P. Saunders

NATIONAL BUREAU OF STANDARDS REPORT

10 049

PASSENGER TRANSPORTATION SCHEDULING

Technical Report
to
Northeast Corridor Transportation Project

NBS

U.S. DEPARTMENT OF COMMERCE
NATIONAL BUREAU OF STANDARDS

NATIONAL BUREAU OF STANDARDS

The National Bureau of Standards [1] was established by an act of Congress March 3, 1901. Today, in addition to serving as the Nation's central measurement laboratory, the Bureau is a principal focal point in the Federal Government for assuring maximum application of the physical and engineering sciences to the advancement of technology in industry and commerce. To this end the Bureau conducts research and provides central national services in four broad program areas. These are: (1) basic measurements and standards, (2) materials measurements and standards, (3) technological measurements and standards, and (4) transfer of technology.

The Bureau comprises the Institute for Basic Standards, the Institute for Materials Research, the Institute for Applied Technology, the Center for Radiation Research, the Center for Computer Sciences and Technology, and the Office for Information Programs.

THE INSTITUTE FOR BASIC STANDARDS provides the central basis within the United States of a complete and consistent system of physical measurement; coordinates that system with measurement systems of other nations; and furnishes essential services leading to accurate and uniform physical measurements throughout the Nation's scientific community, industry, and commerce. The Institute consists of an Office of Measurement Services and the following technical divisions:

> Applied Mathematics—Electricity—Metrology—Mechanics—Heat—Atomic and Molecular Physics—Radio Physics [2]—Radio Engineering [2]—Time and Frequency [2]—Astrophysics [2]—Cryogenics.[2]

THE INSTITUTE FOR MATERIALS RESEARCH conducts materials research leading to improved methods of measurement standards, and data on the properties of well-characterized materials needed by industry, commerce, educational institutions, and Government; develops, produces, and distributes standard reference materials; relates the physical and chemical properties of materials to their behavior and their interaction with their environments; and provides advisory and research services to other Government agencies. The Institute consists of an Office of Standard Reference Materials and the following divisions:

> Analytical Chemistry—Polymers—Metallurgy—Inorganic Materials—Physical Chemistry.

THE INSTITUTE FOR APPLIED TECHNOLOGY provides technical services to promote the use of available technology and to facilitate technological innovation in industry and Government; cooperates with public and private organizations in the development of technological standards, and test methodologies; and provides advisory and research services for Federal, state, and local government agencies. The Institute consists of the following technical divisions and offices:

> Engineering Standards—Weights and Measures — Invention and Innovation — Vehicle Systems Research—Product Evaluation—Building Research—Instrument Shops—Measurement Engineering—Electronic Technology—Technical Analysis.

THE CENTER FOR RADIATION RESEARCH engages in research, measurement, and application of radiation to the solution of Bureau mission problems and the problems of other agencies and institutions. The Center consists of the following divisions:

> Reactor Radiation—Linac Radiation—Nuclear Radiation—Applied Radiation.

THE CENTER FOR COMPUTER SCIENCES AND TECHNOLOGY conducts research and provides technical services designed to aid Government agencies in the selection, acquisition, and effective use of automatic data processing equipment; and serves as the principal focus for the development of Federal standards for automatic data processing equipment, techniques, and computer languages. The Center consists of the following offices and divisions:

> Information Processing Standards—Computer Information — Computer Services — Systems Development—Information Processing Technology.

THE OFFICE FOR INFORMATION PROGRAMS promotes optimum dissemination and accessibility of scientific information generated within NBS and other agencies of the Federal government; promotes the development of the National Standard Reference Data System and a system of information analysis centers dealing with the broader aspects of the National Measurement System, and provides appropriate services to ensure that the NBS staff has optimum accessibility to the scientific information of the world. The Office consists of the following organizational units:

> Office of Standard Reference Data—Clearinghouse for Federal Scientific and Technical Information [3]—Office of Technical Information and Publications—Library—Office of Public Information—Office of International Relations.

---

[1] Headquarters and Laboratories at Gaithersburg, Maryland, unless otherwise noted; mailing address Washington, D.C. 20234.
[2] Located at Boulder, Colorado 80302.
[3] Located at 5285 Port Royal Road, Springfield, Virginia 22151.

NATIONAL BUREAU OF STANDARDS REPORT

NBS PROJECT
2058456

June, 1969

NBS REPORT
10 049

PASSENGER TRANSPORTATION SCHEDULING

by
Dennis R. Young
Stanford University and Technical Analysis Division

Technical Report
to
Northeast Corridor Transportation Project
Department of Transportation

NBS

U.S. DEPARTMENT OF COMMERCE

NATIONAL BUREAU OF STANDARDS

PREFACE

This document differs in origin from the typical National Bureau of Standards Report. It constitutes the author's 1969 Ph.D. dissertation, in Stanford University's Department of Engineering Economic Systems. Dr. Young's research in this field was begun during his stay at the Bureau's Technical Analysis Division, and was recorded both in detailed form ("Scheduling to Maximize Passenger Satisfaction", NBS Report 9569, July 1967) and in a compact open-literature version ("Scheduling a Vehicle between and Origin an a Destination to Maximize Passenger Satisfaction", Proc. 22-nd Nat. Conf.,Assoc. Comp. Mach., 1967; pp. 233-245). He has as a courtesy made his thesis available to us for issue as a Report, so that we could maintain continuity of detailed documentation for our sponsor in this area, the Department of Transportation's Northeast Corridor Transportation Project.

The document has been reproduced verbatim, apart from a few trivial typographical corrections, from the copy supplied by the author. Clarity, suitability of organization and general technical soundness have been checked, but the paper has not been subjected to a full-scale technical and stylistic review, and some possible revisions have been omitted in the interest of prompt issuance.

PASSENGER TRANSPORTATION SCHEDULING

A DISSERTATION

SUBMITTED TO THE DEPARTMENT OF ENGINEERING ECONOMIC SYSTEMS

AND THE COMMITTEE ON THE GRADUATE DIVISION

OF STANFORD UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

By

Dennis Ralph Young

March, 1969

PASSENGER TRANSPORTATION SCHEDULING

Dennis Ralph Young, Ph.D.
Stanford University, 1969

This dissertation presents a method for developing near optimal timetables for the operation of fixed schedule passenger transportation systems. The establishment of efficient operating policies for existing systems is clearly desirable. Furthermore, the ability to compute optimal schedules for drawing-board systems is required in the consideration of planning alternatives for future transportation investments. Here, timetable optimization is accomplished by maximizing an objective function consisting of three basic components: operating costs, revenues, and traveler benefits. Traveler benefits are computed using a consumer preference model that specifies a traveler's willingness to pay for alternative trips, based on the scheduling of those trips. The optimization is based on successive iterations of a dynamic programming algorithm that develops timetables for each vehicle in the system. Solutions are termed optimal to within a "first-order passenger exchange."

Examples are presented to illustrate the utilization of the methodology for computing information relevant to the development of system operating policies and the consideration of planning alternatives. Finally, an analysis is developed for the problem of choosing among proposed alternative transportation systems in the face of uncertainty about traveler demand and preference behavior.

Approved for publication:

By Richard D. Smallwood
Engineering Economic Systems

By Virgil K. Whitaker
Dean of the Graduate Division

ACKNOWLEDGEMENTS

This research has been supported by the National Science Foundation under contract number NSF-DK-1683. The work was initiated at the Technical Analysis Division of the National Bureau of Standards for the Northeast Corridor Transportation Project.

The author wishes to thank his associates at Stanford University and at the National Bureau of Standards for their help and encouragement in the execution of this research. Special thanks are due Dr. Alan Goldman of N.B.S. and Dr. Richard D. Smallwood of Stanford. Dr. Goldman provided invaluable encouragement and support during the formulation stage of this effort. Dr. Smallwood, in his capacity as thesis advisor, ably supervised the development of this work and furnished many valuable suggestions toward the writing of this dissertation.

Finally, the author wishes to thank his wife Rosalie and son Seth for their moral support while this thesis was being written.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

FIGURE PAGE

FIGURE		PAGE
39.	Traveler Population Type P (general public) ------------	176
40.	a) Population B Time Preference Parameters -------------	177
	b) Population P Time Preference Parameters -------------	178
41.	Minimum Valid Fare Levels ------------------------------	180
42.	System Performance Versus Fare -------------------------	182
43.	Variation of Fleet Timetable with Fare -----------------	184
44.	Variation of Schedule Properties with Fare -------------	185
45.	System Performance for Alternate Traveler Populations --	187
46.	Comparison of Timetables for Alternate Traveler Populations --	189
47.	System Performance for Alternate Vehicle Fleets --------	190
48.	Comparison of Timetables for Alternate Vehicle Fleets---	191
49.	Variation of Performance with Fleet Size ---------------	193
50.	Variation of Profit and Cost Under Alternate Objectives-	194
51.	Variation of Patronage Under Alternate Objectives ------	196
52.	Variation of Schedule Properties for Profit Objective --	198
53.	Variation of Timetable Under the Profit Objective-------	200
54.	Non-Stop System --	207
55.	Multi-Node Link System ---------------------------------	207
56.	Network Flow Representation of Benefit Aggregation Along a Link --	210
57.	Min-Cost Flow Benefit Solution -------------------------	212
58.	Networks with Alternate Paths --------------------------	214
59.	Choosing Decision Nodes --------------------------------	214
60.	Alternate Applications of the Scheduling Algorithm -----	222

FIGURE		PAGE

CHAPTER I

INTRODUCTION

This thesis is a study of the problem of scheduling passenger transportation systems. The term scheduling refers to the determination of timetables and corresponding traveler accommodations for a system of transportation vehicles operating over a network. Consideration is restricted to common carrier transportation systems that operate by fixed schedules (e.g. published timetables) as opposed to real time dispatching policies.

In recent years there has been a growing emphasis on transportation planning and research. For example, the Northeast Corridor Transportation Project,* under which this research was initiated, has emphasized the development of methodology relevant to the intelligent selection among proposed alternative transportation system configurations [1]. The Office of High Speed Ground Transportation,* in its report on high priority research [2], has recognized the significance of scheduling to that task. Thus, the problem of developing good schedules is important not only for the generation of day to day operating policy; more importantly, a method for synthesizing schedules that maximize significant measures of system performance is essential to decision makers who must evaluate proposed alternative systems. A method for developing "optimal schedules" for alternative systems permits the proper comparison of these

---

* The Northeast Corridor Transportation Project and the Office of High Speed Ground Transportation are within the jurisdiction of the U.S. Department of Transportation.

alternatives--each in conjunction with its own best scheduling policy. The net benefit associated with the best schedule is the measure of merit for a proposed system.

In this dissertation, the scheduling problem is considered from a "systems" viewpoint, so that scheduling is properly perceived in relation to system objectives, user and operator benefits and costs, available resources (capital equipment), and decision making. This consideration is important for both current operations and planning decisions. With respect to day to day operations, for example, a scheduling methodology enables development of timetables and pricing policies appropriate to the traveler market, the system objective, and the available vehicle inventory and network. In the planning process, scheduling methodology facilitates evaluation of the system-wide effects of employing alternate fleet sizes and vehicle speeds and capacities. In sum, the system perspective is necessary to appreciate the full implications of scheduling.

At this point, it is worthwhile to review the literature in transportation scheduling so that the specific content and contribution of this thesis may be viewed in perspective, later. The literature documents scheduling studies made in conjunction with various modes of transportation (e.g. rail, air). In addition, both fixed schedule and real time dispatching systems have been studied. Bisbee and Kuroda [3, 4], and Meyer and Wolfe [5], have considered the latter. Only fixed schedule systems are considered here. The mode orientation of each contribution to the literature will be mentioned in the course of the following discussion.

The relevant literature may usefully be viewed along the following lines of classification: element of scheduling, relevant factors accounted for, objective to be enhanced, and method of optimization. The element of scheduling refers to the particular subproblem that a study chooses to address; a subproblem, such as constructing a timetable over a route, or allocating vehicles to trips, is an integral part of the overall scheduling problem. Relevant factors, such as costs and traveler delay, are those quantities that a study deems important in the determination of a solution. An objective, such as maximization of profit, specifies the goal for operating the system under study. An optimization method, such as mathematical programming, is a means to compute the solutions corresponding to the specified objective. Contributions to the literature are reviewed according to these four categories, below.

The following elements of transportation scheduling have been investigated by researchers. In the area of evaluation and cost accounting of preconceived operating policies, several efforts have been made. The airline simulation projects documented by Croswell [6], Kingsley [7], and Howard and Eberhardt [8], and the railroad simulation analysis of MacDonald-Taylor [9], all include this aspect of scheduling. Another element that has concerned researchers is the determination of the required vehicle inventory for implementing a given timetable. The paper by Carstens, Baxter, and Reitman [10] is directly concerned with this question, in the context of railroad systems. In addition, Sugiyama [11] has investigated the minimum number of required seats or vehicles to ensure full passenger seating on rail trips. Other efforts such as those by Simpson [12] for airbus systems, and Lampkin and Saalmans [13] for municipal buses, have included computation of minimum vehicle requirements as part of overall schedule determination considerations.

A third element of scheduling that has received attention is the assignment of vehicles and/or passenger flows to given routes and trip departures. Gunn's [14] airline simulation work, Larson's [15, 16] airline scheduling papers, and Miller's [17] book on airline efficiency, are primarily concerned with these questions. Furthermore, the Boeing airline simulation effort [18], and papers by Kushige [19], Taylor [20], Gagnon [21], and Howard and Eberhardt [8], have investigated this area in airline contexts.

A fourth element, closer to the problem of timetable construction, is the determination of frequency of service (i.e., number of trips/hour) levels to meet demands over specified routes. Simpson and Neuve-Eglise [22] and Kushige [19] have analyzed this problem area. Welding [23] has investigated the relationship of service frequency to the stability of timetable execution in high frequency service systems.

Finally, the area of timetable construction has been approached from various directions by different investigators. Several workers such as Heap [24], and Martin-Lof [25], have been concerned with the problem of spacing a set of trips along a route over time, given the overall frequency of service level. Others such as Simpson [12] and Foulkes, Prager, and Warner [26], have addressed the same question as part of their consideration of more comprehensive scheduling problems. More fundamental work has been done by Devanney [27] and Ward [28] to determine optimal timetables for one-way dispatching of vehicles along simple linear networks. Another study, given by the author [29], offers the generation of optimal timetables for a single vehicle operating over a single link network with one-way traveler demand.

There has been some progress towards developing integrated approaches that account for several elements of scheduling in order that high quality, total system solutions may be synthesized. For instance, various comprehensive heuristic methods for developing system operating policies are in current use. Railroads employ automated procedures for route and timetable selection [30], and airlines are well advanced in the art of simulation studies [7, 8, 18]. In another context, Newton and Thomas [31] have developed a computerized semi-analytical approach for route design, vehicle allocation, and timetable development for a school bus system. Lampkin and Saalmans [13] have combined a heuristic algorithm and search procedure to plan routes, timetables, and vehicle requirements for a municipal bus system. All these efforts were directed at particular "real world" problems, and almost by necessity consist in some part of improvised procedures.

On a more theoretical level, the following contributions have been made: The determination of timetables over simple route segments, as discussed in the works of Devanney, Ward, Heap, and Martin-Lof mentioned earlier, suggests the coordination of these route timetables in larger networks. This approach has been put forward by Foulkes, Prager, and Warner [26] for bus service, Bisbee and Devanney [32] in a general passenger transportation context, and is suggested by the analysis of Beckman, McGuire, and Winsten [33] for freight train scheduling.

A dynamic programming method, based on one-shot scheduling of each vehicle, has been given by Hyman and Gordon [34, 35] for developing a total fleet timetable for airline operations. The previously mentioned comprehensive approach taken by Simpson, includes the analytical determination of passenger flows, the fitting of timetables to these flows, and the determination of minimum vehicle fleet requirements to implement the timetables. Finally, Devanney [36] has attempted the optimization of total system timetables through a novel "pseudo-adaptive programming" approach.

In sum, various parts of the passenger transportation scheduling problem have been studied. Some progress has been made towards integration of subproblem solutions into a comprehensive framework. However, additional research in this most important task is still required.

The second category along which research efforts differ involves the factors that are considered relevant to the analysis. Of course, every study must account for certain physical system attributes such as network structure, vehicle characteristics, and fleet size. In the more practically oriented studies [6, 20], other physical requirements such as maintenance, and crew and equipment availability, are explicitly accounted for. In the case of Welding's [23] study, the physical requirement of safety is a primary determining factor in the analysis of minimum headways for high frequency scheduled service.

A second set of factors, more often considered in the less theoretical approaches, consists of "institutional" considerations. For example, some works [8, 18, 21] have directly integrated the effects of competition and government regulation into their analyses. A third group of factors contribute to the measurement of economic efficiency. Most scheduling studies, for example, take revenues and operating cost into consideration. Others emphasize additional indicators such as load factor [12] and total expenditure level [13], as having major significance.

Finally, the most difficult factors to handle are those pertaining to the satisfaction of traveler needs for transport services. The concepts that have been used to understand and quantify this crucial aspect of passenger systems may be divided into two categories--aggregate measures of service and individual traveler properties. The aggregate measures include total waiting time or delay [27], total travel time [13], an aggregate utility index [29], and fare and time elasticities of demand [22]. Individual traveler properties include traveler riding time constraints [31], purpose of travel [8], and passenger preferences among alternate departure and arrival times [29], levels of speed and fare [18], and multi-stop and non-stop trips [21]. In addition, Gunn's [14] concept of "persistence of demand"--the amount of time by which a potential passenger will advance or delay his desired departure in order to use a service offered, is one of the more imaginative indices of traveler behavior.

The set of factors considered in a particular scheduling study is closely related to the specification of system objectives. The diversity of factors catalogued above indicates the heterogeneity of objective formulations found in the literature. Another reason for the diversity of objectives, of course, is that different researchers have investigated different elements of the scheduling problem. For example, if one is interested in assigning vehicles to a given timetable or in computing vehicle requirements to service a specified demand, then it may make sense to minimize vehicle requirements [11] or dollar costs [20]. Such objectives do not apply to the problem of timetable construction, since the timetable with no scheduled trips is obviously the one that minimizes costs or vehicle requirements.

The array of objectives found in the literature may be classified along four lines. First, some works are merely concerned with synthesizing schedules that are feasible, i.e., that meet prespecified demand or service and resource requirements [30]. A second class consists of objectives that are purely economic in nature. These include maximization of revenues [34], minimization of costs, and maximization of profit [19]. A third class of objectives pertains to the enhancement of traveler service indices. These include minimization of total waiting time or delay [26], or total travel time [13], or maximization of a utility index [29] or a service index [7]. A final class of objectives combines economic and traveler service criteria. An example in this class is the minimization of a combination of traveler delay and capacity costs [27, 28]. In general, however, insufficient attention has been paid to formulation of the latter type objective.

The final facet of scheduling research is the method of optimization by which system objectives are extremized in the attainment of solutions. The applicability of particular techniques depends upon the element of scheduling that is considered. This partially accounts for the variety of methods that are found in the literature. The range of techniques includes analytic approaches, simulation, heuristics, and mathematical programming.

The following analytical approaches have been offered for the solution of some aspects of transportation scheduling. Beckman et al. [33] have derived analytic solutions for determining the best freight schedule of a given connectivity structure in simple networks. Carstens et al. [10] give an analytic solution to minimum vehicle requirements in a two node rail network. Sugiyama [11] has presented an analysis to find minimum seat requirements to yield satisfactory accommodation of passenger loads. Welding [23] has contributed an empirical analysis of headway and stability for close interval service. And Foulkes et al. [26] solve the sequencing of buses in a feeder-trunkline network in terms of a set of linear simultaneous equations.

Consideration of the more comprehensive aspects of scheduling has required more complex approaches. Simulation and trial and error improvement techniques have been used in various quarters. The efforts of Carstens et al. [10], Croswell [6], Kingsley [7], MacDonald-Taylor [9], Gunn [14], Howard and Eberhardt [8], and the Boeing Company [18], attest to this fact. Other heuristic techniques have also been found useful. Heuristic methods used by railroads for synthesizing operating policies [30], Taylor's [20] algorithm for achieving maintenance schedules compatible with flight timetables, and Gagnon's [21] allocation procedure for assigning passengers to flights based on behavioral assumptions, are all examples of intuitive procedures designed to achieve reasonably good solutions.

One of the most useful classes of optimization tools for various aspects of the scheduling problem is mathematical programming. For example, linear programming has been exploited by Miller [17] to assign aircraft to routes, and travelers to flights. Aircraft allocation by linear programming has also been done by Boeing [18], and by Kushige [19]. Another area of programming that has received much recent attention is network flow theory. Simpson, and Newton and Thomas each utilize network flow algorithms within a larger computational framework, as described below. Finally, dynamic programming has proved to be one of the most valuable programming approaches. Research work using this method is also reviewed below.

Several research works on scheduling have employed a combination of the previously mentioned techniques to synthesize a comprehensive computational method. For example, Lampkin and Saalmans [13] use a four step procedure to produce municipal bus schedules. Bus routes are chosen via a heuristic procedure that assures that routes have certain desirable attributes. Service frequencies over the routes are found by minimizing total travel time by a random search procedure. Service frequencies are translated into timetables by "conventional" methods to achieve regular service. Finally, linear programming is used to compute the minimum vehicle requirements for implementing the timetables. Another effort that utilizes multiple techniques is the school bus study by Newton and Thomas [31]. In that work, route structure is found by solving a "traveling salesman" network flow problem. The routes are then partitioned into sections for individual buses, so that bus capacity and loading and riding time constraints are met. Finally, bus timetables are fitted to the route segments to ensure feasible stopping and loading times. One of the most imaginative multiple step schemes has been devised by Simpson [12] to compute timetables for an airbus system. The route structure is chosen heuristically. Daily passenger flows on each route segment are found by a minimum-cost network flow algorithm that minimizes total passenger miles, given the demands between city pairs. Frequency of service on each route segment is chosen to accommodate the passenger flows, based on a specified vehicle capacity and load factor. The flights are spaced over time to distribute demand evenly. Finally, timetables are adjusted within small ranges, to minimize vehicle requirements. The value of these multi-stage approaches is that they present feasible procedures for obtaining solutions to the very large computational problem of timetable synthesis. The difficulty is that they are ad hoc in nature, so that there is no clear system objective which the resulting solutions may purport to maximize. More direct approaches are found in the dynamic programming efforts described below.

Dynamic programming has found application in the scheduling literature largely because it is a very flexible methodology for problem formulation and optimization. However, straightforward application of dynamic programming is prohibitive in terms of computer time and memory requirements. Hence, all dynamic programming approaches are either restricted to simple cases, or employ some computational modifications that compromise the optimality of solutions. Devanney's [27] initial research uses dynamic programming to develop optimal one-way timetables for dispatching vehicles on a linear network from an origin to a destination node. Ward [28] implemented Devanney's algorithms to perform computer studies of dispatching policies over three types of one-way linear networks: point to point, line of stations, and a two station loop.

The algorithms of Devanney and Ward are limited in that they are concerned solely with the dispatching times of trips, and not with individual vehicle trajectories. Thus, the coordination of dispatches in alternate directions or on alternate network links is left unsolved.

Bisbee and Devanney [32] have approached this problem by implementing Simpson's heuristic fleet minimization algorithm to coordinate dispatches on connecting links, and to compute vehicle fleet size requirements.

Other dynamic programming approaches have focused on tracking vehicle trajectories. For example, the author's paper [29] on scheduling a single vehicle over a single link with one-way demand, utilizes a dynamic programming iterative equation. Larson [15] has used forward dynamic programming to compute aircraft trajectories for fulfilling a preset flight schedule. His conventional formulation is unable to handle a system with more than a handful of vehicles. Larson's subsequent approach [16], however, integrates the principle of successive approximations with dynamic programming to compute high quality, though not necessarily optimum, trajectories. Incidentally, although Larson addresses the implementation of a specified timetable, his concern is very close to that of timetable synthesis. His solutions allow unscheduled empty "ferry" flights, and modifications of the input schedule. However, he is not concerned with the state of the passenger population in that the returns from each flight are assumed known. Hyman and Gordon [35] have offered an additional vehicle oriented approach in which aircraft are scheduled one at a time. After each airplane is scheduled, the aircraft and the demand that it serves are removed from further consideration.

A very interesting dynamic programming approach is the pseudo-adaptive programming method of Devanney [36]. Devanney formulates the fleet scheduling problem in terms of a conventional dynamic program that provides a complete state space description of the system. Noting the hopeless computational requirements of such a program, Devanney forfeits exact state space information in favor of an aggregated, incomplete description of the system. Consequently, at each stage of the program, scheduling decisions are made in terms of probabilistic estimates of the resulting state transitions. Full development and evaluation of this promising method has not yet been realized.

To preview the contributions of this thesis, it will be helpful to refer to the categories along which the literature has been reviewed. The context of this research is oriented at no particular mode of transportation, but is abstractly concerned with the scheduling of vehicles of specified (average) speeds and capacities, operating over a fixed network. As indicated earlier, the purpose of scheduling is viewed in terms of planning as well as current operations. Synthesis of fleet timetables is the principal element of scheduling that is addressed here. Evaluation of timetables, consideration of vehicle inventory requirements, and the assignment of vehicles and passengers to trips, will all be intimately connected with the approach to timetable construction.

This dissertation contributes to the understanding of factors relevant to scheduling considerations. In particular, a model of traveler preferences for trips of alternative schedules and fares is offered in the analysis of behavior and satisfaction. Furthermore, a contribution is made towards the formulation of system objectives, by encoding traveler benefits in the economic terms of willingness to pay. This enables traveler benefits to be combined with costs and revenues into a flexible formulation that allows an objective function to be chosen for a variety of system orientations.

Of course, this thesis must be heavily concerned with the computational problem of timetable optimization. As indicated above, research efforts have taken various directions in the elusive quest for optimal solutions. Here, a new vehicle oriented approach is offered for the derivation of near optimal timetables.

Finally, this thesis considers the questions of data, uncertainty, and the choice among alternative systems. The organization of the dissertation is described below.

Chapter II presents the development of a spectrum of objective functions for systems operated by agencies with alternate goals. In conjunction with the formulation of objectives, the second chapter presents the development of a traveler preference model that characterizes the traveler as an economic consumer of transportation services.

Chapter III offers a discussion of the computational problem involved in optimizing the schedule of a transportation system. The infeasibility of straightforward optimization methods is demonstrated. Subsequently, a new procedure is developed in which the overall optimization is decomposed into a set of interrelated suboptimizations. This new method is presented in terms of a two station shuttle system. Extensions of the procedure are given in later chapters. Finally, Chapter III presents a discussion of the quality of solutions obtained using the new computational tool.

In Chapter IV, the computational method is supplemented by a framework for decision-making with uncertainty about traveler preferences and levels of potential demand. In particular, a Monte Carlo formulation is designed to be used in a Bayesian mode of analysis for choosing among alternative proposed transportation systems.

Chapter V presents an illustrative case study. Among the questions investigated here are those pertaining to different possible traveler markets, alternative fleet compositions and sizes, varying operator objectives, and alternate fare levels.

In Chapter VI, the computational method heretofore developed for a shuttle system, is extended to more complex cases. Expanded sets of decision options are considered, such as routing, local and express policies, and variation of fares.

Chapter VII presents a discussion of some of the "systems concepts" that have been molded here to synthesize solutions for the passenger transportation scheduling problem. In addition, the particular solution method developed in the context of transportation is characterized more abstractly as a "facility scheduling-user allocation" algorithm. Some illustrative examples that fit this framework, are formulated. Finally, Chapter VII offers a summary of results, and a discussion of future avenues for research and implementation.

Appendix A documents an interview procedure for obtaining data on traveler scheduling preferences. Appendix B displays the computer program used to implement the shuttle system computations.

CHAPTER II

FORMULATION OF OBJECTIVES AND MEASURES OF BENEFIT

2.1 Objectives

An explicit mathematical statement of system objectives is highly useful in the determination of good operating policies. That is, a quantitative measure or criterion of system performance facilitates the task of judging the merit of timetables, passenger accommodations, and fare policies. Since it is the intention here to develop an analytic tool of wide application within the framework of fixed schedule transportation systems, specifying an objective function will require consideration of the spectrum of motivations under which various transportation facilities are intended to operate.

Transportation systems range from purely private operations through publically owned, government-run facilities. Accordingly, system operator goals vary from profit maximization to maximization of user benefits and other social goals. In fact, a complete array of possible objective functions exists. Maximization of patronage* or user benefits minus costs, minimization of costs subject to level of service constraints, or maximization of profits subject to fare and schedule regulations, may all be reasonable objectives for particular systems. This dissertation deals with a wide variety of system objectives, although explicit constraints on costs, revenues, or benefits, are avoided. Aside from computational convenience, the principal reason for this restriction is that

---

* For example, maximization of patronage may be a valid objective for a city concerned with alleviating highway congestion by encouraging transit usage.

the present study is oriented toward the planning process, and the consideration of alternative systems. Hence, artificial constraints are to be discouraged in favor of considering a wide range of systems with corresponding cost and benefit levels. For example, rather than require maximization of patronage subject to a breakeven of revenues and costs, it is wiser to perform unconstrained maximization of patronage and consider the corresponding costs and patronage levels of the alternative systems. It may turn out that a sharp rise in patronage coincides with a small cost deficit, a contingency that may ultimately be judged best.

Two additional comments are in order, with regard to the question of constrained objectives in the context of the computational framework developed here. First, results can often be brought within desirable limits by modifying the parameters of the unconstrained objective. Second, it will sometimes be feasible to modify the computational process directly to include specific conditions.

To construct a flexible objective function, it is necessary to identify the important schedule-related variables that are common to a wide array of systems. Three components--operating costs, revenues or fares, and economic benefits to travelers--are taken here as the building blocks for the set of potential objectives. The first two components--costs and revenues--are self-explanatory. The third, traveler benefits, will be measured in terms of travelers' "willingness to pay" for trips offered in the system timetable.

Before proceeding to the mathematical formulation of objective functions, a short discussion about willingness to pay as a measure of user benefits is relevant here.* The question of measuring public benefits is an important concern of the theory of welfare economics. Two aspects of that theory are germane to this discussion. The first relates to the question of whether individual benefits may be aggregated to compile a measure of overall value. The second aspect concerns the kind of welfare judgement that is implied by using willingness to pay as a measure of benefit.

The question of aggregating individual benefits relates to the theory of utility, and the controversy of whether ordinal or cardinal utility measures are valid [38]. If the concept of cardinal utility is acceptable then the aggregation of the utilities of individuals can be viewed as a proper measure of total benefit. However, current economic thought does not favor the cardinal utility concept. In particular, economists prefer to avoid the dubious premise that the "satisfaction levels" of two or more individuals are additive. Instead, economists prefer to use ordinal utility functions and employ the concept of Pareto optimality** to make social judgements. However, the concept of Pareto efficiency leaves an important question unanswered. Specifically, there are normally an infinite number of efficient points at which an economy may operate, but the Pareto theory does not discriminate among them. Consequently, the thecry provides no help in evaluating alternate efficient points of operation.

---

* See Ref. [37], pp. 65-69, for a more complete discussion of this material.

** An economic allocation is Pareto optimal if it is not possible to make any individual better off without making someone else worse off.

Thus, in order to formulate an objective function that accounts for user benefits, it is necessary to find some index that overcomes the objection to interpersonal comparison of satisfaction levels, yet provides an explicit value scale with which to discriminate among economic alternatives. Willingness to pay is such a measure. Willingness to pay has the additional advantage, of course, that it is measured in the same units, dollars, as the other objective components, costs and revenues.

It is important, however, to be aware of the welfare criterion that is implicit in the use of willingness to pay. A person's willingness to pay does not necessarily correspond to his subjective valuation for a given commodity. As an example, in the context of transportation, a rich man may be willing to pay fifty dollars to take a given air trip from Washington to New York, while a less wealthy man may be willing to pay only ten dollars for the same trip. This does not imply that the trip means more to the rich man than his poorer counterpart. In fact, it may well be that ten dollars is a greater sacrifice to the poorer man than fifty dollars is to the rich man.

When willingness to pay is used as a measure of social benefit, then it is assumed that society values each individual's dollar equally, regardless of his income. This means the distribution of income is accepted as given. Under this assumption, the total benefit of a commodity (trip) is the sum of individual benefits. This is the usual formulation of benefit/cost analysis. In economic theory, this corresponds to the compensation principle which maintains that if net benefits exceed costs, then those who stand to gain from the economic project (trip) can compensate those who suffer from it, and still remain better off.

Note that use of willingness to pay does not preclude the possibility of making alternate welfare evaluations. For example, a person's willingness to pay can be weighted by a factor that depends on his income, so that the aggregation of benefits corresponds to a modified welfare criterion. Such an option could easily be incorporated into the development here.

To proceed to a precise formulation, the objective function will be taken as a linear combination of costs, revenues, and traveler benefits. If there are J trips in the schedule then the value of the objective function will be given by

$$F = \sum_{j=1}^{J} \left(-a_1 OC_j + \sum_{i \in A(j)} (a_2 \pi_j + a_3 WTP_j^i) \right) , \qquad (2\text{-}1)$$

where

J = total number of trips,

OC_j = operating cost of trip j,

π_j = fare charge for trip j,

$A(j)$ = set of passengers aboard trip j,

WTP_j^i = the maximum amount that passenger i is willing to pay for taking trip j,

a_1, a_2, a_3 = constant coefficients.

Thus,

$$F = -a_1 \text{ Total Operating Cost} + a_2 \text{ Total Revenues} + a_3 \text{ Total Willingness to Pay} . \qquad (2\text{-}2)$$

By assigning different values to the coefficients (a_1, a_2, a_3), a spectrum of objective functions can be formed. Using the values 0, $\pm$ 1, a number of objective measures with clear economic interpretations can be obtained. These are illustrated in Fig. 1. However, there is no technical reason to restrict the a's to these values. Any weighted combination of costs, revenues, and willingness to pay can be considered. In fact, adjustment of the a's may facilitate consideration of constraints.* For example, although the objective may be to maximize user benefits, it may be desirable to assign a small positive value to a_1 to discourage excessive costs. With so flexible a formulation, it will be possible to gain instructive insights into the effects of operator orientation on system scheduling policy.

2.2 Traveler Preferences

To attempt to optimize schedules relative to the desired objective, it is necessary to achieve an understanding of traveler behavior with respect to the consumption of transportation services. That is, it is important to learn how travelers choose among alternative services. In particular, if it is known how travelers discriminate among alternative trips, based on the schedule-related characteristics of the trips, then a basis is established for designing good scheduling policies. The first task, therefore, is to identify the schedule-related properties of a trip that consumers consider important.

At this point it is pertinent to review what various research efforts have found about the factors that affect the way a person chooses to travel. Refs. [40] through [47] present a long list of transportation attributes that are found to affect travel behavior. Among these

---

* See Ref. [39], Chapter 1, for a discussion of implementing constraints by changing the weights of component objectives.

(a) $a_1 = 0$

a_3 \ a_2	1	0	-1
0	Total Revenue		
1		Total Willingness To Pay	Consumer Surplus

(b) $a_1 = 1$

a_3 \ a_2	1	0	-1
0	Net Revenue (Profit)		
1		Total Willingness to Pay Minus Costs	Net Willingness to Pay Minus Costs

Figure 1. Array of Economic Objectives

(a) $a_1 = 0$ (b) $a_1 = 1$.

attributes, the following are important in the consideration of scheduling: distance, travel time, cost or fare, speed, schedule convenience and delay, reliability of destination achievement (i.e., the need to arrive on time), frequency of service, and necessity, time, and inconvenience of changing vehicles. In this study, the latter two factors will not receive much direct attention. Frequency of service will be more explicitly considered in terms of schedule convenience. Vehicle transfer will be mentioned briefly in Chapter VI.

Thus, three important schedule-related variables are to be focused on: fare, travel time, and schedule convenience. All but the last factor mentioned above, are implicitly accounted for by these three components. A consumer preference model for alternative trips, based on these "commodities" will now be developed.

The development strategy will be as follows. Willingness to pay will be assumed to be a decreasing function of trip duration (travel time) and schedule inconvenience. The schedule inconvenience of a trip will be defined in terms of its deviation from the traveler's "preferred schedule" for a trip of the same duration. The willingness to pay function will be developed in two steps. First, the contours of equal willingness to pay will be modeled. These contours specify the combinations of trip duration and schedule inconvenience among which the traveler is indifferent; the contours map directly into corresponding contours that specify the trip schedules (departure time, arrival time) among which the traveler is indifferent. Finally, the function to which the equi-value contours correspond, will be specified.

Consider the following notation:

t_d = departure time of a trip (from origin),

t_a = arrival time of a trip (at destination),

(t_d, t_a) = time schedule of a trip,

ϕ = duration (travel time) of a trip,

Δ = "schedule inconvenience" of a trip,

(Δ, ϕ) = composite trip impedance,

$WTP(\Delta, \phi)$ = maximum price the traveler would pay for a trip of impedance (Δ, ϕ),

π = fare charged for a trip.

Note that quantities ϕ and Δ are viewed as impedances to travel. That is, it is assumed that travelers prefer trips of shorter duration, and trips of less schedule inconvenience, and they will pay more for such trips. Furthermore, it is assumed that at some fixed price, π, a traveler will sacrifice some degree of schedule convenience for a faster trip, and, alternately, will give up some degree of trip speed for a more convenient schedule.

For example, suppose a traveler desires to depart on a trip from New York to Washington, at 4 p.m. Assume that he has three trips to choose from in the schedule. The first trip leaves at 4:15 p.m. and takes four hours. The second trip leaves at 4 p.m. and takes four hours. The third trip leaves at 4:15 p.m. and takes three and three-quarters hours. Presumably, the traveler would be willing to pay more for trips 2 and 3 than for trip 1. Further, he would choose among trips 2 and 3 on the basis of his trade-off value between a fifteen minute decrement of schedule inconvenience and a fifteen minute decrement in trip duration. If he valued the decrement of schedule inconvenience more highly than the decrement of trip duration, then he would be willing to pay more for trip 2 than trip 3, and vice versa.

The next step in the development of the consumer model is to define the travel impedances ϕ and Δ, in terms of the coordinates (t_d, t_a) of a trip. The trip duration is given by

$$\phi = t_a - t_d \ . \tag{2-3}$$

Quantification of schedule inconvenience requires a little more work. The inconvenience of a trip will be characterized in terms of its <u>deviation</u> from a traveler's "preferred schedule." It will be assumed that, given a trip of ϕ hours duration, a traveler has a preferred schedule (t_d^*, t_a^*) where

$$t_a^* - t_d^* = \phi \ . \tag{2-4}$$

Thus, the deviation of a trip (t_d, t_a) from a traveler's preferred schedule is just

$$\Delta = (t_a - t_a^*) = (t_d - t_d^*) \ . \tag{2-5}$$

Hence, given a trip (t_d, t_a), its schedule impedance characteristics are found from Eqs. (2-3) and (2-5), once the traveler's preferred schedule for a trip of duration ϕ is known. Note that a person's preferred schedule changes as ϕ changes. In order to understand the way in which a preferred schedule shifts, as a function of trip duration ϕ, the concept of a traveler's relative orientation toward departure time versus arrival time, will now be introduced.

First, consider the following definitions:

t^* = a traveler's preferred departure-arrival time for an "instantaneous" trip ($t^* = t^*_a = t^*_d$ at $\phi = 0$).

w = the relative importance, to the traveler, of departing near his preferred departure time versus arriving near his preferred arrival time, $0 \leq w \leq 1$.

A traveler wishing to adhere closely to some departure time will be assumed to have a value of w near one, while a traveler desiring to adhere to a time of arrival will have a w value close to zero. For example, a commuter on his way to work in the morning may need to adhere closely to a particular arrival time, and is therefore "arrival-oriented," ($w \doteq 0$). On the contrary, a businessman with an appointment in a distant city the next day may be much more concerned about when his flight departs than when it arrives at his destination; hence, he is "departure-oriented," ($w \doteq 1$).

With the above notions, a traveler's preferred schedule is given by

$$t^*_a = t^* + w\phi \ , \tag{2-6a}$$

$$t^*_d = t^* - (1-w)\phi \ . \tag{2-6b}$$

Thus, if a traveler were completely departure-oriented ($w = 1$), his preferred departure time t^*_d would remain fixed at t^*, and his arrival time preference t^*_a would shift according to the change in trip duration ϕ. The reverse occurs if $w = 0$. For values of w between 0 and 1, both t^*_d and t^*_a change with changes in ϕ.

To illustrate the meaning of Eqs. (2-6), consider a departure-oriented traveler who wishes to travel from New York to Washington. Suppose he prefers to leave New York at 10 a.m. If the trip takes four hours, his preferred schedule is $t_d^* = 10$ a.m., $t_a^* = 2$ p.m. If the trip takes only three hours, he prefers a trip schedule $t_d^* = 10$ a.m., $t_a^* =$ 1 p.m. On the other hand suppose that the traveler is arrival-oriented, and prefers to arrive in Washington at 1 p.m. His preferred schedule for the four hour trip is (9 a.m., 1 p.m.); his preference for the three hour trip is (10 a.m., 1 p.m.).

Eqs. (2-6), as written above, do not tell the whole story. It is probably more realistic to consider w, and perhaps even t^*, as functions of trip time ϕ. For example, consider the businessman in New York with an appointment in San Francisco at 11 a.m. (Eastern Standard Time) on the following day. If the flight takes six hours, then the businessman will prefer to leave immediately after work to take a 7 p.m. flight out of New York. He will not be particularly concerned with his time of arrival. Thus he will be departure-oriented, with preferred departure $t_d^* = t^* = 7$ p.m. Suppose, however, that the flight time were reduced to two hours. Under these circumstances, the businessman might re-evaluate his plans. He may prefer to go directly to his appointment the next day, rather than sleep in San Francisco overnight. In this case, he will become arrival-oriented with a preferred arrival time $t_a^* = t^* = 11$ a.m.

For relatively narrow ranges of trip time, however, the assumption of constant values for w and t^* appears reasonable. In any case, the option to consider w and t^* as functions of ϕ is always available. For simplicity, however, these parameters will be treated as constants for each traveler, here.

Using (2-6) and (2-2), the schedule inconvenience can be written

$$\Delta = t_a - t_a^* = t_a - t^* - w(t_a - t_d) \; , \tag{2-7a}$$

or

$$\Delta = t_d - t_d^* = t_d - t^* + (1-w)(t_a - t_d) \; . \tag{2-7b}$$

By defining the "perceived instantaneous schedule,"

$$t' = t_a - w(t_a - t_d) = t_d + (1-w)(t_a - t_d) \; , \tag{2-8}$$

(2-7) can be written more succintly as

$$\Delta = t' - t^* \; . \tag{2-9}$$

A word must be added here with respect to the sign of Δ. From (2-7) or (2-9) it is apparent that Δ may be positive or negative. This does not imply, however, that a negative inconvenience is a convenience. The sign of Δ refers to whether the schedule deviates from the preferred schedule on the "late side" or the "early side." In particular, if the trip leaves later than t_d^*, i.e. $t_d > t_d^*$, then $\Delta > 0$; if $t_d < t_d^*$ then $\Delta < 0$. Since the traveler may feel differently about these two types of inconvenience, it is important to recognize the distinction.

Now that the units of trip impedance have been established, it is possible to formulate the consumer willingness to pay function. The development will be considered in two steps. The first step is to formulate a traveler's "trade-off" behavior with respect to the two (undesirable) commodities Δ and ϕ.

Commodities Δ and ϕ are both measured in units of time. At first glance it may appear that there should be a one-to-one trade-off of travel minutes with schedule inconvenience minutes. That this is not generally the case is related to the fact that the traveler attaches different importance to the two impedances. A precedent for attaching separate values to distinct types of time impedance is given by Cherniack [48]. In his paper, automobile "running time" and "waiting time" are valued differently, for the purpose of computing congestion costs.

The trade-off behavior is modeled by contours of equal willingness to pay that specify the sets of combinations of (Δ,ϕ) values among which the traveler is indifferent. These equal willingness to pay contours are mappings of corresponding contours in the (t_d,t_a) plane that specify the trips (t_d,t_a) among which the traveler is indifferent. Eqs. (2-3) and (2-7) perform the required transformation from the (t_d,t_a) plane to the (Δ,ϕ) plane. Each "indifference curve" corresponds to a particular value of willingness to pay, namely the maximum price that the traveler would pay for each of the (Δ,ϕ) combinations (or equivalently, each of the trips (t_d,t_a)) on the curve. The next step in the construction of the consumer model is to establish the function for which the indifference curves are equi-value contours in the (Δ,ϕ) plane.

The function $WTP(\Delta,\phi)$, once specified, will be used to describe the traveler's preferences among alternative trips. The traveler will prefer that trip for which he is willing to pay the most. Furthermore, it will be assumed that a traveler finds a given trip "acceptable" if his willingness to pay equals or exceeds the fare for that trip.

2.3 Contours of Equal Willingness to Pay

The following definition will be useful for describing traveler indifference curves:

The additional increment of schedule inconvenience $(\delta\Delta)$ that the traveler is willing to accept in order to secure a unit decrease in trip duration, is called the marginal rate of substitution (mrs) of Δ for ϕ.

More precisely, consider a constant level of willingness to pay:

$$WTP(\Delta,\phi) = C \; . \tag{2-10}$$

Thus,

$$\frac{\partial \; WTP}{\partial \; \Delta} \, d\Delta + \frac{\partial \; WTP}{\partial \; \phi} \, d\phi = 0 \; , \tag{2-11a}$$

or equivalently,

$$WTP_{\Delta} d\Delta + WTP_{\phi} d\phi = 0 \; . \tag{2-11b}$$

Thus,

$$\text{mrs} = -d\Delta/d\phi = + WTP_{\phi}/WTP_{\Delta} = -\text{slope of indifference curve at } (\Delta,\phi) \; . \tag{2-12}$$

The mrs is likely to depend on the relative magnitude of ϕ and Δ. In particular, in a region where ϕ is high and Δ is low, the mrs of Δ for ϕ is likely to be large. Alternately, in regions of high Δ and low ϕ, the mrs may be small. Further, it is conceivable that the marginal rate of substitution varies with the absolute magnitudes of ϕ and Δ. That is, the trade-off behavior of the traveler in regions of low trip impedance may differ from his behavior in high impedance regions. For simplicity, however, it will be assumed that the mrs is insensitive to the absolute magnitude of trip impedance.

The use of indifference curves to represent the preferences of consumers is an integral part of classical economic theory.* Usually,

* See Ref. [49], Chapter 2.

indifference curves are drawn for preferences among desirable commodities, apples and pears for example. In such case, the indifference curves are convex from below, as illustrated in Fig. 2. The quantities Δ and ϕ are undesirable, however. Therefore, the indifference curves are convex from above, as in Fig. 3.* Mathematically speaking,

$$-d(mrs)/d\phi = d/d\phi(d\Delta/d\phi) = d^2\Delta/d\phi^2 \begin{cases} < 0, & \Delta > 0 , \\ > 0, & \Delta < 0 , \end{cases} \tag{2-13}$$

whereas

$$d^2A/dP^2 > 0 , \tag{2-14}$$

where A = apples, P = pears.

Eq. (2-13) is the precise characterization for the behavior of the marginal rate of substitution, as described above. In particular, (2-13) requires that the mrs of Δ for ϕ, i.e., the rate at which the traveler would incur an additional unit of inconvenience to secure a unit decrease in trip duration, increases with ϕ. Such behavior corresponds to indifference curves that are convex from above, as in Fig. 3.

Fig. 4 summarizes the discussion of the traveler willingness to pay model, to this point. The figure displays the contours of equal willingness to pay in the (t_d, t_a) plane. In addition, the loci of preferred schedules are shown for arrival and departure oriented travelers as well as a traveler of intermediate orientation. The area enclosed by the willingness to pay equal to fare contour and the "instantaneous trip line," defines the set of acceptable trips. Finally, willingness to pay is

---

* See Ref. [40] for an example of indifference curves for undesirable commodities. The commodities are travel time and operating costs for automobile travelers.

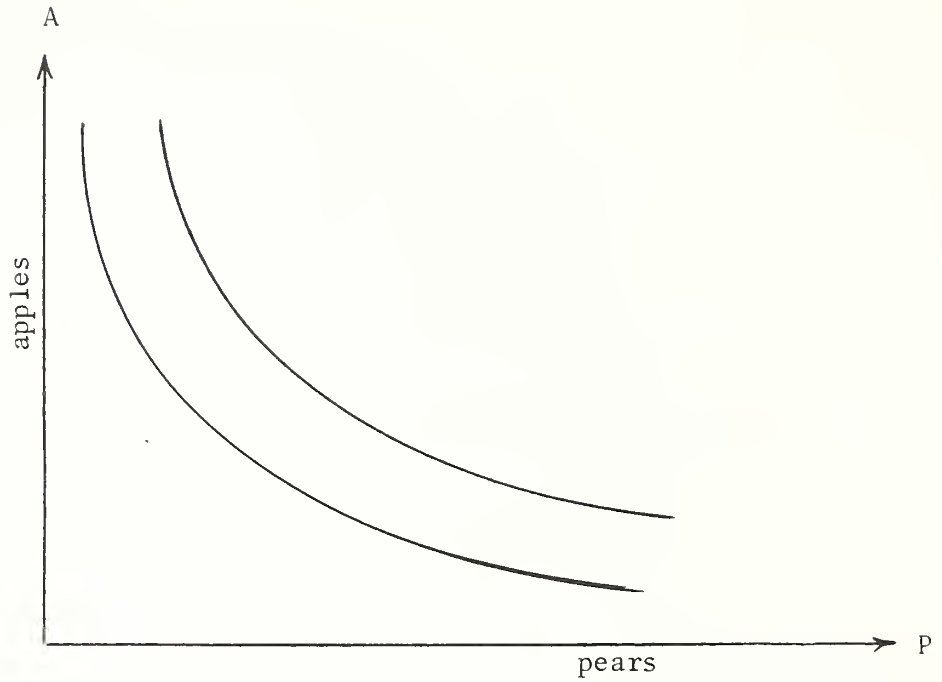


Figure 2. Indifference Curves for Desirable Commodities.

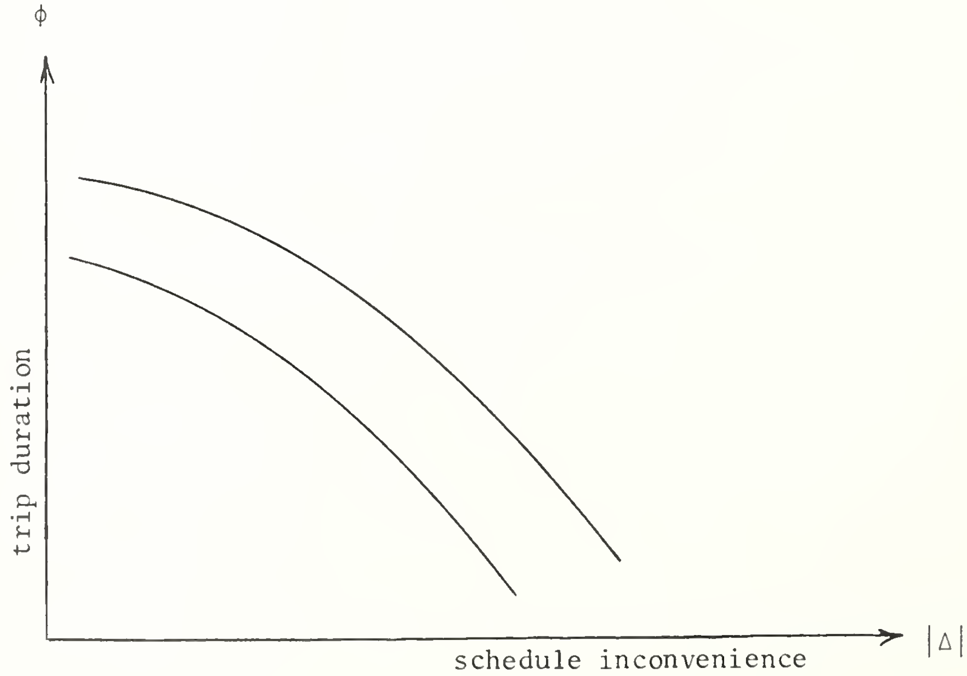


Figure 3. Indifference Curves for Undesirable Commodities.

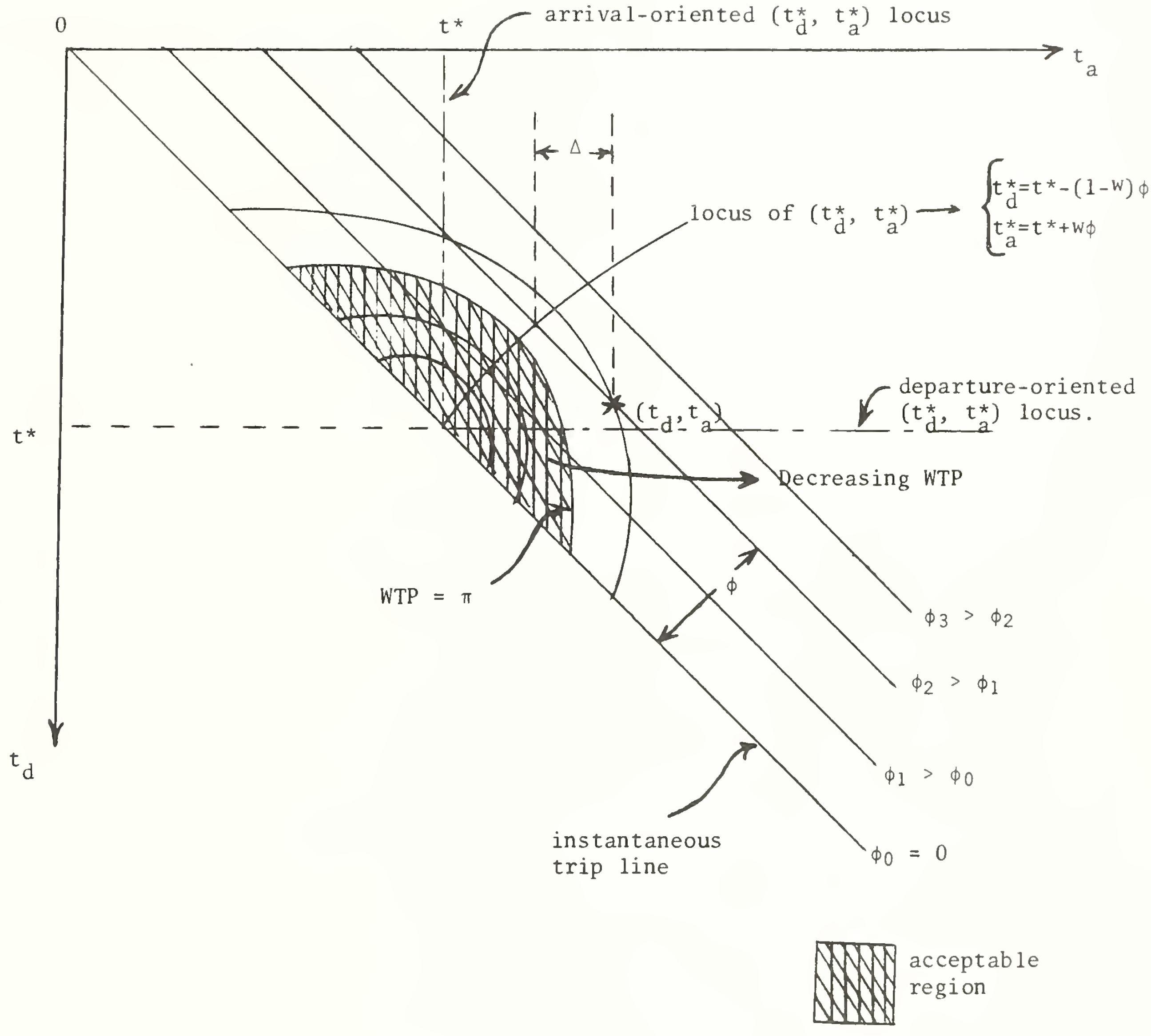


Figure 4. Willingness to Pay Contours in the (t_d, t_a) Plane.

shown to decrease as ϕ increases, and as trips deviate from the preferred schedule locus. Fig. 5 displays the transformation of the indifference contours into the (Δ, ϕ) plane.

In the development that follows, specific mathematical forms are proposed, with which to verify the formulas or calibrate the parameters. A small experimental effort to verify the concepts proposed here is exhibited in Appendix A. Procedures for collecting data are mentioned at the end of this chapter. Further discussion of data and parameter estimation is given in Chapter IV.

Although particular mathematical forms are required in order to proceed with the discussion, the methodology developed does not hinge on them. Thus, this analysis could proceed with any of several indifference curve and willingness to pay formulas. Finally, the mathematical models actually specified here are motivated by current knowledge about consumer behavior with respect to transportation services.

A family of curves that behave according to the notions described above is given by

$$\frac{\Delta^{n+}}{a_+^{n+}} + \frac{\phi^{n+}}{b_+^{n+}} = 1 \qquad \Delta \geq 0 \ , \tag{2-15a}$$

$$\frac{(-\Delta)^{n-}}{a_-^{n-}} + \frac{\phi^{n-}}{b_-^{n-}} = 1 \qquad \Delta < 0 \ , \tag{2-15b}$$

where $n+, n- \geq 1$, and

$$b_+ = b_- \ . \tag{2-16}$$

Eqs. (2-15) distinguish between the two kinds of schedule inconvenience, $\Delta > 0$ and $\Delta < 0$. Presumably, the traveler may trade off

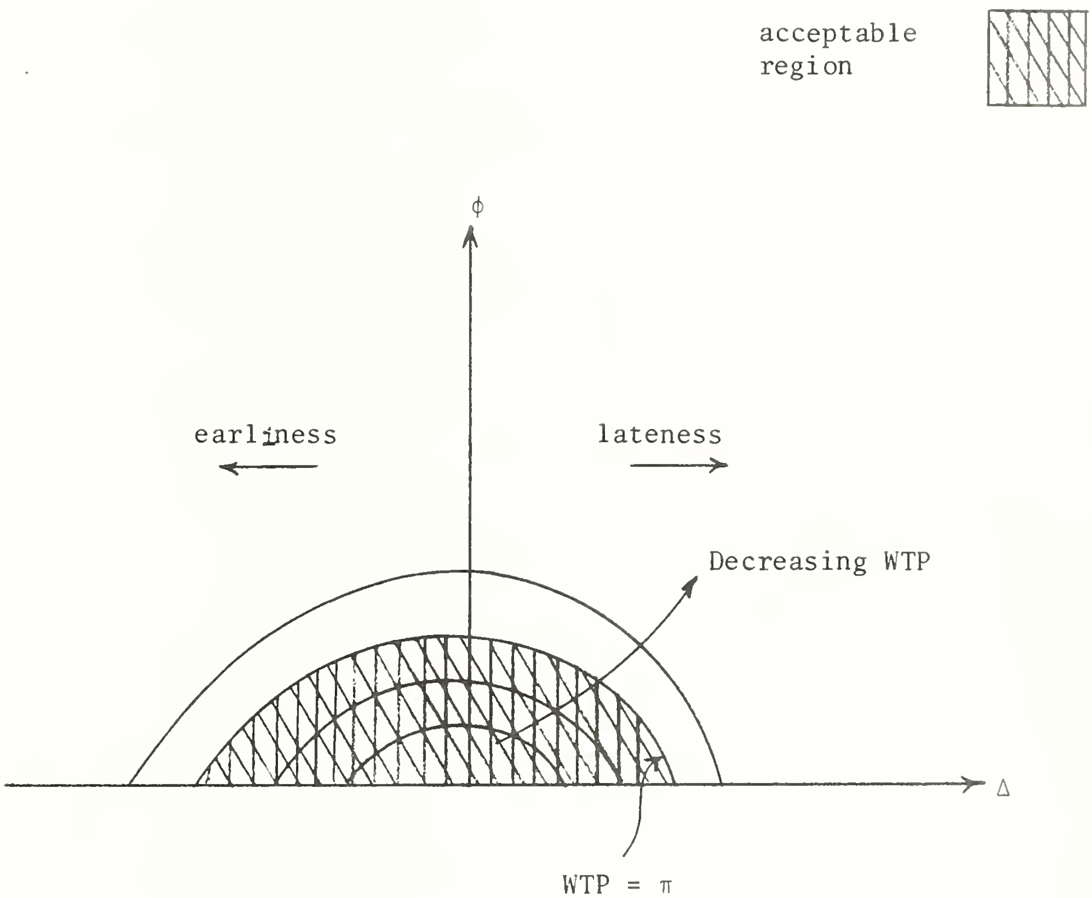


Figure 5. Willingness to Pay Contours in the (Δ, ϕ) Plane.

"earliness" against trip duration differently than "lateness," as indicated in Fig. 5. Eq. (2-16) states the requirement that the curves be continuous at $\Delta = 0$. A sufficient condition that (2-16) be satisfied

Another kind of asymmetry can be considered, in addition to the earliness-lateness dichotomy. In particular, it may be observed that the scheduling of a trip can cause travelers to experience "rushing" and "waiting." Below, the rushing-waiting concept will be formulated mathematically, and it will be shown that such an interpretation of scheduling inconvenience is consistent with the original definition.

As postulated above, for a given trip duration ϕ, the traveler has a preferred schedule (t_d^*, t_a^*). Furthermore, any (imperfect) trip $(t_d, t_a) \neq (t_d^*, t_a^*)$ of length ϕ must either be "early" $(t_d < t_d^*,\ t_a < t_a^*)$ or "late" $(t_d > t_d^*,\ t_a > t_a^*)$, with respect to the traveler's preferences. The earliness or lateness of a trip will cause the traveler to rush or wait, depending on his relative orientation toward departure time or arrival time.

Consider the two extremes of traveler orientation, departure-orientation $(w = 1)$ and arrival-orientation $(w = 0)$. For the departure-oriented traveler, an early trip means that he must rush, and a late trip means that he must wait. For example, suppose a departure-oriented traveler prefers to leave at 8 a.m. If the trip departs at 7:30 a.m., he must eat a hurried breakfast and dash to make the trip. On the contrary, if the trip leaves at 8:30 a.m., he may become impatient at having to leave later than he would like. The reverse behavior holds for an arrival-oriented traveler. If the trip is early, he must wait some time at his destination. If the trip is late, he will be rushed to keep his appointment on time.

Now, not all travelers have extreme departure-arrival orientations. Thus, there is usually an element of rushing and an element of waiting associated with every imperfect trip. To analyze the general case, the degree of waiting at each end of the trip will be specified, and negative waiting will be interpreted as rushing. At the departure end of the trip, the inconvenience is given by,

$$\Delta_d = (t_d - t_d^*) \ . \qquad (2\text{-}17a)$$

For the arrival end of the trip,

$$\Delta_a = (t_a^* - t_a) \ . \qquad (2\text{-}17b)$$

Thus, if $\Delta_d > 0$, then the trip is late and waiting occurs (at the origin). If $\Delta_d < 0$, then the trip is early and rushing obtains (at the origin). The reverse holds for Δ_a (at the destination), in conformity with the previously noted interpretations of early and late trips by departure and arrival oriented travelers. In other words, Δ_d and Δ_a always have opposite signs--if waiting occurs at the origin then rushing occurs at the destination. The relative inconvenience of these components will be given by the orientation weighting parameter w.

At first glance, it may appear that an overall measure of trip inconvenience should be the sum of Δ_d and Δ_a. However, since it is not reasonable to assume that rushing on one end of the trip cancels waiting at the other end, the two components must be considered separately. To examine rushing and waiting more explicitly, let

$$R = -\min(w\Delta_d, (1-w)\Delta_a) = \text{rushing} \ , \tag{2-18a}$$

and

$$W = \max(w\Delta_d, (1-w)\Delta_a) = \text{waiting} \ . \tag{2-18b}$$

Thus, the willingness to pay function may be written as $WTP(W,R,\phi)$, so that waiting, rushing, and trip duration are explicitly considered as separate components of trip impedance. In this representation, equi-value contours of willingness to pay are three dimensional over the positive octant of (W,R,ϕ) space.

To demonstrate the relationship of the (R,W) formulation of schedule inconvenience to the original Δ formulation, consider the behavior of R and W as Δ changes. If the absolute value of Δ increases, i.e., if an early trip is made earlier, or a late trip is made later, then both R and W increase in magnitude (for $0 < w < 1$). Thus, Δ is a travel impedance measure consistent with the rushing-waiting interpretation of inconvenience. In fact, the rushing-waiting interpretation is merely a more refined analysis of the earliness-lateness dichotomy. Specifically, the relative importance of earliness versus lateness becomes a function of the traveler's arrival-departure orientation, since it is that orientation that causes the traveler to interpret earliness-lateness in terms of rushing and waiting.

To demonstrate this mathematically, assume indifference curves of $WTP(W,R,\phi)$ of the form,

$$\frac{\phi^n}{b^n} + \frac{W^n}{a^n} + \frac{R^n}{c^n} = 1 \ , \qquad \phi, W, R, \geq 0 \ . \tag{2-19}$$

Note from the definitions that,

$$\Delta_d = \Delta \ , \tag{2-20a}$$

and

$$\Delta_a = -\Delta \ . \tag{2-20b}$$

Thus,

$$W = \max(w\Delta_d, (1-w)\Delta_a) = \max(w\Delta, -(1-w)\Delta) \ , \tag{2-21a}$$

and

$$R = -\min(w\Delta_d, (1-w)\Delta_a) = -\min(w\Delta, -(1-w)\Delta) \ . \tag{2-21b}$$

Now, consider late trips for which $\Delta > 0$. Then,

$$W = w\Delta \ , \tag{2-22a}$$

and

$$R = (1-w)\Delta \ . \tag{2-22b}$$

Accordingly, Eq. (2-19) becomes,

$$\frac{\phi^n}{b^n} + \frac{(w^n c^n + (1-w)^n a^n)\Delta^n}{(ac)^n} = 1 \ . \tag{2-23}$$

Similarly, for early trips, $\Delta < 0$ and,

$$W = -(1-w)\Delta \ , \tag{2-24a}$$

and

$$R = -w\Delta \ . \tag{2-24b}$$

In this case, (2-19) reduces to,

$$\frac{\phi^n}{b^n} + \frac{((1-w)^n c^n + w^n a^n)(-\Delta)^n}{(ac)^n} = 1 \ . \tag{2-25}$$

Thus, (2-19) reduces to the original formulation (2-15) where $n+ = n- = n$,

$b_+ = b_- = b$, and,

$$a_+ = \frac{ac}{(w^n c^n + (1-w)^n a^n)^{1/n}} , \tag{2-26a}$$

and,

$$a_- = \frac{ac}{((1-w)^n c^n + w^n a^n)^{1/n}} . \tag{2-26b}$$

For simplicity, subsequent discussion will be restricted to symmetric indifference curves of the form

$$\frac{|\Delta|^n}{a^n} + \frac{\phi^n}{b^n} = 1 , \qquad n \geq 1 , \quad \phi \geq 0 . \tag{2-27}$$

Such symmetry ignores the difference between earliness and lateness, and between rushing and waiting. A set of such curves for different values of parameter n is illustrated in Fig. 6. Note that for any indifference curve here, $a = |\Delta|_{max}$ and $b = \phi_{max}$. Hence (2-27) implies that for a given level of willingness to pay, the traveler is willing to endure a maximum level of schedule inconvenience, and a maximum trip duration. Further, $|\Delta|_{max}$ is tolerable only if $\phi = 0$, and ϕ_{max} is acceptable only if $\Delta = 0$.

The case $n = 1$ represents straight line contours, corresponding to a traveler with a constant trade-off value between ϕ and Δ, independent of (Δ,ϕ). For $n > 1$, the mrs behaves as described earlier. In general, the exponent n increases with the degree of sensitivity that the traveler exhibits toward extreme values of Δ or ϕ.

As mentioned earlier, it will be assumed here that a traveler exhibits the same trade-off characteristics in regions of high trip impedance (Δ,ϕ) as he does in regions of low impedance. In other words, it is

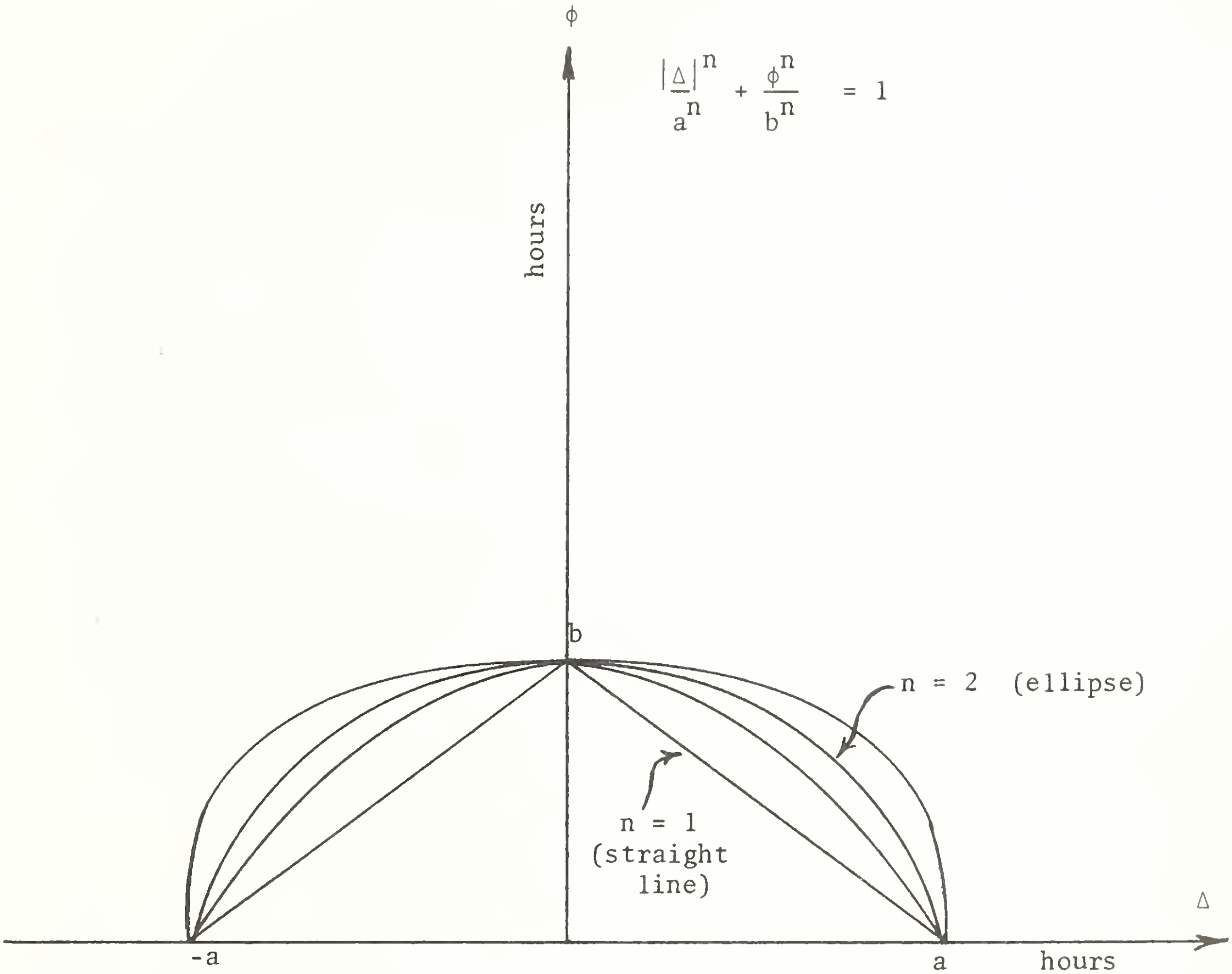


Figure 6. Family of Equal Willingness to Pay Contours.

assumed that the behavior of the mrs is independent of the level of willingness to pay. This assumption is equivalent to requiring that successive willingness to pay indifference curves have the same "shape," in the following sense. Two curves will be said to have the same shape if one curve can be transformed into the other by multiplying each of the variables by the same constant. Such an operation is merely a change of "scale" for the curve. For curves described by Eq. (2-27), the shape is maintained by preserving the value

$$s = b/a \quad , \tag{2-28}$$

for all contours. Thus (2-27) can be rewritten as

$$\frac{|\Delta|^n}{a^n} + \frac{\phi^n}{(sa)^n} = 1 \ . \tag{2-29}$$

To show that contours having different values of a, have the same shape under (2-29), consider the following two curves.

$$\text{(i)} \quad \frac{|\Delta|^n}{a_1^n} + \frac{\phi^n}{(sa_1)^n} = 1 \ , \tag{2-30a}$$

$$\text{(ii)} \quad \frac{|\Delta|^n}{a_2^n} + \frac{\phi^n}{(sa_2)^n} = 1 \ , \tag{2-30b}$$

where $a_1 = Ka_2$. Then curve (i) can be transformed into curve (ii) by multiplying ϕ and Δ by K:

$$\frac{K^n|\Delta|^n}{a_1^n} + \frac{K^n\phi^n}{(sa_1)^n} = \frac{|\Delta|^n}{a_2^n} + \frac{\phi^n}{(sa_2)^n} = 1 \ . \tag{2-31}$$

For the cases $n = 1$ and $n = 2$, it is easy to see that keeping s constant is equivalent to maintaining the slope of the straight lines (n=1) or the eccentricity of the ellipse (n=2):

a. straight lines:

$$\text{slope} = \pm s = \pm b/a \ . \qquad (2\text{-}32a)$$

b. ellipse:

$$e = (a^2 - b^2)^{1/2}/a = (1 - (b/a)^2)^{1/2} = (1 - s^2)^{1/2} \ . \qquad (2\text{-}32b)$$

Thus, constant s implies constant e.

With the formulation of (2-29), the value of willingness to pay attached to each equal-WTP contour in the (Δ,ϕ) plane can be modeled as a function of parameter a. Before doing this, however, it is interesting to see how the contour of equal WTP in the plane of (Δ,ϕ) transforms into the (t_d,t_a) plane. This transformation is shown below, for the straight line case with $t^* = 7$, $w = 0.75$. See Fig. 7.

The transformation is defined by Eqs. (2-3) and (2-7a), which are repeated below, for convenience.

$$\phi = t_a - t_d \ , \qquad (2\text{-}3)$$

$$\Delta = t_a - t^* - w(t_a - t_d) \ . \qquad (2\text{-}7a)$$

The equation to be transformed is

$$\frac{|\Delta|}{a} + \frac{\phi}{b} = 1 \ , \qquad (2\text{-}33)$$

or

$$\frac{\Delta}{a} + \frac{\phi}{b} = 1 \qquad \Delta \geq 0 \ , \qquad (2\text{-}34a)$$

and

$$-\frac{\Delta}{a} + \frac{\phi}{b} = 1 \qquad \Delta \leq 0 \ . \qquad (2\text{-}34b)$$

Substitution of (2-3) and (2-7a) into (2-34a) yields

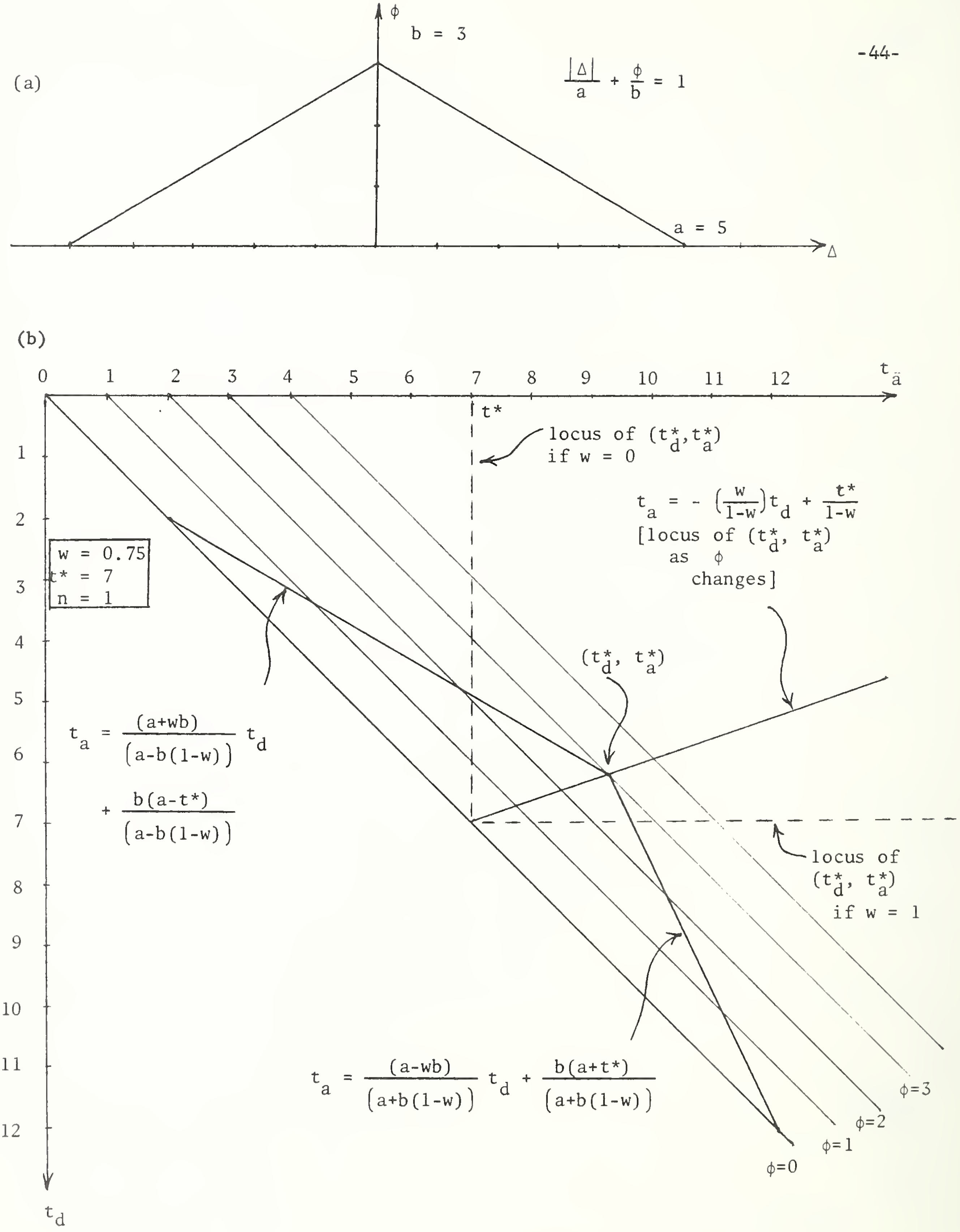


Figure 7. Transformation of a Linear Willingness to Pay Contour
(a) Plane of (Δ, ϕ) (b) Plane of (t_d, t_a).

$$\frac{(t_a - w(t_a - t_d) - t^*)}{a} + \frac{t_a - t_d}{b} = 1 \ , \tag{2-35a}$$

or

$$t_a = \frac{(a - wb)}{(a + b(1 - w))} t_d + \frac{b(a + t^*)}{(a + b(1 - w))} \ . \tag{2-35b}$$

Substitution of (2-3) and (2-7a) into (2-34b) yields

$$\frac{-(t_a - w(t_a - t_d) - t^*)}{a} + \frac{t_a - t_d}{b} = 1 \ , \tag{2-36a}$$

or

$$t_a = \frac{(a + wb)}{(a - b(1 - w))} t_d + \frac{b(a - t^*)}{(a - (1 - w)b)} \ . \tag{2-36b}$$

Lines (2-35b) and (2-36b) intersect where $\Delta = 0$:

$$\Delta = t_a - t^* - w(t_a - t_d) = 0 \ , \tag{2-37a}$$

or

$$t_a = \frac{-w}{1 - w} t_d + \frac{t^*}{1 - w} \ . \tag{2-37b}$$

Eq. (2-37b) describes the locus of points $(t_d, t_a) = (t_d^*, t_a^*)$ of desired schedules, as trip duration ϕ is varied. The point of intersection of (2-35b), (2-36b), and (2-37b), found by solving (2-35b) and (2-36b) simultaneously, and substituting into (2-37b) to check the equality, is given by

$$t_d = t^* - b(1 - w) \ , \tag{2-38a}$$

$$t_a = t^* + bw \ . \tag{2-38b}$$

Eqs. (2-38) define the apex of a triangular contour in the (t_d, t_a) plane.

2.4 Variation of Willingness to Pay

The final step in developing the willingness to pay function is to specify the variation in willingness to pay as the parameter a changes. Recall that a is a parameter that labels equal willingness to pay contours with successively larger values as the curves proceed outward from the origin in the (Δ,ϕ) plane. Thus, a can be viewed as a measure of composite (Δ,ϕ) impedance. The following functional form appears reasonable:

$$WTP(a) = D\exp(-a^2/\alpha^2) \quad . \tag{2-39}$$

The family of functions given by (2-39) is illustrated for various values of the parameter α, in Fig. 8. In that figure, a_{cutoff} is the level at which price equals willingness to pay; higher prices are unacceptable to the traveler. By setting $WTP(a)$ equal to π in (2-39), a_{cutoff} is found to be,

$$a_{cutoff} = \alpha\sqrt{\ell n(D/\pi)} \quad . \tag{2-40}$$

The form (2-39) satisfies the following plausible conditions. First, there is a central region of a-values in which a traveler's willingness to pay is highly sensitive to changes in the level of impedance (a). Further, there are two "saturation" regions, at very low levels and very high levels of impedance, in which a traveler's willingness to pay is relatively insensitive to marginal improvements or decrements in service.

Curves of the shape (inverted S) of those in Fig. 8 have been used before to describe the diversion of travelers from one mode of travel to another, as a function of changes in the relative fares or trip times

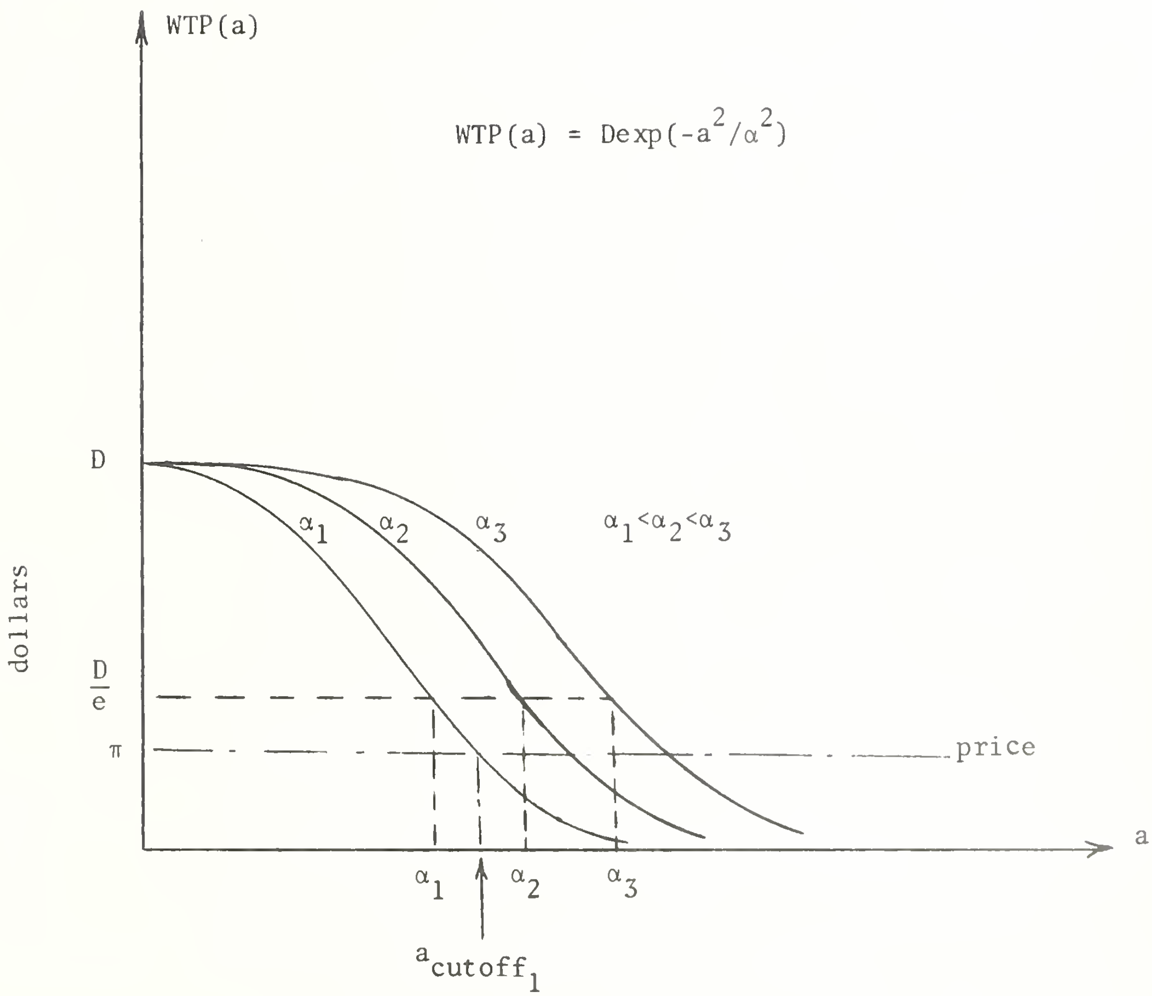


Figure 8. Willingness to Pay as a Function of Composite Impedance, a .

among the modes [50,51]. Since the diversion of travelers to a given mode is directly related to their willingness to pay for service on that mode, the use of curves similar to (2-39) has some precedent.

One problem with the curves of Fig. 8 is that the infinite "tails" imply that if the ride were free, a traveler would be willing to endure an arbitrarily large amount of trip impedance. This contingency is of no concern here, however, because the curves will never be used in that region. In particular, the scheduling algorithm developed in Chapter III will be used only with fares above a positive lower bound.

To complete the discussion, the full functional form of the willingness to pay function, over the variables (Δ,ϕ) and (t_d,t_a), will be obtained by solving for a and substituting into Eq. (2-39). Thus,

$$a = (|\Delta|^n + \phi^n/s^n)^{1/n} , \qquad (2\text{-}41)$$

and

$$WTP(\Delta,\phi) = D\exp(-\alpha^{-2}(|\Delta|^n + \phi^n/s^n)^{2/n}) . \qquad (2\text{-}42)$$

Substituting for Δ and ϕ using equations (2-3) and (2-7b), yields

$$WTP(t_d,t_a) = D\exp\left\{-\alpha^{-2}\left[(t_d - t^* + (1-w)(t_a-t_d))^n + (t_a-t_d)^n/s^n\right]^{2/n}\right\} .$$

$$(2\text{-}43)$$

In summary, it is assumed that Eqs. (2-42) and (2-43) completely specify the traveler's willingness to pay in terms of the trip impedance characteristics, and the trip timetable, respectively.

2.5 Data

This chapter has presented the development of a traveler preference model involving a set of parameters $(t^*, w, D, \alpha, n, s)$. Use of the methods developed in this thesis is contingent on the acquisition of data that facilitate the evaluation of these parameter values. In the present context, data acquisition may be considered from two perspectives. First, one may contemplate estimating parameters from data compiled from observations of actual traveler behavior. This will be called the "observation" approach. On the other hand, data might be obtained experimentally, through carefully designed interview sessions with active or prospective travelers. The latter will be termed the "interview" approach.

Design and execution of a data collection program is outside the scope of this dissertation. However, it is the intention here to discuss the alternative approaches, and to indicate the steps required for their implementation. In the following paragraphs, previous research work concerned with the experimental determination of preference models by interviews, will be reviewed briefly. The interview approach requires the construction of indifference curves and preference functions of individual consumers, by offering the experimental subjects choices among alternative commodity bundles. Consideration of the observation approach is deferred until Chapter IV, since that discussion relies on methods presented later. The utilization of information obtained by both approaches, to estimate the distribution of preference parameter values for the overall traveler population, is also considered in Chapter IV.

Relatively little experimental work has been done on the interview approach to the determination of indifference curves and preference functions. Three efforts, widely spaced in time, constitute the principal contributions to this area. In 1931 Thurstone [52] pioneered the field with experiments to construct indifference curves and preference ordering functions among hats, shoes, and overcoats. In 1950, Rousseas and Hart [53] experimented with preference orderings among alternate combinations of quantities of eggs and bacon, using responses of several individuals to construct composite indifference curves. Recent work (1968) has been done by MacCrimmon and Toda [54], to expand on the Thurstone methodology, with experiments involving dollars, ball point pens, and French pastries.

The Thurstone and MacCrimmon-Toda methods are based on "equivalence-dominance" techniques that require a subject to give his preferences between a given commodity bundle and a "reference" bundle. For example, Thurstone used eight hats and eight pairs of shoes as a reference, and asked his subjects to accept or reject various other combinations (e.g. ten hats and four pairs of shoes) in place of the reference. In this manner, he was able to draw an indifference curve separating the acceptance region from the rejection region. By varying the reference bundle, a series of indifference curves, corresponding to different levels of satisfaction, were obtained for each individual subject. The Thurstone and MacCrimmon-Toda graphical results are impressive in their consistency and conformity to _a priori_ expectations of behavior.

Little has been done along these lines in the sphere of transportation research. In conjunction with this research, the author designed an interview technique to investigate the traveler preference functions. This technique, along with illustrative experimental results, is documented in Appendix A.

There are certainly advantages and disadvantages to both data collection approaches. On the one hand, interview methods may be designed to elicit the exact information that is required for a particular study. In addition, the interview method may explore a traveler's reactions to a wide variety of (existent and non-existent) transportation alternatives. On the other hand, it is questionable whether verbal responses correspond closely with actual behavior.* Data on observable behavior is more valid in this respect. In addition, observable data is probably less expensive to obtain. However, it is frequently difficult to secure actual travel data relevant to many of the transportation situations of interest. For example, no observations can be made to test traveler reaction to "drawing board" systems. In short, "real world" observations must be made in an environment filled with uncontrolled variables whose effects are difficult to discern. All these comments are relevant to the current state of information in transportation research.

Thus, careful consideration should be given to each approach. This author believes that research into both areas is only at the beginning stages. Experimentation is currently proceeding in both the laboratory interview method,** and the real world observational approach.*** Subsequent developments in these areas will likely shed more light on the question.

In summary, empirical investigation of traveler preference behavior

---

* See Ref. [55] for a dissenting view on the validity of experimental determination of indifference curves.

** See Ref. [54].

*** For example, the U.S. Department of Transportation Demonstration Project for the Northeast Corridor.

has been lacking. Appropriate data can conceivably be obtained in two ways--direct observation of actual behavior, and laboratory interviews. The previous discussion has reviewed earlier research on the interview approach, the potential of which is largely unknown. Although the interview method attempts to determine preference curves for individual subjects, the results are to be considered as outputs of a random process that describes the total population of consumers (travelers). The random process itself may vary according to overall shifts or trends in tastes and technology. Use of sample data, obtained either by interview or observation, to estimate the character of the overall random process, is discussed at the end of Chapter IV.

CHAPTER III

THE COMPUTATIONAL METHOD

3.1 Introduction

Stated simply, the computational problem with which this dissertation is concerned, is the following:

Given a population of travelers, a fleet of vehicles with specified speeds and capacities, and a network over which travel takes place, find the vehicle schedules and corresponding passenger assignments that maximize the value of the chosen objective.

This deceptively straightforward proposition may blind the unwary observer to the host of implicit problems that it involves. These problems will become clear as the characteristics required of the solution method are discussed further.

First, it is clear that a flexible solution method is desired. The subject matter of concern here is characterized by a substantial degree of heterogeneity: Travelers exhibit a variety of preferences, vehicles have different speeds and capacities, networks have various topologies, and operators have a number of different motivations and policies by which they run their systems. Hence, a mathematical formulation is sought that is capable of adapting to a reasonably wide array of transportation situations.

The structure of the passenger scheduling problem exhibits several additional mathematical complications, which limit the choice of tools that may be applied for its solution. In particular, the scheduling problem involves non-linearities, discontinuities, constraints, discrete and continuous variables, and a high degree of dimensionality. To see that these characteristics do indeed manifest themselves, it is only necessary to consider the following aspects of the problem:

1. Traveler preferences among trips are, in all likelihood, non-linear with respect to travel time, inconvenience, and cost. It would be highly restrictive and unrealistic to assume otherwise. Further, since a traveler is either willing or unwilling to ride a particular trip, there will be some cut-off point (say a given level of fare at any particular level of trip impedance) beyond which he can receive no benefit from the trip. These two phenomena point to the fact that the objective function is non-linear and discontinuous.

2. Timetables conform to the available number and speed capabilities of the vehicles. Also, each passenger is allowed but one vehicle assignment for his trip. Hence, the problem involves significant constraints.

3. Trip departure times may range over the continuous interval of values within the specified scheduling period (0, T). On the other hand, passenger assignments are discrete (in the mathematical sense). That is, a passenger must be assigned to one particular trip (or no trip at all). Hence, optimization is required over both integer and continuous variables.

4. A timetable must be provided for each vehicle, and an assignment for each passenger. Hence, the dimensionality of the problem is very large.

Finally, in addition to the properties mentioned above, it is important to recognize that the scheduling problem is <u>sequential</u> in nature. For instance, the time at which a vehicle is scheduled to depart a network node affects when it is next available for service. Thus the problem involves significant interdependencies among the variables.

In all, the passenger transportation scheduling problem requires a flexible solution method, capable of optimizing a highly irregular objective function over a complex, many-dimensional, constrained set of solution vectors. The search for such a method is begun below.

3.2 Computational Tools

Having surveyed the properties that characterize the passenger transportation scheduling problem, it is now appropriate to consider the arsenal of tools with which solutions may be approached. In searching for a computational method, it will be found that aspects of various optimal programming methods can contribute to solving particular phases of the problem. However, overall optimization will require a more heuristic approach.

Consider first, the disciplines of linear and non-linear programming. No attempt will be made here to scrutinize these methods. Suffice it to say that there are two principal objections to employing such programming methods as an overall framework. First, it is clear that the workable algorithms of linear or quadratic programming are too restrictive with respect to the mathematical forms they require. However, if the more general programming theory is applied (Lagrange multipliers, etc.) it becomes evident that the size of the problem, in terms of the number of variables and constraints, becomes unmanageable. In any case, the programming methods do not guarantee global (absolute) optimality of solutions except in the rather special case of convex or concave functions, defined over a convex set of solution vectors. Hence, it is well to look beyond these programming methods in the search for an overall computational structure.

However, programming methods do provide a number of techniques and insights that will be helpful in synthesizing a solution method. For example, the solution of large problems by programming is sometimes accomplished by dividing the problem into a group of interrelated smaller problems and performing a coordinated set of suboptimizations. The decomposition principle of linear programming [56] and the manipulation of Lagrange multiplier shadow prices in non-linear problems [57] illustrate this principle. Here, it will be seen that separating the timetables of individual vehicles, and the departure scheduling process from the passenger assignment process, will lead to a feasible solution method.

A second aspect of programming that will prove valuable is the application of specialized integer programming algorithms. In particular, use will be made of optimum assignment and maximum network flow algorithms to compute aggregate passenger benefits.

Next, consider the methodology of dynamic programming. More than a computational technique, dynamic programming is a way of thinking, particularly with respect to sequential decisions. The principle of optimality, embodied in the basic recursive equation, is so general as to allow the present problem to be formulated in such a way that mathematical forms need not be restricted, the problem's sequential character is naturally embodied, constraints are easily included, and solutions can be guaranteed optimal. However, the "hooker" is that conventional dynamic programming quickly becomes computationally infeasible as the dimensionality of the problem increases. In the following discussion, it will be seen that a dynamic programming framework is not adequate in the present context. However, it will be instructive here to attempt such a formulation in order to demonstrate the dimensionality problem, and to set up a structure that will lead to the algorithm for optimizing the timetables of individual vehicles.

Consider the following notation:

$(0, T)$ = time period during which the system is to operate.

t = continuous time variable, $0 \leq t \leq T$.

Let the period $(0, T)$ be divided into K small discrete intervals of width Δt, such that $\Delta t = T/K$. Then

k = discrete time variable, such that $t \doteq k\Delta t$.

k takes on values $(0, 1, 2, \ldots, K)$.

Hence, optimization over the continuous variable t is to be approximated by optimization over a fine mesh discrete time grid.

Let

s_j = station from which the next departure of vehicle j will occur,

k_j = next time at which vehicle j can depart from s_j ,

V = total number of vehicles in the fleet ,

$\tau_j = k_j - k$ = number of time intervals (from present time k) until the j^{th} vehicle can be dispatched from s_j .

The <u>state</u> of the system is described by the vector $\underline{x} = (\underline{s}, \underline{\tau})$ where

$$\underline{s} = (s_1, s_2, \ldots, s_V) ,$$

and

$$\underline{\tau} = (\tau_1, \tau_2, \ldots, \tau_V) .$$

Note that the "state of the passenger population" has not been included in $\underline{x}$. Keeping track of traveler locations directly is, of course, highly infeasible. Some auxiliary method of accounting for passengers would have to be developed to make the conventional dynamic program workable. It will be shown, however, that even if this consideration is ignored, the dynamic program becomes impractical. Hence, further discussion of passenger accounting will be deferred until the synthesis of the final computational method.

The <u>decision</u> (or control) vector is given by $\underline{c} = (c_1, c_2, \ldots, c_V)$, where

$$c_j(k) = \begin{cases} \theta & \text{if vehicle } j \text{ is "enroute" at time } k \text{, and no decision is required,} \\ 0 & \text{if vehicle } j \text{ is at a station at time } k\text{, and is directed to "hold" there, at least until time } k+1 \text{,} \\ m & \text{if vehicle } j \text{ is directed to "dispatch" toward station } m \text{ at time } k. \end{cases} \tag{3-1}$$

Thus, each component $c_j(k)$ may take on one of $(M+1)$ values, where M is the number of stations in the system. Of course, not all decision values are applicable to a vehicle at any particular time. For example,

$$c_j(k) = \theta \quad \text{if } \tau_j > 0 ,$$

and

$$c_j(k) \neq \theta \quad \text{if } \tau_j = 0 . \tag{3-2}$$

Let $s_j = q$, and

L^j_{qm} = number of time intervals (less one) required by vehicle j , to go from its present station (q) to station m .

Define the <u>transition</u> vector $\underline{g}(\underline{x},\underline{c},k) = (g_1, g_2, \ldots, g_V)$ as follows:

$\underline{g}(\underline{x},\underline{c},k)$ = state at time $k+1$ if decisions $\underline{c}$ are made at time k, when the state is $\underline{x}$.

Thus, with notation $x_j = (s_j,\tau_j)$, the situation $s_j = q$ leads to

$$g_j(\underline{x}_j,\theta,k) = (q,\tau_j - 1) \; , \tag{3-3a}$$

$$g_j(\underline{x}_j,0,k) = (q,0) \; , \tag{3-3b}$$

and

$$g_j(\underline{x}_j,m,k) = (m,L^j_{qm}) \; . \tag{3-3c}$$

Let

$b(\underline{x},\underline{c},k)$ = the <u>immediate benefit</u> to the system, if decisions $\underline{c}$ are chosen at time k, when the state is $\underline{x}$.

The immediate benefit is computed as the sum of contributions to the objective function resulting from dispatched trips corresponding to decision vector $\underline{c}(k)$. Hence

$$b(\underline{x},\underline{c},k) = \sum_{j=1}^{V} b_j(\underline{x}_j,c_j,k) \; , \tag{3-4}$$

where

$$b_j(\underline{x}_j,c_j,k) = 0 \qquad \text{if} \quad c_j = \theta \quad \text{or} \quad 0 \; , \tag{3-5a}$$

and

$$b_j(\underline{x}_j,c_j,k) = -a_1 OC_{jm} + \sum_{i\varepsilon A} (a_2 \pi_{jm} + a_3 WTP^i_{jm}) \quad \text{if} \quad c_j = m \; . \tag{3-5b}$$

Notation:

A = set of passengers who would ride vehicle j to station m at time k ,

OC_{jm} = operating cost of the trip to m by vehicle j ,

π_{jm} = fare for trip to m aboard vehicle j ,

WTP^{i}_{jm} = willingness to pay of passenger i for trip to m aboard vehicle j .

Two additional quantities need to be considered before proceeding to the basic iterative equation:

$$J = \sum_{k=0}^{K} b(\underline{x},\underline{c},k) = \text{criterion function},$$

and

$I(\underline{x},k)$ = maximum total benefit it is possible to achieve, from time k through time K, if the system is in state $\underline{x}$ at time k.

The fundamental recursive equation is

$$I(\underline{x},k) = \max_{\underline{c}} \left\{ b(\underline{x},\underline{c},k) + I(\underline{g}(\underline{x},\underline{c},k),k+1) \right\} . \tag{3-6}$$

Given boundary values $I(\underline{x},K)$, Eq. (3-6) may be iterated backwards in time, from $k = K$ through $k = 0$, to obtain $I(\underline{x},0) = \max\left\{J\right\}$. The optimal schedule (trajectory) is found by proceeding in the forward direction, from $k = 0$ to $k = K$, following the maximizing decisions $\underline{c}(k)$ at each stage.

Turning to the basic computational problem, consider the amount of computation and computer storage required to solve equation (3-6). The amount of computation depends primarily on the number of terms to be calculated and compared in finding the maximum at each time k for each state $\underline{x}$. Suppose there are $V(\underline{x})$ vehicles at stations when the system is in state $\underline{x}$. For each such vehicle j , there will be some number of feasible decisions, say $M_{\underline{x}}^{j}$, that can be made when the system is in state $\underline{x}$. $M_{\underline{x}}^{j}$ is the number of stations directly accessible to vehicle j from its current station (in state $\underline{x}$). The value of $M_{\underline{x}}^{j}$ depends on the connectivity of the network, and is less than or equal to M, the total number of stations. The total number of different decision vectors in (3-6) when the fleet is in state $\underline{x}$, is then $\prod_{j=1}^{V(\underline{x})} M_{\underline{x}}^{j}$. The total number of possible decisions, found by summing over all states $\underline{x}$, is $\sum_{\underline{x}} \prod_{j=1}^{V(\underline{x})} M_{\underline{x}}^{j}$. Thus, the total computation time for the backwards iteration is $\left(\sum_{\underline{x}} \prod_{j=1}^{V(\underline{x})} M_{\underline{x}}^{j} \right) Kt_c$, where t_c is the time per single decision alternative computation.

Now consider the computer storage requirements. The results of the maximization in (3-6) for each $\underline{x}$ need to be recorded. To obtain them, it is necessary to save all values $I(\underline{g},k+1)$ at the succeeding stage k+1. Thus, the dominant storage requirement is twice the number of different states. Recall that $\underline{x} = (\underline{s}, \underline{\tau})$ where $\underline{s}$ can take on M^V possible values. Further, let τ_j take on one of L_j+1 values, 0 through L_j. If any two stations are at least L^* time intervals apart, then a lower bound on the number of states is $X = M^V(L^*+1)^V$.

It is apparent that the storage and computation time requirements increase rapidly with the size of the system. Consider the following illustration:

Let $V = 5$ vehicles, $M = 4$ stations, $K = 100$ time intervals,

$t_c = 10^{-6}$ seconds, $L^* = 4$ time intervals.

Then,

$$X = 4^5\ 5^5 = 3.1 \cdot 10^7 \text{ states !}$$

Suppose,

$$M_{\underline{x}}^{j} = 3 \quad \text{for all vehicles } j \text{ at stations when the system is in state } \underline{x},$$

and let the total number of decisions, $\sum_{\underline{x}} 3^{V(\underline{x})}$, be approximated by $3X$.

This estimate corresponds to an average of one vehicle at a station, per state. Then,

$$\text{Computation Time} \doteq 3 \cdot 3.1 \cdot 10^{7} \cdot 100 \cdot 10^{-6} = 9.3 \cdot 10^{3} \text{ sec}$$
$$= 2.59 \text{ hours !}$$

Thus, even for this modest example, the computational requirements are entirely unreasonable. However, as mentioned earlier, the dynamic programming formulation turns out to be highly useful in synthesizing the final computational method. In particular, it will provide the basis for timetable optimization for single vehicles. In addition, a technique known as "successive approximation" [16], often useful in reducing the computational problems of dynamic programs, will help guide the formulation of the final procedure.

It is now appropriate to turn to a class of methods that will be called "general iterative improvement procedures." These methods have one important unifying characteristic. They all start from some initial solution, and monotonically improve on that solution, step by step, until no further improvement appears to be possible. There are several examples of these types of procedures. Often, a method of this kind can be assured to lead to an optimum solution. A few examples of these methods are discussed below. Iterative improvement will provide the basis for the optimization scheme to be developed here.

There is an extensive literature in the area of iterative "search" techniques [58]. Among the best known methods in this field are the "gradient" methods. Gradient methods are based on the fact that the direction of the gradient of a scalar function is the direction of the maximum rate of increase of that function. Hence, gradient methods proceed by following the gradient at each successive solution point. When this is impossible, because of constraints that disallow movements along the gradient, "locally best" directions are chosen. In sum, each solution point produces an improvement in the value of the function, over its value at the previous point. However, except where the problem possesses the appropriate convexity properties, gradient methods cannot be guaranteed to lead to a globally optimum solution, or even to a local optimum. Fig. 9 illustrates the gradient method, using three different initial solutions. Only one of these solutions leads to the global optimum.

Another well known improvement procedure, mentioned earlier in connection with dynamic programming, is called "successive approximation." The idea of this approach is the separation of a multivariable optimization problem into successive single-variable optimizations. Thus, the solution of an n-dimensional problem is reduced to the solution of a sequence of one-dimensional problems. This technique, sometimes called the "one-at-a-time" method [59], is subject to similar pitfalls as the gradient methods. Hence, solutions cannot be assured to be optimal. Fig. 10 illustrates solution by successive approximation, for four different initial solutions. Note here that constraints can cause the solution method to go awry, as in the case of initial solution x_{40}. This is also possible for gradient methods.

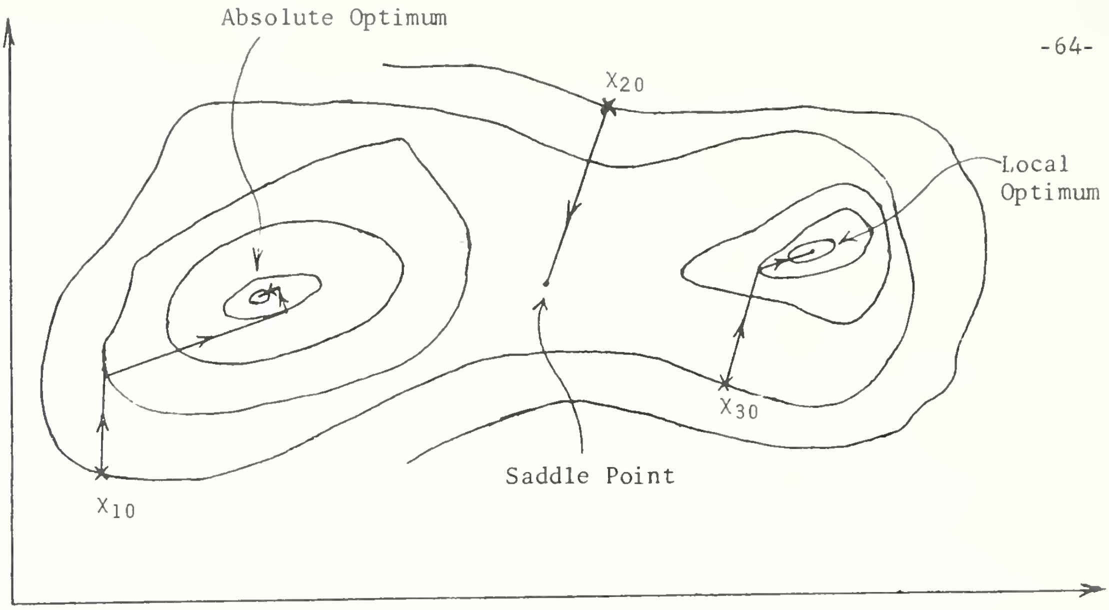


Figure 9. Gradient Method.

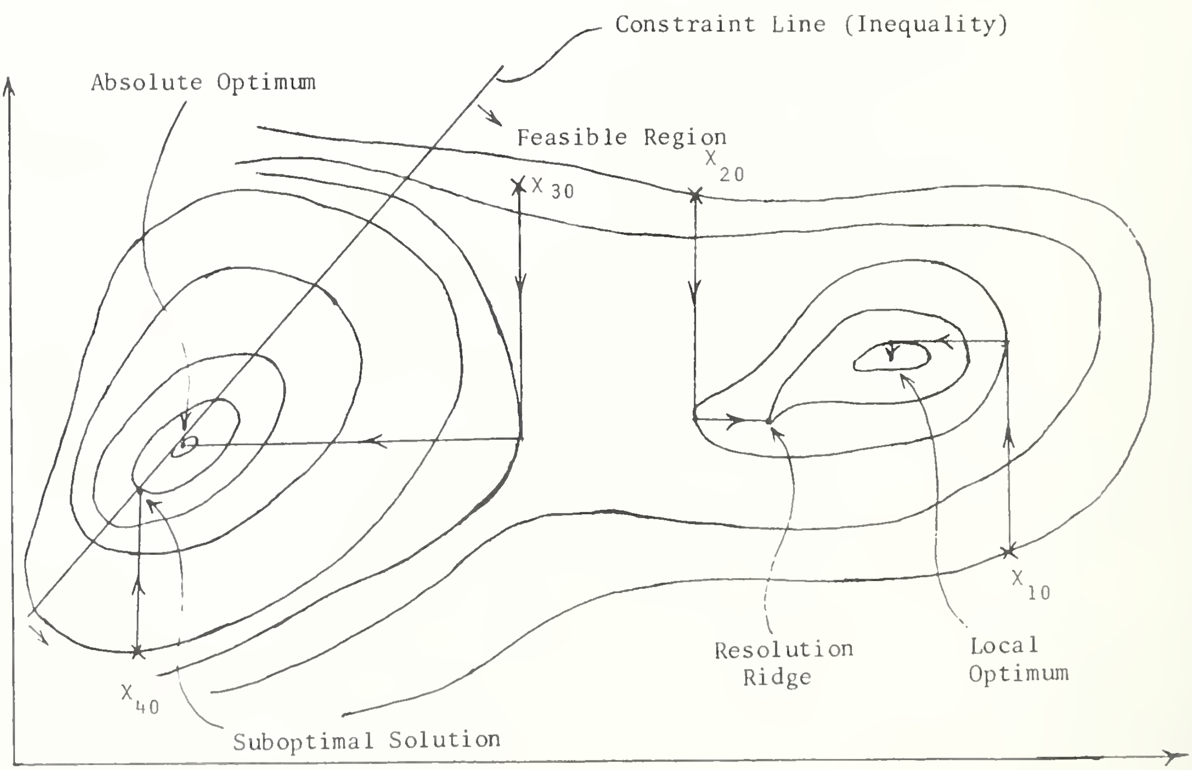


Figure 10. Successive Approximation.

A third interesting iterative improvement algorithm, for determining policies under uncertainties described by a Markov model, has been developed by Howard [60]. Here, a policy is defined as a set of strategies or actions, one for each state of the system. Successive policies are developed by alternating between two steps: 1) Given a policy, compute the objective function values that correspond to it. 2) Using the values corresponding to the present policy, develop a better policy. Howard shows that each successive policy is more valuable than its predecessor, and that the iterative procedure always converges to the optimal policy.

In summary, it is evident that heuristic iterative improvement methods hold out great promise as well as some peril. The basic idea of these methods is of wide application and generality. As long as a way can be found to improve solutions at every step, such a procedure may be tried. However, one must also be aware of the pitfalls that often accompany these methods. Like any serial scheme, the question of convergence must be investigated. Furthermore, as illustrated by Figs. 9 and 10, claims about the optimality of solutions must be carefully scrutinized. As a rule, it may be necessary to try several initial solutions before developing some assurance that the best obtained solution is acceptable.

This concludes the general discussion of optimization methods. Below, a computational procedure for the transportation scheduling problem is developed. The framework is that of an iterative improvement algorithm. The overall problem is analyzed into a series of interrelated subproblems. Several of the methods discussed earlier, are applied to these subproblems.

The dynamic programming formulation developed earlier in this chapter, encountered two major difficulties. First, taking direct account of the "state" of the passenger population was found to be infeasible. Second, it was demonstrated that computational requirements increase rapidly with the size of the system, particularly with the number of vehicles (V). Suppose, however, that attention is restricted to a single transportation vehicle. Furthermore, imagine that a certain portion of the traveler population is "assigned" to this vehicle. In addition, assume that traveler time preferences are sufficiently sharp, relative to vehicle system capabilities, so that it is impossible for any one traveler to find acceptable both of two successive trips by the same vehicle.* Under these conditions, a compact algorithm will be developed for optimizing the schedule of a single vehicle. This algorithm will become an integral part of the general iterative method for scheduling a fleet of vehicles. Developments in this chapter are restricted to a single link, "shuttle" network. See Fig. 11. Extensions to more intricate networks are deferred until Chapter VI.

3.3 The Single-Vehicle Algorithms

Consider the one-way, single vehicle system illustrated in Fig. 12.** Although this system is of little practical consequence, it will be a useful device for developing the important relationships for more complex single vehicle systems.

Let

N = number of travelers wishing to go from the origin to the destination, within the scheduling period $(0,T)$,

F = duration of the forward trip, from origin to destination,

---

* The mathematical statement of this assumption is developed later in this chapter.

** This system is analyzed in the author's paper, Ref. [29].

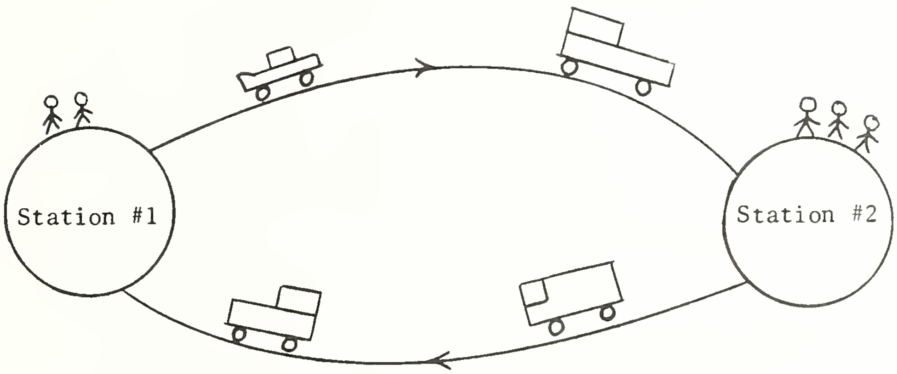


Figure 11. The Two-Station System.

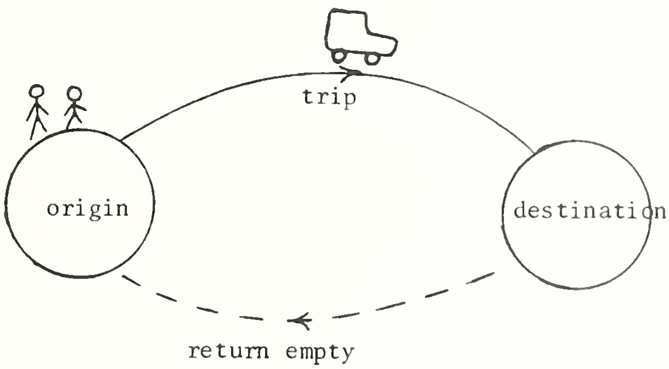


Figure 12. One-way, Single Vehicle System.

R = duration of the return trip (including turnaround time at both ends of the link).

F and R are integer quantities, assumed constant, and measured in discrete time units (Δt). Passengers departing at time k will arrive at time k+F. F+R is the minimum number of time units separating successive departures from the origin.

Recall that the state vector is given by $\underline{x} = (\underline{s},\underline{\tau})$. Here,

$$\underline{x} = (\underline{s},\underline{\tau}) = (s_1,\tau_1) = (1,\tau_1) \Rightarrow \tau_1 = k_1 - k \ , \qquad (3\text{-}7)$$

since there is but one vehicle and one possible station from which departures can take place. To simplify the development, consider the slightly less economical* state description

$\underline{x} = k_1$ = time at which the next departure from the origin is possible.

Thus,

$$\underline{x}(k) = k_1 \geq k \ . \qquad (3\text{-}8)$$

The decision vector becomes

$$\underline{c}(k) = c_1(k) = \begin{cases} \theta & \text{if } k_1 > k, \\ 0 & \text{if } k_1 = k \text{ and "hold" is ordered,} \\ 1 & \text{if } k_1 = k \text{ and "dispatch" is ordered.} \end{cases} \qquad (3\text{-}9)$$

The transition function is given by

$$\underline{g}(\underline{x},\underline{c},k) = g(k_1,c_1,k) = \begin{cases} k_1 & \text{if } k_1 > k \ , \\ k+1 & \text{if } k_1 = k \text{ and } c_1(k) = 0 \ , \\ k+F+R & \text{if } k_1 = k \text{ and } c_1(k) = 1 \ . \end{cases} \qquad (3\text{-}10)$$

The immediate benefit function is

---

* k_1 takes on more possible values than k_1-k .

$$\underline{b}(\underline{x},\underline{c},k) = b(k_1,c_1,k) = \begin{cases} 0 & \text{if } c_1 \neq 1 \,, \\ -a_1 OC' + \sum_i (a_2\pi + a_3 WTP^i) & \text{if } c_1 = 1 \,. \end{cases} \tag{3-11}$$

OC' is the operating cost of the round trip from origin to destination and back, WTP^i is passenger i's willingness to pay for the trip leaving at k and arriving at $k+F$, and π is the fare for that trip. The sum $\left(\sum_i\right)$ is over all passengers i who would leave the origin in the vehicle at time k. The question arises as to which of the N travelers are to be aboard the vehicle for that trip. This is an important point that will be addressed shortly.

Let

$$b_1(k) = -a_1 OC' + \sum_i (a_2\pi + a_2 WTP^i) \,. \tag{3-12}$$

Using the formulation given above, the basic dynamic programming equation becomes

$$I(k_1,k) = \begin{cases} 0 + I(k_1,k+1) & \text{if } k_1 > k \,, \\ \max \begin{cases} 0 + I(k+1,k+1) \\ b_1(k) + I(k+F+R,k+1) \end{cases} & \text{if } k_1 = k \,. \end{cases} \tag{3-13}$$

This equation may be simplified by considering the case $k_1 > k$ more closely:

$$I(k_1,k) = I(k_1,k+1) \qquad \text{if } k_1 > k \,. \tag{3-14}$$

Thus,

$$I(k_1,k) = I(k_1,k+1) = I(k_1,k+2) = \ldots = I(k_1,k_1) \,. \tag{3-15}$$

That is,

$$I(k_1, k) = I(k_1, k_1) \qquad \text{for} \quad k_1 \geq k \ . \tag{3-16}$$

Therefore, states for which $k_1 \neq k$ (i.e., where the vehicle is not at the origin), can be ignored. Consequently, the following simplification in notation is possible:

$$\begin{aligned} I(k_1,k) &\Rightarrow I(k) \ , \\ c(k_1,k) &\Rightarrow c(k) \ , \\ g(k_1,c,k) &\Rightarrow g(c,k) \ , \\ b(k_1,c,k) &\Rightarrow b(c,k) \ . \end{aligned}$$

Hence, Eq. (3-13) can be rewritten as

$$I(k) = \max \begin{cases} 0 + I(k+1) & \text{"hold," } \ c(k) = 0 \ , \\ b_1(k) + I(k+F+R) & \text{"dispatch," } \ c(k) = 1 \ , \end{cases} \tag{3-17}$$
$$\text{for} \quad 0 \leq k \leq K \ .$$

The logic of this equation is simple. The equation states that if the vehicle is at the origin at time k, then there are two options--to "hold" the vehicle at the origin until $k+1$ (at least), or to "dispatch" it at time k. The option that maximizes $I(k)$, is to be picked. If the hold option is taken, there is no immediate benefit, and $I(k)$ is just the maximum potential benefit that could be gained after remaining at the origin until the next time interval $k+1$, namely $I(k+1)$. If the dispatch option is chosen, $I(k)$ is equal to the sum of an immediate benefit $b_1(k)$, plus the maximum benefit achievable after returning to the origin at time $k+F+R$. ($F+R$ is the duration of a round trip.)

Because Eq. (3-17) involves stages $F+R$ units apart, a <u>set</u> of boundary conditions such as

$$I(k) = 0 \qquad \text{for} \quad K \leq k \leq K+F+R \ , \tag{3-18}$$

is required. These particular conditions merely state that there is no further benefit to be achieved if the vehicle returns to the origin after the end of the scheduling period. Other boundary conditions may be used, according to the problem context. Eq. (3-17) may now be used recursively, beginning at $k = K-1$, and working backwards to $k=0$, to obtain $I(0)$, the maximum total benefit of the full scheduling period. At each step in the process, the maximizing decision (hold or dispatch) is recorded.

Subsequently, the optimal vehicle schedule may be retrieved by proceeding in the forward direction, from $k = 0$ to $k = K$, following the optimal decisions. At each time k, if a hold decision is encountered then no departure is scheduled, and consideration is moved to the next time interval, $k+1$. If a dispatch decision is found, however, a departure is recorded, and consideration is advanced to that time $F+R$ time units later, when the vehicle returns to the origin.

Two additional questions concerning the current algorithm remain to be answered. First, as mentioned earlier, a determination must be made as to which of the N passengers are to ride a given trip. This issue arises in conjunction with the computation of $b_1(k)$, the immediate benefit of a departure at time k, in Eq. (3-17). Two separate considerations are involved--"eligibility" and "priority." Recall from Chapter II, that a passenger is termed eligible for a trip if he is willing to pay the fare. A rule must be provided, however, to decide which of the eligible travelers should be boarded, in case the number of these travelers exceeds the vehicle capacity.

From the beginning of this discussion it has been assumed that no traveler may be eligible for each of two trips separated in time by more than one round trip. Hence, there is no concern here about which of two trips by the same vehicle, that a given traveler may prefer. Under this stipulation, the optimal boarding procedure for the single link, single vehicle system is:

Board passengers in decreasing order of the values $(a_2\pi + a_3 WTP)$ for the trip.

Thus, to compute $b_1(k)$, travelers must be ordered by their value levels for the trip leaving at k and arriving at $k+F$. The situation is not so simple for systems with more complex transportation networks. The priority boarding procedure is modified accordingly, in Chapter VI. It should be noted here, of course, that "real world" travelers do not necessarily follow an optimal boarding rule. Thus, the procedure here actually provides an upper bound approximation to the resultant benefit.

The second question concerning the current algorithm is that of developing a mathematical constraint to represent the assumption that no traveler be eligible for two successive trips by the same vehicle. Recall from Chapter II that there is a finite region in the (t_d, t_a) plane in which a passenger is willing to travel. See Fig. 3. That area is bounded by the "instantaneous trip" line,

$$\phi = t_a - t_d = 0 , \tag{3-19}$$

and the willingness to pay contour that corresponds to the price of the trip, π. The speed of the vehicle must be sufficiently restricted, relative to the passenger's "eligible area," such that the passenger cannot be eligible for two consecutive trips by the same vehicle. Thus, the algorithm will not generate schedules based on having passengers board the vehicle more than once.

The minimum time in which a vehicle may complete a round trip (from either node) and prepare to depart again, is given by,

$$r = \phi + \rho + \tau_1 + \tau_2 , \tag{3-20}$$

where

r = round trip time,

ϕ = duration of the forward trip,

ρ = duration of the return trip,

τ_x = minimum turnaround time at station x .

The above variables are continuous. Note that the turnaround times at each node are considered separately here. For the one-way system, Eq. (3-17) ensures that turnaround at the destination station (number 2) is always minimal.

The purpose of this derivation is to ensure that r is always greater than the width of the traveler's eligible area. Suppose that the values ϕ, ρ, τ_1, τ_2, and r are given, and that the fare for the (forward) trip is specified at π. The indifference curve corresponding to WTP = π, is shown in Fig. 13. The intercepts of this curve are (a_π, b_π) where $b_\pi = sa_\pi$. If ϕ is given, then the maximum tolerable schedule inconvenience $|\Delta|_{max}$ is given by,

$$|\Delta|_{max} = (a_\pi^n - (\phi/s)^n)^{1/n} . \tag{3-21}$$

Since

$$|\Delta| = t_d - t_d^* = t_a - t_a^* , \qquad (2\text{-}5)$$

then the "width" of the set of acceptable trips is $2|\Delta|_{max}$, since acceptable trip departures may range from,

$$t_d^+ = t_d^* + |\Delta|_{max} , \qquad (3\text{-}22a)$$

to

$$t_d^- = t_d^* - |\Delta|_{max} . \qquad (3\text{-}22b)$$

Thus, the requirement is

$$2|\Delta|_{max} < \phi + \rho + \tau_1 + \tau_2 , \qquad (3\text{-}23a)$$

or

$$2\left(a_\pi^n - (\phi^n/s^n)\right)^{1/n} < \phi + \rho + \tau_1 + \tau_2 . \qquad (3\text{-}23b)$$

Now, a_π is the intercept of the indifference curve such that,

$$WTP(a_\pi) = D\exp(-(a_\pi/\alpha)^2) = \pi . \qquad (3\text{-}24)$$

Thus,

$$a_\pi = \alpha(\ln(D/\pi))^{1/2} . \qquad (3\text{-}25)$$

Substituting (3-25) into (3-23b) yields,

$$2[\alpha^n(\ln(D/\pi))^{n/2} - (\phi/s)^n]^{1/n} < \phi + \rho + \tau_1 + \tau_2 . \qquad (3\text{-}26)$$

If $\phi = \rho$ and $\tau_1 = \tau_2 = \tau$, then (3-26) reduces to,

$$(\alpha^n(\ln(D/\pi))^{n/2} - (\phi/s)^n)^{1/n} < \phi + \tau . \qquad (3\text{-}27)$$

Fig. 14 is a plot of inequality (3-27), using the values $\alpha = 2$, $n = 1$, $s = 1$, $\tau = 0.1$. The figure illustrates the ranges of fares,

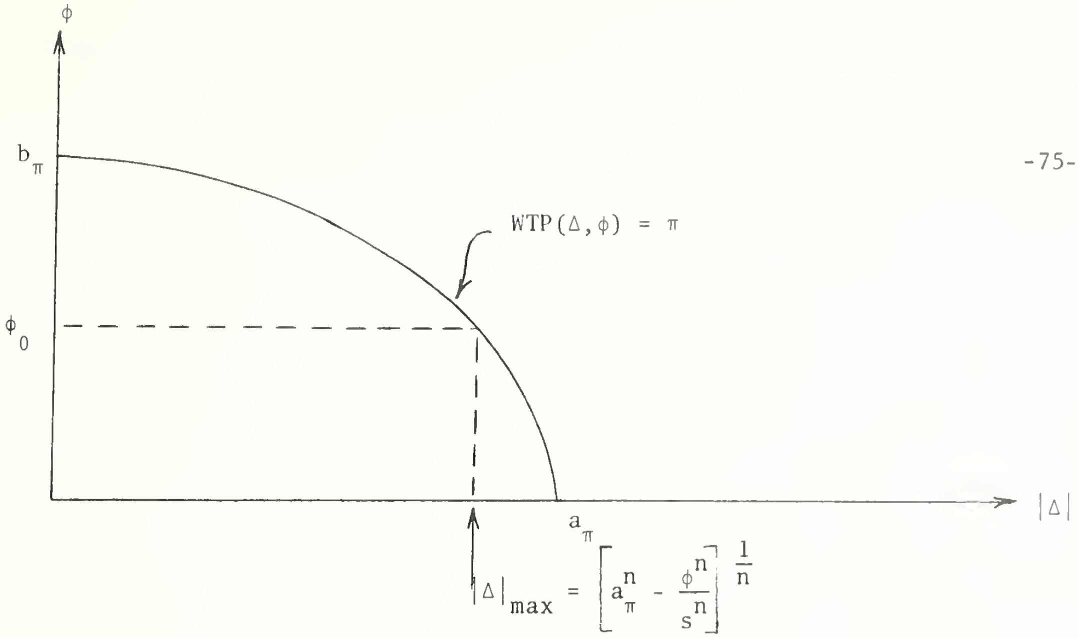


Figure 13. The Willingness to Pay Equals Price Contour.

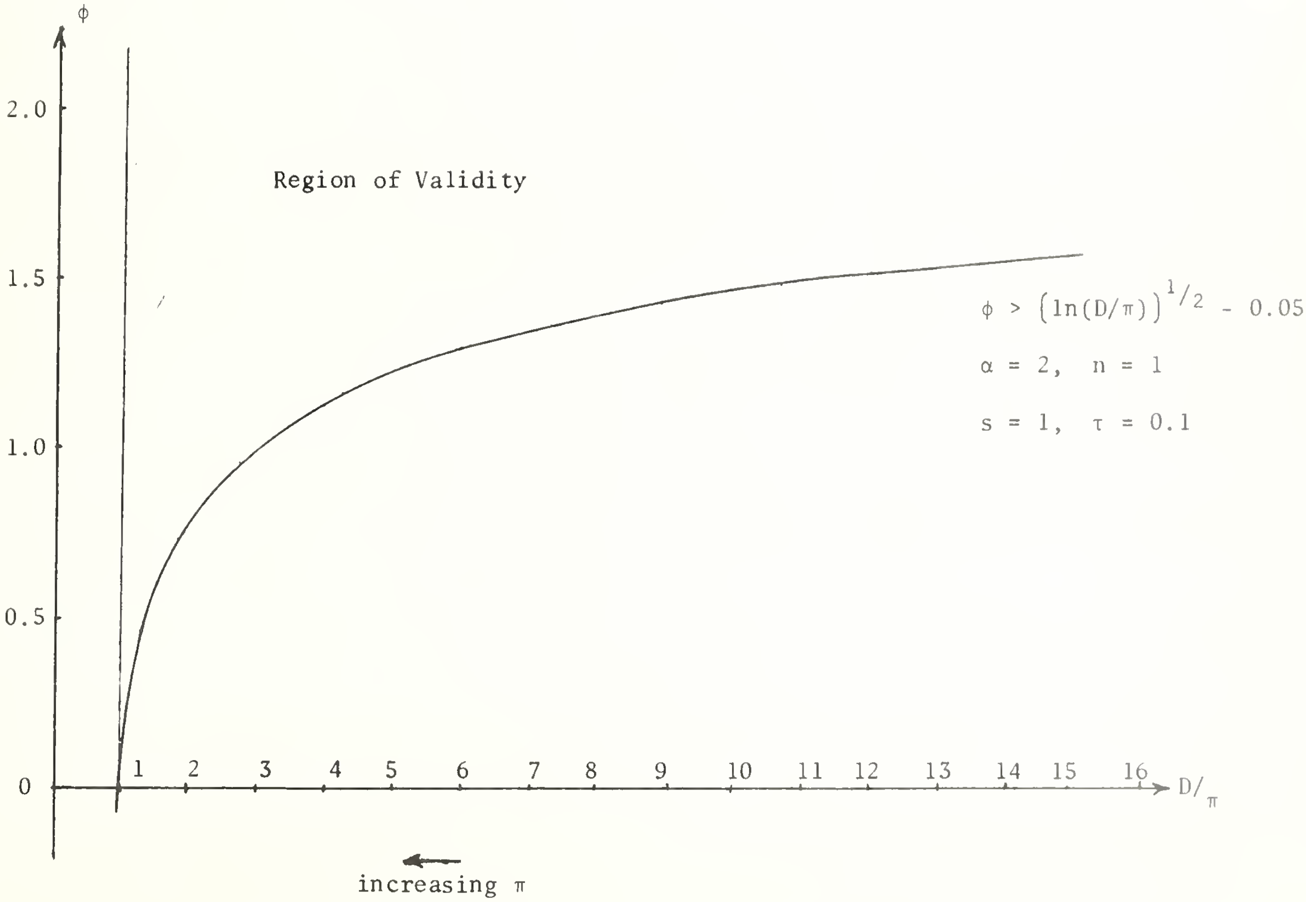


Figure 14. Valid Region for Trip Duration as a Function of Fare.

relative to the trip duration ϕ , valid as input to the algorithm.

From the one-way algorithm of Eq. (3-17) it is an easy transition to the two-way algorithm for scheduling a vehicle to accommodate passenger flow in both directions. Consider the following pair of iterative equations:

$$I(1,k) = \max \begin{cases} 0 + I(1,k+1) & \text{"hold"} , \\ b_1(k) + I(2,k+L_{12}+TR_2) & \text{"dispatch"} , \end{cases} \tag{3-28a}$$

$$I(2,k) = \max \begin{cases} 0 + I(2,k+1) & \text{"hold"} , \\ b_2(k) + I(1,k+L_{21}+TR_1) & \text{"dispatch"} . \end{cases} \tag{3-28b}$$

The terms appearing in these equations have the following definitions:

L_{12} = number of discrete time units that the vehicle requires to make the trip from node 1 to node 2.

TR_1 = turnaround time at node 1, measured in discrete time units.

$b_1(k)$ = immediate benefit achieved by having the vehicle depart node 1 at time k.

The immediate benefit is given by,

$$b_1(k) = -a_1 OC + \sum_{i \in A} (a_2 \pi + a_3 WTP^i) , \tag{3-29}$$

where

OC = operating cost of the trip from node 1 to node 2,

π = fare charged for the trip from node 1 to node 2,

A = set of passengers who would be aboard the vehicle if it departed at time k. A contains no more than C members, where C is the vehicle capacity.

WTP^i = passenger i's willingness to pay for the trip leaving at k , and arriving at $k+L_{12}$.

$I(1,k)$ = maximum total benefit it is possible to achieve, from time k through the end of the scheduling period (time K), if the vehicle were at node 1 at time k .

Parallel definitions apply to L_{21}, TR_2, $b_2(k)$, and $I(2,k)$.

The interpretation of the iterative Eqs. (3-28) is parallel to that of Eq. (3-17). Consider (3-28a), for example:

There are two possible decisions that can be made if the vehicle is at node 1 at time k. The first possibility is to have the vehicle remain at node 1 until at least the next time interval, $k+1$. The second possibility is to dispatch the vehicle at time k. The choice that maximizes $I(1,k)$ is to be chosen. If "hold" is chosen, there is no immediate gain, and $I(1,k)$ is just equal to $I(1,k+1)$. However, if "dispatch" is picked, then $I(1,k)$ is the sum of an immediate gain $b_1(k)$, plus $I(2,k+L_{12}+TR_2)$, the maximum possible benefit achievable after the vehicle reaches node 2 and becomes ready to depart again.

Eqs. (3-28) may be used in conjunction with the following set of boundary conditions:

$$I(1,k) = 0 \qquad \text{if} \quad K \leq k \leq K+L_{12}+TR_2 \;, \tag{3-30a}$$

$$I(2,k) = 0 \qquad \text{if} \quad K \leq k \leq K+L_{21}+TR_1 \;. \tag{3-30b}$$

These boundary conditions state merely that there is no further benefit to be achieved, if the vehicle returns to one of the stations after the end of the scheduling time period.

Eqs. (3-28) are iterated together, in dynamic programming fashion as explained earlier in connection with Eq. (3-17). A flow diagram illustrating the computational process is given in Fig. 15. Finally, it should be noted that the priority boarding rule and the eligibility inequality (3-26) hold for the two-way system as well.

The general iterative method to be developed here will revolve about the use of the single vehicle algorithm to schedule each vehicle in the

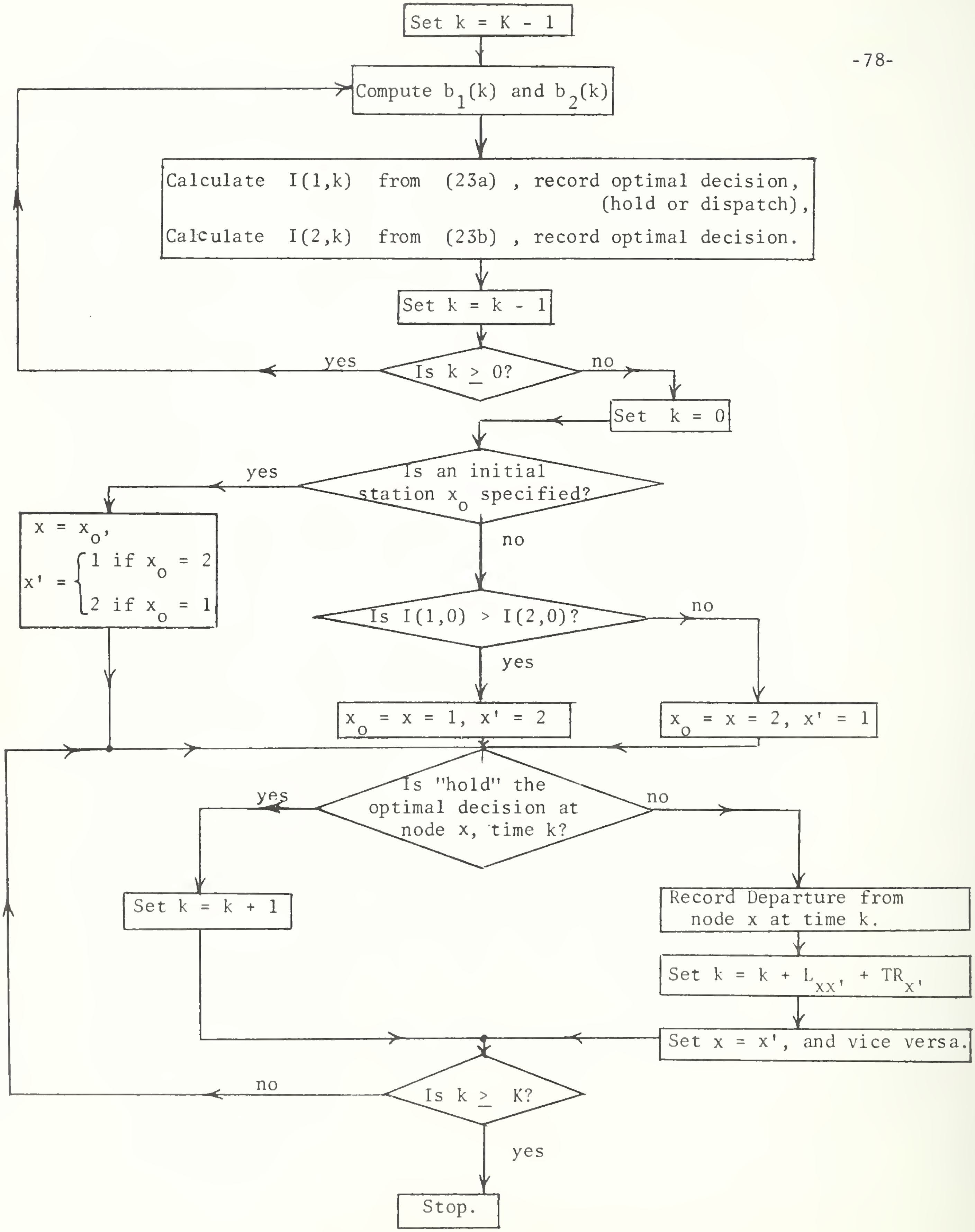


Figure 15. The Two-way Single Vehicle Algorithm.

system. Recall, however, that the latter algorithm requires a subset of travelers as input. Hence, a method is required to allocate travelers among vehicles. In this chapter, attention is restricted to the single link shuttle system illustrated in Fig. 11. In this context, consider the following assignment problem.

3.4 The Matching Problem

In this section, a description of the matching problem, in the context of passenger allocation, is given. An efficient algorithm for obtaining the optimal solution exists and is described in Ref. [61].*

Suppose each vehicle is given an initial timetable from which to operate. Then each vehicle makes a certain set of trips from each of the two stations in the network. Consider the total set of trips made from one of the stations, say station 1, by all vehicles in the system. Let each of these trips be represented by a node n. Tag each node n with a number d_n, such that

$$d_n = C_j, \quad \text{the capacity of the vehicle } j \text{ that makes trip } n.$$

The set of "trip nodes" will be called node set S1. The number d_n is called the "degree constraint" of node n.

Let each traveler (from station 1) be represented by a node m in node set S2. Tag each node in set S2 with $d_m = 1$. Now, consider the node pair nm, where $n \in S1$ and $m \in S2$. Let

$$e_{nm} = \text{the "edge" connecting node } n \text{ with node } m ,$$

---

* The algorithm is based on a "Hungarian" method. Credit belongs to Mr. Jack Edmonds, Applied Math. Division, National Bureau of Standards, for direction in the synthesis of this routine.

and

c_{nm} = the "edge weight" of edge e_{nm} .

Then c_{nm} will be given by

c_{nm} = value that passenger m holds for trip n .

Thus,

$$c_{nm} = \begin{cases} a_2\pi_n + a_3 WTP_n^m & \text{if } WTP_n^m \geq \pi_n , \\ 0 & \text{if } WTP_n^m < \pi_n , \end{cases} \tag{3-31}$$

where π_n is the fare corresponding to trip n, and WTP_n^m is the willingness to pay of passenger m for trip n. If $c_{nm} = 0$, edge e_{nm} will be ignored, i.e., assumed not to exist.

The representation of trips and travelers developed above, transforms the passenger allocation problem into the general matching problem in a "bipartite graph." Fig. 16 illustrates this graph-theoretic representation. A formal statement of the problem requires the following definitions:

> A graph is <u>bipartite</u> if its nodes partition into two sets such that no edge of the graph joins two members of the same set.
>
> A <u>matching</u> in a graph is a subset M of its edges such that no more than d_n edges meet any node n.

The problem statement is given as follows:

> Given a bipartite graph with non-negative edge weights c_e on each edge e, find a matching M such that the sum $\left(\sum_{e \in M} c_e\right)$ is maximized.

Thus, given any feasible timetable for the vehicles, the matching algorithm will find the <u>best</u> assignment of travelers to trips. It is now opportune to combine this routine with the vehicle scheduling algorithm,

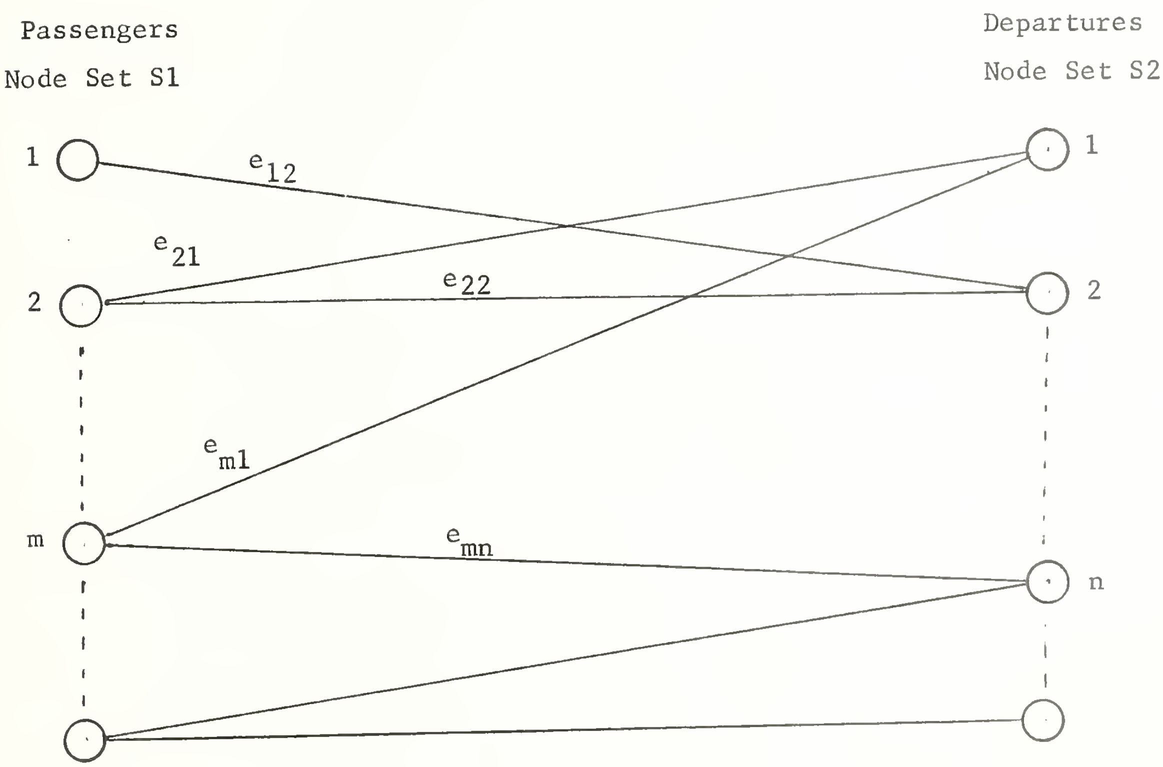


Figure 16. The Matching Problem.

to synthesize the overall computational procedure for scheduling a fleet of vehicles.

3.5 The First Computation Method

Consider the procedure illustrated in the flowchart of Fig. 17, for computing timetables and passenger assignments for the two-way, multi-vehicle shuttle system. Given a passenger to vehicle allocation, this iterative scheme, which will be called by its FORTRAN name MODEL1, computes vehicle timetables by successive use of the single vehicle scheduling algorithm. Alternately, given a set of vehicle timetables, MODEL1 computes the assignment of travelers to trips, using the optimum assignment algorithm. Computations are started by specifying an initial allocation of passengers to vehicles, <u>or</u> an initial fleet timetable. MODEL1 alternates between the passenger assignment computation and the successive vehicle schedule computations, until one timetable is obtained twice in a row. The resulting timetable and corresponding assignment solution constitute the final output.

Important to the discussion of MODEL1 is the observation that there are several aspects of this methodology that require arbitrary specification. First, consider the process of initialization with which computations are begun. Obviously, there is a myriad of possible initial solutions. Unfortunately, it will be found that the initial solution radically affects the output results of MODEL1 computations. This consequence will be discussed later.

Second, careful consideration of the flowchart of Fig. 17 reveals that the assignment algorithm does not completely resolve the passenger allocation problem. In particular, there may be some travelers who remain

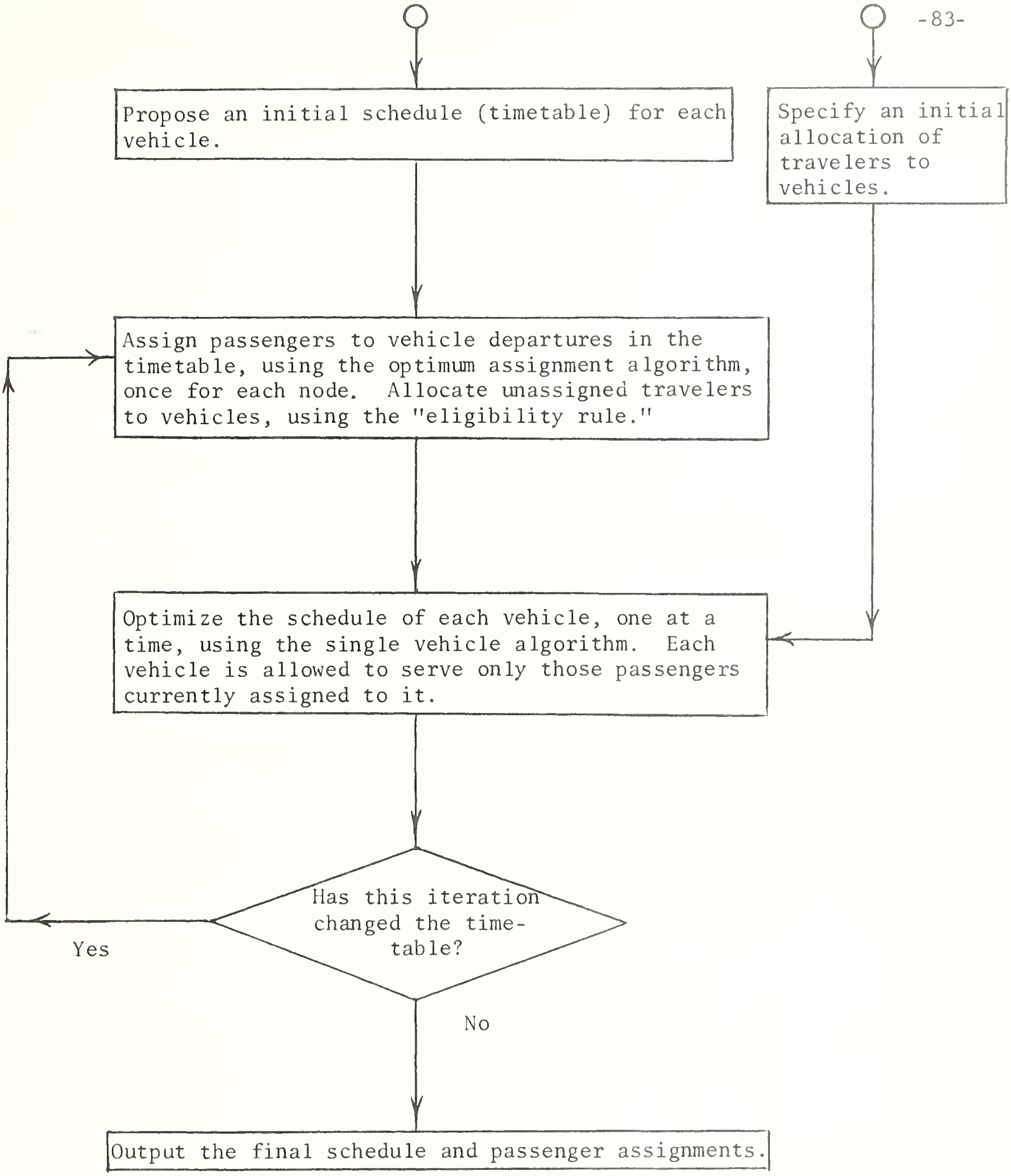


Figure 17. The First Computation Method (MODEL1).

unassigned, either because there are no trips in the current iteration that meet their preferences, or because all acceptable trips are filled to capacity. The question then arises as to the allotment of "temporary" assignments to these travelers, so that their preferences may influence the next round of vehicle schedule computations. This issue may be resolved by specifying an arbitrary "eligibility rule." The following rule was utilized here:

> Allocate each unassigned traveler to the vehicle that makes the trip in the current schedule that he values most.

This rule facilitates the revision of passenger assignments in successive iterations, as schedules shift to the advantage of some travelers relative to others. On the other hand, the rule has the disadvantage that if a traveler cannot be assigned to his favorite trip (because of capacity constraints) he remains ineligible for his second choice, which may have empty seats. An alternate rule was tried, by which an unassigned passenger is made eligible for the vehicle making his favorite "unfilled" departure. This rule proves to be inferior in the sense that it fails to dislodge passengers assigned to vehicles on previous iterations, in favor of new travelers of higher current priority. Of course, many other eligibility rules might be tried. Instead of pursuing such a course, analysis of the weaknesses of the MODEL1 process, including the need for the foregoing eligibility rule, will help lead to the development of a superior methodology later in this chapter.

Before the discussion of MODEL1 computational results, consider the question of convergence. It will be shown that the MODEL1 iterations must converge to a unique solution, for each initialization of the computation. In practice, only a few (three or four) iterations are required for convergence. The proof that MODEL1 converges, emerges from the following five statements:

1. For a finite system of travelers and vehicles, there is a finite maximum value for the stated objective function.

2. MODEL1 produces a non-negative improvement in the value of the objective function at every step of the process.

3. For every passenger to vehicle allocation, MODEL1 computes a unique timetable.

4. For every vehicle fleet schedule, MODEL1 yields a unique set of passenger assignments.

5. For a finite system of travelers and vehicles, there is only a finite number of possible passenger to vehicle allocations.

Statement (1) follows from the fact that each vehicle has a finite capacity and speed. Therefore, within the scheduling period (0,T), only a finite number of passengers may be transported by each vehicle. Given that there are a finite number of vehicles, and that the benefit resulting from the transport of any traveler is finite, then the total system benefit from any schedule, including the "best" one, must be finite. Alternately, it may be argued that a finite traveler population can absorb only a finite benefit.

Statement (2) follows from the properties of the scheduling and assignment algorithms. In particular, given any set of vehicle schedules based on a current passenger assignment, one always attains an assignment at least as good as the current one, by applying the assignment algorithm. Analogously, given any passenger assignment based on a current vehicle fleet timetable, one always obtains a timetable at least as good using the single vehicle scheduling algorithm. These latter two statements follow from the fact that the assignment and scheduling algorithms yield optimal solutions.

Although the assignment and scheduling solutions are optimal, they are not necessarily unique. That is, more than one passenger assignment solution may yield the optimum objective function value. Similarly, more than one timetable may yield the optimal value for a given vehicle. However, the algorithms are programmed such that for any given input, the same optimal solution results each time each algorithm is run with any specific input. Thus, statements (3) and (4) apply. Statement (5) gives an obvious property of a finite system of travelers and vehicles.

Now, statements (1), (3), and (4) imply that MODEL1 cannot improve the solution indefinitely, since at most MODEL1 may test the timetables resulting from every possible traveler-to-vehicle allocation, until the best is reached. Furthermore, statement (2) precludes a decrease in the objective function value on any iteration. Thus, the only remaining roadblock to convergence is the possibility of oscillations at a given level of the objective function. However, statements (3) and (4) eliminate this contingency, i.e., a given timetable must necessarily repeat if no improvement occurs. Thus, convergence is assured.

Finally, it may be observed that once the program has converged, i.e., once a timetable has repeated itself, then it is no longer possible for the MODEL1 procedure to improve the solution through additional iterations. This follows from (3) and (4), which imply that the final timetable would repeat indefinitely if computations were not terminated. Thus, the rule for determining convergence (i.e., timetable repetition) is correct.

MODEL1 was translated into a FORTRAN IV computer program,* to secure a better understanding of this iteration procedure. A few illustrative runs will be presented below.

3.6 Examples for MODEL1

Figs. 18 and 19 display the inputs used in most of the examples of this section. In Fig. 18, the following notations may not be self-explanatory:

$$\text{SLOPE} \longleftrightarrow s\ ,$$

$$\text{DOLLAR} \longleftrightarrow D\ ,$$

$$\text{KFIN} \longleftrightarrow K\ .$$

All travelers are assumed to have the same preference parameters (D,α,n,s) in these examples. The "INITIAL STATION" column specifies the network node (No. 1 or No. 2) from which the corresponding vehicle is required to start at time $t = 0$. The entry 0 indicates that no initial station requirement is imposed. "SPEED" is given in miles per hour. Fig. 19 lists the (t^*,w) values for each potential traveler in the system.

A. The objective function in the first example is "net willingness to pay minus costs," as shown in Fig. 18. The complete set of iterations is displayed in Figs. 20a through 20c. Fig. 20a shows the initial

---

* The computer code utilizes a simpler version of the assignment algorithm than the one described earlier. In particular, the routine programmed here requires all degree constraints, $d_n = 1$. Hence, travelers are matched to vehicle-seat-departures rather than vehicle departures. The reason for this is painfully simple. The more general algorithm was not known to the author at the time MODEL1 was programmed. Furthermore, the exploratory nature of this study did not justify subsequent reprogramming of the routine. However, use of the more general algorithm would certainly improve the efficiency of MODEL1.

THE INPUT PARAMETERS

```
T =  10.000                          100.000 MILES FROM NODE 1 TO NODE 2
                                     100.000 MILES FROM NODE 2 TO NODE 1
TURNARROUND TIME AT NODE 1 =   0.050 HOURS          TURNAROUND TIME AT NODE 2 =   0.050 HOURS
SLOPE =   4.000      DOLLAR =   20.00      ALPHA =   0.500     N =   2.000
A1 =  1.      A2 = -1.      A3 =  1.

KFIN =    120                  DELTA T =  0.083
```

VEHICLE NO.	SPEED	CAPACITY	INITIAL STATION	OPER. COST/MI.	FARE/ONE-WAY
1	75.000	2	0	0.02	5.00
2	75.000	2	0	0.02	5.00
3	50.000	4	0	0.04	5.00
4	50.000	4	0	0.04	5.00

Figure 18. Input Parameters.

LIST OF TRAVELERS

NODE 1

PASSENGER NO.	TSTAR	W
1	0.500	0.0
2	1.000	0.50
3	1.500	1.00
4	2.500	0.0
5	3.000	0.50
6	3.500	1.00
7	4.500	0.0
8	5.000	0.50
9	5.500	1.00
10	6.500	0.0
11	7.000	0.50
12	7.500	1.00
13	8.500	0.0
14	9.000	0.50

NODE 2

PASSENGER NO.	TSTAR	W
1	1.000	0.0
2	2.900	0.0
3	3.000	0.0
4	3.100	0.0
5	3.200	0.0
6	5.000	1.00
7	7.000	1.00
8	7.100	1.00
9	7.200	1.00
10	7.300	1.00
11	9.000	0.50
12	9.100	0.50
13	9.100	0.50
14	9.200	0.50

Figure 19. Traveler Time Preferences.

```
 SCHEDULE FOR VEHICLE   1
DEPARTS NODE    AT TIME     PASSENGERS ABOARD

          1         0.50       2
          2         2.50       0
          1         4.50       8
          2         6.50       0

BENEFIT ACHIEVED BY VEHICLE   1 =      0.0


 SCHEDULE FOR VEHICLE   2
DEPARTS NODE    AT TIME     PASSENGERS ABOARD

          2         1.00       0
          1         3.00       7
          2         5.00       6
          1         7.00      13

BENEFIT ACHIEVED BY VEHICLE   2 =      0.0


 SCHEDULE FOR VEHICLE   3
DEPARTS NODE    AT TIME     PASSENGERS ABOARD

          1         1.50       3
          2         5.00       0
          1         9.00       0

BENEFIT ACHIEVED BY VEHICLE   3 =      0.0


 SCHEDULE FOR VEHICLE   4
DEPARTS NODE    AT TIME     PASSENGERS ABOARD

          2         3.00       0
          1         6.50       0
          2         9.00       0

BENEFIT ACHIEVED BY VEHICLE   4 =      0.0

BENEFIT ACHIEVED IN SCHEDULING PROCESS =      0.0
BENEFIT ACHIEVED IN ASSIGNMENT PROCESS =     36.082

PROGRAM HAS COMPLETED   0 ITERATIONS

THE SCHEDULE HAS NOT YET CONVERGED
```

Figure 20. Illustrative MODEL1 Computation
(a) Initial solution.

```
 SCHEDULE FOR VEHICLE   1
DEPARTS NODE    AT TIME     PASSENGERS ABOARD

          1        0.33       2
          2        1.83       4   5
          1        5.33       9  10
          2        7.25       9  10

BENEFIT ACHIEVED BY VEHICLE   1 =     43.530

 SCHEDULE FOR VEHICLE   2
DEPARTS NODE    AT TIME     PASSENGERS ABOARD

          2        1.58       2   3
          1        3.33       6   7
          2        5.00       6
          1        7.33      12  13

BENEFIT ACHIEVED BY VEHICLE   2 =     41.005

 VEHICLE    3 HAS NO SCHEDULED DEPARTURES


 SCHEDULE FOR VEHICLE   4
DEPARTS NODE    AT TIME     PASSENGERS ABOARD

          2        8.08      11  12  13  14

BENEFIT ACHIEVED BY VEHICLE   4 =      4.823

BENEFIT ACHIEVED IN SCHEDULING PROCESS =     89.358
BENEFIT ACHIEVED IN ASSIGNMENT PROCESS =    109.358

PROGRAM HAS COMPLETED   1 ITERATIONS

THE SCHEDULE HAS NOT YET CONVERGED
```

Figure 20. (b) First Iteration.

```
 SCHEDULE FOR VEHICLE   1
DEPARTS NODE    AT TIME     PASSENGERS ABOARD

         1        0.33        2
         2        1.83        4   5
         1        5.33        9  10
         2        7.25        9  10

BENEFIT ACHIEVED BY VEHICLE   1 =     43.530


 SCHEDULE FOR VEHICLE   2
DEPARTS NODE    AT TIME     PASSENGERS ABOARD

         2        1.58        2   3
         1        3.33        6   7
         2        5.00        6
         1        7.33       12  13

BENEFIT ACHIEVED BY VEHICLE   2 =     41.005

 VEHICLE    3 HAS NO SCHEDULED DEPARTURES


 SCHEDULE FOR VEHICLE   4
DEPARTS NODE    AT TIME     PASSENGERS ABOARD

         2        8.08       11  12  13  14

BENEFIT ACHIEVED BY VEHICLE   4 =      4.823

BENEFIT ACHIEVED IN SCHEDULING PROCESS =     89.358
BENEFIT ACHIEVED IN ASSIGNMENT PROCESS =    109.358

PROGRAM HAS COMPLETED   2 ITERATIONS

THE SCHEDULE HAS CONVERGED
```

Figure 20. (c) Second Iteration.

timetable, with passengers optimally assigned to it. The temporary traveler eligibilities are not shown here. Recall from Fig. 19, that passenger 3 from node 1, for example, is distinct from passenger 3 from node 2.

Fig. 20 displays the following measures of aggregate benefit:

BENEFIT ACHIEVED IN SCHEDULING PROCESS = value of the objective function on the current iteration.

This value is the sum of vehicle benefits shown under each vehicle timetable. Since these benefits are found using the vehicle scheduling algorithm, zero value is specified for the initial (non-computed) timetable iteration of Fig. 20a.

BENEFIT ACHIEVED IN ASSIGNMENT PROCESS = value of benefit accruing to passengers assigned to trips in the current timetable.

The difference between the aggregate scheduling and assignment benefits (after the initial iteration) is the operating cost of the timetable.

Observe that the initial timetable is a rather poor one. Only six of the twenty-eight travelers find acceptable trips. Hence, most travelers are "plugged into" the next iteration via the temporary eligibility procedure. The behavior of vehicle 3 is interesting here. Traveler 3 from node 1 is assigned to the vehicle's initial timetable. However, that traveler is not accommodated in the next iteration. In fact, too few travelers are allocated to vehicle 3 to allow it to run profitably. Hence, the vehicle becomes inactive.

Convergence is achieved quickly. The solution value is $89.358, with $109.358 in traveler benefits, and $20. in operating costs.

B. The following runs are designed to demonstrate the effect of using different initial solutions to solve a given scheduling problem with MODEL1. In these computations, the objective is maximization of consumer surplus ($a_1 = 0$, $a_2 = -1$, $a_3 = 1$). All other inputs are assumed the same as in Figs. 18 and 19.

Figs. 21a and b display four initial schedules and four initial assignments used to generate solutions to the problem. The corresponding results are illustrated in Figs. 22a and b. A few things are apparent from these results. First, although there is visible similarity among the various solutions, it is evident that the initialization radically affects the final results. This is an important point that will receive more attention later.

Second, the figures show that initialization via passenger allocation appears to be more effective than initialization by timetable specification. This results from the fact that unless the initial timetable is chosen rather skillfully, many travelers may find the entire timetable unacceptable. Hence, the heuristic eligibility rule must be relied upon to allocate travelers to vehicles, a task that it may do poorly. On the other hand, guessing at a reasonable initial passenger allocation is relatively easy. For example, assignment A1 is a fairly uniform distribution of travelers among vehicles. A2 is an attempt to group travelers of similar preferences, such that each vehicle is likely to be able to accommodate most of the passengers assigned to it. Assignments A3 and A4 are randomly chosen. That is, each traveler is assigned to a vehicle by the roll of dice. (Each vehicle is equally likely to be chosen.) Even the latter assignments lead to reasonably good solutions.

Schedule	Initial Value	Vehicle No.	Departure Times (- indicate departure from node 2)			
S1	36.1	1	0.50	-2.50	4.50	-6.50
		2	-1.00	3.00	-5.00	7.00
		3	1.50	-5.00	9.00	
		4	-3.00	6.50	-9.00	
S2	28.6	1	2.00	-5.00		
		2	-7.00	9.00		
		3	7.00	-9.50		
		4	-2.00	4.50		
S3	9.7	1	1.00	-3.50	8.00	
		2	1.00	-3.50	8.00	
		3	1.00	-3.50	8.00	
		4	1.00	-3.50	8.00	
S4	97.7	1	-0.01	2.33	-3.92	5.33
			-7.00	8.42		
		2	-1.75	3.33	-5.00	7.33
		3	1.50			
		4	-0.01			

Figure 21a. Initial Schedules.

Assignment	VFS*	Vehicle No.	Assigned Passengers From node 1	From node 2
A1	75.9	1	1,5,9,13	4,8,12
		2	2,6,10,14	3,7,11
		3	3,7,11	2,6,10,14
		4	4,8,12	1,5,9,13
A2	117.2	1	1,2,9,10,14	2,3,9,10
		2	3,4,11,13	6,13,14
		3	8,12	1,4,5,7,8
		4	5,6,7	11,12
A3	61.1	1	1,4,6,7,13	none
		2	9,11,12	8,13,14
		3	3,5,8,10	1,2,4,6,10,11
		4	2,14	3,4,7,9,12
A4	90.8	1	7,9,12	2,9,10,11
		2	1,11,14	1,3,4,5,6,7,13,14
		3	2,4,5,6,10	12
		4	3,8,13	8

* VFS = Value of the first generated schedule.

Figure 21b. Initial Assignments

Schedule	Final Value	Vehicle No.	Departure Times			
S1	116.4	1	0.33	-1.83	5.33	-7.25
		2	-1.58	3.33	-5.00	7.33
		3	2.00	-5.92	8.00	
		4	-3.92	6.00	-8.08	
S2	67.6	1	-0.00	1.42	-5.00	
		2	-7.00	8.42		
		3	-3.92	6.00	-8.08	
		4	-1.08	5.50		
S3	73.2	1	0.33	-1.83	5.33	-7.00
			8.42			
		2	no departures			
		3	-1.92	4.00	-8.08	
		4	no departures			
S4	98.1	1	-0.00	2.33	-3.92	5.33
			-7.00	8.42		
		2	-1.83	3.33	-5.00	7.33
		3	1.50			
		4	no departures			

Figure 22a. Solution Schedules for Initial Schedules.

Assignment	Final Value	Vehicle No.	Departure Times				
A1	108.1	1	-0.00	2.33	-4.08	5.50	-7.25
		2	0.25	-1.67	5.17	-7.00	8.42
		3	1.50	-3.92	6.00	-8.17	
		4	-1.17	4.00	-8.08		
A2	119.8	1	0.25	-1.67	5.33	-7.17	8.58
		2	1.33	-5.00	7.17	-8.58	
		3	-1.17	4.00	-7.08		
		4	-1.42	3.50	-8.08		
A3	91.3	1	-1.83	3.33	-5.92	7.33	
		2	-5.00	6.42	-8.50		
		3	-0.92	4.50	-8.08		
		4	-7.17				
A4	108.2	1	-1.58	3.00	-4.42	5.83	-7.25
		2	-1.83	3.50	-5.00	6.42	-8.50
		3	0.00	-2.42	4.50	-8.08	
		4	2.00	-5.92	8.00		

Figure 22b. Solution Schedules for Initial Assignments.

Finally, it is not clear what "type" of solutions the answers of Fig. 22 represent. Recall from the earlier discussion in this chapter, that iterative procedures may lead to various kinds of suboptimal solutions. The following discussion, which includes the presentation of computer runs designed to "perturb" the solutions obtained above, pursues the question of solution "quality." That discussion, in addition to the earlier observations on MODEL1 solutions, will lead to the development of an improved computational method, MODEL2.

It is clear that MODEL1 has the capability to render non-negative improvement to any input timetable. In this sense, it may be considered a useful computational tool. However, the wide divergence of solutions resulting from alternate initializations casts serious doubt on the ability of MODEL1 to reach an optimal solution. In addition, one rarely obtains identical solutions from different initializations. Hence, even if MODEL1 could be relied upon to generate an optimum, it is difficult to predict how many different initial solutions would be required to do so.

In order to determine whether or not a given solution is "locally optimal" it is usually necessary to test values of the objective function over points in its immediate neighborhood, within the "feasible region" of solutions. In general, it is possible to test for "global" optimality only by finding and comparing all local optima. Points in the neighborhood of a solution may be tested by effecting small perturbations in various directions. In the present context, trip departure times may be varied by plus or minus one grid point, $\Delta t = T/K$.

However, it is not feasible to try all possible such perturbed solutions--there are too many of them. Specifically, three changes to each departure time $(+\Delta t, -\Delta t, 0)$ are possible. If there are d departures per vehicle, and V vehicles, then there is a maximum of 3^{dV} possible perturbations. Actually, fewer possibilities exist since some of these perturbations violate system constraints. However, with $d \doteq 10$ and $V = 4$, as in the current examples, there is a maximum of 3^{40} possibilities. Even if many of these are infeasible, the number of possibilities is still enormous. Hence, a comprehensive perturbation test is not reasonable.

Nevertheless, it is instructive to consider a few perturbed solution runs. The final timetables that resulted from initializations S1, A2, S3, and A3, were perturbed by delaying all departures of even numbered vehicles by Δt, and advancing all odd vehicle trips by Δt. Here, $\Delta t = T/K = 0.083$. These perturbed timetables were input to MODEL1. The purpose is to observe whether the answers would converge back to the original solutions, or move to new solutions. If the original solution is locally optimal then the computation may be expected to converge back to it. Otherwise, if the perturbation happens to be in a direction of potential improvement, an improved answer may be expected. The results are summarized in Fig. 23.

More detailed inspection of runs P2 and P4 reveals that alteration of traveler eligibilities is the important mechanism by which perturbations lead to new answers. That is, small changes in solutions are sometimes sufficient to alter traveler assignments and secondary eligibilities such that improved solutions ultimately result. Although it is clear by now that MODEL1 tends to terminate suboptimally, it is advantageous to pursue the perturbation analysis one step further.

Four runs were performed to determine whether solutions could be further improved by adjusting the eligibilities of travelers left unaccommodated in the final schedules. Adjustments were made by matching (by casual observation) the preferences of unassigned travelers to existing or feasible (supplementary) unfilled departures in the context of the final schedule. Again, the final solutions resulting from initializations S1, A2, S2, and A3 were used. The final passenger assignments, perturbed as described above, were input for the new "adjustment" runs. The results are summarized in Fig. 24.

Manipulation of traveler eligibilities can significantly improve the solutions. Figs. 23 and 24 show that the small perturbations of the timetable or passenger allocation solutions can nudge the computations out of a suboptimal "rut." A graphical consideration of the process will help clarify how MODEL1 may terminate suboptimally.

Consider two "spaces," $\mathcal{A}$ and $\mathcal{S}$:

$\mathcal{A}$ ≡ space of all possible assignments of travelers to vehicles,

$\mathcal{S}$ ≡ space of all possible vehicle timetables.

If spaces $\mathcal{A}$ and $\mathcal{S}$ are represented along separate axes, as in Fig. 25, then MODEL1 is reminiscent of the "one at a time," successive approximation method.* That is, MODEL1 proceeds by fixing the passenger assignment solution and optimizing the timetable, then fixing the timetable and optimizing the assignments. In Fig. 25, these iterations are represented by right-angle movements in ($\mathcal{A}$, $\mathcal{S}$) space. The danger of this procedure is apparent. For example, solution (A_1, S_1) in the figure

* For convenience, Fig. 25 displays assignment space as a continuum, rather than a more appropriate discrete representation. The comments above apply to either case.

Run	Original Solution and Value		New Value	Identical Solution?
P1	S1	116.4	116.4	Yes.
P2	A2	119.8	122.8	No, slightly modified.
P3	S2	67.6	67.6	Yes.
P4	A3	91.3	99.1	No, one trip added to otherwise identican schedule.

Figure 23. Perturbation Runs.

Run	Original Solution and Value		New Value	Changes
AJ1	S1	116.4	118.6	Modification of one vehicle schedule.
AJ2	A2	119.8	119.8	No change.
AJ3	S2	67.6	80.0	Significant modification of two vehicle schedules.
AJ4	A3	91.3	114.4	Trips added to three vehicle schedules.

Figure 24. Adjustment Runs.

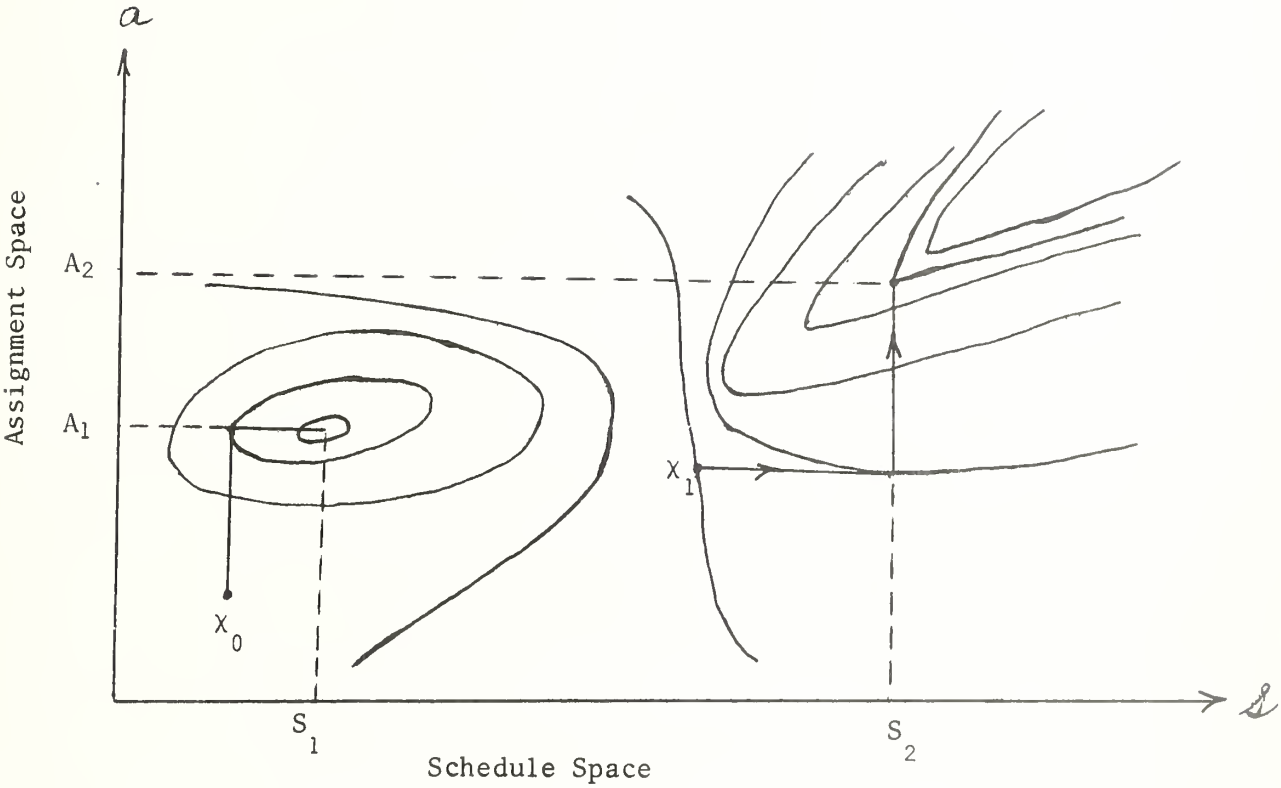


Figure 25. MODEL1 Iterations in Schedule-Assignment Space.

is (locally) optimal; however, a "ridge" solution such as (A_2,S_2) is also a possible output. The examples of Figs. 23 and 24 indicate that ridge-like solutions are occurring. This may be concluded from the fact that small perturbations in the timetables or assignments are found to lead to improvements in each original solution.

Viewing the situation from another perspective, it may be observed that the MODEL1 method does not adequately reflect the interdependencies among interrelated parts of the problem, e.g., the separate vehicle timetables and passenger-to-vehicle allocations. Consider, for instance, the successive vehicle schedule optimizations.

Recall that throughout the successive timetable computations, the traveler to vehicle allocation is fixed. Thus, each vehicle schedule is optimized only with respect to its own traveler subpopulation. No consideration is paid to the potential increase in benefit that a vehicle might provide to travelers outside its domain. It is true that subsequent to the entire vehicle fleet scheduling operation, the passenger assignments are recomputed. This does not appear to be adequate, however. The perturbation examples have shown that manipulation of passenger eligibilities can significantly improve the solutions. Hence, to enhance the iterative method, a technique is required for allowing vehicle schedules to be computed on the basis of benefits to the <u>entire</u> traveler population. Accordingly, optimization of a vehicle timetable must be allowed to consider changes in the passenger assignments. Returning to the viewpoint of Fig. 25, the revised procedure must relax the constraint in MODEL1 and allow movements in more than just the two orthogonal directions in $(\mathcal{a},\mathcal{s})$ space.

3.7 Revised Computation Method

The new computational method, called MODEL2 and illustrated in Fig. 26, operates as follows:

Each vehicle schedule is optimized with respect to the entire traveler population, under the following stipulations: (a) All passengers either currently assigned to vehicle j or currently unassigned are considered fully eligible for vehicle j, during the process of optimizing vehicle j's timetable. (b) A passenger currently assigned to a trip made by a vehicle other than j, is considered eligible for j, with the reservation that the benefit that vehicle j can be said potentially to provide such a passenger, is equal to the net increment that the traveler would achieve by switching to vehicle j from his previous assignment.

Hence, execution of the single vehicle algorithm proceeds essentially as before, with the exception that at each time k, the decision to depart is based on the potential increment in benefit that would accrue to the entire traveler population. That is, the quantity $b(k)$ includes the sum of benefits of passengers previously assigned to j or previously unassigned, plus the increment in benefits accruing to those passengers who would switch to vehicle j from a different vehicle, if vehicle j were to depart at time k.

To be more precise, the immediate benefit $b_n^h(k)$, for trip h leaving node n at time k, is given by

$$b_n^h(k) = -a_1 OC_h + \sum_{i \in H} [a_2(\pi_h - \pi_{h'}) + a_3(WTP_h^i - WTP_{h'}^i)] , \qquad (3\text{-}32)$$

where

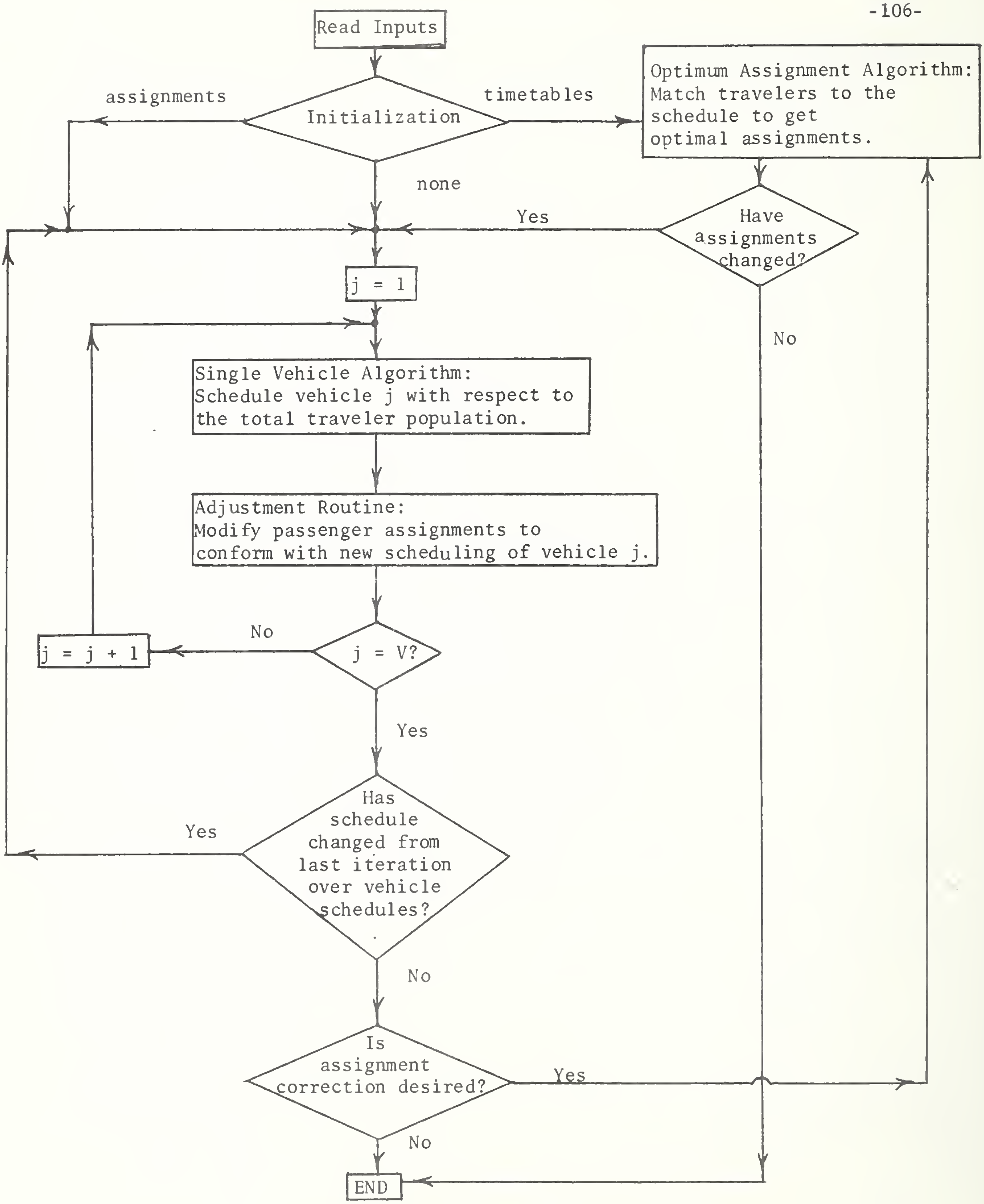


Figure 26. The Revised Computation Method (MODEL2)

OC_h = operating cost for current trip h,

π_h = fare for the current trip h,

$h' = h'(i)$ = trip to which traveler i has been previously assigned,

$\pi_{h'}$ = fare for trip h',

WTP_h^i = willingness to pay of traveler i for trip h,

$WTP_{h'}^i$ = willingness to pay of traveler i for trip h',

H = set of all travelers for whom $WTP_h^i \geq \pi_h$, and

$$a_2\pi_h + a_3 WTP_h^i > a_2\pi_{h'} + a_3 WTP_{h'}^i \ .$$

Let

h' = 0 if traveler i is previously assigned to the vehicle making trip h, or is previously unassigned,

and let

$WTP_{h'}^i = \pi_{h'} = 0$ for $h' = 0$.

Recall that if trip h has limited capacity then a priority boarding rule must be implemented. A slight modification of the MODEL1 rule is required here. Rather than board in decreasing order of values for trip h, travelers are boarded in decreasing order of <u>incremental</u> values for trip h, over the values of previously assigned trips h'. This rule assures that the scheduling of the vehicle making trip h, results in a net gain to the <u>total</u> system benefit.

Subsequent to the optimization of <u>each</u> vehicle schedule, it is necessary to modify the passenger assignments to conform with the revised timetable. One time-consuming way to do this, of course, would be to apply the optimum assignment algorithm after each vehicle schedule computation. A less expensive approach is merely to change the assignments of those travelers affected by the rescheduling of the current vehicle. The latter approach is implemented by an adjustment routine that works as follows.

Suppose that vehicle j's timetable has been optimized, and that the solution is based partially on the incremental benefit that some traveler i would receive, were he to transfer to vehicle j from his previously assigned vehicle. Then before the next vehicle $(j+1)$ timetable is optimized, it is necessary to change traveler i's assignment classification to correspond to the revised timetable. The mechanics of the reclassification are straightforward. It is sufficient here to say that traveler incremental values for the trips in the new vehicle timetable are recomputed and reordered, to implement the optimal boarding rule. Once this is done, those travelers aboard trips different from their original ones, are reclassified accordingly.

Thus, MODEL2 is characterized by successive single vehicle schedulings and passenger assignment modifications. One possibility is that under such a procedure, use of the optimal assignment algorithm might be dispensed with entirely. Unfortunately, the assignment modification procedure described here will not necessarily result in an optimal assignment of passengers to the resulting final schedule. However, the new procedure will significantly reduce the need for using the optimal assignment routine. That is, use of it can be viewed as a "correction maneuver" to prevent the traveler assignments from straying too far from the optimal. Hence, the assignment algorithm can be used only occasionally during the iterations. As a practical matter, the solutions obtained without using the optimal assignment routine will prove to be quite acceptable, and almost as good as solutions obtained with the assignment algorithm option. In any case, one feature of the MODEL2 computer program is an optional call of the assignment algorithm.

A welcome result of developing MODEL2 is the elimination of the arbitrary temporary eligibility rule. That is, under MODEL2 it is no longer necessary arbitrarily to allocate unassigned travelers to vehicles; unassigned travelers are considered fully eligible for the vehicle whose timetable is currently being optimized. Another interesting feature is the choice of three modes of initialization. In addition to initial timetables and passenger assignments, it is feasible to specify no initialization at all. In that case, timetable computations begin with all travelers "uncommitted" and fully eligible for the first vehicle. Incidentally, if passenger assignment initialization is chosen, the first round of vehicle schedule computations are performed with travelers eligible only for their initially assigned vehicles in order that an initial timetable may be constructed from which computations may begin. In other words, an initial assignment specification is interpreted as implying an initial timetable, namely that timetable which implements the initial assignment optimally.

Termination of the MODEL2 process is also more flexible. Once the fleet timetable has converged, an option is available to apply the optimal assignment algorithm. If the option is not chosen, the computation is terminated; otherwise, a failure of the assignment algorithm to modify existing traveler assignments will terminate the process. If such modification occurs, then timetable optimizations begin again. The discussion below will demonstrate that MODEL2 computations must always converge.

Convergence for MODEL2 is assured on a basis similar to that for MODEL1. In particular, statements (1), (2), and (5) of p. 85 hold for MODEL2 as well. Statements (3) and (4) may be restated for MODEL2 as follows:

3. For every possible state of the traveler population, i.e., allocation of travelers to vehicles, the vehicle schedule algorithm is programmed to yield a unique timetable solution.

4. For every fleet timetable, the optimal assignment algorithm yields a unique traveler assignment solution.

As before, statements (1), (3), and (5) imply that MODEL2 cannot improve the solution indefinitely. Statement (2) precludes a decrease in the objective function at any step. Oscillations at a fixed objective level are precluded by (3) and (4) as follows:

First, note that the program will not transfer any traveler from one vehicle to another unless the net system gain is positive. Hence, changes in traveler assignments are impossible if the objective level remains fixed. Second, by (3) successive iterations over vehicle schedules, with passenger assignments fixed, lead to identical output timetables. Thus, convergence within the inner vehicle-timetable optimization loop is assured.

Now, suppose timetable convergence has occurred, and application of the assignment algorithm is desired. If the resulting assignment solution is identical to the current one, termination is automatic. If assignments are altered, timetable computation begins again. However, if the previous timetable has already achieved the terminal objective value, then no further increment is forthcoming, and no further assignment changes will occur. Therefore, the identical timetable will be regenerated, and by (4) final application of the assignment algorithm will produce no additional changes. Hence, convergence obtains.

3.8 Examples for MODEL2

C. To observe the operation of MODEL2, reconsider example A (Fig. 20) used for MODEL1. Figs. 27a through 27c display the successive iterations; the order in which vehicles are read into the program here, is the reverse of Fig. 18. The "optimal assignment" and "no initialization" options are utilized in this computation. As for earlier printouts, identically numbered travelers from different nodes are distinct.

The final solution value is 97.2. The total value of traveler benefits is 129.2. The difference between these numbers is the timetable operating cost.

In contrast to MODEL1, MODEL2 is sensitive to the ordering of the vehicles because the vehicles whose timetables are optimized first have "first shot" at serving unassigned travelers and potential transferees. The allocation of traveler subpopulationsto vehicles under the MODEL1 scheme is not dependent on the vehicle ordering. For uniform fleets, MODEL2's sensitivity to vehicle ordering is of no consequence. For non-uniform fleets, the discrepancy in the final objective value is of the order of that which results from using alternate initial solutions. Fortunately, MODEL2 proves to be much more acceptable than MODEL1, with respect to the divergence of solutions resulting from different initializations. Before proceeding to evaluate the performance of MODEL2, it is advantageous to develop an upper bound algorithm for generating standards against which solution values may be measured. The final result for the original vehicle order of Fig. 18 is shown in Fig. 28.

```
 SCHEDULE FOR VEHICLE   1
DEPARTS NODE     AT TIME      PASSENGERS ABOARD

          2         1.08         4    5
          1         6.00        11
          2         8.08        11   13

BENEFIT ACHIEVED BY VEHICLE   1 =       7.085


 SCHEDULE FOR VEHICLE   2
DEPARTS NODE     AT TIME      PASSENGERS ABOARD

          2         7.17         9   10

BENEFIT ACHIEVED BY VEHICLE   2 =       3.989


 SCHEDULE FOR VEHICLE   3
DEPARTS NODE     AT TIME      PASSENGERS ABOARD

          1         0.25         2
          2         1.67         2    3
          1         3.33         6    7
          2         5.00         6
          1         7.17        12   13
          2         8.58        12   14

BENEFIT ACHIEVED BY VEHICLE   3 =      48.947


 SCHEDULE FOR VEHICLE   4
DEPARTS NODE     AT TIME      PASSENGERS ABOARD

          2         0.00         1
          1         1.42         3    4
          2         3.92         0
          1         5.33         9   10
          2         7.00         7    8
          1         8.42        14

BENEFIT ACHIEVED BY VEHICLE   4 =      35.437

BENEFIT ACHIEVED IN SCHEDULING PROCESS =      95.458
BENEFIT ACHIEVED IN ASSIGNMENT PROCESS =       0.000

THE PROGRAM HAS NOT YET CONVERGED


 THE VEHICLE SCHEDULES HAVE NOT YET CONVERGED -
```

Figure 27. Illustrative MODEL2 Computation.
(a) Initial Iteration.

```
 SCHEDULE FOR VEHICLE   1
DEPARTS NODE     AT TIME      PASSENGERS ABOARD

          2          1.17        4   5

BENEFIT ACHIEVED BY VEHICLE   1 =       0.553


 SCHEDULE FOR VEHICLE   2
DEPARTS NODE     AT TIME      PASSENGERS ABOARD

          2          7.25        9  10

BENEFIT ACHIEVED BY VEHICLE   2 =       0.569


 SCHEDULE FOR VEHICLE   3
DEPARTS NODE     AT TIME      PASSENGERS ABOARD

          1          0.25        2
          2          1.67        2   3
          1          3.33        6   7
          2          5.00        6
          1          7.17       12  13
          2          8.58       12  14

BENEFIT ACHIEVED BY VEHICLE   3 =      56.852


 SCHEDULE FOR VEHICLE   4
DEPARTS NODE     AT TIME      PASSENGERS ABOARD

          2          0.00        1
          1          1.42        3   4
          2          3.92        0
          1          5.33        9  10
          2          7.00        7   8
          1          8.42       14

BENEFIT ACHIEVED BY VEHICLE   4 =      39.248

BENEFIT ACHIEVED IN SCHEDULING PROCESS =      97.222
BENEFIT ACHIEVED IN ASSIGNMENT PROCESS =       0.000

THE PROGRAM HAS NOT YET CONVERGED


 THE VEHICLE SCHEDULE HAS CONVERGED, MATCH IS NOW APPLIED
```

Figure 27. (b) Second Iteration.

THE SCHEDULE

SCHEDULE FOR VEHICLE 1

DEPARTS NODE	AT TIME	PASSENGERS ABOARD
2	1.17	4 5

BENEFIT ACHIEVED BY VEHICLE 1 = 0.553

SCHEDULE FOR VEHICLE 2

DEPARTS NODE	AT TIME	PASSENGERS ABOARD
2	7.25	9 10

BENEFIT ACHIEVED BY VEHICLE 2 = 0.569

SCHEDULE FOR VEHICLE 3

DEPARTS NODE	AT TIME	PASSENGERS ABOARD
1	0.25	2
2	1.67	2 3
1	3.33	6 7
2	5.00	6
1	7.17	12 13
2	8.58	12 14

BENEFIT ACHIEVED BY VEHICLE 3 = 56.852

SCHEDULE FOR VEHICLE 4

DEPARTS NODE	AT TIME	PASSENGERS ABOARD
2	0.00	1
1	1.42	3 4
2	3.92	0
1	5.33	9 10
2	7.00	7 8
1	8.42	14

BENEFIT ACHIEVED BY VEHICLE 4 = 39.248

BENEFIT ACHIEVED IN SCHEDULING PROCESS = 97.222
BENEFIT ACHIEVED IN ASSIGNMENT PROCESS = 129.223

THE SCHEDULE HAS CONVERGED

Figure 27. (c) Final Iteration.

```
                                      THE SCHEDULE


 SCHEDULE FOR VEHICLE   1
DEPARTS NODE    AT TIME     PASSENGERS ABOARD

          1        0.33        2
          2        1.83        4   5
          1        5.33        9  10
          2        7.00        7   8
          1        8.42       14

BENEFIT ACHIEVED BY VEHICLE   1 =      48.755


 SCHEDULE FOR VEHICLE   2
DEPARTS NODE    AT TIME     PASSENGERS ABOARD

          2        1.58        2   3
          1        3.33        6   7
          2        5.00        6
          1        7.17       12  13
          2        8.58       13  14

BENEFIT ACHIEVED BY VEHICLE   2 =      51.517


 SCHEDULE FOR VEHICLE   3
DEPARTS NODE    AT TIME     PASSENGERS ABOARD

          2        7.25        9  10

BENEFIT ACHIEVED BY VEHICLE   3 =       0.569


 SCHEDULE FOR VEHICLE   4
DEPARTS NODE    AT TIME     PASSENGERS ABOARD

          2        8.08       11  12

BENEFIT ACHIEVED BY VEHICLE   4 =       0.505

BENEFIT ACHIEVED IN SCHEDULING PROCESS =    101.346
BENEFIT ACHIEVED IN ASSIGNMENT PROCESS =    129.348

THE SCHEDULE HAS CONVERGED
```

Figure 28. Results for Alternate Vehicle Ordering.

3.9 Upper Bounds

The upper bound will be developed as the minimum of two upper bounds. The first bound will be based on the limitation of available vehicle capability. The second bound will be found as a limit to service that the traveler population can absorb.

A. Bound based on vehicle capability. If a vehicle j is optimally scheduled by itself with respect to the total traveler population, it will achieve a total benefit (value) $\hat{v}_j$, greater than or equal to the component benefit v_j that the same vehicle might provide under any scheduling of the entire vehicle fleet.

To see this, assume that under some scheduling S of the entire fleet, vehicle j achieves a benefit $v_j > \hat{v}_j$. Then $\hat{v}_j$ cannot be the optimal benefit for vehicle j scheduled alone. This is contrary to the definition of $\hat{v}_j$. Therefore, $\hat{v}_j \geq v_j^o$ where v_j^o is the component benefit that vehicle j contributes to the total benefit B^o of the optimal fleet schedule S^o. Hence, an upper bound B^1 can be obtained by summing the $\hat{v}_j$ over all vehicles:

$$B^1 = \sum_{j=1}^{V} \hat{v}_j \geq \sum_{j=1}^{V} v_j^o = B^o , \tag{3-33a}$$

or,

$$B^1 \geq B^o . \tag{3-33b}$$

The values $\hat{v}_j$ may be obtained by applying the single vehicle scheduling algorithm for each vehicle j.

The bound B^1 is weak in the following sense. If the vehicle fleet capacity is very large relative to the number of potential passengers, then B^1 may far exceed B^o. It is easy to see, for example, that B^1

can be arbitrarily increased simply by adding more vehicles to the fleet. Even with a vehicle fleet size well matched to the size of the traveler market, B^1 will closely approximate B^o only under an unusual distribution of passenger time preferences. In particular, the time preference distribution must be such that under optimal fleet scheduling, each vehicle gets the same usage as it would if it alone were (optimally) scheduled with respect to the total traveler population.

To help remedy this difficulty, a new bound B^2 is found, based on a limit to the service that the traveler population can absorb:

B. Bound based on passenger service. The bound B^2 will be found by assigning travelers to a better alternative set of trips than could possibly be available, even under an optimal fleet schedule. This new set is composed of imaginary ideal trips; each such trip is assumed to serve each of its passengers with a perfect schedule (zero inconvenience) relative to the speed of the vehicle that executes the trip. To develop the new set of alternative trips, which will be called the "maximum inventory" of available trips, the maximum number of trips that each vehicle j can make during the scheduling period $(0,T)$ must be determined. Vehicle j can make up to n_j departures from each node, where

$$2n_j = \frac{T\sigma_j}{\lambda} + 1 \quad \text{rounded to next smallest integer ,} \tag{3-34}$$

where

λ = distance of a one-way trip,

σ_j = average speed of vehicle j .

If $2n_j$ is odd, it will be assumed that vehicle j makes $(2n_j+1)/2$

trips from each node (i.e., replace n_j by $(2n_j+1)/2$). Thus, the maximum inventory of trips consists of n_j trips by each vehicle j from each node, or a total of $2Q = 2\sum_{j=1}^{V} n_j$.

It is assumed that every passenger is served by a "perfect schedule" (t_d^*, t_a^*) relative to the speed of the trip in the maximum inventory to which he is assigned. Equivalently, it is assumed that every trip in the maximum inventory serves each of its passengers with a perfect schedule. Finally, operating costs will be considered on a per passenger basis, such that if a passenger is assigned to a trip, he will be allocated a minimal share of the operating expense of his trip. (A minimal share is computed by dividing the total operating cost of a trip by the passenger load, assuming that the vehicle is full.)

To be more precise, passengers will be considered for assignment to trips in the maximum inventory, on the basis of the values

$$v_{ih}^M = -a_1 \frac{OC_h}{C_h} + a_2\pi_h + a_3 WTP^i(\Delta = 0, \phi_h) \ , \qquad (3\text{-}35)$$

where v_{ih}^M is the value or benefit of assigning passenger i to trip h in the maximum inventory of trips. The other symbols are:

OC_h = operating cost of trip h ,

C_h = capacity of the vehicle making trip h ,

π_h = fare charged for trip h ,

ϕ_h = duration of trip h .

It will be shown that if passengers are optimally matched to trips in the maximum inventory such that $B^2 = \sum_i v_{ih(i)}$ is maximized, then that sum constitutes a valid upper bound on the benefit of the optimum schedule. ($h(i)$ is the trip h to which passenger i is assigned.)

First, consider the set of alternative trips in the optimal schedule. Each traveler can be associated with a set of values,

$$v_{ih}^{o} = -a_1 \frac{OC_h}{L_h} + a_2\pi_h + a_3 WTP^i(\Delta_h, \phi_h) , \tag{3-36}$$

where v_{ih}^{o} is the benefit of assigning passenger i to trip h (via vehicle j) in the optimal schedule. Also,

L_h = load or number of passengers on trip h, $0 \leq L_h \leq C_h$,

$\frac{OC_h}{L_h}$ = the share of operating cost that can be allotted to a passenger on trip h,

Δ_h = the schedule inconvenience of trip h in the optimal schedule, with respect to traveler i's preferences.

Note that

$$\min_{L_h} \left(\frac{OC_h}{L_h}\right) = \frac{OC_h}{C_h} .$$

Consider the following summation, which assumes that every traveler i has been assigned to a specific trip (or the null trip) in the optimal schedule:

$$\sum_{i=1}^{N} v_{ih(i)}^{o} = \sum_{i=1}^{N} \left(-a_1 \frac{OC_{h(i)}}{L_{h(i)}} + a_2\pi_{h(i)} + a_3 WTP^i(\Delta_{h(i)}, \phi_{h(i)}\right) , \tag{3-37}$$

where $h(i)$ is the trip to which passenger i is assigned, and N is the total number of passengers.

Now each trip h has L_h passengers on board, according to definition. Thus, there are L_h passengers for whom $h(i) = h$. Therefore, the first term on the right hand side can be written

$$-a_1 \sum_{i=1}^{N} \frac{OC_{h(i)}}{L_{h(i)}} = -a_1 \sum_{h=1}^{H} L_h \frac{OC_h}{L_h} e_h = -a_1 \sum_{h=1}^{H} OC_h e_h , \tag{3-38}$$

where

$$e_h = \begin{cases} 0 & \text{if } L_h = 0 , \\ 1 & \text{if } L_h \neq 0 , \end{cases}$$

and H is the total number of trips in the optimal schedule.

Thus, the overall summation can be rewritten,

$$\sum_{i=1}^{N} v^{o}_{ih(i)} = -a_1 \sum_{h=1}^{H} OC_h e_h + a_2 \sum_{i=1}^{N} \pi_{h(i)} + a_3 \sum_{i=1}^{N} WTP^{i}(\Delta_{h(i)}, \phi_{h(i)}) \ . \tag{3-39}$$

Now by definition, the optimal solution assigns passengers in an optimal way to the optimal schedule. Thus, under the optimal solution, the total benefit above,

$$B^{o} = \max_{\{h(i)\}} \left(\sum_{i} v^{o}_{ih(i)} \right) - a_1 OC_e \quad \text{is maximized,}$$

where $\{h(i)\}$ is the set of traveler to trip assignments and OC_e is the total operating cost of empty trips in the optimal schedule.

Now, for every trip in the optimal schedule, there is a trip in the maximum inventory that is at least as good, since: (a) for every trip by vehicle j in the optimal schedule, there is at least one trip by the same vehicle in the maximum inventory, and (b) that trip in the maximum inventory is assumed to provide a perfect schedule to each of its passengers. Such service is attainable by a trip in the optimal schedule only in the ideal case.

Therefore, consider a subset of trips in the maximum inventory, the elements of which are in a one-one correspondence with those in the optimal schedule. "Pair off" trips in the optimal schedule with trips by the same vehicle in the subset of the maximum inventory. Order these pairs by the trip index h. Then

$$v_{ih}^{M} \geq v_{ih}^{o} \quad \text{all} \quad i,h, \tag{3-40a}$$

or

$$-a_1 \frac{OC_h}{C_h} + a_2\pi_h + a_3 WTP^i(\Delta = 0, \phi_h) \geq -a_1 \frac{OC_h}{L_h} + a_2\pi_h + a_3 WTP^i(\Delta_h, \phi_h) \ , \tag{3-40b}$$

since

$$C_h \geq L_h \quad \text{and} \quad WTP^i(\Delta = 0, \phi_h) \geq WTP^i(\Delta_h, \phi_h) \ .$$

Thus, if we solve for the optimal assignment of passengers to the subset of trips in the maximum inventory, on the basis of the values v_{ih}^{M} , then the resulting total benefit B is greater than or equal to B^o, where B^o is the benefit of the optimal solution. That is,

$$\max_{\{h(i)\}} \sum_i v_{ih(i)}^{M} = B \geq \max_{\{h(i)\}} \sum_i v_{ih(i)}^{o} - OC_e = B^o \ . \tag{3-41}$$

Now, if the optimal assignment of travelers to a subset of the maximum inventory results in a benefit not less than B^o, then the optimal assignment to the <u>total</u> maximum inventory must yield a benefit $B^2 \geq B \geq B^o$. Thus, B^2 is a valid upper bound on B^o .

The optimal assignment of travelers to the maximum inventory of trips can be computed using the optimal assignment algorithm. As a practical matter, however, it is not necessary to actually solve an assignment problem if only one traveler "type" is considered, i.e., if all travelers are assumed to have identical preference parameter values (D,α,n,s). In such a case, all travelers have identical values and relative preferences for alternative trips in the maximum inventory, and it is easy to allocate passengers to trips in the best possible way. All that is required is to "fill up" the trips from each node, in decreasing order of benefit,

$$v_h^M = -a_1 \frac{OC_h}{C_h} + a_2\pi_h + a_3 WTP(0,\phi_h) \ .$$

Thus, reordering the trips by value v_h^M, and using a new index p,

$$B^2 = \sum_{p=1}^{q-1} C_p v_p^M + v_q^M L_q \ , \tag{3-42}$$

where

q = last trip to be loaded before all travelers have been allocated,

L_q = remaining number of passengers loaded onto the trip q.

In the general case where travelers are not all of one type, and it is still desired to avoid having to solve an optimal assignment problem, B^2 can be replaced by an upper bound on the solution to that assignment problem. That is,

$$B^2 \leq \min\left(\sum_{i=1}^{N} \max_h v_{ih}^M, \ \sum_h C_h \max_i v_{ih}^M\right) = B^3 \ . \tag{3-43}$$

The first term within the minimization operation of (3-43) corresponds to assigning each traveler to his most favored (fastest) trip in the maximum inventory. The second term corresponds to filling each trip in the maximum inventory with maximally valued travelers. Each of these terms constitutesan upper bound to the optimal assignment of travelers to trips in the maximum inventory.

By virtue of the service factors that it idealizes, bound B^2 cannot be held as a very "tight" upper bound on the value of the optimal scheduling solution. Recall that B^2 assumes an abundance of trips by all vehicles, perfect schedules, and underestimated operating costs. However, B^2 serves fairly adequately, to measure the limit on service which can be absorbed by the traveler population. Thus, B^2 complements bound B^1 so that the bound $B = \min(B^1, B^2)$ may be reasonably indicative of the maximum achievable benefit.

In summary, two upper bounds have been set forth. The first is based on a limit of vehicle capabilities with respect to the given distribution of passenger preferences. The second is based on a limit to the amount of service that the set of travelers can absorb. A simplified flow chart illustrating the computation of B^1 and B^2 is given in Fig. 29. The upper bound routine to compute $B = \min(B^1, B^2)$ is coded as an optional subroutine to the MODEL1 and MODEL2 programs.

3.10 Results

Fig. 30 displays the set of MODEL2 results for example B. (The objective is maximization of consumer surplus.) The figure shows MODEL2 solution values (with and without the optimum assignment option as designated in Fig. 26) from the nine alternate initializations. The upper bound value and MODEL1 results are also summarized.

Two features of these results are outstanding. First, the quality of MODEL2 solutions, in terms of the achieved level of the objective function, is much superior to that of MODEL1 solutions. Second, MODEL2 solutions obtained from different initializations are in close agreement. These are the two areas of improvement that were sought, in developing MODEL2.

Before discussing these areas further, two other characteristics are to be noted. First, the number of iterations required for convergence is

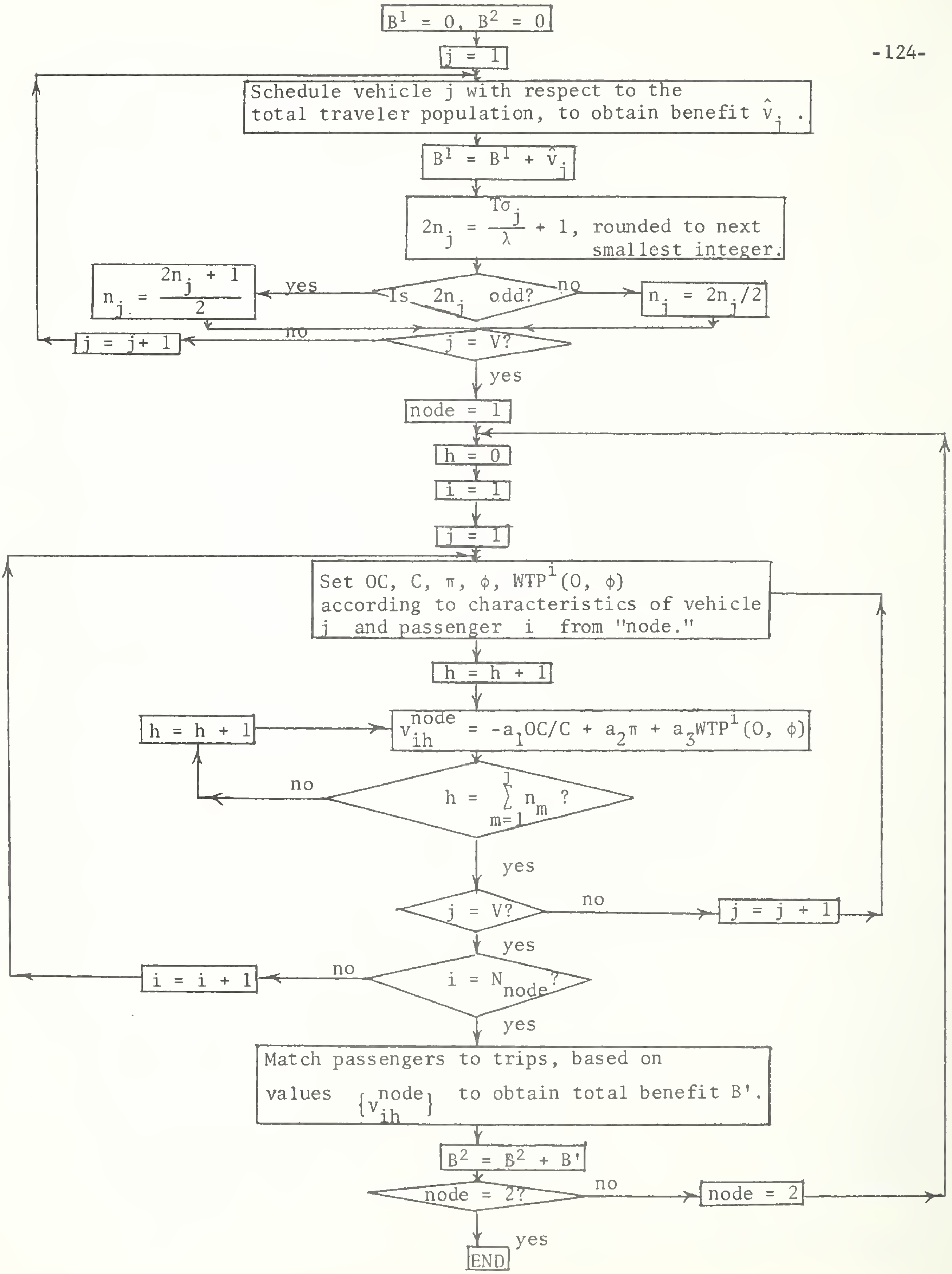


Figure 29. Computation of Upper Bounds.

	MODEL1		MODEL2				
					No. of Iterations*		
Initial Sol'n	Sol'n Value	No. of Iterations	Value without Opt. Assignm't	Value with Opt. Assignm't	W/out Opt. Assignm't	With Opt. Assignm't	Comment
S1	116.4	2	136.4	136.4	2	3	Solutions to S1 and S2 differ on approx. 20% of the timetable.
S2	67.6	2	140.6	140.6	2	3	
S3	73.2	2	140.6	140.6	2	3	Solution identical to S2 solution
S4	98.1	2	134.4	136.4	2	4	Solution identical to S1 solution
A1	108.1	3	135.0	136.4	4	6	"
A2	119.8	3	133.4	133.4	4	5	Very close to S1 solution
A3	91.3	3	139.2	139.2	4	5	Very similar to S2 solution - timetables of slower vehicles diverge.
A4	108.2	4	136.4	136.4	4	5	Solution identical to S1 solution.
None	-	-	140.6	140.6	2	3	Solution identical to S2 solution

Upper Bounds: B1 = 177.0, B2 = 219.1 min(B1,B2) = 177

*An iteration is registered after each fleet timetable computation. MODEL1 takes approx. 12 sec/iteration; MODEL2 takes 18 sec/iteration on IBM 360/67.

Figure 30. Comparison of Solutions by MODEL1 and MODEL2.

greater where initialization (by timetable or assignment specification) leaves few travelers unassigned. In these cases, the program spends substantial time "undoing the mistakes" in the initial solution. This suggests that it is advantageous to specify no initialization, and allow computations to take their own course. Recall that the best solution in Fig. 30 is obtained with no initialization. However, this is not always the case.

Another computationally important point is that use of the optimal assignment algorithm does not appear to improve the solutions significantly. When MODEL2 is extended to the scheduling of vehicles over more complex networks in Chapter VI, the latter point will be highly relevant.

Returning to the question of solution quality, observe that MODEL1 results in Fig. 30 achieve from 38.2 to 67.7 per cent of the upper bound value, whereas MODEL2 values range from 75.4 to 79.4 per cent. In view of the conservativeness of the upper bound and the consistency of MODEL2 solutions, there is some assurance that the solutions are at least close to the optimal one.

The improvement in the consistency of solutions of MODEL2 over those of MODEL1 is just as remarkable as the improvement in solution quality. The variation among MODEL1 solutions is 43.5 per cent of the highest achieved value. The corresponding statistic for MODEL2 is 4.9 per cent. A closer look at the final MODEL2 timetables will allow greater insight into the nature of the solutions.

Fig. 31 displays the four distinct schedules that result from the nine MODEL2 runs. The similarity among the solutions is apparent. For example, eleven of the twenty-five possible trips are found in each

Passengers Aboard*

Departure Times	S1		S2		A2		A3	
I. Fast Vehicles (1&2)								
From node 1	V1	V2	V1	V2	V1	V2	V1	V2
0.33	2	-	2	-	-	2	2	-
1.42	-	-	-	3,4	-	-	-	3,4
3.33	-	6,7	6,7	-	-	6,7	6,7	-
5.33	9,10	-	-	9,10	9,10	-	-	9,10
7.17	-	12,13	12,13	-	-	12,13	12,13	-
8.42	14	-	-	14	-	-	-	14
8.58	-	-	-	-	14	-	-	-
From node 2								
0.00	-	-	-	1	-	-	-	1
1.58	-	2,3	-	-	2,3	-	-	-
1.83	4,5	-	4,5	-	-	4,5	4,5	-
3.92	-	-	-	0	-	-	-	0
5.00	-	6	6	-	-	6	6	-
7.00	7,8	-	-	7,8	-	-	-	7,8
7.17	-	-	-	-	9,10	-	-	-
8.58	-	13,14	13,14	-	-	13,14	13,14	-
II. Slow Vehicles (3&4)								
From node 1	V3	V4	V3	V4	V3	V4	V3	V4
1.50	3	-	-	-	-	3	-	-
1.83	-	-	-	5	-	-	-	-
4.00	-	8	8	-	8	-	-	8
6.00	11	-	-	11	-	11	11	-
From node 2								
0.92	-	-	2,3	-	-	-	2	-
1.00	-	-	-	-	-	-	-	3
3.92	0	-	-	0	-	0	-	-
7.08	-	-	-	-	7,8	-	-	-
7.25	-	9,10	9,10	-	-	-	-	9,10
8.08	11,12	-	-	11,12	-	11,12	11,12	-
Objective Value	136.4		140.6		133.4		139.2	

*Passengers with the same number, but from different nodes, are distinct.

Figure 31. Solutions by MODEL2.

schedule. Thus, the computations from divergent initial points appear to gravitate toward a common solution; however, the process does not always attain the same level. There are suboptimal points at which MODEL2 may stop. This will be illustrated more dramatically, below. The probability of premature termination varies directly with the attained level of the objective function. Thus, it is unlikely that MODEL2 terminates far from the optimal solution.

This behavior is explained by the fact that MODEL2 continues to improve the solution as long as there is some traveler in the system for whom the net benefit of transferring to a new vehicle outweighs the benefit currently received by some other traveler (or empty seat), already assigned to that new vehicle. The probability that such a potential traveler transfer exists, decreases as the solution improves. However, there are suboptimal points, as well as optima, for which no such potential transfer exists.

The probability of suboptimal termination is also a decreasing function of dimensionality. The larger the problem (the more travelers and vehicles), the less likely it is that no travelers can be found for whom a transfer provides a net system benefit. Thus, MODEL2's "freedom to climb" is enhanced by large dimensionality but is mitigated as the summit is approached.

The examples below, illustrate how MODEL2 may terminate suboptimally. Following the examples, this chapter concludes with an explanation and summary of claims on the merits of MODEL1 and MODEL2.

The array of Fig. 32 represents a problem with two passengers and two dissimilar vehicles. Each vehicle has unit capacity. Each box in the array corresponds to a particular passenger-to-vehicle assignment. An entry v_{ij} is the system benefit achieved, as computed by the vehicle scheduling algorithm, if vehicle j serves passenger i (exclusively).

In this context, consider the MODEL2 computational process, applied with no initialization:

1. With both travelers unassigned, vehicle V1 chooses to serve passenger P1, attaining a benefit of 9.

2. Vehicle V2 is faced with the following choice:

   a. Serve passenger P2--net system benefit = 8 ,

   b. Take passenger P1 away from V1--net system benefit = (12-9) = 3 .

   Thus, V2 chooses option (a), to serve P2.

3. Vehicle V1 may:

   a. Continue to serve P1--net benefit = 0 ,

   b. Drop P2, serve P1--net benefit = -8+(12-9) = -5 .

   Thus V2 continues to serve P2.

Therefore, the final solution is (P1 - V1, P2 - V2) with value 9 + 8 = 17. It is clear, however, that the optimal solution is (P1 - V2, P2 - V1), with value 12 + 6 = 18. The latter solution is obtained with MODEL2 if the vehicle order is reversed.

Figs. 33a, b, and c display a more complicated problem, with four travelers and two identical vehicles of capacity two. Fig. 33a illustrates the approximate preferred departure times of the travelers. The passengers may be pictured as traveling from the same station. Fig. 33b tabulates the maximum benefit (computed by the single vehicle algorithm) that could be achieved by a vehicle with each possible passenger

Vehicle / Passenger	V1	V2
P1	9	12
P2	6	8

Capacity of each vehicle is one. Vehicles have different speeds and/or fare.

Figure 32. Two Vehicle-Two Passenger Example of MODEL2 Solution.

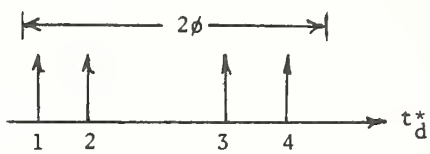


Figure 33a. Representation of Preferred Departure Times of Travelers.

Passenger Combination	Optimal Benefit from Best Vehicle Schedule
1,0	10
2,0	10
3,0	10
4,0	10
1,2	9+9
1,3	7+7
1,4	9+8
2,3	10+0
2,4	6+6
3,4	8+8

Figure 33b. Maximum Benefit for Each Possible Boarding List.

Figure 33. Two Vehicle-Four Passenger Example of MODEL2 Solution.

Aboard V2 \ Aboard V1	1,2	1,3	1,4	2,3	2,4	3,4	1,0	2,0	3,0	4,0
1,2						16 / 18			10 / 18	10 / 18
1,3					12 / 14			10 / 14		10 / 14
1,4				10 / 17				10 / 17	10 / 17	
2,3			17 / 10				10 / 10			10 / 10
2,4		14 / 12					10 / 12		10 / 12	
3,4	18 / 16						10 / 16	10 / 16		
1,0				10 / 10	12 / 10	16 / 10		10 / 10	10 / 10	10 / 10
2,0		14 / 10	17 / 10			16 / 10	10 / 10		10 / 10	10 / 10
3,0	18 / 10		17 / 10		12 / 10		10 / 10	10 / 10		10 / 10
4,0	18 / 10	14 / 10		10 / 10			10 / 10	10 / 10	10 / 10	

Figure 33c. Benefits of Alternate Passenger - Vehicle Arrangements.

combination aboard. For example, the figure shows that if any of the four travelers were boarded alone, a benefit of ten units would accrue. Alternately, the combination of passengers 1 and 4 could best be served by a schedule that affords passenger 1 nine units of benefit, and passenger 4 eight units. Fig. 33c displays the benefits of all feasible passenger-vehicle arrangements. The optimal arrangement is (1,2), (3,4) with a total value 18 + 16 = 34. Consider the MODEL2 computation, beginning with no initialization:

1. V1 chooses to serve (1,2), the traveler combination of highest value--net benefit = 18.

2. V2 chooses to serve (3,4)--net benefit = 16.

Hence, the optimal solution is reached. Suppose, however, that for some reason, perhaps by initialization, the solution (V1-(1,3), V2-(2,4)) is reached. This solution has value 14 + 12 = 26. The following tabulation shows that MODEL2 terminates at this suboptimal point:

1. Vehicle V1 has the following options:

   a. Add passenger 2, drop passenger 1--net gain = (10-6)-7 = -3,

   b. Add passenger 2, drop passenger 3--net gain = (9-6)-7 = -4,

   c. Add passenger 4, drop passenger 1--net gain = (8-6)-7 = -5,

   d. Add passenger 4, drop passenger 3--net gain = (8-6)-7 = -5,

   e. Add passengers 2 and 4, drop passengers 1 and 3--net gain = (12-12)-26 = -26,

   f. Retain passengers 1 and 3--net gain = 0.

Option (f) is best, so no change will occur in the schedule for V1.

2. Vehicle V2 has the following options:

   a. Add passenger 1, drop passenger 2--net gain = (9-7)-6 = -4,

b. Add passenger 3, drop passenger 2--net gain = (8-7)-6 = -5,

c. Add passenger 1, drop passenger 4--net gain = (9-7)-6 = -4,

d. Add passenger 3, drop passenger 4--net gain = (0-7)-6 = -13,

e. Add passengers 1 and 3, drop passengers 2 and 4--net gain = (14-14)-12 = -12,

f. Retain passengers 3 and 4--net gain = 0.

Thus, no change occurs in the schedule of V2, either.

3.11 Concluding Discussion

The examples in this chapter have clearly illustrated that the iterative methods developed here are subject to pitfalls that may lead to suboptimal results. However, the computer runs have demonstrated that MODEL2 produces self-consistent results that are clearly superior to those of MODEL1, and which closely approximate the optimal. The following claims on MODEL2 solutions can be made:

1. Let $F(A,S)$ = the objective function value for assignment solution A and scheduling (timetable) solution S,

and let

(A_o,S_o) = the final solution from a MODEL2 computation.

Then,

$$F(A_o,S_o) = \max_A F(A,S_o) = \max_S F(A_o,S) \ . \qquad (3\text{-}44)$$

That is, given the solution (A_o,S_o), then A_o is the optimal assignment of passengers to vehicle for the solution timetable S_o. Similarly, S_o is the optimal schedule for the passenger allocation A_o. Eq. (3-44) follows from the optimality of the passenger assignment and vehicle scheduling algorithms. To be entirely accurate, it must be assumed that the optimal assignment algorithm is applied to the final timetable. However, the practical consequence of ignoring this requirement is small.

Eq. (3-44) also applies to MODEL1 solutions. In essence, (3-44) characterizes the "one-at-a-time" optimization in $(\mathcal{a}, \mathcal{s})$ space pictured in Fig. 25. Hence, in addition to Eq. (3-44), further claims can be made for the merit of MODEL2 solutions:

2. MODEL2 solutions are optimal to within a "first-order," unilateral exchange of travelers among vehicles. As stated earlier, a property of the solution is that it is impossible simultaneously to transfer passengers <u>to</u> a vehicle, and discharge other passengers from that vehicle, such that a net gain to the system results. Thus, improvement to the solution, subsequent to terminations of MODEL2 iterations, is possible via <u>higher-order passenger trades</u>, only.

Conclusion (2) is intuitively satisfying if one considers that the mode of optimization is a one-at-a-time scheduling of vehicles. MODEL2 is, therefore, a <u>first-order</u> process and does not consider benefits that are attainable by simultaneous adjustments in the operation of two or more vehicles. For instance, two-way trading of travelers among vehicles is a <u>second-order</u> process that must be accomplished through optimization of two or more vehicle schedules, <u>simultaneously</u>. This is not to say, however, that it is unlikely to reach an optimum without higher order operations.

The computational barriers are now clearly defined. In order to synthesize a more accurate algorithm, it is necessary to optimize over two or more vehicles at a time. However, the motivation for single vehicle optimization is the intractibility of straight-forward multi-vehicle optimization. (Recall the dynamic programming formulation and discussion.) The path to improving the algorithm further, is constrained by these opposing factors. Two alternatives may be feasible. One approach is to test a random series of two-way passenger "trades," around the final solution. If a sufficiently large sample of such perturbations fails to improve the solution then one would have some statistical assurance that an optimum has been reached. A second approach is to synthesize a "two-vehicle algorithm" from the dynamic programming formulation given earlier. This approach would require extremely efficient programming to be useful. Although these ideas will not be pursued here, they constitute an area for possible future research. Meanwhile, MODEL2 can be used with the confidence that solutions are likely to be near optimal, and that utilization of several initial points will enhance the likelihood of achieving an optimum.

CHAPTER IV

UNCERTAINTY AND THE CHOICE AMONG ALTERNATIVE SYSTEMS

4.1 Introduction

In Chapter III, a method was developed for computing the schedules and corresponding benefit levels for transportation systems in a completely deterministic setting. In particular, passengers were characterized by preference functions having known parameter values. Given the present state of knowledge of traveler preferences and the inherent random nature of traveler behavior, however, it is wise to consider how to account for uncertainties when MODEL2 is called upon to help choose the best system among a set of proposed alternatives. The only uncertainties to be considered here are those associated with traveler preference parameters, and the levels of potential demand. Uncertainties in physical system parameters such as vehicle speeds or costs, are not analyzed.

The decision-making method presented here is designed to choose among alternatives, based on the objective function components (costs, revenues, and willingness to pay) only. In practice, other factors such as safety, comfort, accessibility, external community benefits, other social objectives related to transportation--literally a host of factors--must be weighed before a decision can be made. The methodology of this chapter can merely aid in decision-making, by helping to choose the alternative that provides the highest operational benefit. Looked at another way, the method here is designed to choose the best system, given that "all other factors are equal."

In order to choose among alternative systems in the face of uncertainty, the following task must be addressed:

Find the schedule (timetable) that maximizes the expected value of the chosen objective, and compute the corresponding expected value of the objective function.

Once the solution to this problem is available for each alternative system, the most desirable system may be chosen on the basis of greatest expected value. However, the above task is not easily accomplished in view of the complex nature of traveler behavior and the difficulty of optimization, even for the deterministic case. In particular, the MODEL2 methodology is not suited to the direct incorporation of uncertainties. Hence, the following mode of analysis will be employed. Traveler populations will be generated at random from the probability distributions that describe the "traveler generating process." Then scheduling solutions will be computed for the random traveler inputs, using the MODEL2 method, and the results of these computations will be analyzed to determine the optimal choice among alternative systems.

Even with the above approach, accomplishing the central task of computing the maximum expected value and corresponding timetable for a given system requires some compromising assumptions. The first such assumption pertains to the manner of determining the "fixed timetable" by which a given system operates best. Specifically, one of the following two premises may be accepted:

P1 (Adaptive Timetables). The timetable is computed only after the actual traveler population (an outcome of the random traveler generating process) is known. Once the (sample) population has materialized, the preference parameters describing each traveler are assumed known with certainty.

P2 (Permanent Timetable). A single fixed timetable is computed on the basis of uncertain knowledge of traveler characteristics. This timetable, which will be called the "optimal stochastic timetable," is permanent as long as the state of knowledge of the random traveler generating process remains unchanged.

The first premise is the easier to implement into a method for choosing among alternatives because it permits the use of MODEL2 to transform random traveler inputs directly into random sample benefits for each proposed system. However, P1 is the less realistic of the two, and diverges from the basic definition of a fixed schedule transportation system. Nevertheless, the analysis for premise P1 will be developed below, both for its own merit and for the opportunity it provides to illustrate the general (Bayesian) framework of analysis that is used in this chapter.

The analysis for choosing among alternative systems under premise P2 requires some additional assumptions or approximations. First, it will be assumed that the levels of potential traveler demand (N) considered here are sufficiently large so that the statistical properties of a random traveler population of size N closely approximate the properties of the distribution from which the population is generated. Thus, an (optimal) timetable computed on the basis of a particular random traveler population sample of size N will be (approximately) optimal for any other such sample population of size N. Hence, only one traveler population need be generated to compute the optimal timetable for any given level of potential demand N.

A second approximation involves the set of timetables from which the optimal stochastic timetable is drawn. Specifically, it will be assumed that the optimal stochastic timetable (which maximizes the expected value of the objective function) is also optimal for some deterministic level of potential demand (N). That is, the analysis here will be restricted to searching over that set of schedules that are optimal for some single level of traveler demand.

Before proceeding to the analyses, a few additional preliminary comments are in order. First, consider the question of scale size in the performance of scheduling computations. Recall from the previous chapter that MODEL2 receives input values for each traveler in the system. In the current chapter, the same requirement holds, although traveler values will be selected randomly. For large systems, with thousands of travelers, the advisability of "full scale" computations is questionable. Thus, it may frequently be necessary to analyze a "toy" system, that is scaled down to manageable size. The scaling is effected by reducing the number of potential travelers, the capacity of each vehicle, and the vehicle operating costs, all in the same proportion.

In general, the larger the scale, the more "accurate" the computation, in the following ways. First, the smaller the scale the more serious are the "round-off" errors. Note, for instance, that travelers in any MODEL2 scheduling computation are treated as integral units, no matter what the scale (i.e., travelers are not allocated to trips in fractional parts). Now, at one-tenth scale there will be only one traveler for every ten at full scale. Suppose that at full scale seventy-five of the one-hundred travelers are accommodated by the solution schedule. Then, presumably, at one-tenth scale either seven or eight of ten travelers will be accommodated. Thus, the resultant benefit level, when rescaled to full size, will be in error. The second effect of scale on computational accuracy concerns the statistical properties of a sample traveler population that is drawn from the overall probability distribution of preference parameter values and used with MODEL2 for computing a schedule. In particular, the larger the size (scale) of the sample traveler population, the better the sample approximates the distribution from which the sample is drawn. To summarize, the analyst is faced with a trade-off between accuracy and computational cost, in choosing an appropriate scale size.

It has been noted above that the methods of this chapter require the selection of sample values from specified (multivariate) probability distributions. The technique of random sampling, a basic element of Monte Carlo methods,† may be used to generate the necessary sample values. The random sampling technique is based on the generation of "random numbers," i.e., numbers that have the same statistical properties as the desired random variables. Here, sample values from the multivariate probability distributions $f_{\underline{N}}(N_1,N_2,\ldots,N_R)$ and $f^r_{\underline{P}|\underline{N}}(t^*,w,D,\alpha,n,s|\underline{N})$, will be required. The first of these distributions is the joint density function over the levels of potential demand (N_r) over each route or trip type (r). (In the shuttle system, for example, there are two trip types, node 1 to node 2, and node 2 to node 1.) The second distribution is the joint density function for the preference parameters $\underline{P}$ of any potential traveler for trip type r, conditional on the demand levels $\underline{N}$.

---

† See Ref. [62] for a text on Monte Carlo techniques.

Subsequent developments in this chapter utilize the ideas of Bayesian statistical analysis. The objective is to provide a logical framework for choosing the best alternative system and processing experimental data on the random traveler process. Some introductory remarks on Bayesian methodology are given below.

4.2 Bayesian Analysis and Decision-Making

Bayesian decision theory provides a framework for making decisions that depend on a real world process that can be modeled as a random process with parameters whose values are uncertain. To describe this framework, several aspects of the formulation must be delineated: The random variables $(\underline{x})$ must be identified. The random process $g_{\underline{x}}(\underline{x}_o)$ that generates the random variables must be modeled in terms of parameters $(\underline{u})$ whose values are uncertain. Finally, the uncertain state of knowledge of the random process parameters must be encoded and the optimal decision and sampling procedures must be determined.

The Bayesian approach requires that prior knowledge on the uncertain parameters $(\underline{u})$ of the random process be encoded in terms of a prior probability distribution $f_{\underline{u}}(\underline{u}_o)$. Data (D), subsequently obtained by sampling observable outputs of the random process $g_{\underline{x}}(\underline{x}_o)$, is used to up-date the state of information on the uncertain parameters, by using Bayes' rule to derive a posterior distribution $f_{\underline{u}|D}(D|\underline{u}_o)$. Bayes' rule is given as follows:

$$f_{\underline{u}|D}(\underline{u}_o|D) = \frac{f_{D|\underline{u}}(D|\underline{u}_o)\ f_{\underline{u}}(\underline{u}_o)}{\int_{\underline{u}} f_{D|\underline{u}}(D|\underline{u}_o)\ f_{\underline{u}}(\underline{u}_o)} . \tag{4-1}$$

The denominator is just a normalization constant, obtained by integrating the numerator over $\underline{u}$. Thus, the posterior distribution $f_{\underline{u}|D}$ is proportional to the prior distribution $f_{\underline{u}}$, multiplied by $f_{D|\underline{u}}$, which is called the likelihood function. The likelihood function is just the conditional probability of the data, given the parameter values $\underline{u}$ of the random process. Since the data is a function of the random variables $\underline{x}$, the likelihood function is derived from the generating function $g_{\underline{x}}$.

It is convenient, though not always possible, to specify the prior $f_{\underline{u}}$ as one of a family of curves, such that the posterior distribution $f_{\underline{u}|D}$, obtained via Eq. (4-1), is a member of the same family. In this event, the family of curves is characterized by a set of parameters, and the process of up-dating (computing the posterior) is reduced to modification of these parameter values. For example, if the random process is normal (Gaussian), and the prior is normal, then the posterior will also be normal, with revised means and variances. The likelihood function and prior are called conjugate functions in such cases. Another example of conjugate functions is used for illustration, later.

The gain or loss resulting from the decision is a function of the decision itself and the values of the parameters of the random process. The decision is made to maximize the expected gain* (or minimize the expected loss) based on the current estimate of the parameters of the random process. The expected value of obtaining additional data on the process may be found by taking the difference between the prior expected gain and the expected posterior gain.

---

* In general, the decision should be made to maximize the expected value of the "utility" of the gain. Maximizing expected gain corresponds to assuming a linear utility function.

4.3 Analysis for Adaptive Timetables

This section develops the framework for choosing among alternative systems under premise P1 which requires that system timetables are computed after the characteristics of the actual traveler population become known with certainty. Here, the phenomena of traveler generation and system scheduling are viewed as a conglomerate random process. That is, a traveler population is viewed as the output of the traveler generation process; subsequently, the transportation system schedule is optimized (deterministically) to meet the requirements of the traveler population. The value of the objective function resulting from the system scheduling is then viewed as the output of the combined traveler generation-scheduling transformation random process. In this perspective, the problem of choosing the best alternative system, to serve traveler populations generated from a particular probabilistic distribution is equivalent to the problem of selecting the "best" of several random processes.

The problem of selecting the best of several processes is well known in statistical literature.* The analysis here is guided by that of Raiffa and Schlaifer [64]. Although the approach is sound, it becomes analytically intractible for very complicated mathematical forms, as in the transportation context here. Hence, implementation of the procedure would likely require the use of numerical methods in place of a closed form analysis.

The original random variables for the current problem are the potential demand levels ($\underline{N}$) and the passenger preference parameters ($\underline{P}$).

---

* See Ref. [63] for a historical summary.

These values undergo a complex transformation, via the MODEL2 computational process, to emerge as a set of new random variables $\underline{b} = (b^1, b^2, \ldots, b^A)$, distributed by $g_{\underline{b}}(b_o^1, b_o^2, \ldots, b_o^A)$, where b^m is the benefit (objective function level) for the m^{th} alternative system. Since it is not analytically feasible to derive $g_{\underline{b}}$ from $f_{\underline{N}}$ and $[f^r_{\underline{P}|\underline{N}}]$, $g_{\underline{b}}$ will be considered the primary random process in this analysis. Thus, random traveler generation-MODEL2 optimization is the overall random process for which the parameter values are uncertain. Sample data (D) is obtained by sampling a population and using MODEL2 to derive sample benefit values of each alternative system. The likelihood function, therefore, is $g_{\underline{b}}(\underline{b}_o)$ for one sample $\underline{b}_o$, and $f_{[\underline{b}_o]|\underline{u}}([\underline{b}_o]|\underline{u}) = K \prod_{j=1}^{n} g_{\underline{b}}(\underline{b}_o^j)$ for n independent samples $[\underline{b}_o^j]$.

Because analytical derivation is not possible, characterization of the generating process is difficult. However, a reasonable argument for picking a multi-normal (multivariate Gaussian) generating process* goes as follows:

The benefit (b^m) of a given system m, is the sum of individual traveler benefits. (Operating costs, if part of the objective, may be reduced to a per passenger basis.) A system serves a relatively large number of travelers. Thus, by a central limit theorem argument,** no matter how the individual traveler benefits are distributed, the system benefit will be normally distributed. It is reasonable to expect, therefore, that the joint density function $g_{\underline{b}}$ is multi-normal.

---

* See Ref. [64] for a description of the multi-normal process.

** See Ref. [65] for a description of the central limit theorem.

There are (at least) two weaknesses in this argument. First, the central limit theorem applies to the sum of independent random variables.* Although the travelers' parameters are sampled independently, their benefits may be correlated, since all traveler benefits are generated through the same transportation system (m). For example, if it is known that travelers B and C have very low benefit levels, possibly on account of high fares or low speed vehicles in the system, then it may be more probable that traveler A has a low benefit level as well. The second problem with the normality argument is that the existence of normally distributed marginal distributions $g_{b^m}(b^m)$ for all m, does not necessarily mean that the joint distribution is multi-normal. Despite these problems, the multi-normal distribution should be a serviceable model of the system benefit generation process. More careful characterization of this process will be possible only after more extensive computational experience has been gained in the generation of schedule benefits for random traveler inputs.

The decision-making framework is based on the criterion to choose the alternative system with the highest expected value. The value of an alternative system m will be given by

$$v_m = -K_m + k_m \overline{b^m} = -K_m + k_m \beta_m \ , \qquad (4\text{-}2)$$

where

$\beta_m = \overline{b^m}$ = mean operating benefit of system m ,

K_m = fixed cost of system m ,

---

* See Ref. [66] for a description of the conditions under which the independence requirement may be dropped.

k_m = proportionality constant to put fixed cost and operating benefit on the same time basis (e.g. per annum).

From (4-2), the mean operating benefits β_m are the uncertain parameters of the generating process that are of direct interest. The uncertainty is encoded in terms of a prior probability density function $f'_{\underline{\beta}}(\beta_1,\beta_2,\ldots,\beta_m)$. By integrating with respect to the prior $f'_{\underline{\beta}}$, the expected values $\bar{v}'_m$ may be found:

$$\bar{v}'_m = -K_m + k_m\bar{\beta}'_m \ , \tag{4-3}$$

where

$\bar{\beta}'_m$ = expected value of β'_m, under distribution $f'_{\underline{\beta}}$.

The optimal decision is to choose the system with the highest expected value. Let r be the index number of the system with the highest expected value, under <u>prior</u> distribution $f'_{\underline{\beta}}$. Then,

$$r = \arg \max_m(\bar{v}'_m) \ , \qquad 1 \leq m \leq A \ , \tag{4-4}$$

so that

$$\bar{v}'_r \geq \bar{v}'_m \ , \qquad 1 \leq m \leq A \ . \tag{4-5}$$

Thus, if the decision were to be made with prior knowledge only, then system r is the best choice.

Now consider the process of obtaining sample values from the random process, to improve the state of information on the process parameters. Sampling and up-dating prior knowledge can be expected to lead to a better decision and a greater expected value. The increase in expected value (i.e., the expected value of sampling information $\overline{VSI}$) less the cost of sampling is the criterion for determining how much sampling information should be purchased. Hence, the optimal number of samples to be purchased (n^*) is found by maximizing the <u>net</u> expected value of sampling information (NEVSI) with respect to the number of samples (n). Thus, n^* is the solution to the following equation:

$$NEVSI(\underline{Q}',n^*) = \max_n NEVSI(\underline{Q}',n), \quad n \geq 0 \text{ and } n \text{ integer }, \tag{4-6}$$

where

$$NEVSI(\underline{Q}',n) = \overline{VSI(\underline{Q}',D,n)} - nc \tag{4-7}$$

and

$\underline{Q}'$ = a set of parameters that characterize the prior distribution $f'_{\underline{\beta}}$,

D = the sample data (a set of n sample system benefits, $\underline{b}$) ,

c = the dollar cost per sample.

Now, $\overline{VSI(\underline{Q}',D,n)}$ is the expectation with respect to the data D, of the value of sampling information, $VSI(\underline{Q}',D,n)$. VSI is a random variable prior to sampling since it is a function of the data D, a random vector. To develop the mathematical expression for VSI, observe that the sampling information will lead to the posterior distribution $f''_{\underline{\beta}}(\underline{\beta}_o)$, yielding the posterior expected values,

$$\bar{v}''_m = -K_m + k_m\bar{\beta}''_m \ . \tag{4-8}$$

The best posterior decision will be to choose the system q such that

$$\bar{v}''_q = \max_m(\bar{v}''_m) \ . \tag{4-9}$$

Recall that the optimal choice among alternative systems, on the basis of prior information, is system r. Subsequent to obtaining sampling information (data), however, the optimal choice is system q. Thus, the value of sampling information (VSI) is just the posterior difference in expected values of systems q and r. Specifically,

$$\begin{aligned} VSI = \bar{v}_q'' - \bar{v}_r'' &= \max(\bar{v}_1'', \bar{v}_2'', \ldots, \bar{v}_A'') - \bar{v}_r'' , \\ &= \max(\bar{v}_1''-\bar{v}_r'', \bar{v}_2''-\bar{v}_r'', \ldots, \bar{v}_A''-\bar{v}_r'') , \end{aligned} \tag{4-10}$$

or,

$$VSI = \max_m \left\{ -(K_m - K_r) + k_m\bar{\beta}_m'' - k_r\bar{\beta}_r'' \right\} , \qquad 0 \leq m \leq A . \tag{4-11}$$

Prior to sampling, the $\bar{v}_m''$ or the $\bar{\beta}_m''$ (and VSI) are random variables. In particular, the $\bar{\beta}_m''$ are functions of the (as yet unknown) parameters $\underline{Q}''$ of the posterior distribution $f_{\underline{\beta}}''$. The posterior parameters are functions $\underline{Q}''(\underline{Q}',D,n)$ of the parameters $\underline{Q}'$ of the prior $f_{\underline{\beta}}'$, the number of samples n, and the sample data D.[†]

The expected value of VSI, required in (4-7) for computation of the optimal number of samples, is found by integrating VSI with respect to the density function describing the prior probability of the data:

$$\overline{VSI(\underline{Q}',D,n)} = \int_D VSI(\underline{Q}',D,n)\; f_D(D|n,\underline{Q}') . \tag{4-12}$$

$f_D(D|n,\underline{Q}')$ is called the "preposterior" distribution, and may be computed from the relationship,

$$f_D(D|n,\underline{Q}') = \int_{\underline{\beta}} f_{D|\underline{\beta}}(D|n,\underline{Q}',\underline{\beta}) f_{\underline{\beta}}(\underline{\beta}|\underline{Q}',n) , \tag{4-13}$$

[†] It is often possible to reduce the sample data D to a set of "sufficient statistics," $\underline{\sigma}$, that retain all the necessary sample information required in the analysis. In such case, $\underline{Q}''(\underline{Q}',D,n) = \underline{Q}''(\underline{Q}',\underline{\sigma},n)$.

where $f_{D|\underline{\beta}}$ is computed from the generating function $g_{\underline{b}}(\underline{b}_\sigma)$ since D consists of random sample benefit vectors $[\underline{b}]$. Function $f_{\underline{\beta}}(\underline{\beta}|\underline{Q}',n)$ is just the prior density function $f'_{\underline{\beta}}$.

Substituting the expected value of sampling information (4-12) into the expression for the net expected value of sampling information (4-7), and computing the optimal number of samples by solving Eq. (4-6), completes the "preposterior" analysis of data acquisition.

A supplementary consideration to the preceding analysis is the option of <u>differential</u> sampling. In the above discussion, samples are obtained by generating benefits of all proposed systems from common population samples. More exotic sampling schemes are easily conceived. In general, consideration may be given to various mixtures of sample benefits, generated from the same or separate (independent) traveler population samples and processed through one or more of the alternative systems. Such sampling methods may be analyzed in a manner analogous to the development above. Only a few general remarks will be made here:

1. More information is obtained by generating alternative system benefits from common traveler population samples. That is, more consistent comparison of system performances is possible when inputs are uniform.

2. Sampling system benefits from a common traveler population is more difficult to analyze analytically than sampling benefits from independent traveler population samples, because alternate system benefits generated from a common population are not independent. Hence, multivariable generating functions $g_{\underline{b}}$ are required, in lieu of single variable functions g_{bm} .

3. It may be more economical to consider procedures in which "clearly inferior" systems are dropped from consideration on the basis of only a few samples. In general, different numbers of samples from alternate systems may be preferable.

4.4 Example

To illustrate the ideas of the foregoing discussion, consider the following decision to choose among two alternative systems. Let the random traveler generation-scheduling benefit process be characterized by the rectangular function,

$$g_{\underline{b}}(b^1,b^2) = \begin{cases} (\gamma\theta)^{-1} & 0 \le b^1 \le \gamma\ , \quad 0 \le b^2 \le \theta\ , \\ 0 & \text{elsewhere}\ . \end{cases} \tag{4-14}$$

Recall that b^m is the (random) benefit of system m .

Thus, γ and θ are the (uncertain) parameters of the random process. The uncertainty about these parameters is encoded by the conjugate prior distribution,

$$f'_{\gamma\theta}(\gamma,\theta) = \begin{cases} K'\gamma^{-n'}\theta^{-m'} & \gamma \ge M_1',\quad \theta \ge M_2'\ , \\ 0 & \text{elsewhere}\ , \end{cases} \tag{4-15}$$

where $n',m' > 0$, $M_1',M_2' \ge 0$, and

$$K' = \frac{(n'-1)(m'-1)}{M_1'^{-n'+1}M_2'^{-m'+1}}\ . \tag{4-16}$$

The likelihood function for n benefit samples of the form (b^1,b^2) is given by

$$\left\{\underline{b}^1,\underline{b}^2/\gamma,\theta\right\} = \begin{cases} \gamma^{-n}\theta^{-n} & \text{for}\ \ 0 \le M_1 \le \gamma\ ,\quad 0 \le M_2 \le \theta\ , \\ 0 & \text{elsewhere}\ , \end{cases} \tag{4-17}$$

where

$$\underline{b}^1 = (b_1^1, b_2^1, \ldots, b_n^1) , \qquad \underline{b}^2 = (b_1^2, b_2^2, \ldots, b_n^2) ,$$

$$M_1 = \max(b_1^1, b_2^1, \ldots, b_n^1) , \quad M_2 = \max(b_1^2, b_2^2, \ldots, b_n^2) ,$$

and

b_k^m = k^{th} sample benefit value from the m^{th} alternative system.

(M_1, M_2) are sufficient statistics for the n pieces of sample data in this example. Note that the prior and the likelihood function are conjugate so that the posterior density function has the same form as the prior. Thus, the posterior density function, with updated parameters, is written,

$$f''_{\gamma\theta}(\gamma,\theta) = \begin{cases} K''\gamma^{-n''}\theta^{-m''} & \gamma \geq M_1'', \quad \theta \geq M_2'' , \\ 0 & \text{elsewhere} , \end{cases} \tag{4-18}$$

where

$$n'' = n' + n, \qquad m'' = m' + n , \tag{4-19}$$

$$M_1'' = \max(M_1, M_1') , \qquad M_2'' = \max(M_2, M_2') , \tag{4-20}$$

and

$$K'' = \frac{(n''-1)(m''-1)}{M_1''^{-n''+1} M_2''^{-m''+1}} . \tag{4-21}$$

The optimal decision, on the basis of prior information, is to pick the system with the highest expected mean value $(\bar{v}_m')$ in conformity with Eq. (4-4). Since $\bar{v}_m' = -K_m + k_m\bar{\beta}_m'$, the expected mean operating benefits $\bar{\beta}_m'$ are the crucial statistics for making the (prior) decision. The mean operating benefits β_1' and β_2' are given by,

$$\beta_1' = \bar{b}_1 = \int_0^\theta \int_0^\gamma (\gamma\theta)^{-1} b_1 db_1 db_2 = \gamma/2 , \tag{4-22a}$$

and

$$\beta_2' = \bar{b}_2 = \int_0^\theta \int_0^\gamma (\gamma\theta)^{-1} b_2 db_1 db_2 = \theta/2 \ . \tag{4-22b}$$

The expected values of the mean system operating benefits are given by,

$$\bar{\beta}_1' = \bar{\beta}_1' (n',M_1') = \int_\theta \int_\gamma 1/2\gamma f'_{\gamma\theta}(\gamma,\theta) = 1/2 \frac{(n'-1)M_1'}{(n'-2)} \ , \tag{4-23a}$$

and,

$$\bar{\beta}_2' = \bar{\beta}_2'(m',M_2') = \int_\gamma \int_\theta 1/2\theta f'_{\gamma\theta}(\gamma,\theta) = 1/2 \frac{(m'-1)M_2'}{(m'-2)} \ . \tag{4-23b}$$

If $-K_2+k_2\bar{\beta}_2' \geq -K_1+k_1\bar{\beta}_1'$ (i.e., $\bar{v}_2' \geq \bar{v}_1'$), the optimal (prior) choice is system 2. If not, system 1 is best.

In analogy to (4-23), the posterior expected values are,

$$\bar{\beta}_1'' = \frac{1/2(n''-1)M_1''}{(n''-2)} \ , \tag{4-24a}$$

and,

$$\bar{\beta}_2'' = \frac{1/2(m''-1)M_2''}{(m''-2)} \ . \tag{4-24b}$$

To compute the value of sampling information, assume that $\bar{v}_2' \geq \bar{v}_1'$, i.e., that the best prior choice is system 2. Then the VSI, found by substituting (4-24) into (4-10), is given by,

$$\text{VSI} = \max\left\{-(K_1-K_2) + k_1\bar{\beta}_1'' - k_2\bar{\beta}_2'', 0\right\} \ , \tag{4-25}$$

$$= \max\left\{(K_2-K_1) + 1/2\left[\frac{k_1(n''-1)M_1''}{(n''-2)} - \frac{k_2(m''-1)M_2''}{(m''-2)}\right], 0\right\} \ . \tag{4-26}$$

The expected value of sampling information, found by integrating VSI with respect to the probability density of the data (M_1,M_2), is given by,

$$\overline{VSI(n,M_1',M_2',n',m')} = \int_{M_1,M_2} VSI(n,M_1',M_2',n',m',M_1,M_2)$$

$$f_{M_1M_2}(M_1,M_2|n,M_1',M_2',n',m') \ . \qquad (4\text{-}27)$$

In (4-27), $f_{M_1M_2}$ is the preposterior distribution, computed from the relationship,

$$f_{M_1M_2}(M_1,M_2|n,M_1',M_2',n',m') = \int_{\theta,\gamma} f(M_1,M_2|\theta,\gamma,n,M_1',M_2',n',m')$$

$$f_{\gamma\theta}(\gamma,\theta|M_1',M_2',n',m',n) \ , \qquad (4\text{-}28)$$

where

$$f_{\gamma_\theta}(\gamma,\theta|M_1',M_2',n',m',n) = f'_{\gamma\theta}(\gamma,\theta)$$

is given by Eq. (4-16), and

$$f(M_1,M_2|\theta,\gamma,n,M_1',M_2',n',m') = f(M_1,M_2|\theta,\gamma,n)$$

is found as follows:

$$f(M_1,M_2|\theta,\gamma,n) = \left\{\frac{n(n-1)}{n(n-1)+n}\, g_{\underline{b}}(b^1=M_1,b^2 < M_2)g_{\underline{b}}(b^1 < M_1,b^2=M_2)\right.$$

$$\left.\left[g_{\underline{b}}(b^1 < M_1,b^2 < M_2)\right]^{n-2}\right\} + \left\{\frac{n}{n(n-1)+n} g_{\underline{b}}(b^1=M_1,b^2=M_2)\right.$$

$$\left.\left[g_{\underline{b}}(b^1 < M_1,b^2 < M_2)\right]^{n-1}\right\} \ . \qquad (4\text{-}29)$$

The two terms in (4-29) correspond to the two mutually exclusive ways of obtaining M_1,M_2. The first possibility is to achieve the values $b^1=M_1$, $b^2=M_2$ in separate samples. There are $n(n-1)$ ways in which that event can occur. The second possibility, that M_1 and M_2 occur on the same sample, can happen n different ways. For the current example,

$g_{\underline{b}}(b^1,b^2) = g_{b^1}(b^1)g_{b^2}(b^2)$. Thus, expression (4-29) reduces to,

$$f(M_1,M_2|\theta,\gamma,n) = g_{b^1}(M_1)g_{b^2}(M_2)\left[g^{\leq}_{b^1}(M_1)\, g^{\leq}_{b^2}(M_2)\right]^{n-1} , \qquad (4\text{-}30)$$

where

$$g^{\leq}_{b^1}(M_1) = \begin{cases} \int_0^{M_1}(\gamma)^{-1}db^1 = M_1/\gamma & M_1 \leq \gamma , \\ 1 & M_1 > \gamma . \end{cases} \qquad (4\text{-}31a)$$

Similarly,

$$g^{\leq}_{b^2} = \begin{cases} M_2/\theta & M_2 \leq \theta , \\ 1 & M_2 > \theta . \end{cases} \qquad (4\text{-}31b)$$

Also,

$$g_{b^1}(M_1) = \begin{cases} \gamma^{-1} & M_1 \leq \gamma , \\ 0 & M_1 > \gamma , \end{cases} \qquad (4\text{-}32a)$$

$$g_{b^2}(M_2) = \begin{cases} \theta^{-1} & M_2 \leq \theta , \\ 0 & M_2 > \theta . \end{cases} \qquad (4\text{-}32b)$$

Thus,

$$f(M_1,M_2|\theta,\gamma,n) = \begin{cases} (\gamma\theta)^{-1}(M_1M_2/\gamma\theta)^{n-1} & M_1 \leq \gamma,\ M_2 \leq \theta , \\ 0 & \text{otherwise} . \end{cases} \qquad (4\text{-}33)$$

Substituting (4-33) and (4-15) into (4-28) and integrating, yields the preposterior distribution, $f_{M_1M_2}(M_1,M_2|n,M_1',M_2',n',m')$. Substituting the preposterior and the VSI (4-26) into (4-27), and carrying through the integration yields the expected value of sampling information, $\overline{\text{VSI}(n)}$. To determine the optimal number of samples, $\overline{\text{VSI}(n)}$ is combined with the cost of sampling, as in (4-7), to yield the net expected value of sampling information, NEVSI(n). Maximizing NEVSI(n) with respect to n yields the optimal number of samples (n^*) to be purchased.

It is evident that the decision analysis presented in the preceding pages may become quite cumbersome in practice. Numerical methods may be required to carry out the procedure. Below, a simplified approach is presented that is more easily implemented, but which requires a new criterion for computing the optimal decision and the value of sampling information.

4.5 Multinomial Process*

In the previous analysis, the random process was viewed as a compound traveler generation-schedule transformation mechanism. In this section, an additional transformation step is applied to the random variables. The following random variable is defined as a function of the random vector $\underline{b} = (b^1, b^2, \ldots, b^A)$:

$$s = \arg \max_m (-K_m + k_m b^m) \; , \tag{4-34a}$$

so that

$$-K_s + k_s b^s \geq -K_m + k_m b^m, \quad \text{for all } m, \; 1 \leq m \leq A \; . \tag{4-34b}$$

The overall process is illustrated in Fig. 34. The traveler population is randomly selected, and input into the alternative systems, 1 through A. The scheduling algorithm produces the random vector $\underline{b} = (b^1, b^2, \ldots, b^A)$. The random variable s is the index of the component b^s of the vector $\underline{b}$, such that (4-34b) holds.

---

* For a more complete description of this process, see Ref. [67].

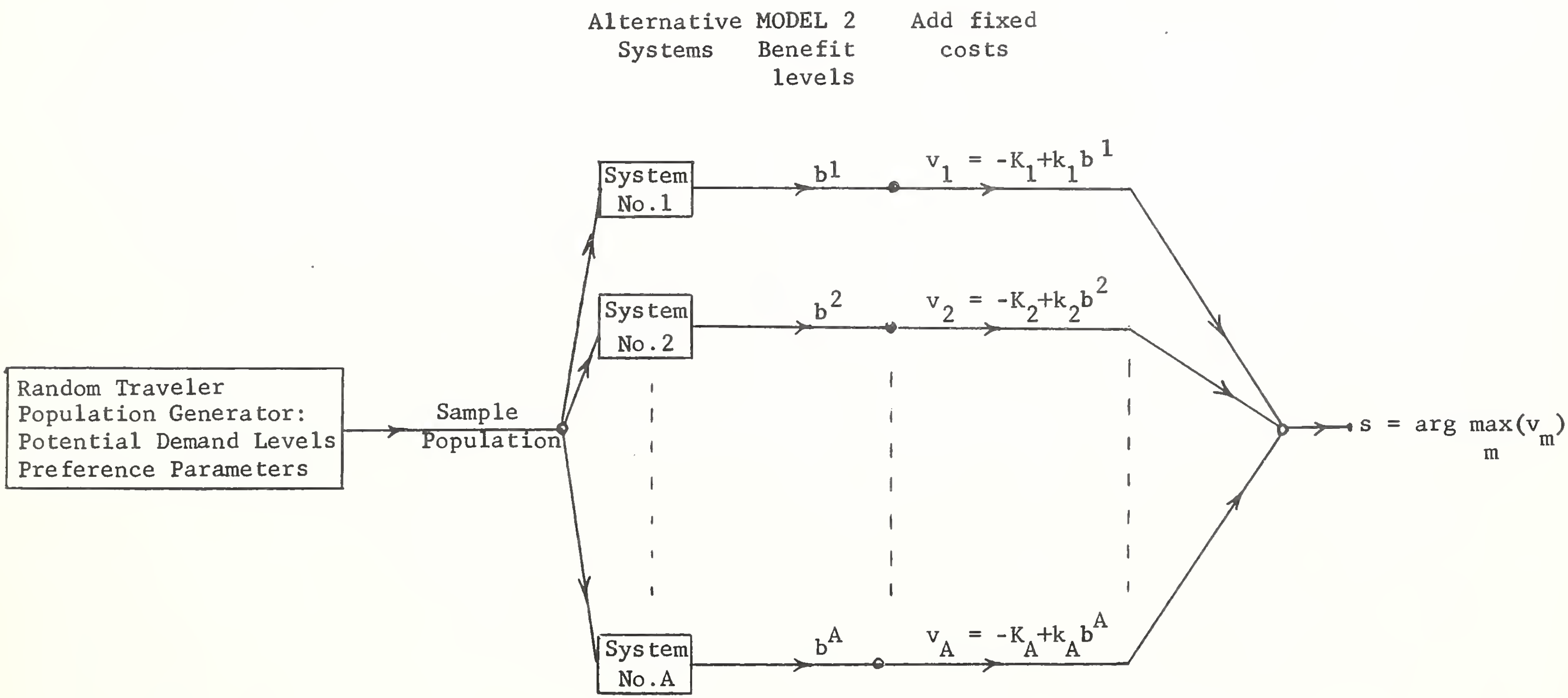


Figure 34. Compound Random Process.

This random process may be characterized by the "multinomial" parameters $(p_1, p_2, \ldots, p_A)$ where,

p_m = probability that the m^{th} system produces the highest benefit value $v_m = -K_m + k_m b^m$,

and

$$\sum_{m=1}^{A} p_m = 1 \ . \tag{4-35}$$

The uncertainty on the values of the p_m may be encoded in terms of a prior from the multivariate "beta" family of distributions:

$$f'_{\underline{p}}(p_1, p_2, \ldots, p_A) = K' p_1^{n_1'-1} \, p_2^{n_2'-1} \ldots p_A^{n_A'-1} \ , \tag{4-36a}$$

where

$$K' = \frac{\Gamma(n_1'+n_2'+\ldots+n_A')}{\Gamma(n_1')\,\Gamma(n_2')\ldots\Gamma(n_A')} \ , \tag{4-36b}$$

and

$$\Gamma(x) = (x-1)! \qquad \text{for integer } x \text{ and } x \geq 1 \ . \tag{4-36c}$$

Suppose that n samples are generated, using n independent traveler populations. The resulting data may be represented as,

$$D = (n_1, n_2, \ldots, n_A) \ ,$$

where

$$\sum_{m=1}^{A} n_m = n \ , \tag{4-37}$$

and,

n_m = the number of times system m's value exceeds the value of all other systems, in the n trials, i.e., the number of "successes" for system m.

The likelihood function is,

$$f_{D|\underline{p}}(D|\underline{p}) = p_1^{n_1} p_2^{n_2} \cdots p_A^{n_A} = \prod_{m=1}^{A} p_m^{n_m} \; . \tag{4-38}$$

The posterior distribution is just,

$$f_{\underline{p}}''(p_1, p_2, \ldots, p_A) = K'' p_1^{n_1''-1} \; p_2^{n_2''-1} \; \cdots \; p_A^{n_A''-1} \; , \tag{4-39}$$

where,

$$n_m'' = n_m' + n_m \; , \tag{4-40}$$

and,

$$K'' = \frac{\Gamma(n_1''+n_2''+\ldots n_A'')}{\Gamma(n_1'')\,\Gamma(n_2'')\ldots\Gamma(n_A'')} \; . \tag{4-41}$$

The decision-making structure will be based on the strategy to minimize the expected cost of choosing the "wrong" system. To formulate the decision criterion, consider the following definition:

w_m = the cost of not choosing system m when in fact system m is the best system.

Each alternative system m is identified with a parameter value w_m. The w_m parameters allow the decision maker to incorporate indirect costs* of making an incorrect decision. For example, suppose a city planning agency attaches a greater cost to rejecting a proposal by a local firm than rejecting a non-local firm's proposal. Then the city agency may compute the costs (w_m's) by weighing the economic and political ramifications of employing local versus non-local talent, against the

---

* Costs (or benefits foregone) not directly attributable to system users or operators.

penalties of failing to choose the best system.

The decision criterion is to choose system $m*$ to minimize the mean value of the expected cost of a wrong decision. Thus,

$$m* = \arg\min_m \overline{(\sum_{j \neq m} w_j p_j)} , \tag{4-42a}$$

$$= \arg\min_m \overline{(\sum_{j=1}^{A} w_j p_j - w_m p_m)} , \tag{4-42b}$$

$$= \arg\min_m (\text{Constant} - w_m \bar{p}_m) . \tag{4-42c}$$

Thus,

$$m* = \arg\max_m (w_m \bar{p}_m) . \tag{4-43}$$

The prior estimate of $\bar{p}_m$ is,

$$\bar{p}'_m = n'_m / \left(\sum_{j=1}^{A} n'_j\right) . \tag{4-44}$$

If all w_m's are equal, then the decision rule is to pick the system corresponding to the highest (prior) mean success probability, $\bar{p}'_m$. Otherwise, the system $m*$ corresponds to $\max_m (w_m \bar{p}_m)$. After sampling has been performed, the posterior estimates are given by,

$$\bar{p}''_m = n''_m / \left(\sum_{j=1}^{A} n''_j\right) . \tag{4-45}$$

The optimal posterior decision is to chose system $m**$ corresponding to $\max_m (w_m \bar{p}''_m)$.

The multinomial formulation yields to the same kind of preposterior analysis that was employed earlier to determine the optimal number of samples to be purchased. The only difference here is that the value of sampling must be gauged in terms of the "cost of a wrong decision" criterion of Eq. (4-43), rather than in terms of the difference in operational benefits of alternative systems, as in the earlier analysis.

In summary, the formulation of this section considerably simplifies the task of choosing among alternative systems. The present formulation provides a clear characterization of the random traveler generation scheduling phenomenon as a multinomial process. Uncertainty about the process parameters (p_m's) is easily encoded in terms of a rich family of multivariate beta distributions. The criterion to minimize the cost of a wrong decision allows discrimination among systems on the basis of highest mean probability of success. However, the formulation does not yield an estimate of the mean operational benefit of the chosen system, nor does it allow determination of sampling policy on the basis of operational benefits.

4.6 Analysis for Permanent Timetables

In this section, the decision-making methodology is developed for the permanent timetable case, corresponding to the premise P2 that a single fixed timetable is computed (for each alternative system) on the basis of uncertain knowledge of traveler characteristics. As explained in the introduction to this chapter, the following two assumptions are made:

1. The possible levels of potential traveler demand (N) are sufficiently large so that the statistical properties of any random traveler population of size N closely approximate the properties of the probability distribution from which the population is generated.

2. The (optimal stochastic) timetable that maximizes the expected value of the objective function is also optimal for some deterministic level of potential demand.

Assumption (1), in conjunction with the Monte Carlo technique for generating random traveler populations, allows a simplification in the consideration of uncertainties. In particular, according to this assumption, an optimal timetable computed on the basis of a particular random traveler population of size N will be (approximately) optimal for any other such sample population of size N. Hence, uncertainty in the traveler preference parameters $(\underline{P})$ need not be explicitly included in the determination of optimal timetables and maximum benefits. Only the uncertainty in demand level need be considered. Actually, as indicated earlier, there is a vector of potential demands $\underline{N} = (N_1, N_2, \ldots, N_R)$, where R is the number of different types of trip. Thus, an additional assumption, that each component demand N_r is related deterministically to a proxy parameter N, will be made to ensure feasibility in the following analysis. N may be regarded as the overall system demand level, and each component N_r as some known percentage of N.

Suppose that N ranges over the interval $[N_{min}, N_{max}]$, and that only a discrete set of values N_{min}, $N_{min} + \Delta N$, $N_{min} + 2\Delta N$, ..., $N_{max} - \Delta N$, N_{max} are assumed possible. (ΔN should be small enough to ensure sufficient accuracy, and large enough to maintain reasonable computational requirements.) Then by generating a random traveler population to correspond to each possible demand level, and by using MODEL2 to generate an optimal timetable for each of these populations, the "optimal set" of timetables (for a given alternative transportation system) may be generated. According to assumption (2), the optimal stochastic timetable is one of the timetables in this optimal set.

To compute the benefit associated with each of the timetables in this set, it must be recognized that the (optimal stochastic) timetable that is eventually selected must serve whatever demand level actually materializes. Thus, the performance of each timetable must be evaluated not only with respect to the demand level for which it is optimal, but also with respect to all other possible demand levels. In the case of a shuttle system, the procedure for performing this evaluation is clear: Use the optimum assignment algorithm to match the sample traveler population at each demand level to each of the timetables in the optimal set. For network systems more complex than the shuttle, there is no easy method for optimally assigning travelers to trips in a timetable. In this case, a simulation may be devised to allocate travelers to acceptable trips. For example, suppose travelers are randomly ordered, and each successive traveler is assigned to his favorite trip in the timetable, unless that trip is full. In the latter case, he may replace some traveler who is already assigned, if a net system gain results. Iterations over travelers would continue until no further improvements are possible. (This procedure parallels the MODEL2 rationale.)

Once the benefit of each demand level is computed for a given timetable, the expected benefit of the timetable may be found by summing up these (component) benefits, each benefit weighted by the probability of its demand level occurring. Thus, for some system m,

$$\bar{b}_{T(m,N_o)} = \sum_{N=N_{min}}^{N_{max}} b_{T(m,N_o)}(N)p(N) \ , \qquad (4\text{-}46)$$

where,

$p(N)$ = probability that demand level N occurs,

$b_{T(m,N_o)}(N)$ = benefit produced by the system with the traveler population of demand level N, from the timetable $T(m,N_o)$ that is optimal for demand level N_o and system m .

$\bar{b}_{T(m,N_o)}$ = expected benefit of the timetable $T(m,N_o)$ that is optimal for demand level N_o and system m.

Once computation (4-46) is performed, the optimal stochastic timetable for system m, $T(m,\hat{N}_m)$, and the corresponding expected benefit level $\bar{b}_{T(m,\hat{N}_m)}$ are chosen such that,

$$\bar{b}_{T(m,\hat{N}_m)} \geq \bar{b}_{T(m,N)} , \quad \text{for all } N \in [N_{min}, N_{max}] . \tag{4-47}$$

($\hat{N}_m$ is the demand level for which timetable $T(m,\hat{N}_m)$ is optimal.)

Once $T(m,\hat{N}_m)$ is found, and $\bar{b}_{T(m,\hat{N}_m)}$ is computed for each proposed system m, the best alternative system m^* may be chosen such that

$$\bar{b}_{T(m^*,\hat{N}_{m^*})} \geq \bar{b}_{T(m,\hat{N}_m)} , \quad \text{for all } m. \tag{4-48}$$

The foregoing procedure for choosing among alternative systems, in the permanent timetable case, is summarized in block diagram form in Fig. 35.

4.7 Preference and Demand Data

Much of the analysis in this chapter is based on Bayesian estimation methodology and decision criteria. The Bayesian techniques are equally applicable to the problem of relating empirical data to the estimation of parameter values of the traveler generation process. Recall that Chapter II presented two modes for collecting information pertaining to the preference behavior of travelers--the "interview" approach and

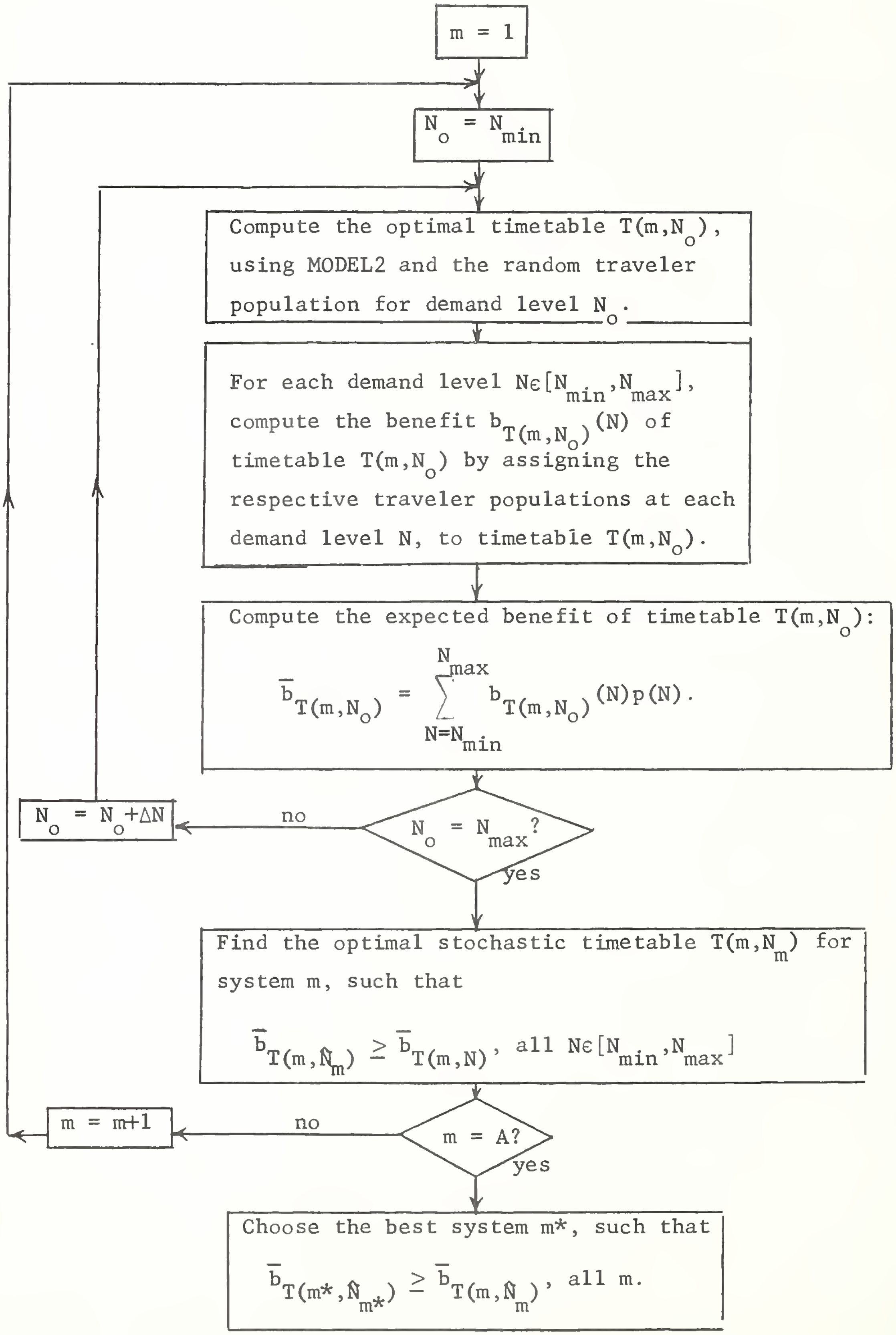


Figure 35. Method for Choosing the Best System in the Permanent Timetable Case.

the "observation" approach. In both cases, the Bayesian scenario may be followed.

First, the (joint) generating function $g_{\underline{N},\underline{P}}$ must be formulated in terms of parameters $\underline{u}$, the values of which are uncertain. ($\underline{N}$ is the vector of potential demand levels; $\underline{P}$ is the vector of preferences parameters--(t^*,w,D,α,n,s).) As an example, $g_{\underline{N},\underline{P}}$ might be such that the "marginal" distributions $g_{\underline{P}}(D,\alpha,n,s)$ and $g_{\underline{N}}(N_1,N_2,\ldots,N_R)$ are multinormal, and $g_{t^*,w}(t^*,w)$ is a curve with a.m. and p.m. "rush hour" peaks. (If $\underline{N},\underline{P}'$, and (t^*,w) are assumed independent, then $g_{\underline{N},\underline{P}}$ is the product of $g_{\underline{N}}$, $g_{\underline{P}'}$ and $g_{t^*,w}$.) The uncertainty over the parameter values of the generating process must be encoded in terms of a prior density function. Next, the likelihood function, which is the probability of the data conditional upon the parameter values of the generating function, must be formulated according to the kind of data actually obtained. It is the latter step that distinguishes the differences in handling the two modes of information collection for the traveler demand and preference estimation process. Finally, the posterior distribution is computed from the prior and sample data, using the Bayesian inference equation (4-1).

In order to clarify the difference in handling the two types of traveler data, and to get a better idea of what the observable behavior data might actually consist of, the steps in the construction of the respective likelihood functions will now be considered. The ensuing discussion presents a superficial treatment, designed to illustrate the basic framework. Substantial refinement would be required to apply the analysis in specific cases.

In the case of interview data, sample values of each subject's willingness to pay and indifference curve parameters $(t^*, w, D, \alpha, n, s)$ may be obtained by "curve fitting" the response data to the formulas (2-42) or (2-43) that specify the traveler's willingness to pay for alternative trips. Data on potential demand over a particular route can be collected by sampling the percentage of the general public that have reason to travel each route on a given day. Thus, the sample data will consist of a set of "response vectors" of the form $\underline{P} = (t^*, w, D, \alpha, n, s)$, plus sample values of the potential demands $\underline{N}$. Suppose that a sample N, and q independent sample preference vectors $[\underline{P}]$ are obtained for a given trip type. In addition, assume N and $[\underline{P}]$ are independent. Then, the likelihood of the data (D) is,

$$f_{D|\underline{u}}(D|\underline{u}') = K \left(\prod_{i=1}^{q} g_{\underline{P}}(\underline{P}_i) \right) g_N(N) \quad , \tag{4-49}$$

where

$\underline{P}_i$ = i^{th} individual sample preference values ,

$\underline{u}'$ = prior values of the generating function parameters.

Naturally, the data handling may be made more sophisticated by adjusting the assumptions regarding the interdependence of N and $\underline{P}$, and the joint sampling of N and $\underline{P}$. However, the updating procedure to improve the state of knowledge of the random traveler generating process $g_{\underline{N},\underline{P}}$ proceeds in accordance with the regular Bayesian inference methods.

Finally, consider the collection of data by recording observed traveler behavior in the context of an operational transport system. Presumably, the following factors may be observed: the available inventory of trip schedules and durations over a given route, the number of travelers aboard each trip, and the fare paid by each traveler. The data consists of a set of observations $D = (N^*, \underline{d})$ where,

N^* = the total number of travelers riding, and

$\underline{d} = (d_1, d_2, \ldots, d_{N^*})$

where

d_i = the decision of traveler i, i.e., the trip that traveler i chose to ride .

Thus,

$$d_i = j\ , \qquad 1 \leq j \leq J\ , \tag{4-50}$$

where each trip j is described by a triplet (π^j, t_d^j, t_a^j). Now, the likelihood of the data is,

$$\left\{D \middle| \underline{u}\right\} = \left\{N^*, d \middle| \underline{u}\right\} = \left\{N^* \middle| \underline{u}\right\} \left\{\underline{d} \middle| N^*, \underline{u}\right\}\ , \tag{4-51}$$

where

$\left\{N^* \middle| \underline{u}\right\}$ = probability that there are N^* travelers for whom at least one of the available trips is satisfactory,

and

$\left\{\underline{d} \middle| N^*, \underline{u}\right\}$ = probability that the N^* travelers make the choices $\underline{d}$.

The two factors in the final expression of Eq. (4-51) must be computed by appropriate manipulation of the generating function $g_{\underline{N},\underline{P}}$. Suppose, for example, that $\underline{N}$ and $\underline{P}$ are independent $\left(g_{\underline{N},\underline{P}} = g_{\underline{N}}(\underline{N}_o)\, g_{\underline{P}}(\underline{P}_o)\right)$, and that each traveler is considered as an independent sample. Then, $\left\{N^* \middle| \underline{u}\right\}$ is found by integrating $g_N(N_o) \prod_{i=1}^{N_o} g_{\underline{P}}(\underline{P}_i)$ over that portion of the N, $\underline{P}$ space for which $N \geq N^*$, and the N^* sets of travelers' parameters fall in a region such that the available trips are acceptable in the observed proportions. The probability $\left\{\underline{d} \middle| N^*, \underline{u}\right\}$ is given by,

$$\left\{\underline{d} \,\middle|\, N^*,\underline{u}\right\} = \prod_{i=1}^{N^*} \left\{d_i \,\middle|\, \underline{u}\right\} , \tag{4-52}$$

where $\left\{d_i \middle| \underline{u}\right\}$ is found by integrating $g_{\underline{N},\underline{P}}$ over the region in $\underline{P}$ space for which the traveler decision d_i is optimal.

Utilization of the likelihood function (4-51) in the context of the Bayesian method facilitates the estimation of the traveler generation process parameters. For either type of data, with corresponding likelihood functions (4-49) and (4-51), Bayes' rule is the mechanism for updating the state of knowledge of the traveler generation process parameters. Estimation of specific values for these parameters can proceed by minimizing the expected value of a suitable loss function. For example, minimizing the square loss function,

$$l(\underline{u},\underline{\hat{u}}) = \sum_i (u_i - \hat{u}_i)^2 , \tag{4-53}$$

where

u_i = i^{th} component of $\underline{u}$, the vector of random process parameters,

$\hat{u}_i$ = the estimated value of u_i,

is equivalent to finding the "least squares" estimate of parameters $\underline{u}$, based on the information contained in the prior distribution and the subsequently collected data.

4.8 Concluding Comment

This chapter has presented the formulation for handling three important interrelated facets of the passenger transportation scheduling problem: the formulation of uncertainty and Monte Carlo generation of traveler populations, the decision analysis for choosing among proposed alternative systems, and the manipulation of empirical data to up-date knowledge of the traveler generation process. Much of the discussion has been theoretical in nature and practical implementation may require the substitution of numerical methods for closed form analysis. However, it is important to understand the logical framework of this chapter, before such numerical techniques can be exploited.

CHAPTER V

ILLUSTRATIVE COMPUTER EXAMPLES

5.1 Introduction

This chapter presents some of the potential applications of the methodology developed in this research. In particular, the scheduling method allows the study of numerous interesting and important questions, the investigation of which fosters more intelligent evaluation of operating policies and planning alternatives. The examples exhibited here include consideration of alternate types of transportation vehicles and passenger markets, as well as the effects of changing fare levels, fleet sizes, and operator objectives.*

The context in which the examples are set, is a helicopter shuttle service between two city centers, forty miles apart.** The scheduling period is assumed to be a twelve hour day (e.g. 8 a.m. to 8 p.m.). Two types of vehicles, corresponding to proposed designs for future operations,*** are considered. Two traveler population distributions are specified (arbitrarily) with the intention of representing contrasting types of travelers. One sample traveler population is picked (randomly) from each distribution, to serve as input to the scheduling computations.

---

* The computer runs in this chapter average approximately two minutes each. The set of examples represents an investment of approximately one hour of computation time on the IBM 360/67.

** Such as Washington, D.C. and Baltimore, Maryland.

*** The vehicle characteristics correspond to the Lockheed CL-1026 and the Sikorsky S-65. See Ref. [68]. The results of this chapter are only illustrative and are not intended to reflect on the relative merits of these vehicles.

(No numerical illustrations of the Bayesian analysis, involving the generation of results from additional sample populations, are presented.) All relevant values (vehicle capacities, operating costs, and numbers of travelers) are scaled down by a factor of six from what might be considered the level of actual operation.

5.2 Inputs

All of the examples were executed under the option to forego use of the optimum assignment algorithm, and with null initialization (i.e., no initial timetables or traveler assignments were specified). The time grid was set at $K = 120$, so that departure decisions were considered at every six minute (0.1 hour) interval.

The alternate vehicle types are described in Fig. 36. Vehicles of type 1 are smaller, faster, and cheaper to operate than vehicles of type 2. The vehicle speeds are such that the forty mile trip takes 10.4 minutes and 13 minutes, via vehicle types 1 and 2, respectively. Minimum turnaround times are assumed to be ten minutes at each city, for each vehicle type. Throughout most of the examples two alternate vehicle fleets will be considered; the first consists of four vehicles of type 1 and the second consists of three vehicles of type 2. These fleets represent roughly commensurate purchase costs of 5.2 million dollars and 5.7 million dollars, respectively.

To generate the traveler populations, the number of potential travelers from each node must be specified. These numbers were selected randomly from the (properly scaled) distribution shown in Fig. 37, via the Monte Carlo selection process mentioned in Chapter IV. The result is 41 travelers from node 1, and 47 travelers from node 2, for these examples.

Vehicle Type / Attribute	1	2
Seating Capacity*	5	10
Average Speed	230 mph	184 mph
Operating Cost*	30¢/mile	40¢/mile
Purchase Cost	$1.3 million	$1.9 million

* Numbers scaled, as previously noted.

Figure 36. Vehicle Characteristics.

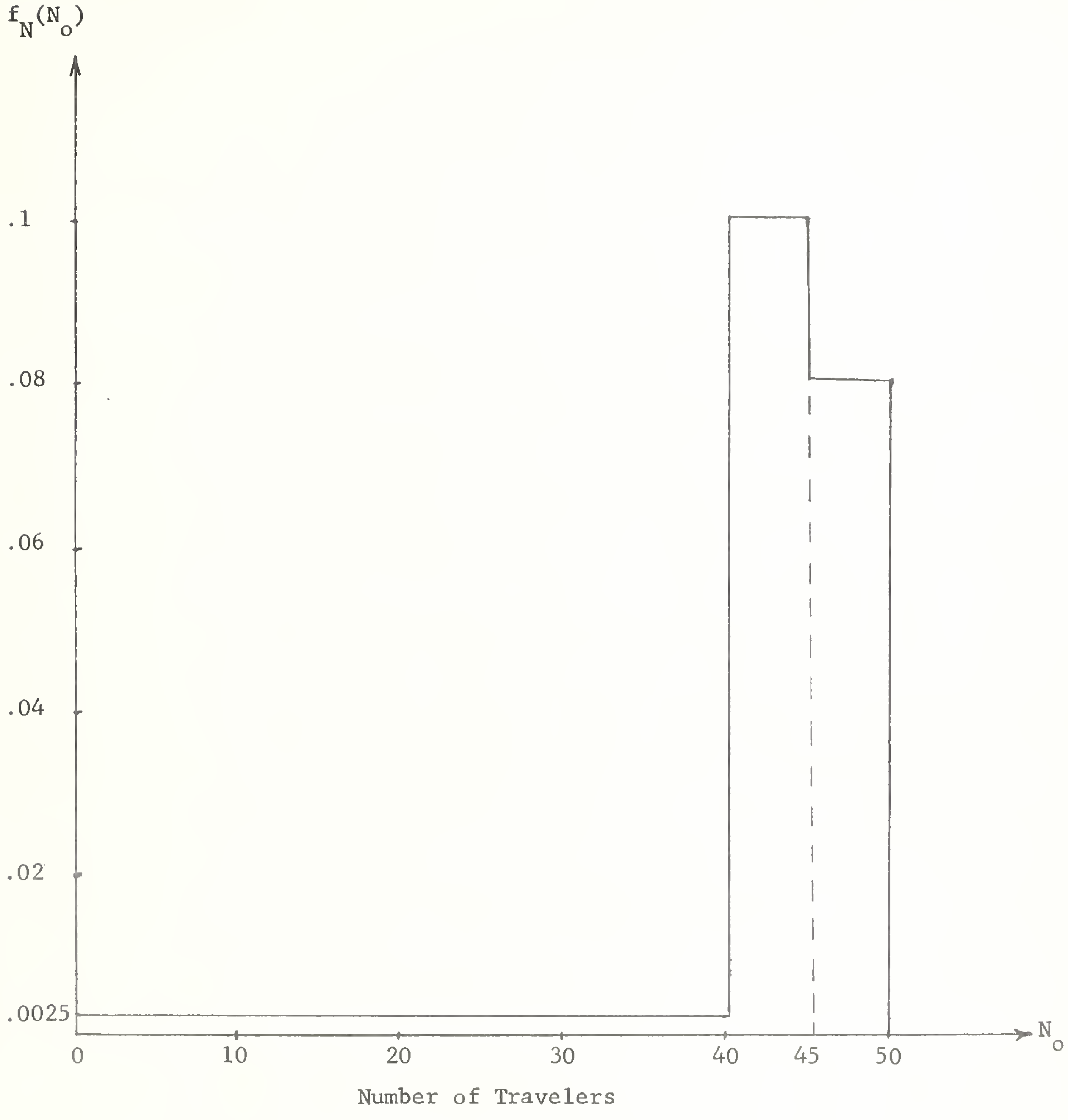


Figure 37. Distribution over the Number of Potential Travelers per Node.

As noted earlier, two types of travelers are considered. To simplify matters, each of these types is characterized by fixed values of the willingness to pay parameters (D,α,n,s). The (t^*,w) time preference values for each traveler were selected randomly from probability distributions discussed below.

The first traveler population, called type B (business), is described by the parameter values $(D = 25., \alpha = 0.467, n = 2, s = 0.67)$. The indifference curve, at willingness to pay level WTP = \$8., is illustrated in Fig. 38. Thus, at a fare level π = \$8., the curve of Fig. 38a represents the set of minimally acceptable trips. The time preference values (t^*,w) for travelers of type B were selected from the probability distributions $f_{t^*}(t^*_o)$ and $f_{w/t^*}(w_o/t^*)$ shown in Figs. 38b and c. These distributions exhibit a.m. and p.m. peaks. In addition, the distribution f_{w/t^*} is specified such that a.m. travelers tend to be arrival-oriented, while p.m. travelers tend to be departure-oriented. Such behavior may be characteristic of daily business trips for which the time of business appointments is the important consideration in the earlier hours, and the time of departure for the return trip is important later in the day. The set of randomly selected type B travelers for these examples, is exhibited in Fig. 40a.

The second traveler population, type P (general public), is given by parameter values $(D = 20., \alpha = 0.697, n = 1, s = 0.5)$. The \$8. indifference curve is shown in Fig. 39a. Note that this traveler type is less sensitive to schedule inconvenience and is willing to trade-off inconvenience for speed in a constant ratio, irrespective of the relative values of Δ and ϕ. Also, observe that a traveler of type P has

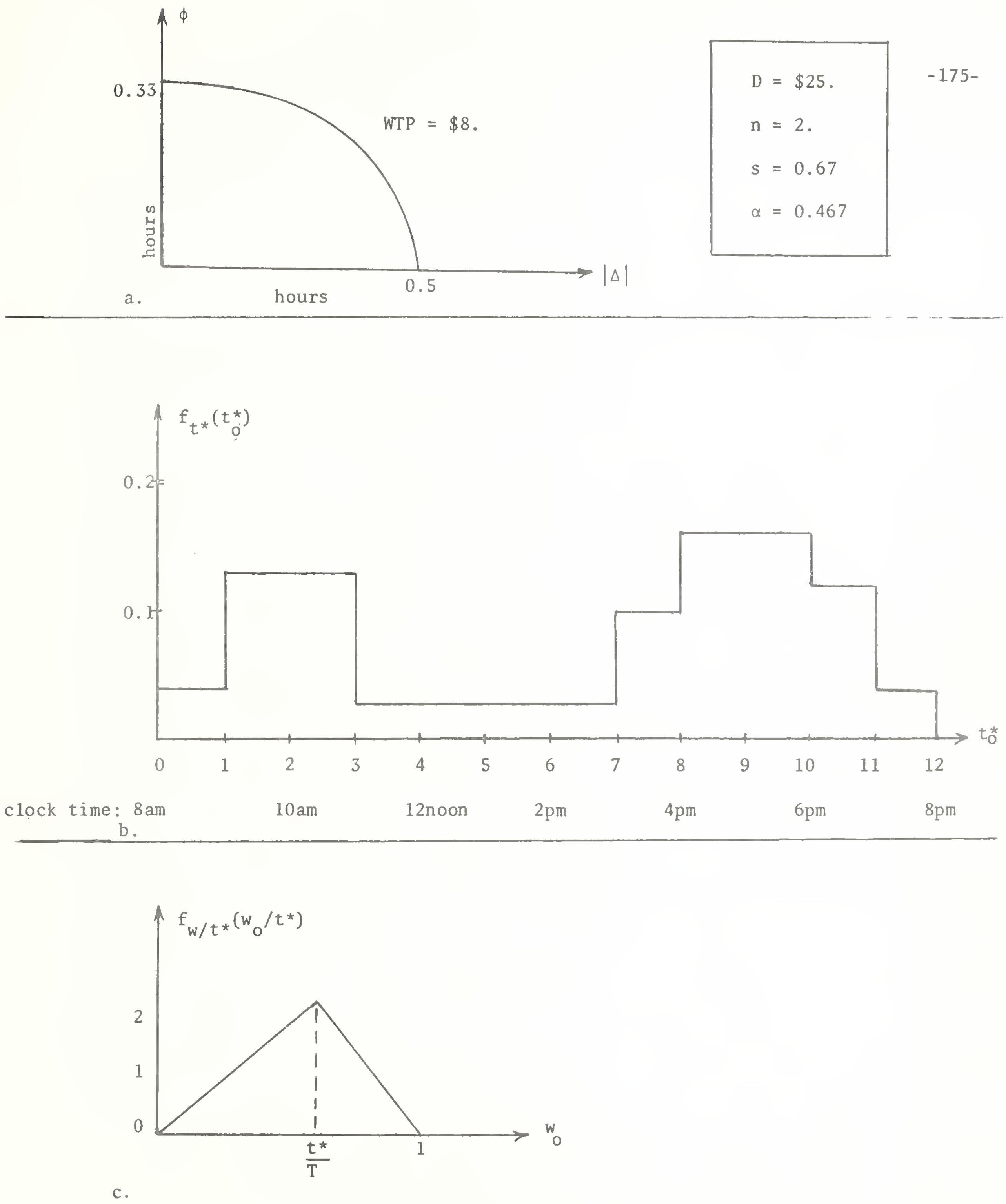


Figure 38. Traveler Population Type B (business).

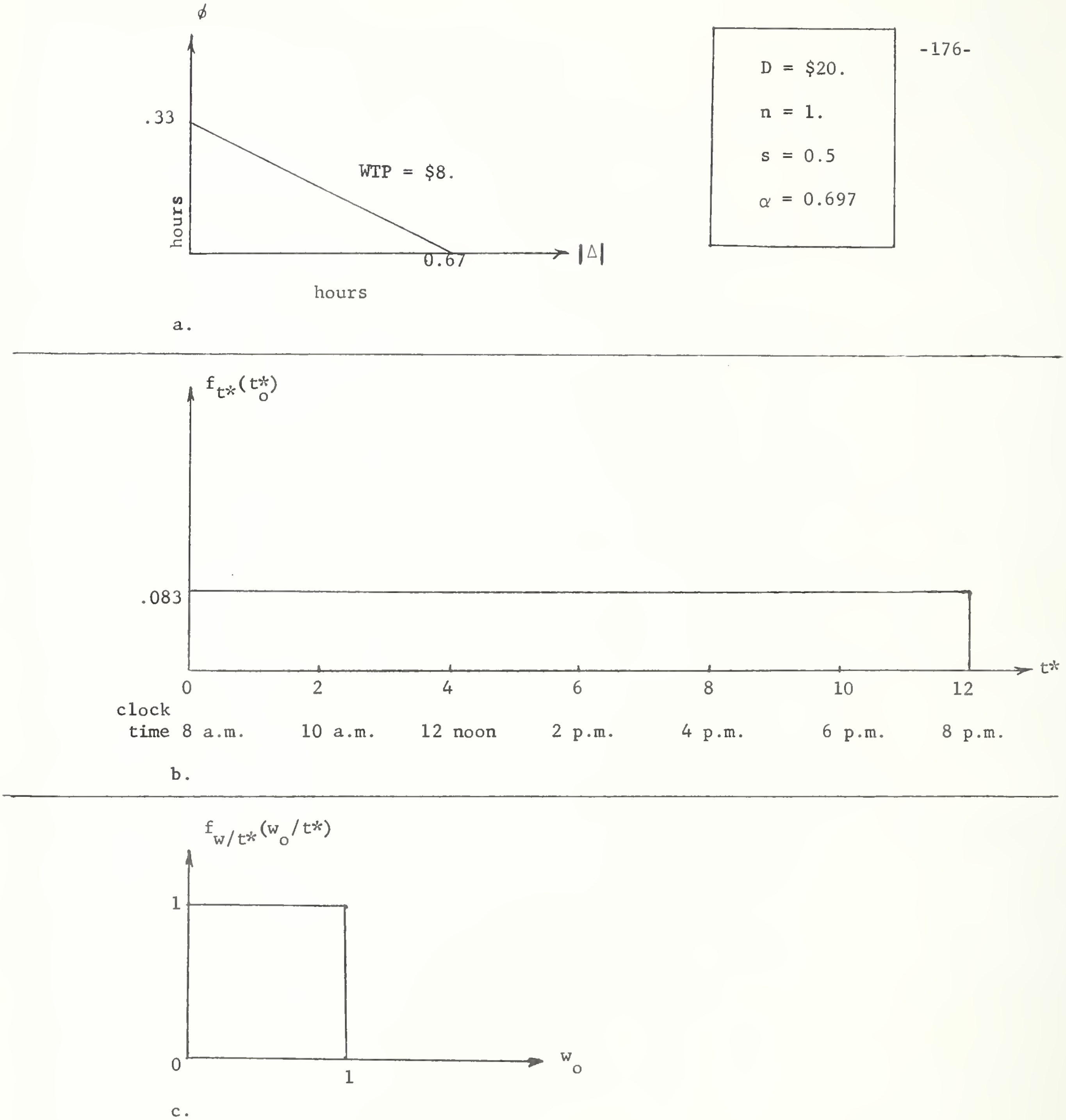


Figure 39. Traveler Population Type P (general public).

LIST OF TRAVELERS

NODE 1

PASSENGER NO.	TSTAR	W
1	1.050	0.33
2	8.410	0.08
3	9.600	0.68
4	7.720	0.69
5	10.840	0.29
6	1.440	0.52
7	8.360	0.75
8	7.930	0.07
9	7.050	0.54
10	1.590	0.21
11	8.760	0.48
12	1.330	0.50
13	8.120	0.75
14	11.400	0.77
15	10.910	0.30
16	1.030	0.24
17	10.100	0.79
18	9.920	0.43
19	10.270	0.47
20	9.250	0.91
21	3.070	0.19
22	1.430	0.24
23	2.570	0.26
24	8.680	0.14
25	7.970	0.77
26	7.160	0.91
27	10.870	0.88
28	8.950	0.46
29	8.780	0.49
30	2.080	0.13
31	11.270	0.94
32	10.100	0.78
33	9.020	0.66
34	2.970	0.46
35	0.630	0.40
36	7.440	0.39
37	3.070	0.45
38	9.970	0.75
39	8.400	0.49
40	9.790	0.78
41	2.080	0.31

NODE 2

PASSENGER NO.	TSTAR	W
1	9.200	0.79
2	7.030	0.41
3	10.660	0.91
4	3.130	0.37
5	0.870	0.06
6	0.180	0.34
7	3.530	0.65
8	1.370	0.13
9	0.350	0.64
10	1.180	0.40
11	0.300	0.46
12	1.690	0.29
13	5.000	0.39
14	11.650	0.75
15	8.390	0.46
16	8.500	0.78
17	8.710	0.43
18	9.860	0.71
19	5.680	0.59
20	9.010	0.87
21	7.580	0.53
22	4.520	0.20
23	8.020	0.36
24	7.080	0.77
25	4.790	0.36
26	9.270	0.61
27	5.470	0.78
28	1.580	0.32
29	9.590	0.58
30	8.330	0.65
31	8.070	0.66
32	1.560	0.63
33	1.020	0.44
34	8.470	0.66
35	3.170	0.30
36	2.490	0.33
37	7.560	0.95
38	9.010	0.32
39	9.190	0.98
40	8.470	0.64
41	1.870	0.46
42	7.990	0.42
43	10.880	0.48
44	11.630	0.80
45	1.770	0.62
46	9.150	0.78
47	5.590	0.29

Figure 40a. Population B. Time Preference Parameters.

LIST OF TRAVELERS

NODE 1

PASSENGER NO.	TSTAR	W
1	0.560	0.50
2	7.040	0.01
3	9.320	0.58
4	5.900	0.73
5	11.290	0.09
6	1.170	0.74
7	6.930	0.79
8	6.150	0.01
9	5.110	0.49
10	1.400	0.28
11	7.700	0.31
12	1.000	0.71
13	6.460	0.80
14	11.710	0.62
15	11.400	0.10
16	0.520	0.37
17	10.220	0.75
18	9.920	0.22
19	10.470	0.26
20	8.640	0.96
21	3.630	0.14
22	1.150	0.35
23	2.930	0.30
24	7.540	0.03
25	6.210	0.85
26	5.230	0.98
27	11.330	0.86
28	8.060	0.28
29	7.750	0.33
30	2.170	0.10
31	11.650	0.93
32	10.220	0.72
33	8.190	0.57
34	3.550	0.62
35	0.300	0.62
36	5.570	0.25
37	3.620	0.59
38	10.030	0.68
39	7.010	0.34
40	9.680	0.74
41	2.160	0.42

NODE 2

PASSENGER NO.	TSTAR	W
1	8.550	0.81
2	5.070	0.29
3	11.030	0.93
4	3.650	0.46
5	0.420	0.05
6	0.090	0.56
7	3.790	0.82
8	1.060	0.14
9	0.170	0.86
10	0.750	0.60
11	0.140	0.71
12	1.550	0.42
13	4.320	0.37
14	11.830	0.58
15	6.990	0.31
16	7.200	0.84
17	7.600	0.25
18	9.800	0.62
19	4.560	0.68
20	8.190	0.94
21	5.730	0.44
22	4.150	0.11
23	6.270	0.19
24	5.130	0.87
25	4.240	0.32
26	8.680	0.48
27	4.490	0.91
28	1.390	0.47
29	9.300	0.42
30	6.880	0.62
31	6.370	0.65
32	1.360	0.84
33	0.510	0.66
34	7.150	0.61
35	3.660	0.33
36	2.800	0.43
37	5.710	0.99
38	8.180	0.14
39	8.530	1.00
40	7.150	0.59
41	1.840	0.65
42	6.230	0.27
43	11.350	0.25
44	11.820	0.67
45	1.680	0.83
46	8.460	0.79
47	4.530	0.19

Figure 40b. Population P. Time Preference Parameters.

less money to spend for the trip and will probably use other means of transit if the fare is too high. In contrast to traveler type B, time preference parameters for travelers of type P are selected from the uniform distributions shown in Figs. 39b and 39c. The set of randomly selected type P travelers is shown in Fig. 40b.

Before considering the computer results, an additional note on input values is necessary. In order that the double-counting inequality (3-26) be properly observed, the fare levels must be maintained above a certain lower bound for each vehicle type-passenger population combination. The appropriate minimum valid fares are tabulated in Fig. 41. Slight violation of these minimum fares incurs only small risk of error. In fact, some computer runs have been included at fare level π = \$10, for the combination of population B with vehicle type 1. There is no evidence of error in this case, although substantial errors do occur for lower fares.

Unfortunately, the region of invalid fares may (alas!) be an interesting one to investigate. For instance, in the succeeding examples, it happens that the fare that maximizes the objective function value often falls below the minimum valid level. In such cases, the MODEL2 program is unable to do more than identify the general region of interest. Although it would not be difficult to present a transportation example in which the double counting inequality holds even for very low fare levels, the current examples serve to identify the limits within which the present methodology is useful.

5.3 Computer Runs

One of the most important areas that a series of computer runs can

Traveler Population \ Vehicle Type	1	2
B	$10.80	$7.81
P	$7.52	$5.08

Figure 41. Minimum Valid Fare Levels.

help investigate is the behavior of the significant system performance measures, over a range of parameter values. To illustrate this function, consider the results of the scheduling computations for population B and vehicle type 1 (4 vehicle fleet),under maximization of user benefit minus cost, over a range of fare values. The important system quantities are plotted in Fig. 42. At each level of fare, the MODEL2 computation produces the value of revenue, patronage (demand), operating cost, profit, and benefit minus cost, for the (optimal) timetable. It is clear that the kind of information developed in Fig. 42 is of direct interest to system planners and operators alike.

In addition to their specific numerical content, curves such as those of Fig. 42 are of interest for the insights they provide into system behavior and scheduling policy formation. Specifically, the behavior exhibited in Fig. 42 clarifies the implications of optimizing the difference between user benefits and operating costs. As the fare is increased, the net benefit that accrues to each passenger decreases. As a result, fewer trips remain "profitable" (in a benefit - cost sense), and fewer travelers are able to find acceptable accommodation. Thus, patronage declines as the fare increases. Costs also decline since fewer trips are executed. Revenue, which is the product of fare and patronage, rises at first and then falls. Operator profit, the difference of revenue and cost, also rises and then falls. Finally, since traveler benefits decline faster than operating costs, as fare increases from ten to fourteen dollars, the objective value (benefit minus cost) declines as well. Although the appropriate region is inaccessible for MODEL2 computations, it is apparent that the objective would be enhanced

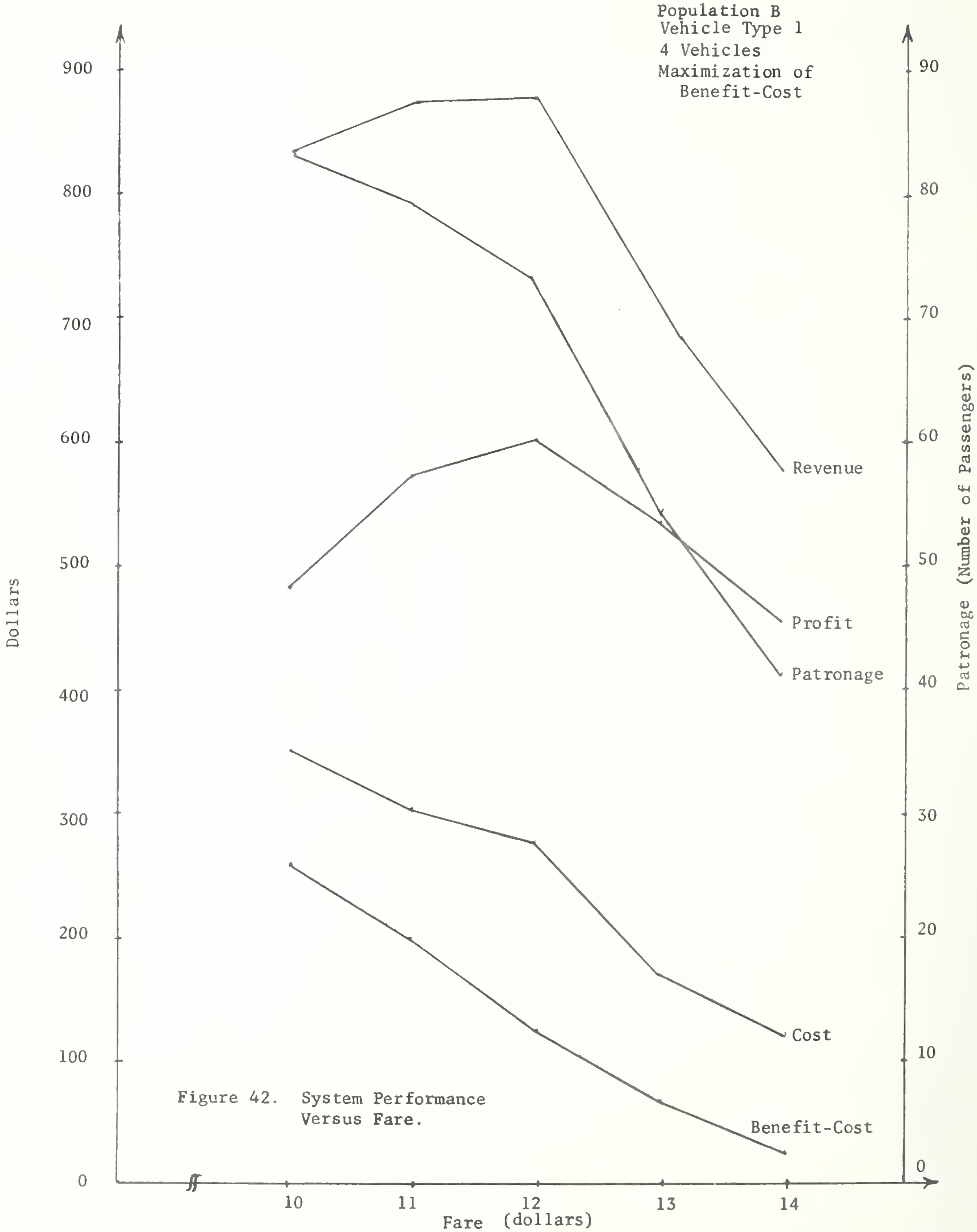


Figure 42. System Performance Versus Fare.

by lowering the fare, below ten dollars, until costs begin to rise faster than traveler benefits.

A second important product of the scheduling computations is the set of timetables corresponding to the alternate fare levels. The successive schedules are displayed in Fig. 43, for fares ranging from eleven to fourteen dollars. In each timetable diagram the line segments above the time axis represent trip departures from node 1. Line segments below the axis correspond to departures from node 2. The integer attached to each line segment specifies the number of passengers aboard the given trip.

The set of timetables in Fig. 43 shows several interesting characteristics. One property is the relative "stability" of the timetables as fare is changed. Although fewer trips are executed as the fare is raised, trips have a tendency to "drop out," leaving the rest of the timetable almost unaffected. In other words, a decline in the number of trips in the (optimal) schedule, resulting from a rise in fare, does not imply a major realignment in the timetable. Under a system that exhibits such behavior, the latter observations could be taken as an operational guideline for adjusting system operations.

Two major properties of the timetable solutions are plotted in Fig. 44. First, the number of trips required in the schedule declines as the fare increases. Second, the load factor, i.e. the percentage of available seats filled by travelers, increases with fare under the benefit-cost objective. Such behavior stems from the decline of traveler benefits, and the number of trips. In short, more travelers must be accommodated on fewer trips to make operations profitable (in a benefit-cost sense) as the fare rises.

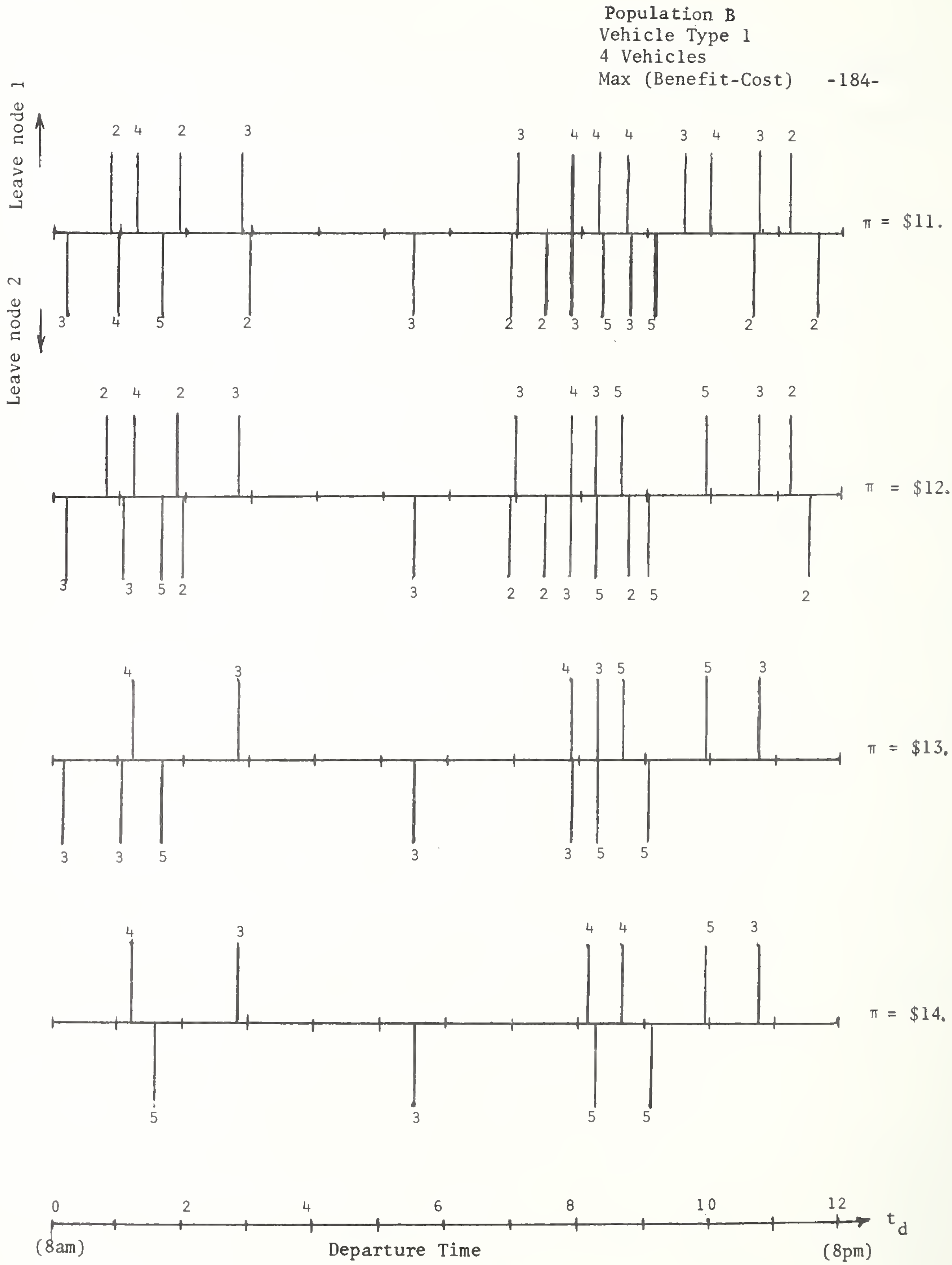


Figure 43. Variation of Fleet Timetable with Fare.

Population B
Vehicle Type 1
4 Vehicles
Maximization (Benefit-Cost)

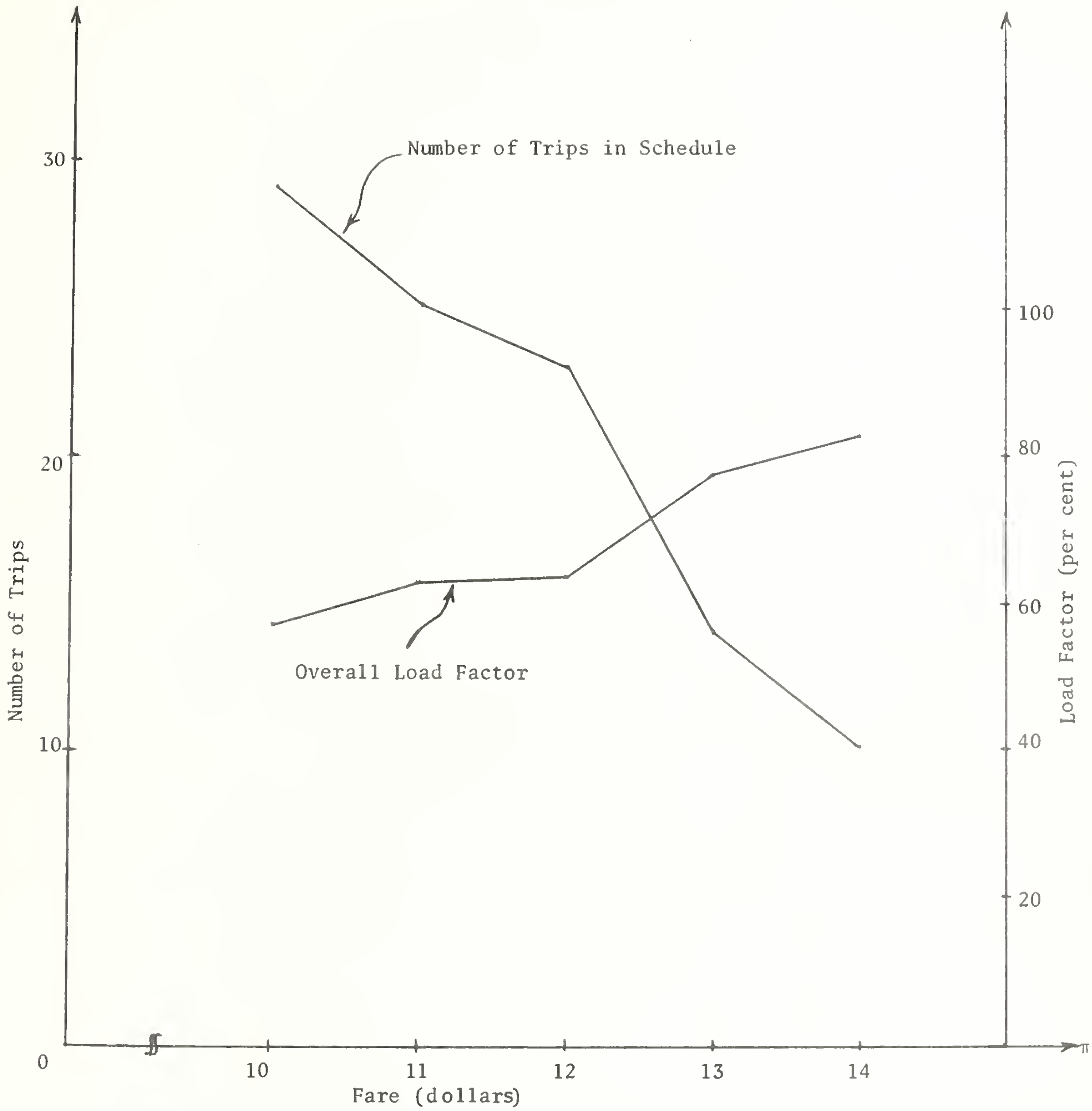


Figure 44. Variation of Schedule Properties with Fare.

5.4 Comparative Studies

In the remaining examples, the outputs of scheduling computations will be illustrated for alternate system combinations of traveler and vehicle types. In addition to variation of fares, parametric changes in fleet size and operator objectives, will be considered.

Fig. 45 displays the patronage, profit, and benefit minus cost curves, plotted against fare, for populations B and P. The fleet consists of four type 1 vehicles, and the objective is maximization of benefit minus cost. It is apparent from the graphs that population P is less effectively served than population B. Several factors are responsible for this. First, a traveler of type P is generally willing to pay less for his trip than a traveler of type B. Hence, at a given level of fare, a traveler of population P is likely to receive a smaller net benefit than his type B counterpart. In addition, the alternate distributions of traveler time preferences for populations B and P, imply different potentials for satisfactory service. In particular, the "peaked" distribution of population B allows the execution of trips at times satisfactory to relatively large numbers of travelers. For these reasons it is more difficult to initiate profitable trips (in a benefit-cost sense) for population P than for population B.

It is apparent from Fig. 45, that the two traveler markets imply different service requirements. For example, the various indices of performance-- patronage, profit, and benefit minus cost-- are found to drop off more quickly with fare for traveler type P. Hence, a lower fare structure is indicated for that market. Naturally, the two traveler markets require substantially different scheduling policies. Fig. 46

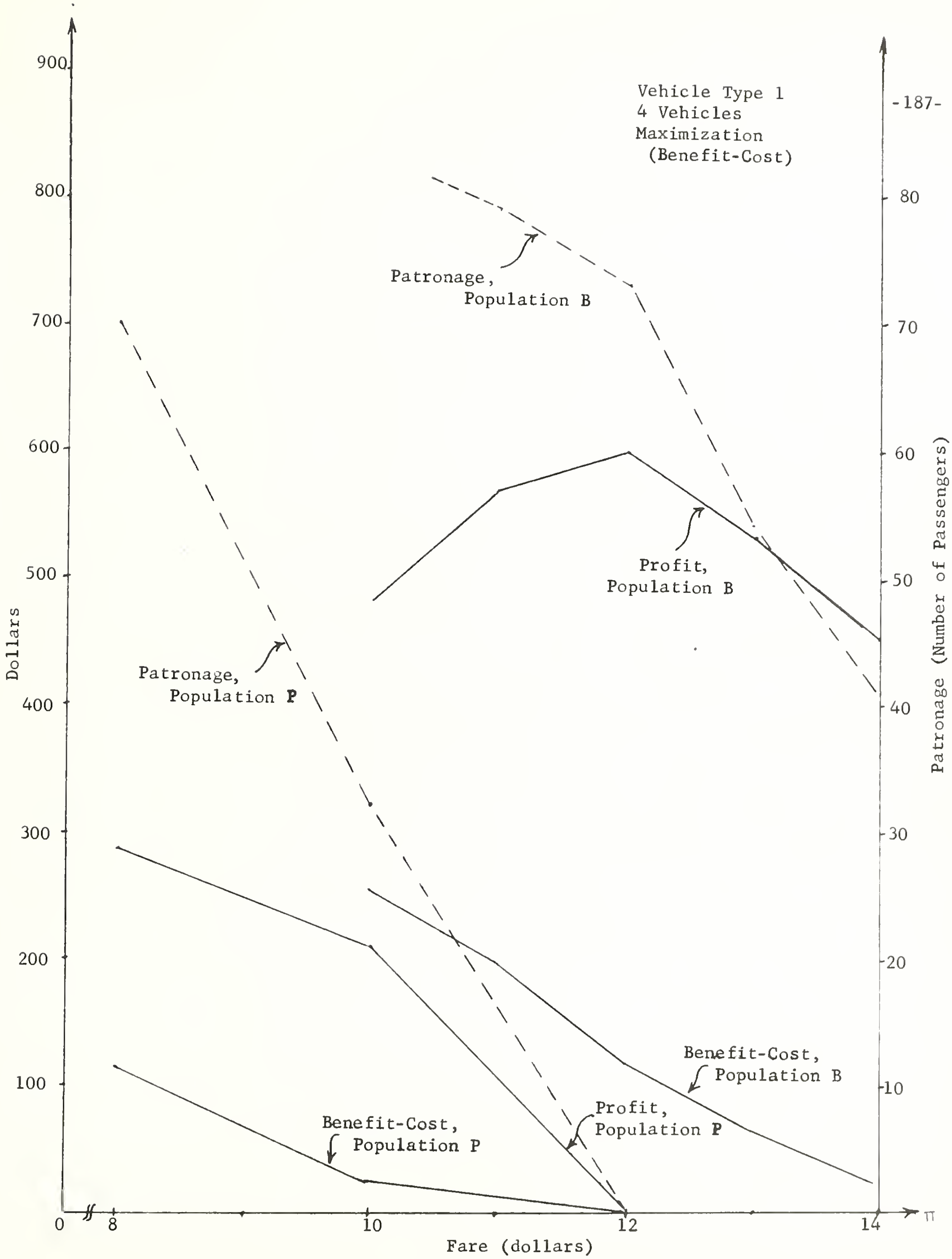


Figure 45. System Performance for Alternate Traveler Populations.

displays the solution timetables for population P at fare π = \$8., and for population B at fare π = \$12. These schedules correspond to similar objective function values. It is clear that the timetables follow the appropriate time preference distributions rather closely.

Fig. 47 compares the performance of alternate vehicle fleets (four type 1 vehicles versus three type 2 vehicles), in the service of traveler population B under benefit minus cost maximization. Over the range of fares considered, the smaller but faster and cheaper type 1 vehicle appears to be the better alternative. This is the result of several factors. First, travelers are willing to pay more for faster trips; hence, vehicle type 1 offers greater benefits in this respect. Second, vehicle type 2 operating costs are higher, so that the "cost" component of the benefit-cost objective is greater for this alternative. In addition, a vehicle type 2 fleet will tend to offer fewer trips since it is more difficult to synthesize "profitable" trips, for the two reasons just given. Finally, the size of the traveler market in the current example is such that the greater capacity of the type 2 vehicle cannot be utilized to full advantage.

As for earlier examples, the vehicle 2 system appears to operate best in the region of lower fare levels. In comparison, the vehicle type 1 system is more efficient than the vehicle type 2 system, at higher fare values. The contrast in vehicle properties also has significant implication with respect to timetable construction. Fig. 48 compares the timetables corresponding to the alternate fleets, at levels of fare that yield similar objective function values. In essence, the faster and smaller type 1 vehicles operate best by executing a high freqeuncy

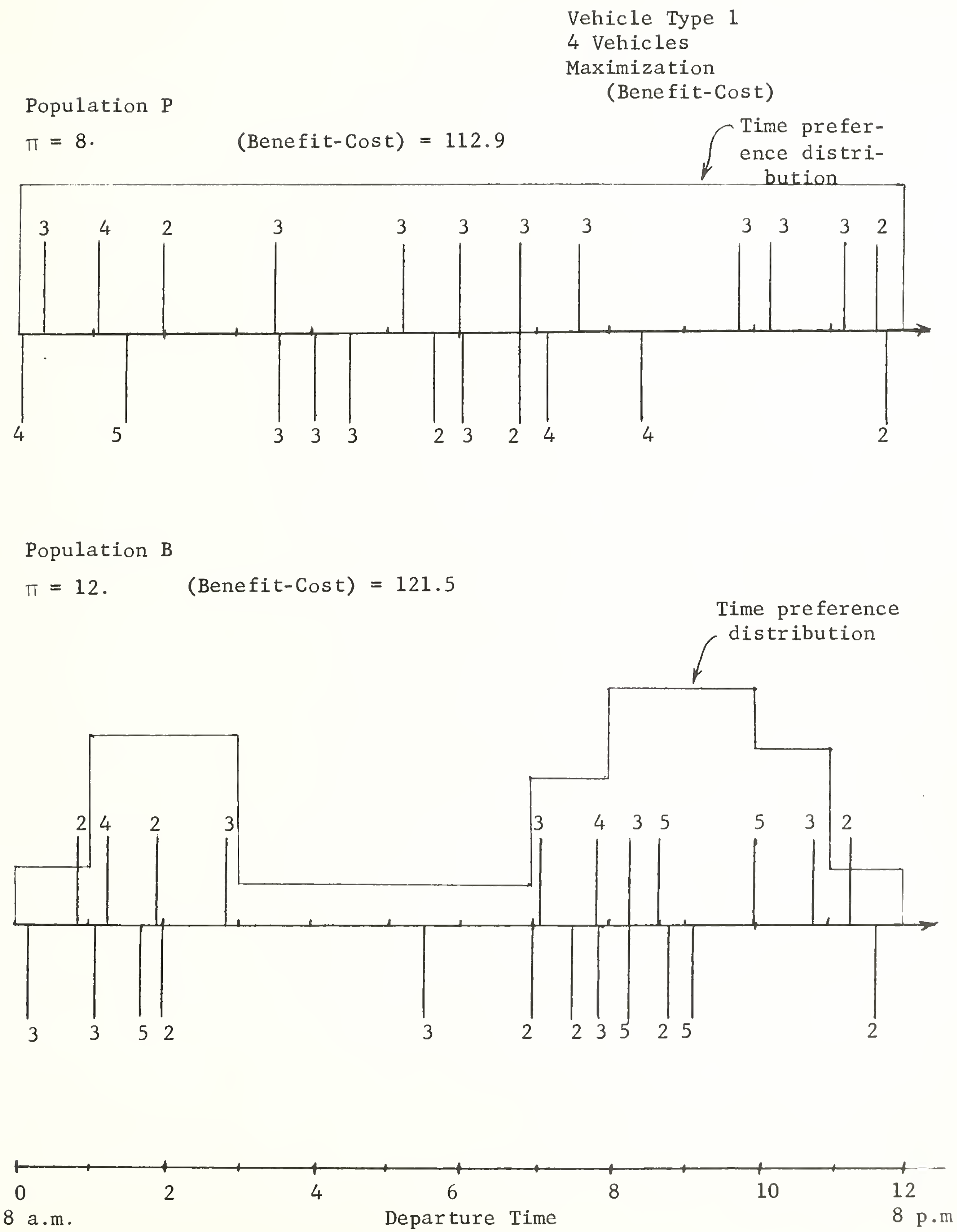


Figure 46. Comparison of Timetables for Alternate Traveler Populations.

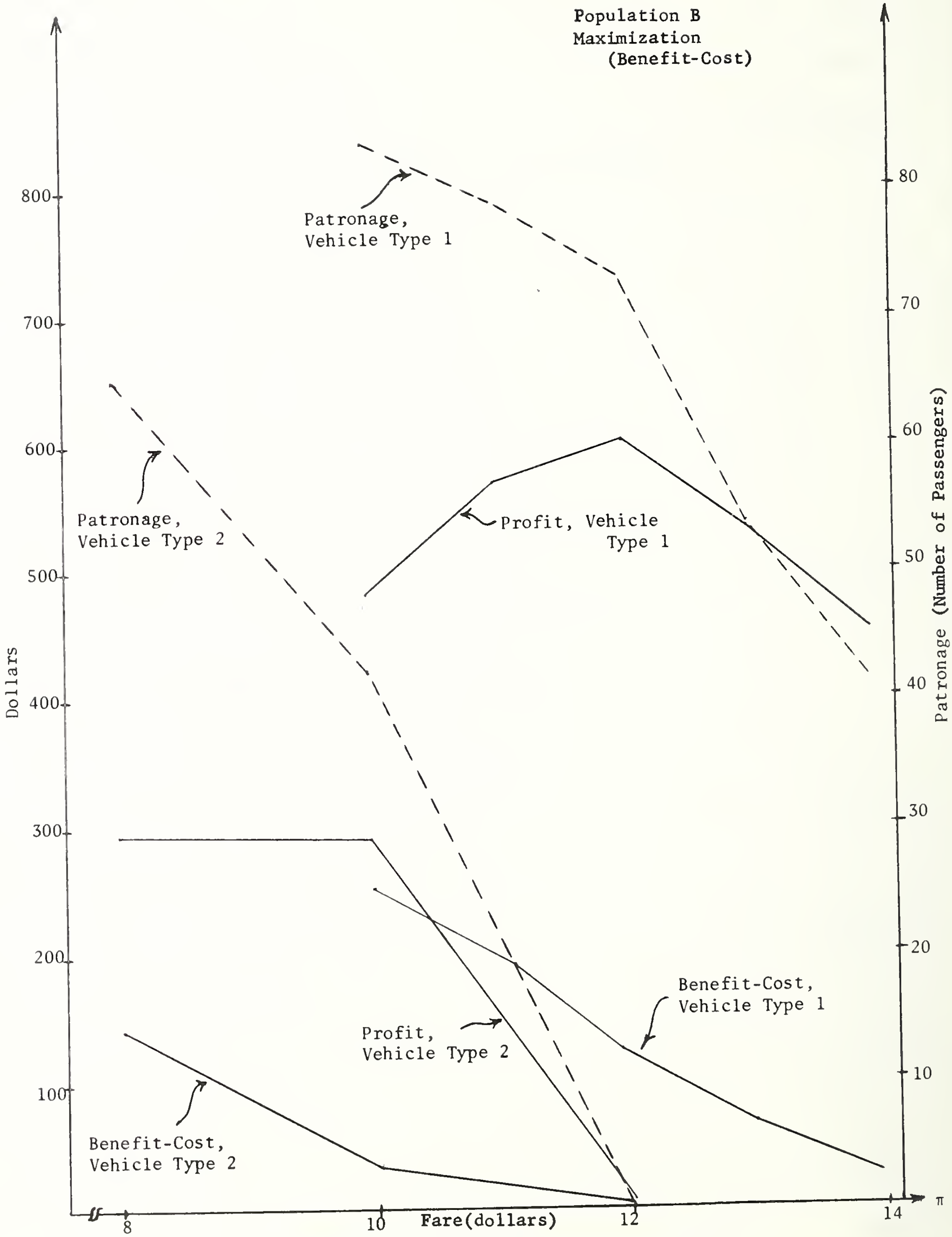


Figure 47. System Performance for Alternate Vehicle Fleets.

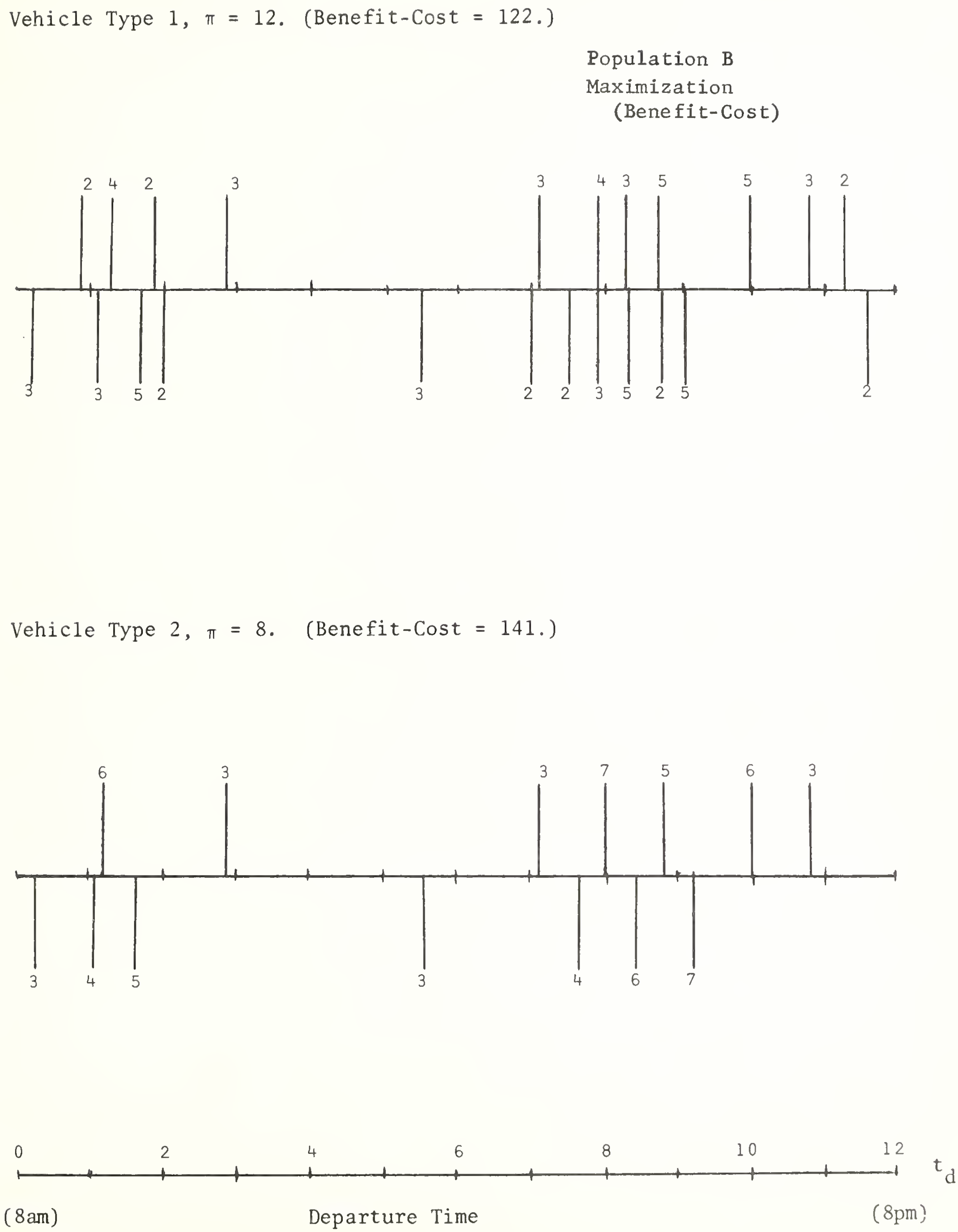


Figure 48. Comparison of Timetables for Alternate Vehicle Fleets.

service. The type 2 vehicles are better suited for a low frequency timetable, with higher passenger loads per trip.

A parameter of great interest to planners is the size of the vehicle fleet required for satisfactory operation. Fig. 49 displays the variation in the level of profits and benefit-cost, as the fleet size is varied from one to five vehicles. The inputs are population B, fare level of ten dollars, and the benefit-cost objective. Both vehicle types are considered. Note that each of the curves exhibits a "saturation" effect. In essence, this indicates that the particular passenger market can be maximally served, with respect to the specified objective, by a limited number of vehicles. Once a certain fleet size is reached, the remaining traveler demand is inadequate to support additional service. In Fig. 49, it is shown that two vehicles of the type 2 variety are sufficient to attain the maximal objective value. In contrast, each of the first four additional vehicles of type 1 makes some positive contribution to the value of the objective. This result is in keeping with the fact that vehicle type 1 is better able to run at low loads. Thus, incremental additions to the objective value are more likely to be feasible for the latter type vehicle.

The discussion thus far has presented the results of computations under the benefit minus cost objective only. Examples using a second important objective, the maximization of profit, will now be considered. Fig. 50 compares the behavior of profit and operating cost, as functions of the fare, under the alternate objectives--benefit minus cost, and profit. Several interesting points emerge from this comparison.

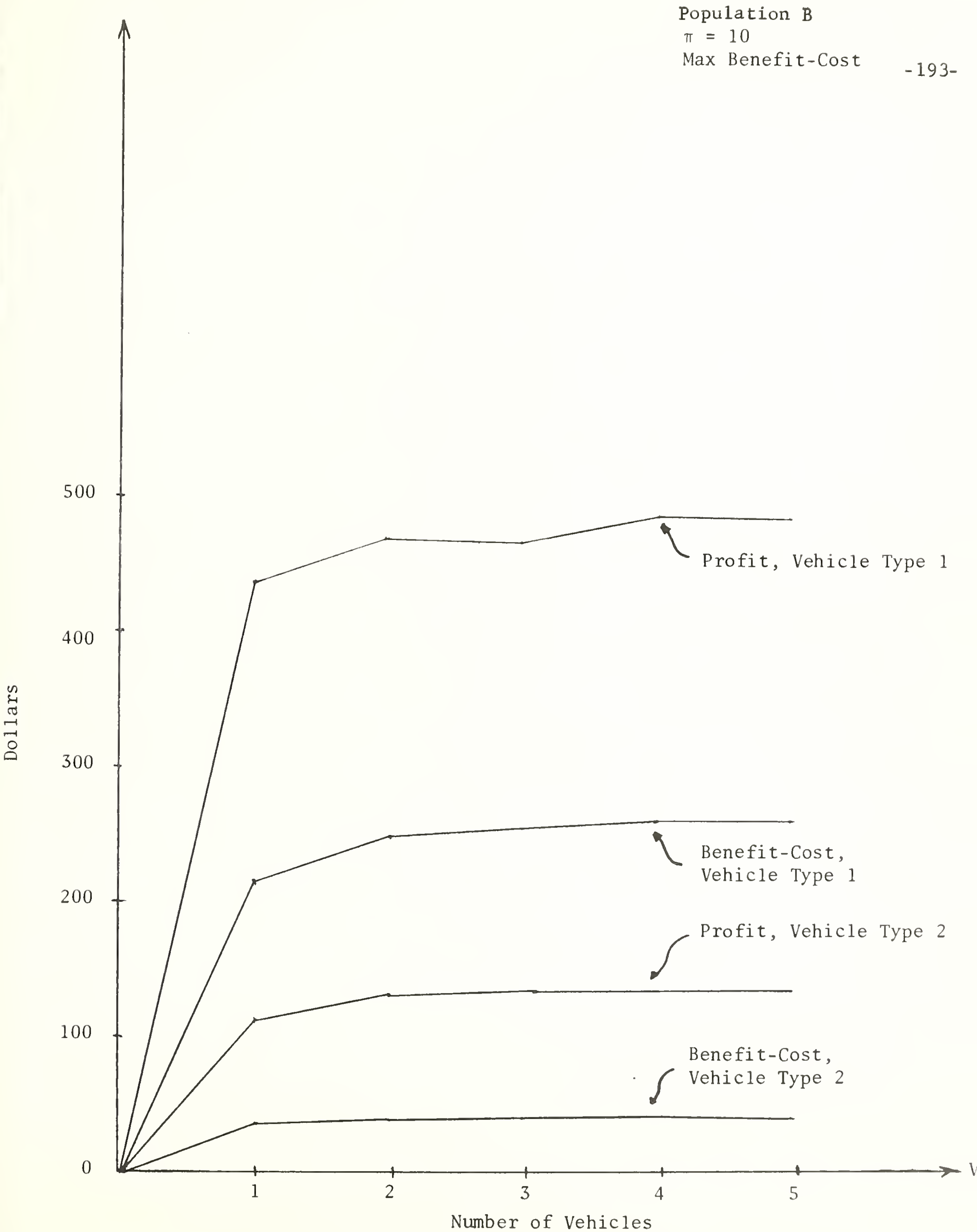


Figure 49. Variation of Performance with Fleet Size.

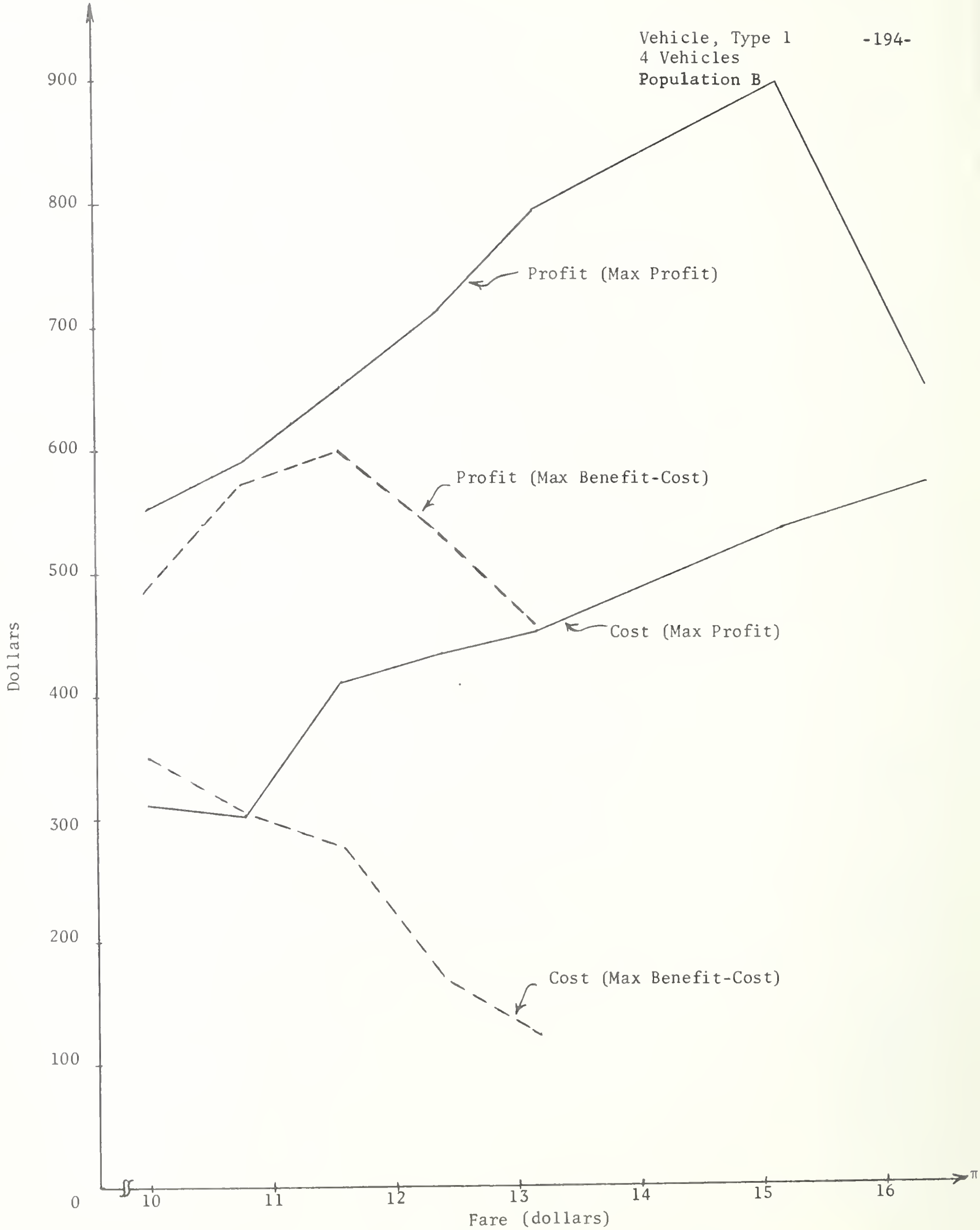


Figure 50. Variation of Profit and Cost under Alternate Objectives.

Naturally, the profit level is higher, at each value of fare, under profit maximization than under benefit-cost maximization. But, in addition, profit is maximized at a substantially higher level of fare under the profit objective than under the benefit-cost criterion. Furthermore, the maximum profit level is significantly greater under the profit maximizing criterion. This behavior is directly attributable to the fact that shifting from benefit-cost to profit maximization implies a transfer of economic benefit from the system users to the system operator. In particular, an increase in the fare, neglecting its effects on patronage or operating policy, represents an increase in profit (revenue) to the system, and a decrease in net benefit (willingness to pay minus fare) to the user. Thus, under profit maximization, the operator will want to increase the level of fare until patronage losses and cost increases combine to make further fare increases unprofitable. Increasing fare in this way implies a transfer of benefits from the traveler to the system.

The two cost functions, corresponding to the alternate objectives, present a striking contrast. As discussed earlier, the operating cost under the benefit-cost criterion is a decreasing function of fare. In contrast, the operating cost under profit maximization is generally increasing. This is a result of the fact that as the fare increases (in the range illustrated) it becomes more profitable to run additional trips. Hence, the increase in cost corresponds to the expansion of the number of trips in the schedule, as fare increases. Of course, if the fare reaches too high a level, passengers will drop out rapidly and service and operating costs will begin to fall.

Fig. 51 displays the variation of patronage with fare, under

Population B
Vehicle Type 1,
4 Vehicles

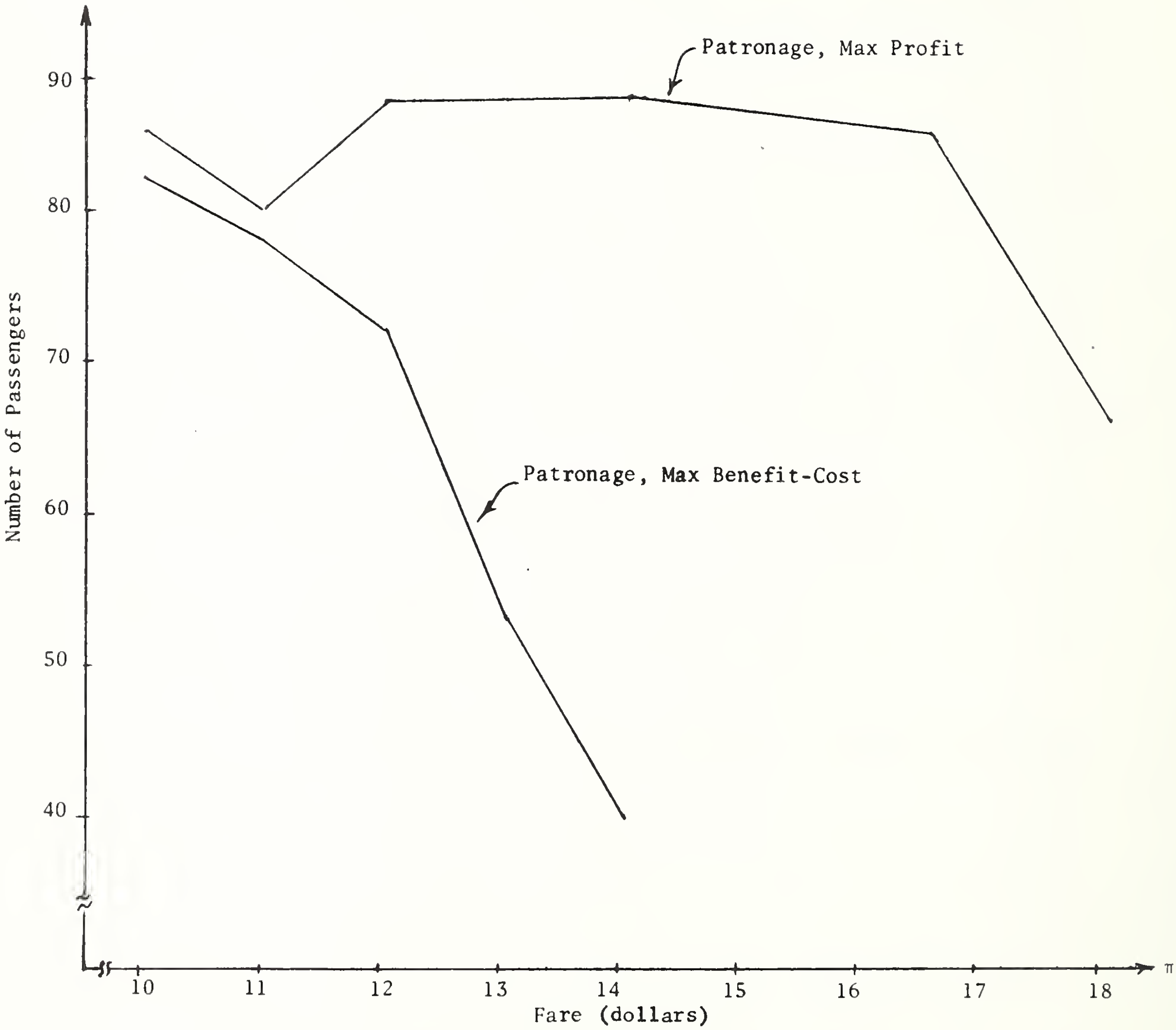


Figure 51. Variation of Patronage under Alternate Objectives.

alternate objective functions. Again, the contrast in behavior is apparent. Under benefit-cost, an increase in the fare implies a decrease in user benefits and a cut-back in the timetable. Hence, a loss in patronage results. Under profit maximization, however, although the increase in fare discourages travelers, the simultaneous enhancement of service tends to draw passengers to the system. As a result, a fare increase may enhance patronage in a region where the increase implies a more than compensatory increase in the level of service. Specifically, in Fig. 51, a fare increase from eleven to twelve dollars increases patronage. From twelve to fourteen dollars, the market is saturated (all potential travelers are served). Above fourteen dollars, the fare increase loses travelers, despite enhancement of service. In the range of higher fares, patronage falls quickly, as the number of travelers willing to pay the price decreases; eventually, at very high fares, the enhancement of service is no longer worthwhile.

Fig. 52 displays the variation in the provided service, as the fare is increased under the profit criterion. As mentioned earlier, the number of scheduled trips increases with fare. In addition, the load factor decreases with fare. The decreasing load factor is attributable to the fact that at higher fares, it becomes profitable to "personalize" service to the tastes of smaller groups of travelers. For example, above the fare of twelve dollars, a trip with only one traveler aboard becomes profitable. In other words, as the fare increases, the "breakeven" load factor decreases, and smaller vehicle loads become profitable. Incidentally, the increase in scheduled service with fare is not necessarily monotonic. Local decreases are possible, as shown, for example, by the

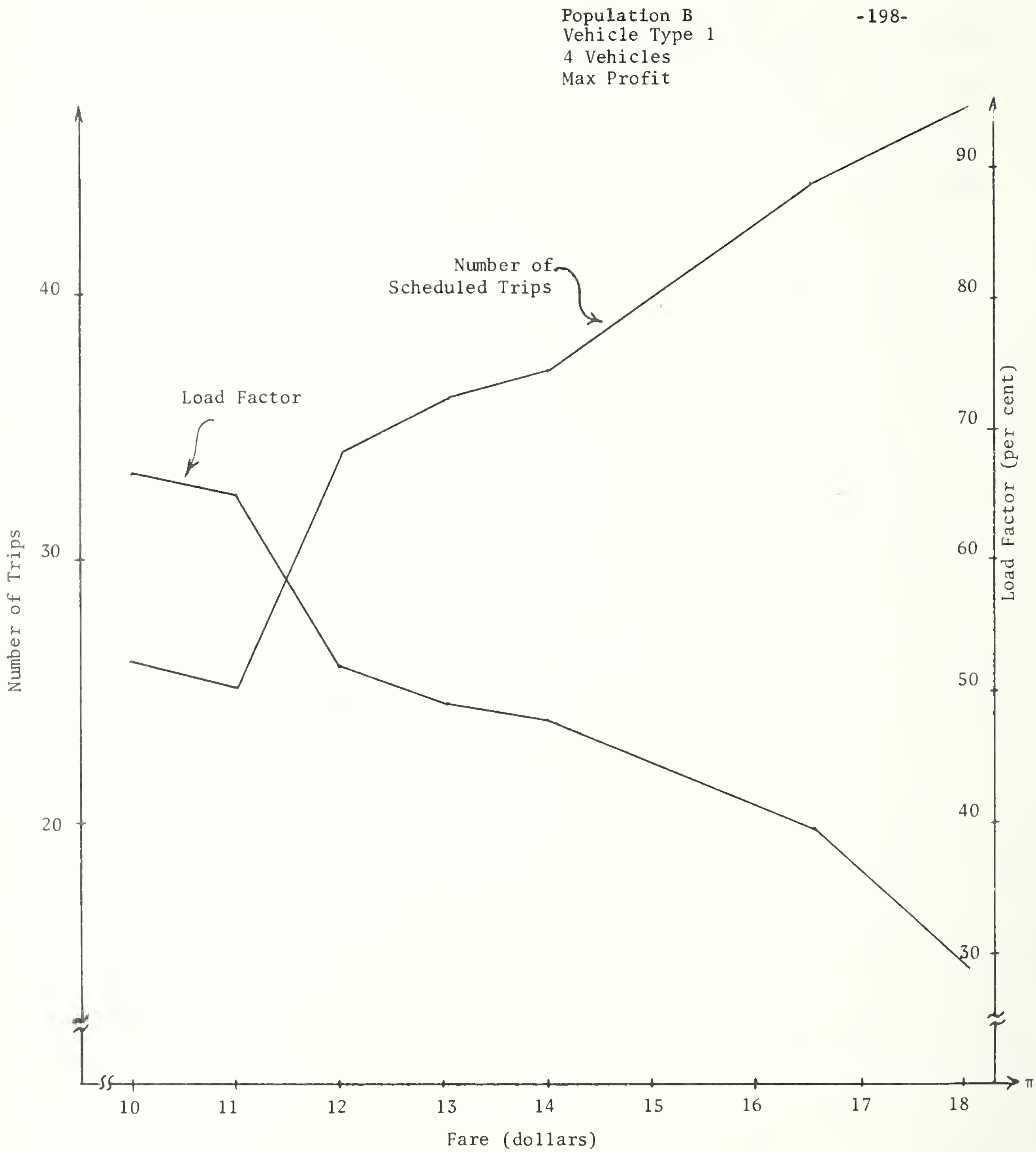


Figure 52. Variation of Schedule Properties for Profit Objective.

behavior of the curves representing the cost and number of scheduled trips, as fare is increased from ten to eleven dollars. In this case, profit is maximized by maintaining a relatively stable load factor, and cutting cost by executing a new timetable with one less trip. Local variations not withstanding, the general behavior in Fig. 52 is in clear contrast to the corresponding performance in the benefit-cost case, shown earlier in Fig. 44. A more detailed view of the timetable behavior under the profit criterion is displayed in Fig. 53. Here, the diminishing load factor, and the proliferation of trips in the timetable, are plainly in evidence.

The foregoing examples have illustrated a few of the interesting and important kinds of studies that the scheduling methodology can facilitate. The most important areas for which valuable information can be developed include the analysis of the behavior of system performance measures, and the discovery of the important mechanisms that account for system behavior under optimal scheduling. Within these two areas, many interesting questions may arise, in addition to those of the previous examples. Some possibilities for future investigation are suggested below.

5.5 Additional Studies

In addition to fleet size and fare, other system parameters offer opportunities for study. For example, the investigation of system behavior under objective functions other than benefit-cost or profit maximization may yield an interesting set of policy guidelines. In addition, parametric variations of vehicle speeds and capacities may aid in the design of vehicles to serve in particular system configurations. Experimentation with the distributions of traveler preference parameters will

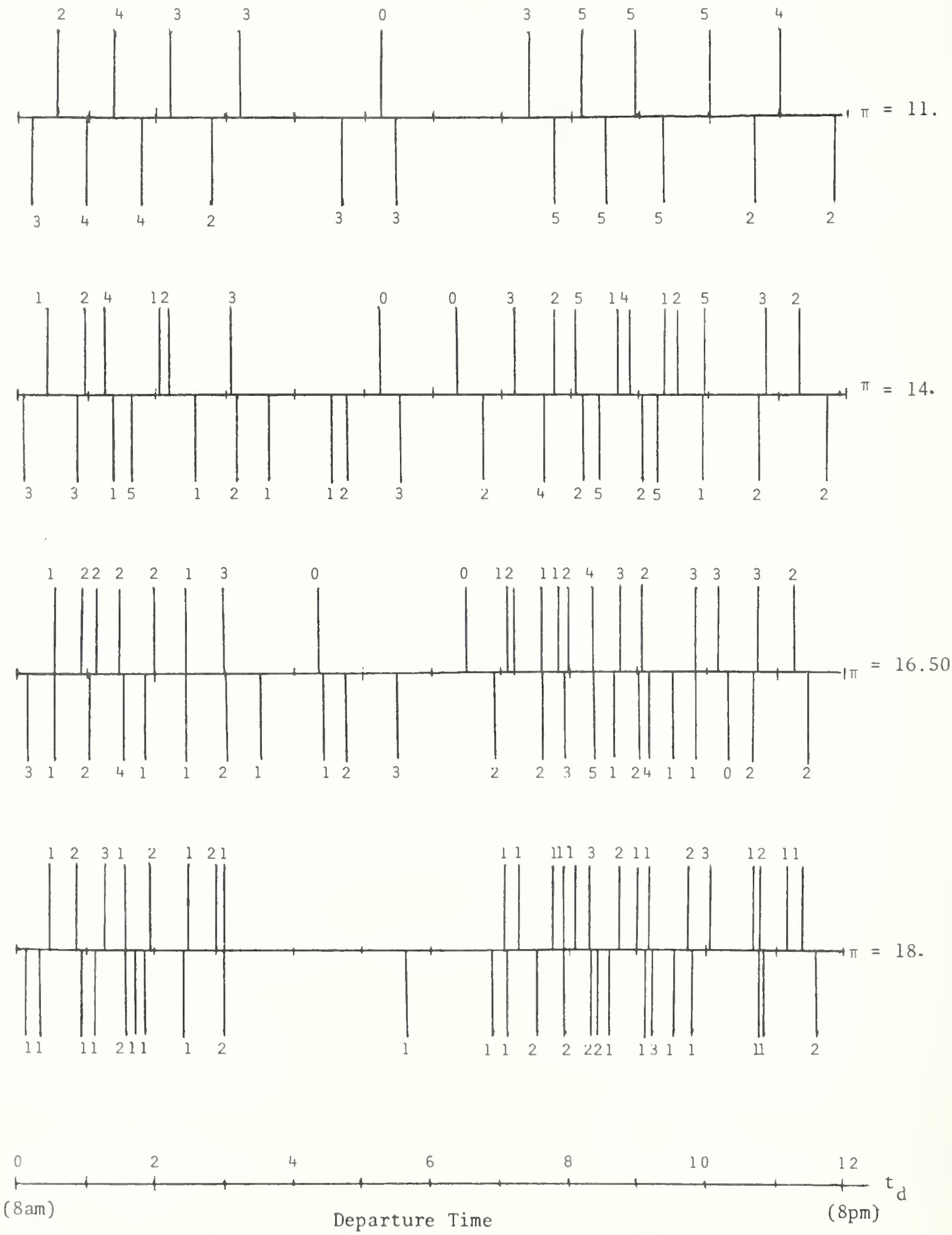


Figure 53. Variation of Timetable Under the Profit Objective.

yield important information regarding the sensitivity of system operation to traveler characteristics. It may be particularly rewarding to analyze more closely, the variations in system performance as traveler time preferences become increasingly "peaked" in their distribution. Perhaps such a study can contribute to the understanding, if not the solution, of the peaking problem of modern transit systems. In particular, the questions of fare policy and the selection among alternative vehicle fleets, may be especially subject to scheduling analysis.

Another important question is the effect of unbalanced demands between system nodes. Parametric variation of the relative levels of potential demand between alternate sets of nodes, may help indicate, for example, what kinds of vehicle systems operate best when demand on outgoing trips is heavy, but demand for incoming trips is light.

Finally, areas suggested in other chapters of this dissertation remain open for investigation. For example, Chapter VI extends the computational method to include routing and scheduling over networks. All of the parametric variations previously suggested, could be applied to determine the important effects on optimal routing behavior. Several areas mentioned in Chapter IV are prime candidates for computational experimentation. For example, it is of direct importance in computational investigations to determine the effect of system scale size and traveler population sample size on the accuracy and consistency of solutions resulting from alternate traveler populations generated from the same distribution. In performing a large set of computations, one would like to choose the smallest scale consistent with reasonable computational accuracy. Hence, a study of these effects is clearly worthwhile.

In addition, the utilization of successive sample traveler populations in the manner suggested in Chapter IV for the decision-making context, is an area of unexplored potential.

In summary, this chapter serves to indicate the variety of questions and analyses to which the scheduling methodology can usefully be applied. Examples have been presented to illustrate a few of the interesting areas and additional investigations have been suggested. The list of appropriate areas for interesting studies has by no means been exhausted, although the author and his resources are more nearly so.

CHAPTER VI

EXTENDING THE COMPUTATIONAL METHOD TO NETWORK OPERATIONS AND VARIABLE PRICING AND SERVICE OPTIONS

6.1 Introduction

The basic computational framework for scheduling a system of transportation vehicles was developed in Chapter III, in the context of a shuttle system. This chapter extends the methodology to more complex transportation networks. Accordingly, the scheduling solutions will involve the routing of vehicles, as well as the timetables and passenger assignments. In addition, other options such as local-express runs and pricing policies will be included in the computational framework.

The basic mode of development here is the expansion of the number of possible decisions at each stage of the single vehicle scheduling algorithm. The single vehicle algorithm will be used in the MODEL2 context for fleet scheduling, as before. In the dynamic programming formulation of Chapter III the decision variable for each vehicle in Eq. (3-1) takes on a value equal to the node number of the next station that the vehicle will visit. This formulation motivates the first extension of the basic single vehicle scheduling algorithm.

6.2 Non-Stop Algorithm

Consider a network consisting of a set of nodes (stations) such that all direct trips (from any node m to any node n) are allowed. Furthermore, assume that all travelers may make direct, non-stop trips only. An example of such a system is illustrated in Fig. 54. The new recursive scheduling equation becomes,

$$I(x,k) = \max_{x'} \left\{ b_{xx'}(k) + I(x',k+L_{xx'}+TR_{x'}) \right\} , \text{ all } x,\ 0 \le k \le K , \tag{6-1a}$$

and

$$L_{xx} = 1 - TR_x , \tag{6-1b}$$

where

x = network node at time k,

x' = next node in vehicle's itinerary,

$L_{xx'}$ = time required to make the trip from x to x' ,

$TR_{x'}$ = turnaround time at node x' ,

and

$$b_{xx'}(k) = -a_1 OC_{xx'} + \sum_{i \in J_{xx'}} (a_2(\pi-\pi_i') + a_3(WTP^i - WTP^{i\prime})) , \tag{6-2}$$

where

π = fare for the (current) trip leaving at k,

WTP^i = passenger i's willingness to pay for the trip leaving at k ,

π_i' = fare for the trip to which passenger i is previously assigned,

$WTP^{i\prime}$ = willingness to pay of passenger i for the trip to which he is previously assigned,

$J_{xx'}$ = the set of travelers going from x to x', for whom $WTP^i \ge \pi$, and for whom the value $(a_2\pi + a_3 WTP^i)$ of the current trip exceeds the value $(a_2\pi_i' + a_3 WTP^{i\prime})$ of the previously assigned trip.

Eq. (6-1b) covers the decision to delay departure from node x for at least one more time interval. In this case (6-1a) reduces to,

$$I(x,k) = b_{xx}(k) + I(x,k+1-TR_x+TR_x) = I(x,k+1) \quad . \tag{6-3}$$

The computation of $b_{xx'}(k)$ in Eq. (6-2) follows from the same boarding procedure developed in Chapter III. The double counting inequality (3-26) for the shuttle system, which precludes double counting of passengers by a particular vehicle, is easily extended to the more general network case.

The assumption that all travelers make direct, non-stop trips is clearly necessary. The algorithm of Eq. (6-1) contains no mechanism for "carrying" travelers through a network node. For example, in Fig. 54, suppose the vehicle makes a trip from node 1 to node 2, and then another trip from node 2 to node 3. Passengers wishing to travel from node 1 to node 3 will not be considered by the separate computations corresponding to the consecutive trips, 1 to 2 and 2 to 3. Of course, if a given passenger <u>prefers</u> to travel to node 3 from node 1, via node 2, then he may be represented as a series of two passengers--the first wishing to travel from 1 to 2, the second from 2 to 3. Two problems are inherent in this procedure, however. First, the procedure may lead to a schedule solution such that only one of the two "pseudo-travelers" is served. Hence, the "real" traveler is accommodated on only one leg of his journey. An equally bad result occurs if the two legs of the journey are scheduled such that the first trip arrives after the second trip has departed! These problems may be minimized by skillful specification of the preference parameters of the pseudo-travelers. Nevertheless, the procedure is intrinsically artificial.

A second problem area is that a passenger's route must be uniquely specified, a priori. In many cases this will be no problem, as there is often a "dominant" path between the two nodes of interest. In addition, route selection decisions, as discussed for the example of Fig. 54, can often be properly incorporated by following the "local-express" or "alternate path" formulations, developed later.

In summary, the algorithm of Eq. (6-1) is useful for determining the routing and scheduling of a single vehicle, in a system for which the assumption of direct, non-stop trips is satisfactory.

6.3 Multi-Node Links

In most transportation networks, some nodes are considerably more important than others. System operations are far more sensitive to routing and scheduling decisions made at Grand Central Station, for example, than at Bronxville, New York. Thus, it is useful to consider a scheduling algorithm in which vehicles are required to make decisions about their itinerary, only when they reach certain "decision nodes" in the network. The network of Fig. 55 has three decision nodes and thirteen minor stations. At each decision node, the vehicle decides which decision node to go to next. For example, if the vehicle is at node 2, it may decide to go to node 1 or node 3, or remain at node 2. The value of choosing a particular destination, say node 1, is found from the benefit afforded to all travelers along the 2-e-d-c-b-a-1 link, that will be aboard on some portion of the journey.

The intention in the following discussion is to extend the methodology to the multi-node link case by developing the appropriate assumptions and modifications of the basic algorithm of Eq. (6-1). First, an assumption must be made corresponding to the earlier non-stop requirement

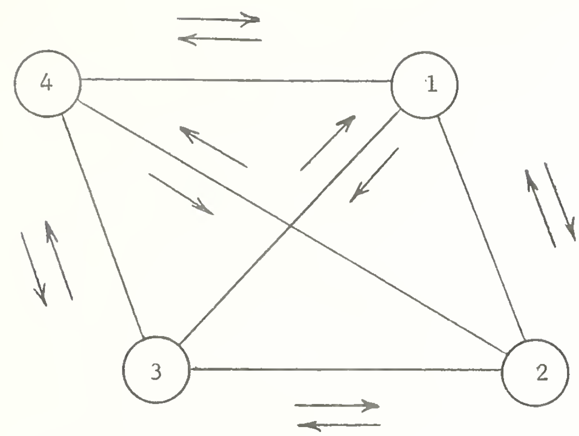


Figure 54. Non-Stop System.

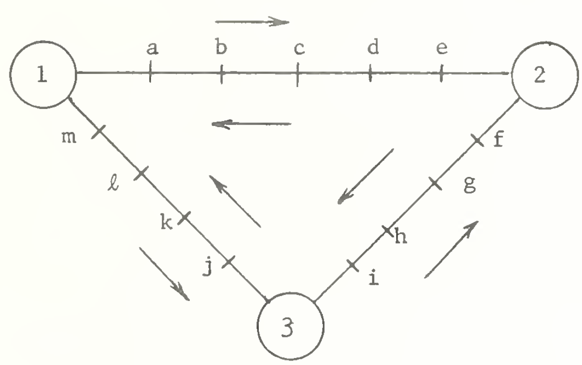


Figure 55. Multi-node Link System.

for passenger trips. Here, passengers will be restricted to direct trips between any two nodes on a single link of the network. (A link is defined as the chain of stops between any two decision nodes.) Thus, in Fig. 55, trips from node m to node j, or node 1 to node e, are acceptable, but a trip from j to h is not. Of course, an appropriate set of pseudo-travelers may be chosen to represent such multi-link passenger trips, although such an option involves the difficulties previously mentioned. In addition it is assumed that there is a unique path between any two decision nodes such that no intervening decision nodes lie in that path.

Two computational modifications are required before the basic algorithm can be applied. First, although dispatching decisions are made from decision nodes only, the benefits of travelers going to or leaving from intermediate stops must be computed using the local time coordinates of the trip. For example, consider a trip scheduled to leave node 1, in Fig. 55, at 10 a.m. and arrive at node 2 at noon. Suppose passenger P rides the segment from station b to station e. If the trip segment from 1 to b takes a half hour, and the trip from e to 2 takes fifteen minutes, then the appropriate trip coordinates with which to compute passenger P's benefit is $t_d = 10{:}30$, $t_a = 11{:}45$, or $\Delta = t_d^* - 10{:}30$, $\phi = 1\text{-}1/4$ hours.

The second modification involves the procedure for choosing the travelers whose aggregated benefits maximize $b_{xx'}(k)$ for a trip along a multi-stop link. Recall that for the two station shuttle, as well as the multi-node, non-stop trip system, the optimal rule is to order passengers by their net value for the trip. That rule is inadequate here because travelers, in general, ride different portions of a route segment between decision nodes. Hence, certain travelers are in competition for seats while others are not. Therefore, the problem reduces to that of finding the "best" mutually exclusive sets of travelers for each seat in the vehicle, such that each set consists of travelers riding non-overlapping route segments.

The problem is illustrated in Fig. 56. In this figure, the arcs represent traveler trajectories. The table displays the travelers' origins, destinations, and values for the trip leaving node 1 for node 2 at time k. Observe, for example, that the trips of travelers 2 and 6 are complementary, whereas travelers 4 and 5 are in competition for seats over route segment a-b.

The negative of the traveler values are attached to the corresponding arcs. Zero-valued arcs are included to represent empty seats. The graph representation suggests that the problem of optimally boarding passengers to the vehicle along the link is analogous to finding the "minimal cost flow" between a source node and a sink node in a graph. Here, the source node (1) is the decision node from which the trip emanates; the sink node (2) is the node where the trip terminates. Each traveler is represented by an edge connecting his origin node to his destination node. Each edge is labeled with a cost (-value) and a "flow capacity." A unit flow capacity is associated with each traveler edge. A flow capacity of C (the vehicle capacity) is associated with zero valued edges. A unit flow from source to sink is equivalent to a set of "connecting" travelers (including "null" travelers, i.e. empty seats) whose trip segments combine to make up a fully occupied seat between successive decision nodes.

Passenger	Origin	Destination	Value
1	1	2	1
2	1	b	6
3	1	a	2
4	a	2	6
5	a	b	5
6	b	2	2

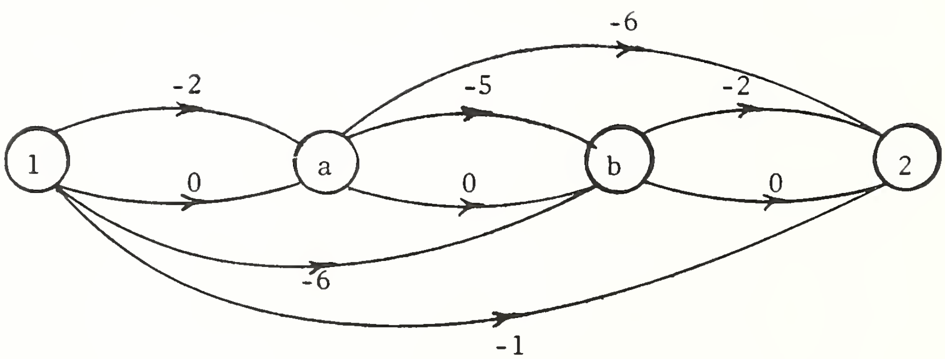


Figure 56. Network Flow Representation of Benefit Aggregation Along a Link.

Thus, the problem of optimal boarding is equivalent to the problem of finding that flow from source to sink of magnitude C, that minimizes cost (maximizes value). Algorithms to solve the min-cost flow problem are well known.* Fig. 57 illustrates the solution to the problem of Fig. 56, for a vehicle capacity C = 2.

Stipulation of the appropriate trip coordinates for each traveler along a link, and utilization of the min-cost flow algorithm for boarding travelers and computing traveler benefits, allows the algorithm of Eq. (6-1) to be used for the routing and scheduling of a vehicle among the decision nodes of a multi-node link network.

6.4 Locals and Expresses

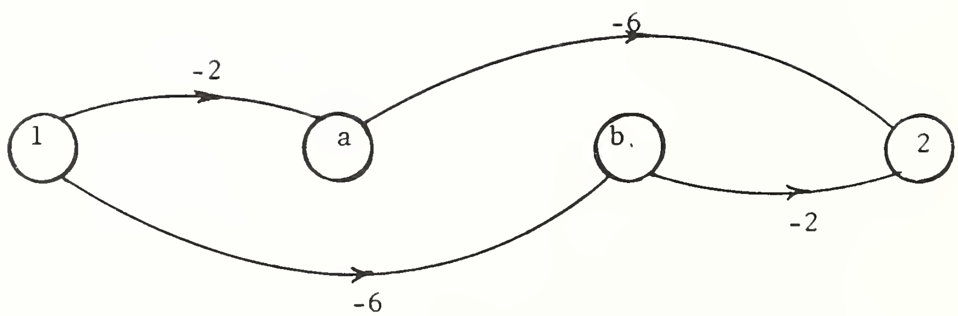
Consider a network with minor stops along the links connecting the decision points, as in Fig. 55. More flexible scheduling policies may be obtained by expanding the decision options to include not only the next destination decision node, but also the "service mode" by which the vehicle travels to that destination, i.e., local or express. If the local mode is chosen, the vehicle makes all minor stops along the link, as before. If express is picked, the vehicle travels non-stop from origin to destination decision nodes. Actually, there is no need to be limited to these two service options. Any additional mode may be considered, such as stopping at odd or even stations only. The new decision variables become,

$$c(x,k) = (x',u) \ , \tag{6-4}$$

where

* See Ref. [69] for example.

C = 2



Passengers: seat #1: 3 and 4,
seat #2: 2 and 6.

Total Value = (2 + 6) + (2 + 6) = 16 .

Figure 57. Min-Cost Flow Benefit Solution.

x' = next decision node,

u = service mode (1 = local, e = express).

The iterative equation becomes,

$$I(x,k) = \max_{x',u} \left\{ b^{u}_{xx'}(k) + I(x', k + L^{u}_{xx'} + TR_{x'}) \right\} . \tag{6-5}$$

The immediate benefit $b^{u}_{xx'}(k)$ is computed differently for alternate service modes u. If $u = e$ (express) then only the benefits of travelers at the origin node who are traveling directly to the destination node, are tallied. If $u = 1$, however, all benefits of travelers along the link are included. Presumably,

$$L^{e}_{xx'} < L^{1}_{xx'} , \tag{6-6}$$

since locals are required to stop at intermediate points along the link.

6.5 Alternate Paths

Suppose the network is set up so that there is more than one way to travel from one decision node to another. Two simple examples of this are shown in Fig. 58. The first example is a network with one decision node. Dispatches from that node (around the loop and back to itself) can go in either of two directions. The same is true for dispatches from the decision nodes of Fig. 58b. In such cases, the decision vector can be expanded to include path decisions, analogous to the service mode decisions of the previous section. If p is the path variable, then the values of $b^{up}_{xx'}(k)$ and $L^{up}_{xx'}$ are computed according to the traveler population and network characteristics for service mode u along path p.

6.6 Variable Fare

Another way to expand the decision options of the vehicle scheduling

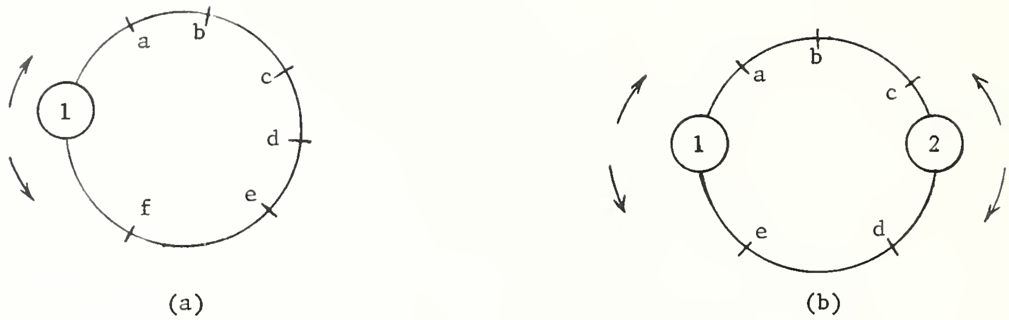


Figure 58. Networks with Alternate Paths.

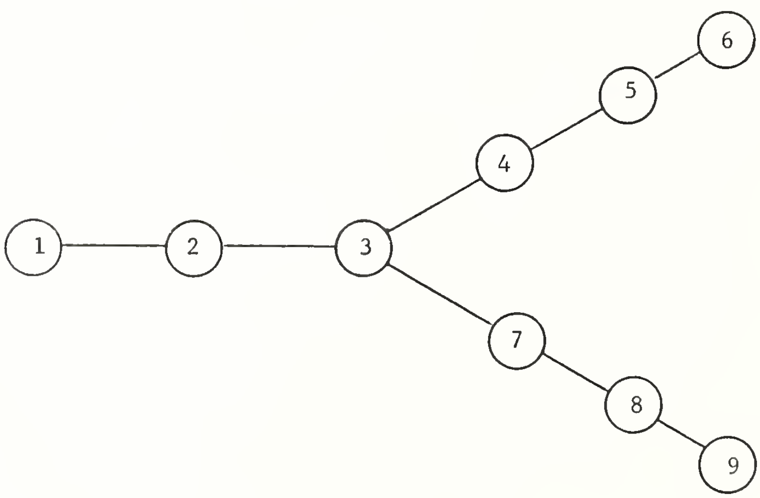


Figure 59. Choosing Decision Nodes.

algorithm is to make the fare for each trip a decision variable. Although fare is not usually considered a scheduling variable, it _is_, in fact, an important determinant of system operation. For example, a fare policy that discourages travelers from riding during hours of very high demand, and encourages them to travel during relatively light periods, will foster more efficient operation.

Modification of the algorithm follows the lines developed in the previous sections. Suppose the fare for each trip is allowed to take on one of F possible values. The basic iterative equation (neglecting service or path options) is,

$$I(x,k) = \max_{x',\pi} \left\{ b^{\pi}_{xx'}(k) + I(x',k+L_{xx'}+TR_{x'}) \right\} \quad , \pi_1 \leq \pi \leq \pi_F \, . \qquad (6\text{-}7)$$

The immediate benefit $b^{\pi}_{xx'}(k)$ is computed by summing (via the appropriate boarding procedure) over those passengers whose willingness to pay exceeds the particular value of fare, π.

Eq. (6-7) must be used with some care, with respect to the double-counting constraint embodied in Eq. (3-26). In particular, a risk of error is run if the fare is allowed to range so low that it becomes possible for a traveler to find two trips acceptable by the same vehicle.

In summary, the algorithm of Eq. (6-7) is designed to study the effectiveness of fare policy in dealing with the problem of "peaked demand," and the consequent uneven utilization of system facilities over time. More generally, the algorithm enables optimization over an important additional scheduling factor, the fare policy.

6.7 Trade-Offs

Setting up the single vehicle algorithm for a particular network arrangement is probably more of an art than a science. The basic questions pertain to the number and placement of decision nodes. In resolving these questions, the goal is to maintain reasonable computational requirements and to limit errors resulting from inadequate representation of multi-link traveler trips without unduly compromising the width of the spectrum of possible scheduling solutions.

The location of decision nodes should harmonize with the character of actual traveler trips, and with the natural structure of the network. For example, it is often appropriate to give decision node status to nodes where many travelers begin or end their trips, and to nodes at which natural route decisions must be made (i.e., a fork in the road). Consider the network of Fig. 59, for example (see p. 214).

If most travel is from the extreme nodes 1, 6, and 9, into and out of the hub node 3, then nodes 1, 3, 6, and 9 may be good choices for decision nodes. However, if trips to and from nodes 1 and 6, 1 and 9, and 9 and 6, are dominant then node 3 should not be a decision node, because travelers along these routes would then have to be represented by series of pseudo-travelers.

Choosing the number of decision nodes requires additional trade-offs. The more decision nodes there are, the greater flexibility the algorithm will exhibit in computing the schedule. For example, in Fig. 59, if node 3 were not a decision node then a vehicle reaching this node would be <u>required</u> to pass through, pausing for the minimal stopping time only. However, increasing the number of decision nodes aggravates two problems. First, the more decision nodes there are, the more travelers there will be who must be represented as series of pseudo-travelers. Hence, there is a greater risk of error.

Second, increasing the number of decision nodes increases the computational requirements, in two ways. First, more decision nodes means more decisions, and thus longer computation time. For example, if there are X decision nodes, and an average of $Y = qX$ $(0 \leq q \leq 1)$ decision nodes accessible from each of the X nodes, then there are approximately $XY = qX^2$ decision computations to be made at each stage k in the algorithm. Thus, the computation time is proportional to the square of the number of decision nodes. The factor of proportionality increases with the accessibility of one node from the others, and also with the number of additional decision options such as local-express, fare, and alternate paths. If there is an average of r options between decision nodes, then the computation time is proportional to qrX^2.

Second, the more decision nodes and decision options, the greater is the computer memory requirement. This stems from the fact that at each stage k, the result of previous computations, $L^{up}_{xx'}$ stages back, must be recalled. The greater the number of possible values of $L^{up}_{xx'}$, the more stages of computation must be stored.

It is apparent that there are several limitations and compromises associated with the algorithms of this chapter. A limitation not mentioned heretofore, is the absence of a traveler transfer option. For example, in Fig. 59, if node 3 is not a decision node, a traveler going from node 6 to node 9 does not have the option of riding a node 1-bound vehicle to node 3, and then a node 1-originating vehicle to node 9. As in this instance, most of the difficulties revolve about the representational accuracy of individual traveler trips. The real concern, however, is to achieve sufficient individual representational accuracy to ensure adequate modeling of travel behavior in the aggregate. As outlined in Chapter IV, the traveler population is to be drawn randomly from a probability distribution over traveler preference parameters values. Thus, the aggregate behavior of the traveler population may be properly represented by developing the appropriate distributions of pseudo-travelers, rather than "true" travelers. The test of representational accuracy is then, not whether individual traveler trips closely reflect real trips, but whether the statistical trip behavior of the population is correct.

In summary, the compromises involved in setting up the single vehicle algorithm for various situations, have been outlined above. Much is left to the engineering judgement of the analyst for each particular case. Below, incorporation of the single vehicle algorithm into the overall iteration, is discussed.

6.8 Applying MODEL2

In the preceding sections, the capabilities of the single vehicle algorithm have been greatly expanded. The next task is to adapt the iterative procedure for scheduling the entire vehicle fleet system. Scrutinizing the MODEL2 flow diagram of Fig. 26 reveals that except for the optimum assignment algorithm, MODEL2 is directly applicable to the more complex transportation systems discussed in this chapter.* The

* Use of MODEL2 without optimum assignment precludes initialization by timetables, unless computations are started by assigning travelers to trips in the initial timetable in a heuristic, suboptimal way.

iterative procedure for successively computing the itineraries of individual vehicles and adjusting the passenger assignments, based on net incremental benefits, needs no fundamental revision. The solutions obtained will be optimal to within a first-order traveler transfer, as before.

Moreover, it was found in Chapter III, that application of the optimum assignment routine did not significantly alter the results. The same behavior can be expected here. This is fortunate because, in general, the equivalent optimal assignment operation over a complex network, is a difficult integer programming problem. Only the "non-stop" case, with links having no minor stops, is directly subject to application by the assignment algorithm. In other cases, the problem is that of finding a matching of mutually exclusive "feasible" groups of travelers, to inter-decision node trips. A feasible group is composed of travelers with non-overlapping trip segments that combine to make a complete trip between the decision nodes. (Empty seats count as null travelers.) Travelers 2 and 6 in Fig. 56, for example, comprise such a group. So does traveler 4 plus an empty seat from node 1 to node a. The task here is to find the matching of traveler groups to trips such that the total value is maximized, and such that no single traveler appears in more than one of the assigned groups. This is equivalent to solving and comparing solutions to Q different assignment problems, where Q is the number of ways that the traveler population along the link can be separated into mutually exclusive feasible groups. Integer programs are notoriously difficult to solve, and this one promises to be no different. The formulation will not be carried further here. The prospective gains of developing such an assignment algorithm do not appear to warrant the effort. Nevertheless, it is an interesting combinatorial problem in its own right.

6.9 Summary

This chapter has extended the domain of the computational method developed in Chapter III. The single vehicle algorithm was expanded to include a wide spectrum of decision options for scheduling transportation network operations. The MODEL2 iterative method was found applicable for use in this wider context.

The algorithms of this chapter have not yet been programmed for a computer. Such implementation, however, holds out great potential for the study and development of scheduling, routing, and fare policies, for a variety of transportation network operations.

CHAPTER VII

CONCLUDING DISCUSSION

This chapter considers the following four areas of interest to round out the discussion in this dissertation: 1) application of the scheduling algorithm outside the context of passenger transportation, 2) the "systems engineering" concepts that have been used to mold the ideas presented here, 3) the specific contributions that this research has offered, and 4) suggested areas for future research in passenger transportation scheduling.

7.1 General Application of the Scheduling Algorithm

The MODEL2 multi-vehicle scheduling algorithm has been developed specifically to fit the context of passenger transportation. This does not preclude the possibility, however, that the method has application in other scheduling contexts. Abstracting from the transportation situation, the method may be viewed as a "facility scheduling - user allocation" algorithm. In that perspective, various applications may be envisioned. Fig. 60 displays some examples for consideration.

In attempting to adapt MODEL2 to new applications, several relevant factors must be recognized. First, it is clear that the preference functions over alternate schedules need not assume the forms proposed in Chapter II for the transportation context. The preference formulation will depend on the particular scheduling application. Second, in order to apply the MODEL2 method an inequality to prevent double counting, analogous to relationship (3-27), must be observed. In brief, that

Facility	Vehicle	Capacity	Duration of Use	Schedule	User	Preferences	Comment
1. Freight Transport	Ship or Truck	Units of Cargo Space	Trip Time	Depart. & Arrival Times	Cargo	Prefer short trips, short depart. delay, arrival deadline.	
2. Communi-cation unit	Channel	One	Message Length	Start & End Times	Person or Group	Prefer long messages, at specified time of day.	Require option to choose message length. Double-counting in-equality may be un-realistic.
3. Restau-rant	Table	Seats/ Table	Meal duration	Mealtime	Groups of Diners	Pref. meal time specif. by diners	Must assume fixed meal duration per table.*
4. Movies	Theatre	Seating Room	Film Length	Show Times	Patrons	Specified by users	
5. Auto As-sembly Plant	Assembly Line	One	Optional	Entry & Exit Times on Line	Car Model, e.g. Falcon Mustang	Specified by sales department accord-ing to consumer demand and inventory requirements	Analogous to other prod. situations: -Manuf. of appliances -Bottling of wines -Packaging of canned goods

* A function of the assigned waiter ?

Figure 60. Alternate Applications of the Scheduling Algorithm.

relationship requires that the overall minimum turnaround time between successive uses of the facility, must exceed the width of the acceptable range of schedule (departure or arrival) times for any given user. In certain applications, such as the second one in Fig. 60, this may be an unreasonable restriction. Finally, MODEL2 is designed under the assumption that the duration of use of the facility (trip length) is a known, fixed number. In some applications this may not hold. For example, in case 3 of Fig. 60 such an assumption may not always be justifiable. In other cases, such as numbers 2 and 5 of Fig. 60, the duration of use may be a matter of choice. In such instances it is possible to modify the single vehicle algorithm, following the pattern of Chapter VI, to allow the option of specifying the desired duration of use. In summary, the foregoing discussion is offered as food for thought. Closer study and refinement will be required to evaluate the utility of the algorithm for various specific scheduling situations.

7.2 Systems Concepts

The underlying theme of this paper has been to apply the methodology of systems engineering to the problem of passenger transportation scheduling. The purpose of the following discussion is to summarize the systems concepts of this analysis in the belief that explicit recognition of these ideas may be helpful to future efforts in transportation research.

The first important systems concept is the clear formulation of a meaningful objective function with which to gauge system operation. Frequently, the problem of timetable construction is viewed merely as the task of "fitting" timetables to travel demands, without relating the schedule to the enhancement of a relevant index of system performance.

Examples in Chapter V have demonstrated the significance of optimizing in accordance with the correct measure of system operation.

In conjunction with the recognition of objectives, the utilization of meaningful economic measures to quantify an objective function is a second important systems requirement. In this vein, modeling of the traveler's willingness to pay provides the required measure of consumer benefit. In addition, integration of the three basic component measures--costs, revenues, and traveler benefits--facilitates the quantification of an array of relevant economic criteria such as profit and benefit-cost.

A third set of systems engineering ideas revolves about the formulation of the computational framework for obtaining scheduling solutions. A number of principles were followed to facilitate the synthesis of a viable algorithm. Owing to the large number of system variables, a primary guideline was to describe the "state" of the system in as economical a fashion as possible. This principle was used in several ways. First, with respect to the single vehicle scheduling algorithm, the derivation in Chapter III, Eqs. (3-7) through (3-17), produced an iteration that dispensed with the consideration of all superfluous vehicle states. Second, direct accounting of the state of the traveler population was suppressed in favor of a passenger to vehicle categorization procedure, in conjunction with an inequality relationship (3-26) to prevent double counting within the single vehicle optimizations.

The size of the overall optimization problem required the use of another computational principle--decomposition. Here, decomposition was implemented in two ways. Vehicle schedules were optimized individually, and passenger allocation was performed separately from timetable optimization. However, closer integration of the vehicle optimization and passenger allocation processes was a prime factor in the development of MODEL2 from MODEL1.

A final computational principle, monotonic improvement by iteration, was used to formulate the comprehensive framework for synthesizing scheduling solutions. The nature of such iterative methods in complex problems does not ensure optimal results. As discussed in Chapter III, however, this methodology does provide an approach to a wide range of complex optimizations.

In addition to the formulation of objectives, the quantification of benefit measures, and the observance of computational principles, another important systems engineering idea is the decision analysis framework for decision making under uncertainty; this was discussed in Chapter IV. In that chapter, the technique for system schedule optimization was woven into the context of the decision to choose among alternative proposed systems in the face of uncertainty about traveler population characteristics. The principal modes of analysis involved Monte Carlo techniques and the Bayesian theory of inference.

7.3 Contributions

The purpose of this section is to pin-point some of the specific products of this research within the context of the systems framework studied here. These contributions fall into three areas--modeling of economic traveler behavior, synthesis of an optimization algorithm, and relating the significance of scheduling to decision making.

Concerning the traveler modeling, Chapter II has developed an economic consumer representation of the traveler in terms of the "commodities," trip duration and schedule inconvenience. The concept of schedule inconvenience was formulated in terms of a traveler's preferred schedule. The preferred schedule was given as a function of the trip duration and the traveler's relative departure time-arrival time orientation. Thus, a <u>new structural representation</u> has been developed for quantifying the attributes of a trip's schedule that affect the traveler's behavior.

The preference ordering and relative levels of economic benefit that a traveler may receive from alternate trips have been encoded in terms of a willingness to pay function over the trip schedule attributes. Alternate assumptions with respect to the values of the time preference and willingness to pay parameters allow study of the system-wide effects of different types of traveler populations (markets). In sum, the economic traveler model, in the context of the comprehensive objective function formulation, allows <u>explicit</u> study of the importance of operator <u>and</u> traveler behavior to overall system operation.

With regard to the difficult computational problem of optimizing a system of vehicles with respect to an arbitrary traveler population, this thesis has contributed what has been termed a <u>first-order optimal</u> method. The solution is based on the repeated use of an optimal single-vehicle algorithm developed in Chapter III for a shuttle service and extended to more complex cases in Chapter VI.

Finally, the present study contributes to the field of transportation planning by stressing the importance of scheduling in the decision to choose among alternative systems. To that end, Chapter IV has developed a framework for utilizing the scheduling optimization method in a decision making context that accounts for the random nature of the traveler population. In addition, the examples of Chapter V illustrate some of the numerous ways in which the scheduling methodology may be used for parametric studies of the many interesting and important questions related to the consideration of transportation alternatives.

7.4 Future Research

In the course of this dissertation several areas have emerged as worthy candidates for continued research. These areas divide into three basic categories--data analysis, computational experimentation, and theoretical development.

As discussed in the second and fourth chapters, there are potential benefits to be reaped from two modes of data acquisition and processing. The first, laboratory interview experimentation, may build on the groundwork laid by earlier research efforts. The second, observation of "real world" traveler behavior, would be enhanced by development of controlled experimental transportation projects. Data acquired by either method may be processed in a Bayesian manner, as suggested in Chapter IV. The development of the appropriate functional forms for such analysis must be an important parallel endeavor. The developments in this thesis clearly suggest the utility of such efforts in data analysis.

With respect to computational experimentation, the scheduling methodology developed here will be useful for investigating a wide array of interesting questions. Some of these questions arose in Chapter V, such as the investigation of system operation for alternate traveler markets, fare policies, vehicle types, fleet sizes, and operator objectives. In addition, the extended algorithms of Chapter VI have not yet been programmed for a computer; such implementation will provide wide opportunity for the study of scheduling and routing over networks, including the investigation of local-express and differential pricing policies.

Finally, the theoretical developments of this dissertation suggest avenues for continued study and refinement. In particular, the following tasks may prove worthy of future effort: a) exploratory research, as suggested in Chapter III, to improve upon the first-order schedule optimization algorithm (MODEL2), b) further analysis to seek the derivation of schedules under uncertainty about traveler characteristics, with less restrictive assumptions than those of Chapter IV, and c) refinement of the ideas of Chapter VI to develop efficient scheduling algorithms for very complex transportation network structures.

In summary, the principal contribution of this research is the framework that it has provided for both the establishment of operational scheduling policies and the consideration of alternative systems within the transportation planning process. Hopefully, the methodology will find use in future transportation system analyses.

REFERENCES

1. Northeast Corridor Transportation Project, Technical Paper No. 5, Transport Systems Planning Division, Office of High Speed Ground Transportation, U.S. Department of Commerce, July, 1966.

2. "High Priority Research Tasks for High Speed Ground Transport," Part II, prepared for the U.S. Department of Commerce, Massachusetts Institute of Technology, Cambridge, Massachusetts, June, 1965.

3. E.F. Bisbee, "Dynamic Programming and Passenger Transportation Scheduling," Department of Civil Engineering, Massachusetts Institute of Technology, February, 1966.

4. E.F. Bisbee and S. Kuroda, "Study on the Dynamic Dispatching Policy and its Application in a Stochastic Model for Vehicle Scheduling," Department of Civil Engineering, Massachusetts Institute of Technology, July, 1967.

5. R.F. Meyer and H.B. Wolfe, "The Organization and Operation of a Taxi Fleet," Naval Research Logistics Quarterly (U.S.), Vol. 8, No. 2, June 1961, pp. 137-150.

6. T.L. Croswell, "Simulation of Airline Operation," 1967 Transportation Engineering Conference, New York, New York, August, 1967.

7. J.D. Kingsley, "Airline Operations and Planning Model," Conference of the International Federation of Operational Research Societies, Massachusetts Institute of Technology, Cambridge, Massachusetts, August, 1966.

8. L.R. Howard and D.O. Eberhardt, "Airline Simulation for Analysis of Commercial Airplane Markets," Transportation Sceince, Vol. 1, No. 3, August, 1967.

9. R. MacDonald-Taylor, "Scheduling Vehicles in a Transportation Network," Technical Memorandum No. 76, Operations Research Group, Case Institute of Technology, January, 1967.

10. J.P. Carstens, R.C. Baxter, and J. Reitman, "Economic Models for Rail Systems," I.E.E.E. Transactions on Systems Science and Cybernetics, Vol. SSC-1, No. 2, December, 1966.

11. T. Sugiyama, "Vehicle Size Scheduling with Consideration of Passenger Comfort," Japanese National Railways, O.R. No. 50, 1964.

12. R.W. Simpson, "Computerized Schedule Construction for an Airline Transportation System," Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, November, 1966.

13. W. Lampkin and P.D. Saalmans, "The Design of Routes, Service Frequencies, and Schedules for a Municipal Bus Undertaking: A Case Study," Operational Research Quarterly, Vol. 18, No. 4, 1967.

14. W.A. Gunn, "Airline System Simulation," Operations Research, 12, 206, 1964.

15. R.E. Larson, "A Dynamic Programming Approach to Airline Scheduling," Proceedings of the Fifth AGIFORS Symposium, Chicago, Illinois, October, 1965.

16. R.E. Larson, State Increment Dynamic Programming, American Elsevier Publishing Company, Inc., New York, New York, 1967, Chapter 12.

17. R.E. Miller, Domestic Airline Efficiency, An Application of Linear Programming, The M.I.T. Press, Cambridge, Massachusetts, 1963.

18. "World-Wide Airline Simulation Model," Boeing Corporation, Renton, Washington, 1967.

19. T. Kushige, "A Solution of Most Profitable Aircraft Routing," Proceedings of the Third AGIFORS Symposium, Val Morin, Quebec, Canada, October, 1963.

20. J. Taylor, "A Common Approach to Aircraft and Aircrew Scheduling," Proceedings of the Sixth AGIFORS Symposium, Killarney, Ireland, September, 1966.

21. G. Gagnon, "A Model for Flowing Passengers Over Airline Networks," Transportation Science, Vol. 1, No. 3, August 1967.

22. R.W. Simpson and M. Neuve-Eglise, "Determining Vehicle Size and Frequency of Service for a Short Haul Air Transport System," Flight Transportation Laboratory, Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1968.

23. P.I. Welding, "The Instability of Close Interval Service," Operational Research Quarterly, Vol. 8, No. 3, September, 1967.

24. J. Heap, "Time of Departure Assignment in Aircraft Scheduling," Douglas Aircraft Company, Long Beach, California, 1968.

25. A.J. Goldman, Internal Memorandum, National Bureau of Standards, April, 1968.

26. J.D. Foulkes, W. Prager, and W.H. Warner, "On Bus Schedules," Management Science, Vol. 1, 1954.

27. J.W. Devanney, "Passenger Transportation Scheduling," M.S. Thesis, Department of Naval Architecture, Massachusetts Institute of Technology, August, 1965.

28. D. E. Ward, "Optimal Dispatching Policies by Dynamic Programming," M.S. Thesis, Department of Civil Engineering, Massachusetts Institute of Technology, November, 1966.

29. D.R.Young, "Scheduling a Vehicle Between an Origin and a Destination to Maximize Traveler Satisfaction," Proceedings of the National Conference of the Association for Computing Machinery, August, 1967.

30. "Calculating Train Path Timings by Computer in the Eastern Region," The Railway Gazette, Vol. 118, No. 6, February, 1963, pp. 156-57.

31. R.M. Newton, and W.H. Thomas, "Design of School Bus Routes by Digital Computer," Department of Industrial Engineering, State University of New York at Buffalo, New York, October, 1966.

32. E.F. Bisbee and J.W. Devanney," Transportation Scheduling Under Multi-Dimensional Demands," Civil Engineering Department, Massachusetts Institute of Technology, Cambridge, Massachusetts, July, 1967.

33. M. Beckman, C.B. McGuire, and C.B. Winsten, Studies in the Economics of Transportation, Yale University Press, 1956, Chapter 11.

34. W. Hyman, and L. Gordon, "Optimal Airline Scheduling," Lockheed-California Company, Burbank, California, 1967.

35. W. Hyman, and L. Gordon, "Commercial Airline Scheduling Technique," Transportation Research, Vol. 2, No. 1, March, 1968.

36. J.W. Devanney, "Transportation Scheduling Under Multi-Dimensional Demands," Ph.D. Thesis, Department of Naval Architecture, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1968.

37. D.G. Haney, "The Value of Time for Passengers Cars: A Theoretical Analysis and Description of Preliminary Experiments," Stanford Research Institute, Menlo Park, California, May, 1967.

38. P.A. Samuelson, Foundations of Economic Analysis, Atheneum,New York, 1965.

39. S.A. Marglin, Public Investment Criteria, M.I.T. Press, Cambridge, Massachusetts, 1967.

40. "The Value of Time for Passenger Cars: Further Theory and Small-Scale Behavioral Studies," Stanford Research Institute, Menlo Park, California, September, 1964.

41. J.F. Mahoney, "A Survey to Determine Factors Which Influence the Public's Choice of Mode of Transportation," Joseph Napolitan Associates, Inc., Boston, Massachusetts, April, 1964.

42. J.B. Lansing, E. Mueller, and N. Barth, "Residential Location and Urban Mobility," Survey Research Center, Institute for Social Research, University of Michigan, June, 1964.

43. M.L. Manheim, "Principles of Transport Systems Analysis" (unpublished), 46th Annual Meeting of the Highway Research Board, Washington, D.C., January, 1967.

44. S.J. Hille, and T.K. Martin, "Consumer Preference in Transportation," Highway Research RECORD, No. 197, Washington, D.C., 1967.

45. M. Wohl, "Another View of Transport System Analysis," Proceedings of the I.E.E.E., Vol. 56, No. 4, April, 1968.

46. "National Survey of Transportation Attitudes and Behavior," Phase I Summary Report, Highway Research Board, Washington, D.C., 1968.

47. "The Metro Flyer," Final Report 1967, Metropolitan Transit Authority of Maryland, 1967.

48. N. Cherniack, "Effects of Travel Impedance Costs," Highway Research Board Special Report 56--Economic Analysis in Highway Programming, Location, and Design, September, 1959.

49. J.M. Henderson and R.E. Quandt, Microeconomic Theory, McGraw-Hill, New York, 1958.

50. "World-Wide Airline Simulation Model," Boeing Company, Renton, Washington, 1967.

51. A.S. Lang, and R.M. Soberman, Urban Rail Transit: Its Economics and Technology, M.I.T. Press, Cambridge, Massachusetts, 1964.

52. L.L. Thurstone, "The Indifference Function," Journal of Social Psychology, Vol. 2, 1931, pp. 139-67.

53. S.W. Rousseas and A.G. Hart, "Experimental Verification of a Composite Indifference Map," Journal of Political Economy, Vol. 59, 1951, pp. 288-318.

54. K.R. MacCrimmon and M. Toda, "The Experimental Determination of Indifference Curves," Working Paper No. 124, Western Management Science Institute, University of California at Los Angeles, February, 1968.

55. W.A. Wallis, and M. Friedman, "The Empirical Derivation of Indifference Functions," Studies in Mathematical Economics and Econometrics, University of Chicago Press, 1942.

56. G.B. Dantzig, Linear Programming and Extensions, Princeton University Press, Princeton, New Jersey, 1963.

57. L.S. Lasdon, "Duality and Decomposition in Mathematical Programming," I.E.E.E. Transactions on Systems Science and Cybernetics, Vol. SSC-4, No. 2, July, 1968.

58. D.J. Wilde and C.S. Beightler, Foundations of Optimization, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1967.

59. G. Nemhauser, Introduction to Dynamic Programming, John Wiley and Sons, Inc., New York, 1966, Chapter 4.

60. R.A. Howard, Dynamic Programming and Markov Processes, M.I.T. Press, Cambridge, Massachusetts, 1960.

61. J. Edmonds and D.R. Young, "General Matching Algorithm in a Bipartite Graph," Lecture Notes (unpublished), National Bureau of Standards, Washington, D.C., Spring, 1967.

62. J.M. Hammersley and D.C. Handscomb, Monte Carlo Methods, Menthuen and Company, Ltd., London, 1964.

63. C.W. Dunnet, "On Selecting the Largest of k Normal Population Means," Journal of the Royal Statistical Society, Series B, 1960.

64. H. Raiffa and R. Schlaifer, Applied Statistical Decision Theory, Harvard University, Boston, Massachusetts, 1961.

65. E. Parzen, Modern Probability Theory and Its Applications, John Wiley and Sons, Inc., New York, 1960.

66. K.L. Chung, A Course in Probability Theory, Van Nostrand, New York, 1955.

67. E.A. Silver, "Markovian Decision Processes with Uncertain Transition Probabilities or Rewards," Technical Report No. 1, Contract Nonr-1841(87), Operations Research Center, Massachusetts Institute of Technology, August, 1963, Appendix B.

68. New York Airways Inc., Direct Exhibits Before the Civil Aeronautics Board for the Washington/Baltimore Helicopter Service Investigation, Docket No. 17665, Washington,D.C., 1967, Exhibit 309.

69. L.R. Ford and D.R. Fulkerson, Flows in Networks, Princeton University Press, Princeton, New Jersey, 1962.

APPENDIX A

INTERVIEW PROCEDURE FOR DETERMINING WILLINGNESS TO PAY CONTOURS

The following procedure is designed to elicit the information necessary to construct the function describing a traveler's willingness to pay for alternatively scheduled trips. The subject of the experiment should be a person that is taking or intending to take a specific trip. The willingness to pay function is to be determined for the particular trip to which the subject is committed. In the sample results given below, the subjects were persons who had made or were about to make reservations for air trips.

The steps in the interview procedure are the following:

1. Pick a value of trip duration ϕ, and have the subject choose his preferred schedule (t_d^*, t_a^*) for a trip ϕ hours long. This trip will be used as the reference trip (R) in the following steps. Note that R is a point on the locus of the traveler's preferred schedules.

2. Hypothesize that R is the only trip available by the current mode of travel (air, rail, etc.). The only alternatives are to travel by another mode or forego making the trip. Set the price for trip R at a very high level relative to existing prices, and ask the subject if he is willing to take trip R for that price. Drop the price by small increments until the subject becomes willing to pay the price. Note the level of willingness to pay. This is the value of willingness to pay associated with the indifference curve derived in the next step.

3. Set up a data sheet consisting of a grid imposed on the (t_d, t_a) plane, as in Fig. A1. Mark R in the box corresponding to the reference trip. Choose a feasible trip (t_d, t_a) at random from the area surrounding the reference trip. Ask the subject if he prefers (0) or does not prefer (X) this trip to the reference trip R. Mark the box (t_d, t_a) with an X or 0, accordingly. Continue this procedure for additional trips in the vicinity of R, until the border separating the region of 0's from the region of X's is clearly delineated. This border outlines the indifference contour passing through R and corresponding to the willingness to pay level derived in step 2.

4. Repeat steps 1, 2, and 3 for several values of ϕ, i.e., several different reference trips R.

Fig. A1 shows a sample data sheet obtained by one pass through the above procedure. The dash (-) entries denote points on the indifference curve. Figs. A2 through A7 illustrate the experimental results of interviewing three different air travelers. Figs. A2, A4, and A6 show the willingness to pay curves in the (t_d, t_a) plane, along with piecewise linear approximations to the preferred schedule loci. Using these linear preferred schedule loci, the experimental points (-) were translated into the (Δ, ϕ) plane, yielding the transformed curves of Figs. A3, A5, and A7.

Too few test cases have been made to allow any general conclusions. However, the three cases shown here are sufficient to illustrate some apparent trends. First, the shapes of the willingness to pay contours, as drawn through the experimental points in Figs. A2 through A7, behave as expected, i.e., they have the proper convexity property of Eq. (2-13).

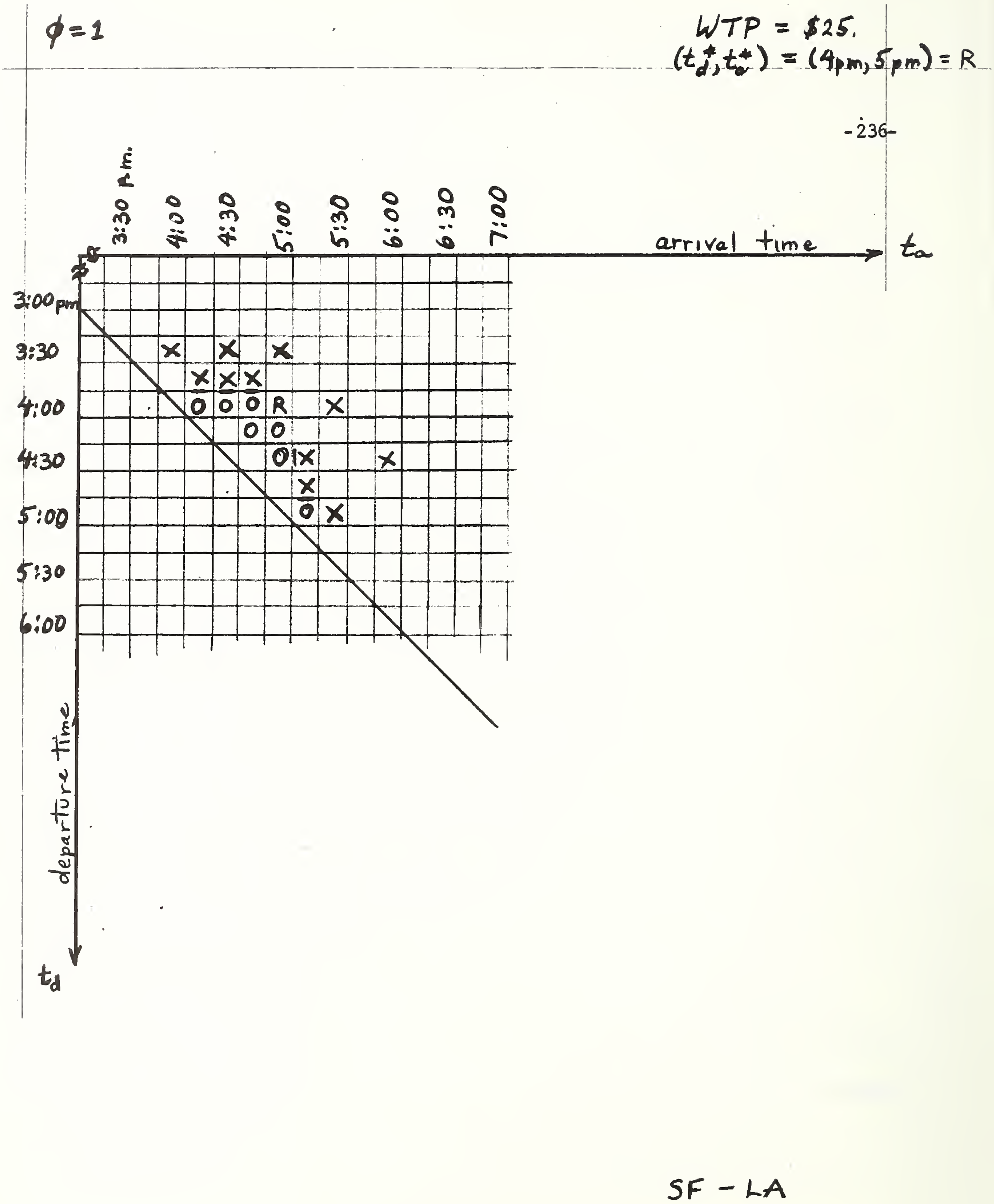


Figure A1. Sample Data Sheet.

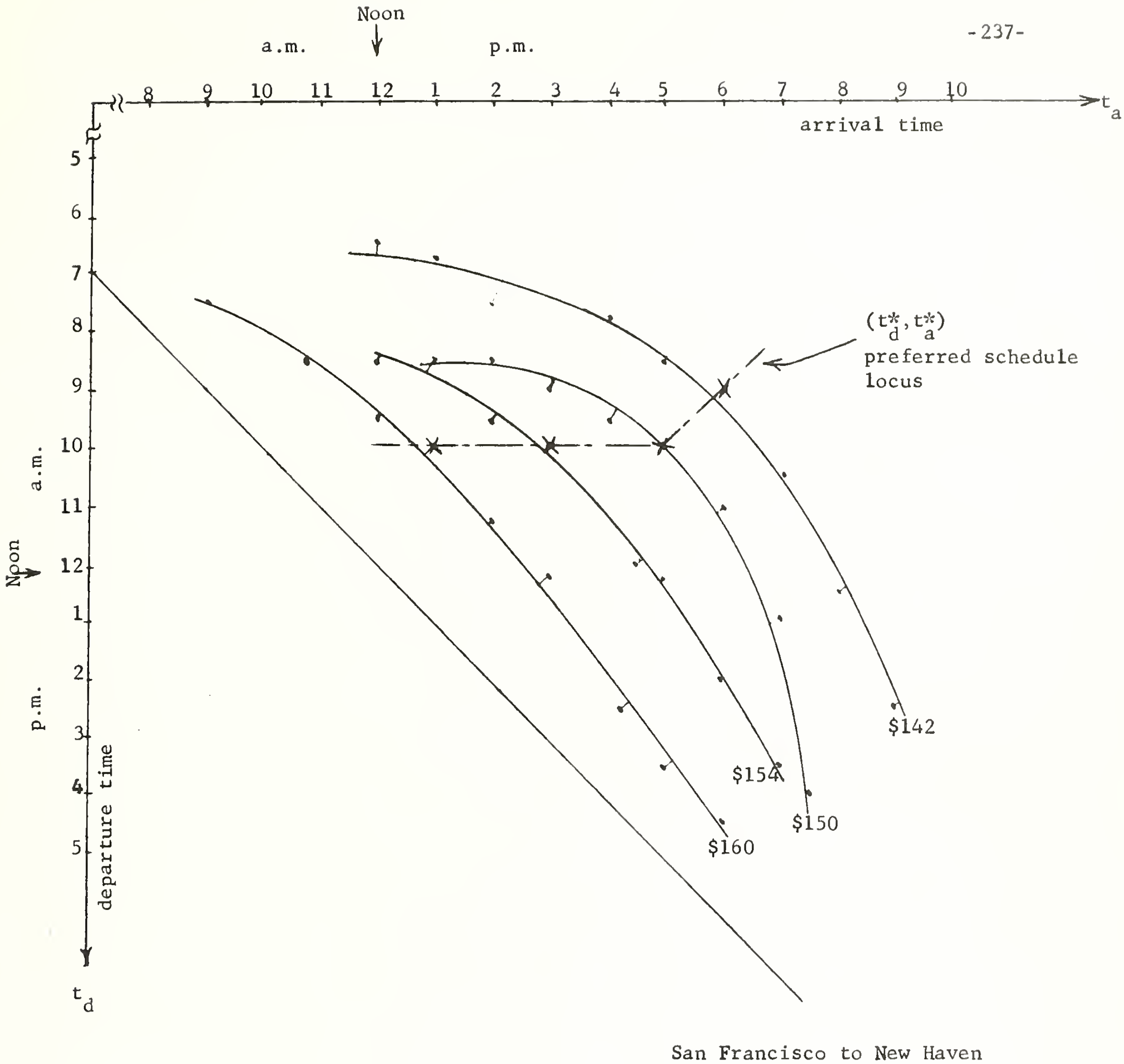


Figure A2. Traveler No. 1, (t_d, t_a) Plane.

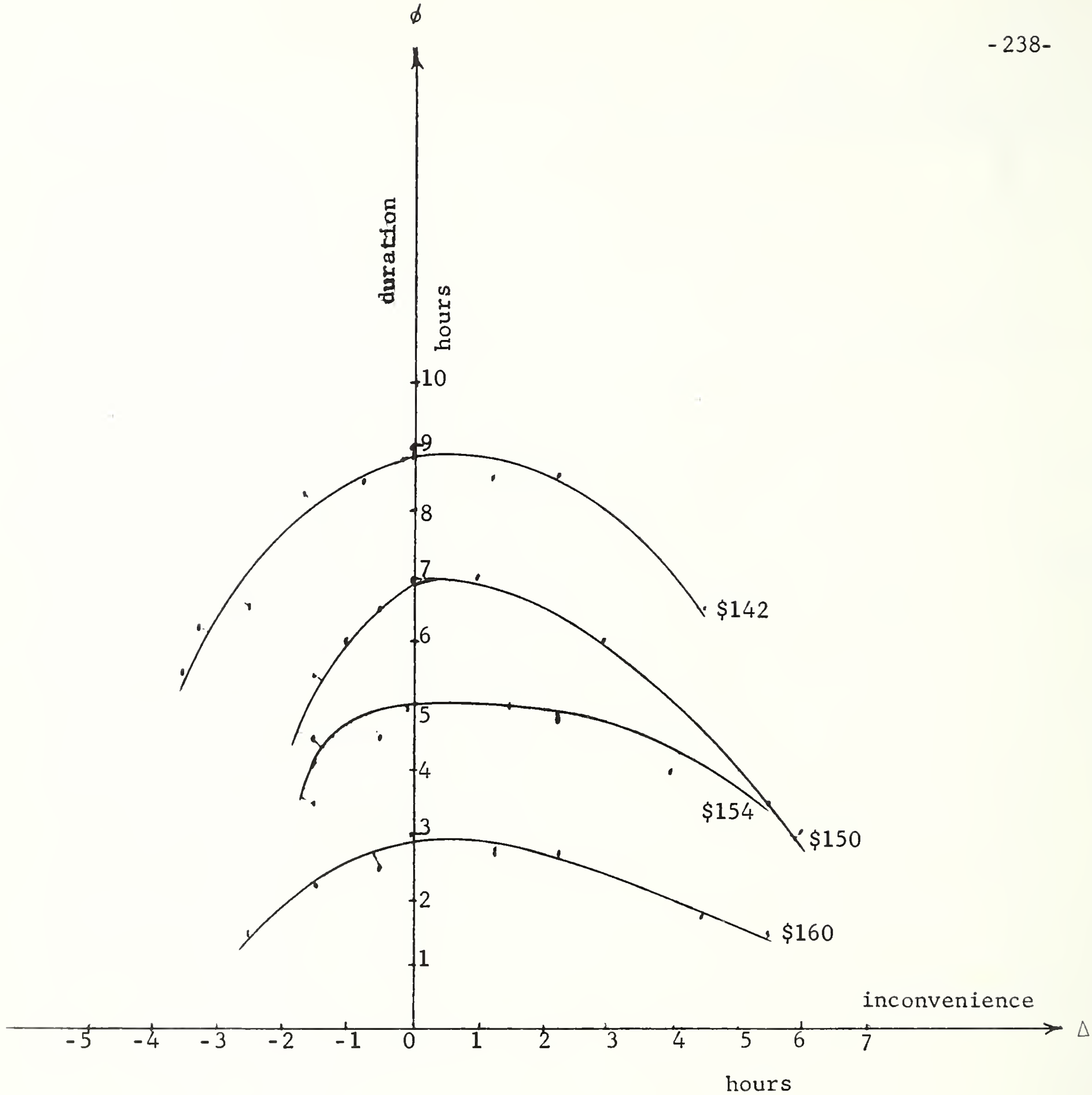


Figure A3. Traveler No. 1, (Δ,ø) Plane.

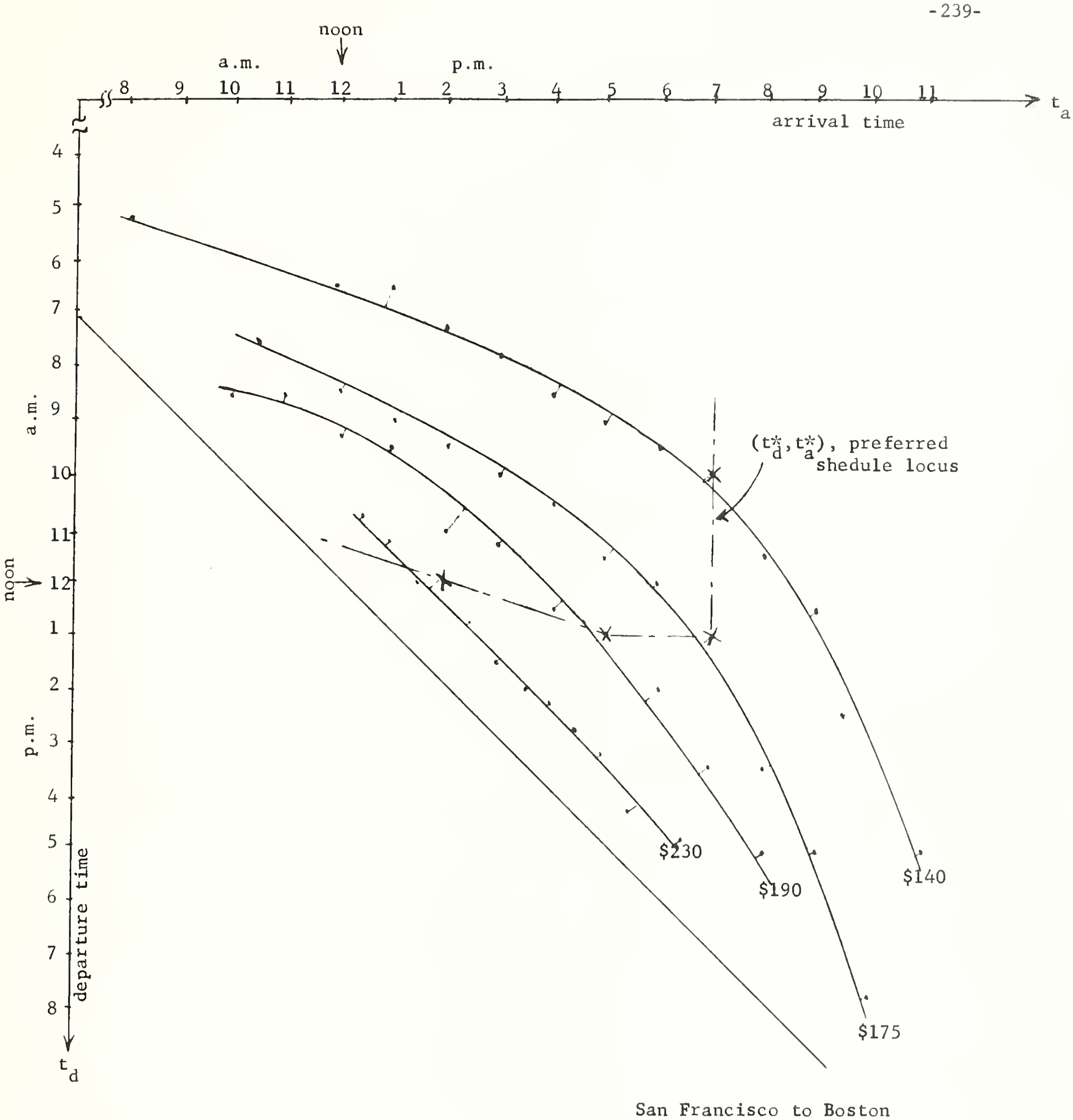


Figure A4. Traveler No. 2, (t_d, t_a) Plane.

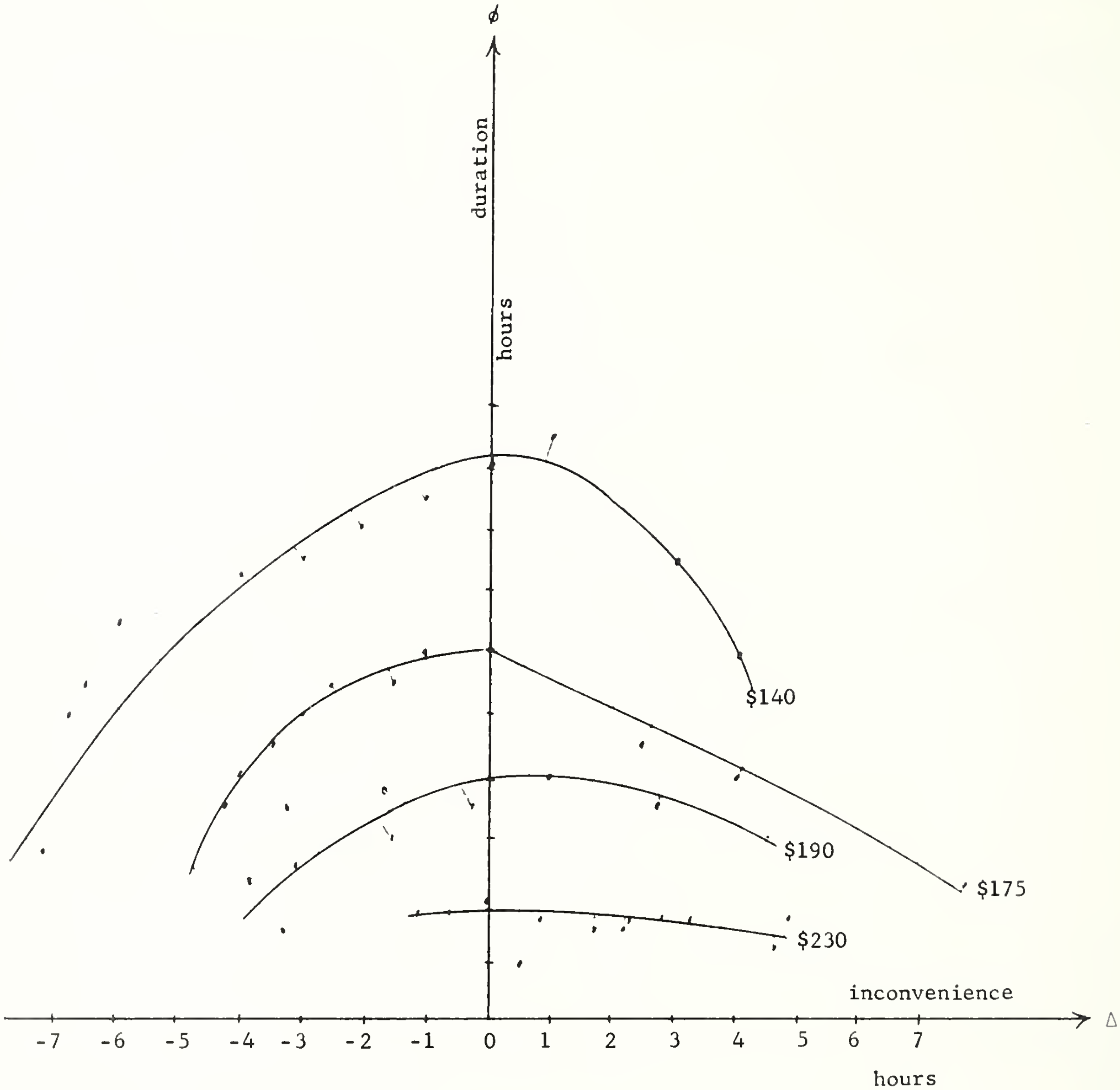


Figure A5. Traveler No. 2, (Δ,ϕ) Plane.

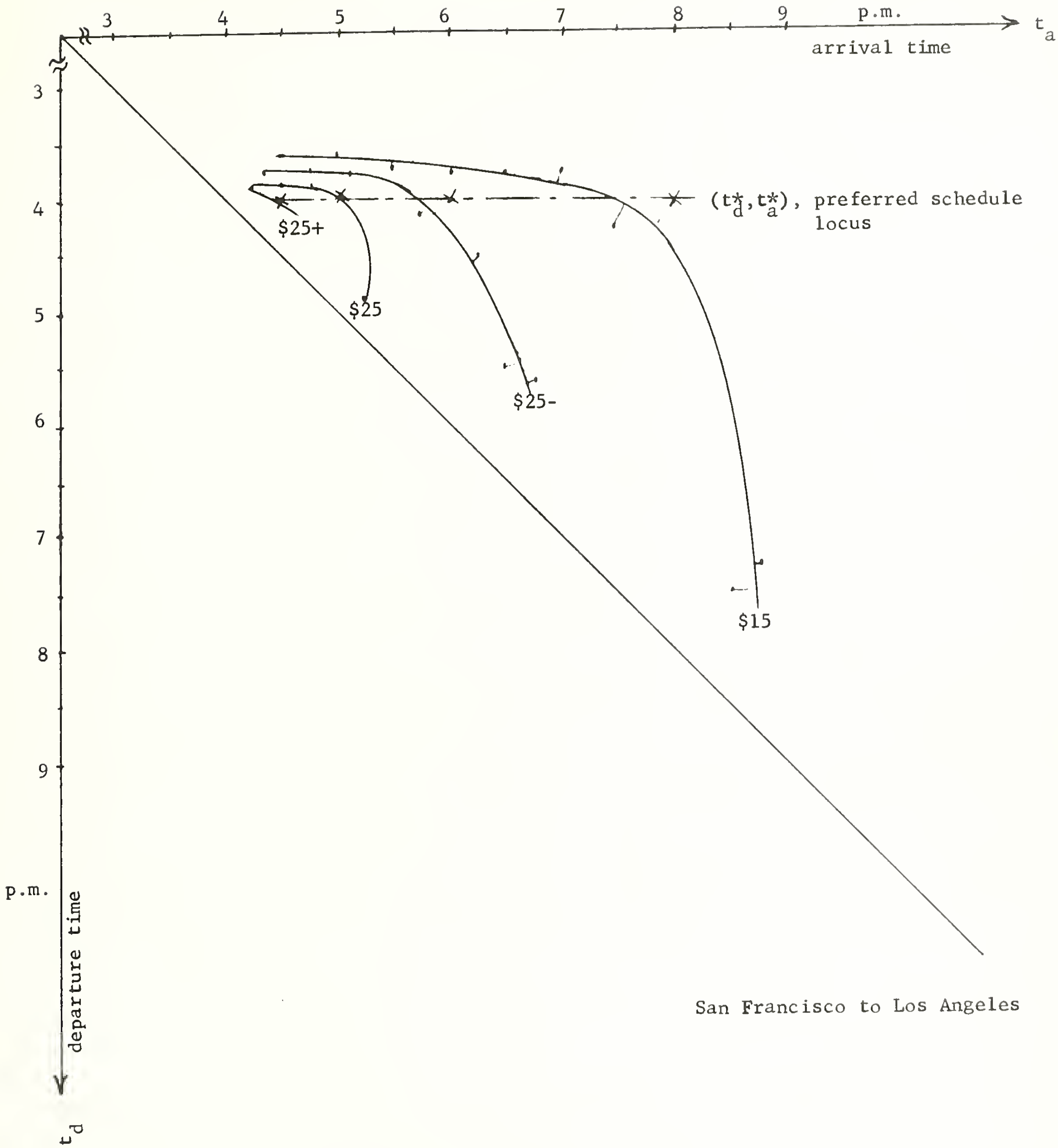


Figure A6. Traveler No. 3, (t_d, t_a) Plane.

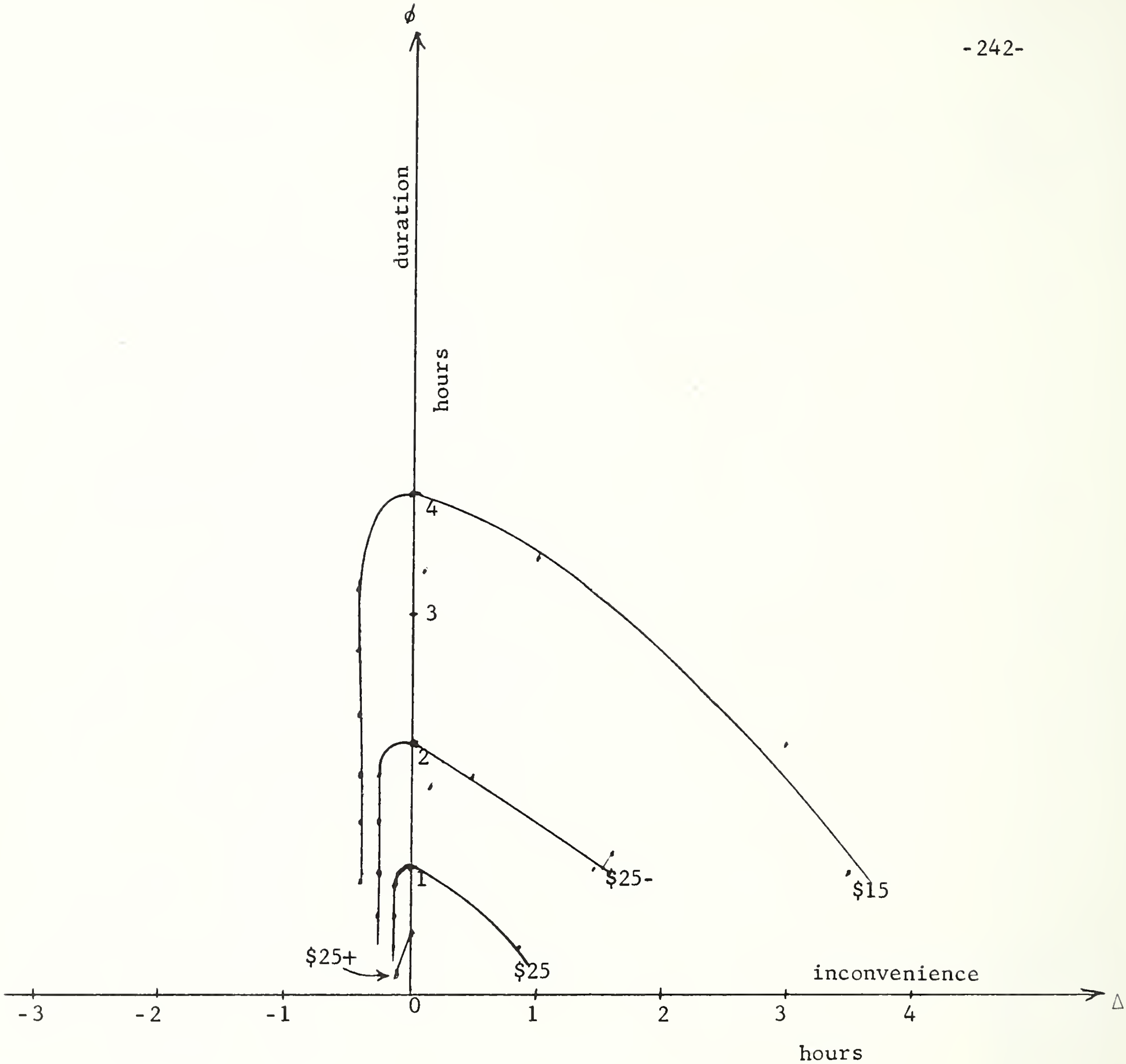


Figure A7. Traveler No. 3, (Δ,ϕ) Plane.

Second, as evidenced by the assymmetry of the contours in the (Δ,ϕ) plane, the distinction between earliness and lateness is an important one. Third, willingness to pay decreases as trips become longer and more inconvenient, as expected. Note, however, that for travelers like traveler number 3 (Figs. A6 and A7), the willingness to pay may be quite insensitive to decreases in trip impedance for trips above a certain "satisfactory" level.

Finally, the characterization of travelers by constant values of w and t^* appears reasonable over moderate ranges of trip duration, but not over very wide ranges. For example, the first traveler (Figs. A2 and A3) is strictly departure-oriented for trips below seven hours in length. Above seven hours, both arrival time and departure time become important. Hence w decreases below one. The second traveler (Figs. A4 and A5) has three distinct preferences regions. He is departure-oriented, with $t^* = 1$ p.m., for trips between four and six hours, and arrival-oriented, with $t^* = 7$ p.m., for trips over six hours. Alternately, he may be characterized, for trips longer than four hours, by $t^* = 1$ p.m. and the following variation of w: w is constant and equal to one for ϕ between four and six hours, and decreases linearly from one to zero as ϕ increases above six hours. For trips shorter than four hours, the traveler's entire orientation changes, such that he now prefers to travel earlier in the day, and can no longer be characterized by $t^* = 1$ p.m. Thus, over very wide ranges of trip duration, w and t^* for travelers 1 and 2 are non-constant functions of ϕ. On the other hand, traveler 3 is described by $t^* = 4$ p.m., $w = 1$, throughout the entire experimental range of trip times, 0.5 hours through 4 hours.

In summary, the experimental interview method appears promising as a research tool for investigating traveler preference behavior. The subjects seem to be quite capable of grappling with questions related to trips of direct concern to them. Test interviews took from forty-five minutes to an hour to administer (to obtain four willingness to pay contours), and no serious experimental difficulties were encountered.

APPENDIX B

COMPUTER PROGRAM FOR THE SHUTTLE SYSTEM (MODEL2)

```
      DIMENSION B(2),NTRAV(2),ELIG(2,100),DIST(2),W(2,100),
     1TSTAR(2,100),OPCOST(20),FARE(20),CAP(20),SPEED(20),EDGEWT(3000),
     2LIST(3000),LOC(200),NODWT1(200),NODWT2(200),IMAX(2,600),TR(2),
     3NZERO(20),TIMDEP(20,36),LASDEP(2,20),ABOARD(50),LOCVEH(2,20),
     4NUMPRE(2,20),PREVDP(20,36),NTRIP(2),PASBEN(2),BENFIT(20)
      INTEGER CAP,ELIG,EDGEWT,CONVER,ABOARD
      REAL IMAX,N
      COMMON IMAX,B,TIMDEP,LIST,LOC,M1,M2,LAST,NODWT1,NODWT2,
     1TR,DIST,SPEED,KFIN,EDGEWT,LASDEP,T,NZERO,ELIG,
     2A1,A2,A3,NTRAV,W,TSTAR,OPCOST,FARE,CAP,SLOPE,DOLLAR,ALPHA,N
     3,NUMVEH,ITER,INIT
      READ(5,4010) T,DIST(1),DIST(2),TR(1),TR(2)
 4010 FORMAT(5F10.3)
      READ(5,4020) SLOPE,DOLLAR,ALPHA,N,A1,A2,A3
 4020 FORMAT(7F10.3)
      READ(5,4030) KFIN,NTRAV(1),NTRAV(2),NUMVEH
 4030 FORMAT(4I5)
      DO 4035 NODE = 1,2
      PASBEN(NODE) = 0.
      LIMIT = NTRAV(NODE)
      READ(5,4040) (W(NODE,I), I=1,LIMIT)
 4035 READ(5,4040) (TSTAR(NODE,I),I=1,LIMIT)
 4040 FORMAT(16F5.2/16F5.2/16F5.2/16F5.2/16F5.2/16F5.2/4F5.2)
      DO 4045 J=1,NUMVEH
      BENFIT(J) = 0.
 4045 READ(5,4050) OPCOST(J),FARE(J),SPEED(J),CAP(J),NZERO(J)
 4050 FORMAT(3F10.2,2I5)
      READ(5,4030) MAXIT,INIT,IBOUND,KORECT
C     THE ABOVE READ STATEMENT INPUTS RUN PARAMETERS - MAXIT IS THE MAX
C     NO. OF PROGRAM ITERATIONS ALLOWED, INIT DETERMINES THE MODE OF
C     INITIALIZATION (SCHEDULES(0), ASSIGNMENTS(1), NONE(2)),
C     IBOUND INDICATES WHETHER UPLIM IS TO BE CALLED, KORECT INDICATES
C     IF MATCH IS TO BE USED (NO(0), YES(1), ONLY ON FIRST ITERATION(2))
      ITER = 0
      CONVER = 0
      MODIF = 0
C     ITER COUNTS PROGRAM ITERATIONS, CONVER = 1 MEANS SCHEDULE HAS
C     CONVERGED, MODIF = 1 INDICATES THAT MATCH HAS MADE NO CHANGE
C     IN PASSENGER ASSIGNMENTS
C     NOW PRINT OUT THE INPUT PARAMETERS
 4945 WRITE(6,4950)
 4950 FORMAT(1H1,40X,'THE INPUT PARAMETERS')
      WRITE(6,4960) T,DIST(1),DIST(2),TR(1),TR(2)
 4960 FORMAT(' T = ',F7.3,25X,F10.3,' MILES FROM NODE 1 TO NODE 2'/
     137X,F10.3,' MILES FROM NODE 2 TO NODE 1'/' TURNAROUND TIME AT NODE
     2 1 = ',F7.3,' HOURS ',10X,' TURNAROUND TIME AT NODE 2 = ',F7.3,
     3' HOURS')
      WRITE(6,4970) SLOPE,DOLLAR,ALPHA,N,A1,A2,A3
 4970 FORMAT(' SLOPE = ',F7.3,'      DOLLAR = ',F7.2,'      ALPHA = ',
     1F7.3,'      N = ',F7.3/' A1 = ',F3.0,'      A2 = ',F3.0,
     2'      A3 = ',F3.0)
      RKFIN = KFIN
      DEL = T/RKFIN
      WRITE(6,4975) KFIN,DEL
 4975 FORMAT(1H0,'KFIN = ',I5,20X,' DELTA T = ',F6.3)
      WRITE(6,4976) INIT,KORECT,MAXIT,IBOUND
```

```
4976 FORMAT(1H0,' INIT = ',I4,' KORECT = ',I4,' MAXIT = ',I4,' IROUND =
    1 ',I4)
     WRITE(6,5000)
5000 FORMAT(1H0,' VEHICLE NO.  SPEED      CAPACITY  INITIAL STATION  OPE
    1R. COST/MI.  FARE/ONE-WAY')
     DO 5015 J=1,NUMVEH
5015 WRITE(6,5020) J,SPEED(J),CAP(J),NZERO(J),OPCOST(J),FARE(J)
5020 FORMAT(6X,I3,3X,F10.3,5X,I3,9X,I3,17X,F6.2,9X,F6.2)
     WRITE(6,5025)
5025 FORMAT(1H1,35X,'LIST OF TRAVELERS'/10X,'NODE 1',64X,'NODE 2'/
    1' PASSENGER NO. TSTAR',5X,'W',46X,'PASSENGER NO. TSTAR',4X,'W')
     ILIM = MIN0(NTRAV(1),NTRAV(2))
     DO 5030 I=1,ILIM
5030 WRITE(6,5050) I,TSTAR(1,I),W(1,I),I,TSTAR(2,I),W(2,I)
5050 FORMAT(6X,I4,2X,F7.3,3X,F4.2,51X,I4,2X,F7.3,3X,F4.2)
     IF(NTRAV(1).LE.ILIM) GO TO 5075
     LIM2 = NTRAV(1)
     ILIM = ILIM+1
     DO 5065 I=ILIM,LIM2
5065 WRITE(6,5070) I,TSTAR(1,I),W(1,I)
5070 FORMAT(6X,I4,2X,F7.3,3X,F4.2)
     GO TO 2000
5075 IF(NTRAV(2).LE.ILIM) GO TO 2000
     LIM2 = NTRAV(2)
     ILIM = ILIM+1
     DO 5080 I=ILIM,LIM2
5080 WRITE(6,5085) I,TSTAR(2,I),W(2,I)
5085 FORMAT(77X,I4,2X,F7.3,3X,F4.2)
2000 IF(INIT.EQ.2) GO TO 7000
C     READ IN THE INITIAL SCHEDULE OR PASSENGER ASSIGNMENT
C     TO READ IN A NULL VEHICLE SCHEDULE, LET LASDEP(1,J)=1,
C     LASDEP(2,J)=0
     IF(INIT.EQ.0) GO TO 4058
     DO 4054 NODE=1,2
     LIMIT = NTRAV(NODE)
4054 READ(5,4053) (ELIG(NODE,I),I=1,LIMIT)
4053 FORMAT(20I4/20I4/20I4/20I4/20I4)
C     PRINT OUT INITIAL ASSIGNMENTS
5088 WRITE(6,5086)
5086 FORMAT(1H1,30X,'INITIAL ASSIGNMENT OF TRAVELERS')
     DO 5091 J=1,NUMVEH
     WRITE(6,5087) J
5087 FORMAT(1H0,10X,'PASSENGERS ASSIGNED TO VEHICLE',I4/)
     DO 5091 NODE=1,2
     LIM = NTRAV(NODE)
     ABOARD(1) = 0
     MM = 1
     DO 5089 M=1,LIM
     IF(ELIG(NODE,M).NE.J*100) GO TO 5089
     ABOARD(MM) = M
     MM = MM+1
5089 CONTINUE
     M2 = MM-1
     MX = MAX0(M2,1)
5091 WRITE(6,5092) NODE,(ABOARD(MN),MN=1,MX)
5092 FORMAT(' FROM NODE',I4,', PASSENGERS  ',25I4)
     DO 5093 J=1,NUMVEH
```

```
      LASDEP(1,J) = 1
      LASDEP(2,J) = 0
      TIMDEP(J,1) = 5.*T
 5093 CONTINUE
      GO TO 4550
C     READ INITIAL SCHEDULE, DEPARTURES FROM NODE 2 ARE TAGGED WITH
C     MINUS SIGNS IN TIMDEP
 4058 DO 4060 NODE = 1,2
 4060 READ(5,4065) (LASDEP(NODE,J), J=1,NUMVEH)
 4065 FORMAT(20I4)
C     FOR NULL SCHEDULES LET TIMDEP(J,1) BE GREATER THAN T
C     AND LET LASDEP(1,J)=1, LASDEP(2,J)=0
      DO 4070 J=1,NUMVEH
      NUMDEP = LASDEP(1,J)+LASDEP(2,J)
 4070 READ(5,4075) (TIMDEP(J,NUM), NUM=1,NUMDEP)
 4075 FORMAT(10F6.2)
      DO 4076 NODE=1,2
      LIM = NTRAV(NODE)
      DO 4076 M=1,LIM
 4076 ELIG(NODE,M) = 0
C     BELOW, CHECK IF MATCH IS TO BE EXECUTED ON THE FIRST ITERATION
 7500 IF(KORECT.NE.0.AND.ITER.EQ.0) GO TO 4077
 5090 WRITE(6,5095)
 5095 FORMAT(1H1,40X,' THE SCHEDULE')
      DO 6064 J=1,NUMVEH
      IF(TIMDEP(J,1).LE.T) GO TO 5099
      WRITE(6,5096) J
 5096 FORMAT(1H0,' VEHICLE ',I4,' HAS NO SCHEDULED DEPARTURES'/)
      GO TO 6064
 5099 WRITE(6,6000) J
 6000 FORMAT(/1H0,' SCHEDULE FOR VEHICLE',I4/' DEPARTS NODE',4X,'AT TIME
     1',5X,'PASSENGERS ABOARD'/)
      LAS = LASDEP(1,J)+LASDEP(2,J)
      NTRIP(1) = 0
      NTRIP(2) = 0
      DO 6060 NUM=1,LAS
      IF(TIMDEP(J,NUM).LE.0.) GO TO 6020
      NODE = 1
      NTRIP(1) = NTRIP(1)+1
      GO TO 6030
 6020 NODE = 2
      NTRIP(2) = NTRIP(2)+1
 6030 MM = 1
      ABOARD(1) = 0
      LIM = NTRAV(NODE)
      DO 6040 I=1,LIM
      IF(ELIG(NODE,I).NE.J*100+NTRIP(NODE)) GO TO 6040
      ABOARD(MM) = I
      MM = MM+1
 6040 CONTINUE
 6045 MAX = MM-1
      MX = MAXO(MAX,1)
      TIM =  ABS(TIMDEP(J,NUM))
      WRITE(6,6050) NODE,TIM,(ABOARD(MN),MN=1,MX)
 6050 FORMAT(8X,I4,6X,F6.2,4X,25I4)
C     IF CAPACITIES GREATER THAN 25 ARE CONSIDERED, EXPAND FORMAT 6050
 6060 CONTINUE
```

```
      WRITE(6,6062) J,BENEFIT(J)
 6062 FORMAT(1H0,'BENEFIT ACHIEVED BY VEHICLE',I4,' = ',F10.3)
 6064 CONTINUE
      BEN1 = 0.
      BEN2 = PASBEN(1) + PASBEN(2)
      DO 4980 J=1,NUMVEH
 4980 BEN1 = BEN1 + BENEFIT(J)
      WRITE(6,6065) BEN1,BEN2
 6065 FORMAT(1H0,'BENEFIT ACHIEVED IN SCHEDULING PROCESS = ',F10.3/
     1' BENEFIT ACHIEVED IN ASSIGNMENT PROCESS = ',F10.3)
      IF(ITER.EQ.MAXIT) GO TO 6500
      IF(KORECT.EQ.1.AND.MODIF.EQ.1.AND.CONVER.EQ.1) GO TO 6500
      IF(CONVER.EQ.1.AND.KORECT.NE.1) GO TO 6500
      WRITE(6,6075)
 6075 FORMAT(1H0,'THE PROGRAM HAS NOT YET CONVERGED')
      GO TO 4550
C     BELOW, TRAVELERS ARE OPTIMALLY MATCHED TO THE CURRENT SCHEDULE
C     ARRAY LOCVEH GIVES THE LOCATION OF THE FIRST SEAT-DEPARTURE IN
C     NODE SET 2 OF MATCH, CORRESPONDING TO EACH VEHICLE J
C     LOCVEH IS FILLED BELOW
 4077 JLIM = NUMVEH+1
      DO 4080 NODE = 1,2
      PASBEN(NODE) = 0.
      LOCVEH(NODE,1) = 1
      DO 4080 J=2,JLIM
      LOCVEH(NODE,J) = LOCVEH(NODE,J-1)+CAP(J-1)*LASDEP(NODE,J-1)
 4080 CONTINUE
C     NOW, CONVERT PASSENGER LIST AND SCHEDULE INTO FORM FOR MATCH
C     INEXT KEEPS TRACK OF INDEX I IN LIST(I) AND EDGEWT(I) AS THESE
C                 ARRAYS ARE FILLED
C     NSET2 KEEPS TALLY OF THE CURRENT NODE IN SET 2 OF MATCH FOR WHICH
C          A PASSENGER'S WTP IS BEING CONSIDERED
      NODE = 1
 4100 M1 = NTRAV(NODE)
      M2 = 0
      DO 4125 J=1,NUMVEH
 4125 M2 = M2+CAP(J)*LASDEP(NODE,J)
C     BELOW, M2 IS INCREASED BY ONE TO CREATE AN ARTIFICIAL NODE TO
C     WHICH INELIGIBLE TRAVELERS CAN BE ASSIGNED IN MATCH.
      M2 = M2+1
      INEXT = 1
      DO 4300 I=1,M1
      NIX = 0
      LOC(I) = INEXT
      DO 4250 J=1,NUMVEH
      NSET2 = LOCVEH(NODE,J)
      LIM = LASDEP(1,J)+LASDEP(2,J)
      TRAVTM = DIST(NODE)
      TRAVTM = TRAVTM/SPEED(J)
      DO 4250 NUMDEP = 1,LIM
      TD = TIMDEP(J,NUMDEP)
      IF(NODE.EQ.2) GO TO 4150
      IF(TD.LE.0.) GO TO 4250
      IF(TD.GT.T) GO TO 4245
      GO TO 4175
 4150 IF(TD.GT.0.) GO TO 4250
      TD = -TD
```

```
4175 WTP = WILPAY(TD,TRAVTM,N,TSTAR(NODE,I),W(NODE,I),SLOPE,
    1             DOLLAR,ALPHA)
     IF(WTP.LT.FARE(J)) GO TO 4245
     NIX = 1
     ILIM2 = INEXT+CAP(J)-1
     DO 4200 II=INEXT,ILIM2
     EDGEWT(II) = (A2*FARE(J)+A3*WTP)*10000.+0.5
4200 LIST(II) = NSET2+II-INEXT
     INEXT = INEXT+CAP(J)
4245 NSET2 = NSET2+CAP(J)
4250 CONTINUE
     IF(NIX.EQ.1) GO TO 4300
     LIST(INEXT) = M2
     EDGEWT(INEXT) = 1
     INEXT = INEXT+1
4300 CONTINUE
     LAST = INEXT-1
     LOC(M1+1) = LAST+1
     CALL MATCH
     MODIF = 1
     DO 4500 M=1,M1
     LIM1 = LOC(M)
     LIM2 = LOC(M+1)-1
     DO 4350 I=LIM1,LIM2
     IF(EDGEWT(I).LT.0) GO TO 4375
4350 CONTINUE
4355 JX = ELIG(NODE,M)/100
     IF((ELIG(NODE,M)-JX).NE.0) MODIF=0
     ELIG(NODE,M) = 0
     GO TO 4500
4375 IF(LIST(I).EQ.M2) GO TO 4355
     WT1 = -EDGEWT(I)
     WT2 = WT1/10000.
     PASBEN(NODE) = PASBEN(NODE)+WT2
     DO 4400 J=1,NUMVEH
     IF(LOCVEH(NODE,J).LE.LIST(I).AND.LOCVEH(NODE,J+1).GT.LIST(I))
    1GO TO 4425
4400 CONTINUE
     WRITE(6,4405)
4405 FORMAT(1H0,'ERROR - PROGRAM PASSED THRU THE 4400 LOOP')
     GO TO 6535
4425 JDEP = (LIST(I)-LOCVEH(NODE,J))/CAP(J)+1
     IF(ELIG(NODE,M).NE.(J*100+JDEP)) MODIF=0
     ELIG(NODE,M) = J*100+JDEP
4500 CONTINUE
     IF(NODE.EQ.2) GO TO 5090
     NODE = 2
     GO TO 4100
7000 DO 7005 NODE=1,2
     LIM = NTRAV(NODE)
     DO 7005 M=1,LIM
7005 ELIG(NODE,M) = 0
     DO 7010 J=1,NUMVEH
     LASDEP(1,J) = 1
     LASDEP(2,J) = 0
7010 TIMDEP(J,1) = 5.*T
     WRITE(6,7015)
```

```
7015 FORMAT(1HO//' NO INITIAL SCHEDULE OF ASSIGNMENT IS GIVEN')
4550 DO 4575 NODE = 1,2
     DO 4575 J=1,NUMVEH
4575 NUMPRE(NODE,J) = LASDEP(NODE,J)
4580 DO 4650 J=1,NUMVEH
     LIM = LASDEP(1,J)+LASDEP(2,J)
     DO 4600 NDEP=1,LIM
4600 PREVDP(J,NDEP) = TIMDEP(J,NDEP)
4605 CALL SCHED(J)
     NSTART = NZERO(J)
     IF(NSTART.EQ.0.AND.IMAX(1,1).GT.IMAX(2,1)) NSTART=1
     IF(NSTART.EQ.0.AND.IMAX(1,1).LE.IMAX(2,1)) NSTART=2
     CALL ADJUST(J)
4650 BENFIT(J) = IMAX(NSTART,1)
     ITER = ITER+1
     WRITE(6,6070) ITER
6070 FORMAT(1HO,'PROGRAM HAS COMPLETED',I4,' ITERATIONS')
     DO 4675 NODE=1,2
     DO 4675 J=1,NUMVEH
     IF(NUMPRE(NODE,J).EQ.LASDEP(NODE,J)) GO TO 4675
     WRITE(6,4655)
4655 FORMAT(1HO,' THE VEHICLE SCHEDULES HAVE NOT YET CONVERGED - MATCH
    1IS NOT APPLIED TO THE FOLLOWING SCHEDULE')
     GO TO 5090
4675 CONTINUE
     DO 4700 J=1,NUMVEH
     LAS = LASDEP(1,J)+LASDEP(2,J)
     DO 4700 NDEP = 1,LAS
     IF(TIMDEP(J,NDEP).GT.(PREVDP(J,NDEP)-T/(2.*RKFIN)).AND.
    1TIMDEP(J,NDEP).LT.(PREVDP(J,NDEP)+T/(2.*RKFIN))) GO TO 4700
     WRITE(6,4655)
     GO TO 5090
4700 CONTINUE
C     CONVERGENCE IS REACHED
     CONVER = 1
     IF(KORECT.EQ.1) WRITE(6,4725)
4725 FORMAT(1HO,' THE VEHICLE SCHEDULE HAS CONVERGED, MATCH IS NOW APPL
    1IED')
     IF(KORECT.EQ.1) GO TO 4077
     WRITE(6,6525)
6525 FORMAT(1HO,'THE SCHEDULE HAS CONVERGED')
     GO TO 5090
6500 IF(ITER.EQ.MAXIT) WRITE(6,6530) MAXIT
6530 FORMAT(1HO,' MAX NO. OF ITERATIONS,',I4,', WERE PERFORMED')
6535 IF(IBOUND.EQ.1) CALL UPLIM
     RETURN
     END
```

```
      SUBROUTINE ADJUST(J)
      DIMENSION IMAX(2,600),B(2),TIMDEP(20,36),LIST(3000),LOC(200),
     1NODWT1(200),NODWT2(200),TR(2),DIST(2),SPEED(20),EDGEWT(3000),
     2LASDEP(2,20),NZERO(20),ELIG(2,100),NTRAV(2),W(2,100),TSTAR(2,100),
     3OPCOST(20),FARE(20),CAP(20),QUEUE(50),PASS(50)
      INTEGER CAP,ELIG,EDGEWT,PASS
      REAL IMAX,N
      COMMON IMAX,B,TIMDEP,LIST,LOC,M1,M2,LAST,NODWT1,NODWT2,
     1TR,DIST,SPEED,KFIN,EDGEWT,LASDEP,T,NZERO,ELIG,
     2A1,A2,A3,NTRAV,W,TSTAR,OPCOST,FARE,CAP,SLOPE,DOLLAR,ALPHA,N
     3,NUMVEH,ITER,INIT
      ICAP = CAP(J)
      LMAX = LASDEP(1,J)+LASDEP(2,J)
      DO 8500 NODE=1,2
      JTRIP = 0
      TRAVTM = DIST(NODE)/SPEED(J)
      LIM = NTRAV(NODE)
C     BELOW, IF CURRENT TRAVELER HAS ELIGIBILITY FOR A TRIP BY J WHICH
C     NO LONGER EXISTS, SET HIS ELIGIBILITY TO ZERO
      DO 8001 M=1,LIM
      JJ = ELIG(NODE,M)/100
      JJTRIP = ELIG(NODE,M)-JJ*100
      IF(J.EQ.JJ.AND.JJTRIP.GT.LASDEP(NODE,J)) ELIG(NODE,M)=0
 8001 CONTINUE
      DO 8475 LL=1,LMAX
      IF((NODE.EQ.1.AND.TIMDEP(J,LL).LE.0.).OR.(NODE.EQ.2.AND.
     1TIMDEP(J,LL).GT.0.)) GO TO 8475
      JTRIP = JTRIP+1
      TD = ABS(TIMDEP(J,LL))
      LOAD = 0
      IFULL = ICAP+1
      DO 8425 M=1,LIM
      JJ = ELIG(NODE,M)/100
      IF(INIT.EQ.1.AND.ITER.EQ.0.AND.JJ.NE.J) GO TO 8425
      JJTRIP = ELIG(NODE,M)-JJ*100
 8005 WTP = WILPAY(TD,TRAVTM,N,TSTAR(NODE,M),W(NODE,M),SLOPE,DOLLAR,
     1ALPHA)
C     BELOW, IF CURRENT TRAVELER IS ELIGIBLE FOR THE CURRENT TRIP, BUT
C     IS UNWILLING TO PAY THE PRICE BECAUSE THE TIME OF THE TRIP HAS
C     BEEN CHANGED SINCE THE LAST ITERATION, SET HIS ELIGIBILITY TO ZERO
      IF(ELIG(NODE,M).NE.(J*100+JTRIP).OR.WTP.GE.FARE(J)) GO TO 8007
      ELIG(NODE,M) = 0
      GO TO 8425
C     BELOW, IF THE CURRENT TRAVELER HAS ELIGIBILITY DIFFERENT FROM,
C     AND IS UNWILLING TO PAY FOR THE CURRENT TRIP, DO NOT MODIFY
C     HIS ELIGIBILITY
 8007 IF(WTP.LT.FARE(J)) GO TO 8425
C     BELOW, IF CURRENT TRAVELER IS ELIGIBLE FOR A DIFFERENT VEHICLE
C     THAN J USE NET INCREMENT RATHER THAN TOTAL BENEFIT
      IF(JJTRIP.NE.0.AND.JJ.NE.J) GO TO 8010
      VALUE = A2*FARE(J)+A3*WTP
      GO TO 8050
 8010 NDEP = 2*JJTRIP
 8015 IF((TIMDEP(JJ,1).LE.0..AND.NODE.EQ.2).OR.
     1(TIMDEP(JJ,1).GT.0..AND.NODE.EQ.1)) NDEP = NDEP-1
C     PASSENGER (NODE,M) HAS BEEN FOUND TO HAVE BEEN ELIGIBLE FOR
```

```
C     THE JJTRIP DEPARTURE FROM NODE BY VEHICLE JJ. ABOVE, AT
C    STATEMENTS 8010 AND 8015,JJTRIP WAS CONVERTED TO NDEP - THE
C     DEPARTURE NUMBER OF THE JJTRIP TRIP FROM NODE
      TDJJ = ABS(TIMDEP(JJ,NDEP))
      TRAVJJ = DIST(NODE)/SPEED(JJ)
      WTP2 = WILPAY(TDJJ,TRAVJJ,N,TSTAR(NODE,M),W(NODE,M),SLOPE,DOLLAR,
     1ALPHA)
      VALUE = A2*(FARE(J)-FARE(JJ))+A3*(WTP-WTP2)
      IF(VALUE.LE.0.) GO TO 8425
 8050 LOAD = LOAD+1
      IF(LOAD.GT.CAP(J)) GO TO 8060
      QUEUE(LOAD) = VALUE
      PASS(LOAD) = M
      GO TO 8425
 8060 QUEUE(IFULL) = VALUE
      PASS(IFULL) = M
C     THE FOLLOWING LOOP MOVES THE SMALLEST COMPONENT IN QUEUE TO THE
C     LAST (IFULL) POSITION, THEREBY DISALLOWING THE CORRESPONDING
C     PASSENGER FROM BOARDING
      DO 8075 NN=1,ICAP
      IF(QUEUE(NN).GE.QUEUE(NN+1)) GO TO 8075
      TEM = QUEUE(NN)
      ITEM = PASS(NN)
      QUEUE(NN) = QUEUE(NN+1)
      PASS(NN) = PASS(NN+1)
      QUEUE(NN+1) = TEM
      PASS(NN+1) = ITEM
 8075 CONTINUE
C     BELOW, IF PASSENGER IFULL WAS ELIGIBLE FOR CURRENT TRIP, SET HIS
C     ELIGIBILITY TO ZERO SINCE HE WILL NOT BE BOARDED
      MX = PASS(IFULL)
      IF(ELIG(NODE,MX).EQ.(J*100+JTRIP)) ELIG(NODE,MX) = 0
 8425 CONTINUE
      MLIM = MIN0(LOAD,ICAP)
      IF(MLIM.EQ.0) GO TO 8475
      DO 8450 MM=1,MLIM
      M = PASS(MM)
 8450 ELIG(NODE,M) = J*100+JTRIP
 8475 CONTINUE
 8500 CONTINUE
      RETURN
      END
```

```
)MPILER OPTIONS - NAME=  MAIN,OPT=01,LINECNT=57,SOURCE,EBCDIC,NOLIST,DECK,LOAD,MAP,NOE
      SUBROUTINE IMBEN(J,K)
      DIMENSION B(2),TRATIM(2),NTRAV(2),ELIG(2,100),DIST(2)
      DIMENSION W(2,100),TSTAR(2,100)
      DIMENSION OPCOST(20),FARE(20),CAP(20),SPEED(20)
      DIMENSION EDGEWT(3000),LIST(3000),LOC(200),NODWT1(200),NODWT2(200)
     1,IMAX(2,600),TR(2),NZERO(20),TIMDEP(20,36),LASDEP(2,20)
      DIMENSION QUEUE(50)
      INTEGER CAP,ELIG,EDGEWT
      REAL N,IMAX
      COMMON IMAX,B,TIMDEP,LIST,LOC,M1,M2,LAST,NODWT1,NODWT2,
     1TR,DIST,SPEED,KFIN,EDGEWT,LASDEP,T,NZERO,ELIG,
     2A1,A2,A3,NTRAV,W,TSTAR,OPCOST,FARE,CAP,SLOPE,DOLLAR,ALPHA,N
     3,NUMVEH,ITER,INIT
      DO 3000 L=1,2
      B(L) = -A1*OPCOST(J)*DIST(L)
      TRATIM(L) = DIST(L)
 3000 TRATIM(L) = TRATIM(L)/SPEED(J)
      R1 = K
      R2 = KFIN
      TD = R1*T/R2
      IFULL = CAP(J)+1
      ICAP = CAP(J)
      DO 3300 NODE = 1,2
      NTRIP = 0
      LOAD = 0
      ILIM = NTRAV(NODE)
      DO 3300 I = 1,ILIM
      WTP = WILPAY(TD,TRATIM(NODE),N,TSTAR(NODE,I),W(NODE,I),
     1SLOPE,DOLLAR,ALPHA)
      IF(WTP.LT.FARE(J)) GO TO 3300
      JJ = ELIG(NODE,I)/100
C     BELOW, IF INITIAL ASSIGNMENTS ARE GIVEN, PASSENGERS ARE ELIGIBLE
C     ONLY FOR THEIR ASSIGNED VEHICLES ON THE INITIAL ITERATION
      IF(JJ.NE.J.AND.INIT.EQ.1.AND.ITER.EQ.0) GO TO 3300
      NTRIP = ELIG(NODE,I)-JJ*100
      IF(NTRIP.EQ.0.OR.JJ.EQ.J) GO TO 3010
      NDEP = 2*NTRIP
      IF((TIMDEP(JJ,1).LE.0..AND.NODE.EQ.2).OR.
     1(TIMDEP(JJ,1).GT.0..AND.NODE.EQ.1)) NDEP = NDEP-1
C     THE FOLLOWING STATEMENTS, UP TO 3010, COMPUTE THE NET INCREMENTAL
C     BENEFIT TO TRAVELERS ASSIGNED TO VEHICLES OTHER THAN J
      TDJJ = ABS(TIMDEP(JJ,NDEP))
      TRAVJJ = DIST(NODE)/SPEED(JJ)
      WTP2 = WILPAY(TDJJ,TRAVJJ,N,TSTAR(NODE,I),W(NODE,I),
     1SLOPE,DOLLAR,ALPHA)
      VALUE = A2*(FARE(J)-FARE(JJ))+A3*(WTP-WTP2)
      IF(VALUE.LE.0.) GO TO 3300
      GO TO 3020
 3010 VALUE = A2*FARE(J)+A3*WTP
 3020 LOAD = LOAD+1
      IF(LOAD.LE.ICAP) GO TO 3250
      QUEUE(IFULL) = VALUE
      DO 3150 NN=1,ICAP
      IF(QUEUE(NN).GE.QUEUE(NN+1)) GO TO 3150
      TEM = QUEUE(NN)
      QUEUE(NN) = QUEUE(NN+1)
```

```
      QUEUE(NN+1) = TEM
 3150 CONTINUE
      B(NODE) = B(NODE)+VALUE-QUEUE(IFULL)
      GO TO 3300
 3250 B(NODE) = B(NODE)+VALUE
      QUEUE(LOAD) = VALUE
 3300 CONTINUE
      RETURN
      END
```

```
)MPILER OPTIONS - NAME=  MAIN,OPT=01,LINECNT=57,SOURCE,EBCDIC,NOLIST,DECK,LOAD,MAP,NOE

      FUNCTION WILPAY(TD,TRATIM,N,TSTAR,W,SLOPE,DOLLAR,ALPHA)
      REAL N
      DEL = ABS(TD+(1.-W)*TRATIM-TSTAR)
      A = (DEL**N+(TRATIM/SLOPE)**N)**(1./N)
      WILPAY = DOLLAR*EXP(-(A/ALPHA)**2)
      RETURN
      END
```

```
      SUBROUTINE SCHED(J)
      DIMENSION B(2),IMAX(2,600),C(2,600),TR(2),ITR(2),DIST(2),SPEED(20)
     1,TRATIM(2),ITIME(2),TIMDEP(20,36),LASDEP(2,20),NZERO(20),
     2EDGEWT(3000),LIST(3000),LOC(200),NODWT1(200),NODWT2(200)
     3,ELIG(2,100)
      COMMON IMAX,B,TIMDEP,LIST,LOC,M1,M2,LAST,NODWT1,NODWT2,
     1TR,DIST,SPEED,KFIN,EDGEWT,LASDEP,T,NZERO,ELIG
      INTEGER C,READY1,READY2,EDGEWT,ELIG
      REAL IMAX
      RKFIN = KFIN
      DELT = T/RKFIN
      DO 3500 L=1,2
      TRATIM(L) = DIST(L)
      TRATIM(L) = TRATIM(L)/SPEED(J)
      ITIME(L) = (TRATIM(L)/DELT)+0.5
 3500 ITR(L) = TR(L)/DELT+0.5
      K1 = KFIN+1
      DO 3550 KK=1,K1
      C(1,KK) = 0
 3550 C(2,KK) = 0
      READY1 = ITIME(1)+ITR(2)
      READY2 = ITIME(2)+ITR(1)
      KKMAX = MAX0(READY1,READY2)+K1
      DO 3600 KK= 1,KKMAX
      IMAX(1,KK) = 0.
 3600 IMAX(2,KK) = 0.
      KK = KFIN
 3650 CALL IMBEN(J,KK-1)
      IMAX(1,KK) = AMAX1(IMAX(1,KK+1),B(1)+IMAX(2,KK+READY1))
      IMAX(2,KK) = AMAX1(IMAX(2,KK+1),B(2)+IMAX(1,KK+READY2))
      IF(IMAX(1,KK).GT.IMAX(1,KK+1))  C(1,KK) = 1
      IF(IMAX(2,KK).GT.IMAX(2,KK+1))  C(2,KK) = 1
      KK = KK-1
      IF(KK.NE.0) GO TO 3650
C     DETERMINE NODE (NZERO) OF INITIAL DEPARTURE
      IF(NZERO(J).NE.0) GO TO 3700
      IF(IMAX(1,1).GT.IMAX(2,1)) GO TO 3675
      NODE = 2
      GO TO 3720
 3675 NODE = 1
      GO TO 3720
 3700 NODE = NZERO(J)
C     BEGIN FORWARD RECOVERY OF THE SCHEDULE
 3720 NUMDEP = 0
 3725 KK=1
 3730 IF(C(NODE,KK).EQ.0) GO TO 3735
      NUMDEP = NUMDEP+1
      RK = KK-1
      IF(NODE.EQ.2) GO TO 3745
      TIMDEP(J,NUMDEP) = RK*DELT+10.**(-4)
      NODE = 2
      KK = KK+ITIME(1)+ITR(2)
      GO TO 3740
 3735 KK = KK+1
 3740 IF(KK.GT.KFIN) GO TO 3750
      GO TO 3730
```

```
 3745 TIMDEP(J,NUMDEP) = -RK*DELT
      NODE = 1
      KK = KK+ITIME(2)+ITR(1)
      GO TO 3740
C     BELOW,CALCULATE TOTAL NUMBER OF DEPARTURES FROM EACH NODE
 3750 LASDEP(1,J) = 0
      LASDEP(2,J) = 0
      IF(NUMDEP.GT.0) GO TO 3760
C     THE FOLLOWING TWO STATEMENTS CONSTITUTE THE REPRESENTATION OF A
C     NULL VEHICLE SCHEDULE
      LASDEP(1,J)=1
      TIMDEP(J,1)=5.*T
      GO TO 3780
 3760 DO 3775 NDEP=1,NUMDEP
      IF(TIMDEP(J,NDEP).GT.0.) LASDEP(1,J) = LASDEP(1,J)+1
      IF(TIMDEP(J,NDEP).LE.0.) LASDEP(2,J) = LASDEP(2,J)+1
 3775 CONTINUE
 3780 RETURN
      END
```

```
OMPILER OPTIONS - NAME=  MAIN,OPT=01,LINECNT=57,SOURCE,EBCDIC,NOLIST,DECK,LOAD,MAP,NOE

      SUBROUTINE MATCH
      INTEGER EDGEWT,EPS,RT,RTSET
      INTEGER ELIG
      DIMENSION EDGEWT(3000),LIST(3000),LOC(200),NODWT1(200),NODWT2(200)
     1,LABEL1(200),LABEL2(200)
      DIMENSION B(2),IMAX(2,600),DIST(2),SPEED(20),TIMDEP(20,36),
     1LASDEP(2,20),NZERO(20),ELIG(2,100),TR(2)
      COMMON IMAX,B,TIMDEP,LIST,LOC,M1,M2,LAST,NODWT1,NODWT2,
     1TR,DIST,SPEED,KFIN,EDGEWT,LASDEP,T,NZERO,ELIG
C                         *********************
C     EDGES ARE ORDERED BY LIST. LOC(M) GIVES LOCATION IN LIST OF THE
C     FIRST PLACE CORRESPONDING TO AN EDGE EMANATING FROM NODE M IN
C     SET 1. NODWT1 AND NODWT2 GIVE WEIGHTS OF NODES IN SETS 1 AND 2.
C     EDGEWT GIVES WEIGHTS OF EDGES IN ORDER OF THEIR APPEARANCE IN LIST
C     THE LABEL ARRAYS LIST THE BACK-NODES OF NODES IN THE TREE
C                         *********************
C     FIRST FIND LARGEST EDGEWT,MAXWT,IN GRAPH
    5 MAXWT = 0
   10 DO 15 I=1,LAST
   15 MAXWT = MAX0(MAXWT,EDGEWT(I))
C     SET ALL NODE WEIGHTS FOR NODES IN SET 1 EQUAL TO MAXWT.
C     SET WEIGHTS FOR NODES IN SET 2 EQUAL TO ZERO.
C     INITIALIZE LABEL1 AND LABEL2
   20 DO 30 M=1,M1
   25 NODWT1(M) = MAXWT
   30 LABEL1(M) = 0
   35 DO 45 L=1,M2
   40 NODWT2(L) = 0
   45 LABEL2(L) = 0
C     FIND NODE RT IN NODE SET RTSET WHICH HAS NON-ZERO NODE WT AND DOES
C     NOT MEET THE MATCHING. (AN EDGE IN THE MATCHING IS TAGGED WITH A
C     MINUS SIGN IN EDGEWT)
C     FIRST SEARCH NODE SET 1
   50 DO 85 M=1,M1
   55 IF(NODWT1(M).LE.0) GO TO 85
   60 ILIM1 = LOC(M)
      ILIM2 = LOC(M+1)-1
      DO 70 I=ILIM1,ILIM2
   65 IF(EDGEWT(I).LT.0) GO TO 85
   70 CONTINUE
      RT = M
      RTSET = 1
   75 GO TO 155
   85 CONTINUE
C     THERE WAS NO NODE IN SET 1 WHICH HAD NON-ZERO NODE WT AND WHICH
C     DID NOT MEET THE MATCHING - SEARCH NODE SET 2
   90 DO 120 L=1,M2
   95 IF(NODWT2(L).LE.0) GO TO 120
  100 DO 115 I=1,LAST
  105 IF(LIST(I).NE.L) GO TO 115
  110 IF(EDGEWT(I).LT.0) GO TO 120
  115 CONTINUE
      RT = L
      RTSET = 2
  116 GO TO 155
  120 CONTINUE
```

```
  125 RETURN
C     IF 125 IS REACHED,NO NODE EXISTS WITH NON-ZERO WEIGHT THAT DOES
C     NOT MEET MATCHING - THUS OPTIMAL MATCHING IS REACHED
C     NODE RT IN SET RTSET IS A TREE. THE OUTER NODES BELONG TO RTSET,
C     RT IS THE ROOT OF THE TREE
C     IN THE NEXT PART THE EDGES IN THE EQUALITY SUBGRAPH ARE SEARCHED
C     TO SEE IF EVERY EDGE IN THE SUBGRAPH WHICH MEETS AN OUTER NODE OF
C     THE TREE MEETS AN INNER NODE OF THE TREE AT ITS OTHER END
  155 DO 235 M=1,M1
  165 ILIM1 = LOC(M)
      ILIM2 = LOC(M+1)-1
      DO 235 I=ILIM1,ILIM2
  170 LL = LIST(I)
      IEDGE = IABS(EDGEWT(I))
      IF(NODWT1(M)+NODWT2(LL).NE.IEDGE) GO TO 235
C     EDGE I IS IN EQ.SUBGRAPH,CHECK IF IT MEETS AN OUTER NODE,AND IF SO
C     WHETHER IT MEETS AN INNER NODE AT ITS OTHER END
  175 IF(RTSET.EQ.1) GO TO 210
C     RTSET IS NODE SET 2
      IF(LABEL2(LL).EQ.0.AND.LL.NE.RT) GO TO 235
C     LL IS AN CUTER NODE OF THE TREE
  195 IF(LABEL1(M).NE.0) GO TO 235
      NUZERO = LL
      MU = M
      GO TO 250
C     RTSET IS NODE SET 1,EDGE I IS IN EQ.SUBGRAPH,CHECK IF IT MEETS AN
C     OUTER NODE,AND IF SO WHETHER IT MEETS AN INNER NODE AT OTHER END
  210 IF(LABEL1(M).EQ.0.AND.M.NE.RT) GO TO 235
C     NODE M IN NODE SET 1 IS AN OUTER NODE OF THE TREE
      IF(LABEL2(LL).NE.0) GO TO 235
      NUZERO = M
      MU = LL
      GO TO 315
  235 CONTINUE
C     EVERY EDGE IN THE EQUALITY SUBGRAPH WHICH MEETS AN OUTER NODE OF
C     THE TREE MEETS AN INNER NODE OF THE TREE AT ITS OTHER END
C     STATEMENT 550 STARTS THE EPSILON REDUCTION PROCEDURE
  237 GO TO 550
C     BELOW,EDGE I MEETS OUTER NODE NUZERO = LIST(I) OF THE TREE,AND
C     NODE MU = M, NOT IN THE TREE
C     CHECK WHETHER MU MEETS AN EDGE IN MATCHING
  250 LIM1 = LOC(MU)
      LIM2 = LOC(MU+1)-1
      DO 254 I=LIM1,LIM2
  252 IF(EDGEWT(I).GE.0) GO TO 254
      NU1 = LIST(I)
      GO TO 265
  254 CONTINUE
C     MU MEETS NO EDGE IN MATCHING. SET LEAF,THE END NODE IN PATH TO RT,
C     EQUAL TO MU
  258 LEAF=MU
  259 LABEL1(MU) = NUZERO
  260 GO TO 400
C     BELOW,MU MEETS EDGE I IN MATCHING. NU1 IS THE NODE WHICH MEETS I
C     AT THE OTHER END
  265 IF(NODWT2(NU1).NE.0) GO TO 270
  266 LEAF = NU1
```

```
      LABEL2(NU1) = MU
      LABEL1(MU) = NUZERO
  268 GO TO 440
C     EXTEND TREE BY INCLUDING MU AND NU1
  270 LABEL1(MU) = NUZERO
  272 LABEL2(NU1) = MU
  275 GO TO 155
C                          * * * * * * * * * * *
C     EDGE I MEETS OUTER NODE NUZERO = M AND NODE MU = LIST(I),NOT IN
C     TREE,AT ITS OTHER END
C     SEE IF MU MEETS AN EDGE IN THE MATCHING
  315 DO 335 MM=1,M1
  320 LIM1 = LOC(MM)
      LIM2 = LOC(MM+1)-1
      DO 335 II = LIM1,LIM2
  325 IF(LIST(II).NE.MU) GO TO 335
  330 IF(EDGEWT(II).GE.0) GO TO 335
      NU1 = MM
      GO TO 365
  335 CONTINUE
C     MU MEETS NO EDGE IN MATCHING-GET NEW MATCHING BY INTERCHANGING
C     EDGES IN PATH FROM MU TO RT
  350 LEAF = MU
  352 LABEL2(MU) = NUZERO
  355 GO TO 440
C                          * * * * * * * * * * *
C     NODE MU MEETS EDGE II IN MATCHING. EDGE II MEETS NU1 AT OTHER END.
C     SEE IF NU1 HAS NON-ZERO WEIGHT
  365 IF(NODWT1(NU1).EQ.0) GO TO 390
C     NU1 HAS NON-ZERO NODE WT. EXTEND TREE  TO INCLUDE MU AND NU1.
  375 LABEL2(MU) = NUZERO
  380 LABEL1(NU1) = MU
  385 GO TO 155
C                          * * * * * * * * * * *
  390 LEAF = NU1
      LABEL1(NU1) = MU
      LABEL2(MU) = NUZERO
C     NODE WT OF NU1=LEAF IS ZERO. INTERCHANGE EDGES IN PATH FROM LEAF
C     TO RT TO GET NEW MATCHING
C     IN THE FOLLOWING PART,A NEW MATCHING IS FOUND BY INTERCHANGING
C     EDGES IN TREE IN PATH FROM LEAF TO RT. STATEMENT 400 BEGINS CASE
C     WHERE LEAF IS IN NODE SET 1.
  400 LIM1 = LOC(LEAF)
      LIM2 = LOC(LEAF+1)-1
      DO 425 I=LIM1,LIM2
  410 IF(LIST(I).NE.LABEL1(LEAF)) GO TO 425
  415 EDGEWT(I) = -EDGEWT(I)
  420 GO TO 430
  425 CONTINUE
  430 LEAF = LABEL1(LEAF)
  435 IF(LABEL2(LEAF).EQ.0) GO TO 480
C     440 BEGINS CASE WHERE LEAF IS IN NODE SET 2
  440 LEAF1 = LABEL2(LEAF)
  445 LIM1 = LOC(LEAF1)
      LIM2 = LOC(LEAF1+1)-1
      DO 465 I=LIM1,LIM2
  450 IF(LIST(I).NE.LEAF) GO TO 465
```

```
  455 EDGEWT(I)=-EDGEWT(I)
  460 GO TO 470
  465 CONTINUE
  470 LEAF=LEAF1
  475 IF(LABEL1(LEAF).NE.0) GO TO 400
C     THE FOLLOWING STATEMENTS DISCARD THE CURRENT TREE
  480 DO 490 M=1,M1
  490 LABEL1(M) = 0
  495 DO 500 L=1,M2
  500 LABEL2(L) = 0
  505 GO TO 50
C                           * * * * * * * * * * *
C     THE FOLLOWING PART FINDS THE MAXIMUM EPS TO BE USED IN NODE
C     WEIGHT REDUCTION
  550 EPS = 10**5
  560 IF(RTSET.EQ.1) GO TO 620
  565 DO 605 M=1,M1
  570 LIM1 = LOC(M)
      LIM2 = LOC(M+1)-1
      DO 605 I=LIM1,LIM2
  575 LL = LIST(I)
      IF(LABEL2(LL).EQ.0.AND.LL.NE.RT) GO TO 605
  580 EPS = MINO(EPS,NODWT2(LL))
  585 IEDGE = IABS(EDGEWT(I))
      IF(NODWT1(M)+NODWT2(LL).EQ.IEDGE) GO TO 605
  590 EPS = MINO(EPS,NODWT1(M)+NODWT2(LL)-EDGEWT(I))
  605 CONTINUE
  610 GO TO 670
  620 DO 665 M=1,M1
  625 IF(LAPEL1(M).EQ.0.AND.M.NE.RT) GO TO 665
  630 EPS = MINO(EPS,NODWT1(M))
  635 LIM1 = LOC(M)
      LIM2 = LOC(M+1)-1
      DO 650 I=LIM1,LIM2
  640 IEDGE = IABS(EDGEWT(I))
      LL = LIST(I)
      IF(NODWT1(M)+NODWT2(LL).EQ.IEDGE) GO TO 650
      EPS = MINO(EPS,NODWT1(M)+NODWT2(LL)-EDGEWT(I))
  650 CONTINUE
  665 CONTINUE
C                           * * * * * * * * * * * *
C     THE NEXT PART DOES THE NODE WEIGHT REDUCTION
  670 LIMEPS = 0
  680 DO 735 M=1,M1
  685 IF(LABEL1(M).EQ.0) GO TO 730
  690 IF(RTSET.EQ.1) GO TO 705
  695 NODWT1(M) = NODWT1(M)+EPS
  700 GO TO 735
  705 NODWT1(M) = NODWT1(M)-EPS
  710 IF(NODWT1(M).NE.0) GO TO 735
  715 LIMEPS = 1
  720 NU = M
  725 GO TO 735
  730 IF(RT.EQ.M.AND.RTSET.EQ.1) GO TO 705
  735 CONTINUE
  740 DO 800 L=1,M2
  745 IF(LABEL2(L).EQ.0) GO TO 790
```

```
  750 IF(RTSET .EQ.2) GO TO 765
  755 NODWT2(L) = NODWT2(L)+EPS
  760 GO TO 800
  765 NODWT2(L) = NODWT2(L)-EPS
  770 IF(NODWT2(L).NE.0) GO TO 800
  775 LIMEPS = 1
  780 NU = L
  785 GO TO 800
  790 IF(RT.EQ.L.AND.RTSET.EQ.2) GO TO 765
  800 CONTINUE
C                             * * * * * * * * * * * *
C     BELOW,IF EPS IS LIMITED BY AN OUTER NODE WEIGHT BECOMING ZERO
C     (LIMEPS=1),AND IF THAT NODE IS RT,THE MATCHING IS RETAINED AND THE
C     TREE DISCARDED. OTHERWISE THE TREE IS KEPT
  805 IF(LIMEPS.EQ.0) GO TO 155
  815 IF(RTSET.EQ.1) GO TO 835
  820 IF(NODWT2(RT).EQ.0) GO TO 480
  825 LEAF = NU
  830 GO TO 440
  835 IF(NODWT1(RT).EQ.0) GO TO 480
  840 LEAF = NU
  845 GO TO 400
      END
```

```
      SUBROUTINE UPLIM
      INTEGER CAP,ELIG,EDGEWT,FILL,CLASS
      REAL IMAX,N
      DIMENSION JTRIP(20),FILL(20),CLASS(20)
      DIMENSION B(2),NTRAV(2),ELIG(2,100),DIST(2),W(2,100),
     1TSTAR(2,100),OPCOST(20),FARE(20),CAP(20),SPEED(20),EDGEWT(3000),
     2LIST(3000),LOC(200),NODWT1(200),NODWT2(200),IMAX(2,600),TR(2),
     3NZERO(20),TIMDEP(20,36),LASDEP(2,20)
      COMMON IMAX,B,TIMDEP,LIST,LOC,M1,M2,LAST,NODWT1,NODWT2,
     1TP,DIST,SPEED,KFIN,EDGEWT,LASDEP,T,NZERO,ELIG,
     2A1,A2,A3,NTRAV,W,TSTAR,OPCOST,FARE,CAP,SLOPE,DOLLAR,ALPHA,N
     3,NUMVEH
      B1 = 0.
      B2 = 0.
      MAXTRP = 0
      NTYPE = 1
      DO 7050 J=1,NUMVEH
      IF(J.NE.1) GO TO 7005
      CLASS(1) = 1
      GO TO 7010
C     STATEMENT 7005 DETERMINES IF VEHICLE J IS IDENTICAL TO VEHICLE J-1
 7005 IF(SPEED(J).EQ.SPEED(J-1).AND.OPCOST(J).EQ.OPCOST(J-1).AND.
     1FARE(J).EQ.FARE(J-1).AND.CAP(J).EQ.CAP(J-1)) GO TO 7040
      NTYPE = NTYPE+1
 7010 DO 7020 NODE=1,2
      LIM = NTRAV(NODE)
      DO 7020 I=1,LIM
 7020 ELIG(NODE,I) = J*100
      CALL SCHED(J)
      NSTART = NZERO(J)
      IF(NSTART.EQ.0.AND.IMAX(1,1).GT.IMAX(2,1)) NSTART=1
      IF(NSTART.EQ.0.AND.IMAX(1,1).LE.IMAX(2,1)) NSTART=2
      WRITE(6,7030) J,IMAX(NSTART,1)
 7030 FORMAT(1H0,'UPPER LIMIT TO BENEFIT OF VEHICLE',I4,' EQUALS',F10.3)
 7040 CLASS(J) = NTYPE
 7050 B1 = B1+IMAX(NSTART,1)
      WRITE(6,7055) B1
 7055 FORMAT(1H0,' UPPER BOUND B1 =',F10.3)
      DO 7070 J=1,NUMVEH
      JTOT = 2.*T*SPEED(J)/(DIST(1)+DIST(2))+1.
      TOT1 = JTOT
      TOT2 = JTOT/2
C     DETERMINE IF JTOT IS ODD
      IF(TOT1/2..GT.TOT2) GO TO 7060
      JTRIP(J) = JTOT/2
      GO TO 7065
 7060 JTRIP(J) = (JTOT+1)/2
 7065 MAXTRP = MAXTRP+JTRIP(J)
 7070 FILL(J) = 0
C     NOW BEGIN ALLOCATING TRAVELERS TO VEHICLES IN THE MAXIMUM
C     INVENTORY OF TRIPS
      N1 = NTRAV(1)
      N2 = NTRAV(2)
 7075 BJMAX = 0.
      NTOP=0
      DO 7090 J=1,NUMVEH
```

```
      IF(FILL(J).EQ.1) GO TO 7090
      TRAVTM = (DIST(1)+DIST(2))/(2.*SPEED(J))
      WTP = WILPAY(1.,TRAVTM,N,1.,1.,SLOPE,DOLLAR,ALPHA)
      CAPAC = CAP(J)
      BJ = (-A1*OPCOST(J)*(DIST(1)+DIST(2)))/(2*CAPAC)+A2*FARE(J)+A3*WTP
      IF(BJ.LE,BJMAX) GO TO 7090
      BJMAX = BJ
      NTOP = CLASS(J)
 7090 CONTINUE
      IF(NTOP.EQ.0) GO TO 7150
      DO 7145 J=1,NUMVEH
      IF(CLASS(J).NE.NTOP) GO TO 7145
      FILL(J)=1
      LIMJ = JTRIP(J)
      DO 7140 MD=1,LIMJ
      IF(N1.LE.0) GO TO 7100
      LOAD = MINO(CAP(J),N1)
      B2 = B2+LOAD*BJMAX
      N1 = N1-CAP(J)
 7100 IF(N2.LE.0) GO TO 7110
      LOAD = MINO(CAP(J),N2)
      B2 = B2+LOAD*BJMAX
      N2 = N2-CAP(J)
 7110 IF(N1.LE.0.AND.N2.LE.0) GO TO 7150
 7140 CONTINUE
 7145 CONTINUE
      GO TO 7075
 7150 WRITE(6,7155) B2
 7155 FORMAT(1H0,' UPPER BOUND B2 =',F10.3)
      RETURN
      END
```

DEPARTMENT OF COMMERCE

NATIONAL BUREAU OF STANDARDS