8}$$

---

† Using the law of proportions,

$$\frac{a}{b} = \frac{c}{d} = \frac{a+c}{b+d}.$$

and the new related optimal dual is

$$\hat{\Lambda} = \hat{\Gamma}_B \hat{B}^{-1} = (0.5, 0.5, 0, -0.5, -0.5), \qquad \text{(D.9)}$$

where

$$\hat{\Gamma}_B \equiv \Gamma_B.$$

Obviously,

$$\hat{\Lambda} \neq \Lambda. \qquad \text{(D.10)}$$

2. Stable Delivery Problem

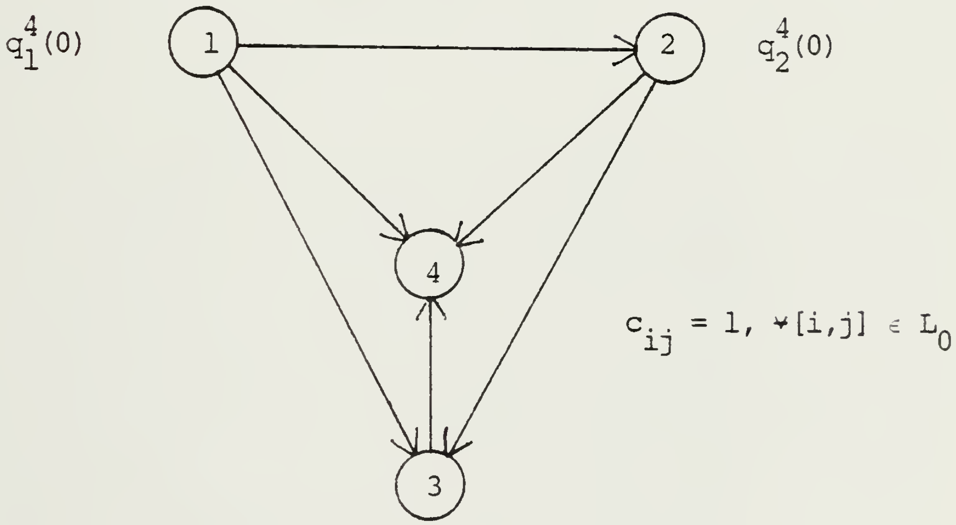


Fig. D.4. Delivery Problem

For the case when $q_1^4(0) = 2q$ and $q_2^4(0) = q$, an optimal constant flow schedule solution can be found by inspection. It is shown in Fig. D.5.

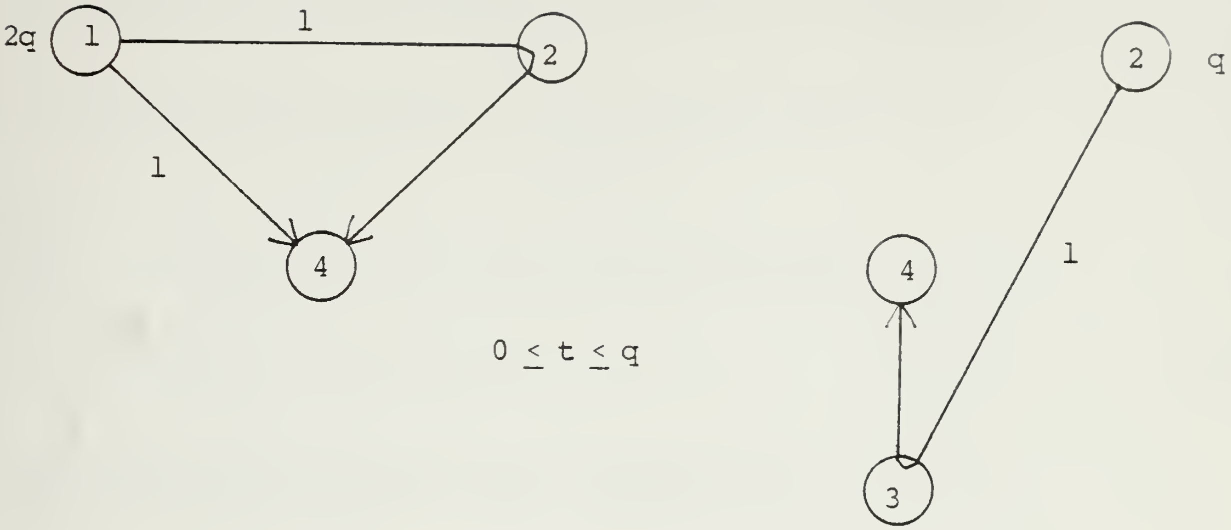


Fig. D.5. Optimal Constant Flow Schedule

Formally, the MTP(1) for this problem can be written as:

MTP(1): $\min t_1$

s.t.

$$\begin{vmatrix} 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{vmatrix} \begin{vmatrix} t_1 \\ u_{12} \\ u_{13} \\ u_{23} \\ u_{14} \\ u_{24} \\ u_{34} \\ s_{12} \\ s_{13} \\ s_{23} \\ s_{14} \\ s_{24} \\ s_{34} \end{vmatrix} = \begin{vmatrix} 2q \\ q \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{vmatrix} \quad \text{(D.11)}$$

$(t_1, u_{12}, u_{13}, u_{23}, u_{14}, u_{24}, u_{34}, s_{12}, s_{13}, s_{23}, s_{14}, s_{24}, s_{34}) \geq \underline{0}$

An optimal basic solution X_B is found to be

$$X_B^T = (t_1, u_{12}, u_{13}, u_{23}, u_{14}, u_{24}, u_{34}, s_{12}, s_{13})$$

$$= (q, q, 0, q, q, q, q, 0, q), \qquad \text{(D.12)}$$

and the related optimal dual solution is

$$\Lambda = \Gamma_B B^{-1} = (\tfrac{1}{3}, \tfrac{1}{3}, \tfrac{1}{3}, 0, 0, 0, -\tfrac{1}{3}, -\tfrac{1}{3}, -\tfrac{1}{3}), \qquad \text{(D.13)}$$

where

$$\Gamma_B = (1, 0, \ldots, 0).$$

Now, suppose that $q_1^4(0) = 2q+\varepsilon$, $\varepsilon > 0$. With a little thought the reader will convince himself that commodity (1,4) must use link chain ([1,3],[3,4]) in addition to ([1,4]) and ([1,2],[2,4]). Therefore an optimal solution has the form shown in Fig. D.6.

Also, it must be true that

$$\hat{t}_1^0 = \frac{2q+\varepsilon}{2+\alpha} = \frac{q}{1-\alpha} = \frac{3q+\varepsilon}{3} = q + \frac{\varepsilon}{3} \qquad \text{(D.14)}$$

We can now calculate α from (D.14)

$$\alpha = 1 - \frac{q}{q + \frac{\varepsilon}{3}} = \frac{\frac{\varepsilon}{3}}{q + \frac{\varepsilon}{3}} \qquad \text{(D.15)}$$

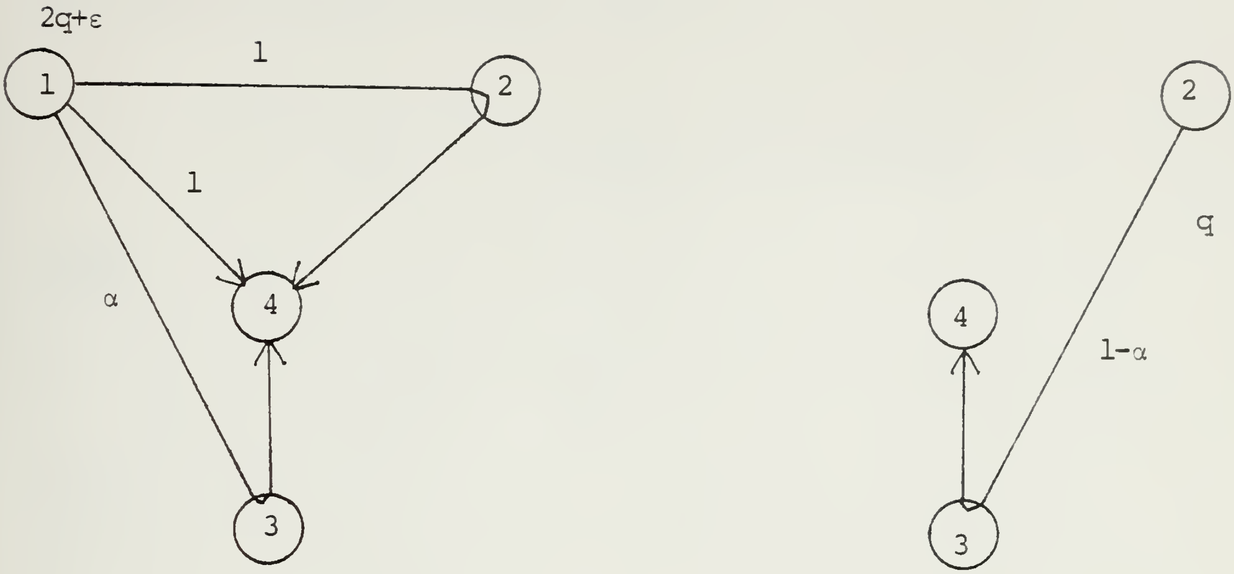


Fig. D.6. Structure of the Optimal Perturbed Flow Schedule

The change Δz in the optimal value of the cost function that was caused by the perturbation is

$$\Delta z \;=\; t_1^0 - \hat{t}_1^0 \;=\; q - (q + \tfrac{\varepsilon}{3}) \;=\; -\tfrac{\varepsilon}{3} \qquad \text{(D.16)}$$

If we use the optimal dual solution (D.14) to evaluate Δz, we obtain

$$\Delta z \;=\; \Lambda \cdot \Delta b \;=\; (\tfrac{1}{3}, \tfrac{1}{3}, \tfrac{1}{3}, 0, 0, 0, -\tfrac{1}{3}, -\tfrac{1}{3}, -\tfrac{1}{3})(-\varepsilon, 0, \ldots,)^T$$

$$= \; -\tfrac{\varepsilon}{3}, \qquad \text{(D.17)}$$

which is the correct value. Actually, the new basic solution is (cf. Fig. D.6 and D.15)

$$\hat{X}_B^T \;=\; (t_1, u_{12}, u_{13}, u_{23}, u_{14}, u_{24}, u_{34}, s_{13}, s_{23})$$

$$= \; (q+\tfrac{\varepsilon}{3}, q+\tfrac{\varepsilon}{3}, \tfrac{\varepsilon}{3}, q, q+\tfrac{\varepsilon}{3}, q+\tfrac{\varepsilon}{3}, q+\tfrac{\varepsilon}{3}, q, \tfrac{\varepsilon}{3}), \qquad \text{(D.18)}$$

and the new related optimal dual is

$$\hat{\Lambda} = \hat{\Gamma}_B \hat{B}^{-1} = (\tfrac{1}{3},\tfrac{1}{3},\tfrac{1}{3},0,0,0,-\tfrac{1}{3},-\tfrac{1}{3},-\tfrac{1}{3}), \qquad \text{(D.19)}$$

where

$$\hat{\Gamma}_B = \Gamma_B.$$

Thus

$$\hat{\Lambda} = \Lambda. \qquad \text{(D.20)}$$

Formally, we have not shown that the original delivery problem in this paragraph is stable (i.e. has a unique optimal dual) but rather have illustrated that although the perturbation causes a change in the optimal basis, the optimal dual solution is not changed, a behaviour which is characteristic of a stable point. It is worth noting that further study of this example would show that the only unstable point here results from a requirement vector b, such that $q_1^4(0) = \frac{1}{2}q$ and $q_2^4(0) = q$. The corresponding flow schedule is shown in Fig. D.7. In general, for a given network we expect a randomly selected requirement vector b (queues sizes) to be stable though the optimal primal solution may be degenerate. This observation is backed up by the experience we have gained in solving a number of delivery problems. This serves as an additional motivation why we are interested more in stability than in non-degeneracy.

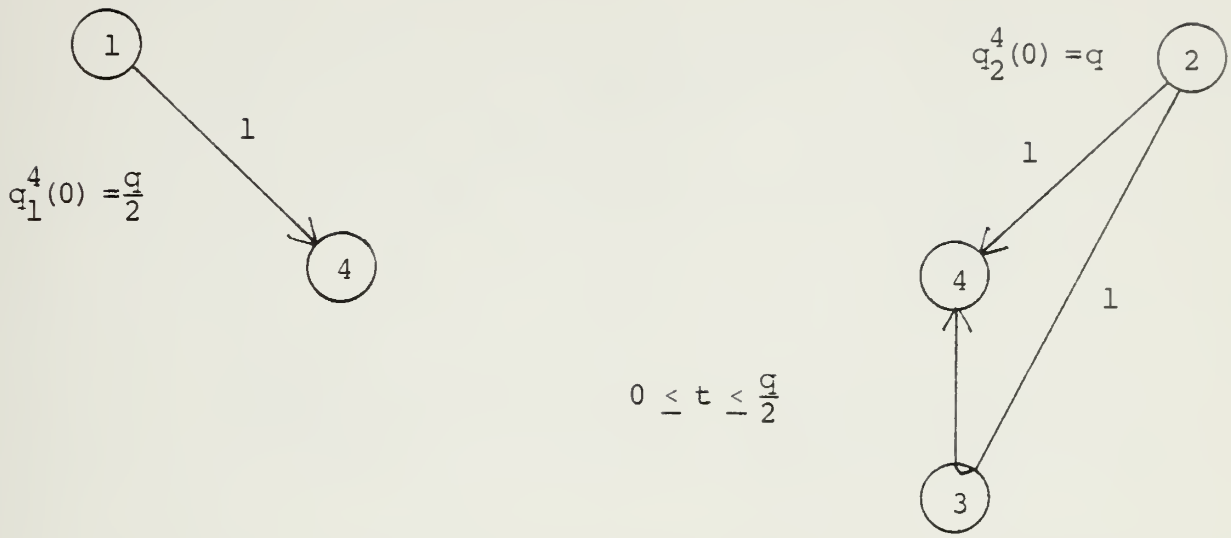


Fig. D.7. Unstable Optimal Flow Schedule

E. OPTIMAL DELIVERY FUNCTION IS PIECEWISE LINEAR

In this section we show that the assumption we made about the piecewise linearity of an optimal delivery function is justified. Assume to the contrary, that there exists a continuous, non-linear optimal delivery function. This function must be convex, since otherwise we could improve on it by generating its convex hull by the method of constant flow substitution (see Appendix I.A). A typical delivery function of this kind is shown in Fig. E.1. The broken line in Fig. E.1 represents the delivery function which corresponds to an optimal solution to MTP(1).

Consider the optimal flow schedule F(t), $0 \leq t \leq t_1^0$ that generates D(t), and in particular the net delivery rates $r_i^k(t)$,

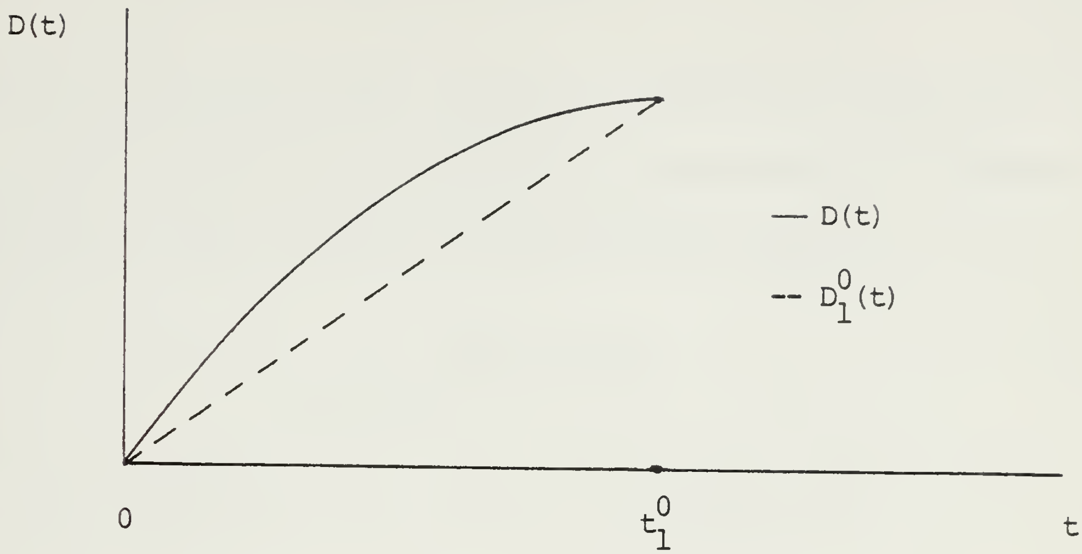


Fig. E.1. Optimal Continuous Non-linear Delivery Function

$\forall (i,k) \in N_0$ at time $t = t_1^0$. Define a perturbation vector ΔQ such that

$$\Delta q_i^k \triangleq \delta r_i^k(t_1^0), \quad \forall (i,k) \in N_1, \tag{E.1}$$

where

$$0 < \delta \leq \delta_1,$$

and δ_1 is the maximal value of δ for which the perturbation ΔQ is acceptable (for simplicity we assume that the problem is at a stable point and thus $\delta_1 > 0$).

The optimal value of the cost function for the perturbed problem is $t_1^0 - \varepsilon$ where (IV.20)

$$\varepsilon = \sum_{(i,k)\in N_1} \sigma_i^k(1)\Delta q_i^k \qquad (E.2)$$

From Thm. IV.4 we have that any flow pattern of commodities in N_1 that saturates the set L_1 (and in particular the flows $R \triangleq \{r_i^k(t_1^0)\}$) must satisfy the perturbation equation, i.e.

$$1 = \sum_{(i,k)\in N_1} \sigma_i^k(1) r_i^k(t_1^0) \qquad (E.3)$$

Using (E.1) and (E.2) with (E.3), we conclude that

$$\varepsilon \equiv \delta \qquad (E.4)$$

Equality (E.4) ensures us that it is possible to deliver the set of queues $\Delta q_i^k = \delta r_i^k(t_1^0)$, $\forall (i,k) \in N_1$ within the period $(t_1^0-\varepsilon, t_1^0]$ by using the set of feasible flows $\{f_{ij}^k(t_1^0)\}$, $\forall (i,k) \in N_1$ of the optimal flow schedule $F(t)$, $0 \leq t \leq t_1^0$. The perturbed (two segment) delivery function is shown in Fig. E.2.

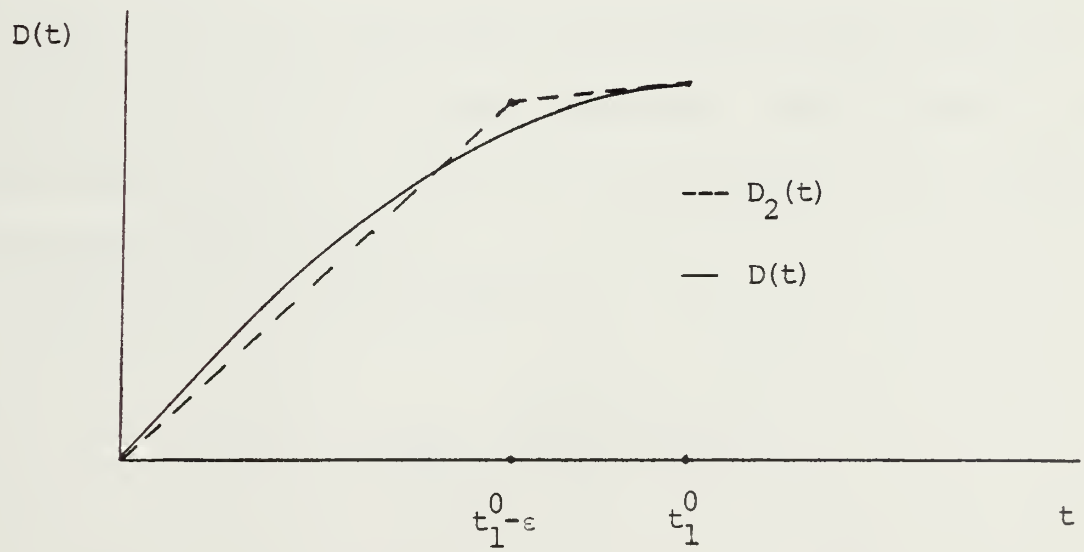


Fig. E.2. Perturbed Delivery Function

Since $D_2(t)$ dominates $D(t)$, the last cannot be an optimal delivery function.

Following the same argument it can be shown that an optimal delivery function must be composed only of linear segments, i.e. it is not possible to have an optimal delivery function as shown in Fig. E.3.

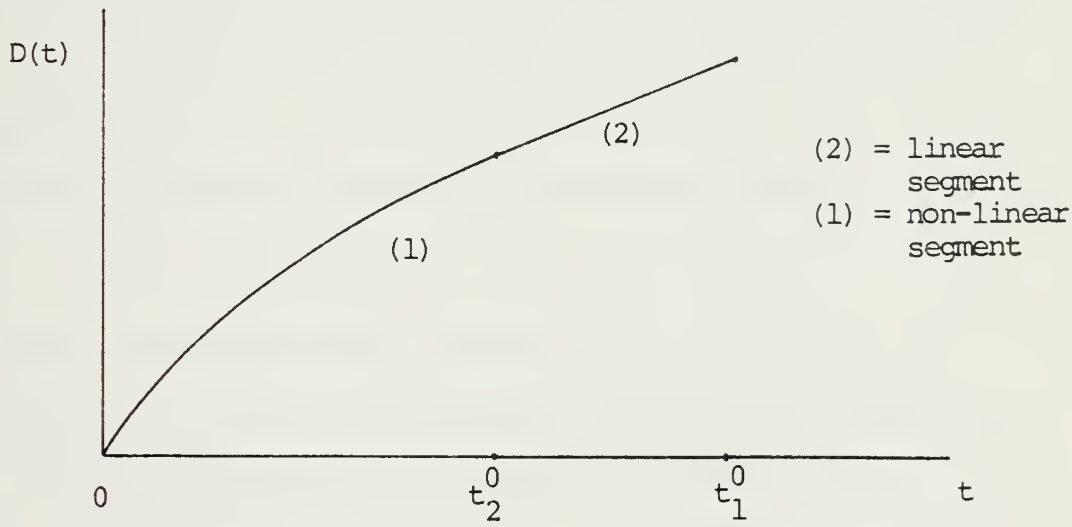


Fig. E.3. Mixed Type Optimal Delivery Function

Basically, this completes our argument about piecewise linearity of optimal delivery functions. We note that in case the optimal solution to MTP (1) is not stable, the argument does not change significantly:

Instead of (E.2) we write

$$\varepsilon = \sum_{(i,k)\in N_1} \sigma_i^k(1,R)\ \Delta q_i^k, \qquad \text{(E.5)}$$

where

$$R \triangleq \{r_i^k(t_1^0)\}, \ \forall (i,k) \in N_1.$$

The dual variables $\sigma_i^k(1,R)$, $\forall (i,k) \in N_1$ in (E.5) correspond to the related optimal dual solution for the perturbed problem (note, $\sigma_i^k(1,R) = \sigma_i^k(1)$, $\forall (i,k) \in N_1$ if the original solution is stable). All other properties remain as before.

We summarize our discussion with the following theorem.

<u>Theorem E.1</u>

An optimal delivery function is piecewise linear.

□

F. ON THE NUMBER OF CORNER POINTS

In Chapter III.B.2 we proposed a conjecture (Conjecture III.1) which upper-bounds the number of corner points of an optimal delivery function with $|N_0|$, the number of non-zero data queues. The conjecture is based on one aspect of our experience with delivery problems, the essence of which may be formalized as follows.

<u>Proposition F.1</u>

At a stable point, let K_m denote the number of commodities for which the dual variables at the m-th corner are equal to their maximal value or are negative, i.e.,

$$K_m = |\{(i,k): \sigma_i^k(m) = \sigma_{max}(m) \text{ or } \sigma_i^k(m) < 0\}|.$$

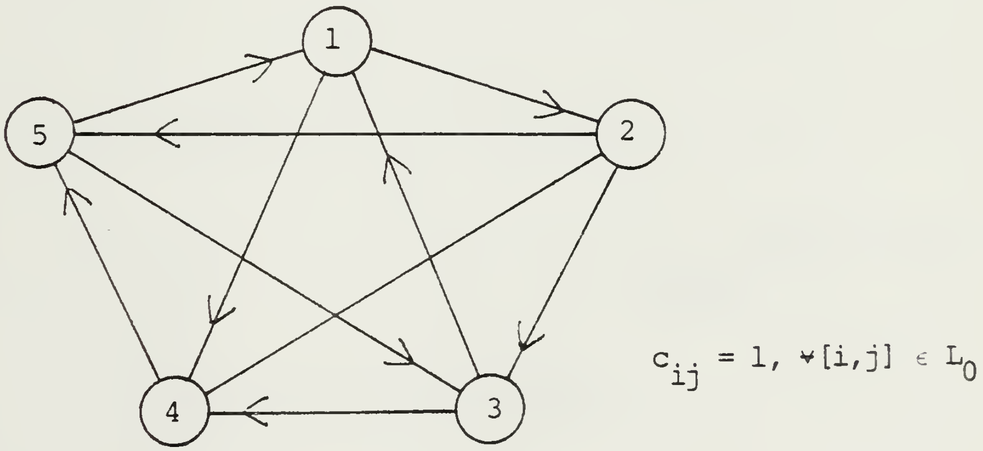
then

$$K_{m+1} > K_m, \quad m = 1,2,\ldots,M.$$

□

Conjecture III.1 is a trivial consequence of Proposition F.1. Actually, we proved a related result (cf. Lemma IV.5) which says that if $K_m = |N_0|$ for some m, then the delivery function cannot be improved any more and thus is optimal. There seems to be a slight difficulty in proving the opposite direction, i.e. that an optimal delivery function implies (at a stable point) equal dual variables. We must leave the proof of this conjecture and of Proposition F.1 as open topics for further research.

G. COMPUTER SOLUTION EXAMPLE OF OPTIMAL DELIVERY PROBLEM

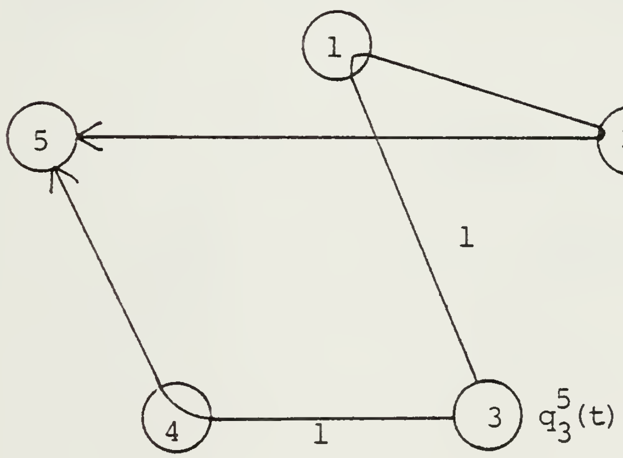


$q_1^5(0) = 70, \quad q_2^4(0) = 35, \quad q_3^1(0) = 50, \quad q_3^5(0) = 25, \quad q_4^2(0) = 25, \quad q_5^3(0) = 35.$

Fig. G.1. Optimal Delivery Problem

The problem in Fig. B.1 was solved using the algorithm of Chapter III. We will not present the partial solutions in detail, but only the optimal flow schedule and its corresponding delivery function. The optimal flow schedule is composed of four segments, the chain flow decomposition of each is shown in the following figures. We append each segment with the relevant information obtained from the solution of the corresponding minimal time problem (recall that in our notation the "first" segment is $(t_2^0, t_1^0])$. Also, we demonstrate the perturbation equation for each one of the flow segments.

MTP(1): $N_1 = \{(1,5),(2,4),(3,1),(3,5)\}$, $L_1 = \{[3,4],[4,5],[2,5],[3,1]\}$, $\sum(1) = (\sigma_1^5(1)=\frac{1}{4}, \sigma_2^4(1)=\frac{1}{4}, \sigma_3^1(1)=\frac{1}{4}, \sigma_3^5(1)=\frac{1}{2}, 0, \ldots, 0)$



$$\sum_{(i,k)\in N_1} \sigma_i^k(1) r_i^k = \frac{1}{2} \cdot 2 = 1$$

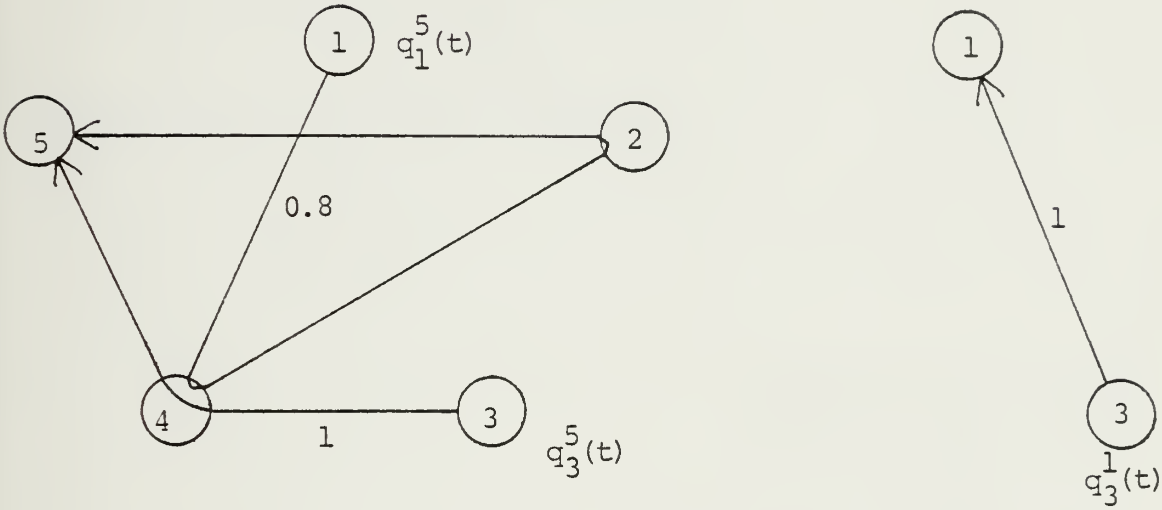
$$\rho_1^0 = 2$$

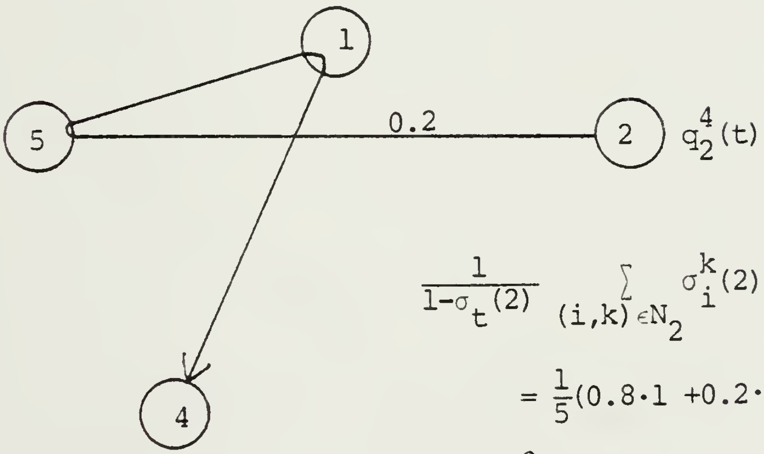
Fig. G.1a. Chain Flow Decomposition for $t \in (50, 51.25]$

MTP(2): $N_2 = N_1, \quad L_2 = L_1$

$\sum(2) = (\sigma_1^5(2)=1, \sigma_2^4(2)=1, \sigma_3^1(2)=2, \sigma_3^5(2)=2, 0, \ldots, 0$

$\sigma_t(t)=-4)$





$$\frac{1}{1-\sigma_t(2)} \sum_{(i,k)\in N_2} \sigma_i^k(2) r_i^k$$

$$= \frac{1}{5}(0.8\cdot 1 + 0.2\cdot 1 + 1\cdot 2 + 1\cdot 2) = 1$$

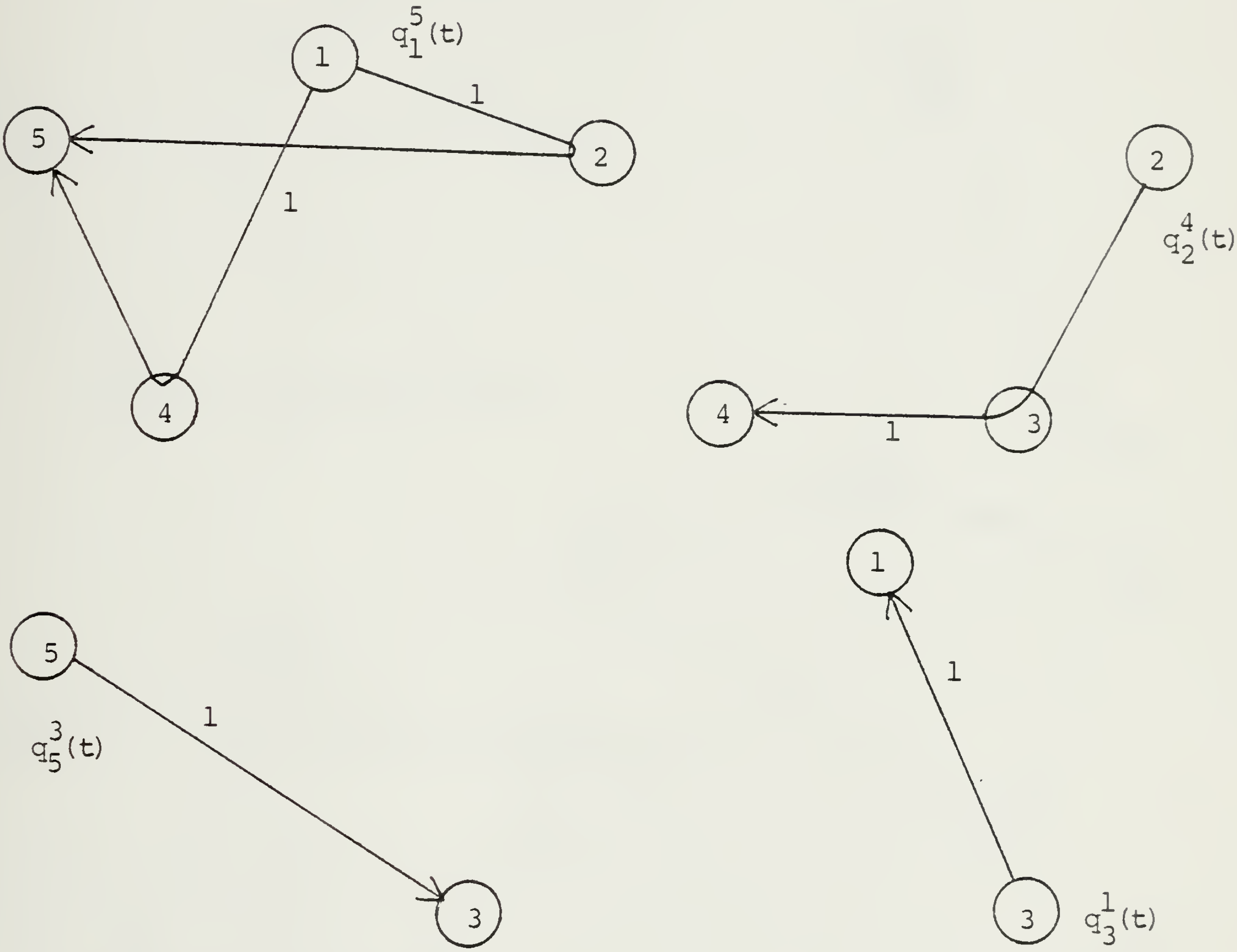
$$\rho_2^0 = 3$$

Fig. G.1b. Chain Flow Decomposition for $t \in (31.25, 50]$

$MTP(3)$: $N_3 = N_2 \cup \{5,3\}$ $L_3 = L_2 \cup \{(2,3),(5,3)\}$

$\sum(3) = (\sigma_1^5(3)=\frac{1}{2}, \sigma_2^4(3)=\frac{1}{2}, \sigma_3^1(3)=\frac{1}{2}, \sigma_5^3(3)=\frac{1}{2}, 0, \ldots, 0,$

$\sigma_t(t)=-1.5)$



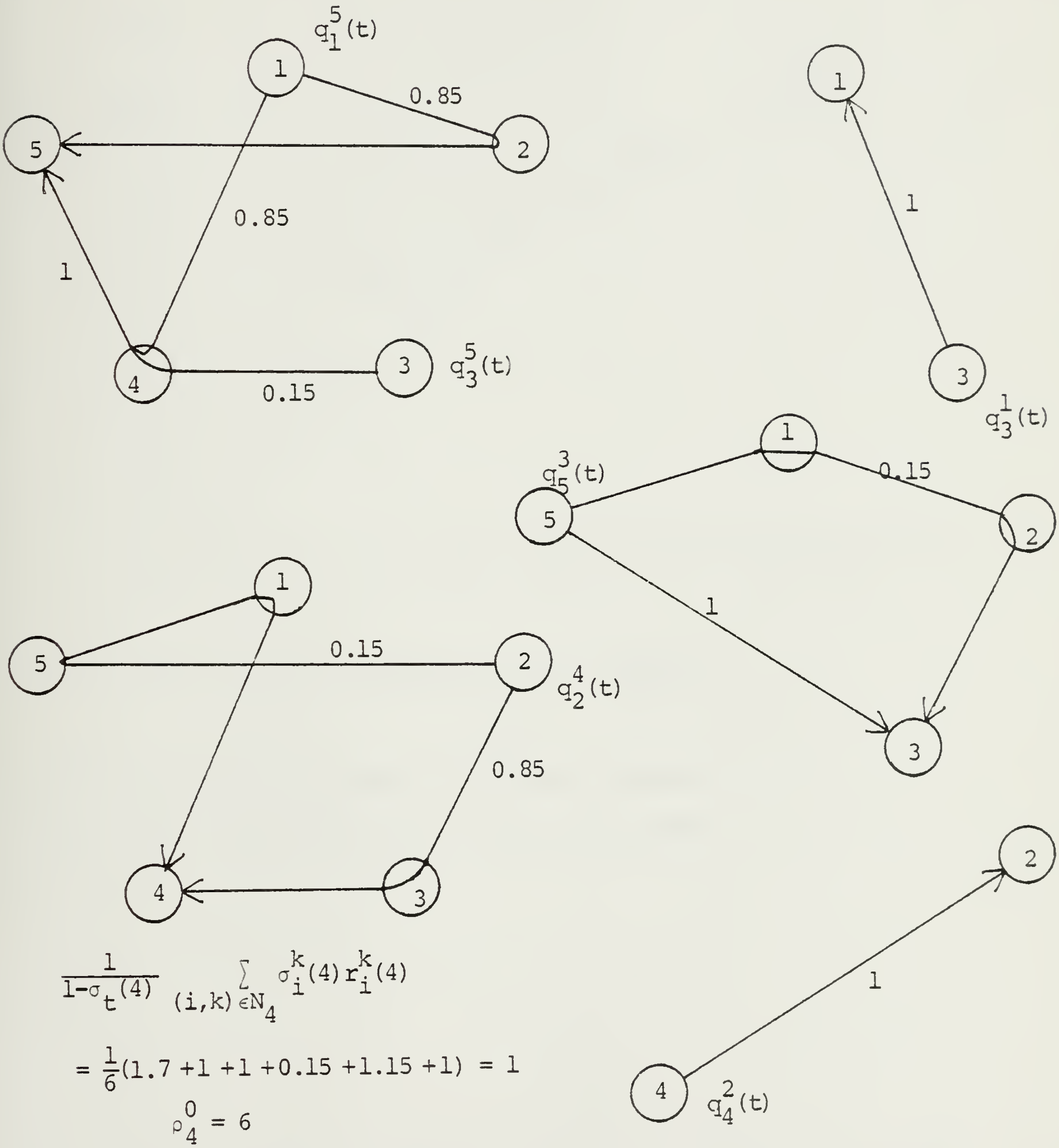
$$\frac{1}{1-\sigma_t(2)} \sum_{(i,k)\in N_0} \sigma_i^k(3)\, r_i^k(3) = \frac{1}{2.5}\left(\frac{1}{2}\cdot 2 + \frac{1}{2}\cdot 1 + \frac{1}{2}\cdot 1 + \frac{1}{2}\cdot 1\right) = 1, \quad \rho_3^0 = 5$$

Fig. G.1c. Chain Flow Decomposition for $t \in (25,31.25]$

MTP(4): $N_4 = N_3 \cup \{(4,2)\}$, $L_4 = L_3 \cup \{[1,2],[1,4],[4,2]\}$

$\sum(4) = (\sigma_1^5(4)=1, \sigma_2^4(4)=1, \sigma_3^1(4)=1, \sigma_3^5(4)=1, \sigma_5^3(4)=1,$

$\sigma_4^2(4)=1, 0, \ldots, 0, \sigma_t(4)=-5)$



$$\frac{1}{1-\sigma_t(4)} \sum_{(i,k)\in N_4} \sigma_i^k(4)\, r_i^k(4)$$

$$= \frac{1}{6}(1.7 + 1 + 1 + 0.15 + 1.15 + 1) = 1$$

$$\rho_4^0 = 6$$

Fig. G.1d. Chain Flow Decomposition for $t \in [0,25]$

The corresponding optimal delivery function is shown in Fig. G.1e.

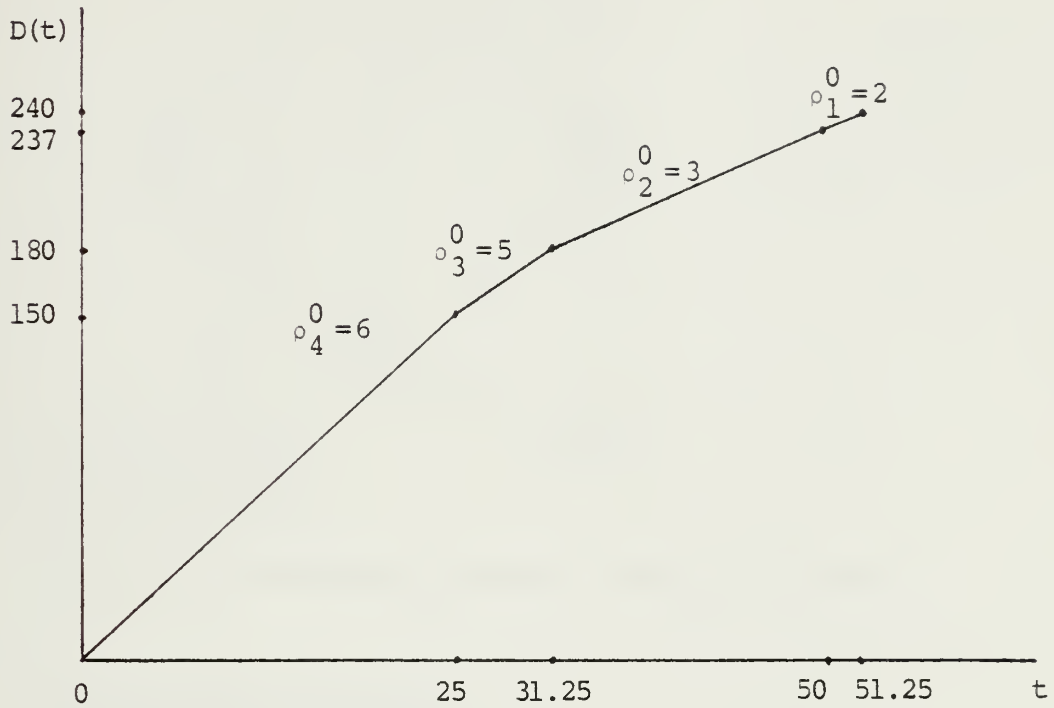
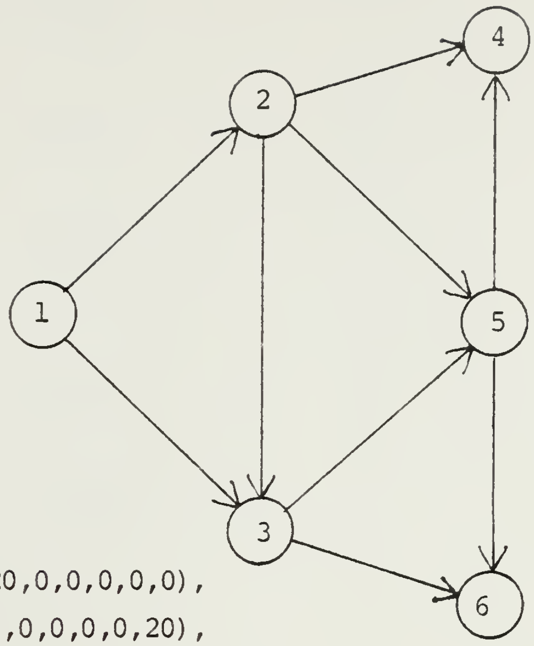


Fig. G.1e. Optimal Delivery Function

H. DETAILED SOLUTION OF THE MDDP EXAMPLE

For completeness we restate the MDDP.



$c_{ij} = 2, \forall [i,j] \in L_0$

$\tau_{ij} = 1, \forall [i,j] \in L_0$

$a_i = b_i = \infty, \forall i \in V$

$Q_0 = (20,0,0,0,0,0)$,
$Q_1 = (0,0,0,0,0,20)$,
$Q_2 = (0,0,0,15,5,0)$.

Fig. H.1. Maximally Delayed Decision Problem

Optimal Flow Schedule:†

$F^1(k)$:

link	period	flow
[1,2]	1	2
"	2	2
"	3	2
[1,3]	1	0
"	2	0
"	3	0

†We assume here that the starting time is k = 1, and not k = 0 as before.

link	period	flow
[2,3]	2	0
"	3	0
[2,4]	2	0
"	3	0
[2,5]	2	0
"	3	2
[3,5]	2	0
"	3	0
[3,6]	2	0
"	3	0
[5,4]	3	0
[5,6]	3	0

$F_1(k)$:

link	period	flow
[1,2]	4	2
"	5	2
"	6	2
"	7	2
"	8	0
[1,3]	4	2
"	5	2
"	6	2
"	7	0
"	8	0
[2,3]	4	2
"	5	2
"	6	2

link	period	flow
[2,3]	7	2
"	8	2
[2,4]	4	0
"	5	0
"	6	0
"	7	0
"	8	0
"	9	0
[2,5]	4	0
"	5	0
"	6	2
"	7	0
"	8	0
"	9	0
[3,5]	4	0
"	5	2
"	6	2
"	7	0
"	8	2
"	9	0
[3,6]	4	0
"	5	2
"	6	2
"	7	2
"	8	2
"	9	2

link	period	flow
[5,4]	4	0
"	5	0
"	6	0
"	7	0
"	8	0
"	9	0
[5,6]	4	2
"	5	0
"	6	2
"	7	2
"	8	2
"	9	2

$\underline{F_2(k)}$:

[1,2]	4	2
"	5	2
"	6	2
"	7	2
"	8	2
[1,3]	4	0
"	5	0
"	6	0
"	7	2
"	8	2
[2,3]	4	0
"	5	0
"	6	2

link	period	flow
[2,3]	7	0
"	8	0
[2,4]	4	2
"	5	2
"	6	2
"	7	1
"	8	0
"	9	2
[2,5]	4	0
"	5	0
"	6	0
"	7	0
"	8	1
"	9	2
[3,5]	4	0
"	5	0
"	6	0
"	7	2
"	8	2
"	9	2
[3,6]	4	0
"	5	0
"	6	0
"	7	0
"	8	0
"	9	0

link	period	flow
[5,4]	4	2
"	5	0
"	6	0
"	7	0
"	8	2
"	9	2
[5,6]	4	0
"	5	0
"	6	0
"	7	0
"	8	0
"	9	0

LIST OF REFERENCES

1. Segall, A, "The Modeling of Adaptive Routing in Data Communication Networks," IEEE Trans. on Comm. Vol. COM-25, No. 1, Jan. 1977.

2. Moss, F.H., "The Application of Optimal Control Theory to Dynamic Routing in Data Communication Networks," Ph.D. Dissertation, M.I.T., 1976.

3. Assad, A.A., "Multicommodity Netowrk Flows: A Survey," Networks, 8, 1, 37-92, 1978.

4. Kennington, T.L., "A Survey of Linear Cost Multicommodity Network Flows," Operations Research, 26, 2, 209-236, 1978.

5. Danzig, G.B., Linear Programming and Extensions, Princeton University, Princeton, 1963.

6. Luenberger, D.G., Introduction to Linear and Nonlinear Programming, Addison-Wesley, 1973.

7. Simmons, D.M., Linear Programming for Operations Research, Holden Day Inc., 1972.

8. Ros, F., "Routing to Minimize the Maximum Congestion in a Communication Network," Ph.D. Dissertation, M.I.T., 1978.

9. Hu, T.C., Integer Programming and Network Flows, Addison-Wesley, 1970.

10. Ford, L.R., Fulkerson, E., Flows in Networks, Princeton University Press, 1962 (Sixth Printing 1974).

11. Shats, S., Segall, A., "Open-loop Solution for the Dynamic Routing Problem," EE Pub. No. 369, Feb. 1980, Technion, Haifa, Israel.

12. Yee, R.J., Personal Communication, 1981.

13. Kleinrock, L., Queueing Systems, Vol. 2, Wiley Science, 1976.

14. Jackson, J.R., "Networks of Waiting Lines," Operations Research, 5, 518-521, 1957.

15. Kleinrock, L., "Communication Nets; Stochastic Message Flow and Delay," Reprint, Dover Publication, 1972.

16. Fratta, L., Gerla, M., Kleinrock, L., "The Flow Deviation Method--An Approach to Store-and-Forward Communication Network Design," Networks, 3, 97-133, 1973.

17. Cantor, D.G., Gerla, M., "Optimal Routing in a Packet Switched Computer Network," IEEE Trans. on Computers, Vol. C-23, Oct. 1974.

18. Defenderfer, J.E., "Comparative Analysis of Routing Algorithms for Computer Networks," MIT, ScD Thesis, March 1977.

19. Vastola, K.S., "A Numerical Study of Two Measures of Delay for Network Routing," R-859, UILU-ENG, 78-2252, University of Illinois, Aug. 1979.

20. Gallager, R.G., "A Minimum Delay Routing Algorithm Using Distributed Computation," IEEE Trans. on Comm., Vol. COM-25, No. 1, Jan. 1977.

21. Yum, P.T., "The Design and Analysis of a Semidynamic Deterministic Routing Rule," IEEE Trans. on Comm., Vol. COM-29, No. 4, April, 1981.

22. Yum, P.T., Schwartz, M., "The Join -Biased-Queue Rule and Its Application to Routing in Computer Communication Networks," IEEE Trans. on Comm., Vol. COM-29, No. 4, April, 1981.

23. Boorstyn, R.R., Livne, A., "A Technique for Adaptive Routing in Networks," IEEE Trans. on Comm., Vol. COM-29, No. 4, April, 1981.

24. Bookbinder, H.J., Sethi, P.S., "The Dynamic Transportation Problem: A Survey," Naval Research Logistics Quarterly, 65-87, March, 1981.

25. Hammer, P.L., "Time-Minimizing Transportation Problems," Naval Research Logistics Quarterly, 18, 487-490 (1971).

26. Szwarc, W., "Some Remarks on the Time Transportation Problem," Naval Research Logistics Quarterly 18, 473-485 (1971).

27. Tapiero, C.S., Soliman, M.A., "Multicommodities Transportation Schedules Over Time," Networks, Vol. 2, 311-327 (1972).

28. Reingold, M.E., Nievergelt, J., Deo, N., Combinatorial Algorithms--Theory and Practice, Prentice Hall, 1977.

29. Wozencraft, T.M., Gallager, R.G., Segall, A., "First Annual Report on Data Network Reliability," ESL-IR-677, MIT, July, 1976.

30. Kirk, E.D., Optimal Control Theory--An Introduction, Prentice-Hall, 1970.

31. Jodorkovsky, M., Segall, A., "A Maximal Flow Approach to Dynamic Routing in Communication Networks," EE Pub. No. 358, Aug. 1979, Technion, Haifa, Israel.

32. Gal, T., Postoptimal Analyses, Parametric Programming and Related Topics, McGraw Hill, 1979.

33. Cullum, J., "Discrete Approximation to Continuous Optimal Control Problems, SIAM J. Control, Vol. 7, No. 1, Feb. 1969.

34. Ford, L.R., Fulkerson, D.R., "Constructing Maximal Dynamic Flows from Static Flows," Operations Res. 6 (1958), 419-433.

35. Merchant, K.D., Nemhauser, L.G., "A Model and an Algorithm for Dynamic Traffic Assignment Problems," Transportation Survey, Vol. 12, No. 3, Aug. 1978.

36. Ho, K.J., Manne, S.A., "Nested Decomposition for Dynamic Models," Mathematical Programming 6 (1974), North Holland Pub. Co.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Technical Information Center Cameron Station Alexandria, Virginia 22314	2
2.	Library, Code 0142 Naval Postgraduate School Monterey, California 93940	2
3.	Department Chairman, Code 62 Department of Electrical Engineering Naval Postgraduate School Monterey, California 93940	1
4.	Professor M. Athans Department of Electrical Engineering and Computer Science Massachusetts Institute of Technology Cambridge, Massachusetts 02139	1
5.	Professor W.B. Davenport, Jr. Department of Electrical Engineering and Computer Science Massachusetts Institute of Technology Cambridge, Massachusetts 02139	1
6.	A. Feit 14/7 Kazan st. Ranana, 43000 Israel	10
7.	Professor R.G. Gallager Department of Electrical Engineering and Computer Science Massachusetts Institute of Technology Cambridge, Massachusetts 02139	1
8.	Professor D.P. Gaver, Jr., Code 55Gv Department of Operations Research Naval Postgraduate School Monterey, California 93940	1
9.	Professor R.W. Hamming, Code 52Hg Department of Computer Science Naval Postgraduate School Monterey, California 93940	1

10. Dr. R.E. Kahn 1
Defense Advanced Research Projects Agency
1400 Wilson Blvd.
Arlington, Virginia 22209

11. Joel Lawson, Code 06T 1
Naval Electronic Systems Command
Washington, D.C. 20360

12. Professor P.H. Moose, Code 62Me 1
Department of Electrical Engineering
Naval Postgraduate School
Monterey, California 93940

13. Professor M.A. Morgan, Code 62Mw 1
Department of Electrical Engineering
Naval Postgraduate School
Monterey, California 93940

14. Professor A. Segall 1
Department of Electrical Engineering
Technion--Israeli Institute of Technology
Haifa, Israel

15. Provost and Academic Dean (Acting), Code 01 1
Dr. D.A. Schrady
Naval Postgraduate School
Monterey, California 93940

16. Dean of Research, Code 012 1
Dr. W.M. Tolles
Naval Postgraduate School
Monterey, California 93940

17. Professor J.M. Wozencraft, Code 74 3
Department of Electrical Engineering
Naval Postgraduate School
Monterey, California 93940

18. Professor J. Ziv 1
Department of Electrical Engineering
Technion--Israeli Institute of Technology
Haifa, Israel

19. Dr. Francisco Ros-Peran 1
Leon Felipe 16, 3º B
Majadahonda (Madrid)
SPAIN

20. Mr. James Yee 1
Room JEC 6038
Rensselaer Polytechnic Inst.
Troy, N.Y. 12181

21. Prof. Kneale T. Marshall, Code 55Mt 1
Chairman, Department of Operations Research
Naval Postgraduate School
Monterey, California 93940

22. Dr. Stuart Starr 1
Director, Long Range Planning & System Evaluation
ODUSD/C3I
Rm 3E182, Pentagon
Washington, D.C. 20301

23. CAPT. M.G. Chlebik 1
M.C.T.S.S.A.
Camp Pendleton, California 90255

24. Office of Research Administration 1
Code 012A
Naval Postgraduate School
Monterey, California 93940

Thesis 197684
F2528 Feit
c.1 Dynamic multicommodity
flow schedules.