DESIGNING BUS ROUTES IN URBAN CORRIDORS

by

PETER GREGORY FURTH

B.S., Massachusetts Institute of Technology (1977)

M.S., Massachusetts Institute of Technology (1980)

SUBMITTED TO THE DEPARTMENT OF CIVIL ENGINEERING IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

July 1981

© Massachusetts Institute of Technology 1981

Signature of Author

Department of Civil Engineering

Certified by

Nigel H. M. Wilson

Thesis Supervisor

Accepted by

G. Milin Cornell
Chairman, Department Committee

JAN 13 1982

LIBRARIES
DESIGNING BUS ROUTES IN URBAN CORRIDORS

by

PETER GREGORY FURTH

Submitted to the Department of Civil Engineering
on July 20, 1981 in partial fulfillment of the
requirements for the Degree of Doctor of Philosophy in
Transportation Systems Analysis

ABSTRACT

Bus service in urban corridors is typically provided with
conventional local routes. Four different routing strategies are
explored and shown to be potentially more efficient than conventional
local service in corridors with moderate to high demand. These
strategies are deadheading some buses in the light direction; local
zonal service, which is a system of local routes of different lengths,
each having its own boarding zone inbound and alighting zone outbound;
local zonal service and express zonal service together; and offering
direct service (local and/or express) to more than one downtown
terminal. The network of streets in a corridor that have bus service
is assumed to be either a simple trunk or a tree.

A model of a bus route is developed, based upon which algorithms
for finding optimal and near-optimal routing configurations and service
frequencies for each routing strategy are developed. The objective
is to minimize a sum of operator and passenger costs, the demand pattern
is general, the service and operating constraints are realistic, and
the data requirements are modest. A closed form solution is found
for the problem of scheduling a route in which some runs deadhead.
The zonal service algorithms are dynamic programs.

Case studies of corridors in Boston and Minneapolis demonstrate
the value of these routing strategies and planning methods. In a high
demand corridor, a routing configuration was found that reduces operating
cost by 29% relative to conventional local service. In moderate and high
demand corridors in which some of these strategies are already implemented
but without the aid of optimal planning methods, configurations were
found that required from 9 to 14 percent fewer vehicles than the existing
configuration.

Thesis Supervisor: Dr. Nigel H. M. Wilson
Title: Associate Professor of Civil Engineering
ACKNOWLEDGEMENTS

This research was funded by the Agency for International Development, U.S. Department of State, as a part of joint program between M.I.T. and Cairo University in technology transfer to Egypt. Although the data that were necessary for applications to Cairo's transit system were not available in time for this thesis, the work was done with Cairo's transit system in mind and applications to Cairo are under way.

I am grateful for the wise guidance of my advisor, Nigel Wilson, who steered me away from doing too little and from doing too much, and whose timely suggestions helped make this thesis something I can be proud of. The rest of my dissertation committee, Ton Magnanti, Amedeo Odoni, Harry Psaraftis, Yossi Sheffi, and Mark Turnquist, who served as an informal outside examiner, also provided useful comments on numerous drafts.

Professors Mohammed El-Hawary, Farouk El-Bary, and Tarek El-Reedy of the Cairo University research team also provided useful review and feedback.

I want to express my thanks to those who provided the data I needed for my case studies. Alan Casteline of the Massachusetts Bay Transportation Authority and Mike Caracasani of the Central Transportation Planning Staff (Boston) were extremely helpful in providing data on Boston's transit system. Larry Englisher and Dave Vozzolo of Multisystems, Inc. and David Lee of the Metropolitan Transit Commission (Minneapolis - St. Paul) were also generous in providing data for the Minneapolis case study. Permission to use some of the Minneapolis data was granted by the U.S. Department of Transportation, under whose auspices they were collected
as a part of the Service and Methods Demonstration program. In the earlier stages of this research, Mark Turnquist of Cornell University, Keith Armstrong of the General Motors Urban Transportation Lab, and Colin Alter of Washington Metro generously offered and provided data.

Susan Feuerbacher, who typed most of this thesis as well as numerous memoranda, working papers, and rough drafts, deserves the credit for the excellent production of this report, as well as for keeping the research work going smoothly over the last two years.

I deeply appreciate my wife Beth, who stood by me through this thesis in its uncertain beginnings, its toilsome development and its joyful completion. And most of all, I want to give honor to God, who is the source of everything beautiful in this world, as well as the source of the ability and strength that made this thesis possible.
# Table of Contents

Chapter 1: Introduction, Problem Description and Thesis Plan .................................................. 8

1.1 Introduction ......................................................................................................................... 8
1.2 The Nature and Context of the Corridor Route Design Problem ....................................... 16
1.3 Development of the Corridor Design Problem and Plan of the Thesis ............................... 20

Chapter 2: Analysis Framework ................................................................................................. 23

2.1 Objectives, Constraints, and Assumptions ....................................................................... 23
  2.1.1 Maximize Service or Minimize Cost? ........................................................................... 24
  2.1.2 Fixed or Variable Demand? ....................................................................................... 26
  2.1.3 Service Policies and Constraints ............................................................................... 28
  2.1.4 Operating Policies and Constraints ......................................................................... 28
  2.1.5 Other Modeling Assumptions .................................................................................... 32
2.2 Data Requirements ............................................................................................................. 35
  2.2.1 Demand Data .............................................................................................................. 35
  2.2.2 Supply Data ............................................................................................................... 38

Chapter 3: Route Profile and the Conventional Local Route ..................................................... 40

3.1 A Typical Route Profile .................................................................................................... 41
3.2 Optimal Headway and Bus Type on a Single Route ......................................................... 50
3.3 Introduction to the Boston Case Study .............................................................................. 55
  3.3.1 Geography and Existing Routes in the Watertown Corridor ..................................... 56
  3.3.2 Data Used .................................................................................................................. 60
  3.3.3 Existing Levels of Service ....................................................................................... 62
  3.3.4 Objectives and Constraints Used in the Case Study ................................................ 62
3.4 Optimal Conventional Local Service in the Watertown Corridor .................................... 66
3.5 Conclusions ....................................................................................................................... 70

Chapter 4: Partial Deadheading ............................................................................................... 72

4.1 Minimum Number of Vehicles Under Fixed Headways .................................................. 73
  4.1.1 Minimum Number of Vehicles With Unevenly Spaced Departures ........................... 74
  4.1.2 Minimum Number of Vehicles with Evenly Spaced Departures: Review of the Literature ................................................................. 76
  4.1.3 Minimum Number of Vehicles With Evenly Spaced Departures: A New Approach ........................................................................... 78
  4.1.4 Some Assumptions Critically Examined .................................................................... 87
4.2 Minimum Number of Vehicles Under Maximum Headway Constraints ........................ 90
  4.2.1 Non-Integer Headway Ratios Allowed ..................................................................... 90
  4.2.2 Integer Headway Ratios Required .......................................................................... 92
4.3 Service Considerations in Partial Deadheading Design .................................................... 93
  4.3.1 Service Level as a Secondary Objective .................................................................... 95
    4.3.1.1 Improving Service in the Light Direction .......................................................... 95
    4.3.1.2 Improving Service in the Heavy Direction ......................................................... 97
### Chapter 5: Local Zonal Service

- 5.1 Different Zonal Service Strategies ........................................ 110
- 5.2 Exclusive Boarding/Alighting Zones ...................................... 113
  - 5.2.1 Optimal Design for a Single Direction ............................... 114
  - 5.2.2 Optimal Design for Both Directions .................................. 123
- 5.3 Overlapping Zones .............................................................. 129
- 5.4 Non-Contiguous Boarding/Alighting Zones ............................... 134
- 5.5 Local Zonal Service Case Study ........................................... 137
  - 5.5.1 Serving the Entire Watertown Corridor ............................... 138
  - 5.5.2 Serving Route 57 Demands ............................................. 141
- 5.6 Conclusions ........................................................................... 143

### Chapter 6: Express/Local Zonal Service

- 6.1 Assumptions About the Express/Local Configuration .................. 148
- 6.2 Optimal Design of Express/Local Zonal Service ......................... 150
- 6.3 Express/Local Zonal Service Case Study .................................. 165
  - 6.3.1 Serving the Entire Watertown Corridor ............................... 165
  - 6.3.2 Serving Route 57 Demands ............................................. 169
  - 6.3.3 Express/Local Zonal Service Without the Turnpike ............... 173
- 6.4 Conclusions ........................................................................... 179

### Chapter 7: Zonal Service on Branching Networks

- 7.1 Local Zonal Service on a Branching Network ............................ 184
- 7.2 Express/Local Zonal Service on a Branching Network .................. 191
- 7.3 Branching Route Case Study .................................................. 198
  - 7.3.1 Description of the Corridor ............................................. 198
  - 7.3.2 Analysis Framework ....................................................... 200
  - 7.3.3 Service Policy Constraints ............................................. 203
  - 7.3.4 Data Sources .................................................................. 204
  - 7.3.5 Number of Buses Needed for Existing Service ...................... 205
  - 7.3.6 Analysis of Demand ....................................................... 206
  - 7.3.7 Minimum Number of Buses Needed with the Existing Configuration .................................................. 210
  - 7.3.8 Proposed Routing Configuration ....................................... 213
- 7.4 Conclusions ........................................................................... 228

### Chapter 8: Serving Corridors with Multiple Downtown Terminals

- 8.1 Single Arterial Corridors with Multiple Downtown Terminals ........ 233
- 8.2 Branching .............................................................................. 236
Chapter 8: Urban Transit Network Design

8.3 Multiple Downtown Terminal Case Study
- Modeling the Corridor Network .......................................................... 238
- Objectives and Constraints ................................................................. 239
- Existing Service Configuration with Optimal Headways .................. 241
- Analysis of Demand ........................................................................ 243
- Selection and Evaluation of Alternative Configurations ................. 245
- Conclusions ..................................................................................... 253

Chapter 9: Summary of Findings and Directions for Further Research

- Summary of Findings ........................................................................ 255
- Research Contribution ...................................................................... 256
- Directions for Further Research ......................................................... 264

References ........................................................................................... 269

Appendix A: Derivation of Vehicle Costs ........................................... 272

Appendix B: Likelihood that an Optimal Partial Deadheading Solution with an Integer Headway Ratio Exists ......................... 274
Chapter 1

Introduction, Problem Description and Thesis Plan

1.1 Introduction

Public transportation in the United States faces a great dilemma in this decade. On one side, public transportation is almost virtually hailed as a remedy to help relieve our problems of oil dependence, street congestion, air pollution, and central city revitalization, and therefore there is popular support for expanding transit in the coming years. On the other side, taxpayers at municipal, state, and federal levels are putting strong pressure on public transit systems to reduce their exploding deficits that must be covered through public subsidy. Whether public transit will be able to expand in the coming years depends critically, therefore, on its ability to bring its costs more into line with its revenues.

It is generally conceded that public transit cannot operate at a profit in most cities. Nevertheless, deficits are probably greater than they have to be. There are a number of reasons for this fact.

One simple reason is that fares are kept low by the political bodies that regulate public transit. Low fares exist in part because of concern for the carless poor; a stronger reason is probably that a naive populace wants to have its cake and eat it, too, by having low fares and still desiring small deficits. A lot can be said about appropriate fare levels, but since concern for the poor and the naivete of the populace will probably be with us for many more years, it is unlikely that the pressure
for reducing (or even holding) the transit deficit will be alleviated solely by fare increases. Thus, whatever changes may be made in the fare structure to increase revenue, pressure will still be strong to reduce the cost of operating the transit system.

On the cost side, much attention has been devoted to the high wages paid to transit workers and to the restrictive work rules that inhibit the efficient use of labor. Generous federal subsidies have been blamed for the high capital and maintenance cost of the transit industry, and for removing the incentive to control costs in general. It is not these problems of a more political nature that are examined in this thesis, but another problem that also holds potential for cost reductions: the problem of efficiency in operations. Some operations problems have to do with day to day management, such as controlling absenteeism and operating when short-staffed. Another set of problems falls into the area of operations planning. One operations problem is labor scheduling, which is an important problem but is outside the scope of this research. Another set of problems falls into the category of service planning. This research effort focuses on one aspect of service planning.

The problem of service planning is choosing what routes to operate and the service frequency of each route during each time period. This problem is naturally decomposed into two parts: selecting routes and setting frequencies. Considerable work has been done in the area of setting frequencies on a given set of bus routes. For example, Mohring [1] and others derive a basic relationship of optimal frequency to ridership on a given route, the so-called "square root rule"; Furth [2], and
Scheele [3] give models for finding optimal frequencies on a set of routes with common resource constraints and variable demand; Hagberg and Hasselstrom [4] find optimal frequencies for a set of routes from a discrete set of acceptable frequencies. These efforts have made it possible for one to find the optimal frequency on a given bus route for a broad range of objectives, constraints, and assumptions. The problem of selecting bus routes, however, is much more difficult, since it is not a one-dimensional problem like finding an optimal frequency. The scope of all the research efforts in bus routing that are reported in the literature has been an entire metropolitan area [5,6,7,8]. In designing bus routes citywide, the driving forces of the problem are providing coverage, minimizing the number of transfers, and minimizing wait time. The only bus route type considered in these studies is the conventional local route, i.e. a bus route with small stop spacing that allows passengers to board and alight at any stop along the route in either direction. Designing conventional local bus routes for demands with origins and destinations dispersed citywide does not allow for more efficient design possibilities that exist in corridors of moderate to high demand where coverage, few transfers, and low wait time are easily achieved.

In the corridor routing problem, demands are oriented along an axis, and there is a dominant destination, usually the downtown area of a city. A city usually has a number of corridors emanating from downtown, each of which is served by a single arterial or branching system of arterials. Demands within the corridor are high enough and concentrated enough that
virtually no trips within the corridor will require a transfer, and coverage is very easily achieved. However, the high demand makes possible special services (i.e., different from conventional routes) tailored to different market segments that can make the overall service more efficient, as both operator experience and this research demonstrate. In the corridor routing problem, then, tries to find the most efficient design for bus service within a corridor, allowing routes that have unconventional operating strategies (such as boarding and alighting restrictions) as well as conventional local routes, and allowing the corridor to be served by a system of complimentary routes, in order to take better advantage of the economies of scale that urban corridors can offer.

The economies of scale in transit, as in other scheduled carrier services, are well known. As demand becomes greater and more concentrated, it becomes possible to serve passengers at a lower average cost to both operator and passenger. For example, in a corridor with low demand, service is infrequent, making waiting time high, and loads per vehicle are low, making operating cost high. In addition, a low demand corridor cannot support express service or direct service to many destinations (without extremely low loads and/or extremely high wait times). Thus, the typical low demand corridor is served by a single conventional local route that terminates at a downtown terminal. As demand increases, it becomes possible to raise loads per vehicle, thereby reducing operating cost, and/or to increase the service frequency, thereby reducing wait time cost. Transit systems have universally exploited economies of
scale in these two ways. To accommodate larger loads without increasing labor cost, vehicle size has been increased when possible; and to make wait times low, vehicles are dispatched at short, regular intervals as much as possible. Increased vehicle loads and reduced wait times are easily attained in a corridor as demands grow without changing the routing configuration in that corridor from the conventional local route: the operator simply lets the increased demand fill the buses to near capacity, and then increase service frequency as demand continues to grow. It may also be possible, in the longer run, for the operator to increase vehicle size.

One vehicle loads are near capacity and average demand is high enough so that wait time is low, the additional economies attainable on a conventional local route as demand grows are quite small. Once vehicle loads are near their maximum (and near-capacity loads are usually attained at moderate levels of demand), there are no more economies in operating cost with respect to increasing demand. There are still some economies in passenger wait time, but they diminish as the service level increases. For example, on a route with an 80 minute run time, if average wait time is half the average headway, then adding a fifth vehicle reduces average wait time by 2 minutes, while adding a fifteenth vehicle reduces average wait time by only 0.2 minutes. Thus, while increasing vehicle size and service frequency on a conventional local route can be quite effective in reducing average cost to both operator and passenger as demand grows from a low to a moderate level, this strategy is ineffective at reducing average operating cost and is only marginally effective at reducing
average passenger cost as demand increases from a moderate level to a high level. The goal of this research is to find routing strategies different from conventional local service that can lower average operating and passenger cost a moderate to high levels of demand.

The basic rationale behind this search for more efficient routing strategies is that since economies of scale on a single route are weak once the route has a moderate level of demand, it may be more efficient to segment the market of a high demand corridor and serve each market segment with a different route, each route being designed to serve its market segment as efficiently as possible. Wait times on the different routes will be greater than wait time would be on a conventional local route that served the entire corridor's demand, to be sure; but if the corridor demand is high, the increase in wait time will be small, and the different routes can be tailored to the demands of their corresponding market segments in such a way that the savings in operating cost, in passenger transfers, and in in-vehicle time more than offset the increase in wait time. For example, in many corridors there is a heavy demand destined for the downtown that can be isolated and served with an express route. These express passengers will have longer waits for the express buses than they would if everyone in the corridor used conventional local service, and the waits of those still using conventional local service will increase somewhat also since the express demand is lost to the conventional local route; but the travel time and operator cost savings on the express route may be great enough to make this strategy profitable. Other routing strategies that are explored besides express service are
deadheading of runs in the light direction, systematic short-turning
runs in long corridors, and provision of service to more than one
downtown terminal.

Interestingly, the transit industry has adopted a way of exploiting
economies of scale at very high levels of demand: a subway line is built.
However, the enormous cost of subway construction has slowed the expansion
of subway service and it is expected that fewer and fewer corridors will
be able to use this solution. Therefore, as transit ridership continues
to increase due to increasing fuel costs and increasing downtown
employment, more and more corridors will exhibit the levels of demand
that make strategies other than conventional local service preferred,
and will be pressed to adopt bus routing strategies such as those described
in this thesis. With the concurrent fiscal austerity that is growing
in this decade, it is important that these strategies be designed in the
most efficient way possible. Therefore the problem of finding optimal
routing configurations for high demand urban corridors is a general
problem with widespread potential application and significant potential
value in meeting both the financial and social goals of our cities and
the energy conservation goals of our nation.

The economies of scale in scheduled carrier service are well known,
and the high demand densities that occur in some urban corridors should
be exploited to the fullest. Unfortunately, traditional transit planning
has usually taken advantage of these economies in only two ways: increasing vehicle size to reduce labor costs, and reducing headways to reduce
crowding and wait time. Other economies that can be attained through the
implementation of routing configurations more complex than conventional
local routes have either been ignored, or where implemented, have not been planned optimally, due to a lack of research in this area. The goal of this research is to study these different configurations to explore their value in improving the efficiency of transit service, and to develop methods for finding the optimal configuration for a given corridor.

Therefore, the problem of designing bus service in urban corridors falls between the problems of optimal design of a single route and design of a citywide route network. Its scope is smaller than the citywide problem, its detail less than the single route problem, but by combining aspects of both it can obtain more efficient transit service designs in some situations than either of the other two approaches. The designs we are looking for are the service specifications for all the routes that are to operate in a corridor: the streets they are to cover, their service frequencies, their restrictions on boarding and alighting. (Sometimes what is called a bus route, for example, "Route 6", is actually a system consisting of many routes: Routes 6A, 6B, 6 Limited, 6 Express, etc. We are looking for the detailed service specifications of each component route in a corridor's route system.)

High density corridors are found in large cities throughout the world. As transit ridership continues to increase due to rising downtown employment and fuel prices, more and more corridors will exhibit the high demand densities that allow for solutions that are more efficient than conventional local routes. The trend to serve these corridors with rail transit is slowing due to the enormous costs of subway construction, making this problem still more relevant. And should the pressure to
reduce costs ever be removed, there will always be the pressure to provide the best possible service for any given cost. Therefore, the problem of optimal routing and scheduling of buses in high density corridors is a general problem with widespread potential application and significant importance in meeting both the financial and social goals of our cities and the energy conservation goals of our nation.

1.2 The Nature and Context of the Corridor Route Design Problem

Other people have studied the problem of designing bus routes over a citywide network. Four large scale models are reported in the literature: Lampkin and Saalmans [5]; Silman, Barzily, and Passy [6]; Mandl [7]; and Hasselstrom of the Volvo group [8]. What need is there for a study of route design of a corridor, a small and simple part of the citywide network? Two reasons must be given. First of all, the citywide routing problem is so complex that it is generally conceded that an optimal design procedure would be computationally infeasible. For this reason, the procedures that have been developed for citywide transit routing are all heuristic. However, the corridor design subproblem, because of its simplicity, can be solved optimally if isolated. Secondly, all of the citywide design models include only one type of route, the conventional local route with frequent stops, no boarding or alighting restrictions, and service symmetric with respect to direction. As many transit operators have experienced, and this research will confirm, optimal route designs may include other types of routes, such as express routes and routes that deadhead in one direction. Allowing greater flexibility in route type
would add to the already enormous complexity of a citywide model, and
the developers of these models have either neglected to consider the
value of these options or have concluded that the computational costs
of the increased flexibility outweigh its benefits. The smaller size
and simplicity of the corridor problem, however, makes explicit con-
sideration of these options possible at low computational cost. Since
heavy demand corridors are the only situations in a city where the opti-
mal solution is likely to include some of these specialized services, it
makes sense to isolate the heavy demand corridors and design service in
them allowing a variety of route types, rather than to design services
covering the entire city allowing the same flexibility in route type
throughout the city.

Solving the smaller corridor design problem allows for other kinds
of flexibility that would be very costly in a citywide model. Many
operators have found it most efficient to offer different services in the
peak and off-peak periods; yet they want most of their routes to be
unchanged. It is hard to accommodate these conflicting desires in a
citywide model. A corridor-based model is ideal for such a design,
however, because the operator is free to specify which high
demand corridors are candidates for specialized peak period services
while leaving the rest of the system unchanged. Another shortcoming of
the citywide models, with the exception of the Volvo model [8], is that
they choose routes and service frequencies sequentially instead of simul-
taneously because of the increased complexity it will bring. The
simpler corridor-based model allows for simultaneous choice of routes,
frequencies, and vehicle types. Operators can have a mix
of vehicle types, and need to decide which vehicle type to deploy on each
route. Only one study, also done by the Volvo group [6],
considers the simultaneous choice of vehicle type and service frequency;
but no study has considered the possible effect on route structure of
employing different vehicle types. Again, it is sensible to solve this
problem for heavy demand corridors only, since the introduction of
articulated buses has made vehicle type a factor that can affect the
design of routes in high density corridors, but not in lower density
areas.

The advantages thus enumerated of isolating heavy demand corridors
and designing service for them independently are considerable. But what
about the disadvantages? When one decomposes a problem and solves each
subproblem independently, even if the subproblems can be solved with more
powerful methods that the overall problem, still the superposition of
the solutions to the many subproblems may be inferior to the best solu-
tion one can find to the overall problem. Whether a decomposition is use-
ful depends on the degree of interdependence between the parts of the
problem.

The various component parts of the citywide network are interdependent
because they compete for the use of scarce resources such as vehicles
and subsidy. These interdependencies are best handled by means of
Lagrange multipliers (shadow prices or dual variables). If the decomposed
design calls for more type 1 vehicles that are available, one need only
impose a price on the use of type 1 vehicles; a few iterations will find
the price that exactly exhausts the supply of these vehicles. (With an
integer constraint, using dual variables is not guaranteed to lead to an
optimum because of the discontinuities of an integer problem. They will
lead to a point very near the optimum, however, which may be preferred to the optimum because it is a more stable solution with respect to small changes in demand and supply data. The true optimum can also be found by using branch-and-bound, as in [4], although this procedure is quite tedious for the small benefits it is expected to yield.

It is not the interdependencies of shared resources that make transit route design difficult, but the interdependencies of passenger routing. A change in service on one route can affect the demand on another route, especially if the two routes are overlapping or in other ways compete for the same market. However, this interdependence almost vanishes when one considers routes that are physically separated by a distance of a mile or more. It is vital to model the interdependence of routes in the same corridor where they may be competing for the same group of patrons; but the route choice of passengers in one corridor can hardly be affected by the design of routes in another corridor, making it reasonable to deal with the corridors independently. Of course, the boundaries of a corridor must be chosen carefully to minimize the inevitable interdependence that will occur at borders; but typical patterns of urban development make defining appropriate corridors a task that is not too difficult.

Thus, the disadvantages of designing each corridor independently are quite small. Indeed, this suggests a new heuristic for citywide route design that includes a corridor decomposition; but that is beyond the scope of this research. It will suffice for us to solve the corridor sub-problem as well as possible, and leave the extensions to later research.
1.3 Development of the Corridor Design Problem and Plan of the Thesis

Corridors can be as simple as a single arterial terminating at a downtown terminal, or more complex, involving branching arterials and multiple downtown terminals. Solution strategies also range from simple to complex. The strategy of this research—and the plan of this thesis—was to begin with the simplest corridors and the simplest strategies, and to build on those results for the more complex problems.

In a corridor design problem as defined in this thesis, only radial demands and routes (i.e., in the direction of the corridor) are considered. Demands for travel across the corridor are ignored, as they are not part of the problem. Passengers who transfer from a crosstown route to a radial route in the corridor are considered, but are treated just as passengers originating from within the corridor. If there are a large number of transfers between a route on a crosstown street and routes in the corridor, one should consider adding that crosstown street to the corridor as a branch in the network.

Physically, a corridor is an area emanating radially from the downtown (or other major attractor) that is served by a small set of arterials. In the model, the corridor is not represented at all as an area, but only as the network of arterials. Including an arterial in the network model reflects the planner's decision that there will be service on that arterial, so that careful judgment must be used in constructing the network. In designing service in low density areas, the choice of which streets will have service is difficult, but in most high demand corridors the choice is simple because there are very few (usually just one) practical alternatives.
Given the demand and supply data, the problem is to meet the demand under the prevailing supply conditions with the routing configuration and schedule that optimizes some objective function. The routing configuration may consist of a number of routes, each of which uses only the streets in the corridor network, has one terminal downtown and the other at an approved uptown terminal, and may have certain operating characteristics such as restrictions on boarding and alighting over certain segments. Four routing strategies are considered in detail: conventional local service; local service with partial deadheading; local zone service; and express/local zone service. Although the simpler strategies are special cases of the more complex ones, for clarity in presentation these four strategies are studied separately. For each strategy a model is developed to find the optimal configuration and schedule for that strategy.

The objective function, the service and operating constraints and policies, and other modeling assumptions are discussed in Section 2.1. Section 2.2 describes the data requirements. The following chapters 3 through 6 study the four routing strategies for the simplest corridor, one whose network consists of a single arterial, and a single downtown terminal. Chapters 3, which discusses conventional local service, also provides an introduction to the other three single arterial routing strategies. Chapter 7 extends the results to a corridor with a tree network. Chapter 8 extends the problem still further by incorporating multiple downtown terminals into the corridor network. Providing direct service to multiple downtown terminals may be considered a fifth routing strategy. This strategy is especially relevant when the downtown is so
large that it is served by a number of distant bus terminals, or when besides downtown there are other major destinations nearby, such as a medical, industrial, or government complex. Chapter 9 presents conclusions and directions for future research.

As each problem and solution strategy is studied, the literature pertaining to that problem is reviewed, a model is developed, and an algorithm for optimal or near-optimal design is presented. An example illustrates each of the solution strategies and gives an indication of its potential benefits. The examples are as realistic as possible, being taken from actual corridors in the Boston and Minneapolis metropolitan areas.
Chapter 2
Analysis Framework

In order to better understand the models developed in the succeeding chapters, and to better understand the contribution of this research in the broader context of transit planning, it is important to set forth the analysis framework upon which the models developed in this thesis are based. In evaluating a model, the more general its objectives, the more realistic its constraints, the more plausible its assumptions, and the more limited its data requirements, the more useful it is. Furthermore, comparing the details of any two models requires that they be similar in their objectives, constraints, modeling assumptions, and data requirements.

2.1 Objectives, Constraints, and Assumptions

The many studies made of transit route and frequency design have employed numerous differing objectives, constraints, and assumptions. One advantage of the methods that have been developed in this research is that virtually any of the objectives, constraints, and assumptions used by other researchers can be included (provided, of course, that they are consistent). This section will go through the various objectives, constraints, and assumptions that one may want to use and show how they can be assimilated into the models. Out of all the possible sets of objectives, constraints, and assumptions, the case studies that appear in Chapters 3-8 employ a consistent set that is described in this section. Do not infer, however, that because the case studies use a
particular set of objectives, constraints, and assumptions that this is the only set that can be used.

2.1.1 Maximize Service or Minimize Cost?

Transit design studies generally either maximize passenger level of service (or patronage) subject to a constraint on operating cost, or minimize operating cost subject to a constraint on level of service. These two problems can be formulated as duals; with the appropriate shadow prices they are identical. Maximizing passenger level of service is a more appropriate objective in the short run when the resources available are known, and minimizing operating cost seems more appropriate in the long run. A general objective function would be to minimize a sum of operator cost and passenger inconvenience (a negative measure of level of service). With the appropriate weights on each objective function component, a socially optimal design would be achieved which included, along with routes and frequencies, the subsidy and number of vehicles that the system ought to have. When subsidy and/or fleet size is constrained, the appropriate weights (i.e., shadow prices) should ideally be found by iteration through interaction with the entire transit system so that no more than the available resources are called for when these shadow prices are applied consistently throughout the network. Since this research is concerned only with corridor design, it is outside the scope of our models to try to find the correct shadow prices. We will therefore use an objective function that is a weighted sum of operating and passenger costs, where the weights are to be determined exogenously.
When an operator is unable to find the correct shadow prices, he can choose a few sets of reasonable values and do the designs using each set. He would then probably feel more comfortable choosing between the different resulting designs with their implicit tradeoffs of level of service measures and operator cost measures than choosing between a few sets of shadow prices. Reasonable shadow price ranges for typical level of service measures are zero to one third of the average wage rate for travel time, zero to fifty cents for transfers. Shadow prices reflecting a shortage of particular vehicle type should make the operating cost of that vehicle type lie somewhere between its true operating cost and the operating cost of the next larger vehicle type; again, a number of reasonable values could be tried and the operator could choose the solution that seemed best.

The tractability of this model depends on the different cost components being separable by route. This is not a major restriction, since passenger costs can easily be separated according to the route on which they are incurred and operating costs can be separated similarly. One shortcoming of this restriction is that it makes interlining of buses infeasible (interlining is when a vehicle switches routes within the design period). But interlining is one of the finer details of scheduling, and these models are not intended to give schedules in full detail, rather they are designed to give a frequency of service during a specified period (e.g. morning peak). It is expected that before implementation an operator will modify the schedules to allow for buses to pull into and out of garages, to accommodate fluctuations in demand within the period, to allow for special runs for schools or industries, and so forth.
It is more appropriate for interlining to be planned at this subsequent stage than at the route/frequency planning stage.

The other major shortcoming of the route independence requirement is that it makes these models unable, in themselves, to design overlapping routes which lack boarding or alighting restrictions. In such configurations, the number of riders on any one route clearly depends on the level of service of all the routes that overlap it as well as on its own level of service. Chapter 5 covers this topic in more detail.

The objective function is therefore very flexible, able to include any operator or passenger cost arising from the operation of a single route. Operating cost, opportunity cost of using scarce vehicles of different types, wait time cost, in-vehicle time cost, crowding cost, transfer cost, reduction in consumer surplus due to a change in demand on a route, and other costs can be included in this framework.

2.1.2 Fixed or Variable Demand?

One of the major assumptions of a transit planning model is the extent of the demand response to the service options chosen. There can be no response, or the response can be modeled by a number of different formulas with different parameters. The user of these models is free to choose whatever formula he wants, provided that the demand response to service on a given route is affected by service on that route alone. For the types of problems with which we are dealing (i.e., coverage is guaranteed and headways will never be very large), this restriction is not severe.
A demand model for route ridership has recently been developed by Hasselstrom [8] and is used in the Volvo route planning model. It assumes the demand curve has a negative exponential form. The other three major route design models assume fixed demand. Another route demand model has recently been developed by González [9] based on a Poisson model of passenger trip-making behavior. Other models used for route demand are described in González.

Care must be taken in making the assumption about demand response consistent with the objective function. Of course, one cannot maximize ridership or passenger miles with demand held fixed. Less obviously, if demand is variable one must not minimize user costs such as total travel time, since they will reflect not only changes in level of service but changes in the number of riders as well. To account for user costs one should maximize consumer surplus. (This is equivalent to minimizing user cost when demand is fixed.) In the Volvo model, the negative exponential demand function makes minimizing consumer surplus equivalent to maximizing ridership. Furth [2] maximizes consumer surplus with a logit demand model in an optimization of route frequencies.

The case studies done in this thesis assume demand is fixed. One reason is for computational simplicity. But beyond that, the main motivation of this research is to show how transit service can be operated more efficiently, and this is perhaps best understood by showing the potential cost savings in serving the same demand. Thirdly, any results that have assumed elastic demand are suspect to many transit operators who mistrust any proposed solution that bases part of its sucess
on generating demand; this was an important consideration since this work is ultimately of no value unless transit operators accept it. And fourthly, Furth [2] has shown that in transit design problems where coverage is not an issue the design resulting from minimizing the costs with a fixed number of users differs little from the design that maximizes consumer surplus with variable demand. This makes sense, since the best way to serve current patrons is usually also the best way to attract and serve new patrons.

2.1.3 Service Policies and Constraints

Some of the constraints in the models arise from generally accepted policies governing level of service. These policies are enumerated below.

A policy headway, the maximum allowable headway on all routes, ensures that wait times will be acceptable.

Maximum crowding levels constrain the headways to be small enough that the average load at the peak load point does not exceed a certain standard. Either this constraint or the policy headway constraint will be superfluous; the lesser of these two is the maximum headway for a given route.

2.1.4 Operating Policies and Constraints

Other constraints and part of the model structure arise from policies governing operations.

Layovers occur at the end of each run. Layover time can be modeled as a constant plus a fraction of the run time.
We may or may not require that the number of buses assigned to each route be an integer. Certainly an operator can only use an integer number of buses at any given time in the schedule, but interlining makes it possible for routes to share vehicles, making the number assigned to any single route non-integer. However, there is a certain inefficiency in interlining, since it is not possible in general to assign to each of a number of routes an arbitrary non-integer number of vehicles and then find an acceptable timetable that achieves these assignments through interlining. The reason for this infeasibility is that interlining requires some extra slack time to keep departures on any given route regularly spaced, and sometimes slack time is needed for deadheading between terminals. On the other hand, compiling and operating a timetable in which the number of buses assigned to each route is an integer is a simple matter, since each route included in the timetable may be scheduled independently. For this reason operators will often prefer a model whose resulting vehicle assignments are integer. In a situation where an operator is willing to interline extensively and therefore accept non-integer vehicle assignments, we must estimate the amount of slack there will be in the interlined schedule. One simple way of modeling this schedule slack is to assume that slack time in an interlined schedule will be a certain fraction of the run time.

All of the case studies in this thesis except one have required integer vehicle assignments. The exception, the Minneapolis case study of Chapter 7, has non-integer vehicle assignments because the routes of the corridor under study are extensively interlined with the routes of another corridor. As the models in the succeeding chapters are
developed, they all reflect the assumption that an integer number of vehicles must be assigned to each route, since it seems that this assumption would most often be preferred. The results for the assumption of real number vehicle assignments can easily be derived from the results given, usually by simply dropping from a formula the integer operator.

Clockface headways, i.e., headways of 5, 10, 15, etc. minutes, are sometimes desired by operators in order to make schedules easier to remember for their patrons. In applying these models one could specify any desired set of acceptable headways. Militating against clockface headways is the desire to minimize unproductive layover time. It seems that in North America and in developing countries this second factor outweighs the concern for schedule simplicity when headways are below 20 minutes, as non-clockface headways abound. For this reason, the case studies do not include this constraint. Nevertheless, they do include a discrete set of dominant headways: those headways corresponding to an integer number of buses and no excess layover, since greater headways that require the same number of vehicles are clearly inferior. (Section 3.2 proves this analytically.) When clockface headways are required, finding the optimal headway requires enumerating over the feasible headways, which is not too burdensome because there will not be many. This approach has been taken in Hagberg and Hasselstrom [4]. When clockface headways are not required, it is not always necessary to enumerate over the dominant headways because of the usual convexity of the objective function. This subject is discussed in detail in Section 3.2.
Regular headways, i.e., dispatches at constant intervals, are required in one of the deadheading models (Section 4.2.3). In all of the zonal service models it is assumed tacitly, although there is no reason under the structure of those models that non-regular headways would ever be superior to regular headways. In practice headways are often not perfectly regular, but are near regular. This is another case, where imposing the constraint (of regular headways) is more reasonable than relaxing it, since a solution that arose without the assumption of regular headways could not be implemented in general without severe crowding and bunching problems.

Only one type of bus is permitted on a given route. This restriction may be imposed by some operators. Nothing intrinsic in the model requires that different types of vehicles not be mixed on the same route, but since the operators studied during the course of this research had this policy, it was adopted in this work. In the only other study that assigns vehicles of different types to routes [4] the same assumption is made.
2.1.5 Other Modeling Assumptions

Stability over time in both passenger arrival rates and in run times was assumed to hold over the period of analysis.

Wait time, which includes the inconvenience of not being able to arrive at one's destination exactly when desired as well as the time actually waiting for the bus, was assumed to be a constant fraction of the headway, given by:

\[ \frac{wt}{bh} = b \]  

(2.1)

where \( wt \) = average wait time

\( h \) = average headway

\( b \) = a parameter, \( b \geq 0.5 \)

The true relationship of wait time to average headway depends on how even the spacing between successive buses is. When buses are evenly spaced, average wait times will be near half the headway, but when they are more irregularly spaced and begin to bunch up, average wait approaches and can even exceed the average headway. The chosen value of the parameter \( b \) should reflect the expected variability in headways.

It can be observed that headway variability is proportionately larger on routes with small headways than on routes with large headways, but this increase in variability is largely attributed to the heavier traffic congestion that small headway routes must go through [10]. If one has reason to believe that this phenomenon will repeat itself in the system being analyzed, one should use a wait time formula that reflects it. One such formula estimated by Holroyd and Scraggs [11] using data from central London bus routes relates observed wait time (not including schedule inconvenience) to headway by the following function:
\[
\frac{w_t}{h} = \left( \frac{h^2 + 70}{2h^2 + 70} \right) h
\]

The increased accuracy from using such a formula must be traded off against the resulting increased computational complexity. Because it is a cubic formula, it makes it necessary to solve the objective function by search rather than in closed form. More importantly, though, it is concave with respect to headway, so that one cannot easily prove that the objective function is convex (and it may be non-convex for some routes). Thus, a search for a global optimum would be even more difficult. Analysis and numerous experiments done as a part of this research effort have shown that it is unlikely for the optimal design to be affected by the choice of wait time formula, since the difference between the two formulas is very small in the relevant range. For this reason the case studies use the simpler formula (2.1).

Expected dwell time at a given stop has two components: a fixed amount of time for deceleration, acceleration, opening and closing doors, and returning to traffic, and the time required for passengers to board and alight. The fixed time for stopping is only incurred if the bus stops, and it will stop only if there are passengers who want to board or alight. If passenger arrivals are modeled as a Poisson process, the probability that there will be no passenger movements at stop \( i \) is given by:
\[ Pr\ (do \ not \ stop \ at \ i) = e^{-(\lambda_i h + \mu_i h)/60} \quad (2.2) \]

where \( \lambda_i \) = boarding rate at stop i in passengers per hour
\( \mu_i \) = alighting rate at stop i in passengers per hour
\( h \) = headway in minutes

Note that \( \lambda_i h/60 \) is the expected number of arrivals at stop i during a headway, and \( \mu_i h/60 \) is the expected number of passengers who arrived upstream during a headway and desire to travel as far as stop i. Because of fare payment, passengers may board at a different rate than they alight, so the dwell time formula is:

\[ \overline{DT} = a_1(1-e^{-(\lambda_i h + \mu_i h)/60}) + a_2 \frac{\lambda_i h}{60} + a_3 \frac{\mu_i h}{60} \quad (2.3) \]

where \( \overline{DT} \) = expected dwell time
\( a_1 \) = fixed time incurred by stopping
\( a_2 \) = delay per passenger boarding
\( a_3 \) = delay per passenger alighting

The fixed time incurred by stopping \( (a_1) \) is a sum of the delay caused by decelerating, the time needed to open and close doors, the wait for a gap in the traffic stream, and the delay caused by accelerating.

Assuming constant rates of deceleration and acceleration, this fixed time can be modeled thus:

\[ a_1 = a_o + \frac{1}{2} \left( \frac{V_o}{acc} + \frac{V_o}{dec} \right) \quad (2.4) \]

where \( a_o \) = time required to open and close doors and return to traffic
\( V_o \) = cruising speed
\( acc \) = acceleration rate
\( dec \) = deceleration rate
Notice that the time required to decelerate and accelerate is \((V_0/\text{acc} + V_0/\text{dec})\), but since some distance is covered in these maneuvers not all of it is delay. Covering this distance at the cruising speed \(V_0\) would require half of that amount of time, so the delay is only half that amount; hence the coefficient of \(\frac{1}{2}\) in equation (2.4).

**Fare effects** are neglected because fare policy is more a political decision than an operations planning issue. The models developed in this research can be used to find the marginal cost of passengers in different market segments, which can be used to set fares; changes in fare policy would then have a feedback effect on demand. But these effects are not modeled since the operator usually does not have the same flexibility to set "optimal" fares that he has to make an optimal schedule.

2.2 **Data Requirements**

Difficult data requirements will severely restrict the usefulness of a transit planning model. It was a goal of this research effort to develop models whose data requirement could be met using generally available data either alone or supplemented by data that could be collected using the kind of manpower and financial resources normally available for data gathering within a transit agency. This goal excluded, therefore, the use of surveys (except perhaps one-question on-board surveys), and called for exploiting to the greatest extent possible existing data and proven methods of inference.

2.2.1 **Demand Data**

The demand data must be in such a form that one can compute the
expected number of passenger movements at each stop for any component route of one of the four possible routing strategies, as well as that route's peak load. Since we are concerned with high density corridors, it is almost certain that routes exist there already, and the bus stops that are part of the model's network will usually be existing bus stops. Thus it is possible to estimate demand from ridership on existing routes.

The form of demand data collected most often by a transit operator is peak load counts, the number of riders on board at a prespecified point thought to have the greatest loading along the route. On/off (characteristic) counts, which give the number of boardings and alightings at each stop, are less commonly available. A proper implementation of these models requires on/off counts. (A procedure suggested by Han [12] could also be used instead of making on/off counts. It relies on point counts being made at a few locations besides the peak load point.) Monitoring a few runs (perhaps 6 to 8 runs) in the period being analyzed would be sufficient to give an accurate distribution of boardings and alightings within the period. For more detail on sampling strategies, the reader is referred to Multisystems [13]. The peak load counts can then be used to expand the on/off counts to yield a good estimate of boarding and alighting rates by stop.

Once boarding and alighting rates are known, a route origin-destination matrix can be inferred using a method developed by
Tsygalnitzky [4]. It uses the assumption that the origin stops of the passengers alighting at a given stop are distributed in proportion to the number of passengers who boarded at each origin stop who are still on board at the alighting stop. A minimum trip distance may also be specified, so that passengers who boarded at stops very close to the alighting stop are not considered. The method has proven very accurate for the routes on which it was tested. The formulas for inferring the O-D matrix are given below.

Define

\[ d_{ij} = \text{number of passengers boarding at } i \text{ destined for } j \]
\[ b_i = \sum_j d_{ij} = \text{number of passengers boarding at } i \]
\[ a_j = \sum_i d_{ij} = \text{number of passengers alighting at } j \]
\[ t = \text{minimum trip distance (in stops)} \]
\[ v_{ij} = \text{number of passengers who boarded at } i \text{ who are still on board just before stop } j \text{ and are eligible to alight at } j \text{ (i.e., } v_{ij} = 0 \text{ if } (j-i)<t \) \]
\[ V_j = \sum_i v_{ij} = \text{number of passengers on board just before stop } i \text{ who are eligible to alight at stop } j \]

Begin at the upstream end of the route, i.e., \( j=1 \). Initialize \( v_{ij} \) and \( V_j \) to zero. Then solve the following equations for each successive stop until the end of the route is reached.

\[ v_{j-1,j} = b_{j-1} \]
\[ V_j = V_{j-1} + b_{j-1} \]
\[ d_{ij} = \frac{v_{ij}}{V_j} a_j \text{ for all } i=1,\ldots,(j-t) \]
\[ v_{i,j+1} = v_{ij} - d_{ij} \quad \text{for all } i=1,\ldots,(j-t) \]
\[ v_j = v_j - a_j \quad (2.5) \]

This procedure is easily programmed or can be solved manually on a tableau. For more detail on this method, see the original reference.

While Tsygalnitzky's method appears to infer O-D matrices quite reliably, a better method could be developed that makes use of one or more point counts that may be available as well as on/off counts. This matter is left to further research.

Thus, using available peak load point counts and making a few on/off counts if none are available, one can infer a route origin-destination matrix. If the corridor studied is served by several routes, this procedure should be followed for each, and the O-D matrices aggregated to yield a corridor O-D matrix. Thus the demand data requirements can be met without a large expenditure.

2.2.2 Supply Data

The supply data that are needed are run times between adjacent stops and run times for express segments that an express route or a deadheading route might use. Operators usually have some run time data (e.g., terminal to terminal) and remaining run times can be estimated by measuring distances from a map and assuming a speed.

The dwell time parameters \( a_0 \), \( a_2 \), and \( a_3 \) used in equations (2.3) and (2.4) must be estimated or guessed. Jordan and Turnquist [15] report that 2.7 seconds were required for each boarding passenger, and that 2.25 additional seconds were consumed while the bus was not moving. The stops they studied had few, if any, alighting passengers. The time required
for each alighting passenger is probably less than the time required for each boarding passenger (unless riders pay as they leave). If two doors are used for passenger movements, rates should be adjusted accordingly and some interaction could be modeled between boarding and alighting passengers.
Chapter 3
The Route Profile and the Conventional Local Route

Everyone who is familiar with public transit is familiar with the conventional local route. It has frequent stops, has no boarding or alighting restrictions, it uses the same path in both directions (or as close to the same path as possible in the presence of one-way streets), and buses come just as frequently in one direction as the other. This route type is the standard building block of bus systems; in any given system, the vast majority of service, if not all, is offered on conventional local routes.

Conventional local routes are well adapted to low demand areas; they are good for providing low frequency coverage of a wide area. Their extensive use in heavy demand corridors, however, is questionable.

The transit industry appears to have made three types of responses to high demand in a corridor. This first is to make service more frequent by using more vehicles. This can reduce passenger wait time and crowding somewhat, but leaves other passenger costs and operating costs (per passenger) unchanged. The second response is to increase the size of the vehicle. This response raises wait time and in-vehicle time, reduces crowding, and reduces operator cost as long as the savings in driver cost outweigh the increase in capital and operating costs. The third response is to build a subway and serve the corridor by train. Subway service is
usually conventional local service as well, except that stop spacing is greater.

While passengers usually benefit from the construction of a subway, it has become obvious that subway construction is no longer a way to reduce operating cost per passenger except under circumstances of unusually high demand. This leaves unanswered the question: are there more efficient ways to serve a high demand corridor than running buses more often and/or using bigger buses? To answer this question it is first worthwhile to examine a typical route profile, i.e., the pattern of demands that a conventional local route in a high demand corridor must serve, as well as the capacity offered throughout the corridor.

3.1 A Typical Route Profile

A convenient tool for illustrating the demand pattern of a route is a route profile. Figure 3.1 contains a typical route profile for a high demand radial route in the morning peak. (Throughout this thesis the morning peak is used for the sake of illustration. Except where otherwise noted, analogous results can be obtained for the evening peak and for off-peak hours, provided demand is high enough.) Its horizontal axis is time, representing the elapsed time since departure from the uptown terminal. The length of the horizontal axis is the round trip run time. The different bus stops along the route in each direction can be keyed to the time axis. The vertical axis is flow in passengers per hour. The curve whose slope varies is the load profile; it represents the number of passengers per hour on board at each point along the vehicle trajectory. The top line, parallel to the horizontal axis, is the capacity
Figure 3.1
Route Profile For a Typical Radial Route

Bus capacity = 70 pax \(^1\) /h

Required frequency = \(\frac{1400 \text{ pax/hr}}{70 \text{ pax/bus run}} = 20 \text{ bus runs/hr}\)

Required \# of buses = \(\left(\frac{20 \text{ bus runs/hr} \cdot 50 \text{ min/run}}{60 \text{ min/hr}}\right)\)

= 27 buses

\(^1\) pax is used as an abbreviation for passengers
profile of a conventional local route. It represents the capacity available (in passenger places per hour) at each point along the vehicle trajectory. It is horizontal because a conventional local route provides the same capacity all along the route in both directions.

The total number of vehicles of capacity C that are needed is roughly the total area under the capacity profile (which is the total number of places needed) divided by C. (The integer constraint on the number of buses makes this measure approximate.) To reduce the number of vehicles needed, then, we would like to reduce the area under the capacity profile.

Notice the shape of the load profile. On the inbound portion of the trajectory (left side of Figure 3.1) it rises to a peak that is near the downtown terminal. This typical load profile occurs because more people are boarding than alighting over the outer portion of the route, and so the vehicles get more and more crowded until the peak load point is reached (the apex of the load profile). The peak load point on radial routes is generally near the downtown terminal, since destinations are concentrated around the downtown. The outbound portion of the load profile (right side of Figure 3.1) displays less peaking, but its important characteristic is that its peak load is far lower than the peak load for the inbound portion of the trip.

The area below the load profile represents used capacity; the shaded area between the load profile and the capacity profile represents unused capacity. It is often the case that more than half of the capacity provided is unused, even if the bus is loaded to capacity at its peak load point. Providing unused capacity greatly lowers the productivity of any
system. Thus, in our search for more efficient routing strategies, it makes sense to direct our first efforts at reducing the amount of unused capacity.

The greatest amount of unused capacity is in the outbound direction. Can it be reduced? Conservation laws require that as many buses per hour go outbound as go inbound if the system is to maintain itself. But it is not necessary that all the outbound buses go in service. Because the outbound demand is so much less than the available capacity, we could have some buses provide service on their way back to the uptown terminal while the remainder deadhead (return empty). The advantage of this strategy is that the deadheading buses could return to the uptown terminal faster, since they would not be subject to the delays of stopping for passengers and they could use the fastest path available. If an expressway were available, the savings could be substantial; if not, the freedom to avoid congested areas, use one-way streets, and take advantage of progressively timed signals could still provide enough time savings to make this strategy worthwhile. The hatched area in Figure 3.2 illustrates the potential savings in required vehicles that this strategy can yield.

The above strategy is called "partial deadheading". The design issues with partial deadheading are to know how many buses per hour to deadhead, and what will be the savings in vehicles. One may also be interested in passenger wait time, which will increase since outbound passengers will no longer enjoy the same service frequency. This topic is taken up in Chapter 4.

The second largest area of unused capacity is in the uptown end of the inbound portion of the route, i.e., the upper left corner of Figure 3.1.
Figure 3.2
Partial Deadheading On a Radial Route

Time saved by deadheading = \( \frac{(80-65) \text{ min}}{60 \text{ min/hr}} \) = 0.25 hr.

Approximate frequency deadheading = \( \frac{(1400-400) \text{ pax/hr}}{70 \text{ pax/bus run}} \) = 14.3 bus runs/hr

Approximate vehicle savings = 0.25 hr \cdot 14.3 \text{ bus runs/hr}
= 3.6 buses
To reduce this excess capacity would mean to not start all of the buses at the uptown terminal, but to start some of them at one or more intermediate terminals. Since loads in the outbound direction are usually smaller than those in the inbound direction, the buses that start at intermediate terminals would have to return only as far as their starting terminal. Thus there are capacity savings at both the upper left and upper right corners of the route profile, as illustrated in Figure 3.3. Still greater capacity savings could be achieved by deadheading some of the buses, as indicated in Figure 3.3 by the greater savings on the outbound portion.

Operating some buses over less than the entire length of the corridor is called "zonal service". When all of the buses operate as local buses (perhaps with some boarding or alighting restrictions), this strategy is "local zonal service". There are a lot of design issues in local zonal service, the primary one being whether the services should overlap, i.e., whether a passenger should be allowed to board any bus that comes by or whether he should be restricted to boarding only the shortest route serving him. Besides this issue, the design questions are what frequency of service should be operated out of each potential terminal with what vehicle type, and what deadheading strategy should be used, in order to minimize a sum of operator and waiting cost. These issues are taken up in Chapter 5.

A third strategy for reducing the total vehicular capacity needed is to serve passengers destined for downtown with express buses. Of course, local passengers must also be served, so there must be local service along side the express service. One motivation for using express
Figure 3.3
Local Zonal Service On a Radial Route

Time savings from zoning and deadheading = \[ \frac{[15-0] + (80-60]}{60 \text{ min/hr}} = 0.58 \text{ hr} \]

Approximate frequency of shorter route = \[ \frac{(1400-800) \text{ pax/hr}}{70 \text{ pax/bus run}} = 8.6 \text{ bus runs/hr} \]

Approximate vehicle savings = 0.58 hr \cdot 8.6 \text{ bus runs/hr} = 5.0 \text{ buses}
service is that express buses can use faster paths, and thus be more productive. Thus it is not surprising to see express routes operating on expressways. A second, not so obvious, motivation for using express service is to overcome some of the problems of zonal service. These problems, which will be described in Chapter 5, can make express service worthwhile even when the express buses use the same streets as local buses.

Designing express and local services together is studied in Chapter 6. The main question there is how to segment the market, i.e., which passengers should be served with express buses and which with local buses. Given a market segmentation, it makes sense to allow both the express and local services to be zonal, and to allow some buses to deadhead. The design is further complicated by the potential existence of transferring passengers when the outer portion of the corridor is served by only local or only express service. Figure 3.4 portrays the route profile in the presence of express service, showing the potential for vehicle savings.

Thus a careful, imaginative look at a route profile suggests three strategies that have the potential to significantly increase the efficiency of transit service in high density corridors. But before getting into the technical aspects of these strategies, which are essentially combinations of a number of single bus routes, we will focus on the issue of optimizing the service level on a single bus route. Since conventional local service is a single bus route, Section 3.2 will give the optimal design strategy for conventional local service and will serve as a component of the design procedure for the more complex strategies covered later.
Figure 3.4
Express/Local Zonal Service On a Radial Route

a) Express Market

b) Local Market

Express buses needed = \left< \frac{650}{70} \cdot \frac{55}{60} \right> = 9

Local buses needed = \left< \frac{400}{70} \cdot \frac{80}{60} \right> + \left< \frac{750-400}{70} \cdot \frac{60-15}{60} \right> = 12

Total buses needed = 21

Vehicle Savings = 6
3.2 Optimal Headway and Bus Type on a Single Route

Given a bus route, whether an ordinary local route covering the entire length of a corridor or a more specialized route that is part of a more complex routing configuration, the design question becomes what type vehicle should be operated on this route, and running at what headway, in order to minimize operating plus passenger cost, appropriately weighted.

The cost of a single route has been modeled in the case studies of this research as a weighted sum of operating cost, total passenger wait time, and total passenger in-vehicle time. One could easily add additional cost components, such as a crowding penalty or an express service bonus (since the express portion of a route is more comfortable, being more free of stops and starts, and the characteristics of the company one would expect to find on an express route make it like riding in a first class compartment).

Operating cost was assumed proportional to the number of vehicles used. This approximation is good, since the dominant labor cost, as well as the capital carrying cost, are proportional to the number of vehicles used. Fuel, maintenance, and depreciation costs are proportional to the number of runs made, which is almost proportional to the number of vehicles used. The "almost" of the last relationship arises because of differences in speed and length of the layover at different headways; but these differences are small, affecting the per hour vehicle cost by only a few percent.
The number of vehicles used at a given headway is given by the basic relationship:

\[ B_y = \left\langle \frac{RT(h)}{h} \right\rangle \quad (3.1) \]

where
- \( y \) = vehicle type chosen
- \( h \) = average headway in minutes
- \( B_y \) = number of vehicles of type \( y \) needed
- \( RT(h) \) = run time in minutes
- \( \langle x \rangle \) = smallest integer greater than or equal to \( x \)

Run time is a function of headway because the headway determines the expected number of passengers who will be boarding and alighting, as well as the expected number of stops that must be made, thus determining the expected total dwell time. Equations (2.3) and (2.4) give the expected dwell time at a stop as a function of headway. In order to make it easier to model run time on the entire route, which requires an aggregation of dwell times at all stops, total dwell time along the route was approximated as a linear function of headway. This approximation was used because the dependence of run time on headway is small, and because in the range of interest their relationship is nearly linear. Thus run time was modeled as:

\[ RT(h) = \alpha + \beta h \quad (3.2) \]

To estimate \( \alpha \) and \( \beta \) for a given route, run time was computed for two headways that were in the range of likely headways (one near the upper end of that range, one near the lower end) using the exact formula for expected dwell time (2.3) and aggregating over stops. Then \( \alpha \) and \( \beta \) were computed as the intercept and slope of the line joining these points.
Thus, the operating cost of a single route using vehicle type \( y \) was given by:

\[
OC_y(h) = C_y B_y
\]
\[
= C_y \left( \frac{RT(h)}{h} \right)
\]
\[
= C_y \left( \frac{\alpha}{h} + B \right)
\]

(3.3)

where \( OC_y(h) \) = cost of operating vehicle type \( y \)
\( C_y \) = cost per hour of using vehicle type \( y \)

Total in-vehicle time for any given route requires knowing the vehicle trajectory \( (t_i(h)) \), i.e., the time a bus leaving the first stop at time 0 will arrive at each stop \( i \). Given this trajectory and the route origin-destination matrix \( ((d_{ij})) \), the total in-vehicle time incurred per hour \( IVT(h) \) is:

\[
IVT(h) = \sum_{i,j} d_{ij} \left[ t_j(h) - t_i(h) \right]
\]

(3.4)

Finding the vehicle trajectory, like the run time, requires using the dwell time formula (2.3). Therefore, a similar approximation was made because of the aggregation problem. Vehicle trajectories and total in-vehicle time were computed for headways in the upper and lower ends of the range of likely headways, and then the line joining those points was taken as an approximation of total in-vehicle time:

\[
IVT(h) = \gamma + \delta h
\]

(3.5)

where \( \gamma \) is the intercept and \( \delta \) the slope of the approximation line.

Wait time per rider, as explained in Section 2.1.5, was modeled as a constant fraction \( b \) of the headway. Therefore, total wait time incurred
per hour on the route is:

\[ WT(h) = Rbh \]  \hspace{1cm} (3.6)

where \( WT(h) \) = total wait time per hour
\( R \) = total ridership per hour

Therefore, the total cost per hour of a single route, given that vehicle type \( y \) is used, is:

\[ z(y,h) = C_y \left( \frac{\alpha}{h} + \beta \right) + C_t (\gamma + \delta h) + C_w Rbh \]  \hspace{1cm} (3.7)

where \( C_t \) = in-vehicle time cost
\( C_w \) = wait time cost
\( z(y,h) \) = cost of operating a route with headway \( h \) and vehicle type \( y \)

As explained in Section 2.1.4, when headways are not restricted to clockface values and number of buses is constrained to be an integer, there is a dominant set of headways, the headways for which \( \left( \frac{\alpha}{h} + \beta \right) \) is an integer. To prove this, suppose \( \left( \frac{\alpha}{h_1} + \beta \right) = k+r \), where \( k \) is an integer and \( r \) is a real number between 0 and 1. Then the number of buses needed is \( k+1 \), and the headway \( h_1 \) is \( \alpha/(k+r-\beta) \). Compare this solution with the headway \( h^* \) for which \( \left( \frac{\alpha}{h^*} + \beta \right) = k+1 \), i.e., \( h^* = \alpha/(k+1-\beta) \). With \( h = h^* \), \( k+1 \) buses are needed, so that vehicle cost is the same in both cases; but because \( h^* < h_1 \), both the in-vehicle time and wait time costs are less. Therefore, any solution \( h_1 \) for which \( \left( \frac{\alpha}{h_1} + \beta \right) \) is not an integer is dominated by the solution \( h^* \), for which \( \left( \frac{\alpha}{h^*} + \beta \right) = \left( \frac{\alpha}{h_1} + \beta \right) \). These dominant solutions are the headways for which there is no excess layover.
When there is a discrete set of acceptable headways, one can find the optimal headway by enumerating them. This approach is taken by Hagberg and Hasselstrom [4], who require clockface headways. When we are not constrained to a particular set of headways, however, there is a computationally more efficient method. We can relax the integer constraint on the number of buses that is implicit in (3.7) by replacing the brackets in the first term by a set of parentheses. It is obvious that this relaxation, call it \( z'(y,h) \), is convex. Since for the dominant set of headways \( \frac{\alpha}{h} + \beta \) is an integer, \( z'(y,h) \) is a convex envelope of the dominant solutions. We must first find the headway \( h_u \) that minimizes the envelope \( z'(y,h) \); then the optimal member of the set of dominant headways must be one of the two members that straddle \( h_u \), provided it does not exceed the maximum headway. If the optimal headway thus found exceeds the maximum headway (the lesser of the policy headway and the headway at which the loading constraint becomes binding), then the optimal dominant headway is the greatest feasible headway.

With the modeling assumptions we have made, \( h_u \) can be solved for in closed form. The first derivative of \( z'(y,h) \) with respect to \( h \) is:

\[
\frac{\partial}{\partial h} z'(y,h) = -\frac{\alpha c_y}{h^2} + C_t \delta + C_w R_b
\]

By setting it equal to zero, we find:

\[
h_u = \left( \frac{\alpha c_y}{C_t \delta + C_w R_b} \right)^{0.5}
\]

Then, since \( h_u \) would require \( \left( \frac{\alpha}{h_u} + \beta \right) = k_u \) buses, the dominant solutions that straddle \( h_u \) are those that require \( k_u \) and \((k_u - 1)\) buses. These two
dominant headways are \( h_1 = \frac{\alpha}{u^{\beta}} \) and \( h_2 = \frac{\alpha}{u^{1-\beta}} \). Then whichever of these two headways minimizes \( z(h,u) \) is the optimal headway, provided the maximum headway constraint is not exceeded. If it is exceeded, then we know that we must use \( k_m = \left( \frac{\alpha}{h_m} + \beta \right) \) buses, where \( h_m \) is the maximum headway, so that the feasible dominant solution closest to \( h_m \), which is the optimum, is \( \frac{\alpha}{k_m^{1-\beta}} \).

Thus, it has been shown how the optimal service level can be found for a given vehicle type. To find the best vehicle type, simply compare the optimal of each vehicle type and choose the best.

3.3 Introduction to the Boston Case Study

Two sites were used for case studies in this research. The Watertown corridor in the Boston metropolitan area served to illustrate routing strategies on a single arterial corridor, both with single and with multiple downtown terminals. The Chicago Avenue corridor in the Minneapolis - St. Paul metropolitan area was used to illustrate routing strategies in a corridor with a branching route network.

The case studies of Chapters 3, 4, 5, 6, and 8 all deal with the Watertown corridor. In order to avoid repetition in each chapter, the Watertown corridor is described in detail in this section. The Chicago Avenue corridor is used only in the case study of Chapter 7 and it will be introduced in that chapter.

In addition to describing the Watertown corridor, this section describes the common factors used in the case studies of the Watertown corridor: the objective functions used, the data used, the vehicle types available, and
some assumptions about markets served that were made for the sake of illustration.

3.3.1 Geography and Existing Routes in the Watertown Corridor

The site of Boston case study is the corridor along the no longer used tracks of the Watertown streetcar line. It begins at Watertown Square, the center of a fairly dense suburb 7 miles west of Boston, passes through Newton Corner (pronounced "kaw'-nuh"), and then goes through Brighton, a working class area of Boston, to the Back Bay, a dense commercial and residential district just west of downtown, and finally to the downtown. Figure 3.5a illustrates the corridor.

The corridor is served by a local route, Route 57, and by three express routes, all of which use the Massachusetts Turnpike from Newton Corner to central Boston. Route 301 begins at Brighton center and proceeds in the reverse direction to Newton Corner, then goes to downtown via the turnpike. Route 304 begins at Watertown Square, proceeds to Newton Corner, then goes to downtown via the turnpike. Route 302 begins at Watertown Square, proceeds to Newton Corner, and then goes along the turnpike to Copley Square, which is in the center of the Back Bay commercial district. Local Route 57 goes from Watertown Square to Kenmore Square, which is on the outer edge of the Back Bay. Riders on Route 57 destined for Copley Square or downtown must board the subway at Kenmore. Downtown is an attractive destination not only because of its own activities, but because transfers to other subway lines all occur in the downtown. Figure 3.5b illustrates the existing services in this corridor.
Figure 3.5
The Watertown Corridor

a. Physical Network

b. Existing Service in the Watertown Corridor
For the purposes of illustrating the routing strategies discussed in this research, this corridor has advantages and disadvantages. Boston routes in general have a disadvantage in that Boston's bus system serves primarily as a feeder to its rapid transit system, and most of the high demand corridors have rail service. Therefore Boston has few high demand corridors that extend into the downtown that are served by bus only. Boston routes, however, have an advantage, which is a survey done in 1978 that produced on/off counts for each route. This availability of data made it worth searching for an appropriate corridor for a case study in the Boston area.

The Watertown corridor, following an abandoned street car line through fairly dense neighborhoods, was chosen because it is probably the Boston area corridor with the greatest demand that is not served by rail, and because the subway station that 57 feeds, Kenmore Square, is close to downtown.

To illustrate some of the routing configurations it was proper to isolate Route 57 and consider improving the efficiency of operations on that route alone. For other applications it was more appropriate to look at the corridor as a whole, including the demand carried by the express routes. In order to make the illustrations as uncomplicated as possible, we assumed in these latter applications that buses would run inbound all the way to downtown, eliminating the need for transfers at Kenmore Square. Figure 3.6 illustrates the load profiles of both Route 57 and of a route that would serve the demands of the entire corridor. The large
Figure 3.6

Load Profiles in the Boston Case Study

a. Route 57: Watertown-Kenmore-Watertown

b. Entire Corridor: Watertown-Downtown-Watertown
shaded areas, indicating unused capacity when conventional local service is employed, indicate the potential for savings through the routing strategies that will be explored.

3.3.2 Data Used

On/off counts were available for a sample of runs on each route. The averages of these counts over the morning peak (7-9 a.m.) determined the trip distribution within the corridor. Recent peak load counts were used to scale the on/off counts up or down so that the peak loads implied by the on/off counts matched observed peak loads. The route O/D matrix was estimated using Tsygalnitzky's method (see Section 2.2). In splitting the passengers alighting at Kenmore Square between those destined for the Back Bay and those destined for downtown and points beyond, we assumed that the fraction of riders choosing one of those destinations was the same as the fraction of express riders boarding between Watertown Square and Newton Corner choosing that destination.

Most operating characteristics and parameters were supplied by the Massachusetts Bay Transportation Authority (MBTA). Run times were available between major points, and run times between minor stops were interpolated in proportion to distance. Layover time was taken as 28.5 percent of run time. Standard values of passenger movement rates were taken, 3 seconds per boarding passenger and 2 seconds per alighting passenger. The fixed delay per stop was taken as 2 seconds. Deceleration was assumed to be 3 mph/second, and acceleration 2 mph/second. Cruising speed on the local parts of the route was assumed to be 25 mph.
Two vehicle types were studied. The standard 40-foot buses that the MBTA operates seat 45 passengers; with a policy load of 1.2 in the peak period [16], these buses have a capacity of 54 passengers. The standard bus costs, using both MBTA figures and assumptions about average speed and vehicle life, are about $32.05 per hour, including wages, fuel, maintenance, and depreciation. The second vehicle type is the articulated bus, seating 65 people (yielding a capacity of 78), and costing $40.03 per hour. Appendix A shows how these vehicle costs were derived.

Including the full cost of vehicle depreciation is not something that United States transit operators do, because capital costs (which, as defined, include vehicle purchase) are financed differently from operating costs. Including capital costs makes the study take on the viewpoint of society, wherein capital costs must be accounted for because they are resources consumed. Operators have a different point of view, and should figure vehicle cost accordingly. Furthermore, an operator who is constrained as to how many vehicles of a given type he may have will need to add a shadow price to the cost of his vehicles. The chosen costs, then, are to serve the purpose of illustration, rather than to offer definitive advice to the MBTA on operations in the Watertown corridor.
3.3.3 Existing Level of Service

The existing scheduled level of service (measured in 1979 to early 1980) revealed that the routes were not as crowded as they could be. On Route 57, with a peak load of 575 passengers/hour and scheduled headways of 4 minutes, average load per bus was planned to be about 38. Actual (observed) headways were closer to 5 minutes, yielding an average load of 47 passengers. Peak loads on the express routes were also about 47 passengers per bus. Given the financial crisis in which the MBTA now finds itself, it is likely that service levels on these routes will be cut, if they have not been already.

3.3.4 Objectives and Constraints Used in the Case Study

For the purpose of illustrating the different service configurations with this case study, two consistent sets of objectives were used throughout. The "austerity objective" gave a tiny weight to passenger level of service and sought to minimize operator cost; passenger service was given a tiny weight so that ties would be broken in favor of the passengers. The "prosperity objective" valued passenger wait time and in-vehicle time at $3/hour and valued the transfer penalty at 20 cents per transferring passenger. Policy headway was taken to be 15 minutes. While the MBTA's policy headway for the peak is 30 minutes, such a long headway would probably be intolerable to riders in a busy corridor who were accustomed to service every 5 minutes.

As the streets in the Watertown corridor are only mildly congested in the morning peak, the average wait time was assumed to be 60 percent of average headway (see equation (2.1)).
The equal weights for wait time and in-vehicle time in the prosperity objective may need justification, since behavioral studies have repeatedly found that travelers consider a minute of wait time more onerous than a minute of in-vehicle time [17,18]. Measuring level of service by the tradeoffs observed in behavioral studies, then, one ought to value wait time greater than in-vehicle time. By same argument, a minute's delay to a high income traveler ought to be valued more than a minute's delay to a low-income traveler, since one can observe that high income travelers demonstrate through their travel choices a higher value of time than low income travelers.

The case studies do not value different levels of service components according to the value passengers seem to put on them (as reflected by their travel choices) because the author believes that level of service should not be measured this way in a service that is as heavily subsidized as transit (farebox revenue covers about one fourth of the MBTA's operating cost) and in which fares are not sufficiently differentiated according to level of service. If rich people value wait time highly, and so the service level is increased in high income areas to reduce the amount of their valuable wait time, then their fares should be raised to cover the entire cost of this service increase, and everyone will be better off. But with existing fare structures and subsidies, the additional cost of saving the rich people's wait time will be borne mostly by the taxpayers in general. But if it is the taxpayers' money that is spent or saved according to the values placed on different level of service components
such as wait time, then the taxpayers should have the prerogative of setting those values. Because it is unlikely that the taxpayers would place a different value on the time of rich and poor travelers, it is unwise to make decisions based on income-differentiated values of time. In the same way, it seems reasonable that taxpayers might be willing to have money taken from them in order to reduce the amount of time people spend traveling, thus increasing general accessibility, while they would not be willing to have money taken from them in order to give people more comfort beyond a comfort standard that is accepted by society. The premium on wait time value is not because one minute of wait time consumes any more time than one minute of in-vehicle time, but because of comfort. For this reason wait time is valued the same as in-vehicle time, since these weights reflect the willingness of population as a whole to be taxed in order to reduce the amount of time needed for travel by people of all income groups.

The values chosen for wait time and in-vehicle time affect the optimal design. Higher values of wait time tend to favor designs with less market segmentations, which have lower average wait times; higher values of in-vehicle time tend to favor solutions that include express routes.

In light of the above argument for equal wait time and in-vehicle time values, the value given as a transfer penalty may need justification, since avoiding transfers seems to be a comfort issue just as wait time. The undesirability of transfers is universally known, and people demonstrate an unusual willingness to give up both time and money to avoid transfers. Including a penalty on transfers can be reasonable even when
other kinds of comfort-related costs are not accepted because of the peculiar nature of a transit network. With no penalty associated with transfers, it is very possible for a transit system to be designed that requires so many transfers that it would be entirely unacceptable to the public. Certain policy decisions can be made that will limit the number of transfers; however, it may not be possible through policy decisions alone to guarantee any given standard for number of transfers in the same way that other comfort factors can be guaranteed. (For example, seat design can guarantee a certain level of seating comfort.) For this reason it will often make sense to put a penalty on transfers in addition to having policies designed to limit transfers. For example, all the routing configurations designed in this thesis assume as a policy that all routes will terminate at a downtown terminal. It may well be possible to terminate some routes a little bit before the downtown, making downtown-bound passengers transfer to other routes, but this configuration is not allowed under the above policy. This kind of policy puts an implicit penalty on transfers of that type. But suppose there are still too many transfers in the system even when this policy is imposed. Then the operator will make other decisions, which have certain costs, to further reduce the number of transfers. In such a case, since resources are already being consumed to help avoid making passengers transfer, it makes sense from an equity and an efficiency point of view to put a penalty on transfers in any new design, as well as to impose the routing policies that limit transfers, in order to maintain consistency in a transit system's effort to offer service that meets an acceptable standard for requiring few transfers.
3.4 Optimal Conventional Local Service in the Waterfront Corridor

Conventional local service, optimized under the assumptions of the models as presented, will be the benchmark for comparing the different strategies. Comparison with the existing level of service on the local route would be inappropriate because the objectives of the MBTA in setting the service level on Route 57 are not clear, and because there may be inefficiencies in the current system.

In order to provide a benchmark for all the configurations to be studied in the coming chapters, two different cases were optimized. The first case requires serving all of the demands in the corridor and models the bus route as extending in all the way to downtown. The second case requires serving only those demands currently served by Route 57 with a route whose inner terminal is Kenmore Square. These cases were optimized for both the austerity and the prosperity objectives. Again, for the purpose of comparison, separate solutions are given for each vehicle type. Table 3.1 shows the solutions for serving the entire corridor and Table 3.2 shows the solutions for serving the Route 57 demands.

To serve the entire corridor at the least operating cost with conventional local service, 34 articulated buses are needed, yielding a headway of 2.5 minutes and an operating cost of $1,361 per hour. The optimal design under the prosperity objective is 46 standard buses for a headway of 1.7 minutes and an operating cost of $1,474 per hour. The buses are still loaded to capacity under the prosperity objective, but the smaller headways that come from using smaller vehicles reduce the wait time and in-vehicle time so that the sum of passenger costs, valued at $3/hour,
Table 3.1

Conventional Local Service For the Watertown Corridor

<table>
<thead>
<tr>
<th></th>
<th>Using Articulated Buses 1/</th>
<th>Using Standard Buses 2/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Headway</td>
<td>2.5 min</td>
<td>1.7 min</td>
</tr>
<tr>
<td>Number of Vehicles</td>
<td>34</td>
<td>46</td>
</tr>
<tr>
<td>Average Wait Time</td>
<td>1.5 min</td>
<td>1.0 min</td>
</tr>
<tr>
<td>Average In-Vehicle Time</td>
<td>22.3 min</td>
<td>20.8 min</td>
</tr>
<tr>
<td>Operator Cost</td>
<td>$1,361</td>
<td>$1,474</td>
</tr>
<tr>
<td>Total Cost (pax time valued at $3/hr)</td>
<td>$4,339</td>
<td>$4,199</td>
</tr>
</tbody>
</table>

1/ Austerity solution; also prosperity solution with articulated buses only

2/ Prosperity solution; also austerity solution with standard buses only
and operator cost is less with the standard buses. It is interesting that even when passenger travel time is valued at $3/hour, it is still better to operate standard buses at capacity rather than operate at less than capacity. This is so because even with full buses the headways are small, and the marginal benefit of adding another bus to reduce headway declines as the headway gets small. With larger vehicles and smaller demands the prosperity objective will often lead to operating at less than capacity in order to reduce passenger time costs.

To serve the Route 57 passengers only, the austerity objective requires 11 articulated buses for a headway of 7.2 minutes and an operator cost of $440 per hour. Under the prosperity objective it is better to use standard buses; 14 are needed at a headway of 5.3 minutes for an operator cost of $449 per hour. For such a small increase in operating cost ($9/hour), passenger time costs, valued at $3/hour, decline by $127/hour by switching to the prosperity solution. Again the prosperity solution has the buses running at capacity since the maximum headway for standard buses on this route, 5.3 minutes, is already quite small; however, if restricted to articulated buses, the prosperity solution uses 2 vehicles more than the minimum.

Recall that existing service on Route 57 is scheduled to operate with standard buses at a headway of 4 minutes, but actually operates with headways closer to 5 minutes. At a 4 minute headway 18 buses are needed, for an operating cost of $577 per hour. Average wait time becomes 2.4 minutes, and average in-vehicle time 11.9 minutes, so that with 1,205 riders total passenger costs are $863 per hour. Thus, the sum of passenger and operator
Table 3.2

Conventional Local Service For Route 57 Demands

<table>
<thead>
<tr>
<th></th>
<th>Using Articulated Buses</th>
<th>Using Standard Buses</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AUSTERITY SOLUTION</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Headway</td>
<td>7.2 min</td>
<td>5.3 min</td>
</tr>
<tr>
<td>Number of Vehicles</td>
<td>11</td>
<td>14</td>
</tr>
<tr>
<td>Average Wait Time</td>
<td>4.3 min</td>
<td>3.2 min</td>
</tr>
<tr>
<td>Average In-Vehicle</td>
<td>13.5 min</td>
<td>12.6 min</td>
</tr>
<tr>
<td>Time</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Operator Cost</td>
<td>$ 440</td>
<td>$ 449</td>
</tr>
<tr>
<td>Total Cost</td>
<td>$1,517</td>
<td>$1,399</td>
</tr>
<tr>
<td>(pax time valued at $3/hr)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PROSPERITY SOLUTION</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Headway</td>
<td>5.8 min</td>
<td>5.3 min</td>
</tr>
<tr>
<td>Number of Vehicles</td>
<td>13</td>
<td>14</td>
</tr>
<tr>
<td>Average Wait Time</td>
<td>3.5 min</td>
<td>3.2 min</td>
</tr>
<tr>
<td>Average In-Vehicle</td>
<td>12.8 min</td>
<td>12.6 min</td>
</tr>
<tr>
<td>Time</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Operator Cost</td>
<td>$ 520</td>
<td>$ 449</td>
</tr>
<tr>
<td>Total Cost</td>
<td>$1,505</td>
<td>$1,399</td>
</tr>
<tr>
<td>(pax time valued at $3/hr)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
costs is $1,440 per hour, whereas with 5.3 minute headways it is $1,398 per hour. The difference is about the cost of operating 1.3 standard buses. Four minute headways would be optimal if passenger time were valued at $6/hour, but that is a very generous figure. Thus, it appears to be good that actual headways are closer to 5 minutes. Operating at 5 minute headways instead of 4 minute saves 3 buses, and still reflects a considerable value of passenger travel time, about $3.50 per hour.

3.5 Conclusions

An examination of the load profile of a typical radial corridor has revealed that there may be more efficient strategies for serving a corridor than conventional local service. Low peak loads in the light direction suggested deadheading some light direction runs; low loads on the outer portion of the route suggested a zonal service strategy wherein some runs cover less than the length of the whole corridor; and a concentration of demand for service to downtown suggested jointly offering express and local service. These three strategies are the subject of the next three chapters.

This chapter has also described the way in which the operation of a bus route is modeled. This model of route operation makes run time and passenger service level sensitive to the number of people boarding and alighting, both in determining the expected number of stops a bus will make and the duration of each stop. The model also accounts for operational factors such as speeds, deceleration and acceleration rates, headway variability, and layover times. This route operation model is a basic component of the models of the different routing strategies that are studied in this thesis.
This chapter then showed how, given the demand pattern and the operating parameters of a single route, one could find the optimal headway, number of vehicles, and passenger level of service. Since conventional local service is a routing strategy that consists of a single route, we therefore have the way to find the optimal service level for conventional local service.

After introducing the site for the Boston case study, the route operations model was applied to the Watertown corridor to find the optimal service level for conventional local service under a couple of different objectives and assumptions. The two objectives were called the austerity objective, in which operator cost alone is minimized subject to constraints on service level, and the prosperity objective, in which a sum of passenger costs (wait time, in-vehicle time, and transfer cost) and operator cost are minimized subject to the same service level constraints. The two different modeling assumptions concerned the market served by the route. One assumption was that the route was to carry the demand of the entire corridor (in order to illustrate a high demand situation); the second assumption was that the route was to carry only the same demands that are now carried on local Route 57 (in order to illustrate a moderate demand situation).

Optimizing conventional local service in the Watertown corridor led to four sets of results due to the two objectives and the two assumptions about the market served. These results will be used as benchmarks for assessing the value of the other routing strategies to be studied in later chapters under the same objectives and market assumptions.
Chapter 4
Partial Deadheading

It is common on radial routes for there to be a large load imbalance favoring one direction during the peak periods. In such situations a conventional local route, in order to provide adequate service in the heavy direction, will offer a much greater level of service than is necessary in the light direction. In such situations it is possible to deadhead some of the buses in the light direction while keeping the crowding and wait time levels in the light direction tolerable; this strategy is called "partial deadheading". The deadheading buses will be able to travel faster than if they were in service, especially if there is a high speed path available to them such as an expressway, and thus there is potential to convert this saving in run time into savings in operator and/or passenger cost. The problem this chapter tries to solve is: how much can be saved through the partial deadheading, and how can one find the vehicle schedule that achieves the optimum savings?

First, in Sections 4.1 and 4.2, we shall try to find the schedule that minimizes the number of vehicles needed. Section 4.1.2 provides a review of the literature addressing this problem. Then, in Section 4.3, we will try to minimize a sum of vehicle and passenger costs. Section 4.4 presents a prototypical design procedure that an operator could follow to design partial deadheading services manually. Section
4.5 applies the deadheading strategy to the Watertown corridor and shows its potential for significantly reducing costs. Section 4.6 offers conclusions regarding this routing strategy.

4.1 Minimum Number Of Vehicles Under Fixed Headways

Given a corridor with its demand pattern, the peak load and the policy headway determine the maximum headway that may be offered in each direction. For convenience let us assume we are dealing with the morning peak, so that the heavy direction is inbound. Then there will be a maximum headway in the inbound direction, $h^m_A$, and a maximum headway in the outbound direction, $h_B^m$ (of course, $h_B^m > h_A^m$). Intuitively it makes sense, if we are trying to minimize the number of buses needed, to provide no more service than these maximum headways dictate. The question is, given that we are to operate at these maximum headways, or any other arbitrarily fixed headways, how many buses do we need?

Let us introduce some notation that we will use throughout this chapter.

\[
\begin{align*}
  h_A & \quad = \text{service headway in the heavy direction} \\
  h_A^m & \quad = \text{maximum service headway in the heavy direction} \\
  h_B & \quad = \text{service headway in the light direction} \\
  h_B^m & \quad = \text{maximum service headway in the light direction} \\
  t_A & \quad = \text{run time (including minimum layover) in the heavy direction} \\
  t_B & \quad = \text{run time (including minimum layover) in the light direction for the in-service vehicles} \\
  t_D & \quad = \text{run time (including minimum layover) in the light direction for deadheading vehicles}
\end{align*}
\]
4.1.1 Minimum Number Of Vehicles With Unevenly Spaced Departures

A basic relationship of scheduled carrier services is that the number of vehicle-hours of service needed per hour is the product of the run time and the service frequency. If run time is $t$ hours per run, and there must be $q$ vehicle runs per hour, we need $(tq)$ vehicle-hours of operation per hour. If we are able to schedule vehicles in such a way that each vehicle could run continuously, i.e., provide up to one vehicle-hour of operation each hour, then the number of vehicles we will need is $(tq)$, where $(x)$ is the smallest integer greater than or equal to $x$. If vehicles can provide only a fraction of an hour of operation each hour, obviously we may need more than $(tq)$ vehicles. There is no way we could get by with less than $(tq)$ vehicles, since vehicles cannot operate any more than an hour in any given hour, and vehicles are indivisible (under an assumption of no interlining).

Thus, if we want to minimize the number of vehicles needed, it seems obvious that we should operate our vehicles as close as possible to continuously. (Minimum layover is considered part of the necessary operation; a vehicle is considered to be not operating only after its layover exceeds the minimum.) Under these assumptions, we can compute the minimum number of vehicles needed on a route by breaking the route into three parts. The inbound portion, with a run time of $t_A$ and headway $h_A$, requires $t_A/h_A$ vehicle-hours of operation per hour (since frequency is the inverse of headway). The outbound in-service portion requires $t_B/h_B$ vehicle-hours of operation per hour. Conservation of flow tells us that the deadheading frequency must be the difference of the inbound frequency and the outbound in-
service frequency, or \((\frac{1}{h_A} - \frac{1}{h_B})\). Thus, the deadheading portion of the route requires \(t_D \left(\frac{1}{h_A} - \frac{1}{h_B}\right)\) vehicle-hours of operation per hour. Under our assumption that continuous vehicle operation is possible, then the number of vehicles needed is:

\[
N_z(h_A, h_B) = \left\langle \frac{t_A}{h_A} + \frac{t_B}{h_B} + t_D \left(\frac{1}{h_A} - \frac{1}{h_B}\right) \right\rangle
\]  

(4.1)

where \(N_z(h_A, h_B)\) is the minimum number of vehicles needed, assuming continuous operation.

What will service be like if vehicles are operated almost continually as assumed above? Suppose \(h_A = 4\) minutes, \(h_B = 9\) minutes, \(t_A = 20\) minutes, \(t_B = 19\) minutes, and \(t_D = 13\) minutes. Then \(N_z(h_A, h_B) = \langle 8.92 \rangle = 9\). One can try to schedule 9 vehicles on this hypothetical route, operating continuously. One will find that while it is possible to achieve average service headways of \(h_A\) and \(h_B\), the actual headways (spacings between departures) will vary widely. Some departures will be spaced very closely; others will have a large gap between them.

Even spacing of departures is critical when buses are operating near their maximum headways. If spacing is uneven, some buses will be overcrowded and others underloaded. Buses will bunch up, and some passengers won't be able to board the first (or perhaps second) bus that comes. For this reason, \(N_z(h_A, h_B)\) is generally an unacceptable solution for passenger service scheduling. (In contrast, a freight carrier may find \(N_z(h_A, h_B)\) acceptable because he can better control how much cargo boards each vehicle and
because crowding and wait time often aren't as important in freight operations.)

For this reason our problem should be restricted to ensure even departure spacings.

4.1.2 Minimum Number Of Vehicles With Evenly Spaced Departures: Review Of The Literature

Given fixed headways, the only thing lacking to know the schedule of all in-service runs is the relative phasing of inbound and outbound trips. Intuitively we would expect the best phasing to be one that has at least one inbound run terminating exactly when an outbound in-service run begins (Section 4.1.3 proves this analytically). We can arbitrarily impose this constraint on the first inbound run; then our schedule of in-service runs is complete. Next our problem is to minimize the number of vehicles needed to meet a given schedule. This problem has been studied in the literature using numerous approaches.

One approach is to construct a network in which each service run is represented as a node, and arcs are drawn between every node pair (i,j) for which it is possible for a bus to perform run j after performing run i. This network is a graphical representation of a partially ordered set. The network can be decomposed into a set of chains, where a chain is a direct path in the network, such that every node belongs to exactly one chain. (A chain may also be a single node.) Because of the way the network has been constructed, a chain represents a series of runs that can be performed by a single vehicle. Thus, minimizing the required number of vehicles is equivalent to finding the decomposition of the network with the minimum number of chains. Dilworth [19] has shown that the
number of chains in a minimal decomposition equals the maximum number of mutually unrelated nodes in the network, where nodes i and j are unrelated if no arc exists between i and j.

The question still arises how to find this minimal decomposition, or equivalently, the maximum number of mutually unrelated nodes. Dantzig and Fulkerson [20] showed that the problem of minimizing the number of vehicles necessary to meet a fixed schedule can be formulated as a linear program in the form of a transportation problem and therefore solved by the simplex method. Later, it was shown that Dilworth's theorem can be deduced from linear programming and duality theory [21].

As a by-product of their work on maximum flow, Ford and Fulkerson showed that the number of mutually unrelated elements in a partially ordered set could be found by solving a maximum flow problem [22]; by Dilworth's theorem, then, they had discovered a way to find the minimum number of vehicles to meet a fixed schedule. The maximum flow problem can be solved using Ford and Fulkerson's maximum flow algorithm [22], or more efficient algorithms that have recently been developed [23]. For the problem of scheduling vehicles on a single route, either the out-of-kilter or another maximum flow algorithm could be executed either manually in a reasonable amount of time, or on a computer at very little cost.

These algorithms are quite efficient, and are general enough to be applied to any fixed schedule. The special form of the partial deadheading schedule, however, makes it possible to solve for the minimum number of vehicles in closed form. This is a considerable advantage over algorithms that require labeling of nodes, updating of labels, balancing of flows, and so on. The derivation of this closed form solution is given in the next section.
4.1.3 Minimum Number Of Vehicles With Evenly Spaced Departures: A New Approach

This approach is best introduced by means of an example. Suppose the parameters of the problem are:

\[ t_A = 25 \text{ min.} \]
\[ t_B = 24 \text{ min.} \]
\[ t_D = 12 \text{ min.} \]
\[ h_A = 6 \text{ min.} \]
\[ h_B = 16 \text{ min.} \]

Without deadheading, i.e., headways of 6 minutes in both directions, this route would require \[ N_n = \left\lfloor \frac{25 + 24}{6} \right\rfloor = 9 \text{ buses; with partial deadheading, but irregular headways, it would require} \]

\[ N_L = \left\lfloor \frac{25}{6} + \frac{24}{16} + 12 \left( \frac{1}{6} - \frac{1}{16} \right) \right\rfloor = 7 \text{ buses.} \]

(Recall that \( \lfloor x \rfloor \) is the smallest integer greater than or equal to \( x \).)

The space-time diagram in Figure 4.1 shows a possible schedule for this route. Inbound and outbound departures are phased so that slack time (in excess of necessary layover) at the turn-around point for the first round trip has been set to zero. The left side of the diagram represents the uptown terminal; the right side, the downtown terminal. There are 5 types of arcs representing bus movements. "Inbound arcs" are directed from left to right at 6 minute intervals, requiring 25 minutes. "Outbound arcs" are directed from right to left at 16 minute intervals, requiring 24 minutes. "Deadhead arcs" are also directed from right to left, departing immediately after every inbound arrival, requiring 12 minutes. "Wait arcs" are directed from every node to the node immediately
Figure 4.1

Network for Solving Partial Deadheading Problem

Note: All arcs are directed forward in time (downward in the diagram).
below it on each side. "Garage arcs" are directed from the garage (source) to the top node on each side and from the bottom node to the garage (sink).

Each arc has a minimum required flow and a maximum required flow, which are the lower and upper bound on the flow, respectively. Each service arc, i.e., each inbound and outbound arc, has a minimum required flow of 1, and an allowed flow of $\infty$. Each deadhead, wait, and garage arc has a minimum required flow of 0, and an allowed flow of $\infty$.

Then the problem of finding the minimum number of buses required to meet this schedule is equivalent to finding the minimum flow in this network that satisfies the flow requirements of every arc. This problem can be solved using the "min-flow max-cut" theorem (analogous to the better known "max-flow min-cut" theorem), which states that the minimum feasible flow in a network with a single source and single sink is equal to the flow requirement of the cut with the greatest flow requirement [24]. In this context, a cut $(X, \overline{X})$ is defined as an imaginary line that separates the set of nodes in the network into two mutually exclusive and collectively exhaustive subsets, $X$ and $\overline{X}$, with subset $X$ containing the source node and subset $\overline{X}$ containing the sink node. If we define the minimum required flow on the arc directed from node $i$ to node $j$ as $\ell_{ij}$, and the maximum allowed flow as $u_{ij}$, then the flow requirement of cut $(X, \overline{X})$, denoted $R(X, \overline{X})$, is defined as:

$$R(X, \overline{X}) = \sum_{i \in X} \ell_{ij} - \sum_{j \in X} u_{ij}$$  \hspace{1cm} (4.2)

(Assume that $\ell_{ij} = u_{ij} = 0$ if there is no arc directed from $i$ to $j$.) The first term of (4.2) implies that the flow requirement of a cut increases by 1 each time it intersects either an inbound or outbound arc as long as the cut puts the start node of that arc into the graph subset containing the source and the end node of that arc into the graph subset containing the sink.
The second term implies that a cut which, if one follows it from left to right, intersects a deadhead arc \((i,j)\) from below cannot be a maximum cut, since it would put node \(i\) into subset \(\overline{X}\) and node \(j\) into subset \(X\), and then its cut requirement would be \(-\infty\). For the same reason, a maximum cut cannot intersect an outbound arc from below or an inbound arc from above (as one follows the cut from left to right). A cut (which happens to be a maximum cut) is shown in Figure 4.1; since it traverses 7 inbound arcs and 1 outbound arc, it has a total flow requirement of 8. The reader can verify that all of the inbound and outbound arcs can be covered with 8 buses, but no less, proving that it is a maximum cut. Verifying that one has a maximum cut is easy, but how can one find the maximum cut?

First, let us show that a maximum cut will lie entirely between adjacent deadhead arcs. Let us now fix the left end of a cut and make the cut lie entirely between two adjacent deadhead arcs, and let it be positioned so that it cuts the maximum number of service arcs (inbound and outbound arcs) that enter the "frame" between those two deadhead arcs, taking care not to cut an outbound arc from below or an inbound arc from above as we follow the cut from left to right. We cannot rotate the right end of the cut upward beyond the upper deadhead arc spanning that frame without intersecting that deadhead arc from below. Suppose that we rotate the right end of the cut downward so that the cut traverses one deadhead arc (from above) and enters the next lower frame. By this move the cut intersects one less inbound arc; and since each frame has a width of only \(h_A = 6\) minutes, and outbound arcs are spaced at intervals of \(h_B = 16\) minutes, the cut can intersect at most one more outbound arc by moving to the next frame. As the cut rotates farther downward on the right side, it intersects one less inbound arc every 6 minutes, and intersects one more out-
bound arc only every 16 minutes. Therefore, the cut flow requirement cannot be increased by rotating downward; and since the cut cannot rotate upward from its original position, a maximum cut must lie entirely between two adjacent deadhead arcs, i.e., entirely within a single frame.

Not every frame, however, has the same cut flow requirement. To find the frame whose cut has the maximum flow requirement, first note that every frame has the same number of inbound arcs that enter it, since both inbound and deadhead arcs are spaced at intervals of \( h_A \) (ignoring the frames near the source and the sink which contain fewer inbound arcs).

The number of inbound arcs that enter each frame, \( n_A \), which is the maximum number of inbound arcs that can be traversed by a cut within a single frame, is easily computed to be

\[
    n_A = \left\lceil \frac{t_A + t_D}{h_A} \right\rceil
\]

which is the number of buses that would be required if all buses deadheaded.

Some service arcs, both inbound and outbound, may traverse only one of the deadhead arcs spanning a given frame. A cut within that frame could perhaps avoid one of those inbound arcs and thereby traverse one or more of those outbound arcs that it could not have traversed if it traversed every inbound arc that entered the frame. However, it can be shown (by the same argument that proved a maximum cut lies within a single frame) that avoiding one inbound arc can make it possible to traverse at most one additional outbound arc, so that nothing is to be gained by avoiding the inbound arcs. Therefore, we may state that the maximum cut will lie entirely within a frame spanned by adjacent deadhead arcs, cutting every
inbound arc that enters that frame. Let us denote such a cut a "single frame cut".

Since each single frame cut traverses the same number of inbound arcs, we should seek to find the single frame cut that traverses the greatest number of outbound arcs. Suppose the upper deadhead arc spanning a frame departs from the right side at time \( \tau_0 \); the lower deadhead arc departs at time \( \tau_0 + h_A \). Then the maximum cut within this frame will not traverse any outbound arc whose departure is later than \( \tau_0 + h_A \). The lower deadhead arc arrives at the left side at time \( \tau_0 + h_A + t_D \). The latest inbound run that departs earlier than \( \tau_0 + h_A + t_D \), from the definition of \( h_A \), departs \( h_A \) minutes after the inbound run that arrives at \( \tau_0 \); hence it departs (from the left side) at time \( \tau_0 - t_A + h_A \) or \( \tau_0 + t_E \), where \( t_E \) is the "effective deadhead time" given by

\[
(t_E = h_A - t_A) \tag{4.4}
\]

\( t_E \) is the deadhead time plus the layover necessary for a deadheading run to get back into phase with the inbound departures; it satisfies the inequality \( t_D < t_E < t_D + h_A \). Since the single frame cut will traverse this inbound arc departing at \( \tau_0 + t_E \), that cut will arrive at the left side of the diagram after \( \tau_0 + t_E \). Thus any outbound arc that the single frame cut traverses must arrive at the left side after time \( \tau_0 + t_E \), so that it must depart from the right side after \( \tau_0 + t_E - t_B \). Therefore, every outbound arc traversed by the single frame cut must depart after \( \tau_0 + t_E - t_B \) and before \( \tau_0 + h_A \), which is the interval of \( t_B - t_E + h_A \) minutes preceding \( \tau_0 + h_A \). To simplify our notation let us define the "effective deadhead premium" \( \pi \) to be
(Page 83 was inadvertently skipped in the pagination of this thesis.)
\[ \pi = t_B - t_E = t_A + t_B - h_A n_A \] 

so that this interval is now \((\pi + h_A)\) minutes long. \((\pi)\) is the deadhead premium, \(t_B - t_D\), minus the layover necessary for a deadheading run to get back into phase with inbound departures; it satisfies the inequality \((t_B - t_D - h_A < \pi \leq t_B - t_D)\).

The number of outbound arcs that can depart within this interval, which is the number of outbound arcs a single frame cut may intersect, is clearly no greater than

\[ U_B = \left\lfloor \frac{\pi + h_A}{h_B} \right\rfloor \] 

(Therefore, if \(U_B = 1\), the number of outbound arcs in the maximum cut, \(n_B\), must be 1.) However, by carefully phasing the inbound and outbound departures we may be able to avoid this upper bound. To find the maximum cut in any given schedule we must find the time \(\tau\) at which an inbound arc arrives so that the number of outbound departures in the \((\pi + h_A)\) minutes before \(\tau\) is maximized. Clearly, this time \(\tau\) will be the arrival time of an inbound arc that most closely follows (but is not coincident with) the departure of an outbound arc. If we wanted to maximize the flow requirement of the maximum cut, we would schedule an outbound arc to depart just moments before an inbound arc arrived. But since in making the schedule we want to minimize the flow requirement of the maximum cut, we would not do this (nor would any scheduler!). Instead we would schedule one outbound departure so that it exactly coincides with an inbound arrival (as we have done in Figure 4.1). Then the \(k'\)th outbound departure after this departure will occur \(kh_B\) minutes after it, preceding
the succeeding inbound arrival by \( h_A(1 - \text{mod}[kh_B/h_A]) \) minutes, where \( \text{mod}[x] \) is the fractional component of \( x \). Therefore, if the \( k \)'th departure is the latest outbound arc traversed by a single frame cut, any earlier outbound arc traversed by that cut must depart later than
\[
(\pi + h_A) - h_A(1 - \text{mod}[kh_B/h_A]) = (\pi + h_A \text{mod}[kh_B/h_A]) \text{ minutes before the } k \text{'th departure. If we include the } k \text{'th departure, the number of outbound arcs that depart in this span is } \langle (\pi + h_A \text{mod}[kh_B/h_A]) / h_B \rangle. \text{ To find the frame with the maximum cut, we should maximize this expression over the possible values of } k. \text{ Using the notation } r = h_B / h_A, \text{ this is equivalent to maximizing } \text{mod } [kr]. \text{ If we express } r \text{ as a ratio of integers whose smallest denominator is } y, \text{ then this maximum is}
\[
g(r) = \max_{k=1,2...} (\text{mod } [kr]) = \frac{y-1}{y} \tag{4.7}
\]

Therefore, the number of outbound arcs traversed by the maximum cut in the optimally phased schedule is
\[
n_B = \langle \frac{\pi + g(r)h_A}{h_B} \rangle \tag{4.8}
\]
and the total flow requirement of the maximum cut, which is the minimum number of buses needed for the route, is
\[
N(h_A, h_B) = n_A + n_B = \langle \frac{t_A + t_D}{h_A} \rangle + \langle \frac{\pi + g(r)h_A}{h_B} \rangle \tag{4.9}
\]

Applying these formulas to our example, we have
\[
n_A = \langle \frac{25 + 12}{6} \rangle = 7
\]
\[
t_E = 7 \cdot 6 - 25 = 17
\]
\[
\pi = 24 - 17 = 7
\]
\[ g(r) = \frac{\frac{2-1}{3}}{3} = \frac{2}{3} \quad \text{since} \quad r = \frac{\frac{h_2}{h_A}}{\frac{8}{3}} \]

\[ n_B = \left\langle \frac{7 + \frac{2}{3}}{16} \right\rangle = 1 \]

\[ N(6,16) = 7 + 1 = 8 \]

showing that the maximum cut has a capacity of 8.

### 4.1.4 Some Assumptions Critically Examined

Before extending the results of the previous section, we do well to critically examine two of the assumptions that are found in deterministic models developed in the previous section, as well as in all of the models reviewed in Section 4.1.2. The first assumption is that run times are assumed to be deterministic, and passenger arrivals and bus loads are not explicitly considered to vary. The second assumption is that run time, even if deterministic, has not been considered to be a function of the service headway.

First we consider the issue of ignoring the randomness in bus operations. Is our model invalid because it assumes deterministic run times when they will in fact vary? In fact, most service planning models make deterministic assumptions like this. Is it because randomness is not recognized? Rather, the randomness is implicitly recognized, and certain assumptions are made implicitly about the nature of this randomness.

For example, a service planner trying to minimize the number of vehicles required on a route will choose headways so that the peak load, computed using a deterministic model, will just equal the maximum allowed.
Certainly he does not expect that every bus will have a load exactly equaling the standard; he knows that due to randomness in bus running time and passenger behavior actual loads will vary. Therefore, if he knows that when loads get to be above 85 passengers, service will rapidly deteriorate, he will not set the peak load standard at 85, but rather at 70 or another value below 85. Then this standard of 70 is not really saying "buses must not carry over 70 people," but rather "under the level of randomness we expect, if the average load stays below 70, actual loads will be tolerable, e.g. not exceed 85 too often." The implicit assumption that is made in using the same standards on different routes is that the degree of randomness on those routes will be about the same. For instance, a route with very little randomness might have loads predominantly within 5 passengers of the mean, while a route with a lot of randomness might have loads that are frequently more than 15 passengers above or below the mean. If the same peak load factor of 70 is used on these two routes, loads on the first will rarely exceed 75 and operations will be smooth, while loads on the second will frequently exceed 85, resulting in poor service. This example shows, then, that using a deterministic model can be adequate for a single route, provided that the standards are set accounting for the expected degree of randomness, and that the same deterministic standards may be applied to other routes as long as they exhibit approximately the same degree of randomness.

Therefore, the fact that randomness exists per se does not invalidate deterministic models of deadheading; the vital question is how one expects the degree of randomness in bus loading to vary with the different possible
deadheading configurations. A single peak load factor is applied to the
routes being designed, which can be interpreted as saying "if the random-
ness on the routes being designed is about the same as the system average,
loading will be tolerable if the average load is designed to be below the
nominal bus capacity."

In our model we required that service departures be evenly spaced;
this requirement resulted in a potential increase in the number of buses
needed. One might ask, why require departures to be evenly spaced on the
schedule, when we know that in practice they will not be evenly spaced?
The reply is that the peak load standard (say of 70) is set expecting
the degree of randomness that normally occurs on routes for which
scheduled departures are evenly spaced; and we have reason to believe
that if scheduled departures are irregular, there will be a much higher
degree of randomness in practice (i.e., a greater variance in peak load)
so that the standard of 70 would no longer be high enough to prevent
intolerable loads. Therefore, to achieve tolerable loads these models
require that a sufficient number of buses be allocated so that a schedule
with evenly spaced departures is feasible, and they should use a peak load
standard that has been designed expecting the randomness in passenger
loads usually observed on routes with evenly scheduled departures.

The validity of a model and its results rests, of course, upon the
validity of its assumptions. If one believes that the actual load
variance of irregularly scheduled service is no worse than that of
regularly scheduled service, he should use a different model. However,
I believe that the assumptions concerning randomness are valid enough to
make the model useful in practical implementation.
The second issue is that run time has not been modeled as a function of service headway. This assumption hardly affects heavy direction run times, since heavy direction headways remain essentially fixed throughout the extensions of this model. Light direction service headway will vary, however, depending on how many vehicles deadhead, and we have noted earlier (Section 2.1.5) that when service headways are greater, dwell times and consequently run times will be greater. Therefore the results should be checked for consistency. For example, if one expects the light direction headway to be 6 minutes, and estimates the light direction run time accordingly, but then the optimal solution has a light direction headway of 8 minutes, the results will be inconsistent. However, run time is rather insensitive to headway, and its relationship to headway is predictable. Therefore one can either incorporate this dependence into the formulas given (by making \( \pi \) a linear function of \( r \)), or iterate with run time recomputed according to the headway of the previous solution. Both of these approaches have been used in the case studies with no difficulty, and the results reported in the case studies are consistent with their run times.

4.2 Minimum Number Of Vehicles Under Maximum Headway Constraints

4.2.1 Non-Integer Headway Ratios Allowed

We have now a formula for the minimum number of buses required when the headways \( h_A \) and \( h_B \) are given. We suggested earlier that setting \( h_A \) and \( h_B \) to their maximum values, \( h_M^A \) and \( h_M^B \), might be a good idea; however, they are not always the optimal headways. Therefore we need to find the optimal values of \( h_A \) and \( h_B \). Because of the integer functions \( \lfloor \cdot \rfloor \) in Equation (4.9), we cannot differentiate \( N(h_A, h_B) \) with
respect to those parameters. However, we would be very surprised if a
decrease in the peak direction headway \( h_A \) would lead to a reduction in
the number of buses needed. Since in practical operations headways are
often required to be integer, we could simply evaluate (4.9) for
\( h_A = h_A^m, h_A = h_A^m - 1, \) and \( h_A = h_A^m - 2. \) Decreasing \( h_A \) from \( h_A^m \) to \( h_A^m - 1 \)
will usually increase \( n_A \) by 1 or more (cf. eq. (4.3)), and it is almost
impossible for \( n_B \) (eq. (4.8)) to be reduced by a greater amount. If we
want to make a reduction smaller than 1 minute that keeps \( n_A \) from
increasing, \( h_A \) may not be reduced beyond \((t_A + t_D)/n_A\). In practical
applications, this will mean a maximum reduction of about 5 seconds.
Therefore in our further discussion we will consider \( h_A \) to be fixed at \( h_A^m. \)

There is potential, however, for reducing \( h_B \) from an
initial value of \( h_B^m \), given that \( h_A \) is fixed, owing to the fact that \( g(r) \)
is not monotonic in \( r \). In our example, with \( h_A = h_A^m = 6, \) if we set \( h_B \)
to 11, we will get \( N(6,11) = 9; \) but if \( h_B \) is reduced to 9, then \( N(6,9) = 8. \)

It is clear that the optimal \( r, r^*, \) lies between \( \text{Int}[r^m] \) and \( r^m, \)
where \( r^m = h_B^m/h_A \) and \( \text{Int}[x] \) is the integer portion of \( x \). This must be,
since for any value of \( r \) for which \( r^m < r < \text{Int}[r^m], N(\langle r \rangle) \leq N(r). \) (Here
\( N(r) \) is equivalent to \( N(h_A, h_B) \), with \( r = h_A/h_B). \) Below an algorithm is
presented for finding the optimal \( r \).

\[(A \, 4.1)\]

1. Let \( r^* = r^m, N^* = N(r^m). \) Let \( r_u = r^m, r_\ell = \text{Int}[r^*]. \)
   If \( N(r_\ell) \leq N^*, \) let \( N^* = N(r_\ell) \) and \( r^* = r_\ell. \)
   Let \( z = 1. \)
2. Let \( z = z + 1 \). If no integer \( x \) exists such that \( r\lambda z < x < r\mu z \), go to 2; otherwise let \( \hat{r} = x/z \).

3. Compute \( N' = n_A + \left( \frac{\pi + h_A g(\hat{r})}{r\mu h_A} \right) \).

   If \( N' \geq N^* \), STOP, the optimal \( r \) is \( r^* \); otherwise let \( r\lambda = \hat{r} \).

4. Compute \( \hat{N} = N(\hat{r}) \). If \( \hat{N} < N^* \), let \( N^* = \hat{N} \) and \( r^* = \hat{r} \). Go to 2.

Throughout the algorithm, \( r\lambda \) and \( r\mu \) are the lower and upper bounds for \( r^* \), the optimal \( r \). The key to the algorithm is Step 3, which computes \( N' \), a lower bound for \( N(r) \) with \( r < r\mu \). If \( N' \) is no lower than the best \( N \) already found, the algorithm stops. Experience has shown that the algorithm will terminate most of the time on the first pass through Step 3; no more than a few iterations are expected for the worst cases.

4.2.2 Integer Headway Ratios Required

When the ratio of the light direction headway to the heavy direction headway, \( r \), is an integer, the dispatch strategy in the light direction is very simple: send the first of every \( r \) buses back in service, and deadhead the remainder. When \( r \) is not an integer an irregular pattern takes place. For example, if \( r \) is 3.4, the first, fourth, seventh, eleventh, and fourteenth of every seventeen buses would be dispatched in service, and the remainder deadheaded. Scheduled layovers will also be much more varied when \( r \) is not an integer. For these reasons (and perhaps others) many operators might prefer to have a more simple, integer-ratio strategy.
To minimize the number of vehicles required, the optimal integer headway ratio is

\[ r^* = \text{Int}[r^m] = \text{Int} \left( \frac{h_B}{h_A} \right) \]  \hspace{1cm} (4.10)

where \( \text{Int} [x] \) is the integer component of \( x \).

Where there are numerous optimal headway ratios (assume the heavy direction headway is fixed and we are choosing the light headway \( h_B \)) and at least one is an integer, we can meet an integer ratio constraint at no cost. The question arises, how often will there be an optimal solution that is an integer?

Whether there will be an integer optimal solution depends, of course, on the parameters. A method for determining the probability of an integer optimal solution existing for certain ranges of parameters was developed as is described in detail in Appendix B. It was found that the probability of there being an integer solution that is optimal is above 90% for a large range of parameter values. This probability rises with the maximum headway ratio \( r^m \) and declines with the effective deadhead premium \( \pi \). Results for specific values of the parameters \( \pi, h_A, \) and \( r^m \) are shown in Table 4.1.

4.3 Service Considerations In Partial Deadheading Design

So far we have had as our objective minimizing the number of vehicles required, with no explicit concern for passenger service level except in maximum headway constraints and the requirement of evenly spaced service departures. What will be the optimal design under a more general objective function that accounts for passenger costs as well as vehicle costs?
Table 4.1

ESTIMATED PROBABILITIES THAT INTEGER SOLUTIONS WILL BE SUB-OPTIMAL

<table>
<thead>
<tr>
<th>Approximate Parameter Values</th>
<th>Probability that Integer Solution is Sub-Optimal</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \pi_0 ) ( ^{2/} )</td>
<td>( h_0 ) ( ^{2/} )</td>
</tr>
<tr>
<td>10.0</td>
<td>2.0</td>
</tr>
<tr>
<td>10.0</td>
<td>2.0</td>
</tr>
<tr>
<td>10.0</td>
<td>2.0</td>
</tr>
<tr>
<td>5.0</td>
<td>2.0</td>
</tr>
<tr>
<td>20.0</td>
<td>2.0</td>
</tr>
<tr>
<td>20.0</td>
<td>1.5</td>
</tr>
</tbody>
</table>

1/ For a detailed explanation of this table, see Appendix B.

2/ \( \pi \) is equally likely to take on any value in the neighborhood of \( \pi_0 \); \( h_A \) is equally likely to take on any value in the neighborhood of \( h_0 \); \( r^m \) is equally likely to take on any value in the range \( k \leq r^m < k+1 \).

3/ Probability that the integer solution will require at least 1 vehicle more than the real solution

4/ Probability that the integer solution will require at least 2 vehicles more than the real solution
Two different ways of accounting for the service level are considered. In the first, the operator's primary objective is still to minimize the number of buses needed, but he also has a secondary objective to maximize passenger service level. That is, given numerous solutions that minimize his vehicular requirements, he will choose the one that provides the best level of service. In the second way, the service objectives are included directly in the objective function.

4.3.1 Service Level as a Secondary Objective

Since wait time, in-vehicle time, and crowding levels all decline as headway declines, the problem we now address is how to minimize headways without increasing the number of vehicles needed above the minimum number needed. We will first look at reducing the headway in the light direction, where there is greater flexibility, and then at reducing the heavy direction headway.

4.3.1.1 Improving Service in the Light Direction

When the heavy direction headway \( h_A \) is given and a number of values of \( h_B \) (headway in the light direction) may minimize the primary objective (number of buses required), we want to choose from among them the headway \( h_B^* \) that provides the best level of service to the passengers traveling in the light direction. This secondary objective is to minimize \( h_B \), or equivalently the ratio \( r = h_B / h_A \) (\( h_A \) is assumed fixed), subject to \( N(r) = N^* \), the minimum fleet size required.

If we are employing the constraint that \( r \) be an integer, we should first set \( r = \text{Int}[r^m] \) and compute \( n_A \) and \( n_B \). Then the minimum integer
that will minimize \( n_B \), from eq. (4.8), is \( r^* = \frac{\pi}{n_B h A} \), or equivalently

\[
h_B^* = \left( \frac{\pi}{n_B} \right)
\]

(4.11)

If \( r \) is not required to be an integer, we need an algorithm similar to (A4.1) to solve this problem. Presented here is an algorithm that finds the minimum headway in the light direction at which the minimum number of buses is required.

(A 4.2)

1. Find the minimum number of buses required, \( N^* \), and a value of \( r, r^* \), for which \( N(r^*) = N^* \) using algorithm (A 4.1). Set \( r_u = r^*, z = 1, r_{\bar{z}} = \text{Int}[r_u] \).

2. If \( N(r_{\bar{z}}) > N^* \), go to 3; otherwise let \( r_u = r_{\bar{z}}, r_{\bar{z}} = r_{\bar{z}} - 1 \), and go to 2.

3. Let \( z = z + 1 \). If no integer \( x \) exists such that \( r_{\bar{z}} z < x < r_u z \), go to 3; otherwise let \( \hat{r} = x/z \).

4. Compute \( N' = n_A + \left( \frac{\pi + g(\hat{r})}{r_u h A} \right) \). If \( N' > N^* \), STOP; the optimal \( r^* = r_u \).

5. If \( N' = N^* \), let \( r_u = \hat{r} \); otherwise let \( r_{\bar{z}} = \hat{r} \).

Go to 3.

Again, \( r_{\bar{z}} \) and \( r_u \) are lower and upper bounds for the optimal \( r \) and \( N' \) is a lower bound for \( N(r) \), \( r_{\bar{z}} < r < r_u \). From the way \( r_{\bar{z}} \) is computed in Steps 1 and 2, when we get to Step 3 we know that \( N(r) > N^* \) for all \( r \leq r \); therefore if \( N' > N^* \), then the optimal \( r \) cannot lie below \( r_u \) and therefore is \( r_u \).
4.3.1.2 Improving Service in the Heavy Direction

So far we have assumed that the heavy direction headway $h_A$ will remain at its maximum value, $h_A^m$. Because of the integer nature of the problem it will usually be possible to reduce $h_A$ slightly without increasing the required fleet size, $N$, thus improving service (wait times, crowding) in the heavy direction. Recalling eq. (4.9), when $h_A$ changes both $n_A$ and $n_B$ are changed. Since the minimum $n_A$ is given by $n_A^* = \left\langle \frac{(t_A + t_D)}{h_A^m} \right\rangle$, $h_A$ can be diminished to

$$h_A^{\text{min}} = \frac{t_A + t_D}{n_A^*}$$

without increasing $n_A$; any further decrease in $h_A$ will increase $n_A$.

In practical situations, however, $(h_A - h_A^{\text{min}})$ will be quite small, rarely exceeding 10 seconds. The effect of $h_A$ on $n_B$ is not as straightforward. As $h_A$ goes down by a few seconds, the effective deadhead time $t_E$ goes down ($t_E = t_D$ when $h_A = h_A^{\text{min}}$), making the deadhead premium $\pi$ go up by $n_A$ times the change in $h_A$; and recalling that $n_B = \left\langle \frac{(\pi + g(r)h_A)}{h_B} \right\rangle$, this tends to raise $n_B$. As $h_A$ decreases, the term $g(r)h_A$ will also tend to decrease, but since $h_A$ will only decrease by a few seconds and $g(r) < 1$, this effect is negligible. The change in $h_A$ will also change $r = h_B/h_A$ and therefore $g(r)$, but since experience shows that most cases have integer solutions and $g(r) = 0$ in such cases, there is little room for decreasing $g(r)$ from a previous optimal solution. Therefore the change in $\pi$ will dominate the other changes, which implies that $n_B$ will not decline but perhaps rise as $h_A$ decreases. Although no rigorous proof has been given, for practical purposes it appears that $h_A$ may not be diminished beyond $h_A^{\text{min}}$ without increasing $N$. 
Since the practical margin for decreasing $h_A$ is so small it does not warrant an optimal search algorithm. It is recommended to use $h_A^m$; the planner may try $h_A^{min}$ or values between them as he likes.

### 4.3.2 Service Level as a Part of the Primary Objective

When the operator is not committed to minimizing the number of vehicles needed on a route, he will want to find the headways that minimize some general cost that includes passenger as well as vehicle cost.

The four basic design variables for partial deadheading are $n_A^r$, $n_B^r$, $h_A$ and $h_B$. Since $h_A$ can vary by only a few seconds given $n_A^r$, it can be assumed a function of $n_A^r$. Then, given $n_B^r$ and $h_A$, Section 4.3.1 shows how to find the optimal $h_B^r$. Therefore there are essentially two degrees of freedom: $n_A^r$ and $n_B^r$. The minimum values of $n_A^r$ and $n_B^r$ are found using the relationships already presented. One can then search over the feasible combinations of $n_A^r$ and $n_B^r$. The objective function will be "almost" convex in both $n_A^r$ and $n_B^r$ (the "almost" is due to small aberrations resulting from the integer nature of the problem), so that one can find the optimal solution without searching over too many values. Beginning at the minimum values of $n_A^r$ and $n_B^r$, a unit increase in $n_B^r$ is much more likely to reduce the total cost than a unit increase in $n_A^r$, and so it is suggested that the search begin by holding $n_A^r$ at its minimum and increase $n_B^r$ until the costs increase; then $n_A^r$ can be increased by 1 and the feasible values of $n_B^r$ scanned similarly; and so on until costs increase as $n_A^r$ is increased (evaluated at the optimal $n_B^r$).
4.4 A Prototypical Manual Design Procedure

The following manual procedure for implementing partial deadheading was developed for the Cairo (Egypt) Transit Authority, an operator whose primary objective is to minimize the number of vehicles needed on each route because of Cairo's severe bus shortage, and whose secondary objective is to minimize the headway. The nature of operations in Cairo are such that service headway ratios should be integer.

(A 4.3)

1. **Select Candidate Routes.** A route is a candidate for partial deadheading only if it meets these criteria:

   a. Its maximum headway, $h^m_A$, is at most half the policy headway.

   b. Its light direction peak load is at most half the heavy direction peak load; i.e., $h^m_B > 2h^m_A$.

2. **Screen Candidate Routes.** Preliminary computations will show whether it is possible to save buses through partial deadheading.

   a. Compute the number of buses needed for conventional local service, $N_n$:

   \[ N_n = \left( \frac{t_A + t_B^m}{h_A^m} \right) \]

   b. From eq. (4.1) compute

   \[ N_\ell = \left( \frac{t_A^m}{h_A} + \frac{t_B^m}{h_B} + t_D \left( \frac{1}{h_A^m} - \frac{1}{h_B^m} \right) \right) \]  (4.1)

   c. Compute $S_\ell = N_n - N_\ell$. If $S_\ell = 0$, no buses can be saved through partial deadheading; delete this route from the candidate list.
d. From eq. (4.3) compute \( n_A = \left< \frac{t_A + t_B}{h_A} \right> \) \( (4.3) \)

If \( n_A \geq N_n - 1 \), no buses can be saved; delete this route from the candidate list.

3. Choose the Light Direction Service Headway. We are assuming \( h_A = h_A^m \), and finding the minimum feasible \( h_B \) for which \( r = h_B / h_A \) is an integer and the number of buses required is a minimum.

a. Compute \( t_E = h_A n_A - t_A \)
b. Compute \( \pi = t_B - t_E \)
c. Compute \( r_m = \text{Int} \left[ \frac{h_B^m}{h_A} \right] \)
d. Compute \( n_B = \left< \frac{\pi}{r_m h_A} \right> \) and \( N^* = n_A + n_B \). If \( N^* = N_n \), no buses can be saved; delete this route from the candidate list.
e. Compute \( r^* = \left< \frac{\pi}{n_B h_A} \right> \)
f. Compute \( h_B^* = h_A^m r^* \)

The optimal design is then:
- heavy direction headway \( h_A = h_A^m \)
- light direction headway \( h_B = h_B^* \)
- required number of buses \( N = N^* \)

A worksheet illustrating how simple it is to apply this procedure is used in the following case study.
4.5 Partial Deadheading Case Study

The Watertown corridor of Boston in the morning peak has good potential for improvement using partial deadheading. Its demand inbound is high, its demand outbound is low, and there is a high speed path available for deadheading using the Massachusetts Turnpike.

Again two cases were analyzed: serving the demands that are currently served by Route 57 alone with a route between Watertown and Kenmore Square; and serving the demands of the entire corridor, from Watertown to downtown.

4.5.1 Serving the Entire Watertown Corridor

The number of vehicles needed to serve the entire Watertown corridor could be substantially reduced through partial deadheading. Table 4.2 shows the results of optimizing deadheading service under the austerity objective. Whereas local service requires 46 standard buses or 34 articulated buses, partial deadheading would require only 38 standard buses or 29 articulated buses, for a savings of 17% with standard buses or 15% with articulated buses. At the assumed vehicle costs, the optimal solution is to operate 29 articulated buses with an inbound headway of 2.5 minutes and an outbound service headway of 10.0 minutes. In this solution the headway ratio is integer. If only standard buses are allowed, the optimal headways are 1.73 minutes inbound and 9.5 minutes outbound, for a headway ratio of 5.5. If an integer headway ratio is required, standard buses can operate at service headways of 1.73 and 10.36 (ratio = 6) at the same operator cost. Exhibit 4.1 shows how the manual procedure of Section 4.4 is executed on a worksheet; it finds the optimal
<table>
<thead>
<tr>
<th></th>
<th>Using</th>
<th>Using</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Articulated</td>
<td>Standard</td>
</tr>
<tr>
<td></td>
<td>Buses</td>
<td>Buses</td>
</tr>
<tr>
<td>Inbound Headway</td>
<td>2.5 min</td>
<td>1.7 min</td>
</tr>
<tr>
<td>Outbound Service Headway</td>
<td>10.0 min</td>
<td>9.5 min</td>
</tr>
<tr>
<td>Headway Ratio</td>
<td>4</td>
<td>5.5</td>
</tr>
<tr>
<td>Number of Vehicles</td>
<td>29</td>
<td>38</td>
</tr>
<tr>
<td>Avg. Wait Time</td>
<td>2.2 min</td>
<td>1.8 min</td>
</tr>
<tr>
<td>Avg. In-Vehicle Time</td>
<td>22.9 min</td>
<td>21.3 min</td>
</tr>
<tr>
<td>Operator Cost</td>
<td>$1,161</td>
<td>$1,218</td>
</tr>
<tr>
<td>Total Cost</td>
<td>$4,298</td>
<td>$4,105</td>
</tr>
<tr>
<td>(pax time valued @ $3/hr)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Operator Cost Savings</td>
<td>$ 200</td>
<td>$ 143 (§256)</td>
</tr>
<tr>
<td>Over Conventional Local Service</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1/ Savings if conventional local service is also restricted to standard buses.
Exhibit 4.1

Partial Deadheading Manual Worksheet

Route: Watertown Corridor

Vehicle Capacity: 45 x 1.2 = 54

Time Period: a.m. peak

I. DATA

\[ h_A = h_A^m = \frac{773}{120} \text{ min} \]
\[ t_A(h_A) = 43.6 \text{ min} \]
\[ t_B(h_A) = 35.3 \text{ min} \]
\[ h_B^m = 13.8 \text{ min} \]
\[ t_B(h_B^m) = 48.6 \text{ min} \]

Guess \( h_B = 12.7 \text{ min} \)

\[ t_B(h_B) = 46.9 \text{ min} \]

\[ t_D = 16.7 \text{ min} \]

II. SCREENING

\[ N_n = \left( \frac{t_A + t_B(h_A)}{h_A} \right) = 46 \text{ buses} \]

\[ q_D = \frac{1}{h_A} - \frac{1}{h_B^m} = 0.505 \]

\[ N_\ell = \left( \frac{t_A + t_B(h_B^m)}{h_A} + t_D q_D \right) = 37 \text{ buses} \]

\[ S_\ell = N_n - N_\ell = 9 \text{ buses} \]

(If \( S_\ell = 0 \), STOP)

\[ n_A = \left( \frac{t_A + t_D}{h_A} \right) = 35 \text{ buses} \]

\[ S_n = N_n - n_A = 27 \text{ buses} \]

(If \( S_n \leq 1 \), STOP)

III. MINIMUM VEHICLES, \( h_B \)

\[ \pi = t_A + t_B(h_B) - h_A n_A = 23.95 \]

\[ r^m = \text{Int} \left( \frac{h_B^m}{h_A} \right) = 7 \]

\[ n_B = \left( \frac{\pi}{r^m n_A} \right) = 3 \]

\[ N_* = n_A + n_B = 38 \text{ buses} \]

(If \( N_* \geq N_n \), STOP)

\[ r^* = \left( \frac{N_A}{n_B h_A} \right) = 6 \]

\[ h_B^* = r^* h_A = 10.38 \text{ min} \]

IV. Check for consistency:

When \( h_B = 10.38; \ t_B(h_B) = 46.1 \); \( \pi = 28.15 \); \( n_B = 3 \); \( r^* = 6 \);

\[ h_B^* = 10.38. \]
deadheading strategy using standard buses with the headway ratio required
to be integer.

The optimal designs under the prosperity objective (passenger wait
and in-vehicle time valued at $3/hour) for the two vehicle types are
shown in Table 4.3. The optimal bus type is the standard bus. The least
cost solution uses one bus more than the minimum needed, but still 7 buses
less than conventional local service. The level of service for outbound
passengers is worse than conventional local service would offer, however,
so that the reduction in operator plus passenger cost from using partial
deadheading is equivalent to the cost of 3.5 standard buses, or about
8 percent of operator cost. This solution has a service headway ratio of
3.75. If an integer headway ratio were required, the best design would
still call for 39 buses and a headway ratio of 4; total cost savings over
conventional local service would then be $8 smaller.

The optimal design for articulated buses under the prosperity objec-
tive also employs one bus more than the minimum required, but still uses
4 fewer vehicles than conventional local service. Although the longer
headway outbound (6.25 minutes) makes total passenger costs a little
greater than they are under conventional local service, total operator plus
passenger costs are still $70 less than the prosperity solution found for
conventional local service using only articulated buses.

4.5.2 Serving Route 57 Demands

Partial deadheading can also be a useful routing strategy instead
of the conventional local service now operating on Route 57. Table 4.4
shows the best partial deadheading designs under the austerity objective.
<table>
<thead>
<tr>
<th></th>
<th>Using Articulated Buses</th>
<th>Using Standard Buses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inbound Headway</td>
<td>2.5 min</td>
<td>1.7 min</td>
</tr>
<tr>
<td>Outbound Service Headway</td>
<td>6.25 min</td>
<td>6.5 min</td>
</tr>
<tr>
<td>Headway Ratio</td>
<td>2.5</td>
<td>3.75</td>
</tr>
<tr>
<td>Number of Vehicles</td>
<td>30</td>
<td>39</td>
</tr>
<tr>
<td>Avg. Wait Time</td>
<td>1.9 min</td>
<td>1.5 min</td>
</tr>
<tr>
<td>Avg. In-Vehicle Time</td>
<td>22.7 min</td>
<td>21.2 min</td>
</tr>
<tr>
<td>Operator Cost</td>
<td>$1,201</td>
<td>$1,250</td>
</tr>
<tr>
<td>Total Cost</td>
<td>$4,269</td>
<td>$4,087</td>
</tr>
<tr>
<td>(pax time valued at $3/hr)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Cost Savings</td>
<td>- (§70) ¹/</td>
<td>$112</td>
</tr>
<tr>
<td>Over Conventional Local Service</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

¹/ Savings if conventional local service is restricted to articulated buses
<table>
<thead>
<tr>
<th></th>
<th>Using Articulated Buses</th>
<th>Using Standard Buses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inbound Headway</td>
<td>7.4 min</td>
<td>5.6 min</td>
</tr>
<tr>
<td>Outbound Service Headway</td>
<td>14.8 min</td>
<td>13.2 min</td>
</tr>
<tr>
<td>Headway Ratio</td>
<td>2</td>
<td>2.33</td>
</tr>
<tr>
<td>Number of Vehicles</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>Avg. Wait Time</td>
<td>5.9 min</td>
<td>4.8 min</td>
</tr>
<tr>
<td>Avg. In-Vehicle Time</td>
<td>14.4 min</td>
<td>13.6 min</td>
</tr>
<tr>
<td>Operator Cost</td>
<td>$ 400</td>
<td>$ 385</td>
</tr>
<tr>
<td>Total Cost</td>
<td>$1,621</td>
<td>$1,493</td>
</tr>
<tr>
<td>(pax time valued @ $3/hr)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Operator Cost Savings</td>
<td>$ 40</td>
<td>$ 55 ($64)</td>
</tr>
<tr>
<td>Over Conventional Local Service</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1/ Savings if conventional local service is restricted to standard buses
The optimal bus type, interestingly, is the standard bus, since partial deadheading can save two standard buses but only one articulated bus relative to conventional local service. This is a case, then, where even when the objective is to minimize operator cost only, it is more efficient to use a smaller vehicle in an innovative routing configuration than to use a larger vehicle. The service headways are 5.6 and 13.2 minutes, for a headway ratio of 2.33. Using articulated buses, the best solution is to operate 10 vehicles at service headways of 7.4 and 14.8 minutes. The articulated bus solution would cost an operator $15 more per hour of operation than the standard bus solution (though it is still $40 per hour cheaper than conventional local service).

The standard bus solution just described is the only solution for which requiring an integer headway ratio would increase the number of vehicles needed. With an integer headway ratio (2), 13 standard buses would be needed. Under the requirement of integer headway ratios, then, the articulated bus solution is optimal.

Under the prosperity objective, conventional local service is a superior strategy to partial deadheading with either vehicle type. The optimal design for conventional local service under the prosperity objective is displayed in Table 3.2.

4.6 Conclusions

Partial deadheading can be an effective routing strategy when there is a strong demand imbalance favoring the peak direction and a high speed path for deadheading vehicles. On a very heavy route, such as a route serving the entire Watertown corridor, it could be 10 to 20 percent less
costly to operate than conventional local service. The number of vehicles required on the existing MBTA Route 57 could be reduced by 14 percent while still offering adequate service. Furthermore, good partial deadheading configurations can be found manually using a simple worksheet such as Exhibit 4.1, so that an operator, using available data (run times and peak loads) could quickly scan all of his routes and design a good partial deadheading configuration on all the routes which can be improved. In fact, finding the optimal solutions (with no requirement of integer headway ratios) manually is not impractical, as all the solutions in the case study just presented were found manually.
Chapter 5
Local Zonal Service

A characteristic of the load profiles of most radial routes brought out in Section 3.1 is that loads are generally low at the uptown end of the route, and increase gradually to a peak near the downtown terminal. With such a load profile, conventional local service (or partially deadheaded service, for that matter) has to provide for more capacity than is needed at the uptown end of the route in order to provide adequate capacity at the peak load point. This skewness in load profile suggests that instead of providing the same capacity level all along the route, the route should begin at its uptown end with a small capacity level, and then add increments of capacity as the flow level increases as the route moves toward the downtown. An analogy can be drawn to designing the capacity of a water main or a sewer main. Unlike a water main, however, a bus route does not provide a constant stream of capacity overtime, but, as a scheduled carrier service, periodically provides discrete amounts of capacity, making the design problem more complex.

Practically speaking, offering capacity levels that increase towards the peak load point implies that along the arterial there should be a number of terminals in addition to the uptown terminal, and that the corridor should be served by a system of overlapping routes, each of which begins at one of these terminals and runs to the downtown terminal and back. It may also be the case that the peak load point lies considerably upstream of the downtown terminal; in such a case it may be worthwhile to
stop some buses at a terminal that lies between the peak load point and the downtown terminal.

This strategy of operating some buses over the entire length of the corridor and operating others along the sections of the corridor with the greatest loads has informally been called "short-turning", because some buses turn back short of the uptown terminal; in this research it is called "zonal service". The segments between the intermediate terminals can be considered zones of demand, and the nature of this strategy gives each zone a particular relationship with the route beginning at the uptown end of that zone, as the remainder of this chapter will show.

5.1 Different Zonal Service Strategies

If the corridor being studied has only express service (i.e. all passengers have a common destination), a zonal configuration is not hard to imagine. Buses that begin at the outermost terminal, say terminal 1, pick up passengers between terminal 1 and terminal 2, then get on an expressway or another fast path and carry their passengers to the destination. Similarly buses beginning at terminal 2 pick up passengers between terminals 2 and 3, and so forth. This zonal configuration has exclusive boarding and alighting zones; that is, buses operating from any given terminal will stop in only one zone for inbound passengers to board and outbound passengers to alight. This configuration is common in commuter rail and elevator banking systems.

Bus service, however, cannot usually confine its market to passengers destined for a single downtown terminal. Passenger destinations in a radial corridor, while heavily concentrated around the downtown during the a.m. peak period, are nevertheless dispersed all along the corridor, and
an adequate level of service must be offered to those whose destinations are not downtown ("local" as opposed to "downtown" passengers). Making them transfer between express routes at every intermediate terminal between their origin and their destination (as suggested by Turnquist [25]) is as unacceptable as making all local passengers ride in to the downtown on one express route and then transfer to another express route that will take them out to their destination (as if they were riding elevators). Dealing with a mix of local and downtown passengers is different from dealing with downtown passengers only, and must be studied as a separate problem. Furthermore, corridor service must be designed with passengers traveling in the light direction in mind, a consideration absent from express route design models.

While local passengers cannot be treated as downtown passengers, downtown passengers can be treated as local passengers, for they can reach their destinations on local routes without making transfers. This chapter confines the routing system to local routes only, forcing downtown passengers to use local service; Chapter 6 will allow both local and express routes to operate together within a corridor.

Three possible strategies then exist to serve a market of local passengers. The first strategy has exclusive boarding zones inbound and exclusive alighting zones outbound. This means that an inbound passenger in the a.m. peak has no choice as to which route he will take; only the route that begins at the upstream end of his zone is authorized to board passengers in that zone. However, a.m. inbound passengers may alight at whatever stop they desire along the corridor. Similarly, an outbound passenger may use only the route whose alighting zone includes his des-
tination, although he may board that route anywhere. Hence the buses, after passing through their respective zones, are not free to use an expressway when operating outside their service zones but must remain on the main arterial of the corridor to let inbound local passengers alight where they need to and to let outbound local passengers board where they need to. This strategy is called "local service with exclusive boarding/alighting zones". It is studied in Section 5.2.

A second zonal service strategy is to impose no boarding or alighting restrictions. The only boarding restriction is that when a bus becomes too full, no one can board it. As soon as someone alights, however, someone else can board. This strategy is called "local service with overlapping zones," since the boarding/alighting zone of each route in the system extends over its entire length. It has the advantage that it keeps buses as full as possible, requiring therefore the least possible number of vehicles, but it also has numerous operational disadvantages that have made local zonal service with exclusive boarding/alighting zones preferred in North American systems. Overlapping zonal service is studied in Section 5.3.

The advantage of the overlapping zonal strategy over the strategy of exclusive boarding/alighting zones is most obvious in the a.m. peak in the inbound direction. With exclusive zones, as full as the buses may be at the downstream end of their route's zone, they will become more and more empty as passengers alight along the rest of the route, since no one is allowed to board there. The same phenomenon occurs in the p.m. peak in the outbound direction. A way of retaining this advantage while still having exclusive boarding/alighting zones is to have a system of routes
with exclusive service zones that are non-contiguous. By non-contiguous we mean a route's service zones may be composed of a number of disjoint pieces, designed in such a way as to balance alighting and boarding all along the route's length. This innovative strategy, "local service with non-contiguous boarding/alighting zones," is discussed in Section 5.4.

The zonal service problems studied in the literature deal with express demands only, with the exception of Bernstein [26]. A few papers have discussed the advantages of zonal service in commuter rail [27,28] but their approaches are either qualitative or too idealized to be applied. Turnquist [25,15] has developed a method for finding the optimal design of a zonal route system for express bus service, for an express bus route, including both optimal zones and optimal service headways. The theory presented in this chapter may be considered an extension of Turnquist's work, incorporating local passengers inbound and outbound in the design along with downtown-bound passengers. (Bernstein [26] has also extended Turnquist's work in an effort to give better treatment to local passengers, but his results are of little use since he ignores light direction passengers, allows only one realistic configuration for local passengers in the heavy direction, and make unrealistic assumptions about the operating characteristics of an overlapping zonal service.)

5.2 Exclusive Boarding/Alighting Zones

Under the strategy of exclusive boarding/alighting zones, inbound passengers may only board their zone's route (i.e. the route beginning at the uptown end of their zone), though they may alight at any stop downstream. Outbound passengers may board their bus at any stop, but must board only the route serving the zone that includes their
destination stop, for only that route is authorized to discharge passengers in that zone. This type of routing configuration exists in numerous transit systems, and passengers are generally able to understand it and use the correct routes.

5.2.1 Optimal Design for a Single Direction

To more easily convey an understanding of the zonal service design problem, let us first ignore the light direction (e.g. by assuming that inbound and outbound service zones and service levels will be symmetric) and find the configuration that serves the heavy direction (inbound in the assumed a.m. peak) with the least number of vehicles of fixed capacity. Let us also assume for now that all passengers are destined for a common downtown terminal. Our problem then, given the arrival rates of inbound passengers at each stop and the travel time characteristics of the roads in the corridor that buses may use, is to choose from among a set of possible intermediate terminals which terminal should be active, and what frequency should be operated out of each active terminal in order to minimize the number of vehicles needed, subject to the standard operating and service constraints discussed in Section 2.3.

This problem is a modification of the problem solved by Turnquist [25]. Rather than derive a solution from his more complex model, a fresh approach is taken.

Let us denote the possible uptown terminals as nodes 1, 2, ..., n where node 1 is the uptown end of the corridor and the intermediate nodes are numbered in the order of their distance from node 1. Let node n+1 be the downtown terminal. We may divide the arterial into n segments, where segment i extends from node i to the stop just before node i+1.
Let us begin by formulating this problem as a capacity scheduling problem. We shall require that the capacity offered on any segment must exceed the total passenger flow on that segment. Since all passengers are destined for downtown, the maximum flow on any segment is the sum of the demands of the stops belonging to that segment and every previous segment. We can express this set of constraints as follows:

\[
\begin{align*}
q_1 & - s_1 = d_1 \\
q_1 + q_2 & - s_2 = d_1 + d_2 \\
q_1 + q_2 + q_3 & - s_3 = d_1 + d_2 + d_3 \\
& \vdots \\
q_1 + q_2 + q_3 + \cdots + q_n & - s_n = \sum_{i=1}^{n} d_i
\end{align*}
\]

where \( q_i \) = frequency of the route beginning at node \( i \) (route \( i \))

\( s_i \) = slack variable associated with constraint \( i \)

\( d_i \) = demand, in busloads per hour, originating on segment \( i \)

We may transform this constraint set into that of a transshipment problem by subtracting each row from the subsequent row:
\[
\begin{bmatrix}
q_1 & q_2 & q_3 & \cdots & q_n & s_1 & s_2 & s_3 & \cdots & s_{n-1} & s_n & d_1 & d_2 & d_3 & \cdots & d_n \\
1 & & & & & 1 & & & & & 1 & & & & \\
1 & -1 & & & & & 1 & & & & & -1 & & & & \\
& 1 & -1 & \cdots & & & & 1 & -1 & & & & & \cdots & & & & \\
& & \ddots & \cdots & & & & & & \ddots & & & & & & & & & \ddots & & & & & & & & & & \ddots & & & & & & & & & & & \ddots & & & & & & & & & & & & 1 \\
-1 & -1 & -1 & \cdots & -1 & & & & & & & & & & & & & & \\
\end{bmatrix}
\]

(5.1)

\[q_i \geq 0 \text{ for all } i=1, \ldots, n\] (5.2)

\[s_i \geq 0 \text{ for all } i=1, \ldots, n\] (5.3)

Equation (5.1) is a general constraint set for a capacity scheduling problem. The production scheduling problem has been formulated in this way by Wagner and Whitin [29] and by Zangwill [30]. In production scheduling, \(d_i\) is demand for the good being produced in period \(i\), \(q_i\) is the amount of the good produced at the beginning of period \(i\), and \(s_i\) is the inventory carried from period \(i\) to period \(i+1\). The constraints of eq. (5.1) represent the conservation formula for inventory, \(q_i + s_{i-1} - s_i = d_i\). Constraints (5.2) ensure that all production is positive, and constraints (5.3) imply that inventories may never be negative, i.e. there may be no backlogging of orders. Zangwill has shown that these constraints describe an underlying network in which each time period is represented by a node and the quantities \(q_i\), \(s_i\), and \(d_i\) are flows into and out of these nodes: \(q_i\), the production at the start of period \(i\), is a flow into node \(i\); \(d_i\), the demand during period \(i\), is a flow out of node \(i\); and \(s_i\), the inventory carried from
period \( i \) to period \( i+1 \), is a flow from node \( i \) to node \( i+1 \). Figure 5.1 illustrates the network described by equation (5.1). Zangwill then used the property of extreme flows to show that if both production and inventory holding costs in each time period are concave, an optimal production schedule (one that minimizes the sum of production and holding costs) will have the property that each node will have at most one positive input; that is, either \( q_i = 0 \) or \( s_{i-1} = 0 \) for all \( i = 1, \ldots, n \). This property leads to an efficient dynamic programming algorithm for finding the cost production schedule that is first derived in [29] and later extended in [30]. The analogy between production planning and bus capacity scheduling can be easily drawn. Just as demands in period \( i \) could be met by production in period \( i \) or by production in earlier periods (but not later periods), so passengers boarding at node \( i \) can be served by the route beginning at node \( i \) or by routes beginning upstream of \( i \) (but not by routes beginning downstream of \( i \)). Analogously to inventory passed from period \( i \) to period \( i+1 \), \( s_i \) is the unused capacity available (empty seats) as buses leave segment \( i \) that then becomes available to passengers originating along segment \( i+1 \).

However, in the bus capacity scheduling problem for a zonal routing configuration with exclusive boarding and alighting zones, costs are not concave due to the minimum frequency constraint. Since passengers with a given origin node can use only one route, each route must have a minimum frequency that will ensure acceptable wait times. Under this minimum frequency constraint, a route may have zero frequency, or the minimum frequency, or any frequency above the minimum. To incorporate this constraint into the route cost function, one must model the cost of operating at any frequency between zero and the minimum as equal to the
Figure 5.1

Network Representation of Zone Scheduling Problem
cost of operating at the minimum frequency. The resulting cost function is illustrated in Figure 5.2. This function is neither concave nor convex due to the minimum frequency constraint; even beyond the minimum frequency it will be a step function if routes are required to have an integer number of buses. This lack of concavity prevents us from extending the result from the production scheduling model that \( q_i s_{i-1} = 0 \) for all \( i = 1, \ldots, n \).

However, by the definition of a zonal route system with exclusive boarding and alighting zones, \( q_i s_{i-1} \) must equal zero, since \( q_i s_{i-1} > 0 \) would imply that passengers boarding at \( i \) have a choice between using route \( i \) and an upstream route. Therefore, the property that \( q_i s_{i-1} = 0 \) holds for the bus routing problem in a zonal route system with exclusive boarding/alighting zones, and hence an efficient dynamic programming algorithm can be used to find the minimum cost production schedule.

Under the zonal routing strategy with exclusive boarding/alighting zones, route \((j,k)\) may be defined as the route whose boarding zone is segments \( j, \ldots, k \) (\( k \geq j \)); its uptown terminal is node \( j \) and its downtown terminal is node \( n+1 \). The presence of a route \((j,k)\) in a routing plan indicates that no route begins at any node \( t, t = j+1, \ldots, k \); hence \( q_{j+1} = q_{j+2} = \ldots = q_k = 0 \). Therefore, \( s_{k-1} = d_k \) (provided that \( j \leq k-1 \)), \( s_{k-2} = d_k + d_{k-1} \) (provided that \( j \leq k-2 \)), and finally \( s_j = \sum_{i=j+1}^{k} d_i \). In other words, if \( j < k \), then route \((j,k)\) must provide capacity at node \( j \) that will be unused in segment \( j \) but will be filled at later segments \((j+1, \ldots, k)\); that amount of unused capacity is \( s_j \). Then since route \((j,k)\) must provide capacity for segment \( j \) demands as well, the total demand carried by route \((j,k)\) is \( \sum_{i=j}^{k} d_i \). Furthermore, the operating pattern of route \((j,k)\) is known: it begins at node \( j \), operates local along segments \( j \) through \( k \), and then travels
Figure 5.2
Route Cost Under a Minimum Frequency Constraint

(At zero frequency, cost is zero; at frequencies between 0 and the minimum frequency the cost is a constant; at frequencies above the minimum frequency, the cost increases as a step function each time an additional vehicle is needed.)
without stopping to pick up passengers from node k+1 to node n+1. Given a route's operating pattern and its demand pattern, we know the route's round trip time and its peak load, from which we know the minimum number of vehicles that route needs. Let \( B_{jk} \) be the minimum number of buses needed by route \((j,k)\). Also define \( Z_{jk} \) as the minimum number of buses needed to serve demands \( d_1, \ldots, d_k \) with routes beginning at nodes 1, \ldots, \( j \) (\( j \leq k \)). Then a dynamic program for finding the routing plan that requires the least number of buses is:

\[
(A5.1)
\]

1. Set \( Z_{1k} = B_{1k} \) for all \( k = 1, \ldots, n \). Let \( j = 1 \).
2. Let \( j = j+1 \). If \( j = n+1 \), STOP. Otherwise, set

\[
Z_{jk} = \min \{ Z_{j-1,k}, (Z_{j-1,k} + B_{jk}) \} \quad \text{for all} \ k = j, \ldots, n
\]

\[
(5.4)
\]

Go to 2.

At the first step in \((A5.1)\) we require that \( q_2 = q_3 = \ldots = q_k = 0 \) and that all demands be served by route 1. At each subsequent stage \( j \) (\( j = 2, \ldots, n \)), we allow \( q_j \) to take on a nonzero value, i.e. we let a route beginning at \( j \) (route \((j,k)\)) be active, which would imply that \( s_{j-1} = 0 \), i.e. routes beginning upstream of \( j \) carry demands \( d_1, d_2, \ldots, d_{j-1} \) only. Therefore, the choice embodied in equation \((5.4)\) is whether route \((j,k)\) is to be inactive, in which case we already know the minimum number of buses needed to serve segments 1, \ldots, \( k \) (which is \( Z_{j-1,k} \)), or active, in which case route \((j,k)\) requires \( B_{jk} \) buses and the upstream segments 1, \ldots, \( j-1 \), which may be served by routes beginning at nodes 1, \ldots, \( j-1 \), require \( Z_{j-1,j-1} \) buses. The problem is solved when we find \( Z_{nn} \). The
entire procedure requires only $O(n^2)$ computations. The main computational burden is in computing $B_{jk}$ for the $n(n+1)/2$ possible combinations of $j$ and $k$; once these are found, algorithm (A5.1) can be executed on a triangular tableau that requires only $n(n+1)/2$ entries.

The remaining models developed in this chapter and in Chapter 6 are extensions of this simple model. They are all based on the binary choice embodied in equation (5.4): shall a route be active or inactive? If the route is to be inactive, we already know the best configuration possible without this route; if the route is to be active, we can compute its cost from the single route model and add it to the already known cost of serving the demands that originate upstream of the new route.

Algorithm (A5.1) may be considered a simplification of the algorithm developed by Turnquist [25], since it has no constraints on the number of zones there may be or on the total number of vehicles to be used, and it can solve the express route zoning problem with a simpler objective that Turnquist's. Yet this algorithm may be used, without modification, to solve the zoning problem for a local route, in which all passengers are not assumed to be destined for the downtown terminal, and with a general objective function that may include any route specific costs and with (as explained in Section 2.1.1). With a general mix of local and downtown bound passengers, route $(j,k)$'s peak load will no longer be $\sum_{i=j}^{k} d_i$ since some passengers may alight before others board, but the peak load may be easily computed from the route O-D matrix. To use a general function, we must simply redefine $B_{jk}$ as the minimum contribution of route $(j,k)$ to the objective function, which is the minimum cost of route $(j,k)$. $B_{jk}$ may be computed using the single route model of Section 3.2, which finds the
optimal headway for a route under a general objective function which may include operator cost, wait time, in-vehicle time, and other route costs. In fact, Turnquist's algorithm, with some redefinitions, could also be used for local route zoning.

Algorithm (A5.1) designs routes considering travel demands in one direction only, and therefore is not realistic enough for designing a zoned system of local routes. Since a local route system must serve outbound passengers as well as inbound, the design should take their demands and their costs into account as well as the inbound passengers' in order to find the system that optimally serves both directions. The design of zonal service for passengers in both directions is the subject of the next section.

5.2.2 Optimal Design for Both Directions

An implicit assumption in the last section was that round trip run time on a route is known, given the headway and the service zone in the inbound direction. This would be true if, for example, we required symmetric service in the light direction. However, usually the phenomenon of load profiles increasing toward one end of the route (the impetus for zonal service) is accompanied by a great imbalance in loads by direction, which means that symmetric service in the light direction will be overcapacitated. This suggests that some deadheading might be warranted. For example, deadheading the buses from all routes except the longest and letting the longest route serve the entire light direction may be a good strategy. There are many possible configurations for serving the light direction demand; we should seek then to find the zonal configuration, both inbound and outbound, that minimizes the
operator plus passenger cost while providing an adequate service level in both directions.

Another possibility that should be considered is partial deadheading on some of the zonal routes. Since in practical implementation operators will usually want service headways in the light direction to be an integer multiple of the heavy direction headway, there will only be a few feasible headway ratios on any given route; and since in zone scheduling heavy direction headways will tend to be longer than if there were no zonal service, there will usually be few possibilities for deadheading in the light direction, and so they may be enumerated without excessive computation.

The algorithm for solving this problem is a straightforward generalization of the algorithm for single direction scheduling, and does not therefore warrant a formal development. Two further assumptions should be made, however, concerning service in the light direction (outbound). As mentioned above, we allow a route's inbound boarding zone to differ from its outbound alighting zone. For example, a route may serve only segment 1 inbound, but outbound it may serve segments 3, 2, and 1; another route may serve segments 2 and 3 inbound, and may deadhead (i.e. serve no segments) outbound. If a route's outbound service area extends as far out as node 4, while its inbound service area begins at node 3, then that route will including a deadheading segment outbound from node 4 to
node 3. This kind of deadheading is permitted. However, we will assume that
a route may not deadhead inbound (in the a.m. peak, that is; in the p.m.
peak routes may not deadhead outbound). That is, its outbound alighting
zone may not extend beyond the start of its inbound boarding zone.
First of all, one could hardly imagine an optimal solution that deadheads
vehicles in the heavy direction, and secondly one can hardly imagine an
operator that would implement such a configuration even if it were opti-
mal. A further assumption is that the ordering of the routes
according to the distance of their service zones from downtown must be
the same outbound as inbound. This would prohibit, for example, the
route beginning at node 2 from having segment 5 in its outbound alighting
zone while the route beginning at node 3 serves segments 3 and 4 outbound.
Again, it is very unlikely that such an arrangement would be optimal or
acceptable to an operator. The first assumption could be relaxed without
increasing the model's complexity, but the second is necessary to maintain
its simplicity.

Let us introduce the following notation:

\[ \text{route } (j,k,g,h) = \text{ route whose boarding zone inbound covers } \]
\[ \text{segments } j \text{ through } k \text{ and whose alighting zone outbound covers segments } g \text{ through } h \]
\[ (j < k, \ j \leq h \leq g+1). \text{ Employ the convention that if } h > g, \text{ the route deadheads outbound.} \]

\[ B_{jkgh} = \text{ minimum cost of operating route } (j,k,g,h) \]

\[ Z_{jkg} = \text{ minimum cost of serving inbound passengers boarding along segments } 1, \ldots, k \text{ and out-} \]
\[ \text{bound passengers alighting along segments } 1, \ldots, g \text{ when the set of routes is restricted to those whose uptown terminal is a node } \]
\[ \text{belonging to the set } 1, \ldots, j. \]
Figure 5.3 is a diagram of route \((j,k,g,h)\), intended to aid the reader visualize the route as it is defined. Outbound segments are defined analogously to inbound segments: outbound segment \(g\) extends from the stop just uptown of node \(g+1\) to node \(g\). \(B_{jkgh}\) is the minimum cost of operating the single route \((j,k,g,h)\); Section 3.2 shows how to find this minimum cost in closed form for the general objective function used in the case studies, and gives the relationships necessary for finding the minimum cost of a zonal route for a wide range of objectives.

If an operator will permit partial deadheading within this route system, Chapter 4 shows how to find the optimal partial deadheading configuration and its cost for a given route.

\(Z_{jkg}\) is the stage variable of the dynamic program. At stage 1 only one route, which begins at node 1, is allowed, so that

\[
Z_{1kg} = B_{1kg} \quad \text{for all } k = 1, \ldots, n \\
g = k, \ldots, n
\]  

(5.5)

At subsequent stages \(j = 2, \ldots, n\) we allow a route to begin at node \(j\) in addition to routes beginning at nodes 1, \ldots, \(j-1\) and make the choice as to whether or not the route beginning at \(j\) should be active. The inbound service zone of a route beginning at \(j\) has to be segments \(j, \ldots, k\), and the inner end of the outbound service zone must be segment \(g\) (for a given \(j,k,g\), but we have flexibility in choosing what the outer boundary of that route's light direction service zone will be. Let \(h\) denote the outermost segment in the light direction service zone of the new route. (Recall the convention that if \(h > g\), the route beginning at \(j\) will deadhead.) Then we may compute the variable \(Z_{jkg}\) using the following recursion:
Figure 5.3

A Zonal Route With Exclusive Boarding/Alighting Zones

Node: 1 2 3 ... n n+1 n ... 3 2 1
Corridor: Inbound Outbound

Route \((j, k, g, h)\) has the following configuration:

Node: \(j j+1 \ldots k k+1 n+1\)

Legend:

\[---\] = service zone (boarding zone inbound, alighting zone outbound)

\[----\] = vehicle path outside service zone with passenger movements (inbound passengers may alight, outbound passengers may board)

\[.....\] = layover and deadheading
$$Z_{jkg} = \min_{j\leq h \leq g+1} \left\{ \min[Z_{j-1,k,g}, (Z_{j-1,j-1,h-1}+B_{jkg})] \right\}$$

for all \( k = j, \ldots, n \)

\( g = k, \ldots, n \) \hspace{1cm} (5.6)

Equations (5.5) and (5.6) constitute algorithm (A5.2).

The binary choice within the brackets reflects the decision as to whether route \((j,h,g,h)\) should be active or not. The enumeration over feasible values of \( h \) is necessary to allow the flexibility of outbound service zones that are chosen independently of the inbound service zone. The minimum cost of serving the corridor is \( Z_{nnn} \). This algorithm requires \( O(n^4) \) computations. Since realistic corridors will rarely have more than 6 potential uptown terminals, this is obviously a very small amount of computation for a large computer (on the order of a few seconds).

Before concluding this section on exclusive boarding/alighting zone service, it is worth mentioning a practical advantage of this type of service. Inequities and inefficiencies in transit service are often blamed on flat fares, the common fare policy of American cities (excluding suburban zones, where little riding takes place, anyway). Distance based fares, which are both more equitable and more efficient in societal resource allocation, are hard to manage with conventional local service unless the bus has a two-man crew. With zoned service, however, one can operate each zonal route as a flat fare route, yet by charging successively higher fares on the longer routes, one achieves a distance based fare. Eisele [22] also finds zonal service with zone-based fares an excellent compromise between the efficiency and equity of distance based fares and the simplicity of flat fares in commuter rail systems that are accustomed to charging a different fare for each station.
5.3 Overlapping Zones

When a route has an exclusive boarding zone, its peak load point must lie within that zone (usually at the last stop in the zone), and therefore will be considerably upstream of the corridor's peak load point, unless the route is the innermost route in a zonal system. As buses continue in from the inner edge of their service zones to the downtown, passengers alight and the buses become more and more empty. Thus a zonal route can pass the corridor's peak load point (usually at the edge of downtown) with quite a number of empty places, even if that route was full at its own peak load point. Thus a zonal system can require a significantly greater frequency of buses entering the downtown than a conventional local route because of empty places on the outer route buses. This phenomenon adds to the cost of zonal service; indeed, it can in some situations make conventional local service more efficient than any zoned local service with exclusive boarding/alighting zones, even when the objective is to minimize the number of buses needed.

Allowing the zonal routes to operate with no boarding or alighting restrictions, acting as overlapping conventional local routes, eliminates the empty seats at the corridor's peak load point that are structured into the exclusive zonal configuration. Then since passengers can ride any one of a number of routes, there is no more need to impose a minimum frequency constraint save on the outermost route. One can tailor the capacity profile by adding one bus to the system here, another bus there to fit the load profile as closely as possible, just as a highway designer might tailor the number of lanes on an urban expressway to envelope the expected flows. This simple approach is taken by Bernstein [26].
This routing strategy, as good as it appears, presents some serious practical difficulties. Suppose some buses start at node 1, and other buses start at node 3. Clearly the node 3 buses will have more room on them for picking up segment 3 passengers than will the node 1 buses, which were probably filled during segments 1 and 2, and have room to pick up passengers only when some riders alight. Thus, our optimal design may call for 10% of the segment 3 passengers to board node 1 buses, and the remainder to board node 3 buses. But suppose the node 1 buses come every 5 minutes, while the node 3 buses come only every 15 minutes. (In Bernstein's example the ratio of the shorter route headway to the longer route headway is even greater than 15:5.) Then we are hoping that 90% of the passengers will board 25% of the buses. Inducing passengers to wait for the buses they are supposed to board, in the absence of a complete restriction in boarding, is difficult. Suppose a node 1 bus stops, discharges one passenger, and then desires to pick up one passenger, while there are six passengers waiting to board. How does the driver say, "One only, please"? On a highway, if some lanes are more full than others, it is easy for the entering traffic to distribute itself according to the space in each lane. But if six passengers want to board a bus, and the first one that stops has room for only one, all six will still try to board that bus because waiting for a bus that isn't full is not as painless as sliding a car into the next lane. "Why, the next bus may be full, the next have room for only two people, and I may have to wait for the fourth bus," reasons a passenger. Thus, if the drivers do not exercise a good deal of control, the longer routes will become overcrowded, and the shorter routes will have empty places. If the drivers do exert a lot of control, they may find eggs, tomatoes, and curses hurled at them or their vehicles. Because it is hard to influence people
to board the buses we want them to board, overlapping routes will not operate as smoothly in practice as they do on paper.

There are ways to influence people to board the bus that has room for them that are more indirect than driver control of the doors. One way is to schedule the vehicles so that there will always be a short headway before the arrival of a long route bus and a long headway before the arrival of a short route bus. This strategy, however, requires that each route have the same frequency, and requires a high degree of schedule adherence, limiting its applicability. An example of this strategy is shown in Figure 5.4. Headways on the overlapped segment are alternately 5 minutes, then 25 minutes, so that $5/6$ of the passengers originating in the shorter segment will ride the shorter route, and only one sixth will try to board the first bus that passes that passengers arrive randomly, and that schedule adherence is perfect, average wait time in this configuration is $\left(\frac{1}{6} \cdot \frac{5}{2} + \frac{5}{6} \cdot \frac{25}{2}\right) = 10.8$ minutes. (This value is considerably higher than half the average headway, which in this case is 7.5 minutes, which is the assumption Bernstein uses.)

Other disincentives could be invented to influence people to not use the longer route. The longer route could charge a higher fare; but not so much higher that it keeps all short distance passengers from using it, or we'll be back to an exclusive zone system. Longer buses could make fewer stops in the inner segments, inconveniencing passengers who must alight in those segments but perhaps improving the service level overall.

All of the problems associated with overlapped zonal service in the inbound direction are magnified in the outbound direction. Going inbound,
Figure 5.4
Staggered Schedule for Routes with Overlapping Routes

![Diagram showing routes A → B → C and B → C]

<table>
<thead>
<tr>
<th>Schedule</th>
<th>Inbound</th>
<th></th>
<th>Outbound</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Depart A</td>
<td>Depart B</td>
<td>Depart C</td>
<td>Depart B</td>
</tr>
<tr>
<td>3:00</td>
<td>3:10</td>
<td>3:30</td>
<td>3:40</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3:35</td>
<td>3:55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3:30</td>
<td>3:40</td>
<td>4:00</td>
<td>4:10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4:05</td>
<td>4:25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4:00</td>
<td>4:10</td>
<td>4:30</td>
<td>4:40</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4:35</td>
<td>4:55</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*(times given in italics for Route 2 buses)*
a passenger would like to board the first bus he sees, but he can wait
for a later bus if the first bus is full. However, going outbound a
passenger going to segment 1 must board the longest route. Suppose his
route only runs every 15 minutes, and the first one that passes him is
full because short distance passengers filled it up? Nothing will keep
the second bus that comes 15 minutes later from being full, either; in
fact we can predict that, in the absence of boarding disincentives, some
long-distance passengers may never be able to board a route that will
carry them all the way to their destination stop. Kulash [31] and Han
[12] have described this phenomenon of "induced transfers"—passengers
who are forced to ride a short distance route to its outer terminal,
then transfer to a long distance route because the long distance route
had no room for them at their origin stop.

The severity of the problems associated with overlapping zonal
routes, and the strengths of their various remedies will vary greatly
in every situation, making it very difficult to predict what the level
of service in such a system is. For example, what is the expected wait
time in such a system? Chiqui and Robillard [32] show what the expected
wait time will be for a passenger who has a choice of riding a number
of independent local routes. However, they assume a passenger will be
able to board the first bus that comes, which will often not be possible
in an overlapped system. Furthermore, the longer the overlapped segment,
the less independent the different routes will be. Or how many induced
transfers will there actually be, or how severe will be the overcrowding
on some buses? Because answers to these questions depend so much on
the local characteristics of the zonal system, no design procedure has
been developed for overlapped zonal service.
If a system has considerable excess capacity, overlapped zonal service may be a good strategy. Another good design option may be to operate overlapped service in the a.m. peak but not in the p.m. peak. On the whole, however, one should not expect such a strategy to be successful without careful planning and monitoring.

5.4 Non-Contiguous Boarding/Alighting Zones

The advantage of exclusive boarding/alighting zones is that since passengers have only one route they can ride, that route can be carefully designed to offer the optimal level of service for the passengers assigned to it. The advantage of routes with overlapping routes is that they do not have to have empty places over the inner portion of the route where demand volumes are the highest. To try to have the advantages of both systems, one could design a set of routes with non-contiguous exclusive boarding/alighting zones, with the zones designed in such a way as to keep buses serving the outer zones full as they traverse the inner zones and some of their riders alight.

An example of non-contiguous exclusive zones is shown in Figure 5.5. Suppose a corridor has 21 stops. This corridor is to be served by two zonal routes. The first route begins at the first stop, the second at the eleventh stop. Suppose the flow carried by the first route just after the tenth stop is 300 passengers/hour. Suppose that the origin-destination pattern of the passengers boarding at the first 10 stops is such that 6 per hour alight at each stop between stops 11 and 20; all the rest alight at stop 21. At each stop between stops 11 and 20, 30 passengers per hour board, all destined for stop 21. If the first route's boarding zone included only stops 1-10, and the second route had
Figure 5.5
Local Service With Non-Contiguous Boarding/Alighting Zones

ROUTE 1

Route 1 Boarding Zone

STOPS

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21

ROUTE 2

Route 2 Boarding Zone
to serve stops 11-20, the first route would need a capacity of 300 passengers/hour, and the second route also a capacity of 300 passengers/hour. However, the peak load on the corridor (just before stop 21) is only 540 passengers/hour; the first route, due to passengers alighting at stops 11 to 20, would have 60 empty places/hour at the corridor peak load point. But suppose we add stops 19 and 20 to the boarding zone of the first route. No more capacity would be needed on the first route, and the peak load on the second route would diminish to 240, reducing the required frequency, and hence the cost, of the second route.

This service strategy is not without its problems, either. This configuration depends on a fine balancing of alighting passengers with boarding passengers, and the day-to-day and run-to-run variations in demand could sufficiently upset this balance as to cause significant delays and aggravation to passengers who can't board the bus that is supposed to serve them because it is full. Another problem is that the small distance between bus stops makes it possible for passengers to change the stop they use in response to the level of service offered. This border effect is small when a contiguous zone has 10 or 20 stops, but it can effect a major change in the demand pattern of an isolated subzone of one or two stops that is served by a different route than the surrounding stops. This effect is even stronger with outbound trips, when a passenger doesn't have to commit himself to the stop at which he will alight until he boards a bus.

The major objection to this strategy will probably be how difficult it is to make the passengers understand which route to use. This is probably not an insurmountable problem, however, as transit riders have learned
to successfully cope with many different subway lines that use the same station and many different bus routes, locals and expresses, leaving a terminal. The possibility is left open, then, for an enterprising transit operator to implement this kind of configuration. Again, no optimal design procedure is given because design will be so dependent on local considerations, and the crucial factor in implementing an innovative strategy like this one is its acceptability to the riding public not its optimality.

5.5 Local Zonal Service Case Study

Experiments were done to assess the value of local zonal service in the Watertown corridor. The routing strategy used called for contiguous exclusive boarding zones inbound and exclusive alighting zones outbound. Solutions were found under the policy of allowing partial deadheading and under the policy of forbidding partial deadheading on a single route, in order to show the possible advantage of combining these strategies. Experiments for both the entire corridor and the market served by the existing Route 57 confirmed the potential of the zonal service strategy to reduce vehicle and passenger costs.

Four potential upstream terminals (nodes) were identified. All four are currently used as route terminals by the MBTA, so that their suitability as bus route terminals (i.e. room for buses to turn around and park) was already proven. These four nodes were Watertown Square, the uptown end of the corridor; Newton Corner, at the entrance to the Massachusetts Turnpike; Oak Square; and Brighton Center. All four locations are shown in Figure 3.1.
5.5.1 Serving the Entire Watertown Corridor

Recall that the demands of the entire Watertown corridor are the demands of local Route 57 and the three express routes serving the corridor. The routes being designed must provide service from Watertown Square to downtown.

Under the austerity objective (minimize operator cost), it was found that a zonal system of 2 routes was the best choice. Routes are to begin at Watertown Square (node 1), and Newton Corner (node 2). The shorter route deadheads, and the longer route serves the entire outbound demand. If partial deadheading is allowed, the operator should employ it on the longer route, offering a service headway outbound three times as great as the inbound headway. These optimal routing configurations with and without partial deadheading are illustrated in Figure 5.6, and the details of these solutions are found in Table 5.1.

With partial deadheading, the corridor can be served with 27 articulated buses at an hourly operator cost of $1,081. This is a savings of $280, or about 21 percent of operator cost, compared to conventional local service, which requires 34 articulated buses. If partial deadheading is forbidden, the operator cost savings is $160, which is 12 percent of operator cost.

Zonal service with partial deadheading is an effective strategy for reducing vehicular requirements in the Watertown corridor. Conventional local service requires 34 articulated buses; partial deadheading alone reduces this figure to 29; and zonal service with partial deadheading requires only 27.
Figure 5.6
Optimal Local Service Zones In The Watertown Corridor
Under the Austerity Objective

(a) With Partial Deadheading

(b) Without Partial Deadheading

Legend:
- service zone (boarding zone inbound, alighting zone outbound)
- segment covered by route outside its service zone
Node 1 = Watertown Square
Node 2 = Newton Corner
Node 3 = Oak Square
Node 4 = Brighton Center
Table 5.1
Optimal Local Zonal Service for the Watertown Corridor
Under the Austerity Objective

<table>
<thead>
<tr>
<th></th>
<th>Segments In Inbound Boarding Zone</th>
<th>Segments In Outbound Alighting Zone</th>
<th>Inbound Service Headway</th>
<th>Outbound Service Headway</th>
<th># Buses</th>
</tr>
</thead>
<tbody>
<tr>
<td>With Partial Deadheading</td>
<td>1</td>
<td>4-1</td>
<td>4.8 min</td>
<td>14.4 min</td>
<td>14 Artic</td>
</tr>
<tr>
<td></td>
<td>2-4</td>
<td>-</td>
<td>4.5 min</td>
<td>-</td>
<td>13 Artic</td>
</tr>
<tr>
<td>Without Partial Deadheading</td>
<td>1</td>
<td>4-1</td>
<td>4.7 min</td>
<td>4.7 min</td>
<td>17 Artic</td>
</tr>
<tr>
<td></td>
<td>2-4</td>
<td>-</td>
<td>4.5 min</td>
<td>-</td>
<td>13 Artic</td>
</tr>
</tbody>
</table>

Service Level and Costs

<table>
<thead>
<tr>
<th></th>
<th>Average Wait Time</th>
<th>Average In-Vehicle Time</th>
<th>Operator Cost</th>
<th>Total Cost (pax time valued @ $3/hr)</th>
<th>Operator Cost Savings Over Conventional Local Service</th>
</tr>
</thead>
<tbody>
<tr>
<td>With Partial Deadheading</td>
<td>3.7 min</td>
<td>21.8 min</td>
<td>$1,081</td>
<td>$4,265</td>
<td>$280</td>
</tr>
<tr>
<td>Without Partial Deadheading</td>
<td>2.8 min</td>
<td>21.1 min</td>
<td>$1,201</td>
<td>$4,187</td>
<td>$160</td>
</tr>
</tbody>
</table>
Zonal service was also found to be a good routing strategy for the Watertown corridor under the prosperity objective. The optimal configuration is the same as it is under the austerity objective. Standard size buses are now preferred on both routes. When partial deadheading was allowed, the longer route used this strategy with a service headway ratio of 2. The results for the Watertown corridor under the prosperity objective are shown in Table 5.2.

As Table 5.2 shows, zonal service with partial deadheading offers a savings of $192 in operator and passenger cost over conventional local service, the equivalent of operating 6 standard buses. Compared to the optimal conventional local service solution, the zonal solution uses 7 fewer standard buses, increases average wait time from 1.0 to 2.2 minutes, and reduces average in-vehicle time from 20.8 to 19.9 minutes. If partial deadheading is forbidden, then the total cost of the best zonal service configuration is only $9 more.

Under the prosperity objective as well, then, zonal service is a further step in efficiency beyond partial deadheading. Partial deadheading reduced total costs relative to conventional local service by $112; zonal service with partial deadheading further reduces total costs by $80.

5.5.2 Serving Route 57 Demands

Serving the existing demand on Route 57 (Watertown Square to Kenmore Square) using zonal service with exclusive boarding/alighting zones was found to be a small improvement over conventional local service, but not as effective as partial deadheading, under the austerity objective. Under the prosperity objective, conventional local service, with its shorter wait
Table 5.2

Optimal Local Zonal Service for the Watertown Corridor
Under the Prosperity Objective

<table>
<thead>
<tr>
<th></th>
<th>Segments In Inbound Boarding Zone</th>
<th>Segments In Outbound Alighting Zone</th>
<th>Inbound Service Headway</th>
<th>Outbound Service Headway</th>
<th># Buses</th>
</tr>
</thead>
<tbody>
<tr>
<td>With Partial Deadheading</td>
<td>1</td>
<td>4-1</td>
<td>3.2 min</td>
<td>6.4 min</td>
<td>21 Std</td>
</tr>
<tr>
<td></td>
<td>2-4</td>
<td>-</td>
<td>3.1 min</td>
<td>-</td>
<td>18 Std</td>
</tr>
<tr>
<td>Without Partial Deadheading</td>
<td>1</td>
<td>4-1</td>
<td>3.3 min</td>
<td>3.3 min</td>
<td>23 Std</td>
</tr>
<tr>
<td></td>
<td>2-4</td>
<td>-</td>
<td>3.1 min</td>
<td>-</td>
<td>18 Std</td>
</tr>
</tbody>
</table>

Service Level and Costs

<table>
<thead>
<tr>
<th></th>
<th>Average Wait Time</th>
<th>Average In-Vehicle Time</th>
<th>Operator Cost</th>
<th>Total Cost (pax time valued @ $3/hr)</th>
<th>Total Cost Savings Over Conventional Local Service</th>
</tr>
</thead>
<tbody>
<tr>
<td>With Partial Deadheading</td>
<td>2.2 min</td>
<td>19.9 min</td>
<td>$1,250</td>
<td>$4,007</td>
<td>$192</td>
</tr>
<tr>
<td>Without Partial Deadheading</td>
<td>1.9 min</td>
<td>19.7 min</td>
<td>$1,314</td>
<td>$4,016</td>
<td>$183</td>
</tr>
</tbody>
</table>
times, was found to be better than any other zonal strategy, just as it was superior to any partial deadheading strategy.

Table 5.3 describes the optimal zone configuration for the Route 57 demand under the austerity objective. It has two routes. The longer route's boarding zone is segments 1 and 2; the shorter route's boarding zone is segments 3 and 4. The shorter route deadheads outbound to its Oak Square terminal, while the longer route serves the entire outbound demand. The optimal solution calls for 13 standard buses, one less than conventional local service if restricted to standard buses. If free to choose bus type, conventional local service would use 11 articulated buses, making the operator cost advantage of zonal service only $24 (about 75% of the cost of operating one standard bus) with its 13 standard buses.

Local zonal service is not as effective as partial deadheading in serving Route 57 demands, requiring one more than the 12 standard buses partial deadheading uses. This is because the zonal configuration, due to the problem of empty places on the longer routes, must have 13.4 buses per hour cross the peak load point while the partial deadheading configuration, like conventional local service, has just 10.7.

5.6 Conclusions

Zonal service in one of the three forms discussed in this chapter is a good alternative to conventional local service in a corridor whose load profile increases toward the downtown. The simplest form in which to implement zonal service, having exclusive boarding/alighting zones that are compact, was found promising for serving the corridor examined in the case study. While its advantages relative to conventional local service
### Table 5.3
Optimal Local Zonal Service for the Route 57 Demand
Under the Austerity Objective

<table>
<thead>
<tr>
<th>Segments In Inbound Boarding Zone</th>
<th>Segments In Outbound Alighting Zone</th>
<th>Inbound Service Headway</th>
<th>Outbound Service Headway</th>
<th># Buses</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-2</td>
<td>4-1</td>
<td>10.7 min</td>
<td>10.7 min</td>
<td>7 Std</td>
</tr>
<tr>
<td>3-4</td>
<td>---</td>
<td>7.7 min</td>
<td>---</td>
<td>6 Std</td>
</tr>
</tbody>
</table>

### Service Level and Costs

<table>
<thead>
<tr>
<th>Average Wait Time</th>
<th>Average In-Vehicle Time</th>
<th>Operator Cost</th>
<th>Total Cost (pax time valued @ $3/hr)</th>
<th>Operator Cost Savings Over Conventional Local Service</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.7 min</td>
<td>12.5 min</td>
<td>$417</td>
<td>$1,510</td>
<td>$24</td>
</tr>
</tbody>
</table>
were often found to be significant, its superiority over the partial deadheading strategy was not established for this case.

It was shown that the optimal routing configuration for the exclusive boarding/alighting zone strategy could be found with very little computation. Two other zonal service strategies, one with overlapping zones and one with non-contiguous zones were also discussed. These strategies hold potential for further reductions in vehicle requirements by making better use of vehicular capacity at the corridor peak load point. Their designs depend on many local factors, and are therefore better left in the hands of local planners with their experience and judgment than an optimizing model involving many questionable assumptions.
Chapter 6

Express/Local Zonal Service

Chapter 5 showed how the optimal local zonal routing configuration could be found for a single arterial corridor. It pointed out that local zonal service could meet the needs of both downtown-bound and local passengers. However, if the downtown demand is large, it may be more efficient to offer express service for the downtown passengers in some parts of the corridor, especially if a high speed road is available for the express leg of the route. The market could be segmented between downtown demand and local demand; zonal express service could then be designed for the downtown market, and zonal local service for the local market.

Joint design of express and local services is not as simple as splitting the market according to destination and then applying the methods described in Chapter 5. It may not be best to offer express service to all downtown passengers. Certainly there could be no express service in the innermost segment of the corridor, for it wouldn't be express at all. Perhaps in other zones as well it may be better to make the downtown passengers use local service. Therefore choosing the optimal market segmentation (deciding which segments will have express service) is not a trivial question.

At the uptown end of the route there may also be interaction between local and express routes. The outer segments of the corridor may have a demand intensity too low to support both express and local routes. In such
a case, all passengers from those outer zones might be served on a local route. It may be that downtown passengers will begin traveling on a local route if no express route is available, and will then transfer to an express route downstream. Or it may be that only express service is offered in the outer zones, so that local passengers would have to use the express route to get to the starting point of a local route and transfer there. These possible interdependencies make the joint design of express and local services a complex problem.

What is the motivation for offering express service? First of all, downtown passengers will appreciate it since it will save them time, especially if the express route uses an expressway or other fast path. Even if the express buses use the same path as the local route, however, the time savings from stopping less and taking advantage of progressive signal timing can still be significant. With express passengers enjoying a better level of service, they will be more willing to pay a higher fare, a definite advantage for financially strapped transit agencies. Express routes are also more effective at attracting new riders than are local routes. Express routes can thus reduce passenger time costs and raise revenue. Furthermore, it costs less to serve a given market with express buses than with local buses, since the express buses can turn around faster and therefore be more productive. Express routes can therefore be a boon to operators, since they can raise ridership and revenue while lowering cost.

Express routes have another advantage in zonal systems. As Section 5.3 brought out, a disadvantage of local zonal service with exclusive
boarding/alighting zones is that buses from the outer zones will have empty places as they cross the corridor peak load point because the places emptied by alighting passengers will not be filled. An express zonal route does not have this problem, since no passengers alight before the downtown. This phenomenon makes express zonal service useful even when there is no high speed road available.

6.1 Assumptions About The Express/Local Configuration

This section will clarify the assumptions that will be made about service in a corridor that has both express and local routes. First of all, we will assume the routes have contiguous, exclusive boarding/alighting zones. Express routes by their nature have exclusive zones; we are therefore excluding from consideration local routes with overlapping and non-contiguous zones. The exclusion of these latter strategies does not mean they are considered infeasible or worthless, but rather that their design is based on the peculiar characteristics of the site, and therefore cannot be found by a mathematical model alone. Suggestions on the design of an express/local system in which the local routes have overlapping zones will be offered here. First, one should segment the market by deciding (based on the planner's judgment) which segments should have express service, and then manually designing the local service to cover the rest of the market. Express service for the express market can be designed using the algorithms of Chapter 5. If one repeats this process for a few promising market segmentations, one can probably find a good solution that is close to optimal.
A second assumption is that the areas served by both local and express routes (the area served by one route type includes all the service zones of that route type) must be contiguous. Thus, for example, one could not offer express along segments 1 and 3 and not along segment 2. And since the innermost segment n can be served only by local service, this means that if any segment i has local service, all the segments between i and n must have local service.

A third assumption is that the outer segments of the corridor may be served by either local routes only, express routes only, or both. If served by local routes only, downtown passengers may transfer downstream to an express route. If served by express routes only, those express routes must offer local service (i.e., allow alightings inbound, boardings outbound) over the area that has no local service, and must stop at the outermost node offering local service. All remaining local passengers on the express routes have to transfer at this node to a local route.

A fourth assumption is that when downtown passengers transfer from a local route to an express route, they will transfer at the outermost node served by an express route.

It is assumed that the light direction will have local service only. Since many buses will probably be deadheading in the light direction, one can argue that it would cost very little for them to carry passengers along the same path they would use for deadheading, thus offering express service. This may be a good idea in implementation, but in design it seems more appropriate to ignore the small potential market for light direction express service and just assign all outbound passengers to local routes.
As with local zonal service, we will require that the ranking of routes by the distance of their service zone from node 1 be the same outbound as inbound. Also, we will again forbid deadheading in the heavy direction, thus preventing a route's outbound service zone from extending out farther than its inbound service zone.

6.2 Optimal Design of Express/Local Zonal Service

As stated earlier, the problem of joint design of express and local routes can be viewed as a problem of market segmentation. If one decides which zones are to be offered express service, then the express service can be designed for its market and the local service designed for its market using the single direction algorithm given in Section 5.2.1 or in Turnquist [25] and the dual direction algorithm given in Section 5.2.2. In taking such a course one must decide whether the light direction is to be served by inbound routes which are local or express (we expect that most planners will choose local); or one could assign some of the light direction demand to routes that are express inbound and some to routes that are local inbound. Certainly this kind of exogenous market segmentation would be the best course if one is designing the route system manually and will be satisfied with a "good" solution.

If one wants to find the optimal market segmentation, however, there are many alternatives to look at if one simply searched over the entire feasible set. Suppose one requires that all outbound demand be served by the routes that inbound are local. Then the market segmentation could be determined by the following three variables:
\[ p = \begin{cases} 
1 \text{ if local service is offered on segment}\, 1 \\
0 \text{ otherwise}
\end{cases} \\
q = \begin{cases} 
\text{the outermost segment with express service if } p=1 \\
\text{the outermost segment with local service if } p=0
\end{cases} \\
s = \text{the innermost segment with express service}
\]

Since \( p \) can take on 2 values, \( q \) can take on \( n \) values, and \( s \) can take on between 1 and \( n \) values depending on \( p \) and \( q \), there are on the order of \( n^2 \) possible market segmentations, without including light direction demands in the market segmentation. Then each possible configuration requires the execution of the dual direction algorithm that takes \( O(n^4) \) computations.

When \( n \) is small, \( O(n^6) \) computations is not too many for a computer to perform within a reasonable budget (for example, \( 6^5 = 46,656 \)). Nevertheless, it would be worthwhile to find a systematic way of searching over the feasible market segmentations, eliminating those that can be proven inferior, in order to make the search more efficient. By incorporating express routes into the dynamic programming framework of Chapter 5 we can do just that. Furthermore, we can free the light direction demand from any \textit{a priori} assignment, allowing it to be served by whatever routes can serve it most efficiently within the constraints we have already specified.

The complexity of the joint express/local configuration requires extensive modification of the dynamic programming algorithm. The stage variable will still be the node at which a new route is allowed to start; that is, at first we will allow routes to begin only at node 1, then at nodes 1 and 2, and so forth. However, there is no longer a binary choice
of having a route or not having a route. One now has four choices: whether to have an express route only, a local route only, both, or neither.

Furthermore, the state of the problem at each stage can no longer be expressed by a single array such as \( z_{jkg} \). At any stage \( i \) the corridor may be in any one of the following conditions with respect to past and future decisions:

1) Only local service is offered at segment \( i \) and at all segments upstream of \( i \)

2) Only express service is offered at segment \( i \) and at all segments upstream of \( i \)

3) Both local and express services are offered at segment \( i \)

4) Both local and express services have been offered upstream of segment \( i \), but only local service is offered at \( i \) and downstream

There must therefore be a vector of state variables that account for all of these possible configurations.

In a joint express/local system, routes may be quite complex. We may define a route by the demands it serves and what transfers they are expected to make. The different route types are described in Tables 6.1 and 6.2. Table 6.1 states exactly which demands are covered by each route type; Table 6.2 describes and illustrates the service zones of each route type for further clarification. Each route type has an associated cost which is the minimum cost of operating that route under the specified objective function. These route costs are found using the relationships given in Section 3.2 since they are each single bus routes; if routes are permitted to use partial deadheading, then Chapter 4 shows how to find their least cost. In defining the demands served by a given route, the following notation is used:
<table>
<thead>
<tr>
<th>Route Type</th>
<th>Route Cost</th>
<th>Local Demands Served</th>
<th>Downtown Demands Served</th>
<th>Outbound Demands Served</th>
<th>Transfers To or From</th>
<th>Restrictions</th>
</tr>
</thead>
<tbody>
<tr>
<td>local</td>
<td>$i_{gh}^{ij}$</td>
<td>$d_1, \ldots, d_j$</td>
<td>$e_1, \ldots, e_j$</td>
<td>$u_h, \ldots, u_g$</td>
<td>none</td>
<td>$j \geq i$</td>
</tr>
<tr>
<td>local</td>
<td>$i_{gh}^{ij}$</td>
<td>$d_1, \ldots, d_j$</td>
<td>---</td>
<td>$u_h, \ldots, u_g$</td>
<td>none</td>
<td>$j \geq i$</td>
</tr>
<tr>
<td>express</td>
<td>$i_{gh}^{ik}$</td>
<td>---</td>
<td>$e_1, \ldots, e_k$</td>
<td>$u_h, \ldots, u_g$</td>
<td>none</td>
<td>$k \geq i$</td>
</tr>
<tr>
<td>local</td>
<td>$i_{gh}^{jk}$</td>
<td>$d_1, \ldots, d_j$</td>
<td>$e_1, \ldots, e_j$</td>
<td>$u_h, \ldots, u_g$</td>
<td>fraction of $e_1, \ldots, e_k$</td>
<td>$j \geq k \geq i$</td>
</tr>
<tr>
<td>local</td>
<td>$i_{gh}^{jt}$</td>
<td>$d_1, \ldots, d_j$</td>
<td>$e_1, \ldots, e_j$</td>
<td>$u_h, \ldots, u_g$</td>
<td>fraction of $e_i, \ldots, e_k$</td>
<td>$t &gt; j \geq i$</td>
</tr>
<tr>
<td>local</td>
<td>$i_{gh}^{ij}$</td>
<td>$d_1, \ldots, d_j$</td>
<td>---</td>
<td>$u_h, \ldots, u_g$</td>
<td>$d_1, \ldots, d_{i-1}$</td>
<td>board at node $i$</td>
</tr>
<tr>
<td>express</td>
<td>$i_{gh}^{ik}$</td>
<td>$d_1, \ldots, d_j$</td>
<td>$e_1, \ldots, e_k$</td>
<td>$u_h, \ldots, u_g$</td>
<td>$d_1, \ldots, d_j$</td>
<td>leave at node $i+1$</td>
</tr>
<tr>
<td>express</td>
<td>$i_{gh}^{it}$</td>
<td>$d_1, \ldots, d_j$</td>
<td>$e_1, \ldots, e_j$</td>
<td>$u_h, \ldots, u_g$</td>
<td>$d_1, \ldots, d_i$</td>
<td>leave at node $i+1$</td>
</tr>
<tr>
<td>express</td>
<td>$i_{gh}^{it}$</td>
<td>---</td>
<td>$e_1, \ldots, e_k$</td>
<td>$u_h, \ldots, u_g$</td>
<td>fraction of $e_1, \ldots, e_{i-1}$</td>
<td>$k &gt; i &gt; 1$</td>
</tr>
<tr>
<td>local</td>
<td>$i_{gh}^{jk}$</td>
<td>$d_1, \ldots, d_j$</td>
<td>$e_k, \ldots, e_j$</td>
<td>$u_h, \ldots, u_g$</td>
<td>none</td>
<td>$j \geq k &gt; i$</td>
</tr>
<tr>
<td>local</td>
<td>$i_{gh}^{jk}$</td>
<td>$d_1, \ldots, d_j$</td>
<td>$e_k, \ldots, e_j$</td>
<td>$u_h, \ldots, u_g$</td>
<td>$d_1, \ldots, d_{i-1}$</td>
<td>board at node $i$</td>
</tr>
</tbody>
</table>

**KEY:**
- $i =$ uptown terminal
- $j =$ innermost segment in the inbound local (and sometimes downtown) boarding zone
- $k =$ innermost segment in the inbound downtown (and sometimes local) boarding zone
- $gh =$ boundaries of the outbound boarding zone
- $t =$ transfer point

See Table 6.2 for explanation of the upper case letters.

---

1/ For every route, $i \leq h \leq g+1$; if $h \geq g$, then no outbound demands are served, i.e., the route deadheads.

2/ Not all passengers from the express zones mentioned must transfer, but only a fraction, since some may desire to remain on the local route.

3/ Local demands alighting upstream of node $i$ are not included.
Table 6.2
Illustration of Service Zones Of Each Route Type

Legend:
- segment
- local demands $d_1, ..., d_n$
- downtown demands $e_1, ..., e_n$
- upstream demands $u_1, ..., u_n$
- inbound demands picked up and retained on this route
- demands picked up on another route and transferring to this route at its uptown terminal
- demands picked up on this route and transferring from it at node $t$
- outbound demands served by this route

<table>
<thead>
<tr>
<th>Route Cost</th>
<th>Description</th>
<th>Service Zones</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i_{BD}^j_{gh}$</td>
<td>Local route carrying 80th downtown and local passengers indiscriminately without transfers</td>
<td><img src="" alt="Diagram" /></td>
</tr>
<tr>
<td>$i_{LO}^j_{gh}$</td>
<td>Local route carrying local passengers only without transfers</td>
<td><img src="" alt="Diagram" /></td>
</tr>
<tr>
<td>$i_{EX}^k_{gh}$</td>
<td>Express route carrying downtown passengers only without transfers</td>
<td><img src="" alt="Diagram" /></td>
</tr>
<tr>
<td>$i_{LD}^k_{gh}$</td>
<td>Local route picking up downtown passengers over some of its boarding zone, as far as segment $k$ (since the longest local route begins at $k$). Downtown pax leave at node $k+1$ to transfer to an express route</td>
<td><img src="" alt="Diagram" /></td>
</tr>
<tr>
<td>Route Cost</td>
<td>Description</td>
<td>Service Zones</td>
</tr>
<tr>
<td>------------</td>
<td>-------------</td>
<td>---------------</td>
</tr>
<tr>
<td>$i_{LT}^{gh}$</td>
<td>Local route picking up both local and downtown passengers over its entire boarding zone, but downtown pax Transfer at node $c$ to the express route that begins at $t$</td>
<td>![Diagram 1]</td>
</tr>
<tr>
<td>$i_{LB}^{j}^{gh}$</td>
<td>Same as route type $L_0$, except that local pax who had to ride express buses to node $i$ and transfer there Board at node $i$</td>
<td>![Diagram 2]</td>
</tr>
<tr>
<td>$i_{EG}^{jk}^{gh}$</td>
<td>Express route picking up local pax over Some of its boarding zone, as far as segment $j$ (since the longest local route begins at $j+1$). Local pax leave at node $j+1$ to transfer to a local route</td>
<td>![Diagram 3]</td>
</tr>
<tr>
<td>$i_{ET}^{j}^{gh}$</td>
<td>Express route picking up both downtown and local pax throughout its boarding zone, but local pax Transfer at node $c$ to a local route</td>
<td>![Diagram 4]</td>
</tr>
<tr>
<td>$i_{EB}^{k}^{gh}$</td>
<td>Same as route type $E_0$, except that downtown pax who had to ride local buses to node $i$ and transfer there Board at node $i$</td>
<td>![Diagram 5]</td>
</tr>
<tr>
<td>$i_{LE}^{jk}^{gh}$</td>
<td>Local route whose presence marks the end of the express service market at segment $k-1$. Picks up downtown pax over the latter part of its boarding zone (after $k$)</td>
<td>![Diagram 6]</td>
</tr>
<tr>
<td>$i_{LE}^{jk}^{gh}$</td>
<td>Same as route type $L_1$, but since it is the first (longest) local route it picks up at $i$ local pax who rode express routes to node $i$ and transferred.</td>
<td>![Diagram 7]</td>
</tr>
</tbody>
</table>

1/ Not all the express passengers from the zones for which transfers are expected must transfer; only a fraction will transfer.
\[ d_i = \text{local inbound demand originating along segment } i \]
\[ e_i = \text{downtown (potentially express) demand originating along segment } i \]
\[ u_i = \text{outbound demand alighting along segment } i \]

Also, in order for the model to be behaviorally accurate, we have to recognize that not all downtown passengers who must board a local bus (because an express bus does not serve their origin segment) will want to transfer to an express route that begins downstream. Some may prefer to stay on the local route. Therefore only a fraction of the downtown passengers in that situation will transfer. This fraction may be determined by a simple all-or-nothing assignment, setting it to 1 if the travel time savings of transferring to the express route exceeds a threshold and 0 if otherwise; or it may be determined by a more complex choice model such as a logit function. The case study examples used the simple all-or-nothing assignment, but the user desiring greater accuracy could use whatever assignment rule he choose.

In addition to the 11 route types defined in Tables 6.1 and 6.2, we must define 7 state variables, as follow:

\[ i_{ZB}^{jk}_g = \text{minimum cost of serving demands } d_1, \ldots, d_j, e_1, \ldots, e_k, \]
\[ \text{and } u_1, \ldots, u_g \text{ with routes beginning at nodes } 1, \ldots, i, \]
\[ \text{given that local service is offered at segments } i, \ldots, j \]
\[ \text{and express service is offered at segments } i, \ldots, k. \]
\[ (j \geq i; n > k \geq i) \]

\[ i_{ZL}^{jk}_g = \text{minimum cost of serving demands } d_1, \ldots, d_j, e_1, \ldots, e_k, \]
\[ \text{and } u_1, \ldots, u_g \text{ with local routes beginning at nodes } 1, \ldots, \]
i, given that an express route will begin at node k+1 and the correct fraction of downtown demands $e_1, \ldots, e_k$ will alight there to transfer. The transfer penalty of those transferring is included, but the cost of transporting them on an express route beyond node k+1 is not. ($j > k > i$)

$$i_{ZE}^{j,k}_g = \text{minimum cost of serving demands } d_1, \ldots, d_j, e_1, \ldots, e_k, \text{and } u_1, \ldots, u_g \text{ with express routes beginning at nodes } 1, \ldots, i, \text{ given that a local route will begin at node } j+1 \text{ and expecting all local demands } d_1, \ldots, d_j \text{ who are destined for segments } j+1, \ldots, n \text{ to alight at node } j+1 \text{ to transfer. The transfer penalty for those transferring is included, but the cost of transporting them on a local route beyond node } j+1 \text{ is not.} \quad (n > k > j > i)$$

$$i_{ZN}^{j}_g = \text{minimum cost of serving demands } d_1, \ldots, d_j, e_1, \ldots, e_j, \text{and } u_1, \ldots, u_g \text{ with routes beginning at nodes } 1, \ldots, i, \text{ given that only local service is offered at segments } i, \ldots, n. \quad (j > i)$$

$$i_{ZX}^{j,k}_g = \text{minimum cost of serving demands } d_1, \ldots, d_{i-1}, e_1, \ldots, e_k, \text{and } u_1, \ldots, u_g \text{ with local routes beginning at nodes } 1, \ldots, i-1 \text{ and express routes beginning at nodes } 1, \ldots, \ell. \quad (n > k > \ell > i)$$

$$i_{YL}^{j,t}_g = \text{minimum cost of serving demands } d_1, \ldots, d_j, e_1, \ldots, e_j, \text{and } u_1, \ldots, u_g \text{ with local routes beginning at nodes } 1, \ldots, i, \text{ given that the outermost node with express ser-}$$
vice is node \( t \) and that the correct fraction of downtown demands \( e_1, \ldots, e_j \) will transfer at \( t \) to an express route. The transfer penalty for those transferring is included, but the cost of transporting them on an express route beyond node \( t \) is not. \((n > t > j \geq i)\)

\[
i_{Ye_{gj}} = \text{minimum cost of serving demands } d_1, \ldots, d_j, e_1, \ldots, e_g,\]

and \( u_1, \ldots, u_g \) with express routes beginning at nodes \( l, \ldots, i \), given that the outermost node with local service is node \( t \) and that all local demands \( d_1, \ldots, d_j \) who are destined for segments \( t, \ldots, n \) will transfer at \( t \) to a local route. The transfer penalty for those transferring is included, but the cost of transporting them on a local route beyond node \( t \) is not. \((t > j \geq i)\)

In order to better illustrate the logic of how the state variables are computed, the following symbols are used:

\[
\begin{array}{c}
\text{segment or node} \\
\downarrow \quad \downarrow \\
\text{local demands } (d_j, \text{ etc.}) \\
\hspace{2cm} \text{downtown demands } (e_j, \text{ etc.}) \\
\hspace{4cm} \text{outbound demands } (u_{ij}, \text{ etc.}) \\
\text{segment or node} \\
\end{array}
\]

\[
\begin{array}{c}
\text{demand served entirely by a local route beginning at node } i
\end{array}
\]
- demand served entirely by an express route beginning at node $i$ for $i_k^g x_k^g$

- demand served entirely by routes beginning upstream of node $i$

- demands picked up by routes beginning upstream of node $i$ of which the correct fraction will transfer to a route beginning at node $i$

- demands picked up on a route beginning at node $i$ of which the proper fraction will transfer at node $t$

The initial values of the state variables are found using the following relationships:

$$l_{ZB}^{jk}_g = \min_{v=1, g+1} \left\{ \min \left[ \left( L^{lj}_g v + E^{lk}_{v-1, 1} \right), \frac{1}{k v g} \right] \right\}$$

for $j=1, \ldots, n$; $k=1, \ldots, n$; $g=\min(j, k), \ldots, n$

(6.1)
\[ l_{ZL}^{jk} = l_{LS}^{jk} \text{ for } j=1, \ldots, n; \ k=1, \ldots, j; \]
\[ g=h, \ldots, n \]  

(6.2)

\[ l_{ZE}^{jk} = l_{ES}^{jk} \text{ for } k=1, \ldots, n; \ j=1, \ldots, k; \]
\[ g=j, \ldots, n \]  

(6.3)

\[ l_{YL}^{jt} = l_{L}^{jt} \text{ for } t=1, \ldots, n-1; \ j=1, \ldots, t-1; \ g=j, \ldots, n \]  

(6.4)

\[ l_{YE}^{jt} = l_{E}^{jt} \text{ for } t=1, \ldots, n; \ j=1, \ldots, t-1; \]
\[ g=j, \ldots, n \]  

(6.5)

\[ l_{ZX}^{k} = l_{ZX}^{k} \text{ for } k=1, \ldots, n-1; \ g=0, \ldots, n \]  

(6.6)

\[ l_{Z}^{k} = \min \left\{ l, l^{-1} l_{Z}^{k} \right\} \]
\[ \min \left( l, l^{-1} l_{Z}^{k} + l_{P}^{k} \right) \]
\[ h=l, g+1 \]  

for \( k=2, \ldots, n-1; \ l=2, \ldots, k; \ g=0, \ldots, n \)  

(6.7)
\[ l_{ZN}^j_g = \min \left\{ \min_{k=1,j-1} \min_{h=1,g+1} \left( l_{Zk}^{h-1} + \frac{l_{LN}^j_{gh}}{g} \right) \right\} \]

for \( j=1,\ldots,n; \ g=j,\ldots,n \)  

(6.8)

The complexity of the variables involved in the above equations renders any attempt at a verbal clarification futile. Carefully examining each term and the symbolic representations given at the right is the best way to understanding the relationships. I will only remark that \( l_{Zk}^i \) has a peculiar structure, allowing express routes to begin at nodes downstream of the stage node \( i \), because this variable is needed in computing \( l_{ZN}^j_g \).

At subsequent stages \( i=2,\ldots,n \), the following recursions update the state variables:
\[
\begin{align*}
    Z_{g}^{i,j,k} &= \min \left\{ Z_{h-1}^{i,j,k} + \min_{v=h,g+1} \left[ \min_{h,v,g} \left( L_{g}^{j,v-1,h} + E_{g}^{k} \right) \right] \right\} \\
    &= \min \left\{ Z_{h-1}^{i,j,k} + \min_{v=h,g+1} \left[ \min_{h,v,g} \left( L_{g}^{j,v-1,h} + E_{g}^{k} \right) \right] \right\}
\end{align*}
\]
\[ i_{iZL^j}^k = \min \left\{ \begin{array}{l}
\frac{i-l_{ZL^j}^k}{g} \\
\min_{h=i,g+1} \left( i_{iY_L^j}^{i-l,h+1} + i_{L^j}^{g} \right) 
\end{array} \right\} 
\text{for } j=i,\ldots,n; \; k=i,\ldots,j; \; g=k,\ldots,n 
\] (6.10)

\[ i_{iZE^j}^k = \min \left\{ \begin{array}{l}
\frac{i-l_{ZE^j}^k}{g} \\
\min_{h=i,g+1} \left( i_{iY_E^j}^{i-l,h+1} + i_{E^j}^{g} \right) 
\end{array} \right\} 
\text{for } k=i,\ldots,n; \; j=1,\ldots,k; \; g=j,\ldots,n 
\] (6.11)

\[ i_{iY_L^j}^t = \min \left\{ \begin{array}{l}
\frac{i-l_{Y_L^j}^t}{g} \\
\min_{h=i,g+1} \left( i_{iY_L^j}^{i-l,h+1} + i_{L^j}^{g} \right) 
\end{array} \right\} 
\text{for } t=i+1,\ldots,n-1; \; j=i,\ldots,t-1; \; g=j,\ldots,n 
\] (6.12)

\[ i_{iY_E^j}^t = \min \left\{ \begin{array}{l}
\frac{i-l_{Y_E^j}^t}{g} \\
\min_{h=i,g+1} \left( i_{iY_E^j}^{i-l,h+1} + i_{E^j}^{g} \right) 
\end{array} \right\} 
\text{for } t=i+1,\ldots,n; \; j=i,\ldots,t-1; \; g=j,\ldots,n 
\] (6.13)

\[ i_{i,i-l_{ZX^k}^j}^g = i_{l_{ZB}}^{i-l,k} \text{ for } k=i,\ldots,n-1; \; g=i-1,\ldots,n 
\] (6.14)
\[ i^l_{ZXg} = \min \left\{ i, \frac{\ell - 1}{Z_Xg} k \right\} \min_{h=\ell, g+1} \left( i, \frac{\ell - 1}{Z_Xh} + \frac{\ell}{E_Xgh} \right) \]

for \( k = i, \ldots, n-1; \ell = i, \ldots, k; g = i-1, \ldots, n \)

(6.15)

\[ i^l_{ZNjg} = \min \left\{ \min_{h=i, g+1} \left( i, \frac{\ell - 1}{ZNh} + \frac{i}{B_Ojgh} \right), \right\} \min_{k=i, j} \left( i, \frac{\ell - 1}{ZBh} + \frac{\ell}{B_Ojgh} \right) \]

\[ i^l_{Zjg} = \min \left\{ \min_{h=i, g+1} \left( i, \frac{\ell - 1}{Zjkh} + \frac{i}{L_Nj,kgh} \right), \right\} \min_{k=i, j} \left( i, \frac{\ell - 1}{ZEh} + \frac{\ell}{L_Fj,kgh} \right) \]

for \( j = i, \ldots, n; g = j, \ldots, n \)

(6.16)

This algorithm requires \( O(n^6) \) computations. The optimal solution that allows both local and express routes is \( n_{ZN} \). The optimal solution for local service only can also be found; it is \( n_{ZL} \).

The form of the solution generated by this algorithm is a set of local and express routes, each with their service zones (inbound local, inbound
downtown, and outbound), headways, vehicle type, number of vehicles, instructions about transfers, and any passenger data one would like such as peak load, total wait time, average crowding level, etc.

6.3 **Express/Local Zonal Service Case Study**

The algorithm for joint design of express and local services was applied to the Watertown corridor of Boston in the same way as the algorithm for local zonal service in Section 5.5. The same four existing terminals were chosen as potential terminals (nodes) in the express/local system: Watertown Square, Newton Corner, Oak Square and Brighton Center. Two cases were studied: serving the entire corridor from Watertown to downtown, and serving the Route 57 demands only, from Watertown to Kenmore Square. Both cases were analyzed under both the austerity and the prosperity objectives described in Section 3.3.4. Then, in addition, the same cases were repeated with the restriction that buses could not use the Massachusetts Turnpike in order to demonstrate the value of the express/local strategy in the absence of a high speed road as well as when one is available.

6.3.1. **Serving The Entire Watertown Corridor**

Because of the existence of the Massachusetts Turnpike, express service improves the efficiency of the Watertown bus service considerably. Under the austerity objective, it was found that the entire corridor could be most efficiently served with a configuration composed of one local route and two express routes. In the inbound direction, all local passengers are served without transfer by the local route, as are the downtown
passengers from the innermost segment. Downtown passengers from the other segments are served without transfer by the express routes, the first covering segment 1 and the second covering segments 2 and 3. The out-bound demand is served by the local route, which offers a service headway twice as large outbound as inbound, deadheading half its runs. This optimal configuration is shown in Figure 6.1.

The above configuration requires 24 articulated buses, whereas the best conventional local service configuration requires 34 articulated buses, yielding a 29% reduction in operator costs. As the optimal express/local configuration has 3 routes, and one of them has partial deadheading, total wait time increases more than threefold; nevertheless, inbound headways are all under 8 minutes and outbound headways under 15 minutes, so that wait times should be acceptable. Average in-vehicle time per passenger is 5.0 minutes less than conventional local service due to travel time savings on the express and zonal routes. The in-vehicle time savings more than offset the increase in wait time, so that average travel time (wait plus in-vehicle time) declines by 1.4 minutes per passenger even while 10 fewer articulated buses are used. The details of this solution are found in Table 6.3. Also found in Table 6.3 is the optimal configuration if partial deadheading is not allowed; it requires one more articulated bus, while improving level of service a bit.

The optimal express/local configuration for the Watertown corridor under the prosperity objective is the same as the configuration found under the austerity objective, except that the route serving the outbound demands does not employ partial deadheading. In response to the value
Figure 6.1

Optimal Express and Local Service Zones in the Watertown Corridor

Under the Austerity Objective

Legend:

0---0 = service zone (boarding zone inbound, alighting zone outbound)

--- --- = express and deadhead route portions
Table 6.3
Optimal Express/Local Zonal Service for the Watertown Corridor
Under the Austerity Objective

<table>
<thead>
<tr>
<th>Route Type</th>
<th>Segments In Inbound Boarding Zone</th>
<th>Segments In Outbound Boarding Zone</th>
<th>Inbound Service Headway</th>
<th>Outbound Service Headway</th>
<th># Buses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inbound</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Local</td>
<td>Local: 1-4</td>
<td>4-1</td>
<td>7.2 min</td>
<td>14.4 min</td>
<td>12 Artic</td>
</tr>
<tr>
<td></td>
<td>Downtown: 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Express</td>
<td>Downtown: 1</td>
<td></td>
<td>6.6 min</td>
<td>6.6 min</td>
<td>6 Artic</td>
</tr>
<tr>
<td>Express</td>
<td>Downtown: 2-3</td>
<td></td>
<td>8.0 min</td>
<td>8.0 min</td>
<td>6 Artic</td>
</tr>
<tr>
<td>Without Partial Deadheading</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Local</td>
<td>Local: 1-4</td>
<td>4-1</td>
<td>7.2 min</td>
<td>7.2 min</td>
<td>13 Artic</td>
</tr>
<tr>
<td></td>
<td>Downtown: 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Express</td>
<td>Downtown: 1</td>
<td></td>
<td>6.6 min</td>
<td>6.6 min</td>
<td>6 Artic</td>
</tr>
<tr>
<td>Express</td>
<td>Downtown: 2-3</td>
<td></td>
<td>8.0 min</td>
<td>8.0 min</td>
<td>6 Artic</td>
</tr>
</tbody>
</table>

Service Level and Costs

<table>
<thead>
<tr>
<th></th>
<th>Average Wait Time</th>
<th>Average In-Vehicle Time</th>
<th>Operator Cost @ $3/hr</th>
<th>Total Cost</th>
<th>Operator Cost Savings Over Conventional Local Service</th>
</tr>
</thead>
<tbody>
<tr>
<td>With Partial Deadheading</td>
<td>5.0 min</td>
<td>17.3 min</td>
<td>$ 961</td>
<td>$3,747</td>
<td>$400</td>
</tr>
<tr>
<td>Without Partial Deadheading</td>
<td>4.3 min</td>
<td>17.1 min</td>
<td>$1,001</td>
<td>$3,675</td>
<td>$366</td>
</tr>
</tbody>
</table>
given to passenger wait and in-vehicle time, however, headways are reduced to such an extent that average wait time is 3 minutes less than it is under the austerity objective. Standard buses are used instead of articulated, so that bus loads at the peak load point are still near bus capacity. Compared to the best configuration for conventional local service under this objective, 11 fewer standard buses are used, a decrease of 24%, and average in-vehicle time per passengers declines from 20.8 minutes to 16.1 minutes while average wait time rises from 1.0 to 2.9 minutes per rider. Thus the express/local strategy reduces the average passenger's trip by 3.2 minutes while reducing the required number of vehicles by 24%. The reduction in total cost (operator plus passenger) is $691.54, the equivalent of over 21 standard buses. Table 6.4 provides the details about this solution.

Express/local zonal service is a significant step in efficiency over simple local zonal service in the Watertown corridor. Under the austerity objective it requires 3 fewer articulated buses than local zonal service, (5 fewer if partial deadheading is forbidden), and under the prosperity objective its total passenger plus operator costs are $499 less, the equivalent of 15 standard buses or 4.0 minutes of travel time per passenger.

6.3.2 Serving Route 57 Demands

Service from Watertown Square to Kenmore Square carrying the existing demand on Route 57 can also be made more efficient by employing an express/local zonal strategy if one seeks to minimize operator cost only; if one also values passenger time at $3/hour, conventional local service is superior.
Table 6.4

Optimal Express/Local Zonal Service for the Watertown Corridor
Under the Prosperity Objective

<table>
<thead>
<tr>
<th>Route Type</th>
<th>Segments In Inbound Boarding Zone</th>
<th>Segments In Outbound Boarding Zone</th>
<th>Inbound Service Headway</th>
<th>Outbound Service Headway</th>
<th># Buses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local</td>
<td>Local: 1-4</td>
<td>4-1</td>
<td>4.8 min</td>
<td>4.8 min</td>
<td>18 Std</td>
</tr>
<tr>
<td></td>
<td>Downtown: 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Express</td>
<td>Downtown: 1</td>
<td>---</td>
<td>4.9 min</td>
<td>---</td>
<td>8 Std</td>
</tr>
<tr>
<td>Express</td>
<td>Downtown: 2-3</td>
<td>---</td>
<td>5.1 min</td>
<td>---</td>
<td>9 Std</td>
</tr>
</tbody>
</table>

Service Level and Costs

<table>
<thead>
<tr>
<th>Average Wait Time</th>
<th>Average In-Vehicle Time</th>
<th>Operator Cost</th>
<th>Total Cost (pax time valued @ $3/hr)</th>
<th>Total Cost Savings Over Conventional Local Service</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.9 min</td>
<td>16.1 min</td>
<td>$1,122</td>
<td>$3,509</td>
<td>$691</td>
</tr>
</tbody>
</table>
The optimal configuration under the austerity objective consists of a local route and an express route. The local route covers the entire length of the corridor, serving all local passengers and segment 4 downtown passengers without transfer. This route deadheads outbound using the turnpike. The express route also begins at Watertown Square, picks up passengers destined for Kenmore Square originating within the first three segments (up to Brighton Center), from where it travels express (without using the turnpike) to Kenmore Square. On its return trip it serves all outbound trips as a local route.

The local route calls for 5 articulated buses while the express route calls for 6 standard buses. The express route's demand is so small that, even if operating at the maximum headway of 15 minutes, it would have 18 empty places at the peak load point if it used articulated buses. (If the policy headway were raised to 20 minutes, then 4 articulated buses would be used on the express route, for an operator cost reduction of $32.)

The operating cost savings of this configuration over conventional local service is $48, the value of 1.5 standard buses. Relative to the best zonal configuration, the savings is $24. Table 6.5 contains the details of this solution.

Under the prosperity objective, conventional local service is superior to any zonal configuration with express and local routes. The demand served by Route 57 is so small (peak load of 575 riders per hour) that segmenting the route into 2 markets makes wait times on these two routes (which must operate at longer headways) too great to offset the reduction in vehicle usage and in-vehicle time the zonal strategy affords.
Table 6.5

Optimal Express/Local Zonal Service for Route 57 Demands
Under the Austerity Objective

<table>
<thead>
<tr>
<th>Route Type</th>
<th>Segments In Inbound Zone</th>
<th>Segments In Outbound Zone</th>
<th>Headway</th>
<th>Buses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local</td>
<td>Local: 1-4</td>
<td>---</td>
<td>12.4 min</td>
<td>5 Artic</td>
</tr>
<tr>
<td></td>
<td>Downtown: 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Express</td>
<td>Downtown: 1-3</td>
<td>4-1</td>
<td>11.5 min</td>
<td>6 Std</td>
</tr>
</tbody>
</table>

Service Level and Costs

<table>
<thead>
<tr>
<th>Average Wait Time</th>
<th>Average In-Vehicle Time</th>
<th>Operator Cost</th>
<th>Total Cost (pax time valued @ $3/hr)</th>
<th>Operator Cost Savings Over Conventional Local Service</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.2 min</td>
<td>12.9 min</td>
<td>$392</td>
<td>$1,607</td>
<td>$48</td>
</tr>
</tbody>
</table>
6.3.3 **Express/Local Zonal Service Without Using the Turnpike**

It was stated earlier that the express/local zonal strategy can be better than the simpler local zonal service strategy even in the absence of a high speed road for the express routes. Using express routes has two advantages. First, some time is saved by not stopping at all outside the service zone, even though the express buses travel along the same path as local buses. Second, using express buses allows for the downtown passengers to be segmented from the local market, leading to routes better tailored to their markets and thus eliminating some of the empty places at the corridor peak load point that are built into the local zonal configuration.

In order to demonstrate the relative merit of the local zonal and express/local zonal strategies in the absence of a high speed road, we required buses to use the same path as the local route even while deadheading or traveling express. Optimal designs were then found for the entire Watertown corridor under both the austerity and the prosperity objective. The optimal design for serving Route 57 demands under the austerity objective was also found. (Since under the prosperity objective the local zonal and express/local zonal strategies could not improve on conventional local service for serving Route 57 demands when buses could use the turnpike, it is obvious that without the turnpike conventional local service would still be the superior configuration under the prosperity objective.)

Table 6.6 shows the optimal designs for the entire corridor under the austerity objective. Compared to conventional local service, local zonal
### Table 6.6

Local Zonal and Express/Local Zonal Service For the Watertown Corridor

In the Absence of the Turnpike:

Austerity Solutions

<table>
<thead>
<tr>
<th>Route Type Inbound</th>
<th>Segments In Inbound Boarding Zone</th>
<th>Segments In Outbound Alighting Zone</th>
<th>Headway</th>
<th># Buses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local</td>
<td>1</td>
<td>3-1</td>
<td>4.8 min</td>
<td>16 Artic</td>
</tr>
<tr>
<td>Zonal Service</td>
<td>2</td>
<td>4</td>
<td>13.5 min</td>
<td>5 Artic</td>
</tr>
<tr>
<td></td>
<td>3-4</td>
<td>---</td>
<td>5.9 min</td>
<td>11 Artic</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Route Type Inbound</th>
<th>Segments In Inbound Boarding Zone</th>
<th>Segments In Outbound Alighting Zone</th>
<th>Headway</th>
<th># Buses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Express</td>
<td>Downtown: 1-2</td>
<td>4-3</td>
<td>5.5 min</td>
<td>12 Artic</td>
</tr>
<tr>
<td>Local</td>
<td>Local: 1-2</td>
<td>2-1</td>
<td>10.4 min</td>
<td>8 Artic</td>
</tr>
<tr>
<td>Zonal Service</td>
<td>Local: 3-4</td>
<td>---</td>
<td>6.1 min</td>
<td>10 Artic</td>
</tr>
<tr>
<td></td>
<td>Downtown: 3-4</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Service Level and Costs

<table>
<thead>
<tr>
<th></th>
<th>Average Wait Time</th>
<th>Average In-Vehicle Time</th>
<th>Operator Cost</th>
<th>Total Cost (pax time valued @ $3/hr)</th>
<th>Operator Cost Savings Over Conventional Local Service</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local Zonal Service</td>
<td>3.9 min</td>
<td>20.6 min</td>
<td>$1,281</td>
<td>$4,340</td>
<td>$ 80</td>
</tr>
<tr>
<td>Express/Local Zonal Service</td>
<td>4.1 min</td>
<td>20.0 min</td>
<td>$1,201</td>
<td>$4,212</td>
<td>$160</td>
</tr>
</tbody>
</table>
service saves 2 articulated buses, and express/local zonal service
saves 4; these are relative reductions of 6 and 12 percent in the required
number of vehicles. The best local zonal configuration has three routes,
beginning at nodes 1, 2, and 3. The best express/local zonal configuration
has an express route serving segments 1 and 2 (Watertown Square to Oak
Square); all remaining passengers are served by two local routes, one of
which begins at Watertown Square and the other at Oak Square. No passen-
gers transfer within the corridor.

Table 6.7 shows the designs for the entire corridor under the pros-
perity objective. Because of the value placed on wait time, both optimal
configurations have just two routes. Under the local zonal strategy, one
route begins at Watertown Square and the other at Oak Square. Under the
express/local zonal strategy, downtown passengers between Watertown Square
and Oak Square are still served by an express route, while the remaining
passengers ride a local route that begins at Watertown Square. The local
zonal strategy reduces total operator plus passenger cost by $55 per hour,
while the express/local zonal strategy reduces total cost by $197 per hour.
These savings are equivalent to the cost of operating 1.7 and 6.1 standard
buses, respectively.

These cases show the benefit of operating express routes even when
express buses cannot use a high speed path. A good express/local configu-
ration can save over 10% in operator costs, or reduce total operator plus
passenger costs by the equivalent of over 10% of operator costs.

Even when the demands are as low as those existing on Route 57, the
express/local strategy can be of value. Table 6.8 shows the optimal
### Table 6.7
Local Zonal and Express/Local Zonal Service For the Watertown Corridor
In the Absence of the Turnpike:
Prosperity Solutions

<table>
<thead>
<tr>
<th>Route Type</th>
<th>Segments In</th>
<th>Segments In</th>
<th>Headway</th>
<th>Buses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Inbound Zone Boarding</td>
<td>Outbound Alighting Zone</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Local Zonal Service</td>
<td>Local</td>
<td>1-2</td>
<td>2-1</td>
<td>2.5 min</td>
</tr>
<tr>
<td>Local Zonal Service</td>
<td>Local</td>
<td>3-4</td>
<td>4-3</td>
<td>4.2 min</td>
</tr>
<tr>
<td>Express/Local Zonal Service</td>
<td>Local</td>
<td>Local: 1-4</td>
<td>4-3</td>
<td>3.2 min</td>
</tr>
<tr>
<td>Express/Local Zonal Service</td>
<td>Express</td>
<td>Downtown: 3-4</td>
<td>2-1</td>
<td>3.7 min</td>
</tr>
</tbody>
</table>

### Service Level and Costs

<table>
<thead>
<tr>
<th></th>
<th>Average Wait Time</th>
<th>Average In-Vehicle Time</th>
<th>Operator Cost</th>
<th>Total Cost (pax time valued @ $3/hr)</th>
<th>Total Cost Savings Over Conventional Local Service</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local Zonal Service</td>
<td>1.9 min</td>
<td>19.9 min</td>
<td>$1,410</td>
<td>$4,144</td>
<td>$55</td>
</tr>
<tr>
<td>Express/Local Zonal Service</td>
<td>2.0 min</td>
<td>19.2 min</td>
<td>$1,346</td>
<td>$4,004</td>
<td>$195</td>
</tr>
</tbody>
</table>
designs for both local zonal and express/local zonal service under the austerity objective. Local zonal service reduces operator cost by $23 relative to conventional local service, while express/local service yields a savings of $40; these savings are equivalent to 5 and 9 percent of operator cost, respectively. Since Kenmore Square does not lend itself as well as downtown to being served by the turnpike, the elimination of the turnpike has a small effect on optimal design. The local zonal configuration is unchanged, because the turnpike was not even used in the unrestricted solution. The express/zonal local configuration changes, but the objective function without the turnpike is only $8 greater than it is with the turnpike. The express/local strategy without the turnpike has the same inbound configuration as it does with the turnpike; however, the advantage of deadheading one route is gone when buses cannot use the turnpike, so that both routes serve a portion of the outbound demand.

Before concluding this case study, it is worth pointing out that none of the solutions produced had passengers transferring between express and local routes in the outer zones. As the beginning of this chapter stated, a configuration that calls for downtown passengers to board a local route and then transfer to an express route, or for local passengers to board an express route and then transfer to a local route, would be expected when the outer zones do not have enough demand to support both a local and an express route. In our case study demands in the outer zones were high owing to relatively dense residential development around Watertown Square and Newton Corner. For this reason applying the express/local strategy
Table 6.8

Local Zonal and Express/Local Zonal Service For Route 57 Demands
In the Absence of the Turnpike:
Austerity Solutions

<table>
<thead>
<tr>
<th>Route Type</th>
<th>Segments In Inbound Zone</th>
<th>Segments In Outbound Alighting Zone</th>
<th>Headway</th>
<th># Buses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local Zonal Service</td>
<td>Local</td>
<td>1-2</td>
<td>4-1</td>
<td>10.7 min</td>
</tr>
<tr>
<td></td>
<td>Local</td>
<td>3-4</td>
<td>---</td>
<td>7.7 min</td>
</tr>
<tr>
<td>Express/Local Zonal Service</td>
<td>Local</td>
<td>Local: 1-4</td>
<td>4-3</td>
<td>12.1 min</td>
</tr>
<tr>
<td></td>
<td>Downtown: 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Express</td>
<td>Downtown: 1-3</td>
<td>2-1</td>
<td>12.6 min</td>
</tr>
</tbody>
</table>

Service Level and Costs

<table>
<thead>
<tr>
<th></th>
<th>Average Wait Time</th>
<th>Average In-Vehicle Time</th>
<th>Operator Cost</th>
<th>Total Cost (pax time valued @ $3/hr)</th>
<th>Operator Cost Savings Over Conventional Local Service</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local Zonal Service</td>
<td>5.7 min</td>
<td>12.5 min</td>
<td>$417</td>
<td>$1,510</td>
<td>$23</td>
</tr>
<tr>
<td>Express/Local Zonal Service</td>
<td>7.3 min</td>
<td>12.6 min</td>
<td>$400</td>
<td>$1,601</td>
<td>$40</td>
</tr>
</tbody>
</table>
always led to having both local and express routes beginning at Watertown Square.

6.4 Conclusions

Operating both express and local service in a corridor with zonal routes can be an efficient strategy for high and medium demand corridors. The case studies in this chapter report savings of 30% in operator costs or savings equivalent to 47% of operator cost in total passenger plus operator cost in a corridor with high demand with an expressway available for express and deadhead runs. In a lower demand corridor, operator cost could be reduced by 10% while keeping service level in an acceptable range.

These case studies also demonstrate, perhaps more importantly, the value of using express and local routes together when there is no expressway available. In a high demand corridor, it was shown that operator costs could be reduced by 12%, or total operator plus passenger costs reduced by the equivalent of 14% of operator cost. In a low demand corridor, operating an express route along with a local route could reduce operator costs by 9%.

Express routes have long been recognized as efficient for long distance routes in a corridor that has an expressway. It has always been assumed, however, that express and local services in the same corridor would be designed separately. This chapter has shown the interaction of express and local route design in a corridor, and has offered a procedure by which the optimal design for both types of service can be found. It
has proven the potential of express routes in both heavy and medium demand corridors, both in the existence and in the absence of a high speed road.
Chapter 7
Zonal Service On Branching Networks

All of the routing strategies covered so far have been applied to the simplest network only: a single arterial. While many urban corridors exhibit route structures that are this simple, another common route structure is a branching (or tree) network. A branching network as defined in this chapter consists of a trunk road emanating from the downtown terminal and roads that branch off from this trunk. The branching roads may themselves have roads that branch off them. It is important, however, that there be no cycles in the network; that is, there must be only one feasible path from any point on the network to downtown. As we assumed in the single corridor network, including a road segment in the network implies that there will be service along that road segment. For the problem we are studying, we must assume that the corridor peak load lies on the trunk. (If peak loads occur on branches before they reach the trunk, then each branch could be designed independently, since zones are designed for the area of a corridor upstream of the peak load point only.) Uptown terminals may exist anywhere on the network upstream of the peak load point; of course, terminals must exist at each extreme point in order for all road segments to be covered. Figure 7.1 shows a typical branching network for which we might want to find the optimal routing configuration.

A branching network cannot be served by a conventional local route, but must be served by a set of routes, each of which will have its own boarding and alighting zones. These routes may have overlapping zones,
Figure 7.1
Branching Network

1 -- 2 -- 3 -- 4 -- 5 -- 6
    |      |      |      |
   8 -- 9 -- 7 -- 10
      |      |
    11 12  Downtown
for example, when routes serving two different branches both include
the trunk in their boarding zone. The routes may also have exclusive
boarding and alighting zones, imposing boarding/alighting restrictions
along the trunk and possibly on other road segments as well. Hybrid
routing systems may also exist, having some boarding/alighting restric-
tions and having some zone overlap as well. Express routes may belong
to the system, and routes in the system may partially or completely dead-
head. We seek, in this chapter, to find the system of routes that mini-
mizes a sum of operator and passenger cost.

Chapter 5 discussed how even in a single arterial network it is hard
to find an optimal system that includes routes with overlapping zones
because of the indeterminancy of passenger path choice. Design of routes
with overlapping zones requires more than choosing a service frequency,
but must consider a range of things such as boarding disincentives, phasing
of departures, and the possibility of "induced transfers". For these
reasons models were not developed for overlapping route systems on single
arterial network. The problems of overlapping zones are the same in
branching networks, and therefore we will leave this routing strategy out-
side the scope of problems for which we seek an optimal solution algorithmi-
cally. Overlapping zonal service is nevertheless a good routing strategy,
perhaps the best in some situations, and by careful application of the
design principles set forth in this thesis one can manually design a good
solution.

We are left then with the problem of finding the optimal design for
the local zonal and express/local zonal strategies under the policy of
exclusive boarding/alighting zones. After discussing these two problems
in Sections 7.1 and 7.2, a case study of a branching route system in Minneapolis will be used in Section 7.3 to illustrate the process of designing a branching route system.

7.1 Local Zonal Service on a Branching Network

Because there are no cycles in a branching network, the capacity scheduling formulation of the zonal service design problem can be applied to a branching network as well as to a single arterial network and still yield a network problem that solves in a spanning tree solution. To illustrate this capacity scheduling formulation, let us use the simple branching network of Figure 7.2. We shall consider inbound flows only, and assume for now that all demands are destined for the downtown terminal.

Let $q_i$ be the frequency, in runs per hour, of the route that operates between node $i$ and downtown. Let $d_i$ be the demand, in busloads per hour, originating along the segment $i$, which is the segment between node $i$ and the next downstream node. The loading constraint at the downstream end of each segment requires that the total capacity provided at that point, which is the sum of the frequencies of all the routes originating upstream of that point, must exceed the load at that point, which is the sum of the demands of that segment and every upstream segment. If we arrange all of the loading constraints for Figure 7.2 in a matrix, we obtain:
Figure 7.2

a. A Simple Tree Network

b. A Network Representation of the Zonal Branching Route Planning Problem
\[
\begin{bmatrix}
q_1 & q_2 & q_3 & q_4 & q_5 & s_1 & s_2 & s_3 & s_4 & s_5 & d_1 & d_2 & d_3 & d_4 & d_5 \\
1 & -1 \\
1 & 1 & -1 \\
1 & 1 & -1 \\
1 & 1 & 1 & 1 & -1 \\
1 & 1 & 1 & 1 & 1 & -1 \\
\end{bmatrix}
= 
\begin{bmatrix}
1 \\
1 \\
1 \\
1 \\
1 \\
1 \\
\end{bmatrix}
\]

Again, \( s_i \) is the slack variable for the \( i \)th constraint. Subtracting rows from one another we transform the constraint set to the constraint set of a transshipment problem:

\[
\begin{bmatrix}
q_1 & q_2 & q_3 & q_4 & q_5 & s_1 & s_2 & s_3 & s_4 & s_5 & d_1 & d_2 & d_3 & d_4 & d_5 \\
1 & -1 \\
1 & 1 & -1 \\
1 & 1 & -1 \\
1 & 1 & 1 & 1 & -1 \\
-1 & -1 & -1 & -1 & -1 & 1 \\
\end{bmatrix}
= 
\begin{bmatrix}
1 \\
1 \\
1 \\
1 \\
1 \\
-1 \\
\end{bmatrix}
\]

(7.1)

Constraints (7.1), together with the objective of minimizing transportation cost, can be interpreted as a transshipment problem on the network described in Figure 7.2b. This transshipment network illustrates how at extreme nodes such as nodes 1 and 3 the entire demand of segment \( i \) (\( d_i \)) must be carried by the route beginning at node \( i \). At node 2, which is neither an extreme node nor a junction node in the physical network of Figure 7.2a,
one can see that the demand of its segment can either be served by its own route (i.e. \( q_2 \geq d_2 \) and \( s_1 = 0 \)), or by a route that began upstream (i.e. \( q_2 = 0 \) and \( s_1 \geq d_2 \)), as in the single arterial problem. (Since we impose the constraint of no zonal overlap, as we did in Chapter 5, either \( s_1 \) or \( q_2 \) must equal zero.) At node 4, which is a junction node in the physical network, the transshipment network shows how the demand of segment 4 can either be served by route 4 (\( q_4 \geq d_4 \)), in which case \( s_2 = s_3 = 0 \), or by one of the upstream branches. To avoid zonal overlap, it is necessary that the demand of segment 4 be served by at most one upstream branch; therefore of the flows \( s_2 \) and \( s_3 \), at most one of them may be nonzero.

The structure of the transshipment network of Figure 7.2b indicates that this small problem can be solved by twice decomposing the network into two subgraphs and then applying the single arterial dynamic programming algorithm to each subgraph. The first decomposition is achieved by setting \( s_2 = 0 \). This separates nodes 1 and 2 from nodes 3, 4, and 5, thereby leaving one branch (node 3 and segment 3) "connected" to the trunk and the other branch "unconnected". When a branch is "connected" to a trunk, this means that segments of the branch and of the trunk may belong to the same boarding zone; when a branch and its corresponding trunk are "unconnected", the branch segments may not belong to the same boarding zone as trunk segments. It is obvious that these two resulting subgraphs are network representatives of two single arterial problems, and therefore can each be optimized using the single arterial algorithms of Chapter 5. The second decomposition, created by setting \( s_3 = 0 \), separates nodes 3 and 6 from the rest of the network. This decomposition connects the branch of segments 1 and 2 to the trunk.
and leaves the other branch unconnected to the trunk. The optimal solution
to this branching network problem is the decomposition for which the sum
of its two component single arterial subproblem solutions is minimized.

If junction node 4 had m branches coming into it, a feasible solution,
which could have no zonal overlap, would have to be a decomposition of the
network in which only one branch was connected to the trunk and the other
(m-1) branches were not. There are m such feasible decompositions, one
for each branch that may be connected to the trunk. Each of these decom-
positions consists of m single arterial problems, so that by solving m
single arterial problems m times, we could find the optimal local zonal
design for this problem of 1 trunk and m branches.

With a simple branching network such as Figure 7.2a, the above approach
is fine, since solving $m^2$ single arterial problems is not too costly as
long as m is small. This approach will work not just when all demands are
destined for the downtown or when the objective is just to minimize operator
cost, but for any objective that is separable by route (including both opera-
tor and passenger cost) and when passengers may alight at any stop down-
stream.

The above "brute force" approach can be improved upon, however, since
it involves a lot of repetitive computation. Instead of solving $m^2$ single
arterial problems, we need to solve only 2m: 1 for each branch when it is
connected to the trunk, and one for each branch when it is not connected to
the trunk. All m decompositions can be found as combinations of these 2m
single arterial problems. We can think of this procedure as a dynamic pro-
gram at a level higher than the single arterial problem: we find the
optimal design for each branch under the conditions connected to the trunk
and not connected, and then choose the combination of these optimal solutions to be the optimal solution for the trunk and branches together.

This dynamic programming approach can be extended to a network with many junction nodes, such as Figure 7.1. There are at least two extreme nodes upstream of each junction node, and we shall call the paths from these extreme nodes to the junction node the branches of that junction. To apply the dynamic programming approach to such a network, we would begin at the extreme nodes and progress toward downtown. Each time we reach a junction we decide which one of its branches should be connected to its trunk. Like the brute force approach, this approach can solve the local zonal design problem with a general demand pattern inbound and outbound, and a general objective. In designing service to cover both inbound and outbound demands, this approach requires one restriction, however: the branches and trunk of each junction node must be connected the same way for outbound service as they are for inbound service. This would prohibit, for example, segments 2 and 4 in Figure 7.2 from belonging to the inbound service zone of a route that begins on branch (1,2) while segments 3 and 4 belong to the outbound alighting zone of a route that begins at node 3. It seems that most operators would want this restriction to apply, in order to make the zonal structure more easily comprehensible to their patrons. Since the heavy direction demands are the primary factors in zonal service, lifting this restriction on the structure of the light direction service zones is unlikely to increase efficiency very much, any-

way.
Let us first define the following variables:

\( z^e_j \) = minimum cost of serving the inbound and outbound demands of path \((e,j)\) including segment \(j\), with routes that begin at nodes that lie on path \((e,j)\).

\( E_j \) = the set of extreme nodes upstream of segment \(j\)

\( U_{ej} \) = the set of segments immediately upstream of any node on path \((e,j)\), excluding segments belonging to path \((e,j)\).

\( V_j \) = minimum cost of serving segment \(j\) and all segments upstream of \(j\).

The following algorithm finds the optimal design of local zonal service in a branching network for inbound and outbound demands, under the same restrictions as the single arterial local zonal service design problem (Section 5.2, algorithm (A5.2)), with outbound alighting zones allowed to differ from inbound boarding zones, but requiring that the connections between branches and trunks at each junction node be the same for the inbound boarding zones as for the outbound alighting zones.

\( (A7.1) \)

1. Construct a reduced network by eliminating all nodes except extreme nodes, junction nodes, and the downtown terminal. Segment \(j\) is defined on this new network as the segment between node \(j\) and the next downstream node on the reduced network. Number all of the nodes in the network, beginning with the extreme nodes, so that if node \(i\) is upstream of node \(j\), node \(i\) will have a numerically lower ranking than node \(j\).

2. For each extreme node \(e\), treat the path on the original network from \(e\) to the downtown terminal as a single arterial and apply the single arterial local zonal service algorithm (A5.2). For each segment \(j\) of the reduced network, let \(i\) be the component segment
of $j$ in the original network that is the farthest downstream. Then as the intermediate solutions $Z_{jkg}$ of (A5.2) are found, set

$$Z_j^e = Z_{iij}$$

(7.2)

for every segment $j$ downstream of node $e$ (including segment $e$).

3. For each extreme node $e$, set

$$V_e^e = z_e^e$$

(7.3)

Let $j$ be the junction node with the least numerical ranking.

4. Set

$$V_j = \min \left\{ z_j^e + \sum_{k \in U_{ej}} V_k \right\}$$

(7.4)

5. Let $j = j + 1$. If node $j$ is the downtown terminal, STOP; otherwise go to 4.

Thus the optimal design for local zonal service with exclusive boarding/alighting zones for a branching network with $N_e$ extreme nodes can be found by making $N_e$ applications of the single arterial design algorithm, and then by updating $V_j$ according to equation (7.4). Solving equation (7.4) requires $O(N_e^2)$ computations, and in the worst case the number of junction nodes is $O(n^2)$. Therefore the solution to most practical problems can be found at very low cost on a computer since $N_e$ will rarely exceed 8 in realistic applications.

7.2 Express/Local Service on a Branching Network

Operating complimentary express and local services in a radial corridor can be the most efficient routing strategy, both when an expressway exists
in the corridor and when express buses must remain on local streets. Chapter 6 showed how express/local zonal services could be designed on a single arterial; we wish now to extend these results to a branching network.

Recall that in designing express/local services the basic issue is to segment the market into demands that will be served by express routes and demands that will be served by local routes. Given a market segmentation, each market's service can be designed independently using the local zonal service algorithms, both in a single arterial corridor and in a branching route corridor as well. Chapter 6 gave two ways of finding the optimal market segmentation. The first was to simply enumerate the feasible market segmentations. The second way used a dynamic programming approach to the entire problem, thus avoiding some repetitive computation and avoiding some segmentations that could be proven inferior. The dynamic programming approach was rather complicated, however, requiring 7 state variables due to the many possible paths along which one could route the passengers. For example, downtown passengers could board a local bus, then later transfer to an express; similarly, local passengers could first board an express and later transfer to a local bus.

The possibilities for interroute transfers increase in a branching network structure. Suppose a junction node has 3 branches. Downtown passengers originating along the first two segments may have to board local routes and then transfer at the junction node to an express route that began on the third branch. This kind of transferring is not considered in the single arterial express/local algorithm. Considering all the
possible combinations at each junction node would require enumerating \( m^{2m-1} \) state variables at each junction node, where \( m \) is the number of branches at a junction node, just for the state variables that account for downtown passengers transferring to an express route. (For example, if a junction node has 3 branches, each branch may have to pick up at the junction node the downtown demands of both of the other 2 branches, or of either one of them, or of neither of them.) To incorporate the more complex transfer patterns that can occur in a branching route network into the dynamic programming approach then would require updating an exponentially growing number of state variables. Enumerating all of the feasible market segmentations would require even more computation. Thus finding the optimal express/local zonal routing configuration on branching networks requires a great deal more computation than the single arterial problem except for the simplest branching networks.

There are two non-optimal approaches to this problem that seem promising, however. The first is to put a restriction on the freedom of branches and trunks to be connected at junction nodes. If we require that the branches and trunk at each junction node be connected in express service zones in the same say as local and outbound service zones, then we are back to simply finding the optimal connection at each junction node. We can solve this problem using algorithm (A7.2), a modification of algorithm (A7.1) that accounts for express service, which shall be presented below. Operators may like having this restriction imposed in order to make the service zones more easily understood by their patrons; however this restriction will perhaps be less generally acceptable than the similar restriction we made on light
direction service zones. A further drawback of algorithm (A7.2) is that while it requires that local service areas and express service areas along a set of connected segments be contiguous, it imposes no restriction on service area contiguity between unconnected segments.

To find the optimal express/local design under the restriction that at each junction node local service areas, express service areas, and light direction service areas must be connected in the same way, first define the following state variables:

\[ y^e_j = \text{minimum cost of serving local, downtown, and outbound demand of all segments on the path (e,j) including segment j with routes that begin along the path (e,j).} \]

\[ W_j = \text{minimum cost of serving all demands (local, downtown, outbound) of segment j and all segments upstream of j.} \]

Then execute the following algorithm.

(A7.2)

1. Construct a reduced network, following Step 1 of algorithm (A7.1).

2. For each extreme node e, treat the path on the original network from e to the downtown terminal as a single arterial and find its optimal express/local zonal design using algorithm (A6.1). If segment n is the segment farthest downstream, set \( n_{ZB_n}^{nn} = \infty \) (since we do not allow n to be served by an express route). For each segment j of the reduced network, let i be the component segment of j in the original network that is farthest downstream. Then as the intermediate solutions of (A6.1) are found, set

\[ y^e_j = \min\left(i_{ZB_i}, i_{ZN_i}\right) \]  

(7.5)
3. For each extreme node \( e \), set

\[ W_e = y_e^e \tag{7.6} \]

Let \( j \) be the junction node with the least numerical ranking.

4. Set

\[ W_j = \min_{e \in E_j} [y_j^e + \sum_{k \in U_{ej}} W_k] \tag{7.7} \]

5. Let \( j = j + 1 \). If node \( j \) is the downtown terminal, STOP; otherwise go to 4.

The second non-optimal approach to designing express/local zonal services on a branching network is to separate the problem into two parts, market segmentation and service design for each market, and to do the market segmentation manually. By market segmentation we mean deciding which segments should have local service, which express, and which both. One must also decide whether the outbound demands of each segment are to be served by a route which inbound is local or express. Given the service areas for each service type, the local market and express markets are known, and each service type may then be designed independently of the other. Both service types may be designed using algorithm (A7.1). This kind of design would call for an interactive computer program, whereby the planner could then try as many market segmentations as he wanted until he was satisfied with a "good" solution.

This semi-manual planning procedure is perhaps more useful than a procedure that is entirely automated because it allows the planner to enter
more information than the automated procedure. For example, the operator may be under a constraint of equity, that if he offers express service in one neighborhood he ought to offer it in another. He may have constraints of historical precedent, or may have non-quantifiable feelings about service in some areas. He may be unsure of the weights he wants to give to various service level components such as the transfer penalty or the crowding penalty, and may therefore want to look at the solutions for many different market segmentations rather than just be told the optimal segmentation. It is the market segmentation that determines the service characteristics that are most difficult to compare quantitatively, such as what service types are offered where, and how many transfers will take place and under what circumstances, so that it makes sense for market segmentation to be done by direct input from the planner. Given the market segmentation, the route design for each service type primarily determines operating cost and wait time, and these factors are much more easily traded off against each other in a consistent manner throughout the system. Therefore, it makes sense for the route design for each service type to be performed automatically, while the market segmentation is done manually.

How close will this semi-automated procedure come to an optimal solution? It is not possible to say precisely without knowing what the optimal solution is in any given situation, and it depends also on the skill of the planner. But in general I would expect the semi-automatic procedure to yield a solution that requires at most two vehicles more than the optimal solution. The planner, by using common sense, can eliminate a vast number of inferior alternatives that an automated procedure would evaluate. He is also able to identify most of the promising alternatives. And it is
the nature of this problem that there will usually be several "good" solutions that require an operating cost within two buses of the optimal solution. For example, in a system with a number of routes, there will usually be a number of different ways to serve the outbound demand at the same cost. A good planner will most likely include among his promising candidates at least one of these "good" solutions; for this reason I believe the semi-automated procedure to be a good heuristic. In a small network the number of feasible market segmentations may be small enough that a planner could evaluate them all; then he will know he has the optimal solution.

Nothing can substitute for the planner's common sense, experience, and intuition in choosing alternative market segmentations in executing this semi-automated procedure. As he gains experience with the model he will learn to find good solutions quickly. The planner who is just beginning to use this model, however, may find it helpful to use a system of service warrants. In such a system, if the downtown demand of a segment and its distance from downtown exceed a certain threshold, that segment warrants express service. Warrants can be developed based on demand intensity, distance from downtown, and run time savings of express over local service. The values of these warrants depend on vehicle capacities and the relative weights given to operator and passenger costs; with a little bit of experimentation, appropriate values for a given situation can be found.

This section has shown the complexity of jointly designing express and local zonal services in a corridor served with a branching route network when the structure of the express, local, and outbound service areas are independent. If one restricts these service areas to be connected
in the same way at each junction, the optimal solution can still be found by dynamic programming. To find the best design without this restriction on service area structure, the suggested method is a semi-automated heuristic, wherein the planner chooses the segments that are to be served by express routes and then an optimal configuration for this market segmentation is found automatically using a dynamic programming algorithm that takes as input the solutions of the single arterial design algorithm, which is itself another dynamic program.

7.3 Branching Route Case Study

7.3.1 Description of the Corridor

The case study for this chapter is route design in the Chicago Avenue corridor of the Minneapolis - St. Paul metropolitan area. This corridor has a length of 4.3 miles, extending southward from downtown Minneapolis to the suburbs of Richfield and Bloomington. Illustrated in Figure 7.4, it has a main trunk 5 miles long in the city of Minneapolis (Chicago Avenue), and then has three branches: a one mile branch at the southern edge of the city terminating at Front Street and 15th Avenue, and two long, parallel branches, roughly 1/2 mile apart, that extend southward through the suburbs for about 6 miles. The Portland Avenue branch terminates at 104th Street and 3rd Avenue in Bloomington; the 12th Avenue Branch terminates at Old Shakopee Road and 10th Avenue in Bloomington. Peak hour demand on the trunk is quite high, with existing service providing an average headway of under 5 minutes, while demand on the branches is quite low. Peak hour flows in the light direction are also low relative to the heavy direction.
Figure 7.4
The Chicago Avenue Corridor

- local street
- expressway
- expressway interchange
- bus terminal

MINNEAPOLIS

BLOOMINGTON

RICHFIELD

1 mile
Figure 7.5 illustrates the existing routes operated by the Metropolitan Transit Commission (MTC) that serve the Chicago Avenue corridor. Local service is provided through a group of branching routes that belong to the Route 5 system. The different branching routes serving the Chicago Avenue corridor that make up the Route 5 system, in order of increasing length, are 5B, 5C, 5G, 5H, and 5D. These routes operate as overlapping routes (except for one run of route 5H in the a.m. peak and another in the p.m. peak that operates with exclusive boarding/alighting zones). The Route 5 system includes branches north and south of downtown, with buses from the southern branches being routed through the downtown and continuing north out of the city to 5 different terminals in the northwestern suburbs.

The MTC also provides express service in the corridor during the peak hours with three routes. These express routes use the expressway I-35W which parallels Chicago Avenue and empties into the heart of downtown. Route 35G covers the Bloomington part of the corridor; Route 35E covers the part of the corridor that is in Richfield, as well as a bit in southern Minneapolis. Route 35D covers a little more of the southern Minneapolis part of the corridor. Altogether 2.2 miles of the 5.6 miles of the corridor that lie within the city of Minneapolis have express service, as well as local service; the remaining part of the corridor within Minneapolis has only local service.

7.3.2 Analysis Framework

The case study will be concerned with route design in the morning peak. Design in the off-peak hours is not as complex since there is no express service then and branch demands are so low that hourly frequency is provided
Figure 7.5
Existing Routes in the Chicago Avenue Corridor

= express portion of a route
to most branches. It is the peak periods, with their high demands, that strain the transit system's ability to provide a sufficient number of vehicles, and that have the highest marginal costs. The objective taken in this case study is to find the routing configuration that minimizes the number of vehicles needed to adequately serve the demands that are now being served by 8 existing routes in the Chicago Avenue corridor during the morning peak period.

Travel demand in Minneapolis has a rather sharp peaking pattern in the morning. The prime arrival time downtown is the interval 7:40 – 8:10 a.m. The downtown arrival rate is also high from 7:10 to 7:40 a.m., and it remains high until 8:30. Since it is these peak hour demands that determine the vehicle requirements of the route system, this case study used demand rates that were averages of demands that are served by runs arriving and leaving downtown during the interval 7:10 – 8:30 a.m.

In order to reduce its complexity, the case study separates the southern half of the Route 5 system from the northern half. With this simplification we can simply try to minimize the number of vehicles needed to operate the southern half of the Route 5 system along with the express routes in our corridor, leaving the design of the northern half of the Route 5 system outside the study. It is still assumed, however, that buses will be interlined between the northern and southern halves of the system as they now are. We assume, however, that the interlining pattern will be modified to best match north and south runs, minimizing unnecessary layover and unused capacity. Because there are several north side branches, there is considerable flexibility in matching north and south runs, so
the assumption that the degree of efficiency that now exists in interlining north and south runs will not significantly change in a modified design does not seem implausible.

Since the local buses will interline, it is not necessary that every separate route have an integer number of vehicles. We must have some slack in the schedule, however: some to allow late arrivals the recovery time they need to get back on schedule, and some because of the degree of inefficiency in the interlining process. In estimating the degree of slack in the existing system, it is not possible to separate these two effects. Neither is it necessary, however, to separate these effects, since our case study will simply require the same degree of slack that exists in the present system, both for recovery time and for schedule inefficiency, in any new configuration we design.

7.3.3 Service Policy Constraints

The existing degree of slack in the schedule was estimated by examining the printed schedules and headway sheets of the routes as they were in the winter of 1979. The routes were aggregated into 3 categories: short-distance, local routes, long-distance local routes, and express routes. It was found that both categories of local routes had slack times that were on the average 12% of run time, while the express routes had an average slack of 20% run time. These figures were used for average slack in the new designs.

The MTC has no stated policy on minimum frequency of local service. However, local demands on all the city and suburban branches could be served with frequencies of less than one run per hour, while existing frequencies
are considerably higher. Therefore, it seems reasonable to interpret the existing frequencies as reflecting an implicit policy headway. The following policy headways for local service were guessed based on the existing schedule: suburban branches get two runs per hour in the peak period peak direction, and one run per hour otherwise; city branches get four runs runs per hour in the peak period peak direction and two runs per hour otherwise. No frequency standards were imposed on express service; however there is an impetus to keep the express service frequency competitive with the parallel local service frequency in order to keep express passengers from switching to local routes.

The buses operated by the MTC have 50 seats, and their policy headway forbids average peak loads on local routes exceeding 70. Design loads are not 70, however, to allow slack for both missed runs and demand growth. The case study made the maximum load on local buses 57.5 passengers per bus during the peak hour for a design load factor of 1.15. Express buses are not supposed to have standees; their design load was set to 45, for a design load factor of 0.9.

7.3.4 Data Sources

The source of the demand data for passengers currently served by local routes was a number of point counts taken at various points in the system as a part of a Service and Methods Demonstration Project of the U.S. Department of Transportation. These point counts were taken at the peak load point as well as at the northern end of each branch, so that demands for each corridor could be estimated. Counts were taken on 8 weekdays in
winter of 1979, and averages of these 8 days were used.

No demand data was available for the express routes, but as each express route has only a few runs each peak hour, and since the runs are scheduled to achieve peak loads near the bus capacity of 50, it was assumed that each express run carried a peak load of 45.

Run times were taken from headway sheets and printed schedules. When a new route was designed, run times were estimated based on the difference between the new route and the most similar existing route. Based on existing run times, speeds were estimated to be 40 mph on the expressway inbound and 45 mph outbound, 13.7 mph on local streets in Minneapolis inbound and 15.9 mph outbound, and 19.8 mph on suburban streets inbound and 22.5 mph outbound. Vehicles deadheading outbound were assumed to travel 17 mph on city streets, 25 mph on suburban streets. The approach between the end of the expressway and the downtown terminal for express and deadheading buses was estimated to take, in either direction, 8.5 minutes for vehicles in service and 7.5 minutes for vehicles deadheading.

7.3.5 Number of Buses Needed for Existing Service

One cannot ascertain the number of buses needed for the existing services in the Chicago Avenue corridor from the number of buses actually used because the buses used in that corridor are interlined with routes in the northern half of the Route 5 system. Therefore, the number of buses needed was computed in accordance with the analytical framework that will be used for estimating vehicle requirement for proposed systems. The number of buses needed on each route is then the run time, expanded to include the proper amount of slack time, divided by the average headway during the period of
greatest demand (arriving at downtown from 7:10 to 8:30 a.m.). As reported in Section 7.3.3, slack time was taken as 12% of run time on local routes, 20% of run time on express routes.

Table 7.1 gives the run times, headways, and vehicular requirements for each of the routes currently serving the Chicago Avenue corridor. To serve the 5 local routes, 18.8 buses are needed; to serve the 3 express routes, an additional 8.7 buses are needed, for a total vehicular requirement of 27.5 buses.

7.3.6 Analysis of Demand

Before designing or analyzing any new configuration, it is necessary to be able to estimate the peak loads of any new route. One sufficient means is to have an estimated stop-by-stop origin-destination matrix of transit trips in the corridor. This was the approach taken in the Boston case study. However, demand data in Minneapolis was not available at the stop level. The primary data source was a set of point counts at the end of each branching segment and at two locations on the trunk. Since these point counts are the averages of counts made on eight different days, the relative magnitude of the demand rates should be quite accurate. The point counts were supplemented by an on-board survey in which passengers were asked, among other things, their origin and destination. The results of this survey were used to distribute the destinations of these branch and trunk demands among the various segments of the trunk. A few minor assumptions were also made in order to disaggregate the data to the desired level. These assumptions were as follows:
<table>
<thead>
<tr>
<th>Route</th>
<th>Run Time In (min)</th>
<th>Run Time Out (min)</th>
<th>Expanded Run Time (Slack Included) (min)</th>
<th>Average Headway (min)</th>
<th>Vehicles Required</th>
</tr>
</thead>
<tbody>
<tr>
<td>5D</td>
<td>52.0</td>
<td>46.0</td>
<td>109.8</td>
<td>24.3</td>
<td>4.3</td>
</tr>
<tr>
<td>5H</td>
<td>50.0</td>
<td>44.0</td>
<td>105.3</td>
<td>24.7</td>
<td>4.3</td>
</tr>
<tr>
<td>5G</td>
<td>37.0</td>
<td>34.0</td>
<td>79.5</td>
<td>12.9</td>
<td>6.2</td>
</tr>
<tr>
<td>5C</td>
<td>36.0</td>
<td>33.0</td>
<td>77.2</td>
<td>45.0</td>
<td>1.7</td>
</tr>
<tr>
<td>5B</td>
<td>32.0</td>
<td>29.0</td>
<td>68.3</td>
<td>30.0</td>
<td>2.3</td>
</tr>
<tr>
<td>35G</td>
<td>36.0</td>
<td>25.5</td>
<td>73.8</td>
<td>25.3</td>
<td>2.9</td>
</tr>
<tr>
<td>35E</td>
<td>29.0</td>
<td>19.5</td>
<td>58.2</td>
<td>16.0</td>
<td>3.6</td>
</tr>
<tr>
<td>35D</td>
<td>25.5</td>
<td>19.4</td>
<td>53.9</td>
<td>24.7</td>
<td><strong>2.2</strong></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>27.5</strong></td>
</tr>
</tbody>
</table>
1) The relative distribution of demands among destinations was the same for the two branch segments that did not have express service operating on them.

2) The relative distribution of demands among destinations was the same for the three branch segments that had express service operating on them as well as local service.

3) The differences in demand between the two long suburban branches were attributed entirely to there being a small demand from the Portland Avenue branch to downtown, since both branches have very similar markets, both in size and in population characteristics, but Portland Avenue has express service available while 12th Avenue does not.

Table 7.2 presents the estimated zonal origin-destination matrix for inbound peak hour flows in the Chicago Avenue corridor. From this matrix the peak loads of any newly designed local routes can be computed.

In order to ensure an adequate service level in the light direction, it is important to know the demand in the outbound direction as well. Demands destined to the branches are so tiny (less than 10 per hour on each branch) that service frequency on the branches will be dictated by policy headway and by the need to recycle buses and not by bus loading constraints. The outbound peak load on the trunk, however, is significant, 186.2 passengers per hour. Most of this outbound demand does not originate in the Chicago Avenue corridor but in areas north of downtown and is destined for a commercial area that includes several hospitals just south of downtown. Maintaining a relatively high outbound frequency is important to providing direct service between the northern branches of the Route 5 system and this commercial area.

The demands carried by existing express routes was estimated by simply assuming that each express run carries 45 passengers on the average. Since the boarding zones of the existing express routes do not coincide with
Table 7.2

Origin-Destination Matrix for Trips Served by Local Routes in the Chicago Avenue Corridor

Zone 1: 104th and 3rd Ave. to 68th and Portland
Zone 2: Old Shakopee and 10th Ave. to 66th just east of Portland
Zone 3: 66th and Portland to 62nd and Portland
Zone 4: 60th and Portland to 58th and Chicago
Zone 5: Front and 15th Ave. to 57th and Chicago
Zone 6: 56th and Chicago to 39th and Chicago
Zone 7: 38th and Chicago to just before Lake and Chicago
Zone 8: Lake and Chicago to just before 8th and Chicago
Zone 9: Points north of 8th and Chicago

Interzonal Passenger Movements

(peak hour passengers/hr)

<table>
<thead>
<tr>
<th>From</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.9</td>
<td>1.9</td>
<td>9.2</td>
<td>5.4</td>
<td>19.4</td>
</tr>
<tr>
<td>2</td>
<td>2.9</td>
<td>1.9</td>
<td>9.2</td>
<td>29.7</td>
<td>43.7</td>
</tr>
<tr>
<td>3</td>
<td>0.3</td>
<td>0.2</td>
<td>0.9</td>
<td>0.5</td>
<td>1.9</td>
</tr>
<tr>
<td>4</td>
<td>0.4</td>
<td>0.6</td>
<td>1.9</td>
<td>1.1</td>
<td>4.0</td>
</tr>
<tr>
<td>5</td>
<td>1.3</td>
<td>0.9</td>
<td>4.2</td>
<td>13.5</td>
<td>19.9</td>
</tr>
<tr>
<td>6</td>
<td>6.7</td>
<td>10.1</td>
<td>90.2</td>
<td>242.1</td>
<td>349.1</td>
</tr>
<tr>
<td>7</td>
<td>-</td>
<td>3.4</td>
<td>25.8</td>
<td>126.0</td>
<td>155.2</td>
</tr>
<tr>
<td>8</td>
<td>-</td>
<td>-</td>
<td>58.1</td>
<td>157.2</td>
<td>215.3</td>
</tr>
<tr>
<td>Total</td>
<td>14.5</td>
<td>19.0</td>
<td>199.5</td>
<td>575.5</td>
<td></td>
</tr>
</tbody>
</table>
the zones of the local route system, it was necessary to disaggregate
the demands of two of the express routes by origin segment. For this
purpose, it was assumed that production rates (passengers/hour per mile)
were uniform along the length of each route, except for Route 35E which
has a portion in Richfield and a portion in Minneapolis. For this route
it was assumed that the production rate in Richfield was 10% higher than
the production rate in Bloomington. The resulting demand rates for each
segment are shown in Table 7.3.

7.3.7 Minimum Number of Buses Needed with Reduced Frequencies with the
Existing Configuration

In order to make a fair evaluation of the benefits of a new configura-
tion, it is important to compare any new design against the existing
configuration with its frequencies adjusted to their optimal values. As
currently operated by the MTC, the Route 5 system has considerably more
capacity than needed. At the assumed loading standard of 57.5 passengers
per bus, the Route 5 system provides room for 741 passengers per hour
during the morning peak, while the peak point flow is 576 passengers per
hour. Thus the average load at the peak point is 45 passengers per
bus.

Table 7.4 shows the operating design of each route in the existing
configuration when frequencies are adjusted to minimize the number of buses
required. By assumption, the express routes are already operating at
minimum frequency, so the reduction in vehicle requirements comes entirely
from the local routes. By raising the average peak load per vehicle to
57.5 passengers, 4.0 buses are saved, making the total vehicular requirement
Table 7.3
Origin-Destination Matrix for Trips Served by Express Routes in the Chicago Avenue Corridor

Zone Definitions
Zone 11: 104th and Portland to 79th and Portland
Zone 12: 77th and Portland to 64th and Portland
Zone 13: 61st and Portland to 59th and Portland
Zone 14: 57th and Portland/Chicago to I-35W near 55th
Zone 15: 53rd and Chicago to 44th and Chicago
Zone 16: 42nd and Cedar to 52nd St. and 12th Ave.

<table>
<thead>
<tr>
<th>Origin Zone</th>
<th>Pax/hr Destined for Downtown</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>107</td>
</tr>
<tr>
<td>12</td>
<td>68</td>
</tr>
<tr>
<td>13</td>
<td>40</td>
</tr>
<tr>
<td>14</td>
<td>60</td>
</tr>
<tr>
<td>15</td>
<td>52</td>
</tr>
<tr>
<td>16</td>
<td>58</td>
</tr>
<tr>
<td>Route</td>
<td>Design Peak Load (pax/hr)</td>
</tr>
<tr>
<td>-------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>5D</td>
<td>115</td>
</tr>
<tr>
<td>5H</td>
<td>115</td>
</tr>
<tr>
<td>5G</td>
<td>230</td>
</tr>
<tr>
<td>5C</td>
<td>(does not operate)</td>
</tr>
<tr>
<td>5B</td>
<td>115</td>
</tr>
<tr>
<td>35G</td>
<td>107</td>
</tr>
<tr>
<td>35E</td>
<td>168</td>
</tr>
<tr>
<td>35D</td>
<td>110</td>
</tr>
<tr>
<td>Total</td>
<td></td>
</tr>
</tbody>
</table>
23.5 buses. Average headway on the trunk becomes 6 minutes, compared to
4.7 minutes as service currently exists.

Under this design, the branches are all served at their policy headways
(30 minutes in the suburbs, 15 minutes in the city). The trunk is served
by 10 buses per hour, but they cannot be spaced at regular 6 minute inter-
vals without making the branch headways quite uneven. Average wait time
on the trunk, then, accounting for schedule irregularity and lack of sche-
dule adherence, will probably be around 4 to 4.5 minutes, compared with
approximately 3 minutes in the existing system.

7.3.8 Proposed Routing Configuration

Now we come to the task of the case study: finding another routing
configuration by which the Chicago Avenue corridor can be served with
fewer buses while still meeting the service constraints. This problem
is one of route design in a branching network with both local and
express services. Section 7.2 pointed out that in designing both express
and local services on a branching network it is best to decide manually
which road segments will have express service, which will have local
service, and which will have both. Another consideration in the design
procedure for the Chicago Avenue corridor is that operating the local routes
with overlapping boarding/alighting zones, as in the existing system, seems
preferable to operating the local routes with exclusive boarding/alighting
zones. The reason that overlapping routes appears to be the superior
strategy is that branch demands are so low. If the three major branches
operated with exclusive boarding/alighting zones at their minimum frequen-
cies, these branches would need 11.7 buses to serve them, and which would
provide a capacity of 480 places per hour, but carry only 50 passengers
per hour past the corridor peak load point. Not one of the branches has
a demand great enough to warrant an exclusive boarding zone, even at
30 minute average headways. (In the existing system, one run of Route 5H
is operated with an exclusive boarding zone in the morning peak. A single
run like this could be included in the detailed design of a new system,
when the individual runs are scheduled, but at the level of detail appro-
priate to the route design problem exceptions like this one are ignored.)
Low branch demands alleviate much of the operational difficulties of an
overlapping system that are discussed in Section 5.3. Since the corridor
seems to exhibit few or no problems with its current system of overlapping
routes, it seems reasonable to go ahead with overlapping routes in the
proposed system. Overlapping zonal routes are best designed manually
rather than automatically. Thus the design of the proposed system will
be performed manually, not using any of the algorithms presented in this
thesis as such, but relying on the insights and some of the analytical
results found in the previous chapters.

The results of the previous chapters have revealed five principles
for efficient routing design in a high demand corridor:

1) Eliminate empty seat miles by using a zonal route system for both
local and express service.

2) Reduce turnaround time by deadheading in the light direction.
Where possible, complete deadheading is preferable to partial
deadheading.

3) Offer express service wherever it requires fewer buses to serve
the same demand, provided it will still leave the local service
market big enough to offer an adequate level of service.
4) Where both express and local service are being offered, the express service should be zoned as much as possible. The local service in such a case will often not be served very efficiently by zonal service with exclusive boarding/alighting zones, so consider offering conventional local service or zonal service with overlapping zones.

5) When a road segment's demand is so low that it doesn't warrant both local and express service, consider offering only one type of service and making some passengers transfer.

Let us apply these principles to the Chicago Avenue corridor. Zonal routing systems are already in effect there: there are three express routes which have successively more distant service areas. There are five local routes, all of different lengths, which have service zones that overlap along the trunk. Our new design will also have zonal express routes and overlapping zonal local routes. In applying the first principle of corridor route design, we might like to establish additional intermediate terminals at which short routes could begin, thereby reducing the amount of capacity a longer route must provide. Establishing an intermediate terminal for local service along the trunk between the 56th Street terminal and downtown seems to be a good idea. A good site is at 38th Street and Chicago Avenue. It is readily accessible to I-35W, making it efficient for deadheading. It is about 10 minutes running time north of the 56th Street terminal, so that locating an intermediate terminal south of 38th Street would so reduce the run time savings as to make the potential vehicle savings very small. Locating an intermediate terminal north of 38th Street would not be fruitful, either, since the corridor load a mile north of 38th Street is almost as high as the peak load. By replacing some runs from 56th Street or 60th Street with runs from 38th Street some vehicles could be saved while still providing sufficient capacity both south and north of 38th Street.
The first principle also suggests to us that the branch routes operate at minimum frequency. We will follow this suggestion in the proposed design. No additional intermediate terminals seem viable for express service because the demand on each express route is quite low. The present zoning system of 3 routes should at least be maintained, however.

The principle of deadheading has potential in the corridor because of the existence of I-35W and branching expressways. Some of the local service runs could be deadheaded without straining the outbound service level. Deadheading of express routes is understood. We will apply deadheading in both the local and express route design as they develop.

In applying the third principle, one can see that express service is already offered on the entire Portland Avenue branch and on the trunk as far north as 46th Street. The low demand levels for downtown trips on the other two branches (30 passengers/hour and 14 passengers/hour) preclude expanding express service there. It seems worthwhile to extend the express service area northward along the trunk, however, to the next expressway entrance at 35th Street. Using the expressway to get downtown from 35th Street cuts travel time by 7 minutes. The extra 11 blocks that would be covered can be expected to generate about 64 express passengers, based on production rates estimated from the demand for Route 35D. Serving these passengers on an express route rather than on a local route would yield a tiny savings in buses, since 64 people isn't very many and since the 7 minute savings in run time is offset by the lower load factor and higher slack time factor of express routes. However, express passengers
pay a premium fare, making them more profitable to the MTC; those using
the express route will enjoy a better level of service; and the additional
64 passengers might make it possible to subdivide the express service
into four zonal routes instead of three, yielding a greater overall efficiency.

The fourth principle tells us to zone the express service as much
as possible and to consider overlapping zonal routes for the local service.
It has already been mentioned how a branching system lends itself to an
overlapping zonal system of routes, and the low branch demands and the
history of the Route 5 system suggest that overlapping zonal service will
work well in the Chicago Avenue Corridor. We will therefore use an over-
lapping zonal system for local service.

To increase the extent to which the express service is zoned, we
should take advantage of the 64 additional express passengers per hour
we can divert from local service to add a fourth zone to the express ser-
vice route structure. By extending the express service area up to 38th
Street, Route 35D can be realigned. First of all, Route 35D operates for
1.25 miles along Cedar Avenue, which either borders or is within one
block of a large park for that entire 1.25 miles. Moving that portion of
the route a quarter of a mile west to Bloomington Avenue would enhance
passenger accessibility, as well as reduce the round trip distance by
half a mile (for a run time savings of about 2.5 minutes). Then with the
express service area reaching to 35th Street, we could serve the express
demands originating on Chicago Avenue south of 44th Street with a new
express route and let Route 35D serve Chicago Avenue between 42nd Street
and 35th Street only. This modified express route, which we will call
E4, will then begin at 52nd Street and Bloomington Avenue, go north on Bloomington to 42nd Street, go west on 42nd to Chicago, north on Chicago to 35th, and west on 35th to I-35W and downtown. This alignment requires 2.6 minutes less run time inbound than Route 35D, and its uptown terminal can be reached from downtown by a deadheading bus just as quickly as Route 35D's terminal. Route E4 has a slightly higher demand than Route 35D does, so that it will offer lower wait times as well as better walk accessibility.

A new express route, Route E3, will have as its boarding zone the Chicago Avenue segment between 56th Street and 46th Street. Thus it will serve all of the Chicago Avenue demands now served by Route 35D. In addition, it will take some demand away from Route 35E, which we will also modify. The modified Route 35E, which we call Route E2, will now access I-35W near 60th Street, instead of near 55th Street. Before we establish the southern boundary of Route E2's boarding zone, and the boarding zone of Route E1, let us first go on to the fifth principle.

The fifth principle is to consider offering only express or only local service on a segment of low demand and have some passengers transfer. In the current system, the 12th Avenue branch has only local service, so that passengers who want to take an express bus must either transfer from Route 5H to route 35E, or walk to Portland Avenue (which is only 1/2 mile from 12th Avenue) where express service is offered. We will keep offering local service only on the 12th Avenue corridor because of the small downtown demand it exhibits and because of the proximity of the Portland Avenue express service.
The existing configuration offers both local and express service along the Portland Avenue corridor. This is a questionable strategy, since Portland Avenue south of 66th Street generates only 19 passengers per hour for local route 5D, and we estimate that about 5 of them are going to downtown and could be served by express buses. Providing two buses an hour, with capacity of 115 passengers per hour, for only 14 passengers is quite an inefficient use of resources. Replacing those local runs with runs beginning at 60th Street (since their capacity is needed on the trunk) would save 35 minutes of round trip run time per run; expanding this to include the 12% slack, we find that eliminating local service on Portland Avenue south of 66th Street should save about 1.3 bus-hours of operation per hour.

The few local passengers that originate on Portland Avenue south of 66th Street could be picked up by the express buses and then made to transfer at 66th Street to the local route serving the 12th Avenue branch, Route 5H. Because 5H will be operating at the policy headway of 30 minutes, providing these local passengers with a timed transfer at 66th and Portland would make their transfer much less unpleasant. Having a timed transfer between an express route and a local route requires, however, that the frequency of one route be a multiple of the other route's frequency. The existing frequency on the southern end of Portland Avenue is 2.4 buses per hour; in our proposed system we suggest reducing this frequency to 2, equal to the modified frequency of Route 5H. In order to have this lower frequency, however, this route serving the southern part of Portland Avenue, which we shall call Route EI, needs to have a service area that will generate only 90 express passengers per hour.
so that its average load per bus on its express segment will not exceed 45. (We will allow the load on the express bus to go as high as 57.5 south of 66th Street, as long as it is 45 or less when the bus accesses the expressway. Thus, the few local passengers that there are can be carried with no extra capacity provided). The boarding zone of Route E1 should therefore extend from 104th Street and 3rd Avenue to about 86th Street and Portland Avenue. Route E1 buses would then continue north on Portland after 86th Street, dropping off local passengers at 66th Street to transfer, and accessing the expressway near 62nd Street. The run time of Route E1 is actually a bit less than the run time of Route 35G, for although it accesses the expressway system two miles farther north, its route length, due to the shape of the expressway system, is 2.4 miles less.

The final express route in the new design, Route E2, must begin around 83rd Street and Portland, and go north to 60th and Portland, from where it goes west to I-35. Its service area extends from 83rd Street to 60th Street. The demand level for this route requires a frequency of 2.8 runs per hour, meaning that timed transfers at 66th Street could not occur every 30 minutes. However, local passengers who would have to use this route (those originating on Portland Avenue between 83rd and 68th Streets) could transfer at 60th Street, which is in Minneapolis, where the service frequency will be at least 4 buses per hour. Average wait time for these transferring passengers at 60th Street, with careful scheduling, could be kept under 3 minutes.
Having decided what service will be provided on each branch, we are now able to finalize the design of the local routes. Route L1, which is the same as Route 5H, will have the minimum frequency of two runs per hour. In order to have a 15 minute headway at 60th Street and Portland Avenue, a local route must start at the 60th Street terminal with a frequency of two runs per hour; this is Route L2. Route L3, the same as Route 5G, must have 15 minute headways. Routes L1, L2, and L3 combined provide a capacity of 460 places per hour on the trunk. The load just before 38th Street, when we allow for passengers originating north of 46th Street who switched to express route E4, is about 430 passengers/hour, so there is no need to start any more runs south of 38th Street. The corridor peak load is 512 passengers/hour, so that a route that began at 38th and Chicago would have to provide a capacity of 52 passengers per hour. Serving 52 additional passengers warrants a frequency of only about one run per hour. However, it is not a good operating strategy to have a low frequency short route that is overlapped by longer routes, since we can't expect a large proportion of passengers who are boarding in that shorter route's boarding zone to use the shorter route since it only comes every 60 minutes. If a route were operated from 38th Street every 60 minutes, it would probably only be half filled, while the other routes would be overcrowded. Therefore, it seems better to let the extra peak load capacity that is needed be provided by a route beginning at 56th and Chicago, since the demands upstream of 56th Street are very low and therefore the 56th Street route's buses should fill almost as easily as the buses originating on the branches. Route L4 therefore will begin at 56th and Chicago, and will need to have a frequency of 0.9 runs per hour. Figure 7.6 illustrates the proposed routing configuration.
Meeting the policy headway constraints for outbound service requires that one of the L1 runs, one of the L2 runs, and two of L3 runs per hour return in service. Although the peak load outbound could be carried with only 3.3 runs per hour, in order to provide adequate service for passengers from the northern branches of Route 5 who are destined for the commercial area at the northern end of Chicago Avenue, we should probably insist on at least twice that frequency for outbound direction service. The cheapest way to provide this level of outbound service is to offer one L1, one L2, four L3, and the 0.9 L4 runs per hour in service, and deadhead one L1 and one L2 run per hour. Similarly one E1 run per hour must return in service, covering Portland Avenue from 60th Street to 104th Street, while the remaining express runs deadhead. A timed transfer would occur at 60th Street for outbound passengers who must use Route E1 who originated south of the downtown.

Table 7.5 gives the design parameters of the four express routes in the proposed system, including peak load, run time, and number of vehicles needed. Inbound headways are 30 minutes on Route E1, 21.6 minutes on Route E2, 24.1 minutes on Route E3, and 22.1 minutes on Route E4. Average headway per express passenger is 24.0 minutes in this new design, compared with 21.1 minutes in the existing system. While the new design requires 7.4 buses for the express routes, 0.7 more than the existing system, it carries 17% more express passengers and provides local service on a 6 mile suburban branch.

The design parameters of the 4 local routes of the proposed configuration are displayed in Table 7.6. Compared to the existing configuration with headways optimized (Table 7.4), inbound headways on the branches are the same, and average inbound headway on the trunk is 0.6 minutes greater (6.6 minutes versus 6.0 minutes). Outbound headways are substantially
### Table 7.5

Express Routes in the Proposed Configurations

<table>
<thead>
<tr>
<th>Route</th>
<th>Boarding Zone</th>
<th>Peak Load (pax/hr)</th>
<th>Average Headway (min)</th>
<th>Run Time Inbound (min)</th>
<th>Run Time Outbound (min)</th>
<th>Buses Needed</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1</td>
<td>104th and 3rd Ave. to 86th and Portland</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>a. return deadheading:</td>
<td>45</td>
<td>60</td>
<td>35.2</td>
<td>25.5</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>b. return in service:</td>
<td>45</td>
<td>60</td>
<td>35.2</td>
<td>30.4</td>
<td>1.3</td>
</tr>
<tr>
<td>E2</td>
<td>83rd and Portland to 60th and I-35W</td>
<td>125</td>
<td>21.6</td>
<td>28.5</td>
<td>21.8</td>
<td>2.8</td>
</tr>
<tr>
<td>E3</td>
<td>56th and Chicago to 46th at I-35W</td>
<td>112</td>
<td>24.1</td>
<td>18.9</td>
<td>16.9</td>
<td>1.8</td>
</tr>
<tr>
<td>E4</td>
<td>52nd and Bloomington to E35th at I-35W</td>
<td>122</td>
<td>22.1</td>
<td>23.5</td>
<td>19.4</td>
<td>2.3</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>9.4</strong></td>
</tr>
</tbody>
</table>

1/ Local passengers originating south of 66th and Portland transfer at 66th and Portland to a local route. A timed transfer will occur every 30 minutes.

2/ Local passengers originating south of 66th and Portland transfer at 60th and Portland to a local route.
Table 7.6
Overlapping Local Routes in the Proposed Configuration

<table>
<thead>
<tr>
<th>Route</th>
<th>Uptown Terminal</th>
<th>Design Peak Load (pax/hr)</th>
<th>Average Headway (min)</th>
<th>Run Time Inbound (min)</th>
<th>Run Time Outbound (min)</th>
<th>Buses Needed</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td>Old Shakopee and 10th Avenue</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>a. return in service 57.5</td>
<td>60.0</td>
<td>52.0</td>
<td>46.0</td>
<td>1.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>b. return deadheading 57.5</td>
<td>60.0</td>
<td>52.0</td>
<td>25.8</td>
<td>1.5</td>
<td></td>
</tr>
<tr>
<td>L2</td>
<td>60th and Portland</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>a. return in service 57.5</td>
<td>60.0</td>
<td>36.0</td>
<td>33.0</td>
<td>1.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>b. return deadheading 57.5</td>
<td>60.0</td>
<td>36.0</td>
<td>16.2</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>L3</td>
<td>Front and 15th Ave.</td>
<td>230</td>
<td>15.0</td>
<td>37.0</td>
<td>34.0</td>
<td>5.3</td>
</tr>
<tr>
<td>L4</td>
<td>56th and Chicago</td>
<td>52</td>
<td>66.3</td>
<td>32.0</td>
<td>29.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Total | 11.9
increased on the branches, while on the trunk outbound headway increases from 6.0 minutes to 8.6 minutes. The number of vehicles needed for the local routes is 11.9, compared with 14.8 under the existing configuration. Part of the savings is due to deadheading two runs per hour, part is due to diverting 64 passengers per hour to express service, and part is due to eliminating the costly runs along the Portland Avenue branch.

The vehicular requirements and service level of the existing system, the existing configuration with optimal headways and the proposed configuration are summarized below. The existing system requires 27.5 vehicles, while rationalizing the headways reduces the requirements to 23.5 vehicles. In rationalizing the headways, average headway on the suburban branches increases from about 24.5 minutes to 30 minutes, affecting about 65 inbound and 7 outbound passengers per hour; average headway on the city branches increases from about 12.5 minutes to 15 minutes, affecting 24 inbound and 4 outbound passengers per hour; and average headway on the trunk increases from 4.7 minutes to 6 minutes. The total increase in wait time resulting from rationalizing headways is about 32 hours, while operator cost savings, at $30 per bus-hour, are $120. As long as passenger wait time is valued at less than $5.45, then rationalizing headways as suggested is a good idea.

The proposed configuration requires 21.3 vehicles. Its impacts on passenger level of service, compared with the service level under the existing configuration with the headways rationalized, are varied. The new configuration denies direct local service to 19 suburban passengers per hour who now have it. We expect that 14 of them will make a transfer, but proper timing of transfer can keep their average transfer time to about 1.5 minutes. We expect that the other 5 will prefer to
pay a 10c premium and use an express route, which will give them a considerable time savings. Average headway for inbound passengers in the suburbs who presently use local routes will be unchanged except for an 8.2 minute reduction in headway to about 10 passengers per hour on the Portland Avenue corridor who will use Route E2. Outbound headway in the suburbs will increase from 30 minutes to 60 minutes, affecting 7 passengers per hour. Headways on the city branches will be unchanged inbound, and increased from 15 minutes to 30 minutes outbound, affecting 4 outbound passengers per hour. Average inbound headway on the trunk rises from 6.0 to 6.6 minutes, affecting 655 passengers per hour, and average outbound headway rises from 6.0 to 8.7 minutes, affecting 185 passengers per hour. On the express routes, average headway rises from 21.1 minutes to 24.0 minutes, affecting 385 passengers per hour. The total increase in wait time the new configuration would cause is about 16 hours, including the wait time of transferring passengers.

Because the new configuration serves more passengers on express routes and because it makes some of the express routes quicker, it yields an in-vehicle time savings of about 12 hours.

Walk access to local routes is unchanged in the new configuration. Walk access to express routes is somewhat improved in the area around 42nd Street and Bloomington Avenue, while access is a little worse between 56th Street and 60 Street. The overall impact is probably positive but small.

MTC revenue should rise by about $7 per hour under the new configuration, since more passengers will use express service. On the other hand, fare collection will be complicated a bit by the need to carry local passengers on express routes to points at which they will transfer
to local routes.

Because of the many impacts on passenger service the new configuration would cause, one would expect, some opposition to its implementation. However, the opposition should not be too great, since the large service cuts affect only branches with tiny levels of demand.

Whether or not the new configuration should be implemented depends on the MTC's objectives and financial resources. Given abundant financial resources, the operator may be reluctant to change service levels as the new configuration requires. Given a need to reduce its operating budget, however, the new configuration offers the MTC a way to save an additional 2.2 peak hour buses while raising average wait time in the corridor by less than one minute, reducing average in-vehicle time by about half a minute, and causing only about 14 transfers per hour. This small change in level of service, which could lead to the savings of 2.2 peak vehicles, should be viewed as a valuable opportunity by any operator that is not in an expansion mode.

7.4 Conclusions

This chapter has shown how the same principles that were found useful in the design of single arterial corridors can also be applied in a corridor with a branching route network. It has shown how a branching network route design can be decomposed into a sum of single arterial designs, provided one is willing to impose restrictions on the way branches and trunks are connected and provided one is designing service with exclusive boarding/alighting zones. It has given a simple algorithm for efficiently finding the optimal decomposition. When the restrictions necessary to apply these efficient algorithms are too severe, one may employ a manual or semi-manual procedure that is based on the same design principles.
A case study performed on a corridor in Minneapolis clearly laid out and applied these manual design principles to a route system that includes express and local routes, with the local routes having overlapping service zones. It showed how careful planning could reduce the fleet requirement in a corridor from 27.5 to 21.3 vehicles; it also showed that even if headways were at their optimal levels, changing the routing configuration and operating strategy in accordance with the design principles set forth in this thesis could yield additional reductions in fleet requirements of around 10 percent without straining the level of service much. In an era in which service cuts are being forced on operators due to reduction of subsidies, routing configurations that employ the strategies discussed in this thesis can be a valuable way of economizing without great pain.
Chapter 8

Serving Corridors With Multiple Downtown Terminals

All of the corridors studied in the previous chapters, whether they had single arterial or branching networks, were modeled as having a single downtown terminal shared by all the routes in that corridor. This single downtown terminal model is relevant in many situations, but in other situations there may be more than one downtown terminal to which buses serving a given corridor should be routed because of the spatial distribution of destinations. This chapter is concerned with finding the best routing configuration for a corridor in which there is more than one downtown terminal which buses should serve.

Multiple downtown terminals often exist when downtowns are so large that a direct path from the corridor to a single downtown terminal does not provide adequate coverage of the downtown, and when congestion makes circulating buses through the downtown impractical. In such situations a corridor may have service to a north terminal and a south terminal, at opposite ends of the downtown, for example. Even if the downtown could be served with a single terminal at the far side of the downtown, if it takes a long time to travel from the near side of the downtown to the far side because of traffic congestion, it may be worthwhile to establish a terminal on the near side of downtown and have some runs terminate there. A third reason for the existence of multiple downtown
terminals is that there may be employment centers near the downtown, such as a medical or government complex, which generate a considerable demand. These employment centers may not warrant as much direct service as the downtown, but part of the economies of scale associated with high demand corridors is being able to offer direct service to more destinations.

When there are a number of spatially distinct destinations with considerable demand concentrations, each with a corresponding bus terminal, a transit operator has many options in how to serve each one from a given corridor. He may serve some destinations with local routes, or express routes, or both. He may be able to serve two or more of these major destinations with the same route, if their alignment permits. He may require passengers destined for a certain destination to transfer between routes; that transfer may take place on the outer end of the corridor, or its inner end. There are also all the additional issues of route zoning, deadheading, and the extent of the express service area to be dealt with in designing multiple downtown terminal service.

Any configuration that serves a number of downtown terminals can be decomposed into a number of single terminal routing configurations. These single terminal configurations have two sources of interdependence. First, local passengers (those whose destination is in the corridor and could be served by routes going to any terminal) may be able to choose whether they will take a route that is destined for one terminal or another. In this kind of situation, the service level to one terminal obviously will affect the demand on routes that serve other terminals. Second, passengers originating on low demand segments may first board a bus that goes to one terminal, and later transfer to a bus going to another terminal. In this case again the demand for routes going to one terminal depends on the
service area of routes that go to the other terminals. Then each single
terminal configuration may itself be decomposed into single arterial
configurations if it is a branching network, or into a local service
network and an express service network, as discussed in Chapter 7.

The interdependencies between the different single terminal
configurations can be determined when the market segmentation is given.
Market segmentation here means assigning each trip in the corridor to
the single terminal configuration(s) by which it will be served. Given a
market segmentation, then, we can decompose the route design problem into
a number of single terminal routing problems and solve each independently.
(If some local trips are assigned to a set of routes with overlapping service
zones from which passengers may choose any one and these overlapping routes
go to different terminals, there will still be some interdependence between
the different terminals' configurations that cannot be determined exogenously.
Such a system is still easily solved; one must simply guess the market
share of those local trips that each terminal will get, optimize the networks
of each terminal, then determine again the market share of those local
trips that each terminal will get based on service frequencies, and iterate
until an equilibrium of supply and demand is found.)

Our approach to the multiple downtown terminal corridor route
design problem will then be to search for the best market segmentation,
since given an alternative market segmentation, we can use the procedures
developed in the previous chapters to find the best service configuration
for each terminal. We will have to search therefore over a number of
different market segmentations. In some corridors the number of feasible
segmentations is small enough that we may search exhaustively. In other
corridors there will be too many possible market segmentations to allow enumeration; for such corridors planners must use their judgment to select a reasonable number of promising market segmentations and evaluate them.

8.1 Single Arterial Corridors with Multiple Downtown Terminals

Let us first examine the simplest multiple downtown terminal corridor, a single arterial corridor with two downtown terminals. Let one downtown terminal be called A and the other B. The arterial has nodes and segments 1, ..., n, as in previous chapters. Figure 8.1 illustrates this corridor. Local passengers are those whose destinations could be served by an A route or a B route; A passengers are those whose destination is served only by an A route; B passengers are those whose destination is served only by a B route. It is not necessary that node 1 have routes going to both downtown terminals since we may want to require some passengers to transfer. However, we shall require that if direct service to terminal t is offered at segment j, direct service to t must also be offered at segments j+1, ..., n.

For reasons given in Chapter 5, letting the local routes have overlapping service zones is not usually as good a strategy on single arterial corridors as letting them have exclusive boarding and alighting zones. Previous chapters have also shown the advantage of isolating non-local passengers and serving them with express routes. For these reasons we will require non-overlapping service zones in the succeeding discussion. Therefore local passengers boarding on a given segment will have no choice as to which route they use. In the region of the corridor that is served by routes that go to only one terminal, clearly the local
Figure 8.1

Corridor with Two Downtown Terminals
market must be either served by that terminal's routes alone, or, if
that terminal has no local route in that region, local passengers must
transfer downstream to a route that goes to the other terminal. In the
region that has service to both terminals, we shall require that the
entire local market be served by routes that go to the same terminal.
Probably most operators would want to impose this restriction since
without it the system would probably be very confusing to passenger.

Given the above restrictions on market segmentations, the following
variables will uniquely determine a market segmentation:

\[
p = \text{terminal served from node 1; } \bar{p} \text{ is the other terminal}
q = \text{outermost node with service to terminal } \bar{p}
\]

\[
r = \begin{cases} 
0 & \text{if all local passengers use routes going to } q \\
1 & \text{if local passengers boarding upstream of } q \text{ use routes}
\text{ going to } p \text{ and local passengers boarding at } q \text{ and}
\text{downstream use routes going to } \bar{p} \\
2 & \text{if local passengers boarding upstream of } q \text{ board routes}
\text{ going to } p \text{ and those whose destination is downstream}
\text{ of } q \text{ transfer at } q \text{ to a route going to } \bar{p}, \text{ and local}
\text{passengers boarding at } q \text{ and downstream use routes}
\text{going to } \bar{p}.
\end{cases}
\]

Since \( p \) can take on 2 values, \( r \) can take on 3 values, and \( q \) can take
on \( n \) values, there are at most \( 6n \) feasible decompositions. Since \( n \) will
usually be small in a single arterial corridor, enumerating every
alternative is a computationally tractable way of solving this problem.
A planner could also screen out some of the alternatives for different
reasons and thereby lighten the computational load.

This simple network just studied has only 2 downtown terminals. What
about single arterial networks with \( m \) downtown terminals, \( m = 3, 4, \ldots \)?
Unfortunately, the number of feasible decompositions grows exponentially with \( m \), and enumeration becomes too burdensome a method to be practical. With more than two downtown terminals, then, the planner's judgement in screening and choosing alternatives becomes more important. The role of the planner in selecting alternative market segmentations is discussed further in the following section.

8.2 Branching Corridors with Multiple Destinations

In Chapter 7 we showed that with certain restrictions the number of feasible decompositions of express/local service on a branching network into a single arterial configuration was manageable. These restrictions required that the way branches and trunks are connected must be the same in the configuration of all three service types: local service, express service, and outbound service. Adding another downtown terminal to the network can be thought of as adding one or more service types to the system: express service to the second terminal, and possibly local service to the second terminal and outbound service from the second terminal. If we require that branches and trunks also be connected in the same way for these service types as for the original three service types, then a similar algorithm to the dynamic programming algorithm presented in Section 7.2 can be used for multiple downtown terminal corridors.

However, we showed that the restrictions on the feasible market segmentation implicit in that algorithm would probably be too severe for many operators, and that for various reasons that cannot be modeled operators will usually prefer to determine manually what service types are offered on each segment. When the network is further complicated
by the addition of a second or third downtown terminal, operators will likely want still more control over the service types offered on each segment. Therefore, both from a computational point of view and from the operator's point of view, a procedure that lets the operator select the market segmentation manually would be the most valuable. For this reason the following semi-manual procedure is suggested for all but the simplest multiple downtown terminal corridors.

The semi-manual procedure requires that the planner choose a number of different market segmentations. A market segmentation is determined by specifying the service types that will be offered on each segment. A particular market segmentation allows the problem to be decomposed into separate networks for each service type, and these networks are then optimized using the algorithms of Chapters 4 through 7. The superposition of these solutions is then the solution for that particular market segmentation. Comparing the solutions of the alternative market segmentations allows the planner to choose the segmentation that best meets the operator's objective.

Five principles of route design in branching corridors were summarized in Section 7.3.8. These principles apply when there are multiple downtown terminals as well as when there is a single downtown terminal. Because of the added complexity of the multiple destination problem, another principle should be added.

6) Some downtown terminals should be marked out for primarily express service. In general, one terminal should be chosen as the terminal for local service, and the other terminals should have only express service. The factors that make a terminal better suited for express service only are: having ready access to an expressway or high-speed path, having a demand pattern that suggests offering peak hour service only, and being located such that the major demands of the
inner segments of the corridor, which usually have local service only, could not be served by a route going to that terminal (such as a terminal on the near side of downtown, or in a commercial area outside the downtown).

The planner will find that choosing a small number of promising market segmentations is not difficult. With a little experience, one should be able to come very close to the optimal segmentation.

8.3 Multiple Downtown Terminal Case Study

The Watertown corridor in metropolitan Boston will again serve as a case study to illustrate the routing strategies discussed in this chapter. While the previous case studies of the Watertown corridor required abstracting certain aspects of the corridor in order to more plainly illustrate the routing strategies then under consideration, we have now arrived at the point in our study of corridor route design where we can model all of the features of the corridor realistically and design the best routing configuration for the bus patrons of the corridor. This case study will therefore allow a fair comparison of the configurations suggested by the design procedures developed in this research and the existing configuration.
8.3.1 **Modeling the Corridor Network**

The physical network of roads along which buses serving the Watertown corridor will operate and the existing routes in the corridor are shown in Figure 3.1. The road network is essentially a single arterial, with a parallel expressway available for express service and deadheading. The corridor has three downtown terminals: Kenmore Square, Copley Square, and downtown (near South Station).

Figure 8.2 depicts the network as it is modeled for this chapter. It has four potential uptown terminals, as did the models of Chapters 5 and 6. Unlike previous chapters, the arterial is subdivided into six segments: Watertown Square to Newton Square; just after Newton Corner to just before Oak Square; Oak Square to just before Brighton Center; Brighton Center; just after Brighton Center to Linden Street; and just after Linden Street to downtown. The extra segments allow for more flexibility in defining service areas for the express routes.

Expressway access is at Newton Corner and near Linden Street. Existing express routes do not extend their service area inward beyond Brighton Center, but the existence of expressway access near Linden Street suggests trying to offer express service to those boarding between Brighton Center and Linden Street.

The Watertown corridor is actually part of a branching network. A little over a mile out of Kenmore Square, the Watertown corridor branches off the Commonwealth Avenue corridor, which is served by streetcar trains which go underground at Kenmore Square, forming a part of the subway service from Kenmore Square to downtown. Because of the high service frequency on Commonwealth Avenue and because the streetcars offer
Figure 8.2

The Watertown Corridor with Multiple Downtown Destinations

--- local arterial
----- expressway
--- subway

Nodes:
1. Watertown Square \textsuperscript{1,2}/
2. Newton Corner \textsuperscript{1,2}/
3. Oak Square \textsuperscript{1}/
4. Brighton Center \textsuperscript{1,2}/
5. Just after Brighton Center
6. Linden St.
7. Kenmore Square \textsuperscript{3}/
8. Copley Square \textsuperscript{4}/
9. Downtown

\textsuperscript{1}/ Potential uptown terminal for local service
\textsuperscript{2}/ Potential uptown terminal for express service
\textsuperscript{3}/ Downtown terminal for local service
\textsuperscript{4}/ Potential downtown terminal for express service
direct service to downtown, the segment shared by the Watertown and Commonwealth Avenue corridors is, in the existing service configuration, part of the Commonwealth Avenue corridor's service zone. That is, the Watertown buses do not allow inbound passengers to board or outbound passengers to alight along the shared segment.

Because of the streetcar line linking the segment shared by the Watertown and Commonwealth Avenue corridor, this case study assumed that the existing boarding and alighting restrictions on that shared segment on the local Watertown corridor buses would remain in force in any new configuration. With this assumption, which was made implicitly in the previous chapters, the Watertown corridor can be analyzed independently, as it was in the previous chapters.

8.3.2 Objectives and Constraints

Designs were sought for two different objectives, as in the previous case studies of the Watertown corridor, the "austerity objective" of minimizing operator cost, and the "prosperity objective" of minimizing operator plus passenger cost, with passenger travel time valued at $3/hour and a transfer penalty of 20¢ per passenger.

Operating policies and constraints were largely the same as in the previous case studies, as described in Section 3.3.4, with the following exception. Only one bus type was allowed, the standard bus, so that benefits due to a better routing strategy will be not be confounded with benefits due to using a different size vehicle.

In order to properly compare alternative configurations, some costs of subway service had to be estimated. Some designs might carry more express passengers, thus requiring less passengers to transfer at Kenmore
Square to a subway and ride the subway to either Copley Square or
downtown; other designs might carry less express passengers, putting a
greater load on the subway. Accurately estimating marginal operator
costs of peak hour subway service from Kenmore to Copley and from Kenmore
to downtown was impossible due to lack of data. The MBTA was not able
to give figures for these costs, so proceeding with the study required
guessing reasonable values based on available information. The cost of
extra capacity on the Green Line (the subway line that serves Kenmore
Square) was estimated as 60¢ per passenger per hour. Extra passengers
from Kenmore to downtown would certainly require additional Green
Line capacity. Extra passengers destined for Copley perhaps would and
perhaps would not require extra capacity, since the peak load point of
the Green Line is near Copley Square. For this reason additional
capacity to Copley was costed at 30¢ per passenger per hour. One might
argue that a benefit of not providing express service to Copley Square is
that some passengers, faced with the choice of transferring at Kenmore to
the subway and walking from Kenmore to their destination, will prefer to
walk. Not accounting for these walking passengers would bias the analysis
in favor of solutions that include more express service to Copley. In order
to avoid this bias, then, it was assumed that 20 percent of the Copley-
bound passengers would walk from Kenmore rather than transfer to subway.
This rather high fraction was used because of the heavy concentration of
employment between Kenmore Square and Copley Square, most notably the
Prudential Center which is about a 12 minute walk from Kenmore Square.
Therefore the marginal operator cost incurred on the the subway by the
average Copley-bound passenger who rides a bus to Kenmore Square was
assumed to be 24¢.
Passenger costs incurred by subway riders also had to be estimated. It was assumed that the average downtown-bound passenger who had to transfer to a subway at Kenmore Square would spend 3 minutes walking to the subway platform and waiting for a train, and then 10 minutes in the vehicle. The in-vehicle time to Copley Square was taken to be 4 minutes for those who transfer to subway at Kenmore Square; for those who walk from Kenmore Square an incremental walk time cost of 10 minutes was imputed. Each transferring passenger also incurs a transfer penalty.

8.3.3 Existing Service Configuration with Optimal Headways

In order to have a fair base of comparison against which to measure the benefits of any new routing configuration, headways under the present configuration were optimized under both the prosperity and austerity objectives. The two objectives, in fact, call for the same headways. Table 8.1 presents the details of this solution. It requires 33 standard buses, and has an hourly operator cost of $1158, of which $101 is incurred on the subway. The objective function under the prosperity objective, which includes passenger travel time valued at $3/hour and a 20¢ transfer penalty for 242 transferring passengers as well as operator cost, is $3603, with average wait time (including walk time for Copley-bound passengers who walk from Kenmore Square to their destination, since walk time, like wait time, is out-of-vehicle travel time) of 4.0 minutes, average in-vehicle time of 15.2 minutes, and 242 passengers transferring at Kenmore Square per hour.

8.3.4 Analysis of Demand

In order to evaluate a configuration that offers either more or less express service, it is important to know the demands from each segment to each express terminal. For segments with existing express service, these demands were known from the peak load counts. Currently
Table 8.1

Optimal Headways With the Existing Service Configuration
(for Both Austerity and Prosperity Objective)

<table>
<thead>
<tr>
<th>Bus Routes</th>
</tr>
</thead>
<tbody>
<tr>
<td>In-Town Terminal</td>
</tr>
<tr>
<td>Downtown</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Downtown</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Copley Sq.</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Kenmore Sq.</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Service Level and Costs

<table>
<thead>
<tr>
<th>Average 1/ Wait Time</th>
<th>Average In Vehicle Time</th>
<th># Pass.</th>
<th># of Transferring</th>
<th># of Buses</th>
<th>Operator Subway Cost</th>
<th>Operator Cost (Total)</th>
<th>Operator &amp; Pass. Cost (Total)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.0 min</td>
<td>15.2 min</td>
<td>241.6</td>
<td>33</td>
<td></td>
<td>$101</td>
<td>$1158</td>
<td>$3603</td>
</tr>
</tbody>
</table>

1/ Includes walk time of Copley-bound passengers who walk from Kenmore
passengers boarding on segments lacking express service must ride Route 57 to Kenmore Square and then transfer to the subway. The task of the demand analysis was then to disaggregate the Kenmore Square demands coming from segments lacking express service to either Copley or downtown into demands for downtown, demands for Copley Square, and demands for Kenmore Square. These destination splits were estimated based on existing demands to the various destinations on similar segments in the corridor that have direct service to two or more of the terminals. It was estimated that on segments lacking express service to Copley and downtown, for every passenger destined for Kenmore Square there were 9.7 passengers destined for downtown and 2.9 passengers destined for Copley Square.

8.3.5 Selection and Evaluation of Alternative Configurations

The procedure we shall employ in trying to find a better routing configuration is the semi-manual heuristic in which the planner chooses a number of alternative market segmentations, each specifying the type of service to be offered on each segment. Each alternative is then evaluated by decomposing it into a separate network for each service type, and then optimizing each service type for its given market using the single arterial, single destination route design procedures developed in Chapters 3-6. A planner must use his judgment in selecting a set of alternative market segmentations. In this case study, 9 promising market segmentations were chosen and evaluated. The reasoning that led to the selection of these 9 alternatives is given below.

First of all, it is the general policy of the MBTA not to run local buses into the downtown when there is a parallel subway service. This
policy implies that local bus routes must terminate at Kenmore Square. Requiring downtown-bound passengers to transfer to subway at stations like Kenmore is common in the MBTA's system, so configurations that require this kind of transfer were allowed.

Secondly, the high demand for downtown express service along the small segment from Watertown Square to Newton Corner (554 passengers/hour) and the access to the turnpike at Newton Corner suggests that any promising alternative ought to offer express service to downtown along that segment.

Thirdly, the proximity of Oak Square to its neighboring terminals, Newton Corner and Brighton Center, the relatively low demand intensity between Newton Corner and Oak Square, and the location of Oak Square with respect to turnpike access suggested that it was very unlikely for a good solution to include an express route with a terminal at Oak Square. The feasible uptown terminals for express service were therefore Watertown Square, Newton Corner, and Brighton Center. No such restriction was put on the local service configuration, however.

Fourthly, it was observed that downtown is a superior downtown terminal to Copley Square. The downtown attracts a far greater demand than Copley Square; furthermore, sending an express bus to downtown rather than to Copley Square only requires an extra minute of run time while it costs considerably more to carry passengers on the subway to downtown than to Copley Square. These facts suggest that any market segmentation that is worthy of consideration should have the downtown express service area at least as great as the Copley express service area.

Given the above suggestions and limitations, there remain nine
feasible market segmentations. There are four segments that can make up an express service zone: 1) Watertown Square to Newton Corner; 2) just after Newton Corner to just before Brighton Center; 3) Brighton Center; 4) just after Brighton Center to Linden Street. A market segmentation is uniquely determined by specifying the segments that are to receive service to the two express terminals. For each market segmentation the number of Copley-bound and downtown-bound passengers who must use a local route and then transfer to subway at Kenmore Square can be computed. The 9 alternative market segmentations with the resulting number of transfers are then:

<table>
<thead>
<tr>
<th>Alternative</th>
<th>Downtown(^1)</th>
<th>Copley(^2)</th>
<th>Downtown pax/hr Transferring at Kenmore</th>
<th>Copley pax/hr Transferring at Kenmore</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>-</td>
<td>570</td>
<td>450</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>570</td>
<td>184</td>
</tr>
<tr>
<td>3</td>
<td>1-3</td>
<td>-</td>
<td>94</td>
<td>450</td>
</tr>
<tr>
<td>4</td>
<td>1-3</td>
<td>1</td>
<td>94</td>
<td>184</td>
</tr>
<tr>
<td>5</td>
<td>1-3</td>
<td>1-3</td>
<td>94</td>
<td>41</td>
</tr>
<tr>
<td>6</td>
<td>1-4</td>
<td>-</td>
<td>0</td>
<td>450</td>
</tr>
<tr>
<td>7</td>
<td>1-4</td>
<td>1</td>
<td>0</td>
<td>184</td>
</tr>
<tr>
<td>8</td>
<td>1-4</td>
<td>1-3</td>
<td>0</td>
<td>41</td>
</tr>
<tr>
<td>9</td>
<td>1-4</td>
<td>1-4</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

\(^1\)Segments in Downtown Express Service Area

\(^2\)Segments in Copley Express Service Area

In every alternative, local service will be offered between Watertown Square
and Kenmore Square; and it will carry all local demands as well as downtown and Copley demands that are not carried by the express routes.

Evaluating these nine alternatives required designing express service to downtown for three different markets: segment 1, segments 1-3, and segments 1-4; designing express service to Copley Square for the same three markets; and designing local service for the nine different markets, one for each alternative. The solution for any alternative could then be found as the combination of the corresponding local, Copley, and downtown designs. The above procedure was executed for both the austerity objective and the prosperity objective.

Table 8.2 displays the objective function values for the nine alternatives under the austerity objective. The best alternative is the eighth, providing express service to downtown on segments 1-4 and express service to Copley and segments 1-3 at an hourly operator cost of $1036. This alternative requires one bus fewer than the existing configuration, and saves $91 in hourly subway operating cost since it carries 237 more passengers per hour directly to their destinations in express buses. Its total operator cost savings over the existing configuration is $122, which is over 10% of the current operator cost.

The seventh and ninth alternatives were almost as good as the eighth, having total hourly operator costs of $1,038 and $1,058 respectively. The next best alternative, the fourth, is the existing market segmentation, but it requires two buses fewer than the existing configuration because it employs partial deadheading on the local route.

Table 8.3 displays the results of optimizing the designs of the nine alternatives under the prosperity objective. The optimal configurations for each alternative market segmentation under the prosperity objective
Table 8.2

Optimal Designs for Alternative Market Segmentations

Under the Austerity Objective

<table>
<thead>
<tr>
<th>Alternative</th>
<th>Express Buses</th>
<th>Local Buses</th>
<th>Total Buses</th>
<th>Subway Operating Cost</th>
<th>Total Operating Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>24</td>
<td>31</td>
<td>$450</td>
<td>$1,444</td>
</tr>
<tr>
<td>2</td>
<td>11</td>
<td>19</td>
<td>30</td>
<td>$386</td>
<td>$1,348</td>
</tr>
<tr>
<td>3</td>
<td>15</td>
<td>17</td>
<td>32</td>
<td>$164</td>
<td>$1,190</td>
</tr>
<tr>
<td>4</td>
<td>19</td>
<td>12</td>
<td>31</td>
<td>$101</td>
<td>$1,094</td>
</tr>
<tr>
<td>5</td>
<td>22</td>
<td>11</td>
<td>33</td>
<td>$66</td>
<td>$1,123</td>
</tr>
<tr>
<td>6</td>
<td>16</td>
<td>16</td>
<td>32</td>
<td>$108</td>
<td>$1,134</td>
</tr>
<tr>
<td>7</td>
<td>20</td>
<td>11</td>
<td>31</td>
<td>$44</td>
<td>$1,038</td>
</tr>
<tr>
<td>8</td>
<td>23</td>
<td>9</td>
<td>32</td>
<td>$10</td>
<td>$1,036</td>
</tr>
<tr>
<td>9</td>
<td>24</td>
<td>9</td>
<td>33</td>
<td>0</td>
<td>$1,058</td>
</tr>
</tbody>
</table>
Table 8.3

Optimal Designs for Alternative Market Segmentations

Under the Prosperity Objective

<table>
<thead>
<tr>
<th>Alternative</th>
<th>Avg. Wait Time (min.)</th>
<th>Average In-Vehicle Time (min.)</th>
<th>Buses</th>
<th>Subway Operating Costs</th>
<th>Total Operator and Passenger Costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.1</td>
<td>17.8</td>
<td>33</td>
<td>$450</td>
<td>$4,444</td>
</tr>
<tr>
<td>2</td>
<td>3.6</td>
<td>16.4</td>
<td>33</td>
<td>$386</td>
<td>$4,088</td>
</tr>
<tr>
<td>3</td>
<td>3.6</td>
<td>16.7</td>
<td>34</td>
<td>$164</td>
<td>$3,888</td>
</tr>
<tr>
<td>4</td>
<td>4.0</td>
<td>15.2</td>
<td>33</td>
<td>$101</td>
<td>$3,603</td>
</tr>
<tr>
<td>5</td>
<td>4.0</td>
<td>14.8</td>
<td>36</td>
<td>$66</td>
<td>$3,596</td>
</tr>
<tr>
<td>6</td>
<td>3.7</td>
<td>16.9</td>
<td>33</td>
<td>$108</td>
<td>$3,815</td>
</tr>
<tr>
<td>7</td>
<td>3.9</td>
<td>15.3</td>
<td>33</td>
<td>$44</td>
<td>$3,535</td>
</tr>
<tr>
<td>8</td>
<td>4.1</td>
<td>15.0</td>
<td>35</td>
<td>$10</td>
<td>$3,536</td>
</tr>
<tr>
<td>9</td>
<td>4.1</td>
<td>15.1</td>
<td>35</td>
<td>0</td>
<td>$3,528</td>
</tr>
</tbody>
</table>
require one to three buses more than the corresponding austerity configurations, but they save travel time. The optimal configuration for the existing market segmentation is the existing routing configuration, with its objective function value of $3603. The best market segmentation, the ninth, offers express service to downtown and Copley from the entire area between Watertown Square and Linden Street; its objection function has a value of $3528, a savings of $75, which is 6.5% of the existing operator cost. Reducing the area with express service to Copley yields two other market segmentations (alternatives 7 and 8) whose objectives are also very low, within $8 of the ninth alternative.

Under both the austerity and the prosperity objectives, the last three alternatives were close to the best. In light of these results, it seems clear that for any objective lying between the austerity and the prosperity objective it would be good for the MBTA to extend the downtown express service area to Linden Street, while it would make little difference whether the Copley express service area remained unchanged or were extended to either Brighton Center or Linden Street. Alternative 7 in particular looks promising, since it would cause the smallest change in service offered among the three best alternatives. Given the uncertainty in the operating cost parameters for subway, the seventh alternative may also be the best since its low objective function value is the least dependent on savings in subway costs. Under the prosperity objective, the best configuration for alternative 7 has four routes. Two express routes serve the downtown terminal, one serving the area from Watertown Square to Newton Corner (just as Route 304) and the other serving the area from Newton Corner to Linden Street. This second route is a simple modification of Route 301, and would operate at a headway
slightly less than the optimal headway of Route 301. The express route to Copley is the same as Route 302, serving the segment between Watertown Square and Newton Corner. The local route operates between Watertown Square and Kenmore Square, without deadheading, at a service headway of 5.7 minutes. The austerity configuration for alternative 7 has the same three express routes as the prosperity solution. Its local route uses two less buses, however, and deadheads half its runs in the light direction on the Massachusetts Turnpike, offering a service headway of 6.3 minutes inbound and 12.6 minutes outbound. The austerity configuration for alternative 7 requires 2 buses fewer than the present configuration with its headways optimized, and reduces subway operating costs by $67, thus reducing total operator cost by over 10%. The prosperity configuration for alternative 7 uses the same number of vehicles required by the existing configuration, but by reducing overall travel time and sparing 94 passengers per hour the need to transfer at Kenmore Square, it reduces total passenger cost by $68, which is equivalent to saving the cost of operating two buses or of reducing travel time for each peak hour passenger in the Watertown corridor by a little over half a minute.

It is important to note that designs were made and selected in this case study without regard to fare and revenue impacts, because fare policy is usually outside the scope of the transit planner in Boston, being politically determined. Nevertheless fare and revenue impacts can be important in determining which alternative is preferable to the operator. For example, an operator, under a certain political climate, may shy away from alternatives that force passengers to pay a higher fare. Since none of the good alternatives found in this case study would deny local service to any current patrons of local service or
express service to any current patrons of express service, this factor is not a consideration in this case study. Under another political climate, an operator might have a strong preference for alternatives that raise more revenue. This consideration should be important in the Boston case study, since Boston's transit system is in a financial crisis; however, depending on the fare structure an alternative could either increase or decrease the revenue. In 1979 an express passenger on a Watertown corridor route paid a higher fare than the local bus fare plus subway fare, while in 1981 it is 10¢ cheaper to use an express route than to use a local route and transfer to subway. If the express route fares stay low, the operator will favor alternatives that do not increase the express route ridership; however, if express fares rise (as expected), the operator may prefer alternative 9 which gets as many passengers as possible using express routes.

8.4 Conclusions

In going from single downtown terminal corridors to corridors with multiple downtown terminals, especially corridors with branching networks, we leave the realm of route design problems that can be efficiently solved through an entirely automated procedure. Instead, such problems are best solved by a semi-manual procedure in which the planner identifies a number of different market segmentations, and an optimal configuration for each segmentation is then found using the automated procedures developed in the earlier chapters that allow for route zoning, partial deadheading, and parallel local and express service. Because of planning considerations that cannot be modeled analytically and because the well-behaved nature
of the problem makes identifying good market segmentations relatively easy, it is expected that the semi-manual procedure will yield as good a solution (in the eyes of the planner/operator) as a fully automated procedure (if one existed) most of the time.

A case study of the Watertown corridor of Boston showed that even in a corridor that already uses express service and route zoning, a more careful, systematic application of the routing strategies discussed in this thesis can lead to routing configurations that require 10% less of the operator's resources.
Chapter 9

Summary of Findings and Directions for Further Research

9.1 Summary of Findings

This research has two sides: exploring innovative routing strategies, and developing analytical planning methods for using these strategies. One goal of this thesis was to show the value of routing strategies that are seldom employed in transit systems. Even if the mathematical models that were developed are forgotten or ignored, this research can still have a significant impact upon the efficiency of transit service by making planners aware of these routing strategies, principles for applying them, and their potential benefits. The planning methods that have been developed are also valuable, however, since it is important to know the best way to apply any strategy in a given corridor. In corridors where some of these strategies have already been implemented, applying these planning tools can help one find better ways to employ those strategies and combine them with other routing strategies to yield still greater improvements in efficiency, as some of the case studies have demonstrated.

The first strategy that was explored was deadheading runs in the light direction. Peak period demands in urban corridors are usually much heavier in one direction than in the other, so that the light direction requires a service frequency much lower than the heavy direction. If a high speed path is available, some runs could deadhead in the light direction and take advantage of this quick path. A route can deadhead all of its runs, provided it is part of a route system in which there are other
routes that can carry the light direction demands. Partial deadheading, as opposed to complete deadheading, is a more innovative strategy; it has some of the runs of a route return in service in the light direction and other runs deadhead.

The benefit of deadheading is that the more runs use the faster paths in the light direction, the lower the average turnaround time will be and the fewer vehicles will be needed. The disbenefit of deadheading is that it lowers the service level in the light direction. Light direction loads on existing peak period routes with no deadheading are often very low, however, so that the light direction service level can justifiably be cut. The factors that help make deadheading beneficial are the availability of a high-speed path, high service frequencies in the heavy direction, and low demands in the light direction.

A case study on an existing route in the Boston area showed that 2 out of 14 vehicles on the route could be saved through partial deadheading.

Existing planning methods are adequate for obtaining a lower bound of the number of vehicles needed to meet a schedule that includes partial deadheading. This lower bound allows departures to be irregularly spaced. A closed-form solution was found for the problem of minimizing the number of vehicles needed to meet a given schedule if departures must be regularly spaced. In addition, a procedure that can be executed manually was developed for finding the optimal partial deadheading configuration, since the number of vehicles needed is not a monotonic function of service headway.

The second strategy that was explored was route zoning. Previous studies have been made of route zoning in commuter rail and in express bus
service, but this is the first study (to our knowledge) of zoning on local bus routes. A zonal system is composed of two or more routes, wherein the outer segments which have smaller passenger flows have a smaller service frequency than the inner segments because some runs are systematically short-turned. Routes belonging to the zonal system may completely or partially deadhead in the light direction. Three types of zonal service were discussed. In zonal service with exclusive boarding/alighting zones, each segment has a unique route that is authorized to pick up inbound passengers there and a unique route authorized to discharge outbound passengers there. Outside a route’s boarding/alighting zone, it will stop only to discharge inbound passengers and pick up outbound passengers. In corridors with small local demands, a zonal route with exclusive boarding/alighting zones can be almost as effective as an express route in reducing passenger travel time, since the buses serving the outer zone will have to make few stops between their service zone and the downtown. In zonal service with overlapping service zones, the routes comprising the zonal route system have no boarding or alighting restrictions. The effectiveness of an overlapping zonal system depends on getting long distance passengers to use the shortest route that can serve them and getting short distance passengers to fill the extra capacity on all the routes. If passengers can be induced to ride on the routes we want them to use, an overlapping zonal system can be more efficient than a zonal system with exclusive service zones since it can keep all routes near their capacity at the corridor peak load point. Inducing passengers to ride the route we want them to ride is difficult, however; Section 5.3
discussed various incentives and disincentives one might employ. Overlapping zonal service on a single arterial corridor is difficult to implement effectively, because the shorter routes are entirely overlapped by the longer routes, which tends to make the longer routes overcrowded and the shorter routes underutilized. In a branching network, however, overlapping zonal service can often be more easily executed, particularly if the different branches can be designed so as to generate roughly the same load per bus. The third type of zonal service has non-contiguous exclusive boarding/alighting zones. This service type tries to combine the advantages of exclusive service zones with the advantages of allowing just enough boarding in the morning peak (alighting in the evening peak) outside the primary service zone to keep bus loads near capacity at the corridor peak load point. This strategy appears promising on paper, but its success is more subject to fluctuations in demand and passenger stop choice than the other zonal service strategies; furthermore, passengers can be expected to have difficulty understanding this type of service. For these reasons, this strategy was not studied beyond the discussion stage.

Factors that contribute positively to the efficiency of a zonal service are the length of a corridor, a load profile that rises gradually from the uptown end of the corridor, a corridor peak load point that is near the downtown, the availability of intermediate terminals within the corridor, and a small local demand relative to the downtown demand. The benefits of route zoning are a reduced need for vehicles and reduced travel times for longer distance travelers. The disbenefit of route zoning with exclusive boarding/alighting zones is increased wait times.
With overlapping zones, the time passengers wait for the first bus to pass them will not change much, but passengers will not always be able to board the first bus. The confusion and overcrowding that can result from the overlapped system can make the service level considerably worse, although careful phasing of departures and implementation of boarding incentives and disincentives can alleviate some of these problems.

A case study of a heavy demand corridor demonstrated the potential savings a zonal system can yield relative to conventional local service. Whereas serving the morning peak demands of the entire Watertown corridor with conventional local service would require 34 vehicles, a system of two zonal routes with exclusive boarding zones, one of which employs partial deadheading, would require only 27.

An inexpensive dynamic program was developed that can find the optimal zonal configuration with exclusive boarding/alighting zones under a general objective function that can include both operator and passenger costs. The configuration is designed to serve both inbound and outbound demands in the most efficient way. Because of the need to pay attention to boarding incentives and disincentives in the design of zonal service with overlapping zones, no analytical model was developed for this strategy.

The third strategy for efficient routing in high-demand corridors is the joint design of express and local service. By segmenting the market between express and local passengers, the express passengers can be served much more efficiently, particularly if a high speed path is available. Even in the absence of a high speed path, segregating the express demand from the local demand and zoning each service type can lead
to a more efficient design than simply zonal local service for the entire market.

Express service can both reduce the vehicular requirements of a corridor and can substantially reduce travel time for the express passengers. An additional benefit is that express riders will bear a higher fare than local riders, making express service a potential source of added revenue. Segregating express from local passengers increases wait time for the local passengers, since their service frequency will be cut. The factors that positively contribute to the efficiency of an express/local zonal system in a corridor are a high demand for downtown service and the availability of a high speed path. The existence of high local demand also tends to make express/local service more efficient than local zonal service. Of course, all of the factors that contribute to the efficiency of the local route zoning and the deadheading strategies will also aid in the success of an express/local system, since both local route zoning and deadheading can be incorporated in an express/local configuration.

A case study showed that in a high demand corridor with an expressway available, a demand that required 34 vehicles for conventional local service and 27 vehicles for local zonal service with extensive deadheading could be served with 24 vehicles in an express/local zonal system consisting of two express routes and one local route. The case study also showed that if the expressway were not available, the best local zonal configuration would require 32 vehicles while an express/local zonal configuration required only 30 vehicles.

Finding the optimal express/zonal configuration is complicated by the
possibility that outer segments of the corridor may have too low a demand to support both a local and an express route, and consequently transfers may be required between local and express routes. It is also not known a priori how far in the express service area should extend. A dynamic programming model was developed that finds the optimal routing configuration accounting for all of these complexities, serving demands in the light direction as well as the heavy direction. The model requires $O(n^6)$ computations, where $n$ is the number of potential uptown terminals in the corridor. In realistic corridors $n$ will usually be small, making the program inexpensive to run; performing the model for the Watertown corridor with $n = 4$ required on 6 seconds of CPU time on an IBM 370/168.

Applying the local zonal and express/local zonal strategies to branching networks is not a new strategy in itself, but requires different planning methods because of the added complexity of the network. A dynamic programming procedure that embeds the single arterial zonal service algorithms developed earlier was developed to find the optimal configuration in a branching network under certain restrictions. Because it was expected that these restrictions would be too severe for many operators, a semi-manual heuristic procedure was developed in which the planner manually selects a set of alternative market segmentations (each of which is equivalent to a network decomposition), and then each market's service is designed using the procedures for single arterial route design.

A case study in a Minneapolis corridor illustrated the principles that a planner should follow in selecting a market segmentation. The analysis for this case study was entirely manual, since the local zonal routes had overlapping zones and since the express service had to be
designed to facilitate a timed transfer with the local service. The existing configuration in that corridor, which already included express zonal routes, local zonal routes with overlapping service zones, and some deadheading, could be operated with 23.5 vehicles; a routing configuration was designed in accordance with the principles set forth in this chapter that reduced the vehicular requirement by 9 percent.

The final strategy that was explored for improving the efficiency of corridor transit service was operating routes to more than one downtown terminal. Additional downtown terminals could be located at the near side of downtown, saving some runs the costly trip through the downtown, or at distant ends of the downtown or at peripheral employment centers near the downtown, saving both operator and passengers the cost of transferring and traveling on another route.

A semi-manual heuristic procedure was also suggested for designing service to multiple downtown terminals. The planner must select a set of alternative market segmentations, each specifying the service types available on each segment; the optimal configuration for each market segmentation can be found using the previous methods developed in this thesis, and then the best market segmentation can be selected. In a case study of a Boston corridor, 9 alternative market segmentations were evaluated. Although the existing configuration is quite efficient, including three zonal express routes that deadhead, a configuration was found that required 2 fewer buses and required 94 fewer passengers per hour to transfer from the local route to the subway, reducing operator cost by 10 percent.

Together, the array of models and procedures developed in this thesis
offer the transit planner some very effective tools for designing service in high demand corridors, whatever his objective or resource constraints may be. With these tools, the planner is equipped to find optimal or near-optimal designs that incorporate four powerful strategies for improving service efficiency, whether his corridor is as simple as a single arterial with a single downtown terminal or has a branching network uptown and must serve more than one downtown terminal.

The case studies performed in Chapters 3-8 have illustrated the potential of these four routing strategies for improving transit efficiency in high demand corridors. In a high demand corridor where only conventional local service is currently offered, a potential reduction in operator cost of over 25 percent should not be considered unrealistic. And the two corridor case studies that allowed direct comparison with an existing service configuration showed that the careful application of these tools in designing a more efficient service configuration can yield operator cost savings of 10 percent, even when the existing configurations already make use of these strategies. If an operator seeks to improve passenger service as well as reduce operator costs, these tools and strategies can still yield considerable efficiency improvements, although the savings will be somewhat less since segregating the market, as these strategies do, tends to increase wait time.

9.2 Research Contribution

This research has contributed to the body of knowledge in transit planning in three ways. First of all, it has systematically brought to light strategies for routing buses in urban corridors different from conventional local service. Most of these strategies have been known and variously applied, although this is the first mention of
partial deadheading in the literature (to the author's knowledge); however, this is the first time these strategies have been assembled, combined, and compared with local service and with each other with the goal of finding the best routing configuration for a corridor. This research has exposed a shortcoming in all route design methods that do not consider routing strategies other than conventional local service, since conventional local service was demonstrated inferior to other routing strategies in many situations.

Secondly, this research has developed methods for analyzing and planning routing configurations that employ these different routing strategies. Except for the work of Turnquist [25,15] and Bernstein [26], no planning methods have been previously developed for the efficient design of these innovative strategies in transit systems. This research offers the planner methods for designing local zonal, joint express/local zonal, and deadheading service on both single arterial and branching corridors with both one and multiple downtown terminals.

Third, this research has demonstrated through numerous case studies the potential of these routing strategies, if applied with the planning methods developed, to serve urban corridors more efficiently than they are now being served. Case studies of two high demand corridors indicated that in each an operator who was under financial pressure could reduce operating costs by 10% while still maintaining an adequate level of service; similarly, an operator with resources to spend on improving service could reduce a sum of operator and passenger cost by about 6% of operator cost.

9.3 Directions for Further Research

There appear to be two major directions of research into which this work may be extended. The first is in the area of implementation,
studying whether these strategies and tools are as valuable in practice as they appear on paper, and refining them based on the experience of implementation. The second is in the area of further design of planning methods, both in developing methods for strategies that were overlooked in this research and in integrating the results of this research with the overall transit service planning process.

On the implementation side, it would be worthwhile to try to get some changes in route configuration implemented and to monitor them to see if the promised efficiency gains are realized. Parallel to this would be an effort to identify practical problems and other factors in designing complex route configurations that may have been overlooked in this research. It is certain that a study such as this one could be improved after some practical experience.

Another means of achieving the same end would be to study routing configuration changes that have occurred in the past. If sufficient data are available, this kind of study could yield the same kind of insights as practical experience.

Turning to the need for more research into methods development, one strategy for which no formal procedure was presented in this research was overlapping zonal routes. Developing a procedure for design of overlapping routes is difficult because of the indeterminacy of the system in not knowing which route passengers will use, and because the way that indeterminacy is resolved depends so much on local factors such as the strength of boarding incentives and disincentives. Nevertheless overlapping zonal service is still probably a good strategy in many situations, particularly developing countries where boarding and alighting restrictions are harder to enforce and high speed paths for express service
are scarce. Further research into this area would probably prove useful.

A strategy that was touched on in one of the case studies but otherwise overlooked is the planning of timed transfers in a routing configuration. Timing transfers can make them far less onerous to passengers, and thereby can allow the operator a little more freedom in designing his routes. However, implementing timed transfers imposes restrictions on service frequencies, making the joint design of routes and frequencies a difficult problem.

A third issue for further study is the relationship of the scheduling of runs in the corridor for which routes are being designed to the scheduling of runs in the remainder of the system. The Boston case studies employed one approach to this issue, the Minneapolis case study another. In the Boston case studies we required that each route be scheduled to require an integer number of buses. Under such a restriction there is no interdependence in scheduling either among the different routes in the corridor or between those routes and the rest of the transit system, at least for a period as short as the morning peak. The Minneapolis case study, on the other hand, assumed extensive interlining could take place, both among the routes in the Chicago Avenue corridor, and between those routes and other routes in the transit system, particularly the routes that comprise the Route 5 north system. Because of this assumption, it was not necessary to assign an integer number of buses to a route. However, to suppose that x bus-hours per hour can be performed by exactly x buses and still maintain regular headways is unrealistically optimistic. In the Minneapolis case, we assumed that every bus-hour of
operation per hour would require a little more than one bus; a "productivity factor" was defined as the ratio between vehicles needed on a route and vehicle-hours of operation needed per hour on a route. A productivity factor was estimated for three classes of routes, and the number of vehicles assigned to each route was then modeled as the number of vehicle-hours of operation needed per hour times the corresponding productivity factor for that route.

Research into run scheduling could determine whether either or both of these methods of dealing with scheduling relationships are "good." A "good" method is defined as follows. Analysis of a certain routing change may predict it will save a certain number of vehicles, but when that change is implemented and integrated into the schedule of the entire transit system, the change may save fewer vehicles than predicted, or more. A "good" method of modeling scheduling interrelationships is one that accurately predicts the number of vehicles that a particular routing configuration will require when it is implemented. Better methods of modeling scheduling interrelationships would be valuable for route planning in general.

A fourth direction for further research into planning methods which was touched on in the first two chapters is the incorporation of these results into citywide route planning models. Existing citywide route planning models design conventional local routes only, and this research has demonstrated their inefficiency in high demand corridors. As suggested in the introductory chapters, the citywide network can be decomposed into corridors, and then the methods for corridor design employed. Rules for decomposition and ways to iterate between the
subproblems and the master problem could be developed that would hopefully yield a better procedure for citywide transit route planning.
References


Appendix A

Derivation of Vehicle Costs

The cost of a standard MBTA bus was computed as a sum of driver, fuel, maintenance (including parts), overhead, and capital costs, using the most recent costs available as of May, 1981.

Average driver wage was $11.06 per hour. Benefits and overhead are costed at 47% of the average driver wage. Thus, driver and overhead costs together were $16.26 per hour.

Average fuel consumption for the MBTA's buses in 1980 was 4.1 mpg. Current fuel prices are $1.16/gal, so that fuel cost is $0.28 per mile.

The average maintenance cost per mile at the MBTA during 1980 was 50 cents; inflated by 9%, this yields a maintenance cost of $0.545 per mile.

The cost of a new bus similar to those now operated by the MBTA is about $130,000. The annual capital cost, including finance charges, was assumed to be 14% of the purchase cost, or $18,200. Expecting a vehicle to run 30,000 miles a year, the per mile capital cost is $0.61.

The per mile costs (fuel, maintenance, capital) are converted to per hour costs by assuming an average speed (which includes layovers). Average speed was assumed to be 11 mph.

Thus, the total cost of a standard vehicle was computed as $32.05 per hour.

Driver and overhead costs for an articulated bus were assumed to be the same as for a standard bus. According to a recent UITP survey of
operators worldwide [33], articulated buses consume 21% more fuel than standard buses, yielding a fuel cost of $0.34 per mile. Maintenance costs (maintenance labor, maintenance parts, and tires) are 28.6% higher than those of a standard bus, or $0.701 per mile. Life expectancy and vehicle miles traveled per year of articulated buses were about the same as those of standard buses, so that at a purchase price of $240,000 the capital cost per mile is $1.12. Combining these costs using the same average speed yields a total cost of $40.03 per hour.

Assuming 45 seats on a standard bus and 65 on an articulated bus, the cost per seated passenger per hour is $0.71 on a standard bus and $0.61 on an articulated bus.
Appendix B

Likelihood that an Optimal Partial Deadheading Solution
with an Integer Headway Ratio Exists

As described in Chapter 4, an operator may want to require that the ratio of outbound headway to inbound headway on a partially deadheaded route be integer because both scheduling and operating are easier under this restriction. If there exists an integer headway ratio that minimizes the number of vehicles needed, then this restriction costs the operator nothing; however, if a solution with a non-integer headway ratio requires fewer buses than the best integer headway ratio solution, the operator may not want to impose this restriction. It is the goal of this appendix to determine how likely it is that an integer optimal solution will exist.

Recall that the number of vehicles required on a partially deadheaded bus, N, is modeled as the sum of n_A and n_B (see eq. (4.9)). Section 4.2.1 showed that, given the parameters of a route including the maximum headway inbound and maximum headway outbound, the optimal value for the inbound headway is within a couple of seconds of its maximum; for practical purposes, then, the inbound headway h_A is fixed at its maximum. With h_A fixed, n_A is determined, so the only decision variable left to be chosen is the outbound headway h_B, which is equivalent to choosing the headway ratio \( r = \frac{h_B}{h_A} \). The variable r determines n_B by eq. (4.8):

\[
    n_B(r) = \left\lfloor \frac{\pi + g(r)}{rh_A} \right\rfloor
\]  

(8.1)
where \(\langle x \rangle\) is the smallest integer greater than or equal to \(x\), \(g(r)\) is the fraction \(\frac{y-1}{y}\) where \(y\) is the least integer denominator of \(r\) when \(r\) is expressed as a fraction, and \(\pi\) is the effective deadheading premium as defined in eq. (4.5). The upper bound on \(h_B^m\), \(h_B^m\), can be transformed into an upper bound on \(r\), which is \(r^m = h_B^m/h_A\). Then the minimal \(n_B\) without an integer restriction on \(r\), \(n_B^*(r^m)\), is defined as

\[
  n_B^*(r^m) = \min_{r \leq r^m} \left\lfloor \frac{\pi + g(r)}{rh_A} \right\rfloor
\]

(B.2)

and the minimal \(n_B\) with an integer solution, \(n_B^i(r^m)\), is

\[
  n_B^i(r^m) = \left\lfloor \frac{\pi}{\text{Int}[r^m]h_A} \right\rfloor
\]

(B.3)

We want to find, then, the likelihood that \(n_B^*\) will be less than \(n_B^i\) for different ranges of values of the different parameters.

Given the input parameters \(\pi\), \(h_A\), and \(r^m\), there is no need to look for a probability of the existence of an optimal integer solution; we can simply compute the integer solution and the real solution, compare them, and then know with certainty whether the integer solution is optimal. Suppose, however, that we don't give specific values to all of the input parameters, but specify only that \(\pi\) is to be in the neighborhood of \(\pi_0\) and that \(h_A\) is to be in the neighborhood of \(h_0\). Then for any given value of \(r^m\) we can't compute the integer and real solutions, so we shall look for a way of estimating how likely it is that an integer solution is optimal.

Given the functions \(\langle a \rangle\) and \(\langle b \rangle\), if \(a\) and \(b\) are equally likely to take on any real value over a broad range, the probability that \(\langle a \rangle\) exceeds \(\langle b \rangle\) given that \(a \times b\) is easily shown to be

\[
  \Pr(\langle a \rangle > \langle b \rangle) = \begin{cases} 
  a - b & \text{if } a - b < 1 \\
  1 & \text{if } a - b \geq 1
\end{cases}
\]

(3.4)
Similarly,
\[
\Pr \left( \langle a \rangle - \langle b \rangle > 1 \right) = \begin{cases} 
0 & \text{if } a-b < 1 \\
\begin{cases} 
a-b-1 & \text{if } 1 \leq a-b < 2 \\ 
1 & \text{if } a-b \geq 2 
\end{cases}
\end{cases}
\tag{B.5}
\]

If \( \pi \) and \( h_A \) are variables for which any value over a considerable range is equally likely, then from eq. (B.4)
\[
\Pr \left( n_B^* (r^m) > n_B (r^m) \right) = \begin{cases} 
v(r^m) & \text{if } v(r^m) < 1 \\
1 & \text{if } v(r^m) \geq 1 
\end{cases}
\tag{B.6}
\]
where
\[
v(r^m) = \min_{r \leq r^m} \left( \frac{\pi + g(r)h_A}{rh_A} \right) - \frac{\pi}{\text{Int}[r^m]h_A}
\tag{B.7}
\]

Not knowing the exact values of \( \pi \) and \( h_A \), we cannot compute the exact value of \( v(r^m) \); however, a good estimate of \( v(r^m) \), given that \( \pi \) is in the neighborhood of \( \pi_0 \) and \( h_A \) is in the neighborhood of \( h_0 \), is
\[
v'(r^m) = \min_{r \leq r^m} \left( \frac{\pi_0 + g(r)h_0}{rh_0} \right) - \frac{\pi_0}{\text{Int}[r^m]h_0}
\tag{B.8}
\]

Substituting \( v'(r^m) \) in eq. (B.6) gives an estimate of the probability that the integer solution will be inferior to the real solution:
\[
\Pr \left( n_B^* (r^m) > n_B (r^m) \right) \approx \begin{cases} 
v'(r^m) & \text{if } v'(r^m) < 1 \\
1 & \text{if } v'(r^m) \geq 1 
\end{cases}
\tag{B.9}
\]

Similarly, from eq. (B.5) the probability that the integer solution exceeds the real solution by 2 or more vehicles is approximately
\[
\Pr(n_B^*(r^m) - n_B(r^m) > 1) \approx \begin{cases} 
0 & \text{if } v'(r^m) < 1 \\
\begin{cases} 
v'(r^m) - 1 & \text{if } 1 \leq v'(r^m) < 2 \\
1 & \text{if } v'(r^m) \geq 2 
\end{cases}
\end{cases}
\tag{B.10}
\]

Given values of \( \pi, h_0 \), and \( r^m \), then, equations (B.9) and (B.10) estimate the probability that the integer solution will require at least one vehicle more and at least two vehicles more than the real solution.
So far we have let $\pi$ and $h_A$ be variables while $r^m$ is still assumed to be given. Since we are not solving a problem for a particular route but are trying to estimate probabilities for a large number of routes with different parameters, $r^m$ should be a variable as well. The function $v'(r^m)$ is very sensitive to the proximity of $r^m$ to an integer; for example, if $r^m$ is an integer or just greater than an integer, $v'(r^m)$ will equal zero, while $v'(r^m)$ rises monotonically to a local maximum just before the next integer. Therefore if we want to find the probability of an optimal integer solution existing when $r^m$ is in a certain range, we should define the range of $r^m$ as lying between the two neighboring integers in order to get an estimate that is as precise and unbiased as possible. Thus we will define $r^m$ as being a random variable equally likely to take on any value over the range $k \leq r^m < k + 1$, where $k$ is an integer. Integrating equations (8.9) and (8.10) over this range of $r^m$, we can find the probabilities that an integer solution will require at least one vehicle more and at least two vehicles more than the real solution, given that $\pi$, $h_A$, and $r^m$ are in the neighborhood of $\pi_0$, $h_0$, and $k + 1/2$, respectively.

A pocket calculator program was written to perform the above-mentioned integrations numerically. Taking as input different values of $\pi_0$, $h_0$, and $k$, this program estimated the probabilities that an integer solution with parameters in the corresponding ranges would be optimal, would require at least one vehicle more than the optimal solution, and would require at least two vehicles more than the optimal solution. The results of this program, reported in Table 4.1, indicate that for a large range
of parameter values one can expect integer solutions to be optimal more than 90% of the time. As the maximum headway ratio \( r^m \) increases, so does the likelihood of optimality; as the effective deadhead premium \( \pi \) increases, this likelihood decreases. As the reader will notice, the case study in Chapter 4 required finding optimal headway ratios 8 times, and in only one of those cases did an integer headway ratio solution require more vehicles than a non-integer solution.