R75-27

ADVANCED DIAL-A-RIDE ALGORITHMS: INTERIM REPORT

Nigel H. M. Wilson

Richard W. Weissberg

B. Trevor Higonnet

John Hauser

July, 1975

MIT

DEPARTMENT
OF
CIVIL
ENGINEERING

SCHOOL OF ENGINEERING
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Cambridge, Massachusetts 02139

R75-27

ADVANCED DIAL-A-RIDE ALGORITHMS: INTERIM REPORT

By

Nigel H.M. Wilson
Richard W. Weissberg
B. Trevor Higonnet
John Hauser

Massachusetts Institute of Technology
Cambridge, Massachusetts   02139

July, 1975

TABLE OF CONTENTS

TABLE OF CONTENTS

iii

LIST OF TABLES

LIST OF FIGURES

CHAPTER I
THE ADVANCED ALGORITHMS RESEARCH PROJECT:
AN INTRODUCTION

During the period from 1967 to 1971 significant research was devoted to investigation of the potential use of computers in the control of demand responsive transportation systems. Two of the more tangible outputs of these efforts were

1. A computer simulation model which was developed to test alternative computer control algorithms and to predict system performance, and

2. A recommended set of computer control procedures resulting in the immediate assignment of each request to the current "tour" of the best vehicle, based on

a) Feasibility conditions, under which each user receives service within specified bounds, and

b) Minimization of total service times for current and future passengers.

These control procedures were tested in the simulation model environment and were found to perform well on intuitive grounds (i.e., by examining individual assignments and comparing them with judgment) and to compare favorably with other proposed algorithms. However, since no optimal solution algorithm had been developed (nor has one been developed since), absolute statements about their performance were impossible.

One result of this research program was the decision to mount a demonstration project of the concept in Haddonfield, New Jersey, in order to obtain a market test of the service concept and to obtain data on the potential of computer dispatching. The system (which has been extensively described elsewhere[1]) has just terminated

(1) Report UMTA-NJ-06-0002-74-7, Haddonfield Dial-A-Ride Project Final Report, February, 1975.

its demonstration project phase having provided valuable data in both these areas. The computer control system used in the latter stages of the Haddonfield project was developed by the MITRE Corporation using the control algorithms previously developed at M.I.T.

M.I.T. is now the recipient of a University Research and Training Grant from the Urban Mass Transportation Administration to develop advanced dial-a-ride control procedures based on the experience gained in Haddonfield, and to explicitly investigate the problem of controlling integrated dial-a-ride fixed route transit services. This presents a rare opportunity to evaluate academic research in light of subsequent operational experience, specifically to validate the simulation model and to analyze and improve the operation of the total system. An additional benefit of the Haddonfield experiment has been the collection of extensive data on a similar manual system (identical from a user viewpoint), thus making possible a comparison of the quality of computer assignment with that of manual assignment.

The research objectives of the Advanced Dial-a-Ride Algorithms Research Project can be subdivided into four tasks:

1. Evaluation of simulation effectiveness and upgrading of simulation capabilities so as to ensure the availability of an effective method of assessing the likely effects of changes to the dial-a-ride control techniques or operating environment prior to, or in lieu of, field testing.

2. Evaluation of the present dial-a-ride control algorithm (used in Haddonfield, New Jersey), and identification of shortcomings and areas for improvement.

3. Development of advanced computer control algorithms in the context of single module dial-a-ride systems incorporating better use of contraints, control of service extremes to individual passengers, and scheduling of deferred and periodic demands.

4. Definition, description, and evaluation of roles for computer scheduling in the context of co-ordinated dial-a-ride systems which incorporate interfaces to each other and to existing conventional modes of transportation.

This report describes work accomplished to date on these tasks. It comprises observations, conclusions, and new work based on the Haddonfield demonstration project, as well as more abstract research and improvements which are a logical continuation of prior work not directly connected with Haddonfield. Work on Tasks 1 and 2 has been essentially completed. Chapter 2 discusses the work performed under Task 1. Comparisons are drawn between computer controlled Haddonfield operation and simulation of Haddonfield operation. Also described are modifications to the simulation model to facilitate these tests. Chapter 3 deals with Task 2. In it operational experience with computer dispatching is reviewed from an algorithm viewpoint, and comparisons are drawn between manual and computer dispatching as observed in Haddonfield. Finally shortcomings in the algorithm and areas where the algorithm is being or might be improved are discussed.

Tasks 3 and 4 represent work in progress. Significant progress in eliminating the drawbacks of constraints and improving service extremes to individual passengers has been made through development of the quadratic objective function. Scheduling of deferred and periodic demands is at this time not yet a solved problem, but is a prime target of current research efforts. Encouraging results have been obtained, however, in the area of scheduling of fixed-time stops, which, it is thought, is a necessary preliminary to being able to schedule advanced requests. Chapter 4 discusses the quadratic objective function along with relevant utility theory, and mentions some preliminary results in the area of scheduling of fixed stops.

3

Most of the work accomplished thus far has been con-
cerned with single module systems. Some preliminary ideas
on co-ordinated systems have, however, been formalized
(under Task 4) and are discussed in Chapter 5.

Finally Chapter 6 presents a summary of work accom-
lished and work to be accomplished under the Advanced
Algorithms Research Project.

Five appendices are also included in this document.
Appendix A investigates the distribution of interarrival
times for demands for dial-a-ride service in Haddonfield.
Appendix B investigates algorithm performance as a func-
tion of varying demand levels. In Appendix C are discussed
a proposed procedure to assess consumer preference and a
method for determining parameter settings for the quadra-
tic objective function. The derivation of the quadratic
objective function appears in Appendix D. Finally,
Appendix E describes the system (loosely based on Haddonfield)
which was used in investigating the performance of different
objective functions.

CHAPTER 2
VERIFICATION OF THE SIMULATION MODEL

2.0 Introduction

With the ultimate goal of development of dial-a-ride
assignment algorithms (or automatic control strategies)
that provide the best possible service to the riding public,
a secondary problem arose - that of designing an accurate,
cost effective model. Only by using a mathematical char-
acterization can the analyst perform experiments to evalu-
ate control procedures which would be difficult and/or
expensive to carry out on the real system. The primary
raison d'etre of the simulation model which has evolved to
fill this need is to analyze and test algorithm effective-
ness. It is important, therefore, that the model be a
valid one.

The dial-a-ride computer simulation model developed
at M.I.T. was completed in 1968, and for the past seven
years it has undergone significant expansion and modifica-
tion. A simulation model is necessary to evaluate al-
ternative control procedures because the dial-a-ride system
is too complex to model analytically at this time. The
model is an event-driven simulation which was implemented
in FORTRAN to facilitate implementation of complex decision
rules and compatibility with a wide range of computers.

The major events in the model are the occurrence of
a service request, and a vehicle's making a stop. At
the time each request is generated, a record is established
including request time, origin and destination. As soon
as the request occurs it is assigned by the algorithm to
the future tour of a vehicle (It is this assignment that
is the raison d'etre of the objective function, which is
perhaps the heart of the dial-a-ride control algorithm.).
Whenever a vehicle makes a stop the record for the passen-
ger served (either picked-up or delivered) is updated

with the time of the stop. The vehicle's travel time to
its next scheduled stop is then estimated based on the
direct distance between the two points, a street adjust-
ment factor, and assumed vehicle speed. When a passenger
is delivered, its wait, travel and total service times
are computed and added to the statistics of all other
passengers served, and its record is erased.

2.1 Experience in Haddonfield, New Jersey

Unique opportunities to test, calibrate and improve
the purely simulation aspects of preceding M.I.T. work
were presented by the existence of the Haddonfield demon-
stration project, since the algorithmic procedures em-
bodied in the computer system used in Haddonfield were
derived directly from earlier research at M.I.T. Specifical-
ly, with a minor exception in the area of advanced (or
periodic) requests and an addition in the area of vehicle
in and out of service times, the algorithm used was taken
directly from the previous M.I.T. work. Complete data
describing travel demands in Haddonfield and the conse-
quent performance of the system were gathered, making full
and valid comparisons possible.

Focusing on the design of the simulation model, it
should be noted that, as with any model, numerous assump-
tions and simplifications of the real world were required.
The model, as developed under the previous DOT grant,
was designed to provide the analyst with the ability
to simulate a wide range of systems. To this end the
input parameters include area dimensions, demand rate,
demand pattern, number of vehicles, vehicle size, vehicle
speed, etc. However, there were two major assumptions
of the original model which warranted further investi-
gation in light of Haddonfield operating experience:

    1) A constant number of vehicles are in continuous
       service throughout the simulated period.

2. The demand rate is constant over the simulated period -- although the time between successive demands is selected from a user specified distribution.

2.2 Changes to the Model

In order to investigate the validity of the model two new options have been implemented which allow the two above mentioned assumptions to be relaxed by the analyst. The first model extension provides the analyst with two complementary ways of comparing predicted and observed results:

a) perform a 'pure' simulation by having the model generate as well as assign demands and simulate vehicle movement, and

b) perform a 'replication' by using actually observed demands (in both time and space), limiting the simulation to assignment and vehicle movement.

The primary difference between these two techniques is that (a) requires the user to provide a demand function in the form of statistical distributions for times and end-of-trip coordinates while (b) eliminates the demand function completely, replacing it with a specific set of trip requests. Because the algorithms used in Haddonfield and the M.I.T. model are virtually identical, case (b) amounts to a test of the manner in which the model simulates vehicle movement.

The second option allows vehicles to enter or leave service at any times specified by the analyst or to utilize a constant, continuous supply of vehicles. These options provide significant flexibility and power in validating the simulation model.

2.3 Findings on Modelling Assumptions

Once these options were implemented, simulation experiments were run representing the Haddonfield system using real and simulated demand and vehicle input. Because of the close similarity between the Haddonfield algorithm and the algorithm in the M.I.T. model, most

7

parameters were rather easy to set inasmuch as they had
identical counterparts -- except vehicle speed, the one
major parameter for which there is no direct correspondence.

2.3.1 Variation in Vehicle Speed

For validation it is in fact vitally important to
have an accurate representation of the speeds at which
vehicles actually travel. In order to determine this,
a program was written to extract this information from
Haddonfield transaction tapes. This program, which
computed the speeds of individual vehicles as well as
the overall speed, showed that there was great variation
both among drivers and among days. To illustrate, the
following results were obtained for two days of actual
computer operation:

	19 September 1974	4 October 1974
MEAN	15 mph	12 mph
MINIMUM	12 mph	10 mph
MAXIMUM	18 mph	15 mph

Though simulation results indicate that the average
speed will suffice if it is accurate (all vehicles in
the simulation have the same nominal speed), these numbers
suggest that the speed of individual vehicles could have
a great effect on service for some individual customers.

2.3.2 Vehicles' Entering and Leaving Service

It was found that the assumption of a constant supply
of vehicles in continuous service resulted in significant
over-estimation of vehicle productivity and/or over-estima-
tion of the quality of service which can be provided.
The reason for this is that when a vehicle enters (leaves)
service it is significantly under-utilized in the period
immediately following (preceding) the change. The greater
the number of changes in vehicle status the greater the
overall impact. Moreover, since fully demand responsive
operations occur in the base period of the schedule,

8

vehicle status changes are frequent because of shift changes and driver lunch breaks.

It was found to be difficult to approximate Haddonfield results using the basic unmodified simulation model with the constant number of vehicles equal to the average number of vehicles actually operating. However, by using actual vehicle in-service times, it was possible to closely approximate actual Haddonfield quality of service. Table 2-1 shows the summary statistics for one day of Haddonfield operations. Table 2-2 shows results of four different simulation experiments to reproduce this system. The first and second set of results utilize a constant number of vehicles in continuous service, and demonstrate that similar service could be provided with about 20% fewer vehicles if they were in continuous service. The third case is a simulation using actual vehicle in-service times and shows close correspondence with the actual operation.

TABLE 2-1:

ANALYSIS OF OPERATIONS (9 a.m. - 3 p.m.) 9/19/74

NUMBER OF PASSENGERS	262
VEHICLE PRODUCTIVITY (PASS/VEH/HR)	5
NUMBER OF VEHICLES IN SERVICE	9 - 11
NUMBER OF TIMES VEHICLE ENTERED (LEFT) SERVICE	34

ACTUAL QUALITY OF SERVICE

	MEAN	STD.DEV.	MAX
WAIT TIME (MINS)	9.5	6.0	34
RIDE TIME (MINS)	9.5	5.4	32

9

TABLE 2-2:

SIMULATED <u>QUALITY</u> OF SERVICE

CASE 1:   CONSTANT 8 VEHICLES

	MEAN	STD.DEV.	MAX
WAIT TIME	9.7	6.3	36
RIDE TIME	11.6	7.0	34

CASE 2:   CONSTANT 7 VEHICLES

	MEAN	STD.DEV.	MAX
WAIT TIME	12.0	8.0	54
RIDE TIME	12.5	8.5	59

CASE 3:   VEHICLES IN AND OUT OF SERVICE

	MEAN	STD.DEV.	MAX
WAIT TIME	9.6	8.1	47
RIDE TIME	12.5	8.1	47

CASE 4:   VEHICLES IN AND OUT OF SERVICE
          REPLICATION

	MEAN	STD.DEV.	MAX
WAIT TIME	8.2	7.5	56
RIDE TIME	10.7	7.3	42

Once the variable number of vehicles feature was installed, the model performed well in all cases, with excellent results in replication of Haddonfield (Case 4). Clearly, this indicated that the two significant factors in modelling vehicle movement, i.e., the absence of a street network and the mechanism to simulate the uncertainty of vehicle travel times, were functioning well.

2.3.3 Comparison of Replication and Simulation

Results using generated demands, although moderately good, were not as good as those for replication. In light

of what had been learned about the effects of transient
availability of vehicles, this was only to be expected,
since the inter-arrival rate for generated demands was
based upon only one distribution. It is not a surprising
conclusion that demand in suburban Haddonfield from the
hours of 9 A.M. to 3 P.M. cannot be described with one
probability density function (for a further discussion of
this see Appendix A). Thus, even though performance of
the model was adequate, it was thought desirable to make
more complex specifications of demand rates possible.
To this end the model was modified to allow a user to
specify one or several inter-arrival rate distributions
which are all added together. This solution almost cer-
tainly does not make it possible to closely simulate actual
demand patterns, but it does greatly increase existing
capabilities and specifically permits the user to add
spikes which were previously impossible to simulate.

Other tests gave a preliminary indication that results
were less sensitive to the use of actual spatial data as
opposed to spatial data generated by the model itself.
This can be ascribed to two factors: 'the use in an
approximate way of observed Haddonfield patterns, and the
ability of the trip generator to accept the existence of
different zones with different demand rates (as opposed
to the one distribution for inter-arrival times). Because
only seven zones were used to generate trips, the use of
observed Haddonfield data does not vitiate this result.

Thus it was determined by comparing an actual demand
stream simulation (as obtained from Haddonfield transaction
tapes) with random demands (based on approximations of
the Haddonfield demand pattern) that approximate and random
demands are quite satisfactory for the prediction of sys-
tem performance. This suggests that estimating the
appropriate spatial distribution of demand and level of
demand is sufficient to predict future performance in a

demand responsive transportation system. This is fortunate,
since if this assumption were not valid, prediction of
future systems' performance would have been infeasible.

2.4   Summary and Conclusions

In this chapter, the following conclusions were drawn
concerning the validity of the simulation model:

1. The initial assumption of vehicles' being con-
tinuously in service resulted in significant over-
estimation of productivity and service quality. In
Haddonfield the error due to this assumption repre-
sents a twenty percent change in vehicle requirements.

2. The assumption that an actual street network is
not a necessary element of the model was verified.

3. The mechanism which simulated uncertainty of
vehicle travel times is apparently functioning
well.

4. Estimation of the appropriate spatial distribution
of demand and level of demand is sufficient to
predict future performance for a demand responsive
transportation system.

Although the simulation model was sophisticated by
any standard, it was not, as originally designed, realis-
tic enough to provide reliable estimates of productivity
and service quality. At the time the simulation model was
developed, not enough was known about the transient
behavior of the system to recognize this as a significant
factor. The assumption of a constant and continuous
supply of vehicles resulted in overestimation of quality
of service.

More attention must therefore be directed to system
performance under transient supply conditions, and the
improved model should be used in planning new systems in
conjunction with expected vehicle in-service times. In-
deed the model might now be considered a valuable tool
in planning driver shifts and in investigating ways to
increase productivity by utilizing driver shift information

Such techniques as allowing the driver to be relieved while
the vehicle is in active service should be considered.

CHAPTER 3:

THE ASSIGNMENT ALGORITHM

3.0   Introduction

The computer assignment algorithm which was used in
Haddonfield was the result of the previous research project
at M.I.T. The algorithm, which was first developed in
1968, focuses on the assignment of immediate requests for
service[1]   In the algorithm each passenger is assumed to
have a utility function which is linear in service time
up to a specified constraint. During assignment the origin
and destination of the new request are inserted in the
future tour of each vehicle in service -- all such assign-
ments are considered. An assignment is termed feasible
if as a result no service constraints are violated for
any passenger currently on the system. In selecting
the best assignment any feasible assignment is preferred
to any infeasible assignment, and between alternative
feasible (infeasible) assignments the one which minimizes
the sum of total service time for all current customers
and change in system resources used is selected. Clearly,
depending on the demand level and the setting of the
constraints, constraints on occasion will be violated,
but their purpose is to cut off the tail of long service
times without significantly affecting other measures of
service. In the Haddonfield implementation no diversions
were permitted on the current (first) leg of a vehicle's
tour, no passenger re-assignment was considered, and
there was no empty vehicle policy.

In general the algorithm used in Haddonfield performed
well. Preliminary evaluation indicates that the quality of

(1) Wilson, Nigel, et al, Scheduling Algorithms for a Dial-
A-Ride System, M.I.T., Department of Civil Engineering,
Report Number USL TR-70-13, March, 1971.

14

service provided under computer control was at least as good as that under manual decision-making, and probably somewhat better. This chapter first compares manual and computer dispatching in Haddonfield and then addresses the shortcomings of the computer control algorithm.

3.1 Comparison of Manual and Computer Dispatching

Manual and computer dispatching can only be compared for that service which both have provided: many-to-many service from the hours of 9 A.M. to 3 P.M. on weekdays. From the customers' point of view these two dispatching methods are identical: except inasmuch as different qualities of service may be provided, customers otherwise do not know whether the computer is being used or not. Effective comparison of the data for manual and computer dispatching is limited by three factors:

1. When manual operation was the only mode of operation, the service area and number of vehicles were not the same as when the computer became operational.

2. Once the computer did become operational, manual dispatching was used only when the computer could not function due to hardware or software failures.

3. The computer reached full and effective operational status in early September 1974.

Therefore the statistics to be compared in this report represent every 'immediate' trip (i.e., excluding every deferred trip) for weekdays from 16 September 1974 to 8 October 1974 inclusive (excluding 2 October) for computerized operation, and 11 days of manual operation in July 1974 when the computer was not used because of various air conditioning and hardware problems. Because of the manner in which trips were manually dispatched, manual statistics do not include any trips originating at the PATCO speedline station. In addition there are a small number of trips made between 9 A.M. and 10 A.M. on Satur-

15

days as well as from 9 A.M. to 3 P.M. on Sundays. Because
of the preponderence of weekday trips, it is thought that
the following figures are valid for weekday 9 A.M. to 3 P.M.
trips. Table 3-1 shows the manual many-to-many data by
hour of day, and Table 3-2 shows the data resulting from
computer dispatching for both scatter and many-to-many
trips. It should be stressed that both sets of data refer
to actual trips taken in Haddonfield, not simulation re-
sults. Before discussing these results it should be noted
that both the demand rates and the numbers of vehicles in
service were essentially identical for both manual and
computer operations (hence the productivities[2] of the two
periods were similar).

TABLE 3-1:

MANUAL MANY-TO-MANY STATISTICS

	WAIT TIME			TRAVEL TIME	
time	trips	mean	st.dev.	mean	st.dev.
9-10	402	16.5	9.2	9.9	6.6
10-11	370	15.6	10.8	9.6	6.0
11-12	359	16.7	11.1	10.7	5.8
12-1	431	19.7	13.3	11.4	7.5
1-2	451	19.6	12.0	11.6	8.7
2-3	624	18.4	10.6	12.4	8.6
Total(9-3)	2637	17.9	11.3	11.1	7.5

TABLE 3-2:

COMPUTER DISPATCHING RESULTS

	WAIT TIME			TRAVEL TIME	
	trips	mean	st.dev.	mean	st.dev.
from PATCO	853	10.6	9.3	13.0	8.3
others	2658	14.9	11.1	11.3	8.2

(2) Productivity is here measured in passenger trips per
vehicle hour.

It can be seen in the two tables that the service provided by the computer to passengers not starting from PATCO (i.e., the service for which manual statistics are available) has a wait time significantly lower than that provided by manual dispatchers. In all other respects computer and manual dispatching performance are very similar. It is interesting to note the excellent service provided passengers originating at PATCO (figures for manual dispatching are not available for these passengers): The mean wait time for such passengers was a full four minutes lower than that for other trips, and standard deviation was also lower. The slightly higher travel time for passengers starting at PATCO is due to a greater average trip length than that of the remaining passengers: 1.56 miles versus 1.36 miles (air-line distance). These results are encouraging for the continued use of the computer in demand responsive systems. In particular it is now clear that the computer can perform at least as well as manual dispatchers, even in small systems (10-12 vehicles) operating at low demand levels (40-60 demands per hour). Furthermore, as discussed in the following sections, it is clear that further improvement in computer dispatching is feasible which would yield clearly superior performance by computer dispatching.

3.2 Improving the Algorithm

Based on operational experience in Haddonfield improved performance might be achieved through concentration on the following problem areas:

1. Inflexibility of hard constraints

2. Objective function not a true reflection of customer utility

3. Handling of advanced and periodic requests

4. Inability to constrain vehicle position at future time

5. Inability to restrict certain vehicles to given zones

6. Re-assignment capability

7. Problems scheduling at start and end of driver/ vehicle shift

8. Algorithm not geared to under-utilized system.

Each of these areas are described briefly here in addition to more complete discussions of some of them elsewhere in this report.

3.2.1 Inflexibility of Hard Constraints

The algorithm was designed to minimize total service time (for current and future passengers) within fixed constraints on wait, travel, and total service times. Any assignment in which no constraint is violated is preferred to any assignment involving a violation, independent of the value of the objective function. This policy was developed to reduce the number of passengers experiencing "unreasonably long" service times, with the acknowledged and expected effect of some increase in the mean service times (see Figure 3-1). To achieve this goal the constraints must be set about 100% above the mean service times. In practice two problems arise from this approach:

a) Because the short-run demand rate varies widely over the course of the day, and because mean service times are very sensitive to the recent demand rate, a constraint set correctly for some time of the day may be incorrect for many other times of the day. The problem is that the constraints are not dynamically set as a function of the number of passengers currently on the system and the number of vehicles currently in service. This problem could be solved by using a short-memory heuristic to compute the current constraint set.

b) More basic is the problem that assignments which may be far superior from the objective function's viewpoint will be rejected if a constraint is violated. This introduces a perturbation in performance and can lead to short-sighted decisions which tend to

18

FIGURE 3-1:

RELATIONSHIP BETWEEN MAXIMUM SERVICE TIME
AND MEAN SERVICE TIME AS A FUNCTION OF CONSTRAINT VALUE

MAXIMUM TIME

MEAN TIME

CONSTRAINT VALUE
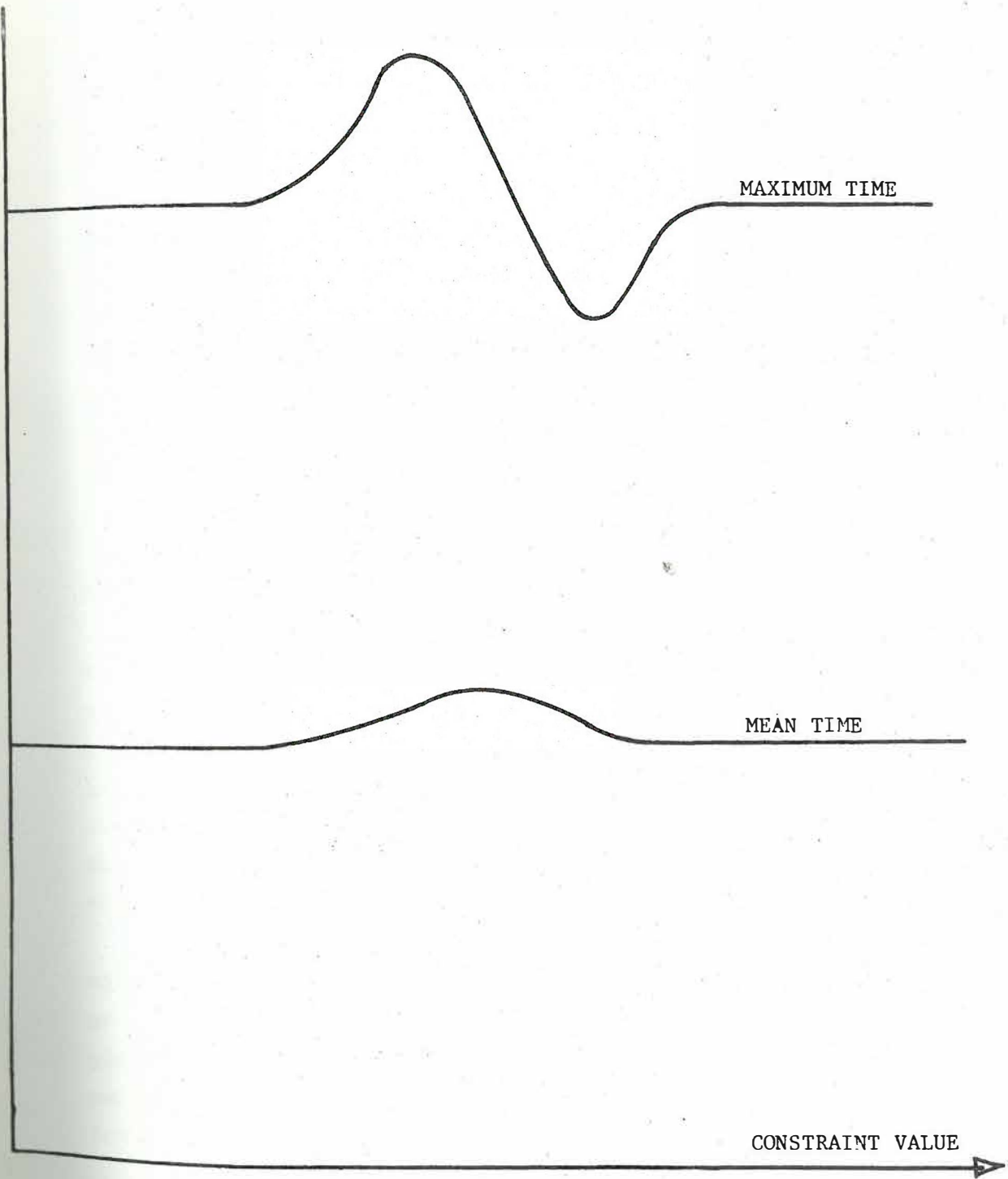
waste system resources. This problem cannot be solved by any useful setting of the constraints, and its existence argues for a reduction in the role of constraints in future algorithm development work. This is possible only if the individual customer utility function can be equally or better represented by some other construct (see below).

3.2.2 Objective Function Not a True Reflection of Customer Utility

The current objective function implies that users of the system associate with the service a utility function which is linear in service time. This may be an inaccurate and simplistic representation of actual passenger satisfaction, and hence its use might result in customer dissatisfaction. Although the actual utility function associated with dial-a-ride service has not yet been identified, it is clear that for the distribution of service time other measures than the mean are also important -- e.g., standard deviation. It is probable that the uncertainty in service is also an important characteristic. One measure of this is the difference between estimated and actual pickup and delivery times. Once again the means and standard deviations of these distributions should be considered.

It is clear that different customers will have different utility functions. For example, someone who is going to work or transferring to a scheduled bus will be very conscious of the latest arrival time. Another person arriving from a scheduled bus or leaving work will be very conscious of earliest pickup time. Thus a range of different passenger utilities should be able to co-exist simultaneously in the algorithm.

It is highly likely that actual utility functions will vary not only from customer to customer, but from area to area. For these reasons it is important that the next generation of algorithms incorporate a richer mix of elements in the objective function and provide the user (operator) with ways to manipulate the objective function

to achieve desired service characteristics. With a more
representative objective function, the service constraints
can be used more as a means to reduce computation (by elim-
inating unpromising assignments early), rather than as
an integral part of the algorithm. This problem is addressed
in the next chapter.

3.2.3 Handling of Advanced and Periodic Requests

At present advanced requests (this term also being
used to include periodic requests) are assigned a fixed
period before their desired pickup time with a special
set of (tight) constraints, using a modified objective
function which attempts to minimize the time between
expected and desired pickup time. All subsequent assign-
ments to a tour including an advanced request are made
as if the tour consisted of only immediate service requests.
This results in service for the advanced request being no
better than service for immediate requests -- an unsatis-
factory state of affairs since advanced requests should be
easier to schedule and serve than immediate requests.
This is an important area of future work.

3.2.4 Inability to Constrain Vehicle Position at Future
       Time

The present system was designed for the dynamic many-
to-many case for which scheduled and/or repetitive demands
on the system are not a major factor. In actual opera-
tion there will frequently be a need for vehicles to make
regularly scheduled or one time appearances at specific
locations, even though no originating service requests
have been made, e.g., PATCO station in Haddonfield for
scatter operations. This capability has been implemented
in the simulation model and is currently being evaluated
(see Section 4.6).

21

3.2.5 Inability to Restrict Certain Vehicles to Given Zones

For ease of use at high-density demand generators, it may be desirable to specify service zones so that passengers know immediately which vehicle serves their destinations -- each vehicle can then post one or more zone numbers. For this operational technique to be compatible with computer dispatching, the computer system must be able to restrict a vehicle to serve only limited origin/destination pairs. This capability does not exist in the Haddonfield system, but recently M.I.T. has implemented a scheme whereby vehicles can be restricted in terms of the origins and/or destinations served in the simulation model.

3.2.6 Re-assignment Capability

The Haddonfield computer system does not have a passenger reassignment capability except in the situation where a vehicle breaks down, in which case the tour (including both collected and uncollected passengers) is shifted to the end of the tour of the vehicle which can be the first to reach the breakdown point. Passenger reassignment is an element of the algorithm which was investigated previously by M.I.T. and found to be of only marginal benefit. This work was performed in the context of constant demand rates and vehicle supply (not to mention the absence of other external transients such as vehicle failures, etc) as discussed in Section 2.1. Now that the importance of transient effects has been realized, this area should be re-examined, if only for the handling of vehicles which break down or suffer unusual delays en route.

3.2.7 Problems Scheduling at the Start And End of Driver/ Vehicle Shift

As discussed in connection with driver lunch breaks and the starts and ends of shifts (Section 2.3.2) there is a need for the computer to be able to build up tours efficiently and stop further assignments at specific times so as to

22

maximize system productivity.

3.2.8 Algorithm not Geared to Under-utilized System

The previous algorithm development research was geared
heavily to system (and hence algorithm) performance at or
near the point of maximum system utilization. This resulted
in higher vehicle productivities than typically observed
in Haddonfield, and therefore the algorithm has been oper-
ating at much lower productivities than previously studied.
It now appears, both from observations in Haddonfield and
from simulation experiments, that the algorithm may not
perform most effectively in this situation. Specifically,
the increase in the tour length term in the objective func-
tion may lead to significant imbalances in utilization of
vehicles. There is a high probability of a new request's
being assigned to an already highly utilized vehicle, and
moreover, once a vehicle becomes unassigned it tends to
remain so. It must be recognized that the best objective
function may well be a function of the current utilization
of the system. This is discussed in detail in Appendix B.

3.3 Summary

From operational experience gained in Haddonfield with
the existing computer control algorithms the following gen-
eralizations may be made:

1. Computer dispatching in its current state can be
more effective than manual dispatching even for small
systems (e.g., 10 square miles) operating at low
productivities (about 5 passenger trips per vehicle
hour).

2. Further improvements can be made in computer
control techniques without exceeding computational
constraints.

In Chapter 4 two approaches are examined which have been
implemented to improve the existing control procedures.

23

CHAPTER 4:
IMPROVING EXISTING COMPUTER CONTROL TECHNIQUES

4.0 Introduction

Under the current DOT grant M.I.T. is investigating
means of further improving the basic dial-a-ride control
algorithm used in Haddonfield as suggested by the discussion
in the previous chapter. In this chapter two approaches
will be described which have received most attention
to date on the research project. The first is the use of a
quadratic rather than linear objective function within the
basic algorithm framework. Most of the chapter will be
devoted to this research (Sections 4.1 - 4.6). Additionally
the concept of the fixed time stop and its potential use
to improve service for certain types of passengers is in-
troduced (Section 4.7).

The area of investigation most advanced to date is
the use of a quadratic objective function aimed at elimina-
ting the inflexibility of hard constraints, thus providing
a more desirable service and eliminating the performance
sensitivity to system loading. In Sections 4.1 - 4.6 the
relationship between individual utility functions and the
algorithm objective function is explored and preliminary
results from simulation experiments using a quadratic
objective function are presented.

The choice of the objective function is one of the
more important design considerations in the implementation
of Wilson's[1] dial-a-ride routing algorithm. The objective
function, which is used to synthesize various measures of
performance associated with each insertion of a demand in a
vehicle subtour, establishes a criterion for selecting the
"best" assignment.

(1) Wilson, Nigel, et al, Scheduling Algorithms for a
     Dial-A-Ride System, M.I.T. Department of Civil Engineering,
     Report Number USL TR-70-13, March, 1971.

4.1 Form of the Quadratic Objective Function

The quadratic objective function is a "second generation" equation which was developed in an attempt to eliminate weaknesses discovered during the first actual implementation of the original - linear - objective function in Haddonfield (see Section 3.2). Based on the theory that customer dissatisfaction increases quadratically rather than linearly as a function of time delay and/or estimation error, the quadratic objective function incorporates terms in wait time, travel time, wait time deviation and delivery time deviation. The quadratic objective function consists of four separate equations of similar form -- each corresponding to one level-of-service parameter -- and a fifth equation concerned with conservation of system resources for the service of future requests. These equations (which are derived in Appendix D) are presented below. The value of a specific assignment is the weighted sum of the values of these five functions.

1. Wait Time

$$Z(WT) = (NP(P))(dP)^2 + (NP(D))(dP + dD)^2 + 2(dP)(TWT(P)) + 2(dD)(TWT(D)) + (WT(NEW))^2$$

where

$NP(P)$ = Number of pickups between new pickup and delivery insertion

$NP(D)$ = Number of pickups after new delivery insertion

dP = Detour due to insertion of new pickup

dD = Detour due to insertion of new delivery

$TWT(P)$ = Total wait time for pickups after new pickup

$TWT(D)$ = Total wait time for pickups after new delivery

$WT(NEW)$ = Wait time for new passenger based on this assignment.

2. Wait Time Deviation

$$Z(WTD) = (NP(P))(dP)^2 + (NP(D))(dP + dD)^2 + 2(dP)(TWD(P)) + 2(dD)(TWD(D))$$

TWD(P) = Total wait time deviation for all pickups after new pickup

TWD(D) = Total wait time deviation for all pickups after new delivery.

3. Travel Time

$$Z(TRT) = (NO(P))(dP)^2 + (NO(D))(dD)^2 + 2(NB(P,D))(dP)(dD)$$
$$+ 2(dP)(TRT(P)) + 2(dD)(TRT(D)) + (TT(NEW))^2$$

where

NO(P) = Number of passengers on board at new pickup

NO(D) = Number of passengers on board at new delivery

NB(P,D) = Number of passengers on board at both new pickup and new delivery

TRT(P) = Total travel time for those on board at new pickup

TRT(D) = Total travel time for those on board at new delivery

TT(NEW) = Travel time for new passenger

4. Delivery Time Deviation

$$Z(DTD) = (ND(P))(dP)^2 + (ND(D))(dP + dD)^2 + 2(dP)(TDD(P))$$
$$+ 2(dD)(TDD(D))$$

where

TDD(P) = Total delivery time deviation for deliveries after new pickup

TDD(D) = Total delivery time deviation for deliveries after new delivery

ND(P) = Number of deliveries after new pickup but before new delivery

ND(D) = Number of deliveries after new delivery

5. System Resources

$$Z(SR) = (dP) + (dD)$$

The total objective function is then:

$$Z = A(Z(WT)) + B(Z(WTD)) + C(Z(TRT)) + D(Z(DTD)) + E(Z(SR)).$$

The parameters A, B, C, D, and E determine the relative weights of the five level-of-service measures. As might be expected, different sets of relative weights are appropriate for different classes of users. The following sections discuss the development of the quadratic objective function and the reasoning behind the form described above. As mentioned above, the purpose of an objective function is to synthesize various measures of performance associated with each insertion of a demand in a vehicle subtour. It seems appropriate therefore, to begin the discussion with an attempt to identify what "ideal" performance criteria might be.

4.2 Ideal Criteria

One potential criterion might be "to maximize some measure of goodness for the current and future passengers of the service." More rigorously, the algorithm would compute for each current passenger and estimate for each future rider a scalar measure of goodness, i.e., utility with the property that the larger the utility for a particular option, the more that option is preferred. It would then synthesize these individual utilities into a group utility. Keeney[2] shows that under reasonable assumptions such a group utility can be computed by summing individual utilities, i.e.,

$$U_g = \Sigma \, k_i u_i$$

where

U_g = group utility
k_i = weight for individual i
u_i = utility for i^{th} individual

(2) Keeney, Ralph. <u>A Group Preference Axiomatization with Cardinal Utility.</u> Presented at the 46th Annual Operations Research Society of American Conference in San Juan, Puerto Rico, October 16-18, 1974.

Another criterion might be "for a given cost, maximize patronage." In other words, determine how the consumer makes his choice of travel and provide service in such a way as to make dial-a-ride as favorable as possible within the constraints. Hauser and Urban[3] show that an important step in determining consumer response is to identify individual utility functions which compact the measures of performance into a scalar measure of goodness. Again under reasonable assumptions, the probability of choice of a mode increases as the utility of that mode increases. The overall objective is to maximize the market share of dial-a-ride, but an approximation would be to maximize the increase in average utility, where

$$\bar{U} = \sum u_i / N$$

and N = number of individuals
     U = average utility

Both criteria discussed above indicate the need to identify individual utility functions. Such (multi-attributed) utility functions combine the performance measures, such as wait time, travel time, pickup deviation, dropoff deviation, into a single number. In doing so, it explicitly incorporates risk, tradeoffs, and interdependencies among the performance measures. Various utility theorists such as Keeney, Fishburn, and Ting identify conditions under which a multi-attributed utility function can be decomposed into a sum of uni-attributed utilities. I.e., let

$x_1, x_2, \ldots, x_K$ = performance measures

$u(x_1, x_2, \ldots, x_K)$ = multi-attributed utility function

$u_k(x_k)$ = uni-attributed utility for performance measure x_k

then

$$u(x_1, x_2, \ldots x_K) = \sum_{k=1}^{K} \lambda_k u_k(x_k).$$

In this particular form, the λ_k's identify the tradeoffs among the performance measures and the risk preference (reliability) is incorporated in the uni-attributed utilities.[4]

(3) Hauser, John and Glen Urban. A Normative Methodology for Modeling Consumer Response to Innovation. M.I.T. Operations Research Center Technical Report 109, May, 1975.
(4) An additive form assumes no interdependency among the performance measures.

Since most people are risk averse, i.e., prefer reliability, the functions, $u_k(x_k)$, are strictly concave. In other words, they have the following shape:

$u_k(x_k)$
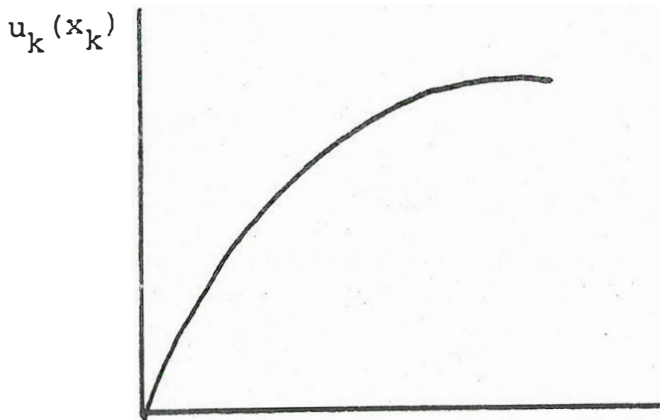
FIGURE 4-1:

CONCAVE UNI-ATTRIBUTED UTILITY FUNCTION

In essence, the above theoretical arguments imply the following:

    (1) An ideal objective function would calculate individual utilities for current passengers, estimate them for future passengers, and determine an overall measure of goodness by summing the individual utilities.

    (2) Individual utilities can be approximated by summing transformations of performance measures such as wait time, travel time, etc.

    (3) These transformations, $u_k(x_k)$, are to have the shape indicated in Figure 4-1 (strictly concave).

4.3 Consideration of Computer Constraints

The preceding section described what an ideal objective function would be if there were no limitations on computer capabilities. Because assignments and insertions must be made in "real time" there are severe limitations on the complexities of the individual utility functions. These limitations imply that individual utilities cannot be computed separately for all passengers, but must be

29

inferred from some aggregate measure which is simply up-
dated for each potential assignment and insertion.

An example will help to clarify this concept: Suppose
that one is concerned only with waiting time, and suppose
that the insertion under consideration increases everyone's
wait time by dw. Then the new utility sum must be easily
computed from the old utility sum, i.e.,

$$u_{old} = \sum_{i=1}^{N} u(w_i)$$

where

$$w_i = \text{individual i's wait time}$$

becomes

$$u_{new} = \sum_{i=1}^{N} u(w_i + dw).$$

In other words, $u_{new} - u_{old}$ must be easily computed:

$$u_{new} - u_{old} = \sum_{i=1}^{N} (u(w_i+dx) - u(w_i))$$

Wilson[5] shows that if $u(w_i) = b(w_i) + c$ then this
constraint is satisfied. Later work in connection with
the Advanced Dial-A-Ride Algorithms Research Project has
shown that a more general requirement is:

$$u(w_i+dw) - u(w_i) = (aw_i+b)dw_i$$

This implies that:

$$u(w_i) = a'w_i^2 + bw_i + c$$

(Note that this reduces to Wilson's objective function
when a' = 0.)

This objective function is the quadratic objective

(5) Wilson, Nigel, et al, Scheduling Algorithms for a
    Dial-A-Ride System, M.I.T. Department of Civil
    Engineering, Report Number USL TR-70-13, March, 1971.

function, which is believed to be the most complex form
possible within the current limits on computing cost.

4.4 Advantages of the Quadratic Objective Function

The previous sections argued for an objective function
which approximated a group utility function and showed that
a quadratic form is the most general form useable at reason-
able costs in the existing algorithm. The question arises:
"Can a group utility function be reasonably approximated
by a quadratic objective function?" Consider the three
criteria given at the end of Section 4.2 (1,2, and 3).

Criterion 1 (summing individual utilities) and criterion
2 (additive individual utilities) are automatically satis-
fied by any functional form for $u(x_j)$ and are naturally
satisfied for quadratic objective functions. Criterion 3
(strictly concave) is satisfied for quadratic objective
functions if a' is negative.

In the interest of making the model more intuitively
understandable, we change the uni-attributed utility
function to a more familiar form -- disutility:

disutility = -utility

Performing this transformation and defining the scale such
that no wait time is equal to zero disutility we have the
following uni-attributed "cost" function, where now the al-
gorithm tries to minimize total consumer "cost."

$$c_k(x_k) = -u_k(x_k) + constant$$
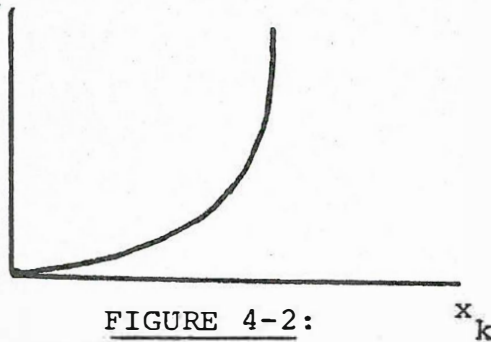
FIGURE 4-2: x_k
QUADRATIC COST FUNCTION

31

Interpreting this curve we note that it has the property that large deviations from an "ideal" wait time are much more onerous (more heavily weighted) that small deviations. This property, which is equivalent to risk aversion in the utility function, tends to favor reliability at the expense of average travel time. In other words, it would tend to favor the "reliable" distribution of wait times in Figure 4-3a over the "unreliable" distribution in Figure 4-3b.
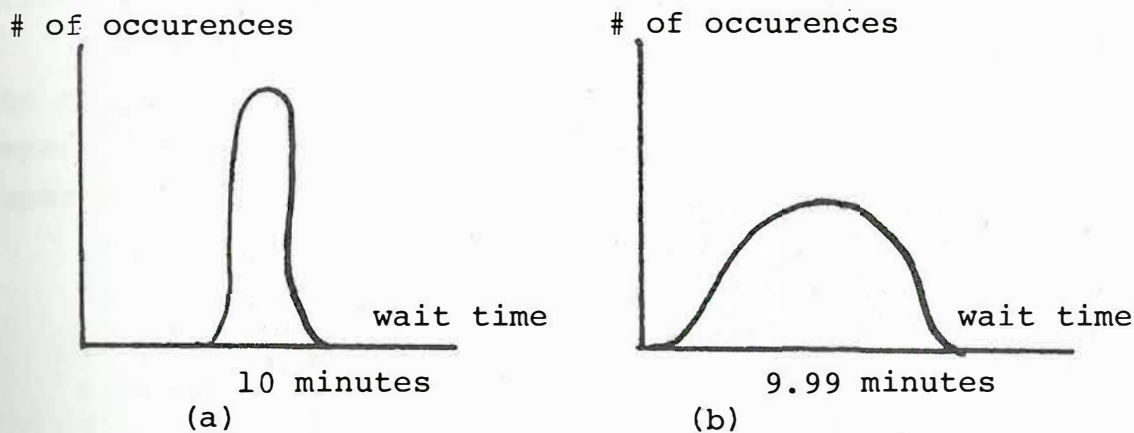
FIGURE 4-3:

<u>RELIABLE AND UNRELIABLE DISTRIBUTIONS OF WAIT TIME</u>

Notice that a linear cost function[6], would ignore the spread (measured by the variance) in the performance measures and always chose the alternative with the smallest mean. Given the distributions in Figure 4-3, it would choose the unreliable distribution.

<u>4.5 The Quadratic Objective Function: Summary</u>

Thus far we have suggested that (1) a quadratic objective function is a good approximation to a true consumer utility function, (2) a true utility function would be the ideal objective function based on social welfare and maxi-

_____

(6) A linear function is concave, but not strictly concave.

patronage criteria, (3) the quadratic objective function can increase reliability (and thus avoids providing extremely bad service to a few individuals) and (4) a quadratic objective function can be implemented in the existing computer model.

The issue of determining parameter settings for the quadratic objective function is addressed in depth in Appendix C. In the next section some preliminary experiments with the quadratic objective function are described.

4.6 Testing the Quadratic Objective Function

In order to gain familiarity with the properties of the quadratic objective function a series of simulation experiments were made. The primary objectives of the experiments were

1. To compare the performance of the linear and quadratic objective function.

2. To determine the extent to which the type of service provided can be controlled by different settings of its parameters.

3. To determine the sensitivity of the best parameters for the quadratic at different demand levels.

4. To compare the performance of the linear and quadratic objective function under varying system loads.

Before describing the results obtained, the characteristics of the quadratic objective function and the specific system simulated should be described. For this first series of experiments, two simplifying assumptions were made. First, only one set of weight coefficients (one class of passenger) was assumed. Second the objective function included only the pure quadratic terms in wait time and travel time and the term in increase in tour length. The terms in wait time deviation and delivery time deviation were not included because they require specification of the advertised pickup and delivery times at time of assignment. Prediction of advertised times adds significant complications to the

assignment process and so has not been addressed to date. This problem will be addressed in later stages of this research and at that time the deviation terms will be considered for inclusion in the objective function. It should be stressed that their exclusion at this time in no way invalidates results obtained with the quadratic; we should bear in mind, however, that further improvements may be achieved when the deviation terms are also included.

The parameters which controlled the setting of the quadratic objective function were defined as:

A = weight given to quadratic term in wait time

C = weight given to quadratic term in travel time

E = weight given to increase in tour length

In all experiments with the quadratic form the value of C was set equal to 0.1 and the values of A and E varied.

In order to obtain results which would provide the most insight into the performance of different objective functions all simulation experiments utilized similar input conditions except for the demand rate and the objective function. The system simulated, which is described in detail in Appendix E, is loosely based on Haddonfield. Specifically, the spatial distribution of demand was taken directly from Haddonfield although vehicles were assumed to be in continuous service and continuous knowledge of vehicle location was also assumed. While the results to be presented are accurate for this situation, it is not certain that similar results would be obtained in different situations. Work to validate these preliminary results will be completed in the remainder of this project.

In the absence of individual utility functions estimated as described in the previous section, settings of the parameters in the quadratic objective function were selected to reflect approximately equal weighting of

passenger wait and travel time.

Two demand levels were used to determine the best parameter settings for the quadratic objective function: approximately 40 demands per hour, which closely approximated the Haddonfield demand rate, and 60 demands per hour. Tables 4-1 and 4-2 present the results obtained for a range of parameter settings for the quadratic objective function as well as for the linear form for both demand levels. The following conclusion can be drawn from these results:

1. In all cases the quadratic objective function reduced the standard deviation of service times over the linear form.

2. In some cases at the lower demand level the quadratic objective function produced improvements in the mean service times.

3. In all cases the quadratic objective function produced lower deviation in pickup and delivery times.

4. The performance of the quadratic objective function is not very sensitive to the setting of its parameters.

5. The best settings of the parameters of the quadratic objective function are not very sensitive to demand level.

6. Significant control of the ratio of wait time to travel time can be achieved through varying the parameter settings for the two terms in the quadratic objective function without significant degradation in the overall quality of service.

7. No significant improvement in the overall quality of service can be expected from the quadratic form over the linear form.

8. At higher demand rates the linear form seems to perform as well as the quadratic form.

To further investigate the performance of the quadratic objective function, a replication of Haddonfield was run with both linear and quadratic objective functions, no continuous knowledge of vehicle location and vehicles entering and leaving service. These replication results

35

TABLE 4-1: PARAMETERS FOR QUADRATIC OBJECTIVE FUNCTION: DEMAND RATE=40/hr

Objective Function	WAIT M	WAIT ST.D.	TRAVEL M	TRAVEL ST.D.	TOTAL M	TOTAL ST.D.	PICKUP DES M	DELIV DES M
LINEAR	8.6	6.1	10.9	6.6	19.5	9.9	2.1	4.2
A=.13 E=2	8.9	5.1	10.0	4.8	18.9	7.5	1.8	4.0
A=.16 E=2	8.8	5.2	10.5	5.3	19.3	8.1	1.9	4.0
A=.19 E=2	8.6	5.3	10.6	5.8	19.2	8.2	2.0	3.4
A=.16 E=1	10.1	5.8	10.6	5.8	20.7	8.6	2.1	3.5
A=.16 E=2	8.8	5.2	10.5	5.3	19.3	8.1	1.9	3.0
A=.16 E=2.5	8.9	5.1	10.4	5.4	19.3	8.0	1.9	3.2
A=.16 E=3	8.6	5.0	10.8	5.8	19.4	8.0	2.0	3.4

TABLE 4-2: PARAMETERS FOR QUADRATIC OBJECTIVE FUNCTION: DEMAND RATE = 60/hr.

Objective Function	WAIT M	WAIT ST.D.	TRAVEL M	TRAVEL ST.D.	TOTAL M	TOTAL ST.D.	PICKUP DES M	DELIV DES M
LINEAR	12.0	8.2	12.9	7.4	24.9	11.3	2.6	4.8
A=.07 E=2	15.0	7.6	12.0	6.0	27.0	9.4	2.3	3.6
A=.1 E=2	14.8	8.1	13.0	6.5	27.8	10.3	2.4	3.8
A=.13 E=2	13.3	7.7	13.7	6.9	27.0	10.4	2.2	4.1
A=.16 E=2	13.0	7.1	14.1	7.1	27.1	9.8	2.3	4.3
A=.19 E=2	12.6	7.2	15.0	7.8	27.6	10.5	2.3	4.5

TABLE 4-3: REPLICATION OF HADDONFIELD

	WAIT M	WAIT ST.D.	TRAVEL M	TRAVEL ST.D.	TOTAL M	TOTAL ST.D.	PICKUP DES M	DELIV DES M
Actual Operation	9.5	6.0	9.3	5.4	18.8	---	---	---
Linear Form	8.2	7.5	10.7	7.3	18.9	11.	2.1	3.8
Quadratic Form: A=.13,C=.1,E=2	8.9	6.3	10.1	5.8	19.0	9.9	1.8	3.3

TABLE 4-4:

CONSTRAINT PERFORMANCE AT DEMAND RATE = 40/hr.

	WAIT			TRAVEL			TOTAL			PICKUP	DELIVERY
	M	ST.D.	MAX	M	ST.D.	MAX	M	ST.D.	MAX	DEV.	DEV.
Unconstrained Linear	6.6	5.0	29.5	11.6	6.7	37.8	18.2	8.9	46.6	1.7	1.8
Constrained Linear*	7.0	5.2	23.5	14.3	6.5	32.9	18.2	8.3	43.2	1.8	2.5

*Wait Time Constraint = 25 mins; Travel Time Constraint = 1.6 * Direct Travel Time +8;
Total Time Constraint = 1.7 * Direct Travel Time +20.

TABLE 4-5:

PERFORMANCE AT VARYING DEMAND RATE

DEMAND RATE = 40/80/40 PER HOUR

	WAIT			TRAVEL			TOTAL			PICKUP	DELIVERY
	M	ST.D.	MAX	M	ST.D.	MAX	M	ST.D.	MAX	DEV.	DEV.
Unconstrained Linear	10.8	9.9	57.0	13.4	9.3	50.8	24.3	14.0	74.0	2.4	4.8
Constrained Linear*	13.5	12.2	68.7	13.9	9.4	59.5	27.4	16.0	98.6	2.9	5.2
Quadratic with A=.13 C=.1 E=2	12.1	9.4	46.5	13.5	8.2	59.1	25.6	13.6	77.0	2.2	4.5

are presented in Table 4-3 together with the results of actual operation. These results support the previous conclusion that the quadratic form performs slightly better than the linear form actually used in Haddonfield.

A final series of experiments was run to compare the performance of the linear and quadratic forms in a situation where the demand rate varied over the course of the experiment. In this case experiments were also run on the linear form with constraints set effectively for the lower demand rate to determine the extent to which poorly set constraints could affect service during a period of variable demand. The simulation was for a six hour period in which the first two and last two hours had a demand rate of 40 per hour and the middle two hours had a demand rate of 80 per hour.

First the best setting of the constraints for 40 demands per hour was determined by running unconstrained and constrained experiments at that demand rate. The results showing the level of constraints selected and the impact on service are shown in Table 4-4 with the maxima also shown. As expected the constraints can be effective in reducing the maximum service times without adversely affecting other measures of service.

Table 4-5 shows the results of constrained linear, unconstrained linear and the quadratic under the varying demand rate condition. The results show that the unconstrained linear and the quadratic perform very similarly, with the linear having a slightly lower mean and the quadratic having a slightly lower standard deviation. As before, the quadratic is superior in both pickup and delivery time deviations. Also as suspected the constrained linear with constraints geared to the lower demand rate does significantly worse than either of the other two -- underlining the danger of using constraints without an adaptive process for changing them as the state of the system changes.

38

4.7 Fixed Time Stops

During operation of the Haddonfield computer control system, it became clear that a control capability was needed which would guarantee a vehicle to be present at a specified location at a given time. This capability has been implemented and tested in the simulation model with preliminary results described in this section. As currently implemented fixed time stops to be used in a simulation experiment must be specified at the start of an experiment: they are not dynamically assigned like a normal service request. With this restriction fixed time stops may be used to encourage the development of certain types of tours which might be expected to improve the overall quality of service. Any vehicle can be given many fixed stops (at different times) at the same or different locations, the requirement on subsequent tours being that the vehicle must be available to leave the fixed stop at the specified time.

Fixed time stops which are pre-assigned might be used to

1. Ensure that passengers transferring from a fixed route bus to dial-a-ride would always have a dial-a-ride vehicle waiting.

2. Operate zonal service in which all vehicles meet at a common transfer point at regular intervals.

3. Ensure that passengers in a remote part of the service area frequently have a vehicle available.

If fixed time stops are also allowed to be dynamically scheduled then they might also be used to

4. Ensure that dial-a-ride passengers transferring to a scheduled fixed route service arrive in time to catch the bus.

5. Ensure that advanced (pre-booked) requests are not picked up before their earliest pickup time.

39

Before describing some initial results obtained from the use of fixed time stops it should be noted that in general use of fixed time stops will result in some idle vehicle time at every such stop. Thus extensive use of fixed time stops may well result in increased unproductive vehicle time. For this reason fixed time stops will be used only sparingly and where desired service characteristics dictate their use.

The rather dramatic improvement in level of service that is achievable by incorporating fixed time stops in the dial-a-ride control algorithm is illustrated in Table 4-6. Two simulation runs were made, both using vehicle and demand characteristics similar to those of Rochester, N.Y., to determine the effectiveness of using prescheduled fixed time stops to improve service for passengers transferring from fixed route service to dial-a-ride. Service demands from transferring passengers are considered advanced requests. In Run #1, there are no fixed time stops. In Run #2, each vehicle has a scheduled fixed time stop at the line-haul transfer point at the time it is to come in and out of service, if that time is near to the time of one of the line haul departures. Additional fixed stops were added to ensure that every line-haul vehicle would be met by at least one dial-a-ride vehicle.

TABLE 4-6:* EFFECT OF FIXED TIME STOPS

| | Transferring Passengers | | | | Many-to-Many | |
| | Fixed Route-DAR | | DAR-Fixed Route | | | |
	Wait	Travel	Wait	Travel	Wait	Travel
Run #1	11.1	16.8	10.0	16.6	13.1	19.3
Run #2	0.6	14.2	10.5	13.9	13.5	18.8
Percentage of Total Passengers	10%		10%		80%	

*All figures are means in minutes.

In this experiment, a significant improvement in level of service to passengers transferring from fixed route service to dial-a-ride is achieved by use of fixed time stops <u>without significantly reducing service to other</u> customers. It is believed that further experimentation with fixed time stops will lead to improved service for passengers transferring from dial-a-ride to fixed route service as well.

Work is currently underway to evaluate the potential of fixed time stops in the other situations mentioned previously.

CHAPTER 5

THE CONTROL OF INTEGRATED DIAL-A-RIDE, FIXED ROUTE SYSTEMS

5.0 Introduction

To date, most dial-a-ride systems have been implemented
in relatively small geographic areas, but as dial-a-ride
becomes more popular, there will be a demand for services
which encompass entire counties. Although it is possible
that the best system for such areas would be a simple single
module dial-a-ride, there are other, coordinated systems
worth investigating. On the one hand control problems
are more complex, but on the other hand it becomes possible
to consider more capital intensive control options because
of the larger number of vehicles involved.

Investigation of coordinated systems is one of those
projects on the frontier of advanced algorithms research.
The simulation program has been modified to facilitate
simulation of certain simply coordinated systems, but no
substantive testing of algorithms has yet been performed.
In this chapter an initial control algorithm for a basic
coordinated system is presented.

The basic philosophy behind coordinated systems is to
serve "local" trips directly by dial-a-ride and to serve
longer distance trips by a sequence of dial-a-ride and
line haul. To do this, the geographic area is partitioned
into dial-a-ride zones and a separate line haul system is
provided for travel between zones, as shown in Figure 5-1.

Although a continuum of zonal and/or line haul options
are possible, three __basic__ zonal options and four __basic__
line haul options can be identified. They are:

Zonal Options

(1) Autonomous dial-a-ride zones.

(2) Overlap: some sharing of zones and/or communica-
tion between zones.

(3) Rendezvous: some trips between contiguous dial-a-ride zones to be served by rendezvous of dial-a-ride vehicles.

Line Haul Options

(1) Fixed route and fixed schedule between fixed zonal terminals.

(2) Fixed route between zonal terminals with dynamic dispatch from route terminals.

(3) Dynamically scheduled and routed between zonal terminals. Routes restricted to arterial network.

(4) Other:

(a) dynamically scheduled and routed with possible route deviation to serve single demands.

(b) dumbbell tours: i.e., some tours allowed with all origins in one zone and all destinations in another.

(c) dynamically scheduled and routed with no network or terminal points defined.

This chapter defines and presents a general algorithm for the least complex and most basic option (and, thus, the easiest to implement in a real-world situation): autonomous dial-a-ride zones/fixed route-fixed schedule line haul.
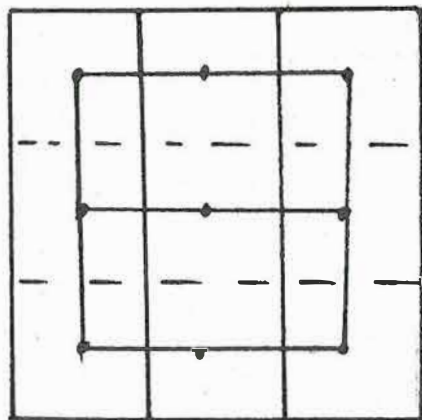

FIGURE 5-1:

COORDINATED SYSTEMS

5.1 System Definition for the Basic Coordinated System

The zones in this option are autonomous -- that is, they are mutually exclusive (no overlap) and collectively exhaustive (together cover the entire area). The control center for each zone communicates with the master control center regarding interfaces with the line haul system, but it retains complete control over all dial-a-ride trips within the zonal boundaries. Furthermore, it is unaware of the state of the system in other dial-a-ride zones.

The line haul system is a fixed (pre-planned) route and fixed scheduled bus system operating on an arterial road network connecting fixed terminals in each zone. To avoid degradation in line haul travel time, the number of terminals within each zone is limited preferably to one or two. This limitation in number also allows the algorithm to examine each terminal more carefully when routing interzonal demands.

5.2 A General Algorithm

For ease of display and understanding, various notational definitions are used in presenting the algorithm. These are summarized below. Besides this notation, four subroutines are referred to in the algorithm; a short discussion of each appears after the algorithm.

The basic algorithm assumes that each demand requests service by phone. This restriction was made for ease of presentation. Other options are allowed, such as park and ride or hailing of the line haul vehicles. Explanations of how each of these alter the algorithm are presented after the discussion of the subroutine.

Finally, to encompass a wider variety of situations, two options are provided in step A. Option A1: if the telephone exchange boundaries are coincident with dial-a-ride zone boundaries, an automatic telephone switching sys-

44

tem can be utilized to direct the request for service directly to operators at zonal centers. Option A2: requires telephone operators at a central location rather than in each zone.

Notation Summary

Z_o = origin zone, i.e., the zone in which the demand originates

Z_d = destination zone, i.e., the zone to which the demand is going

i = index for terminals in the origin zone

j = index to terminals in the destination zone

$C(Z_o)$ = controller for zone Z_o

$C(Z_d)$ = controller for zone Z_d

MC = master control

(x_o, y_o) = coordinates of the point at which the demand originates

(x_d, y_d) = coordinates of the point at which the demand terminates

(x_i, y_i) = coordinates of the ith terminal in the origin zone

(x_j, y_j) = coordinates of the jth terminal in the destination zone

t = time at which request for service is made

$t_o(i)$ = feasible time at which a demand from (x_o, y_o) can arrive at terminal i, i ε Z_o. [estimated by $C(Z_o)$]

$t_{1h}(i,j)$ = the time <u>from</u> $t_o(i)$ <u>until</u> arrival at terminal j, j ε Z_d, given that the demand arrives at terminal i, i ε Z_o at $t_o(i)$. [estimated by MC]

$t_d(j)$ = the estimated time from arrival at terminal j, j ε Z_d until arrival at (x_d, y_d), given that the demand arrives at terminal j at $t_o(i) + t_{1h}(i,j)$. [estimated by MC]

45

t_s, T_s = "safety" times, constants provided for the algorithm

$S(i)$ = a measure of system slack used up if $C(Z_o)$ assigns demand to terminal i, $i \in Z_o$ [estimated by $C(Z_o)$].

C = generalized notation for constraints
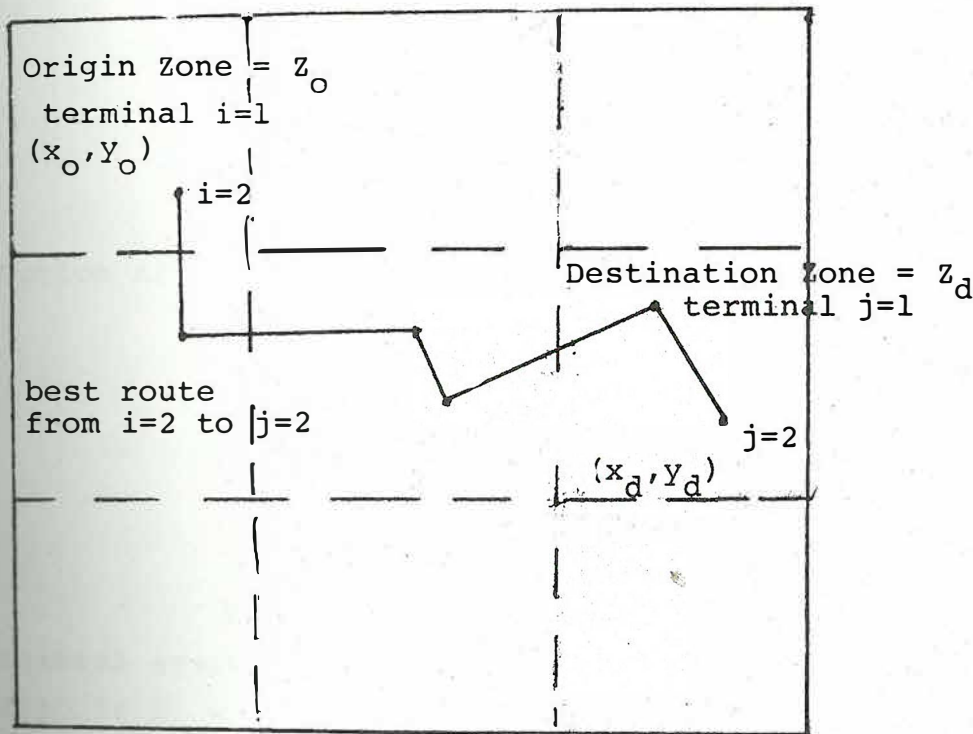
FIGURE 5-2:
NOTATION
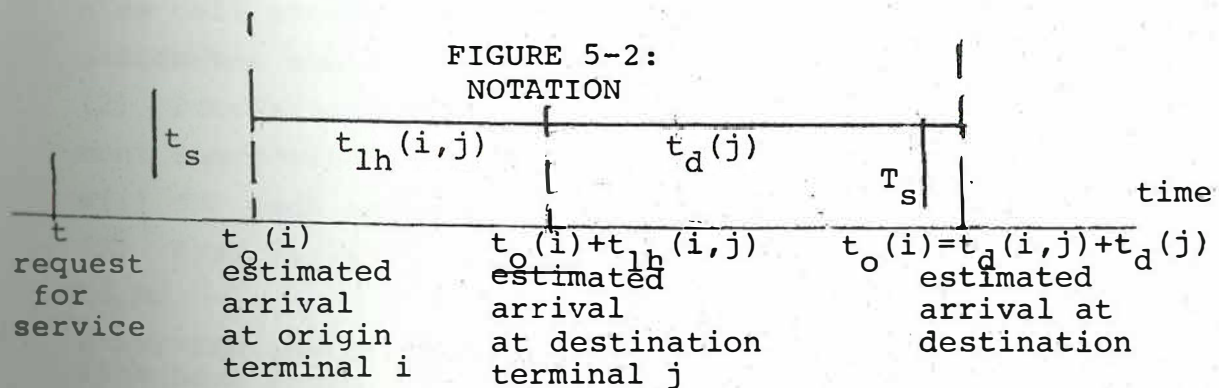
FIGURE 5-3:
TIME AXIS

Algorithm Description

	Controller
Demand with $(x_o, y_o, x_d, y_d, t, c)$	

A. Call reception

Option A1: (1) Automatic telephone switching system transfers call directly to $C(Z_o)$ based on telephone exchange number   *telephone switching then $C(Z_o)$*

(2) Operator at Z_o's control center records data and inputs it into $C(Z_o)$

(3) $C(Z_o)$ determines Z_d

Option A2: (1) Operator at master control center receives call, records $(x_o, y_o, x_d, y_d, t, c)$ and inputs data into MC computer   *MC*

(2) MC determines Z_o and Z_d then transfers data to $C(Z_o)$

B. Initial assignment in origin zone   *$C(Z_o)$*

(1) If $Z_o = Z_d$ then go to step D, else call provisional dial-a-ride assignment subroutines.

(2) Provisional dial-a-ride assignment subroutine returns $t_o(i)$ and $s(i)$ for each terminal i, $i \varepsilon Z_o$.

(3) Pass $(x_d, y_d, Z_d, c, t_o(i), s(i))$ to MC

C. Determination of optimal terminals and line haul route   *MC*

(1) For each terminal i in zone Z_o and for each terminal j in zone Z_d,

47

determine the line haul assignment
that gets the demand to j at the
shortest possible time after $t_o(i)$.
The assignment is constrained be-
cause it cannot pick up the demand
before $t_o(i)$. Thus, estimate
$t_{lh}(i,j)$.

(2) For each terminal j in zone
Z_j, estimate $t_d(j)$.

(3) Determine "best" terminals in
Z_o and Z_d where the "best" set
of terminals minimize F:[1]

$$F = \min_{i,j} f(s(i),t_o(i),t_{lh}(i,j),t_d(j),t,\text{etc})$$

(4) Pass the following data to $C(Z_o)$

    (a) (x_o,y_o,c)

    (b) optimal terminal i

    (c) $t_o(i)$: arrival time constraint

    (d) consumer information including

        -line haul vehicle to be met

        -destination terminal

        -expected arrival times at

         terminal and at destination

(5) Pass the following data to $C(Z_o)$

    (a) (x_d,y_d,c)

    (b) terminal j

    (c) arrival time: $t_o(i)+t_{lh}(i,j)$

D. Permanent assignment in origin zone

(1) If $Z_o = Z_d$, then call permanent
dial-a-ride assignment with

(1) One example of an objective function is a linear function
of the variables, i.e., $F = \min_{i,j} \{ a * s(i) + b * [t_o(i) + t_{lh}(i,j) + t_d(j) - t] \}$.

48

origin   = (x_o, y_o)

destination = (x_d, y_d)

constraints = t, c

else remove provisional assignments to
all terminals except for "best" i,
and call permanent dial-a-ride assign-
ment subroutine with

origin   = (x_o, y_o)

destination = (x_i, y_i)

constraints = t, c

      hard constraint that arrival
      time at terminal i be $\leq t_o(i)$
      $- t_s$

(2) If $Z_o = Z_d$, then end;
else go to D(3).

(3) If dial-a-ride assignment in
D(1) is feasible, then go to E.
Else return to B(2) and update
$t_o(i)$ for all $i \in Z_o$.

E. Initial assignment in destination zone

(1) Call provisional dial-a-ride
assignment subroutine with

origin   = (x_j, y_j)

destination = (x_d, y_d)

constraints = c

      arrival at (x_j, y_j) to be after
      $t_o(i) + t_{1h}(i, j)$

(2) Pass earliest possible pick-up
time to MC.

F. Monitor line haul for late arrival

(1) If it appears assigned line haul
will arrive at j later than the time
given in E(2), then notify $C(Z_d)$ for
update.

Else at time $t_o(i) + t_{1h}^*(i,j) - T_s$,

MC is sure of arrival time at

$j(t_{1h}^*(i,j)$ is updated version)

(2) Return either updated time

or final time to $C(Z_d)$.

G. Permanent assignment in destination zone

If time passed from F is a final

time, then change the provisional

assignment to a permanent assign-

ment and end;

Else remove the old provisional

assignment and call provisional

dial-a-ride assignment subroutine

with origin = (x_j, y_j)

destination = (x_d, y_d)

constraints = c

   arrival at (x_j, y_j) to be after

   $t_o(i) + t_{1h}^*(i,j)$

and return to F.

5.2.1 Subroutines

(1) __Permanent dial-a-ride assignment__: This is simply the normal dial-a-ride assignment package currently in use.

(2) __Provisional dial-a-ride assignment__: This subroutine is very similar to the normal dial-a-ride assignment package except that it flags origin-destination insertions for easy removal at a later time. Since it is only provisional, it can use some heuristics which are more approximate than the permanent dial-a-ride assignment subroutine.

(3) Determination of $t_{lh}(i,j)$: This subroutine is given the origin terminal i, the destination terminal j, and an arrival constraint at the origin of $t_o(i)$. It searches the schedules of the line haul system to determine the routing that gets the demand to j as soon after $t_o(i)$ as possible, but under the constraint that the demand cannot be picked up before $t_o(i)$.

(4) Determination of $t_d(j)$: This subroutine is given the destination terminal j and the destination (x_d, y_d). It estimates the time from arrival at j to arrival at (x_d, y_d) in one of two ways:

(4-1) Call provisional dial-a-ride assignment with origin (x_j, y_j) and destination (x_d, y_d).

(4-2) $C(Z_d)$ maintains a service time matrix, $ST_d = \{st_{j1d}\}$, and a dynamic service speed value, s_d, where:

$$st_{j1d} = \frac{1}{n} \sum_{k=1}^{n} \quad (k^{TH} \text{ previous service time from terminal j to subzone } Z_{1d}).$$

$$s_d = \frac{1}{n} \sum_{k=1}^{m} \quad (k^{TH} \text{ previous "speed" incurred in zone } Z_d)$$

and

- Z_{1d} are mutually exclusive, collectively exhaustive partitions of Z_d.

- speed = the time from pick-up by dial-a-ride vehicle in zone Z_d to drop-off by dial-a-ride vehicle in zone Z_d divided by distance travelled.

51

The MC maintains a ST_d matrix and S_d value for all zones which are updated at finite time intervals. Then

$$t_d(j) = a * st_{jld} + b * S_d * \text{(distance from } (x_j, y_j)$$
$$\text{to } (x_d, y_d)) \text{ where } (x_d, y_d) \in Z_{ld}, \ a + b = 1,$$
$$\text{and } a, b > 0.$$

5.2.2 Alterations to Algorithm

I. <u>Request for service originates at origin zone terminal:</u>

To encourage such requests, a toll-free telephone should be placed at each terminal with a direct line to operators either at $C(Z_o)$ or MC. The caller states his destination, the terminal he is currently at, and any constraints.

The destination zone, Z_d, is determined, the algorithm is entered at C, and $t_o(i)$ is set as follows:

$$t_o(i) = \begin{cases} t & \text{if } i = \text{terminal he is at} \\ +\infty & \text{otherwise} \end{cases}$$

Otherwise the algorithm remains the same. (Step D is of course skipped.)

II. <u>Request for service originates at destination terminal:</u>

In some circumstances a rider will plan his own trip via the line haul facility and request service directly from the toll-free telephone at the destination terminal. In this case the algorithm simply calls the permanent dial-a-ride assignment subroutine with origin (x_j, y_j) and destination (x_d, y_d).

II. <u>Request for service originates by hailing line haul vehicle:</u>

If there is no on-board communication, the person must wait until he arrives at the destination terminal to request ervice. If there is on-board communication the driver communicates the destination to the MC. The algorithm is entered at C with the following changes:

(a) Z_d must be determined before entry.

(b) $t_{lh}(i,j)$ is re-defined to be the time required to travel from the pick-up point to terminal j, j ε Z_d.

(c) Since the origin terminal need not be determined, step C3 is changed to:

Determine the "best" terminal in Z_d where the best terminal minimizes G:

$$G = \min_{j} \quad g(t_{lh}(i,j), t_d(j), t, \text{etc.})$$

(d) Consumer information no longer requires origin information.

(e) Step D is skipped.

IV. **Alteration in line haul terminals due to delay in line haul vehicles:**

Clearly, if delays are incurred on line haul vehicles, it may be optimal in terms of F to alter the assignment of terminals. This addition is not included in the algorithm because it is felt that such changes in terminal assignments would have devastating effects on consumer perception of the service. Thus the line haul vehicle is not monitored for late arrival at the origin terminal nor is any change considered in the destination terminal assignment once it is made. (Steps F and G)

CHAPTER 6:

SUMMARY AND FUTURE RESEARCH

6.1 Simulation Modelling

The simulation capability which existed at the outset of the Project has been strengthened as follows:

1. For the first time the simulation model has been validated by comparing actual operating results from an existing dial-a-ride system with results of simulations of that system.

2. The model now operates with vehicles' entering and leaving service and with demand rates and patterns varying over the course of the day as in real operation.

6.2 Advanced Algorithms

To date much of the research on the project has been directed to this area with the principal results being

1. Full evaluation of the performance of the linear objective function operating in the Haddonfield Dial-A-Ride Demonstration Project.

2. The development, implementation, testing and evaluation of quadratic objective functions designed to eliminate the use of constraints in the scheduling algorithm.

3. The development, implementation, and testing of a new algorithm component designed to require a vehicle to be at a specified location at a stated time (fixed time stops).

4. The development of a procedure for estimating an individual's utility function for dial-a-ride service.

5. The development of a procedure for facilitating transfers between fixed routes and dial-a-ride using fixed time stops.

6. The development of a procedure for assigning advanced requests for service.

7. The development of a control procedure for integrated service including several dial-a-ride areas interconnected by fixed routes.

6.3 Future Research

In the remainder of this project stress will be on the completion of these areas of algorithm design which have been started but not yet carried to evaluation:

1. The quadratic objective function will be completed with incorporation of the wait time deviation and delivery time deviation terms. It will be tested in a range of situations to verify the tentative conclusions reached to date.

2. The use of fixed time stops will be further explored, particularly as a means to facilitate one-to-many service from either transfer points or high activity generators.

3. Use of special objective functions for transfer passengers will be explored as a means to further improve their service.

4. Use of special objective functions, fixed stops and hard constraints will be used for improving service to advanced request passengers.

5. Direct dial-a-ride vehicle to dial-a-ride vehicle transfers will be evaluated as a means for providing service between adjacent dial-a-ride service areas.

Looking further ahead it is now becoming clear that further attempts to improve service through optimization of the objective function are rapidly reaching the point of diminishing returns. It will become increasingly important to investigate other approaches to improving service

such as

1. Stop resequencing, passenger reassignment and passenger transfer between vehicles.

2. Incorporation of expected (but not yet known) future requests in the assignment process.

3. Obtaining bounds on best possible service to determine how close the algorithm is to optimal service.

4. Adaptive control techniques for selecting algorithm parameters as the system state changes.

5. Implementation and evaluation of control procedures for integrated service including several dial-a-ride areas inter-connected by fixed route bus.

With substantial research accomplished since the end of the Haddonfield Dial-a-Ride Demonstration Project it is now important to obtain real operational experience with these new techniques. In the forthcoming Rochester Integrated Demonstration Project a real opportunity to get this necessary experience will be provided and it is expected that the new approaches developed in the Advanced Algorithms Research Project will be implemented and evaluated in the demonstration. The Haddonfield Project demonstrated that computer control could be superior to manual dispatching, and the fruits of that experience should result in further improvement and expansion of capabilities in Rochester.

ACKNOWLEDGEMENT

57

TESTING THE DISTRIBUTION OF INTERARRIVAL TIMES

The existence of substantial amounts of data on ob-
served demand patterns in Haddonfield provided an oppor-
tunity to test some of the modelling assumptions used
in the simulation model. In particular, this data made
possible a testing of the commonly used Poisson assump-
tion for the distribution of demands through time.
Additionally, if the Poisson assumption appeared untenable,
this data provided the basis for the building of a more
sophisticated model of demand arrivals.

A sample of four days' demands was selected for
testing, two days from June, 1974, and two from early
September. For the present purposes, these will simply
be referred to as Day 1 through Day 4.

The method of testing was to compare distributions of
interarrival times (times between demands) with exponential
distributions having the same mean as the collected data.
If and only if the distribution of demands is Poisson,
the distribution of interarrival times is exponential.
Figures A-1 through A-4 illustrate the observed distribu-
tions of interarrival times, grouped into intervals of
width 30 seconds. Superimposed over these histograms
are the probability density functions of exponential
random variables with mean equal to the observed sample
mean in each case. Even this sort of crude visual
testing indicates that there are differences in the
degree to which the days conform to the Poisson assumption.
This observation is confirmed by the results of a
Kolmogorov-Smirnov (K-S) test to exponentially distribute
interarrivals. Table A-1 summarizes the K-S results,
where $P(x|exp)$ is the probability that an interarrival
distribution which is truly exponential would give rise
to the observed data.

TABLE A-1:

SUMMARY OF K-S TEST RESULTS FOR THE FOUR DAYS

Day	# Demands	Mean Interarrival Time (mins)	$P(x|exp)$
1	218	1.59	∿.4
2	195	2.37	∿.1
3	180	1.91	<.01
4	196	1.84	∿.1

While the Possion model is a reasonable one for Day 1, it is quite unreasonable for Day 3, and questionable for Days 2 and 4. A glance at Figures A-1 through A-4 indicates that the major discrepancies are in the first two frequency classes, i.e., those arrivals less than one minute apart. While very few calls are taken less than 30 seconds apart, a large portion of the calls are between 30 seconds and 60 seconds apart. This sort of observation contradicts the Poisson model, but may be more related to the "buffering" between the actual time of call and time of entry to the computer than to the interarrival process itself.

An additional point illustrated by the results in Table A-1 is that the mean arrival rate varies significantly from day to day.

Apart from the Poisson assumption, which is the most common way of modelling random arrival patterns, a typical characteristic of demand on transportation systems is periodic variation through the day. With this motivation, a series of tests for periodic components in the interarrival times was conducted. This series was largely unsuccessful with respect to finding significant periodic components in the variation of inter-arrival times, with the simple exception of Day 2, which exhibited statistically significant components with

periods 1, 2, 10, and 20 hours. Even for this day,
the periodic components do not explain a great deal
of the variation in interarrival times (nor would one
expect them to), but they are statistically significant.

A number of other possible models for the interarrival
process were tested including lognormal, gamma, autore-
gressive, and (in conjunction with the periodic model)
a model proposed by Fishman and Kas.[1] None of these
models fit the data significantly better than did the
Poisson, in general, although in certain specific cases
they did provide better fits.

The conclusions to be drawn from these comparisons
are:

1. The interarrival rate distributions and mean
arrival rates differ from day to day.

2. The Poisson model is probably adequate for
most purposes. (Results of simulation of
Haddonfield operations indicate low sensitivity of
model results to the distribution of arrivals.)

(1) Fishman, G.S., and E.P.C. Kas, "Arrival Generators for
Queueing Simulations," Yale University Technical Report
28-5, July, 1974.

FIGURE A-1:

DISTRIBUTION OF INTERARRIVAL TIMES: DAY 1

Observations = 218
Minimum Interarrival Time = 0.00
Maximum Interarrival Time = 13.40
Mean Interarrival Time = 1.59

Observations = 195
Minimum Interarrival Time = 0.00
Maximum Interarrival Time = 20.10
Mean Interarrival Time = 2.37

FIGURE A-2:

DISTRIBUTION OF INTERARRIVAL
TIMES: DAY 2

Frequency

Observations = 180
Minimum Interarrival Time = 0.00
Maximum Interarrival Time = 12.50
Mean Interarrival Time = 1.91

FIGURE A-3:

DISTRIBUTION OF INTERARRIVAL
TIMES: DAY 3

Interarrival times

63

Frequency

Observations = 196
Minimum Interarrival Time = 0.00
Maximum Interarrival Time = 10.50
Mean Interarrival Time = 1.84

FIGURE A-4:

DISTRIBUTION OF INTERARRIVAL
TIMES: DAY 4

Interarrival times
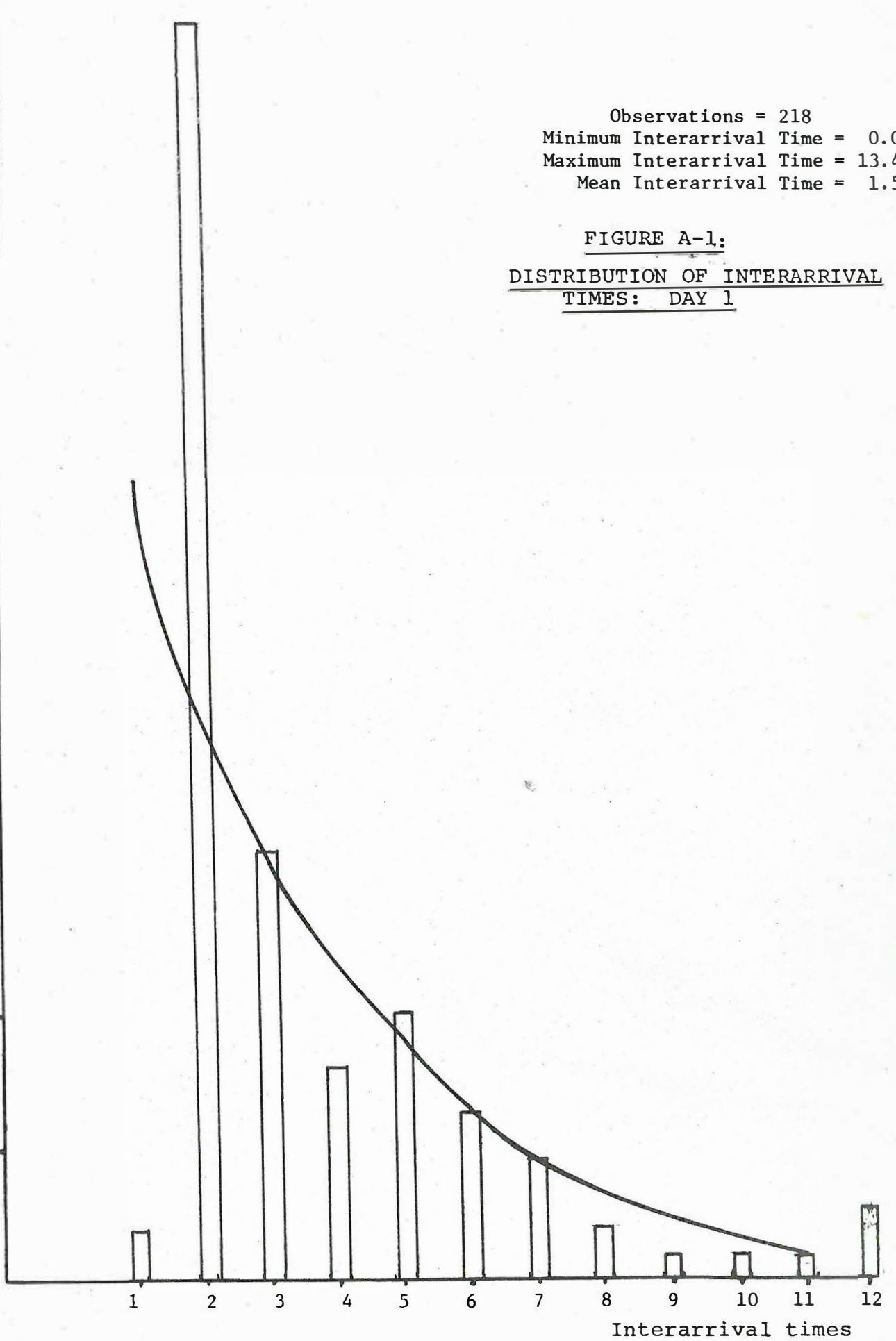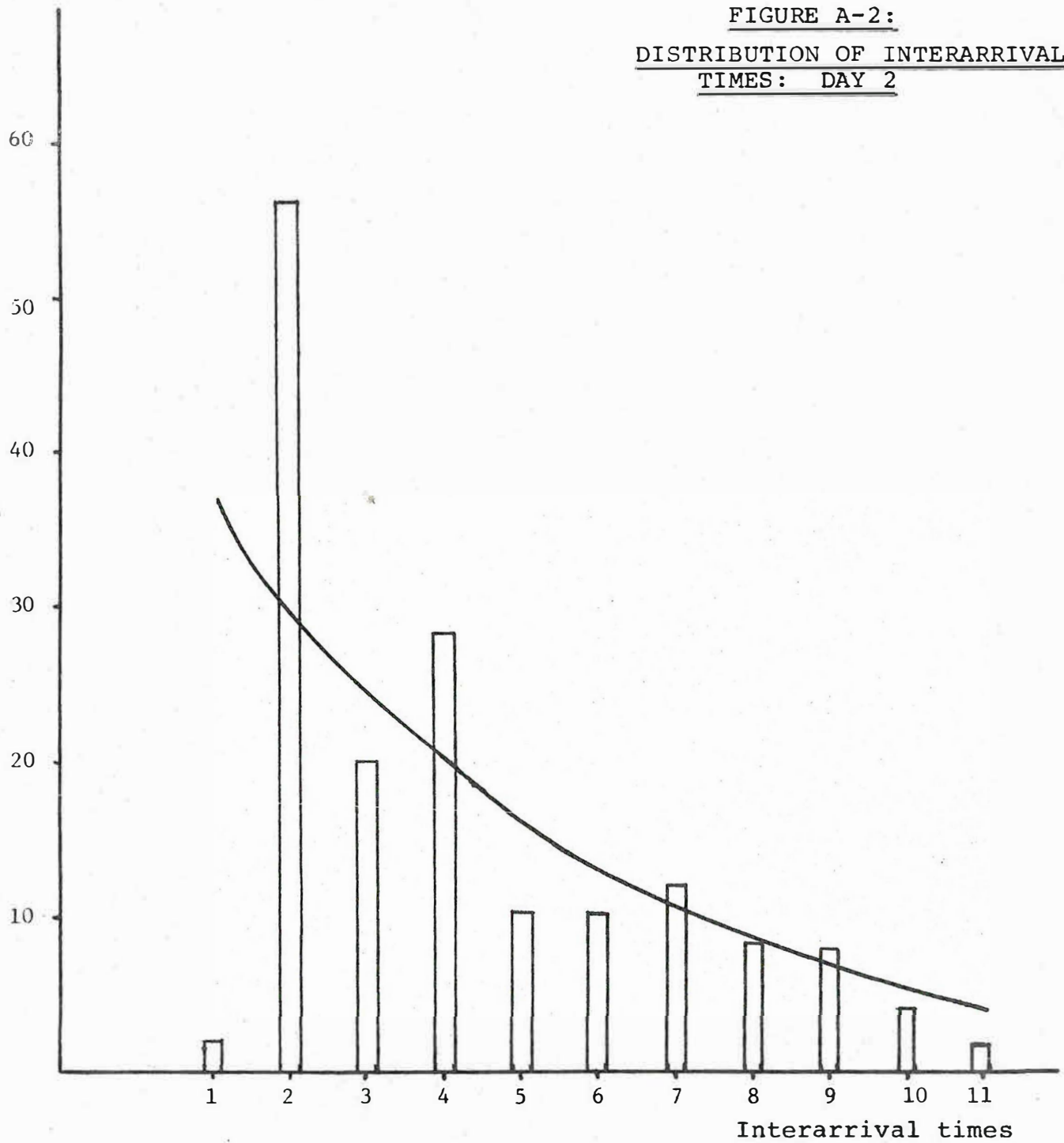
64

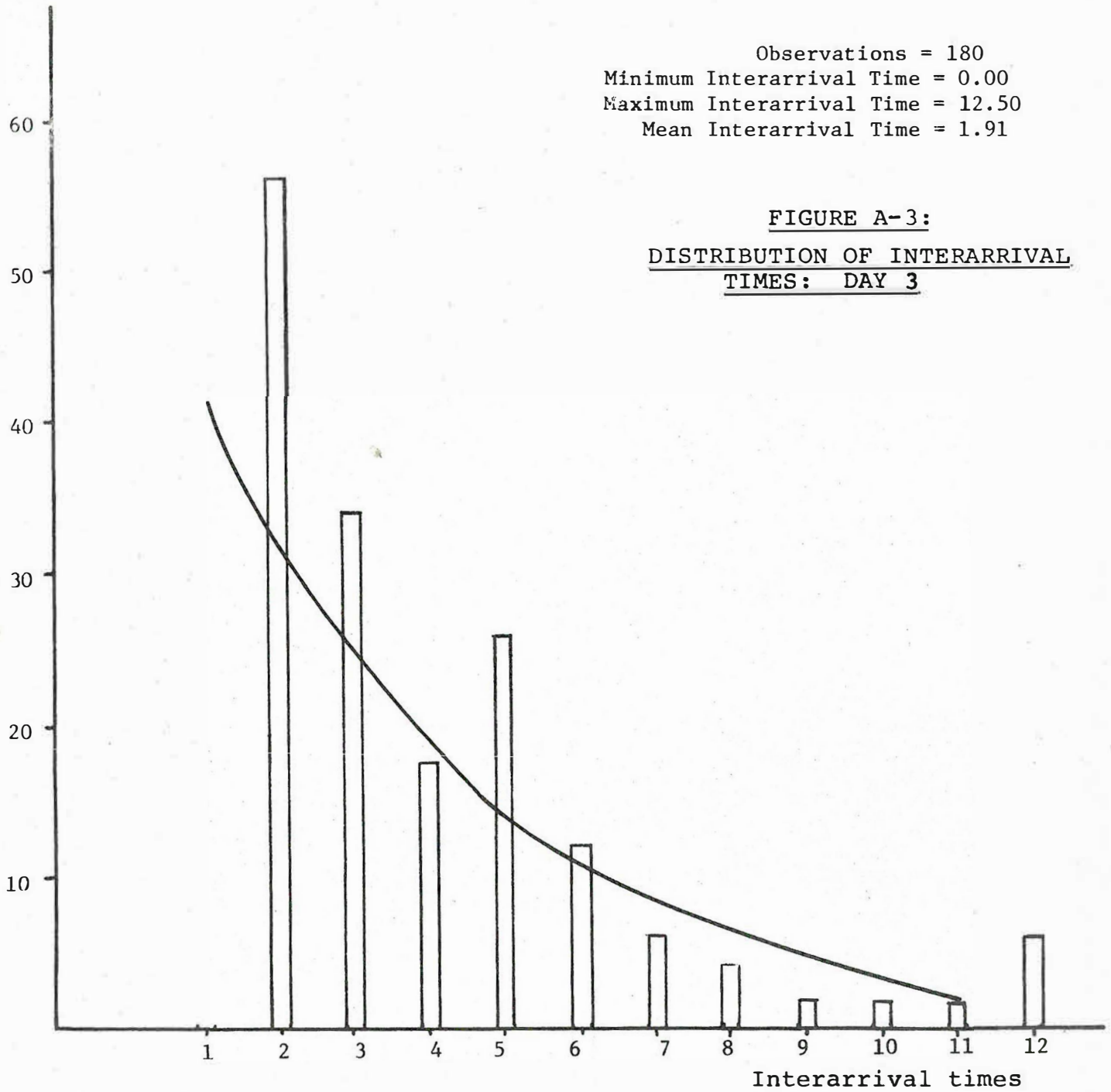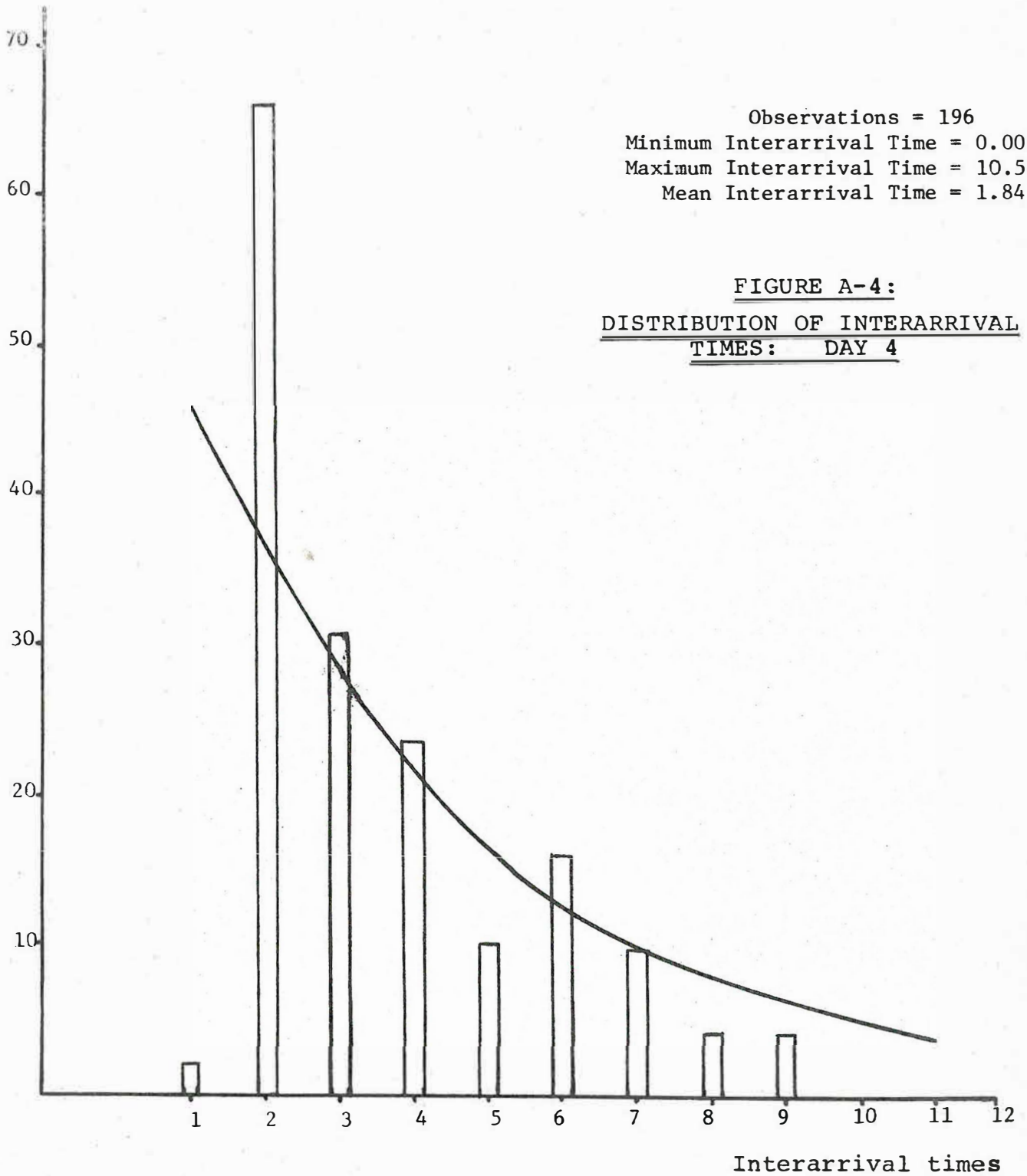THE EFFECTIVENESS OF DISPATCHING POINTS AND OBJECTIVE
FUNCTIONS IN IMPROVING SERVICE AT VARYING DEMAND LEVELS

A series of experiments were made to determine the
extent to which system performance could be improved by
using modified algorithms at different demand levels.

Initially six runs were made simulating a Haddonfield-
like situation. Three experiments used no dispatching
points,* and three used four dispatching points. Each
input file was run at 30, 50, and 70 demands per hour.
The general result, as can be seen in Table B-1, was
that at low demand levels, the dispatching points
improved service while at high demand levels the dis-
patching points had little effect.

This is because at higher demand levels vehicles are
sent to dispatching points less often (only unassigned
vehicles can be sent to dispatching points). The output
further suggests that dispatching points are useful at
lower demand rates because they anticipate future vehicle
needs.

More particularly, at a demand rate of 30 per hour
the dispatching points improved mean total time by 1.1
minutes (from 15.5 to 14.4) and reduced the standard
deviation from 7.5 to 5.3. (The maximum total time was
also reduced to 32.6 from 42.8 minutes.) One drawback
of the dispatching points was that they seemed to produce
an uneven distribution of work, with maximum time con-

---

*Dispatching points are used only when empty and unassigned
 vehicles occur in the simulation -- such a vehicle will be
 sent to the closest dispatching point where it will wait
 for assignment.

TABLE B-1:
SIMULATION RESULTS--VARYING DEMAND LEVELS

	MEAN WT TIME	MEAN TR TIME	MEAN TOT TIME	STANDARD DEVIATION	MAX TOT TIME
NDEM=30					
DP3	5.672	8.451	14.123	5.098	26.631
DP4	4.950	10.033	14.983	5.097	27.418
NDEM=40					
DP3	6.275	9.250	15.525	5.961	29.308
DP4	6.187	9.998	16.185	6.774	34.969
NDEM=50					
DP3	9.509	10.840	20.349	8.333	43.982
DP4	7.307	10.586	17.893	6.801	35.340
NDEM=70					
DP3	12.756	11.823	24.579	10.847	54.338
DP4	9.772	12.575	22.347	8.946	49.307

NDEM refers to the number of demands per hour.

DP indicates the use of dispatching points.

tinuously unassigned varying from 9.8 to 109.7 minutes
between vehicles, whereas without dispatching points
this range was from 12 to 78 minutes. Also, the pickups
and deliveries were much better distributed without dis-
patching points. In connection with this, two vehicles
were taken out of circulation (bringing the total to
eight vehicles still at thirty demands per hour) with
essentially unchanged results.

When the demand rate is increased to 40 per hour the
dispatching points algorithm was still substantially
better than the version without dispatching points.

At 50 demands per hour, an intermediate demand level,
the situation changes drastically. The only advantage
that the dispatching points provided was a somewhat
better distribution of work: maximum unassigned time
varied from 2 to 21 with dispatching points, whereas
without dispatching points some vehicles were idle
for as long as 48 minutes. As far as mean total time
and variability are concerned, the dispatching points
were no help: the mean total time with dispatching
points was 17.9 compared to 16.8 without, and the standard
deviation was unchanged.

At 70 demands per hour, the no-dispatching points
version slightly outperformed the dispatching version in
most respects. The work load was slightly better dis-
tributed without dispatching points; and   the range
of percent time unassigned was smaller in the version
without dispatching points, as was the composite per-
cent time empty. Concerning service provided, the no-
dispatching points version had a mean total time of 21.9
compared to 22.3 for the dispatching points version.
The dispatching points seemed generally unimportant at
this high demand level.

At this point it seemed clear that dispatching points
offered better service at low demand levels but that there

was room for improvement in the area of work load distribution: apparently once a vehicle was unassigned for a while the system was reluctant to work that vehicle back in, leaving some drivers unassigned for as long as 98 minutes. To counteract this, experiments were performed using the objective function with the increase in tour length term eliminated (this will be referred to as the modified objective function).

From Figure B-1 it seems clear that the modified objective function works very well with and without dispatching points at low demand levels. At intermediate and high demand levels, however, it performs abysmally. (One interesting feature of this poor performance is the fact that the modified objective function keeps the vehicles much busier even as it does a poor job.)

More specifically, at a demand rate of 30, dispatching points and the modified objective function clearly outperform all other combinations in terms of mean, standard deviation, and maximum service time, as well as work distribution.

At 40 demands per hour, there is again a remarkable improvement with the modified objective function. With the dispatching points, mean total time drops from 16.7 to 15.4, while standard deviation goes from 7.3 down to 6.1.

At 50 demands per hour, however, the modified objective function abruptly ceases to be effective. With dispatching points mean total time rose from 17.9 to 20.4 with the modified objective function, while the standard deviation rose to 8.3 from 6.8. In the no-dispatching points version, mean total time rose from 16.8 to 21.1 and standard deviation from 6.5 to 8.7.

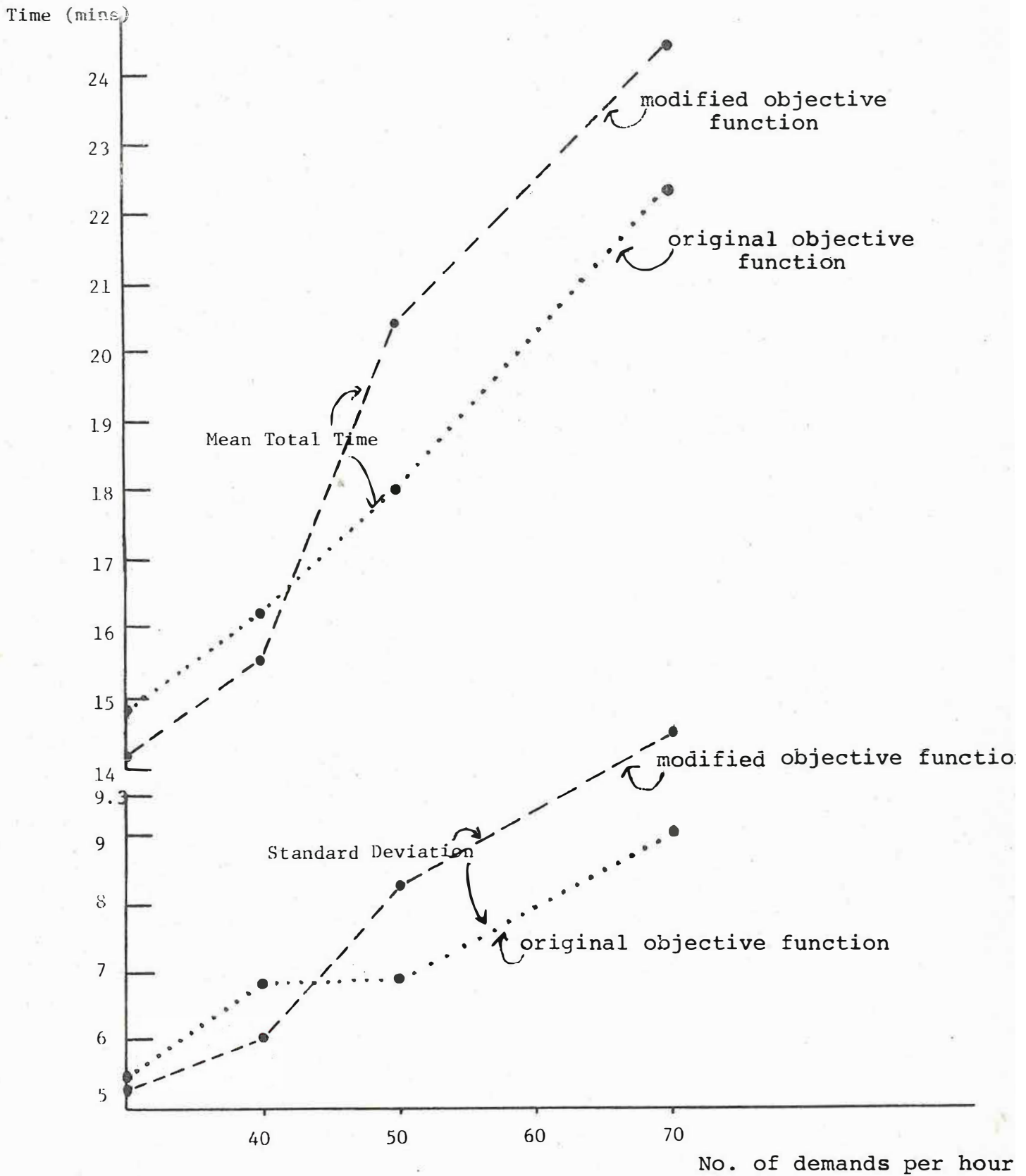At still higher demand levels the modified objective function continues to perform poorly.

FIGURE B-1: PERFORMANCE UNDER VARYING LOADS

APPENDIX C

UTILITY THEORY AND THE QUADRATIC OBJECTIVE FUNCTION

C.1 Determining Parameter Settings for Quadratic Objective
     Functions

An objective function is a means to compute a single
measure by which potential assignments are ranked, and
as such defines an "optimum" assignment. Section 4.1
showed that an ideal objective function would take the
form of an "average" utility function, and it was shown
that such an objective function is consistent both
with maximizing social welfare and with maximizing patron-
age. It was then shown that within current computer and
economic constraints, the best approximation to a utility
function is a quadratic objective function.

The problem remains: "What is the 'best' quadratic
objective function?" To implement the dial-a-ride al-
gorithm, parameters must be selected for the quadratic
objective function which are, in some sense, best.
Clearly the best settings are those which most nearly
approximate consumer preferences. Consider a hypothetical
system in which a particular choice of assignment only
affects one person, and all feasible alternatives
have identical system costs. The best assignment will
be that which the   person favors, and it will only
occur if consumer preferences are incorporated in the
objective function. To do this, individual utility
functions must either be inferred from actual choice
behavior or directly assessed.

Standard econometric techniques such as logit analysis[1]
might be used on revealed choice to infer preference par-
ameters. Unfortunately, such an analysis requires exten-

_____
(1) Ben-Akiva, M. Structure of Passenger Travel Demand
Model, Ph.D. Thesis, Department of Civil Engineering, M.I.T.,
1973.

sive collection of data and must be performed in a com-
munity which already has a dial-a-ride system. There are
problems in interpretation due to multi-collinearity among
the performance measures, due to extrapolation beyond the
range of service available in the existing dial-a-ride
system, and due to transfer from a community with a dial-
a-ride system to one without.

A more feasible approach is to directly assess the
utility functions of a sample of consumers from the community
in which the dial-a-ride system is to be implemented. Recent
advances such as

    (1) the identification of certain mathematical simplifica-
tions,[2]

    (2) the development of computer programs to compute the
utility parameters from simple questions,[3] and

    (3) the development of consumer measurement techniques
to assess utility functions by questionnaire[4]

have made it feasible to determine individual consumers'
utility functions directly.

This approach is currently being used at M.I.T.[5,6] to
determine consumer preferences and utilities for health
maintenance organizations and to relate these preferences
to design decisions. The next section presents a proposed
procedure for applying this approach to dial-a-ride.

(2) Keeney, R.L, "Multiplicative Utility Functions,"
Operations Research XXII, January, 1974, pp. 22-23.
(3) Sicherman, Alan. An Interactive Computer Program
for Assessing and Using Multi-attributed Utility Functions,
Master's Thesis, Operations Research Center, M.I.T.,
June, 1975.
(4) Hauser, John and Glen Urban. A Normative Methodology
for Modeling·Consumer·Response to Innovation, M.I.T.
Operations Research Center Technical Report 109, May, 1975.
(5) Hauser, ibid.
(6) Hauser, John. Modeling Decisions of Choice Among
Finite Alternatives: Applications to Marketing and to
Transportation Demand Theory, Operations Research Center
Report OR 038-74, M.I.T., October, 1974.

C.2 Proposed Procedure to Assess Consumer Preference

The proposed procedure consists of two phases, (I) exploratory and (II) actual consumer assessment. The purpose of the exploratory phase is to identify those performance measures important to the consumer in his choice of dial-a-ride which are operationally meaningful to the design team. The design team will first generate potential performance measures based upon their professional judgement and experience. These will then be tested in a consumer survey to identify those which are relevant to the consumer choice process and to develop a set of semantics which adequately describe these performance measures to consumers.

In the consumer assessment phase a survey based on the identified performance measures will be developed, pre-tested, and implemented to directly assess consumer utility functions. This survey, which will allow consumers to explicitly consider risk, tradeoffs, and interdependencies will have questions similar to the following:

Tradeoff Questions

1. Suppose the mode of transportation you are using costs $1.00. Suppose you can expect a waiting time of 10 minutes and a travel time of 20 minutes.

A more reliable mode is offered which also costs $1.00. This mode guarantees only a 5 minute waiting time. What is the maximum travel time you would accept and still prefer this more reliable mode?

Risk Questions

2. Suppose the mode of transportation you are using costs $1.00. Suppose you are not sure of the waiting time, in fact, it is as if someone flipped a coin: heads mean you had to wait 5 minutes, tails meant 25 minutes. In other words, an average time of $(1/2)(5+25) = 15$ minutes.

A more reliable mode is offered which also costs $1.00.

This mode can guarantee a fixed wait time. What is the maximum guaranteed wait time you would accept and still prefer this more reliable mode?

Independence Questions

3. If both the existing and the new reliable mode cost only $.50 would your answers to questions 1 and 2 change? If so, what would they now be?

The analysis of the survey results will yield parameters of individual utility functions. These parameters will be synthesized to determine an "average" utility function and finally this "average" function will be approximated with a quadratic objective function.

C.3 Relationship to Demand Prediction

The direct utility assessment described above can also be used to make initial predictions of demand. The techniques will not be described here, but with a slightly enlarged questionnaire, enough information can be gathered to calibrate a direct utility/bayesian demand model as described in Hauser[7]. This model would then be used to predict individual trial probabilities if the data is for a future system, or repeat probabilities if the data is for an existing system.

The demand model is mathematically complex but basically works as follows:

1. A distribution of the performance measures are estimated with the simulation model for dial-a-ride and are statistically determined for existing models.

2. The utility functions are used to compute scalar measures of goodness for dial-a-ride and the existing modes.

3. A Bayesian probability model is used to transform

(7) Hauser, John. Modeling Decisions of Choice Among Finite Alternatives: Applications to Marketing and to Transportation Demand Theory, Operations Research Center Report OR 038-74, M.I.T., October, 1974.

73

the scalar measures of goodness into individual choice probabilities.

4. The individual choices are aggregated into market share projections.

This section has proposed direct utility assessment of consumers as a technique to determine "optimum" parameter settings for quadratic objective functions, and has indicated a methodology to do this. An added benefit of this study would be a dial-a-ride demand model.

APPENDIX D
DERIVATION OF QUADRATIC OBJECTIVE FUNCTION

The premise that individual passenger utilities are proportional to the weighted sum of the squares of the following four level of service attributes:

1. wait time

2. difference between expected and actual wait time (wait deviation)

3. travel time

4. difference between expected and actual delivery time (delivery deviation)

can be used to develop an objective function to select the best assignment of a new passenger.

Definitions:

$NP(P)$ = number of pickups between new pickup and delivery insertion

$NP(D)$ = number of pickups after new delivery insertion

dP = detour due to insertion of new pickup

dD = detour due to insertion of new delivery

$TWT(P)$ = total wait time for pickups after new pickup

$TWD(D)$ = total wait time for pickups after new delivery

$WT(NEW)$ = wait time for new passenger based on this assignment

$TWD(P)$ = total wait time deviation for all pickups after new pickup

$TWD(D)$ = total wait time deviation for all pickups after new delivery

$NO(P)$ = number of passengers on board at new pickup

$NO(D)$ = number of passengers on board at new delivery

$NB(P,D)$ = number of passengers on board at both new pickup and new delivery

$TRT(P)$ = total travel time for those on board at new pickup

$TRT(D)$ = total travel time for those on board at new delivery

$TT(NEW)$ = travel time for new passenger

$TDD(P)$ = total delivery time deviation for deliveries after new pickup

$TDD(D)$ = total delivery time deviation for deliveries after new delivery

$ND(P)$ = number of deliveries after new pickup but before new delivery

$ND(D)$ = number of deliveries after new delivery

$WT(i)$ = wait time for pickup at stop i, for all pickups after new pickup, $i=1,\ldots,NP(P) + NP(D)$

$TT(i)$ = travel time for delivery at stop i, for all passengers on board at new pickup or new delivery; $i=1,\ldots,NO(P)-NB(P,D),\ldots NO(P),\ldots,NO(P)+NO(D)-NB(P,D)$

$DD(i)$ = delivery time deviation for all deliveries i after new pickup, $i=1,\ldots,ND(P),\ldots,ND(P)+ND(D)$

WAIT TIME

$$Z(WT) = dU(WT) = \sum_{i=1}^{NP(P)} (WT(i)+dP)^2 - \sum_{i=1}^{NP(P)} WT(i)^2$$

$$+ \sum_{i=NP(P)+1}^{NP(P)+NP(D)} (WT(i)+dP+dD)^2$$

$$- \sum_{i=NP(P)+1}^{NP(P)+NP(D)} WT(i)^2 + WT(NEW)^2$$

$$\therefore Z(WT) = \sum_{i=1}^{NP(P)} WT(i)^2 + 2dP \sum_{i=1}^{NP(P)} WT(i) + NP(P)dP^2$$

$$- \sum_{i=1}^{NP(P)} WT(i)^2 + \sum_{i=NP(P)+1}^{NP(P)+NP(D)} WT(i)^2$$

$$+ 2(dP+dD) \sum_{i=NP(P)+1}^{NP(P)+NP(D)} WT(i)+NP(D)(dP+dD)^2$$

$$- \sum_{i=NP(P)+1}^{NP(P)+ND(D)} WT(i)^2 + WT(NEW)^2$$

Now $TWT(P) = \sum_{i=1}^{NP(P)+NP(D)} WT(i)$ and $TWT(D) = \sum_{i=NP(P)+1}^{NP(P)+NP(D)} WT(i)$

Then by elimination of terms:

$$Z(WT) = NP(P) * dP^2 + NP(D) * (dP+dD)^2 + 2dP * TWT(P)$$
$$+ 2dD * TWT(D) + WT(NEW)^2$$

WAIT DEVIATION

$$Z(WTD) = dU(WTD) = \sum_{i=1}^{NP(P)} (WTD(i)+dP)^2 - \sum_{i=1}^{NP(P)} WTD(i)^2$$

$$+ \sum_{i=NP(P)+1}^{NP(P)+NP(D)} (WTD(i)+dP+dD)^2$$

$$- \sum_{i=NP(P)+1}^{NP(P)+NP(D)} WTD(i)^2$$

$$\therefore Z(WTD) = 2dP \sum_{i=1}^{NP(P)} WTD(i) + NP(P) * dP^2 + 2(dP+dD)$$

$$\sum_{i=NP(P)+1}^{NP(P)+NP(D)} WTD(i) + NP(D) * (dP+dD)^2$$

$$\text{Now } TWD(P) = \sum_{i=1}^{NP(P)+NP(D)} WTD(i) \text{ and } TWD(D) = \sum_{i=NP(D)+1}^{NP(P)+NP(D)} WTD(i)$$

$$\therefore Z(WTD) = NP(P) * dP^2 + NP(D) * (dP+dD)^2 + 2dP * TWD(P)$$
$$+ 2dD * TWD(D)$$

TRAVEL TIME

$$Z(TRT) = dU(TRT) = \sum_{i=1}^{NO(P)-NB(P,D)} (TT(i)+dP)^2 - \sum_{i=1}^{NO(P)-NB(P,D)} TT(i)^2$$

$$+ \sum_{i=NO(P)-NB(P,D)+1}^{NO(P)} (TT(i)+dP+dD)^2$$

$$- \sum_{i=NO(P)-NB(P,D)+1}^{NO(P)} TT(i)^2$$

$$+ \sum_{i=NO(P)+1}^{NO(P)+NO(D)-NB(P,D)} (TT(i)+dD)^2$$

$$- \sum_{i=NO(P)+1}^{NO(P)+NO(D)-NB(P,D)} TT(i)^2 + TT(NEW)^2$$

$$Z(TRT) = (NO(P)-NB(P,D)) * dP^2 + 2dP \sum_{i=1}^{NO(P)-NB(P,D)} TT(i) + NB(P,D) *$$

$$(dP+dD)^2 + 2(dP+dD) \sum_{i=NO(P)-NB(P,D)+1}^{NO(P)} TT(i)$$

$$+ (NO(D) - NB(P,D)) * dD^2$$

$$+2dD \sum_{i=NO(P)+1}^{NO(P)+NO(D)-2NB(P,D)} TT(i) + TT(NEW)^2$$

Now $TRT(P) = \sum_{i=1}^{NO(P)} TT(i)$ and $TRT(D) = \sum_{i=NO(P)-NB(P,D)+1}^{NO(P)+NO(D)-NB(P,D)} TT(i)$

$$\therefore Z(TRT) = NO(P) * dP^2 + NO(D) * dD^2 + 2 * NB(P,D) * dP * dD$$
$$+2dP * TRT(P) + 2dD * TRT(D) + TT(NEW)^2$$

DELIVERY TIME DEVIATION
$$Z(DTD) = dU(DTD) = \sum_{i=1}^{ND(P)} (DD(i)+dP)^2 - \sum_{i=1}^{ND(P)} DD(i)^2$$

$$+ \sum_{i=ND(P)+1}^{ND(P)+ND(D)} (DD(i)+dP+dD)^2$$

$$- \sum_{i=ND(P)+1}^{ND(P)+ND(D)} DD(i)^2$$

78

$$\therefore Z(DTD) = ND(P) * dP^2 + 2dP \sum_{i=1}^{ND(P)} DD(i) + ND(D) * (dP+dD)^2$$

$$+ 2(dP+dD) \sum_{i=ND(P)+1}^{ND(P)+ND(D)} DD(i)$$

$$\text{Now } TDD(P) = \sum_{i=1}^{ND(P)+ND(D)} DD(i) \quad \text{and } TDD(D) \sum_{i=ND(P)+1}^{ND(P)+ND(D)} DD(i)$$

$$\therefore Z(DTD) = ND(P) * dP^2 + ND(D) * (dP+dD)^2 + 2dP * TDD(P) +$$
$$2dD * TDD(D)$$

SYSTEM RESOURCES

$$Z(SR) = dP + dD$$

OBJECTIVE FUNCTION

The general objective function is then:

Minimize: $Z = A * Z(WT) + B * Z(WTD) + C * Z(TRT) +$
$D * Z(DTD) + E*Z(SR)$

APPENDIX E

DESCRIPTION OF SIMULATION INPUT FOR
OBJECTIVE FUNCTION TESTS

E.1 Characteristics of Service Area

The service area is divided into six (6) zones for the purpose of reproducing a Haddonfield-like spatial distribution of demand. The simulated service area is as shown in Figure E-1. The various zones have relative weights for origin and destination of trips as follows:

| ZONE	ORIGIN WEIGHT	DESTINATION WEIGHT
1	21	7
2	21	123
3	53	37
4	53	32
5	53	31
6	33	11

E.2 Demand Characteristics

Demands are assumed to arise as a Poisson process through time, with mean interarrival time of 1.59 minutes ($\sim$38 demands/hour) in the base case. Note that the mean demand rate was changed for some of the experimental runs to a mean interarrival time of 1.0 minutes (60 demands/ hour). Each demand is assumed to be exactly one passenger.

The trip length distribution consists of seven steps of width .5 miles. The weighting factors governing the choice of a given step are as follows:

STEP	WEIGHT
0.0-0.5mi.	0
0.5-1.0mi.	78
1.0-1.5mi.	58
1.5-2.0mi.	45
2.0-2.5mi.	37
2.5-3.0mi.	20
3.0-3.5mi.	1

FIGURE E-1:
SIMULATED SERVICE AREA, SHOWING SIX ZONES
(SCALE: 1 INCH = 1 MILE)
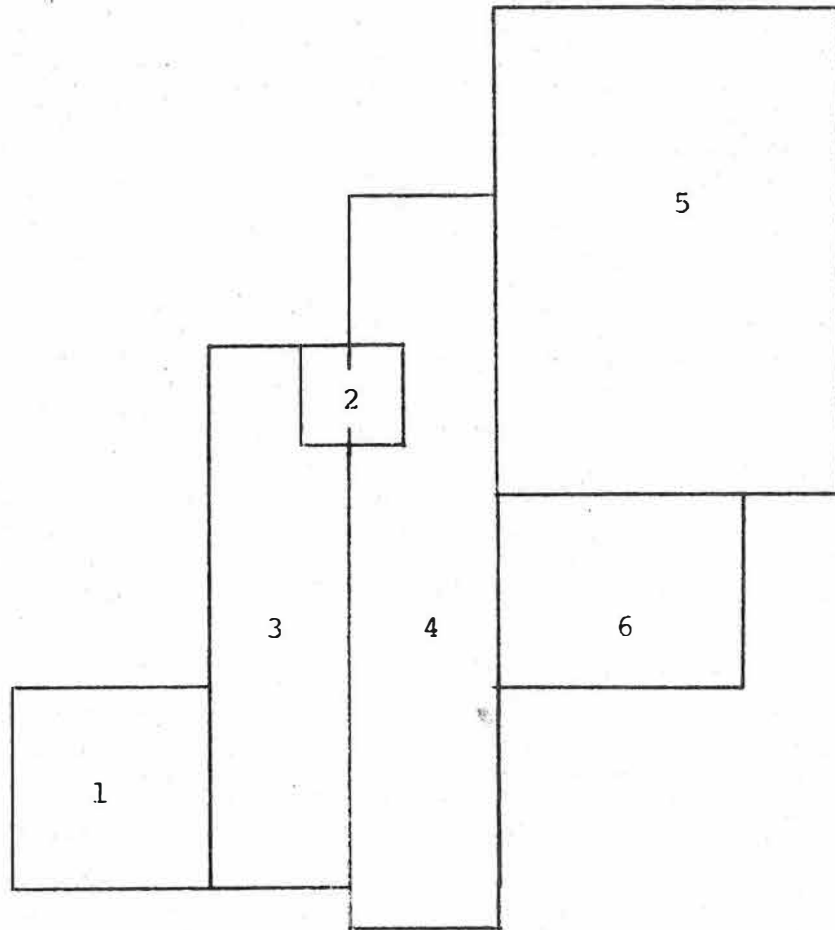
This distribution implies a mean trip length of 1.47 miles.

E.3 Vehicle Characteristics

There are eight (8) vehicles in service, each with
capacity of 20 passengers, and each of which travels at
an average speed of 15 miles/hour when moving. These
eight vehicles begin the simulation run empty, and at
random locations in the service area.

Continuous vehicle communication is used, with updates
at 30 second intervals.

E.4 Service Characteristics

The waiting time constraint is set at 60 minutes when
using the linear objective function. The travel time and
total time constraints are set to

3.5 * direct driving time + 60 (minutes).

The times required to pickup and deliver passengers
are uniformly distributed between .25 and .50 minutes
per passenger.

The algorithm options (objective function and parameter
values) are as described in the test results in Chapter 4
of this report.

A factor of 1.4 is used to convert straight-line
distance to actual distance travelled for the purposes
of travel time computation.