TRANSIT NETWORK DESIGN

by

TAL SADEH
S.B., Technion, Israel Institute of Technology
(1984)

Submitted in partial fulfillment
of the requirements for the degree of

Master of Science in Transportation

at the

Massachusetts Institute of Technology
(October, 1987)

Signature of Author ____________________________

Department of Civil Engineering, October 9, 1987

Certified by ____________________________
Nigel H. M. Wilson, Thesis Supervisor

Accepted by ____________________________
Chairman, Departmental Committee on Graduate Students
TRANSIT NETWORK DESIGN

by

TAL SADEH

Submitted to the Department of Civil Engineering on October 9, 1987 in partial fulfillment of the requirements for the degree of Master of Science in Transportation

ABSTRACT

The Transit Network Design problem is probably the single most important component of the transit planning process, and yet it is one for which research results have not been widely implemented. This is mainly due to the high cost involved in implementing the few methods developed so far, and the complexity of the problem.

This work describes the Transit Network Design problem in the context of the transit planning process and summarizes the different approaches that have been proposed for its solution. The problem is formulated as a mathematical program, and a new heuristic approach is proposed and implemented as a route generation algorithm. This algorithm is based on the concept of generating a small set of routes as opposed to selecting the set of routes from a much larger set of feasible routes. A small network example is solved both optimally, using the MINOS nonlinear optimization computer package, and applying the heuristic algorithm. The results are explained and analyzed and recommendations for future research are made.

Thesis supervisor: Nigel H. M. Wilson
Title: Professor of Civil Engineering
ACKNOWLEDGEMENTS

First, I would like to thank my advisors, Nigel Wilson and Avi Ceder, for their excellent advice, helpful guidance and sincere encouragement throughout this work.

I would also like to express my deepest gratitude to my parents, Ruth and Avraham Sadeh, who have always given me endless support and made all this possible.

To my dear friend, Marji Berkman, who spent a few nights typing the most complicated chapter of this thesis, I give my sincere appreciation and a case of dos-equis.

Last but not least, my very much loved wife, Orly. It was her love, understanding and patience that helped me through the tough moments and gave me the encouragement needed to get the job done.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Abstract</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acknowledgements</td>
<td>3</td>
</tr>
<tr>
<td>Table of Contents</td>
<td>4</td>
</tr>
<tr>
<td>List of Figures</td>
<td>7</td>
</tr>
<tr>
<td>Notation</td>
<td>8</td>
</tr>
</tbody>
</table>

**Chapter 1**

<table>
<thead>
<tr>
<th>INTRODUCTION</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 The Transit Planning Process</td>
<td>12</td>
</tr>
<tr>
<td>1.2 The Problem</td>
<td>13</td>
</tr>
<tr>
<td>1.3 Literature Review</td>
<td>14</td>
</tr>
<tr>
<td>1.3.1 Demand</td>
<td>14</td>
</tr>
<tr>
<td>1.3.2 Objective Function</td>
<td>14</td>
</tr>
<tr>
<td>1.3.3 Constraints</td>
<td>15</td>
</tr>
<tr>
<td>1.3.4 Passengers Behavior</td>
<td>15</td>
</tr>
<tr>
<td>1.3.5 Solution Techniques</td>
<td>16</td>
</tr>
<tr>
<td>1.4 The Approach Taken in this Work</td>
<td>17</td>
</tr>
<tr>
<td>1.5 Thesis Contents</td>
<td>18</td>
</tr>
</tbody>
</table>

**Chapter 2**

<table>
<thead>
<tr>
<th>PROBLEM FORMULATION</th>
<th>19</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Notation</td>
<td>19</td>
</tr>
<tr>
<td>2.2 The Solution Structure</td>
<td>21</td>
</tr>
<tr>
<td>2.3 The Objective Function</td>
<td>23</td>
</tr>
<tr>
<td>2.4 The Constraints</td>
<td>28</td>
</tr>
<tr>
<td>2.5 Formulation as an Optimization Problem</td>
<td>31</td>
</tr>
<tr>
<td>2.5.1 The MINOS Optimization Package</td>
<td>31</td>
</tr>
<tr>
<td>2.5.2 Problem Formulation for MINOS</td>
<td>32</td>
</tr>
<tr>
<td>2.5.2.1 The Objective Function</td>
<td>33</td>
</tr>
<tr>
<td>2.5.2.2 The Constraints</td>
<td>33</td>
</tr>
<tr>
<td>2.5.3 Problem Size</td>
<td>36</td>
</tr>
</tbody>
</table>
Chapter 3

PROBLEM FORMULATION ........................................... 39
  3.1 The Structure ................................................. 39
    3.1.1 Part 1 - Generating Shortest Routes ............... 43
    3.1.2 Part 2 - Connecting Disconnected Nodes .......... 44
    3.1.3 Part 3 - Trip Assignment ............................. 47
    3.1.4 Part 4 - Route Network Improvement ............... 49
  3.2 The Route Evaluation Function ............................. 54
  3.3 The Transfer Path Evaluation Function .................... 55
  3.4 The Evaluation Functions Revisited ....................... 57

Chapter 4

ROUTE GENERATION ALGORITHM .................................... 58
  4.1 Input ....................................................... 58
    4.1.1 Network ................................................. 58
    4.1.2 O-D Demand Matrix .................................... 59
    4.1.3 Route Bounding Parameters ........................... 60
  4.2 The Route Generation Algorithm ............................ 61
    4.2.1 Part 1 - Generating Shortest Routes ............... 61
    4.2.2 Part 2 - Connecting Disconnected Nodes .......... 63
    4.2.3 Part 3 - Trip Assignment ............................. 64
    4.2.4 Part 4 - Route Network Improvement ............... 66

Chapter 5

APPLICATION ..................................................... 71
  5.1 The Example Network ....................................... 71
  5.2 The Problem Parameters .................................... 71
  5.3 The Route Generation Algorithm ............................ 74
    5.3.1 Part 1 - Generating Shortest Routes ............... 74
    5.3.2 Part 2 - Connecting Disconnected Nodes .......... 78
    5.3.3 Part 3 - Trip Assignment ............................. 86
    5.3.4 Part 4 - Route Network Improvement ............... 96
  5.4 The MINOS Optimal Solution ................................ 100
  5.5 Comparing the Solutions ................................... 102
<table>
<thead>
<tr>
<th>Chapter 6</th>
<th>SUMMARY, CONCLUSIONS AND FUTURE RESEARCH</th>
<th>106</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1</td>
<td>Summary</td>
<td>106</td>
</tr>
<tr>
<td>6.2</td>
<td>Conclusions and Future Research</td>
<td>108</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>References</th>
<th>109</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Appendix A</th>
<th>DIVIDING THE DEMAND ON A COMMON SEGMENT BETWEEN ROUTES</th>
<th>111</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appendix B</td>
<td>THE ROUTE EVALUATION FUNCTION</td>
<td>120</td>
</tr>
<tr>
<td>Appendix C</td>
<td>DETAILED FORMULATION OF THE NETWORK EXAMPLE FOR MINOS</td>
<td>130</td>
</tr>
<tr>
<td>Figure</td>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>3.1</td>
<td>The structure of the heuristic approach</td>
<td>40</td>
</tr>
<tr>
<td>5.1</td>
<td>The example network</td>
<td>72</td>
</tr>
<tr>
<td>5.2</td>
<td>Route 1 the first active route</td>
<td>75</td>
</tr>
<tr>
<td>5.3</td>
<td>Route 2 when routes 1 and 2 are active</td>
<td>77</td>
</tr>
<tr>
<td>5.4</td>
<td>Route 3 when routes 2 and 3 are active</td>
<td>80</td>
</tr>
<tr>
<td>5.5</td>
<td>Route 2 when routes 2 and 3 are active</td>
<td>81</td>
</tr>
<tr>
<td>5.6</td>
<td>Route 4 when routes 1 and 4 are active</td>
<td>83</td>
</tr>
<tr>
<td>5.7</td>
<td>Route 1 when routes 1 and 4 are active</td>
<td>84</td>
</tr>
<tr>
<td>5.8</td>
<td>Route 1 after dividing the demand between routes 1 and 4</td>
<td>88</td>
</tr>
<tr>
<td>5.9</td>
<td>Route 4 after dividing the demand between routes 1 and 4</td>
<td>89</td>
</tr>
<tr>
<td>5.10</td>
<td>Route 1 after assigning transfers 1 and 2</td>
<td>91</td>
</tr>
<tr>
<td>5.11</td>
<td>Route 4 after assigning transfers 1 and 2</td>
<td>92</td>
</tr>
<tr>
<td>5.12</td>
<td>Route 1 with transfers and demand divided between 1 and 4</td>
<td>94</td>
</tr>
<tr>
<td>5.13</td>
<td>Route 4 with transfers and demand divided between 1 and 4</td>
<td>95</td>
</tr>
<tr>
<td>5.14</td>
<td>Adding route 5 to the network</td>
<td>98</td>
</tr>
<tr>
<td>5.15</td>
<td>The MINOS solution</td>
<td>101</td>
</tr>
<tr>
<td>5.16</td>
<td>Comparing the solutions</td>
<td>103</td>
</tr>
<tr>
<td>B.1</td>
<td>Trade-offs between density and frequency</td>
<td>122</td>
</tr>
<tr>
<td>B.2</td>
<td>Comparison of two routes</td>
<td>126</td>
</tr>
</tbody>
</table>
NOTATION

$G = \langle N, A \rangle$ - A street network with $|N|$ nodes and $|A|$ arcs.
The set of nodes $N$ is divided into two subsets:

$M = \{m\}, m \in N$ - The set of terminal nodes at which a route can start or end.

$S = \{s\}, s \in N$ - The set of service nodes that must be served but are not terminals ($M + S = N$).

$K$ - The set of all feasible routes, $k \in K$.

$y_k$ - A binary variable: $\begin{cases} 1 \text{ if route } k \text{ is active in the solution.} \\ 0 \text{ otherwise.} \end{cases}$

$R$ - The set of all feasible transfer paths, $r \in R$.

$T = \{t_{ij}\}$ - The matrix of travel times [minutes].

$T^m = \{t^m_{ij}\}$ - The matrix of shortest riding times on the network [minutes].

$T_k$ - The total riding time on route $k$ [minutes].

$T_u$ - An upper bound on the route length [minutes].

$T_l$ - A lower bound on the route length [minutes].

$t_{ij}$ - The riding time on arc $i-j$ [minutes].

$t_{ijk}$ - The riding time between nodes $i$ and $j$ on route $k$ [minutes].

$T_{ijk}$ - The total travel time, including both waiting and riding times, between nodes $i$ and $j$ on route $k$ [minutes]. $T_{ijk} = \phi_k w_{ij} + t_{ijk}$.

$t^m_{ij}$ - The shortest possible riding time between nodes $i$ and $j$ on the network [minutes]. This is the riding time from node $i$ to node $j$ if there was a route along the shortest path.

$t^r_{ij}$ - The riding time on transfer path $r$ between nodes $i$ and $j$ [minutes].

$t^r_k$ - The riding time on route $k$ for a passenger using transfer path $r$ [minutes].

$T^r_{ij}$ - The total time spent by a passenger on transfer path $r$ [minutes].

$D = \{d_{ij}\}$ - The O-D demand matrix. This matrix represents the transit
demand between each pair of nodes on the network during a certain time period. For simplicity, but without loss of generality, the time period throughout this work is assumed to be one hour.

\( d_{ij} \) - The total demand between nodes \( i \) and \( j \) on the network [passengers/hour].

\( d_{ijk} \) - The demand between nodes \( i \) and \( j \) satisfied directly by route \( k \) [passengers/hour]. Passengers on transfer paths are not included.

\( d^T_{ij} \) - The portion of the demand between nodes \( i \) and \( j \) satisfied by transfer \( r \) [passengers/hour].

\( d^n_{ijk} \) - The demand between nodes \( i \) and \( j \), satisfied by a transfer via node \( n \) on route \( k \) [passengers/hour].

\( x_{ijk} \) - The number of passengers traveling on arc \( i-j \) on route \( k \), including passengers on transfer paths [passengers/hour].

\( q_k \) - The maximum arc load on route \( k \) [passengers/hour].

\( P_k \) - The total number of passengers on route \( k \) [passengers].

\( PH \) - The total riding time for all passengers on the network [passenger-hours].

\( \Delta PH \) - The total excess riding time, over the shortest possible riding time, spent by all passengers on the network [passenger-hours].

\( \Delta PH^T \) - The difference between the time spent by all passengers using transfer \( r \) from \( i \) to \( j \) and the shortest possible riding time between nodes \( i \) and \( j \) [passenger-hours].

\( \Delta PH_k \) - The difference between the actual riding time spent by all passengers traveling between all node pairs on route \( k \) and the shortest possible path connecting the same nodes. This is the excess time spent on route \( k \) due to the fact that it is not the shortest possible route connecting the same nodes [passenger-hours].

\( WH \) - The total waiting time for all passengers on the network [passenger-hours].

\( UH \) - The total number of unutilized-space-hours on the network [unutilized-space-hours].

\( VH \) - The total number of vehicle-hours required to satisfy all the demand
on the network [vehicle-hours].

PH\textsubscript{k} - the total riding time for all passengers traveling on route k [passenger-hours].

UH\textsubscript{k} - The number of unutilized-space-hours on route k [unutilized-space-hours].

PH\textsuperscript{m} - The minimum possible riding time for all passengers on the network, when each one is traveling on the shortest path between his origin and destination [passenger-hours].

C - Desired maximum vehicle occupancy [passengers/vehicle].

f\textsubscript{k} - The service frequency on route k [vehicles/hour].

f\textsuperscript{F}_{ijl} - The service frequency on the first route on transfer path \( r \) between nodes \( i \) and \( j \) [vehicles/hour].

F\textsuperscript{m}\textsubscript{k} - The minimum required frequency on route k [vehicles/hour].

w\textsubscript{k} - The waiting time per passenger on route k [minutes].

\( \delta_d \) - The maximum allowable excess riding time, over the shortest possible riding time, between two nodes on a direct route, expressed as a decimal fraction of the shortest possible riding time between these nodes on the network plus one.

\( \delta_t \) - The maximum allowable excess travel time, over the shortest possible riding time, between two nodes on a transfer path, expressed as a decimal fraction of the shortest possible riding time between these nodes on the network plus one.

Z\textsubscript{k} - The value of the route evaluation function [hours].

Z - The value of the objective function used to evaluate the quality of the route network [capacity-hours].

Z\textsubscript{u} - Upper bound on the acceptable route evaluation function [hours].

n\textsubscript{r} - The number of routes on transfer path \( r \).

Q - The available fleet size [vehicles].
CHAPTER 1

INTRODUCTION

The transit route design (or redesign) process is often initiated in practice by identifying a single route that has to be changed. Sometimes the changes in a single route will cause changes in interacting routes. Transit properties rarely consider restructuring the overall transit network. In fact, many North American properties have not done an overall re-evaluation of their transit network for more than 50 years. During this period metropolitan areas have grown and experienced redistribution of population and employment. Even though changes have been made the overall network structures usually remained the same and no overall evaluation of the network was undertaken. In many cases there is a need for re-evaluation and possible redesign of the overall network.

Prior work on transit network design includes attempts to develop transit route planning methods. None of these methods however has been accepted and implemented within a planning unit of a transit property, principally due to the complexity and/or high cost of these methods. Consequently there is a need for a simple, efficient transit route planning system that will be understandable and simple enough so that transit properties will accept and implement it.

Even though the costs (both the analysis cost and the data collection effort) involved in the redesign process are high, the benefits from this process may well be worth the effort especially since it will not be done very often.
1.1 The transit planning process

The transit planning process includes five stages with the output of one stage being part of the input to the next stage:

Stage 1: Data collection and demand estimation. In this stage information on the region in which the transit system is to be operated is gathered. This is done for a new transit system as well as for an existing system that has to be evaluated for possible redesign. The road network is established and the O-D demand matrix is estimated.

Stage 2: Transit network design. The input to this stage is the output from stage 1, namely the road network and the O-D demand matrix, and route design and level of service parameters. The output of this stage is a route network (either new or revised routes).

Stage 3: Frequency and timetable determination. The demand estimated in stage 1, and the route network generated in stage 2, are used to determine the service frequencies on each route on the network. The frequencies are then used to create a timetable.

Stage 4: Vehicle scheduling. The departure and arrival times created in stage 3, along with operation parameters, are used to generate the vehicle schedules.

Stage 5: Crew scheduling. Once the vehicle schedules are determined drivers can be assigned and the crew schedules can be determined.

While much work has been done on stages 4 and 5 of this process, the network design stage has not been fully explored, and those methods that
have been developed have not been widely implemented by transit properties. For this reason this work concentrates on the network design stage, in an attempt to develop an algorithm that will be easily accepted and implemented by transit properties.

1.2 The problem

The objective of the transit network design process is to create a set of transit routes that will satisfy the demand for public transportation while maintaining a certain level of service for the passengers, and minimizing the operator's and passengers costs.

The main difficulties in solving the transit network design problem are:

1. The problem is NP-hard meaning that the computation time grows exponentially with the problem size. The computation cost becomes prohibitive when problem size even approaches that of typical transit networks.

2. Due to its complexity, the problem cannot be formulated as a clean and concise mathematical program. Consequently, attempts to solve the problem optimally have not yet been successful.

3. The solution to the problem, a set of transit routes, is highly dependent on the O-D demand matrix, which is one of the inputs to the process. This demand for public transportation varies greatly at different times of the day, different days of the week and throughout the year. This variation in demand makes the planning process even more difficult. Furthermore the estimation of the demand is a complex and expensive process and the results are, many times, far from accurate. As a result the O-D demand matrix, on which the route generation procedure is based, does not always fully represent the actual demand for public transportation.
1.3 Literature review

This section presents prior approaches dealing with actual routes. Other approaches, dealing with idealization of the network [Newell, 1979; Kocur and Hendrickson, 1982], are not discussed here because they are designed primarily for policy analysis. These methods are used for estimating design parameters and not for detailed route network design.


1.3.1 Demand

Most prior approaches, except Dubois, Hasselstrom and Uhry, assumed fixed demand. Hasselstrom generated a desired demand matrix based on an ideal route network with high quality service. This matrix is then updated based on the actual route network. Dubois derived the transit demand matrix from the total trip matrix based on expected transit times. Uhry uses a general demand function in which the number of passengers between each node pair is a function of the travel time between these nodes.

1.3.2 Objective function

Most of the objective functions in prior work are based on minimizing generalized cost [Sharp, Silman, Ceder and Wilson], or generalized time [Lampkin, Rosello, Dubois, Mandl, Sonntag]. Hasselstrom uses the
objective of maximizing consumers' surplus. Rea does not have an explicit objective function. Instead the solution has to satisfy certain constraints including service parameters determined by the operator.

1.3.3 Constraints

Almost all the methods optimize the chosen objective function subject to a constraint on the number of vehicles available. Sharp includes the operator's cost in the objective function, and also constrains the vehicle capacity. This constraint may force passengers to travel on routes that they would prefer not to use, but if the capacity constraint exists, the evaluation of the network will be more realistic. Rea constrains every link in the transit system so that the service frequency on each link is a function of the flow on the link. Sonntag does not consider route frequencies, but imposes an upper bound on the sum of the route lengths.

1.3.4 Passenger behavior

Some methods use the system optimizing approach, where passengers are assumed to behave in a way that optimizes the system performance. This means that some passengers will be traveling on routes that are not best for them. This approach [Sharp], requires special measures, such as pricing, to motivate the passengers to use those routes. This pricing system presents many practical difficulties, like determining the prices, and is very hard to implement. The user optimizing approach, taken in all the other methods, assumes that all passengers are maximizing their own benefit, without considering the network as a whole.

Multiple path assignment is used in most approaches. This means that passengers that can take more than one route from origin to destination are divided among all acceptable routes. Most methods assign passengers
to these routes according to the probability of the passengers choosing each route. These probabilities are estimated under the assumption that a passenger will take the first vehicle to arrive [Lampkin, Silman, Dubois] or calculated as a function of the general cost of the route [Rosello].

1.3.5 Solution techniques

Because of the complexity of the complete problem all the methods divide the problem into two parts, route generation and frequency determination. One approach [Lampkin, Silman, and Dubois] is to build skeleton routes with a few major nodes between pairs of acceptable terminals. Each route is defined later by adding nodes between these major nodes. The selection of routes to be included in the set of routes is based on the marginal contribution of each route to the objective function. Some methods generate more routes than will actually be activated [Silman]. Routes are then added or deleted and the final route set is determined by the frequency determination process. Mandl and Rea look at individual links first and build routes out of the acceptable ones. Hasselstrom assigns desired trips to all possible links on the network. The routes are built using those links according to parameters like passengers flow, route length and terminal locations. Ceder and Wilson generate a large set of routes, with a restriction on the route length that reduces the number of feasible routes to be considered.

Of all the aforementioned methods, only the work done by Hasselstrom has been extensively implemented within the Volvo transit planning package. While Hasselstrom’s work is the most sophisticated and comprehensive of all the approaches suggested so far, it is also expensive to implement in terms of the data needed and the effort involved in its execution. Thus, there is a need for an efficient algorithm for solving the Transit Network Design problem that will be acceptable and inexpensive to imple-
ment by transit agencies.

1.4 The approach taken in this work

The demand matrix is assumed to be fixed and independent of the network design. This is both because there is no appropriate, proven demand model for public transit network design, and in order to reduce the complexity of the problem. The demand, with the new network, can be re-estimated and the network design algorithm can be applied again if the new demand matrix is significantly different from the previous one. The design is based on the demand matrix for the peak period, and the network can be adjusted to other time periods by changing the route frequencies and making minor changes to routes.

The objective function is based on minimizing both the passengers and the operator's costs. The passengers' cost is expressed in terms of the waiting time, and the excess time, over the shortest riding time, spent on the network. The operator's cost is assumed proportional to the required vehicle hours. The objective function is constrained to satisfy all the demand, by a number of level of service constraints and by the total fleet size.

The passengers are assumed to minimize their cost on the network. Passengers that can travel on more than one route (or transfer path) between origin and destination, are divided among the appropriate routes according to the service frequencies on the routes. This is done in order to make the design process as realistic as possible.

Most of the aforementioned approaches are route selection methods in which the routes are selected from a large set of routes. The approach in this work is focusing on generating a small set of shortest routes, one route at a time. Once the initial set of routes is generated, an attempt is made to improve the network by altering the routes.
1.5 Thesis contents

Chapter 2 presents a formulation of the Transit Network Design problem as a mathematical program. This formulation is used to demonstrate the size of the problem, and to solve an example network problem both optimally and by applying a heuristic algorithm.

Chapter 3 describes the approach taken in this work to the solution of the Transit Network Design problem, and discusses its properties. The heuristic route generation algorithm is outlined and explained in detail.

Chapter 4 presents the heuristic approach to the transit route generation problem. The route generation algorithm is stated in a step by step manner.

Chapter 5 contains solutions of a small network example. The application of the heuristic algorithm is detailed, the optimal solution obtained by a non-linear optimization computer package is presented and the solutions are compared.

Chapter 6 includes a summary of the work as well as conclusions and recommendations for future research.

It is hoped that this work will be the first step towards the development of a heuristic approach for the solution of the Transit Network Design problem that will be accepted and implemented by transit properties.
CHAPTER 2

PROBLEM FORMULATION

This chapter presents a formulation of the Transit Network Design problem as a mathematical program. The chapter begins with a description of the notation used in the formulation. Section 2.2 discusses the first part of the solution process in which a set of routes, $K$, is generated and a set of transfer paths, $R$, is defined using these routes. The objective function and the constraints, which are used in the second part of the solution process to choose the best subset of routes, are discussed in sections 2.3 and 2.4 respectively. The chapter ends with a discussion on the formulation and size of the Transit Network Design problem as an optimization problem.

The objective function is based on the concept of minimizing the total cost (expressed in time units) to both the passengers and the operator. The solution is constrained to satisfy all the demand and meet several level of service constraints.

2.1 Notation

$K$ - The set of all feasible routes, $k \in K$.

$y_k$ - A binary variable: $\begin{cases} 1 & \text{if route } k \text{ is active in the solution.} \\ 0 & \text{otherwise.} \end{cases}$

$R$ - The set of all feasible transfer paths, $r \in R$.

$T_k$ - The total riding time on route $k$ [minutes].

$T_u$ - An upper bound on the route length [minutes].
$T_1$ - A lower bound on the route length [minutes].

$t_{ij}$ - The riding time on arc i-j [minutes].

$t_{ijk}$ - The riding time between nodes i and j on route k [minutes].

$T_{ijk}$ - The total travel time, including both waiting and riding times, between nodes i and j on route k [minutes]. $T_{ijk} = \phi^*_w + t_{ijk}$.

$t_{ij}^m$ - The shortest possible riding time between nodes i and j on the network [minutes]. This is the riding time from node i to node j if there was a route along the shortest path.

$t_{ij}^r$ - The riding time on transfer path r between nodes i and j [minutes].

$w_k$ - The waiting time per passenger on route k [minutes].

$x_{ijk}$ - The number of passengers traveling on arc i-j on route k, including passengers on transfer paths [passengers/hour].

$d_{ij}$ - The total demand between nodes i and j on the network [passengers/hour].

$d_{ijk}$ - The demand between nodes i and j satisfied directly by route k [passengers/hour]. Passengers on transfer paths are not included.

$d_{ij}^r$ - The portion of the demand between nodes i and j satisfied by transfer r [passengers/hour].

$d_{ijk}^n$ - The portion of the demand between nodes i and j satisfied by a transfer via node n on route k [passengers/hour]. This variable is a simplified version of $d_{ijk}^r$ and is used in the MINOS formulation.

$q_k$ - The maximum arc load on route k [passengers/hour].

$P_k$ - The total number of passengers on route k [passengers].

$PH$ - The total riding time for all passengers on the network [passenger-hours].

$\Delta PH$ - The total excess riding time, over the shortest possible riding time, spent by all passengers on the network [passenger-hours].

$WH$ - The total waiting time for all passengers on the network [passenger-hours].

$UH$ - The total number of unutilized-space-hours on the network [unutilized-space-hours].

$VH$ - The total number of vehicle-hours required to satisfy all the demand
on the network [vehicle-hours].

$\text{PH}_k$ - the total riding time for all passengers traveling on route $k$ [passenger-hours].

$\text{UH}_k$ - The number of unutilized-space-hours on route $k$ [unutilized-space-hours].

$\text{PH}_k^{\text{m}}$ - The minimum possible riding time for all passengers on the network, when each one is traveling on the shortest path between his origin and destination [passenger-hours]:

$C$ - Desired maximum vehicle occupancy [passengers/vehicle].

$f_k$ - The service frequency on route $k$ [vehicles/hour].

$f_{ij1}^r$ - The service frequency on the first route on transfer path $r$ between nodes $i$ and $j$ [vehicles/hour].

$F_k^{\text{m}}$ - The minimum required frequency on route $k$ [vehicles/hour].

$\delta_d$ - The maximum allowable excess riding time, over the shortest possible riding time, between two nodes on a direct route, expressed as a decimal fraction of the shortest possible riding time between these nodes on the network plus one.

$\delta_t$ - The maximum allowable excess travel time, over the shortest possible riding time, between two nodes on a transfer path, expressed as a decimal fraction of the shortest possible riding time between these nodes on the network plus one.

$n_r$ - The number of routes on transfer path $r$.

$Q$ - The available fleet [vehicles].

2.2 The solution structure

The solution of the Transit Network Design problem as a mathematical program is a route selection process which can be divided into two parts. In part one a set of feasible routes between terminal nodes is generated and a set of feasible transfer paths defined. The generation of routes can be done in different ways:

1. Consider all possible paths between all terminal nodes.
2. Consider, initially, only the shortest paths between terminal nodes.

3. Consider more than one path between each pair of terminal nodes but not all the paths (for example the shortest and the second and third shortest paths).

These methods for generating routes can be done manually by the planner, or a computer may be used to generate the set of routes. The general formulation presented in this chapter is independent of the method by which the set of routes is generated, and is in fact a route selection tool. The heuristic approach, that will be presented in chapter 3, includes both route generation and route selection procedures.

A feasible route is one that satisfies the following constraints:

1) Route length is within the route length bounds:

\[ T_l \leq T_k \leq T_u \quad \forall \ k \in K \] \hspace{1cm} [2.1]

where \( T_l \) and \( T_u \) are the lower and upper bounds on route length, and \( T_k \) is the length of route \( k \) [minutes].

2) The excess riding time, over the shortest possible riding time, between all node pairs, served directly with a single route, is below the maximum allowable excess direct riding time, \( \delta_d \):

\[ t_{ijk}/t_{mij} \leq \delta_d \quad \forall \ (i,j) \in N \] \hspace{1cm} [2.2]

where \( t_{ijk} \) is the riding time between nodes \( i \) and \( j \) on route \( k \) [minutes], and \( t_{mij} \) is the shortest possible riding time between nodes \( i \) and \( j \) on the network [minutes].
A transfer is considered initially feasible if the number of routes included in it is not greater than the maximum number of routes allowed. In order to simplify the discussion, but without loss of generality, a transfer path can include up to three different routes, meaning that a passenger is allowed to switch routes twice. A constraint on the length of the transfer paths is introduced later when the service frequencies on the routes are determined and the mean waiting time on each route is calculated.

The objective function is then used to select the best subset of routes out of the initially feasible routes, subject to satisfying all the demand and a number of level of service constraints described in section 2.4.

2.3 The objective function

The general objective adopted here is to minimize the total cost, including both the passengers and operator's cost. The objective function is comprised of three components: the first two are measures of passenger costs, namely riding and waiting times, and the third component is a measure of the operator's cost or the efficiency of the route network in terms of equipment utilization. All three components are expressed in time units.

The components of the objective function are:

1. The total excess time, over the shortest possible riding time, spent on the network by all passengers, \( \Delta PH \), [passenger-hours]. This is the difference between the actual riding time spent by all passengers on the network, and the shortest possible riding time the passengers would experience had each one of them been able to travel along the shortest path between origin and destination:
\[ \Delta P_H = \sum_{k \in K} P_{Hk} - P_{Hm} \]  \hspace{1cm} [2.3]

where \( K \) is the set of all feasible routes, and \( P_{Hk} \) is the total riding time spent by all passengers on route \( k \) [passenger-hours], defined as:

\[ P_{Hk} = (1/60) \sum_{(i,j) \in k} (x_{ijk} \cdot t_{ij}) \]  \hspace{1cm} [2.4]

where \( x_{ijk} \) is the number of passengers traveling on arc \( i-j \) on route \( k \), including passengers on transfer paths, and \( t_{ij} \) is the riding time on arc \( i-j \) [minutes]. The expression \( (i,j) \in k \) means "...over all arcs \( i-j \) on route \( k \)".

The term \( P_{Hm} \) represents the minimum possible riding time spent by all passengers on the network, when each one travels on the shortest path between his origin and destination [passenger-hours]:

\[ P_{Hm} = (1/60) \sum_{(i,j) \in N} d_{ij} \cdot t_{mij} \]  \hspace{1cm} [2.5]

where \( d_{ij} \) is the total demand between nodes \( i \) and \( j \) [passengers], \( t_{mij} \) is the shortest possible riding time between nodes \( i \) and \( j \) on the network, and the expression \( (i,j) \in N \) means "...over all \( (i,j) \) node pairs on the network".

2. The total waiting time for all passengers, \( WH \), [passenger-hours]:

\[ WH = (1/60) \sum_{k \in K} P_k \cdot w_k \]  \hspace{1cm} [2.6]

where \( w_k \) is the waiting time on route \( k \) [minutes], and \( P_k \) is the total number of passengers on route \( k \) [passengers], defined as:
\[ P_k = \sum_{(i,j) \in k} \left( d_{ijk} + \sum_{r \in k} d^r_{ij} \right) \]

where \( d^r_{ij} \) is the portion of the demand between nodes \( i \) and \( j \) satisfied by transfer \( r \) [passengers/hour]. The expression \( r \in k \) means "...over all transfer paths that include route \( k \)".

3. The measure of the operator's cost used here is the total number of vehicle-hours required to satisfy all the demand on the network, \( VH \), defined as:

\[ VH = \left( \frac{1}{60} \right) \sum_{k \in K} f_k T_k \]

where \( T_k \) is the total riding time on route \( k \) [minutes], and \( f_k \) is the service frequency on route \( k \) [vehicles/hour], defined as:

\[ f_k \geq \max \left( \frac{q_k y_k}{C}, \frac{F^m_k y_k}{C} \right) \]

where \( y_k \) is a binary variable indicating whether route \( k \) is active in the solution \((y_k=1)\) or not \((y_k=0)\), \( C \) is the desired maximum vehicle occupancy [passengers/vehicle], \( F^m_k \) is the minimum required frequency on route \( k \) [vehicles/hour], and \( q_k \) is the maximum arc load on route \( k \) [passengers/hour], defined as:

\[ q_k = \max \left( \sum_{(i,j) \subseteq k} x_{ijk} \right) \]

The term \( VH \) is actually the capacity of the network, defined as the sum of the utilized capacity [passengers-hours] and the unutilized capacity [space-hours], divided by the desired maximum vehicle occupancy. The capacity of the network can be expressed as:
\[
(\frac{1}{60})C \sum_{k \in K} f_k T_k = \sum_{k \in K} PH_k + \sum_{k \in K} UH_k \quad [2.11]
\]

where \(PH_k\) is the utilized capacity on route \(k\) defined in equation [2.4] [passenger-hours], and \(UH_k\) is the unutilized capacity on route \(k\) defined as the difference between the capacity of the route [space-hours], and the actual load on the route [passenger-hours]:

\[
UH_k = (\frac{1}{60})C f_k T_k - PH_k \quad [2.12]
\]

Substituting equation [2.11] into equation [2.8] the term vehicle-hours becomes:

\[
VH = (\frac{1}{C})(\sum_{k \in K} PH_k + \sum_{k \in K} UH_k) \quad [2.13]
\]

Since \(C\) is a constant and the sum of \(PH_k\) over all \(k \in K\) is already minimized in the first component of the objective function, \(\Delta PH\), the vehicle-hours term can be included in the objective function by adding the term \(UH\) as the third component and changing the weight for the first term, so that it represents part of the operator's cost. Thus, the third component in the objective function is the total number of unutilized-space-hours on the network:

\[
UH = (\frac{1}{60})C \sum_{k \in K} f_k T_k - \sum_{k \in K} PH_k = \sum_{k \in K} [(\frac{1}{60})C f_k T_k - PH_k] = \sum_{k \in K} UH_k \quad [2.14]
\]

The objective function to be minimized is the weighted sum of these components:

\[
Z = \omega \Delta PH + \beta WH + \gamma VH
\]

\[
= \omega (\sum_{k \in K} PH_k - PH^m) + \beta (\sum_{k \in K} P_k w_k) + \gamma (\sum_{k \in K} UH_k + \sum_{k \in K} PH_k) \quad [2.15]
\]
where \( \omega, \beta \) and \( \gamma \) are the weights for \( \Delta PH, WH \) and \( VH \) respectively.

Since the term \( PH^m \) is fixed for the network it can be dropped from the objective function. The first and third components can be modified so that the weight on the first component includes part of the operator's cost, namely the utilized capacity \( PH \). Thus, the objective function used in this work is:

\[
Z = \alpha \Delta PH + \beta WH + \gamma UH \\
= \alpha \sum_{k \in K} PH_k + \beta \sum_{k \in K} (1/60)P_kV_k + \gamma \sum_{k \in K} UH_k
\]

[2.16]

where \( \alpha \) is the modified weight for the first component, \( \alpha = \omega + \gamma \).

The weights are included in the objective function for the following reasons:

1. Different components of the objective function may have different significance in the network evaluation process. Waiting time, for example, may be more important to passengers than riding time, so it should receive more weight in the function. Similarly, the operator's measure, \( UH \), may be of greater importance than both waiting and riding times. In order to capture these trade-offs between the components, weights, to be determined by the operator, must be included in the objective function.

2. Each component in the objective function represents costs to the passengers, the operator, or both. The first component represents both the passengers (who would like to minimize their excess riding time) and the operator (because this component together with the third component represents the operator's cost). The second component is another measure of the passengers cost (waiting time), and the third
component is another measure of the operator's cost since it is part of the expression representing the total number of vehicle-hours required on the network.

2.4 The constraints

1) The demand between nodes i and j satisfied directly by route k is determined by the frequency on route k according to the frequency share model:

\[ d_{ijk} = \left( \frac{f_k}{\sum_{k^*} f_k} \right) * d_{ij} \quad [2.17] \]

where the expression \( k^*(i,j) \) means "...over all routes that provide acceptable direct service between nodes i and j".

The frequency share model is based on the assumption that a passenger will board the first bus to arrive, on any of the acceptable routes in the set \( K \), that provide service between his origin and destination. A passenger, traveling from node i to j, will take the first bus to arrive if its riding time is shorter than the total travel time on the best route (the one that will minimize his expected arrival time) that provides service between nodes i and j. The frequency share model is valid when the service is purely random with bus arrivals as in a Poisson process. It is also based on the assumption that all passengers can board the first arriving bus.

2) The portion of the demand between nodes i and j satisfied by transfer path r is determined by the service frequency on the first route on the transfer path, according to the frequency share model:

\[ d^r_{ij} = \left[ f^r_{ij1} * \frac{y_k}{\sum_{k^*} (f^r_{ij1} * y_k)} \right] * d_{ij} \quad [2.18] \]
where \( d_{ij}^r \) is the portion of the demand between nodes \( i \) and \( j \) satisfied by transfer path \( r \). The term \( f_{ij}^r \) is the service frequency on the first route of transfer \( r \) between nodes \( i \) and \( j \). The expression \( r=_{ij} \) means "...over all acceptable transfer paths that provide service between nodes \( i \) and \( j \)." The product of all \( y_k \)'s that belong to the transfer path is needed here to make sure that only transfer paths on which all the routes are active will be considered.

The frequency share model is used here, because it is assumed that a passenger waiting for a bus on the first route of a transfer path, will board the first bus to arrive on an acceptable transfer path that provides service between his origin and destination.

3) All the demand between each pair of nodes is satisfied:

\[
\sum_{k \in K} d_{ijk} + \sum_{r=_{ij}} d_{ij}^r = d_{ij} \quad \forall (i,j) \in N \tag{2.19}
\]

4) The excess travel time (both waiting and riding times), over the shortest possible riding time, between all node pairs, served with a transfer, is below the maximum allowable excess transfer time, \( \delta_t \):

\[
t_{ij}^r/t_{ij}^m \leq \delta_t \quad \forall r \in R \tag{2.20}
\]

where \( t_{ij}^r \) is the travel time between nodes \( i \) and \( j \) on transfer path \( r \) [minutes].

5) The frequency on every route is determined by the maximum load arc or the policy headway, whichever yields the higher frequency:

\[
f_k \geq \max \{ q_k y_k / C, F_k^m y_k \} \quad \forall k \in K \tag{2.21}
\]
This constraint means that the level of service can be higher than the one required by the load provided that it minimizes the value of the objective function.

6) The flow of passengers on arc \( i-j \) on route \( k \) is the sum of all the passengers, including transfer passengers, traveling on this arc on route \( k \):

\[
x_{ijk} = \sum_{(s,t) \in (i-j)} \left( d_{stk} + \sum_{r \in k} d_{rst} \right)
\]

where the expression \( (s,t) \in (i-j) \) means "...over all the paths between nodes \( s \) and \( t \) that include the arc \( i-j \)".

7) Maximum arc load constraint:

\[
q_k = \max_{(i-j) \in k} \{ x_{ijk} \}
\]

8) Fleet size constraint. In this constraint the required fleet size is estimated to be the total number of vehicle-hours, \( f_k T_k \). This is multiplied by 2 to capture the fact that there is demand in both directions:

\[
(2/60) \sum_{k \in K} f_k T_k - Q \leq 0
\]

where \( Q \) is the available fleet expressed in vehicle-hours.

9) Nonnegativity constraints:

\[
y_k = \begin{cases} 
1 & \text{if route } k \text{ is active in the solution.} \\
0 & \text{otherwise.}
\end{cases}
\]
\[ d_{ijk} \geq 0 \quad \forall (i,j) \in N \quad \text{and} \quad \forall k \in K \quad [2.26] \]

\[ d^{r}_{ij} \geq 0 \quad \forall r \in R \quad [2.27] \]

\[ f_k \geq 0 \quad \forall k \in K \quad [2.28] \]

\[ x_{ijk} \geq 0 \quad \forall (i,j) \in N \quad [2.29] \]

\[ q_k \geq 0 \quad \forall k \in K \quad [2.30] \]

2.5 Formulation as an optimization problem

In order to demonstrate the complexity and size of the Transit Network Design problem and to test the efficiency of the heuristic algorithm, a small example was solved both optimally, using a non-linear optimization computer package (MINOS), and by applying the heuristic. This section starts with a short description of the MINOS computer package. Next the MINOS formulation of the Transit Network Design problem, is presented followed by a discussion on the problem size. The optimal solution of a small network problem is presented in appendix C and compared with the heuristic results in chapter 5.

2.5.1 The MINOS optimization package

MINOS (Modular In-core Nonlinear Optimization System) is a Fortran-based computer program developed at Stanford University by Bruce Murtagh and Michael Saunders to solve large-scale optimization problems.

The Transit Network Design problem contains a nonlinear objective function and both linear and nonlinear constraints. In this case MINOS uses a projected augmented Lagrangian algorithm with the solution
process summarized as follows. A sequence of major iterations is performed, each one requiring the solution of a linearly constrained subproblem. The subproblems contain the original linear constraints and bounds, as well as linearized versions of the nonlinear constraints.

MINOS is designed to find solutions that are locally optimal. The nonlinear functions in a problem must be smooth (have first derivatives) but need not be separable. Integer restrictions cannot be imposed directly.

2.5.2 Problem formulation for the MINOS package

The formulation starts from the general Transit Network Design problem formulation presented earlier in this chapter. The following sections describe the modifications necessary for the MINOS package. These modifications are needed mainly because it is not possible to formulate the following constraint in MINOS:

\[ q_k = \max \{x_{ijk}\} \quad (i-j) \in K \]

Since \( q_k \) can not be constrained to exactly equal the maximum arc load, the variable \( x_{ijk} \) is not needed. Instead \( q_k \) is constrained to be at least as large as the maximum arc load on route \( k \). This means that \( q_k \) can be assigned values that are higher than the actual maximum arc load as long as the other constraints are satisfied and the objective function minimized. The other modifications are due to implementation issues. Some of the constraints (for example constraint 3) are formulated slightly different because the small size of the example network allows for somewhat simpler formulation. This is done without changing the spirit of the original general formulation. The following section states the objective function and the constraints for the MINOS opti-
mization formulation.

2.5.2.1 The objective function

The objective function is expressed in terms of $f_k$:

$$Z = \alpha \sum_{k \in K} \sum_{(i,j) \in k} d_{ijk} t_{ijk} + \beta \sum_{k \in K} \left( \frac{1}{2f_k} \right) \sum_{i,j \in k} d_{ijk} + \gamma \left( \frac{1}{60} \right) \sum_{k \in K} f_k C_k T_k$$  \[2.31\]

2.5.2.2. The constraints

The input to the MINOS package is a set of all feasible routes, that is routes that have riding time between the upper and lower bounds. Since the package cannot impose integer restrictions, the $y_k$ variables are allowed, initially, to have values between 0 and 1, and an additional set of constraints on the demand forces the $y_k$'s to be either 0 or 1 exactly. The following constraints are imposed in MINOS with constraint numbers corresponding to the similar constraints in the general formulation, except as specified.

1) The demand between nodes $i$ and $j$ satisfied directly by route $k$ is determined by the frequency on route $k$ according to the frequency share model:

$$d_{ijk} = \frac{f_k y_k}{\sum_k f_k y_k} d_{ij}$$  \[2.32\]

2) The portion of the demand between nodes $i$ and $j$ satisfied by transfer path $r$ is determined by the service frequency on the first route on the transfer path, according to the frequency share model:
\[ d_{ijk}^n = \left[ f_{ijl}^r y_k / \sum_{k \in \mathcal{R}} (f_{ijl}^r y_k) \right] d_{ij} \]  

where \( d_{ijk}^n \) is the portion of the demand between nodes \( i \) and \( j \) satisfied by a transfer path via node \( n \) on route \( k \). The term \( f_{ijl}^r \) is the frequency on the first route of transfer \( r \) between nodes \( i \) and \( j \).

3) Since the example network is small, the number of routes on any transfer path is limited to two. This constraint requires that the number of passengers leaving node \( i \), on all active transfer paths from \( i \) to \( j \) via node \( n \), is equal to the number of passengers arriving at node \( j \) on all the transfer paths between nodes \( i \) and \( j \):

\[ \sum_{kl \in (i,n)} d_{ijk}^n y_k = \sum_{k2 \in (n,j)} d_{ijk}^n y_k \]  

where the expression \( kl \in (i,n) \) means "...over all routes that provide service between nodes \( i \) and \( n \) (which is the transfer node)". This constraint simply means that a passenger waiting at the transfer point, \( n \), for the second part of his transfer trip, can travel on any route that provides this service.

The following constraints 4 and 5 correspond to constraint 3 in the general formulation.

4) Demand constraints (without \( y_k \)'s) to make sure all the demand is satisfied:

\[ \sum_{k \in \mathcal{K}} d_{ijk} + \sum_{kl \in (i,n)} d_{ijk}^n = d_{ij} \quad \forall (i,j) \in \mathcal{N} : i \neq j \]  

[2.35]
5) Demand constraints (with \( y_k \)'s to make sure they are 0 or 1):

\[
\sum_{k \in K} d_{i,j} y_k + \sum_{k \in K} \sum_{l \neq (i,n)} d_{i,l} y_k = d_{i,j} \quad \forall (i,j) \in \mathcal{N} : i \neq j \quad [2.36]
\]

6) Minimum arc load constraints. The maximum arc load on each route, expressed as a function of the service frequency on the route, \( q_k = f_k \times C \), should be exactly equal to the highest load on any arc on the route. This can not be achieved in the MINOS package because it is not known in advance which arc will have the highest load, and it is not possible to constraint the maximum arc load to be equal to the load on more than one arc at the same time. Instead these constraints make sure that the maximum arc load on any route is at least equal to the load on any arc on the route. This means that the package can assign to \( f_k \), values that are higher than the minimum service frequency required due to the actual maximum arc load on the route, if such values will decrease the value of the objective function. In this case the level of service on route \( k \) will be higher:

\[
f_k - \frac{1}{C} \sum (d_{stk} + d_{stk}^n) \geq 0 \quad \forall \text{ arc } i \rightarrow j \in k \text{ and } \forall k \in K \quad [2.37]
\]

Constraint 6 replaces constraint 7 in the general formulation.

7) Minimum frequency constraints to make sure some required minimum level of service is provided:

\[
f_k - F_{mk}^\text{min} y_k \geq 0 \quad \forall k \in K \quad [2.38]
\]

This constraint replaces constraint number 5 in the general formulation.
8) Fleet size constraint. In this constraint the required fleet size is estimated to be the total number of vehicle-hours, $f_k T_k$. This is multiplied by 2 to capture the fact that there is demand in both directions:

$$\left(\frac{2}{60}\right) \sum_{k \in K} f_k T_k - Q \leq 0 \quad [2.39]$$

where $Q$ is the available fleet expressed in vehicle-hours.

9) Nonnegativity constraints:

$$y_k = \begin{cases} 
1 & \text{if route } k \text{ is active in the solution.} \\
0 & \text{otherwise.} 
\end{cases} \quad [2.40]$$

$$d_{ijk} \geq 0 \quad \forall (i,j) \in N \text{ and } \forall k \in K \quad [2.41]$$

$$d_{ijk}^n \geq 0 \quad \forall (i,j) \in N \text{ and } \forall k \in K \quad [2.42]$$

$$f_k \geq 0 \quad \forall k \in K \quad [2.43]$$

Constraint 4 in the general formulation is not imposed here because of the small size of the example network. It is assumed that all transfer paths with two routes satisfy the transfer path excess time constraint.

2.5.3 Problem size

The unknown variables in this program are $d_{ijk}$, the demand between nodes $i$ and $j$ satisfied by route $k$, $d_{ij}^n k$, the transfer demand between nodes $i$ and $j$ via node $n$ satisfied by route $k$, $y_k$, the variable that indicates whether route $k$ is in the solution, and $f_k$, the service frequency on route $k$. 

36
The number of variables depends on the number of routes $k$ and the number of origin destination pairs. For a network with $N$ nodes the number of O-D pairs is $(1/2) N(N-1)$. The number of possible routes on the network is assumed here to be the number of terminal pairs. Even though this assumption is very restrictive, and in practice the number of routes is likely to be higher, the size of the problem is demonstrated here to be very large. For a network where half the nodes can be terminals the number of possible routes, $K$, is:

$$K = (1/2)T(T-1) = (1/2)(N/2)[(N/2)-1] = (1/8)N(N-2) \quad [2.44]$$

It is assumed that a typical route serves, on the average, not more than 10 nodes in large networks, and half the number of nodes in the network in small ones.

In addition to the direct $d_{ijk}$ variables there are O-D pairs that can be served via transfers. The number of such variables depends on the number of possible transfers which is assumed to be half the number of routes. Since each transfer requires at least two variables, in case of a transfer with two routes, it is like having $K$ transfers. Each transfer is assumed to have 5 nodes on each side, a total of 20 variables per transfer and $20K$ variables in total.

Since there are $K$ routes and each one has an $f_k$ variable, there will be additional $K f_k$ variables, and $K y_k$ variables, a total of $2K$ variables.

The following table shows the increase in the number of variables, $V$, as the network becomes large:
<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>10</td>
<td>50</td>
<td>100</td>
<td>500</td>
</tr>
<tr>
<td>K</td>
<td>10</td>
<td>300</td>
<td>1225</td>
<td>31125</td>
</tr>
<tr>
<td>D</td>
<td>100</td>
<td>13500</td>
<td>55125</td>
<td>1400625</td>
</tr>
<tr>
<td>Q</td>
<td>20</td>
<td>600</td>
<td>2450</td>
<td>62250</td>
</tr>
<tr>
<td>T</td>
<td>200</td>
<td>6000</td>
<td>24500</td>
<td>622500</td>
</tr>
<tr>
<td>V</td>
<td>320</td>
<td>20100</td>
<td>82075</td>
<td>2085375</td>
</tr>
</tbody>
</table>

where \( D \) is the number of direct \( d_{ijk} \) variables \( D = 45 \times K \) in large networks when \( N > 10 \) or \( D = 10 \times K \) when \( N \leq 10 \), \( T \) is the number of \( d_{ijk}^n \) variables using transfer, \( T = 20 \times K \), and \( Q \) is the number of \( f_k \) and \( y_k \) variables, \( Q = 2 \times K \). \( V \) is the total number of variables and is the sum of \( D \), \( Q \) and \( T \).

It is clear that when the number of variables is growing this way with the problem size, the computation cost becomes prohibitive very soon. The solution time is discussed in chapter 5. A heuristic approach is taken in this work in an attempt to reduce the problem size while achieving a good solution. This approach is described in chapters 3 and 4. A simplified version of this MINOS formulation is presented in appendix C for a small network example.
CHAPTER 3

THE STRUCTURE AND PROPERTIES OF THE HEURISTIC APPROACH

This chapter describes the approach taken in this work to the solution of the Transit Network Design problem, and discusses its properties. The structure of the route generation algorithm is described first in section 3.1, followed by a discussion of the route evaluation function, which is used as a measure of route quality, in section 3.2. The chapter ends with a description of the transfer path evaluation function which is used to rank the transfer paths in decreasing order of quality.

The general approach is to generate a relatively small set of good routes, one route at a time, as opposed to considering all possible routes on the network. Routes are generated by considering shortest routes between pairs of terminal nodes having the highest demand. The shortest route generation procedure is repeated until all the demand is satisfied or until all terminal node pairs have been considered. After the initial set of shortest routes is generated, some nodes may still be disconnected from the network. These nodes are connected to the network so that all the demand can be satisfied. Next transfer trips are assigned where no direct route exists. The final phase of the algorithm is a procedure to improve the route network, with all the demand assigned to it.

3.1 The structure

The algorithm consists of four major parts, shown in figure 3.1.
part 1: generating shortest routes
1. choosing a pair of terminal nodes.
2. choosing a path between the nodes.
3. shortest route activation procedure.

part 2: connecting disconnected nodes
4. connecting by adding a node.
5. connecting by replacing a node.

part 3: trip assignment
6. dividing the flow on common segments.
7. transfer trip assignment.

part 4: route network improvement
8. replacing transfer paths with direct service.
9. adding a node to a route.
10. replacing a node on a route.

figure 3.1: the structure of the heuristic approach
In the first part feasible shortest routes between terminal nodes are evaluated, using the evaluation function described in section 3.2, and the best ones are activated. This is done iteratively considering one route at a time. Once a route is accepted, the demand satisfied by it is deleted from the O-D demand matrix. This route by route generation approach is taken in an attempt to reduce the size of the problem by not considering all possible routes on the network, but rather looking at a much smaller subset of good routes. As a result, the final route network will depend on the order in which routes are examined. The routes considered first are likely to be assigned more demand than those examined later, and therefore are more likely to be accepted. For example, the first route that provides service between a pair of nodes \((i,j)\) will be assigned all the demand between these nodes. If later another route also provides service between the same pair of nodes, it will be assigned only half the demand between them, and the other half will be satisfied by the first route. The first route will not be re-evaluated until a later stage in the algorithm. It is possible to use different route examination strategies to generate different initial route sets from which the best one can be chosen. Different route examination strategies can include considering the highest demand generating nodes first, or any other method of choosing a pair of terminal nodes.

Since the first part of the algorithm deals only with shortest routes between terminals, it is likely that after this phase there will be some nodes that are disconnected from the route network. This means that these nodes are not served by any of the shortest routes accepted in part 1. In the second part of the algorithm these nodes are connected to the route network.

In the third part of the algorithm, acceptable transfer paths between nodes that have no direct service are found, based on the total expected travel time on the transfer path (where travel time includes
both waiting and riding times). The demand between each pair of nodes is then divided among all acceptable transfer paths serving these nodes, according to the frequency share model, considering the frequency of the first route on each transfer (see section 3.1.3 for a detailed discussion of this procedure).

The fourth part of the algorithm includes route improvement procedures. The set of routes generated in part 1, and altered in part 2 to include any disconnected nodes, is improved by adding or replacing nodes. This procedure has two purposes:

1. To try and provide more direct service by altering the routes in a way that will eliminate the need for some of the transfer trips. This will reduce the waiting and riding times passengers experience on the network, and may reduce the number of vehicles required to satisfy the demand.

2. To try and alter the routes so that the operator's cost is reduced more than the level of service, or the level of service is increased more than the operator's cost.

The route improvement procedure could also be performed in part 3, instead of part 4, before the transfer trip assignment. In this case the network would not yet be completely loaded and the routes may change in part 4, when transfer trips are assigned to the network. Furthermore, the improvement procedure, when done before the transfer trip assignment, would be much longer and involve looking at many nodes for improving each route. If the network is already loaded with all the demand, it will be easier to detect the weak spots (bad routes and transfer paths) and therefore less work will be required to improve the route network.

Assigning the transfer trips at the end of the algorithm may result in early rejection of some routes, just because the transfer trips were
not assigned to them. On the other hand, the improved set of routes may include better transfer possibilities that were not available before the improvement phase. Furthermore, in some cases the need for a transfer may be eliminated because the improved set of routes offers direct service between a pair of nodes that were not previously served directly.

It is probably better to assign transfer trips to the network before the improvement phase, because in this way the planner can get a better assessment of the network quality and problems before beginning the improvement procedure.

The following sections describe the general structure of the algorithm and presents its main properties. A formal step by step statement of the algorithm will be given in chapter 4.

3.1.1 Part 1 - Generating a set of shortest routes

Part 1 of the algorithm involves the generation of an attractive set of initial routes. First a pair of terminals and a shortest path route between them are chosen. The shortest routes are considered because in this way the time spent by all passengers on the network is minimized (each passenger is traveling on the shortest path from origin to destination). The operator's cost is also likely to be low because of the shortest path nature of all routes. These shortest routes will be modified in subsequent parts of the algorithm. The quality of a route depends greatly on the number of passengers using it. Routes that serve more passengers will usually be better than those with fewer passengers, because the service frequency on these routes is likely to be high (shorter waiting times). For this reason the terminals and route chosen in steps 1 and 2 are those satisfying the most demand. This is done by first selecting the terminal that generates the most demand (has the highest row total in the Origin-Destination demand matrix) as the origin. Next the terminal that has
the most demand to and from the origin terminal (highest entry in the origin terminal's row in the O-D demand matrix) is chosen as the destination. In this way routes that have many potential passengers, will be considered before routes that have less potential passengers. As mentioned in the previous section, different strategies may be used to generate different route networks from which the best network will be chosen.

In step 3 of part 1 the demand is assigned to the route chosen in step 2. If several routes share a common segment, the flow on this segment is initially divided equally among the routes. For example, if three routes provide service between a pair of nodes (i,j), one third of the demand between these nodes will initially be assigned to each route. After the assignment of transfer trips to the route network, in part 3 of the algorithm, the frequency share model will be used to divide the flow of passengers on common segments more realistically among the appropriate routes.

Part 1 of the algorithm is repeated until all the demand has been satisfied or until all terminal pairs considered. It is unlikely that all the demand will be satisfied upon completion of part 1. Usually unserved nodes, in which the demand has not yet been satisfied, will remain either because they have no direct route between them or because they are disconnected from the route network. A disconnected node is a node that is not served by any of the routes accepted in part 1.

3.1.2 Part 2 - Connecting disconnected nodes

At this point there is a set of shortest path routes, usually with some disconnected nodes. Since the demand can not be fully satisfied as long as there are disconnected nodes, this problem is addressed in part 2 of the algorithm. Two methods are suggested for
connecting disconnected nodes.

The methods deal with one disconnected node at a time. The existing routes are altered by adding a disconnected node to a route, or replacing a node on a route with a disconnected node, whichever minimizes the increase (or maximizes the decrease) in the value of the route evaluation function (see appendix B for a detailed discussion of this function). This procedure will usually yield routes that are longer than the shortest path between the termini. As a result some passengers may have excess riding time over the shortest possible riding time. On the other hand, since the number of passengers on the route will usually increase, the waiting time may decrease. The following sections describe these two methods in detail:

The highest demand generating disconnected node is considered each time. This is done because in the beginning of this phase there are more routes to which nodes can be connected. As nodes are connected to routes the riding time on the routes increases, and less routes are available for adding nodes because their riding time approaches the upper bound on the route riding time. As a result the nodes considered first may have more possible connections. Once a disconnected node is chosen, an appropriate route to connect it to has to be found. This is done as follows:

An attempt to add the node to an existing route is made first because in this way more demand is satisfied (as opposed to replacing the node with an existing node in which case some demand, that is already satisfied directly, may have to be served with a transfer). A feasible connection is one in which the riding time on the route, after adding the new node, does not exceed the route riding time upper bound.
Now an attempt is made to replace a node on an existing route with the disconnected node. A feasible connection in this case is one in which the node on the existing route does not become disconnected, and the riding time on the route is still within the route riding time bounds.

The value of the route evaluation function (see section 3.2) is calculated for all the routes involved in every possible connection. This can be done in one of the following ways:

1. Assigning only the demand that is served directly between the added node and all the nodes on the route.

2. Assigning the demand satisfied directly and the transfer trips from the added node to the route. This is possible because since this is the only connection of this node to the route network, all the flow out of it must travel on this route. The problem with this approach is that at this stage the rest of the transfer paths, beyond the first route, are unknown so the waiting time and unutilized-space-hours for the route can not be calculated accurately.

Thus only the demand satisfied directly with the new connection will be assigned.

The connection that caused the minimal increase (or the maximal decrease) in the sum of the values of the route evaluation functions of the routes effected by the connection is activated. In this way the total increase in the network objective function is likely to be minimized.

If more than one connection has the same change in the value of the route evaluation function, the one that satisfies more demand is
chosen. For example, consider disconnected node \( i \), which can be connected to either route A or B. The change in the route evaluation functions of both routes A and B due to the connection is the same. If the demand between node \( i \) and all the nodes on route A is higher than the demand between node \( i \) and all the nodes on route B, the connection to route A will be chosen.

This is done until the network is fully connected, meaning that there is at least one route serving each node. Even though an attempt is made to connect disconnected nodes to the route network without making any route worse, it may not always be possible.

3.1.3 Part 3 - Trip assignment

In this part the frequency share model is used to divide the flow of passengers on common segments more realistically among the appropriate routes. This is done because a passenger that can use more than one bus line for the same trip, is likely to board the first bus to arrive. On the average, more of the flow on the common segment will be found on the route with the higher frequency. This procedure is performed before the assignment of transfer trips to the route network, because the transfer trip assignment depends on the route frequencies that may change as a result of using the frequency share model to divide the demand. The following sections briefly describe the procedure:

The demand on the common segment, \( i \rightarrow j \), is denoted \( d_{ij} \). \( d_{ijk} \) is the portion of the demand between \( i \) and \( j \) satisfied by route \( k \), operating at a frequency \( f_k \). \( d_{ijk} \) is calculated by equation 2.17.

For example if the route being considered is number 1, and there are two other routes, 2 and 3, that share a common segment \( i \rightarrow j \), with route 1:
\[ d_{ij1} = \frac{f_1}{(f_1+f_2+f_3)} \times d_{ij} \]
\[ d_{ij2} = \frac{f_2}{(f_1+f_2+f_3)} \times d_{ij} \]
\[ d_{ij3} = \frac{f_3}{(f_1+f_2+f_3)} \times d_{ij} \]

Now the route frequencies are modified according to the new demand assignment, and \( d_{ijk} \) is re-calculated. This procedure is repeated iteratively until there is no change in the frequencies from one iteration to the next. (See Appendix A for a detailed discussion of this procedure).

Transfer trip assignment occurs on a node pair basis. This means that for each pair of nodes having unserved demand, the best transfer path is found. A feasible transfer path between two nodes must satisfy the following constraints:

* The travel time between these nodes, including the transfer time, must be within the travel time bounds.

* The excess travel time over the shortest possible riding time between these nodes on the network, is below the upper bound on transfer excess travel time.

* The number of transfers does not exceed the maximum number of transfers allowed.

If there is more than one feasible transfer path which satisfies the constraints, the demand is divided among all the feasible transfer paths according to the frequency share model considering the frequency on the first route of each transfer path (see equation 2.18).

This approach is taken because the passengers are assumed to minimize their expected arrival time at their destination.
As a result of the assignment of transfer trips, the frequencies of some routes may change. The frequency share model is used here again to divide the demand on common segments among the appropriate routes.

Some times it will not be possible to satisfy the demand between a pair of nodes without violating one of the constraints mentioned above. In such cases the best transfer path will be accepted and it will be noted which constraints are violated by which trip.

3.1.4 Part 4 - route network improvement

At this stage the network is connected, and all the demand is loaded on it. Part 4 of the algorithm attempts to satisfy the demand in a better way. The weak parts of the route network (bad transfer paths and routes) are identified and a systematic attempt is made to improve them.

To review the parts of the algorithm presented so far; first a set of shortest path routes was generated in part 1. Part 2 connected disconnected nodes to the route network in a way that minimized the increase in the value of the objective function. In part 3 transfer trips were assigned to the route network, which means that the number of passengers on some of the routes, to which transfer passengers were assigned, increased.

The route network can be improved in three ways:

a. By reducing the total excess riding time, ΔPH, the passengers spend on the network, and/or

b. By reducing the total waiting time, WH, on the network, and/or
c. By reducing the number of unutilized-space-hours, UH, on the network.

The best way to improve the route network is to try and eliminate as many transfer trips as possible, by replacing them with direct service. In this way the excess riding time is reduced because usually a direct route between two nodes will be shorter than a transfer path. The waiting time will also be reduced because passengers that had to wait for the bus twice (or more) on a transfer, have to wait only once when traveling on a direct path from origin to destination. As for the number of vehicle-hours, if the passengers spend less time on the network, the total number of vehicle-hours required to satisfy the demand is likely to decrease.

There is no point in considering all node pairs served with a transfer. Some node pairs will require a transfer anyway, just because they are too far away from each other, and there is no feasible direct route between them (the shortest path between them exceeds the route length upper bound). Therefore, the transfer paths considered for replacement with a direct service, are those on which the shortest riding time between origin and destination does not exceed the upper bound on route riding time. Once these transfer paths are identified, they are evaluated using the transfer path evaluation function discussed in section 3.3. This function is based primarily upon the excess time over the shortest possible riding time that the passengers on the transfer path spend on the network. The transfer paths are ranked according to the value of the transfer path evaluation function in decreasing order (the worst transfer path, with the highest value of the function, first). An attempt is made to replace the transfer paths with direct service starting with the first (worst) transfer path on the list.
As mentioned before the improvement of the service between a pair of nodes, provided by a transfer path, is attempted by providing it with more direct service. This means that an attempt is made to reduce the number of routes on the transfer path. Since the existing transfer path is the best one found on the route network (in part 3 of the algorithm), the improvement process should seek new routing for the passengers. These new routing possibilities are evaluated in the route network context and not as individual routes. If the route network is improved (the value of the objective function reduced) the improvement is accepted. The following sections describe the improvement process for transfer paths which include two routes. Multiple route transfer paths are discussed later. A distinction is made between two cases:

1. Both the origin and destination are terminal nodes. In this case the direct shortest route between these terminal nodes is considered again, even though it was rejected in part 1 of the algorithm. The demand is assigned to the route and the distribution of passengers on the route network is updated. If the value of the objective function is reduced the new route is accepted.

2. One (or both) of the nodes is not a terminal node. The shortest direct route between the nodes is still considered and an attempt is made to connect the node that is not a terminal node to the nearest terminal. If this is possible and the new route is within the route length bounds, the demand is assigned to it and the passengers flow on the route network is updated. As in case 1 the change is accepted if the value of the objective function is reduced.

3. If the transfer path is bad because the waiting time on the second route is high, a new route overlapping the segments of the original routes can be tried. This is possible only if the origin
and destination are terminals.

In cases where the transfer path includes more than two routes, the improvement procedure is as follows:

a. An attempt is made to replace the transfer path with a single direct route between origin and destination in the same way as for the two routes transfer paths mentioned above.

b. If a. is not possible, each pair of routes on the transfer path is considered separately for replacement with the appropriate direct route. For example, consider a transfer path between nodes 2 and 7 that includes routes A, B, and C. The trip is via nodes 2 on route A, 5 on routes A and B, 12 on routes B and C, and 7 on route C. The direct route between nodes 2 and 12 is considered instead of routes A and B, and the direct route between nodes 5 and 7 is considered to replace routes B and C.

As for improving individual routes, it may be possible to achieve some improvement by adding nodes to a route or replacing a node on a route with another node. This procedure involves considerable effort in finding the right routes and nodes, and the improvement, if any, is likely to be very small for the following reasons:

a. Adding or replacing a node on a route will usually increase the length of the route, which means more excess riding time.

b. Even though the number of passengers on one route will increase (which may decrease the waiting time on this route), the number of passengers on another route will decrease by the same amount (the total number of passengers on the network is fixed), and the waiting time on that route may increase.
c. Since the passengers will only be distributed differently among routes, and the riding time will not necessarily decrease, the number of vehicle-hours required to satisfy the demand across the network is not likely to decrease much (if at all).

Nevertheless, the routes, like the transfer paths, are ranked according to the value of the route evaluation function in decreasing order. The routes considered for improvement are only those for which the value of the route evaluation function increased above the threshold used in part 1 for accepting a route.

An attempt is made to modify these routes by adding or replacing nodes. The node to be added to a route has to satisfy the following conditions:

a. The route length, with the added node is within the route length bounds, and

b. It is the node that has the highest demand to and from the other nodes on the route.

The node to replace an existing node on the route has to satisfy two additional conditions:

c. The node it replaces will not become disconnected from the route network, and

d. The demand between the new node and the other nodes on the route is greater than the demand between the existing node and the others on the route.

Only one node (the one with the highest demand to and from the route) is considered for adding or replacing (not necessarily the
same node). This is because the improvement of the route is mainly due to the increase in the number of passengers. If adding the largest number of passengers possible did not improve the route, trying another node, with less passengers, is not likely to improve it either. Each time a route is modified, the passengers flow on the route network is updated and the change is accepted only if there is a decrease in the value of the objective function.

Unlike the procedure of connecting disconnected nodes, no change is accepted if it does not improve the route network (reduce the value of the objective function). This is done until no route can be further improved. The frequency share model is used again to divide the demand on common segments among routes on which the service frequency was changed as a result of the improvement process.

3.2 The route evaluation function

The purpose of the route evaluation function is to capture the quality of a transit route from both the passengers and the operator's perspectives. It does not take into account any transfer demand assigned to the route. Transfers are assigned to the route network and evaluated after the initial set of routes is generated and the network is connected. The most important measure from the passenger's point of view is the time spent on the network, including both the waiting and riding times. The operator, on the other hand, is most interested in efficient utilization of the available equipment.

This route evaluation function is used in the route generation algorithm to eliminate bad initial routes (routes with very low Z values) in the generation of shortest path routes in part 1. It is used again in part 2 to choose the best connection of a disconnected node according to the value of the route evaluation function of all the routes involved.
In part 1 of the algorithm only shortest routes between terminal nodes are considered, one route at a time. The demand satisfied directly is assigned to the routes, and the value of the route evaluation function is calculated. The route is activated if this value is below the threshold for accepting a route.

The route evaluation function is very similar to the objective function described in chapter 2. While the objective function is used to evaluate the whole route network, with all the demand loaded (which means that the number of passengers is fixed), the route evaluation function is used to evaluate a single route. Therefore, the number of passengers using the route is an important factor in the route evaluation (see appendix B for a detailed discussion of the route evaluation function).

Thus, the route evaluation function in this work is:

\[ Z_k = \frac{\alpha \Delta PH_k + \beta \cdot WH_k + \gamma \cdot UH_k}{P_k} \]  \hspace{1cm} [3.1]

where \( Z_k \) is the value of the route evaluation function for route \( k \), \( \Delta PH_k \), \( WH_k \), and \( UH_k \) are the excess riding time, waiting time, and unutilized-space-hours on route \( k \) respectively. The terms \( \theta \), \( \mu \), and \( \varphi \) are the weights for \( \Delta PH_k \), \( WH_k \), and \( UH_k \) respectively and \( P_k \) is simply the total number of passengers traveling on route \( k \).

3.3 The transfer path evaluation function

The transfer path evaluation function is used to rank the transfer paths according to their quality. The quality of a transfer path between two nodes is measured by the excess time (both travel and waiting time) the passengers have to spend on it over the shortest possible riding time between these nodes on the network.
The total time spent by a passenger on transfer path $r$ between nodes $i$ and $j$ is:

$$T_{ij}^r = \sum_{k \in r} (w_k + t_k^r)$$  \[3.2\]

where $t_k^r$ is the travel time spent by a passenger using transfer $r$ on route $k$ [minutes].

The transfer path evaluation function, $Z_r$, is actually the total excess time that the passengers using transfer $r$ spend on the network, $\Delta PH_r$, is [passenger-hours]:

$$\Delta PH_r = (1/60) * d_{ij}^r * (T_{ij}^r - t_{ij}^m)$$  \[3.3\]

where $t_{ij}^m$ is the shortest possible riding time between nodes $i$ and $j$ on the network [minutes], and $d_{ij}^r$ is the demand between nodes $i$ and $j$ satisfied by transfer $r$ [passengers/hour].

Since the total excess time passengers on transfer $r$ spend on the network includes both riding and waiting times, the number of routes on the network is used only as a tie breaker. If two transfer paths have the same evaluation function value the one that includes more routes will be considered first. This is done because a transfer path that includes less routes is more desirable than a transfer path with many routes.

The value of this function is calculated for all node pairs served by a transfer path, between which the shortest possible riding time does not exceed the upper bound on riding time. The transfer paths are ranked in decreasing order and the improvement process begins with the first one (the one with the highest value).
3.4 The evaluation functions revisited

The heuristic algorithm includes three evaluation functions. The route evaluation function is used in Part 1 to evaluate the shortest path routes between terminal nodes. In this part a route is accepted if the value of its route evaluation function is below the threshold for accepting a route. It is also used to rank the routes, according to their quality, for the route improvement procedure. In part 2 of the algorithm the function is used, when connecting a disconnected node to the route network, to evaluate different connections and choose the best one.

The objective function is used to evaluate the whole route network with all the demand assigned to it. This is done in Part 4 to evaluate an improvement of the route network.

The transfer path evaluation function is used to rank the transfer paths according to the excess time passengers spend on them for the route network improvement.
CHAPTER 4

ROUTE GENERATION ALGORITHM

This chapter presents the heuristic approach to the transit route generation problem as a step by step procedure. The input required for the algorithm is described first, followed by the algorithm itself.

4.1 Input

The input includes three major components, the street network, the Origin-Destination demand matrix and the route bounding parameters.

4.1.1 Network

The network $G = (N,A)$, with $|N|$ nodes and $|A|$ arcs, includes the following information:

* A set of terminal nodes, $m \in N$ - nodes at which routes can be initiated and/or terminated. All routes must connect two terminal nodes.

* A set of service nodes, $s \in N$ - nodes that must be served but can not be terminal nodes.

* Each arc represents a possible path between two nodes in the network, and is defined by:
1. Nodes at each end.

2. Average bus travel time.

3. Distance.

4.1.2 O-D Demand Matrix

The following assumptions are made with regard to the O-D demand matrix:

a. The demand matrix is treated as deterministic. Even though the demand for public transportation is stochastic, meaning that it has some distribution even when all the conditions are equal, the expected value of this distribution is taken as the actual value of the demand between each origin destination pair.

b. Demand may vary greatly between different hours of the day, different days of the week and different seasons throughout the year. One way to take into account this variation is to generate different demand matrices for different times. For example a matrix for the morning peak period and a different matrix for the off-peak period on a weekday. However routes are usually designed based on the peak period demand matrix. These same routes will generally be used during the off-peak periods with different operating strategies (like short-turns and deadheading) and different frequencies. Accordingly the peak period demand matrix will be used as the input to this process.

c. The demand represented by the O-D demand matrix is on a per time unit basis. An hourly demand matrix, for example, repre-
sents the average number of passengers per hour between each node pair. That is, for a peak period of three hours, a demand of 100 passengers between a certain pair of nodes, implies a demand of 300 passengers, between these nodes, during the entire period. For simplicity, but without loss of generality, the O-D demand matrix is assumed to be per hour.

4.1.3 Route Bounding Parameters

The route bounding parameters are specified by the user of the algorithm. The purpose of most of them, such as bounds on route travel time, is to eliminate undesirable routes at an early stage of the algorithm and avoid wasting time evaluating such routes. Other parameters are technical in nature such as vehicle capacity, or policy oriented such as minimum required frequency.

The following route bounding parameters are included in the algorithm:

a. Route travel time upper bound $T_u$ and lower bound $T_l$. all routes must have travel times between $T_l$ and $T_u$.

b. Maximum allowable excess travel time over the shortest travel time between a pair of nodes, for direct service, $\delta_d$, and for service provided on a transfer path, $\delta_T$. This bound is expressed as a percentage, in decimal fraction, of the shortest possible riding time between the nodes.

c. An upper bound on the value of the route evaluation function $Z$ (see appendix B for a detailed discussion of the route evaluation function).

d. An upper bound on the number of transfers allowed in one
passenger trip.

e. Minimum required frequency, \( F^m \).

f. Desired maximum occupancy on a vehicle, \( C \).

4.2 The route generation algorithm

4.2.1 Part 1 - Generating a set of shortest routes

In this part the shortest routes between terminal node pairs are
considered and the best ones are activated until all the demand
is satisfied or all the terminal node pairs considered.

**Step 1: Choosing a pair of terminals.**

a. Choose the initial terminal as that generating the highest
demand, i.e. highest row or column total in the O-D demand
matrix. If two terminals generate the same amount of demand,
choose the pair of terminals that have the highest demand
between them (highest entry in the rows corresponding to the
terminals in the O-D demand matrix).

b. Choose the second terminal as that attracting the highest
demand from the initial terminal, i.e. the highest entry in
the row or column for the initial terminal in the O-D demand
matrix.

c. Check that the shortest path between the selected terminals
is within the route travel time bounds, if it is not choose
another terminal by returning to step 1.b.

**Step 2: Choosing a path between the O-D pair.**
If there is more than one path between the origin and destination with the same shortest travel time, choose the path that satisfies more demand.

Step 3: Demand assignment and shortest route activation procedure.

a. If it is the first route in the network:

1. Assign all demand between all node pairs belonging to the route to this route.

2. Calculate the value of the route evaluation function (see equation 3.3).

3. Accept the route if the value of the route evaluation function $Z$, is below the user-specified upper bound, $Z_U$.

4. If the route is accepted delete all the demand satisfied by the route from the O-D demand matrix.

5. Go to step 1.

b. If it is not the first route in the network:

1. Assign the demand between all node pairs on the route. If a segment of the route is common with other routes, divide the flow on the common segment equally among the appropriate routes.

2. Calculate the value of the route evaluation function.

3. Accept the route if the value of the route evaluation
function Z is below the user-specified upper bound.

4. If the route is accepted delete all the demand satisfied by it from the O-D demand matrix.

5. If all the demand is satisfied or all terminal node pairs were considered, go to step 4. Otherwise go to step 1.

4.2.2 Part 2 - Connecting disconnected nodes

This part deals with one disconnected node at a time. First an attempt is made to add the node to an existing route. The same node is then considered for replacement with a node on an existing route. The best connection of these two is activated.

Step 4: Adding a disconnected node to an existing route.

a. Find the highest demand generating disconnected node, h.

b. Find the route with the highest demand to and from node h, on which the riding time, after adding the node, does not exceed the riding time upper bound.

c. If there is more than one way to add the node to the route, choose the one that minimizes the riding time on the new route.

d. Assign the demand between all the nodes on the route and node h to the route.

e. Calculate the value of the route evaluation function.
Step 5: Replacing a node on an existing route with a disconnected node.

a. Find a node 1, on the route with the highest demand to and from node h, found in step 4.a, that node h can replace so that:

1. When node 1 is removed from the route, it will not become disconnected.

2. The route, with node h instead of 1, meets the route riding time constraint.

b. Delete the demand between node 1 and all the nodes on the route, and assign the demand between node h and the nodes on the route.

c. Calculate the value of the route evaluation functions of both routes that are involved in the connection.

d. Accept the connection for which the increase in the value of the route evaluation functions is minimal (or the decrease is maximal).

e. If all the nodes are connected, go to step 6. Otherwise go to step 4.

4.2.3 Part 3 - Trip assignment.

In this part the frequency share model is used to divide the flow of passengers on common segments more realistically among the appropriate routes. This is done both before and after the transfer trip assignment, and repeated until there is no change in the routes frequencies (see appendix A for a detailed discus-
sion of this procedure).

Step 6: Dividing the flow of passengers on common segments among routes.

a. Find all the routes that share a common segment i-j.

b. Divide the demand on the common segment among the appropriate routes, until there is no change in the route frequencies, according to the frequency share model (see equation 2.17).

Step 7: Transfer trip assignment.

a. Find the pair of nodes with the highest transfer demand between them.

b. Find a transfer path using active routes, between the nodes found in step 7.a, that meets the following constraints:

1. The travel time between these nodes, on the transfer path, is within the travel time bounds.

2. The excess travel time over the shortest riding time between these nodes, on the transfer path, is below the upper bound on transfer excess travel time.

3. The number of transfers does not exceed the maximum number of transfers allowed.

c. If the only transfer path between these nodes does not meet one or more of the aforementioned constraints, accept it, assign the demand to the appropriate routes, and note which constraint is violated.
d. If there is only one feasible transfer path between these nodes, assign the demand to the appropriate routes.

e. If there is more than one feasible transfer path between the nodes found in step 7.a:

1. Identify the best transfer path - the one on which the expected time of arrival at the destination is minimized.

2. Find other acceptable transfer paths on which the expected arrival time, conditioned on boarding the first bus to arrive, is smaller than the expected arrival time on the best transfer path found in step e.1 above.

3. Divide the demand among the acceptable transfer paths according to the frequency share model, considering the frequency on the first route of each transfer path (see equation 2.18).

f. If all the demand on the network is satisfied, go to step 8. Otherwise, go to step 6.

4.2.4 Part 4 - Route network improvement.

In this part attempts are made to improve the route network. This is done first by trying to replace transfer trips with direct service and then trying to improve the routes by adding or replacing nodes.

Step 8: Replacing transfer trips with direct service.

a. Rank all the active transfer paths in increasing order of quality (the worst one first), according to the transfer path
evaluation function (see section 3.3 for a detailed discussion of this function).

b. Find the first (worst) transfer path on the list. If the list is empty, go to step 9.

c. When the transfer path includes two routes and the shortest riding time between origin and destination is within the route riding time bounds, there are two cases:

I. Both the origin and destination are terminal nodes:

1. Delete the demand from the routes on all transfer paths between origin and destination.

2. Assign the demand to the shortest path between origin and destination.

3. Use the frequency share model to update the passengers flows on common segments among routes.

4. Calculate the value of the objective function for the network (see equation 2.16).

5. Accept the improvement if the value of the objective function for the network is reduced.

6. Delete the considered transfer path from the list, and go to step 8.b.

II. Either the origin or destination (or both) is not a terminal node:
1. Try to connect the shortest path between origin and destination so it starts and ends at a terminal node.

2. If a connection is found perform step 8.c.I. Otherwise delete the considered transfer path from the list and go to step find 8.b.

d. When the transfer path includes more than two routes:

1. If the shortest path between origin and destination is within the route riding time bounds perform steps 8.c.I.1 through 8.c.I.5 if both nodes are terminal nodes, and step 8.c.II if at least one of them is not. Otherwise, or if the improvement was not accepted, go to next step.

2. Try to replace each pair of consecutive routes on the transfer path with the appropriate direct route. Use step 8.c.I if both nodes on the routes are terminal nodes and step 8.c.II otherwise.

Step 9: Adding a node to a route.

a. Rank all the routes for which the value of the route evaluation function is above the threshold for accepting a route, in decreasing order (the worst route first).

b. Find the first (worst) route on the list. If the list is empty, stop.

c. Find a node to add to the route that satisfies the following conditions:

1. The route length, with the added node is within the route
length bounds, and

2. It is the node that has the highest demand to and from the other nodes on the route.

d. Assign the demand between the added node and all the other nodes to the route.

e. Calculate the value of the objective function for the network with the added node.

f. Accept the improvement if the value of the objective function, after adding the node, decreased. Otherwise, go to step 10.

g. Delete the route from the list, and go to step 9.b.

**Step 10: Replacing a node on a route.**

a. Find a node to replace an existing node on a route that satisfies the following conditions:

1. The route length, with the added node is within the route length bounds, and

2. It is the node that has the highest demand to and from the other nodes on the route.

3. The node it replaces will not become disconnected from the route network, and

4. The demand between the new node and the other nodes on the route is greater than the demand between the existing node and the others on the route.
b. Assign the demand between the new node and all the other nodes on the route, instead of the demand between the node it replaced and the other nodes.

c. Calculate the value of the objective function for the network with the new node.

f. Accept the improvement if the value of the objective function, after replacing the node, decreased.

g. Delete the route from the list, and go to step 9.b.
CHAPTER 5

APPLICATION

In order to test the efficiency of the algorithm presented in chapter 4, a small example was solved both optimally, using the MINOS nonlinear optimization computer package, and applying the algorithm. The formulation of the Transit Network Design problem for the MINOS package was presented in chapter 2. In this chapter the example network is presented along with both the optimal and the heuristic solutions. Finally these solutions are compared with the aim of learning how the heuristic could be improved.

5.1 The example network

The example network includes five nodes, with three terminal nodes and two service nodes. There are seven arcs on which travel is possible in both directions. The O-D demand matrix is expressed in terms of passengers/hours. Both the network and the O-D demand matrix are shown in figure 5.1.

5.2 The problem parameters

This section presents the parameters that are used in both the heuristic and optimal solutions:

1. The riding time upper and lower bounds are 50 and 20 minutes respectively.
figure 5.1: THE NETWORK (riding time in minutes)

T-TERMINAL NODE

THE 0-D DEMAND MATRIX
(PASSENGERS/HOUR)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>30</td>
<td>20</td>
<td>30</td>
<td>10</td>
<td>90</td>
</tr>
<tr>
<td>2</td>
<td>30</td>
<td></td>
<td>30</td>
<td>25</td>
<td>20</td>
<td>105</td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td>30</td>
<td></td>
<td>15</td>
<td>25</td>
<td>90</td>
</tr>
<tr>
<td>4</td>
<td>30</td>
<td>25</td>
<td>15</td>
<td></td>
<td>15</td>
<td>85</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>20</td>
<td>25</td>
<td>15</td>
<td></td>
<td>70</td>
</tr>
</tbody>
</table>
2. The maximum allowable excess riding time for a direct route is chosen to be $\delta_d = 2.0$. This means that a path between two nodes is not acceptable, if its riding time is more than twice the shortest riding time between the nodes.

3. Since the network is small and the paths are relatively short, the maximum allowable excess time constraint for transfer service is not imposed. This means that all the transfer paths within the maximum number of routes bound can be considered.

4. The upper bound on the value of the route evaluation function (above which routes are not accepted in part 1) is $Z_u = 0.75$.

5. The upper bound on the number of routes on a transfer path is 2.

6. The minimum required frequency constraint is not imposed because the number of passengers is relatively small, and the frequency is not likely to exceed 2 vehicles/hour. This does not affect the comparison between the optimal and the heuristic solutions.

7. The desired maximum vehicle occupancy is 50 passengers/vehicle.

8. The weights for the objective function are as follows:

Let $T_{ijk}$ be the total travel time, including both waiting and riding times, between nodes $i$ and $j$ on route $k$:

$$T_{ijk} = \phi w_k + t_{ijk}$$  \[5.1\]

where $\phi$ is a coefficient greater than 1 that represents the fact that passengers value waiting time more than riding time [Quarmby, Friedlander]. The value of $\phi$ is estimated to be 2.5. This means that a passenger is indifferent between riding time of
2.5 minutes and waiting time of 1 minute, but would rather spend 2 minutes riding the bus than one minute waiting for it, even if this means arriving later at the destination. The relation between riding time [passenger-hour] and vehicle time [vehicle-hours] in terms of cost to the operator is assumed to be 1:10. This means that if the weight for waiting time $\beta = 1.0$ the other weights will be $\omega = 0.4$ and $\gamma = 4.0$ so $\alpha = 4.4$.

Since these numbers are valid for the route level too, the weights for the route evaluation function are the same:

$\alpha = 4.4; \beta = 1.0; \gamma = 4.0$.

5.3 Applying the heuristic algorithm

The application of the heuristic algorithm will be demonstrated in this section in a step by step manner.

5.3.1 Part 1 - Generating a set of shortest routes

This part examines the shortest path routes between terminal nodes and selects the ones for which the value of the route evaluation function is below the threshold for accepting a route.

First route (figure 5.2)

Step 1: Choosing a pair of terminal nodes. In this case both terminals 1 and 3 are generating the same demand, 90 passengers/hour, so the choice is made according to the terminal attracting the highest demand from the first terminal (the highest entry in the two rows, 1 and 3). Thus, the first pair of terminals is (3,5).

Step 2: Choosing a shortest path between the terminal nodes. The
figure 5.2: ROUTE 1 (riding time in minutes)

T-TERMINAL NODE

THE 0-D DEMAND MATRIX
AFTER ACTivating ROUTE 1
(PASSENGERS/HOUR)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>30</td>
<td>20</td>
<td>30</td>
<td>10</td>
<td></td>
<td>90</td>
</tr>
<tr>
<td>2</td>
<td>30</td>
<td></td>
<td>25</td>
<td>1</td>
<td>20</td>
<td>55</td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td>1</td>
<td>15</td>
<td></td>
<td>25</td>
<td>35</td>
</tr>
<tr>
<td>4</td>
<td>30</td>
<td>25</td>
<td>15</td>
<td></td>
<td>15</td>
<td>85</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>1</td>
<td>1</td>
<td>25</td>
<td>15</td>
<td>25</td>
</tr>
</tbody>
</table>

1 30 - demand between nodes 2 and 3 satisfied by route 1.

\[ P = 75 \text{ passengers} \]
\[ Z = 0.543 \]
shortest path between nodes 3 and 5 is 3-2-5 with riding time of 20 minutes.

Step 3: Calculating the value of the route evaluation function:

**Route 1** (figure 5.2)

\[ \Delta PH_1 = PH_1 - PH_{M1} = 0 \quad \text{[passenger-hours]} \]

\[ WH_1 = P_1*G/(2*q_1) = 75*50/(2*55) = 34.09 \quad \text{[passenger-hours]} \]

\[ UH_1 = (1/60)*q_1*T_1 - PH_1 = (1/60)*[55*20 - 10*(55+45)] = 1.66 \quad \text{[space-hours]} \]

\[ Z_1 = (\theta*\Delta PH_1 + \mu*WH_1 + \varphi*UH_1)/P_1 = (1.0*34.09 + 4.0*1.66)/73 = 0.543 \]

The value of the objective function is below the threshold for accepting a route, so the route is accepted.

**Second route** (figure 5.3)

Step 1: Choosing a pair of terminal nodes. The highest demand generating terminal is now terminal 1 (90 passengers). The terminal attracting the highest demand from terminal 1 is terminal 3 (20 passengers). The shortest path between terminals 1 and 3 is 10 minutes, below the lower bound on route riding time, so this terminal pair is not acceptable. The next highest demand attracting terminal from terminal 1 is terminal 5 with shortest riding time of 30 minutes. Thus, the next terminal pair to be examined is (1,5).

Step 2: Choosing a shortest path between the terminal nodes. The shortest riding time between terminals 1 and 5 is 30 minutes
figure 5.3: ROUTE 2 (riding time in minutes)

THE 0-D DEMAND MATRIX
AFTER ACTIVATING ROUTES 1 & 2
(PASSENGERS/HOUR)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>30</td>
<td>2</td>
<td>20</td>
<td>30</td>
<td>20</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>1</td>
<td>15</td>
<td>15</td>
<td>25</td>
<td>15</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>15</td>
<td>2</td>
<td>15</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>4</td>
<td>30</td>
<td>25</td>
<td>15</td>
<td>15</td>
<td>85</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>10</td>
<td>1</td>
<td>12</td>
<td>15</td>
<td></td>
</tr>
</tbody>
</table>

15 - demand between nodes 2 and 3 satisfied by route 1.

P = 98 passengers
Z = 0.66
and there are two paths that have this riding time, 1-2-5 which serves 50 passengers and 1-3-2-5 which serves 98 passengers. The second route serves more passengers so it is chosen.

Step 3: Calculating the value of the objective function. Since the segment 3-2-5 is common with route 1 the flow on this segment is initially divided equally between routes 1 and 2. For example, both routes 1 and 2 will be assigned 15 passengers each between nodes 3 and 2:

Route 2 (figure 5,3)

\[ \Delta PH_2 = 0 \quad \text{[passenger-hours]} \]

\[ WH_2 = \frac{98 \times 50}{(2 \times 68)} = 36.03 \quad \text{[passenger-hours]} \]

\[ UH_2 = \frac{1}{60} \times [68 \times 30 - 10 \times (60 + 68 + 33)] = 7.16 \quad \text{[space-hours]} \]

\[ Z_2 = \frac{1.0 \times 36.03 + 4.0 \times 7.16}{98} = 0.66 \]

The value of the objective function is below the threshold for accepting a route, so the route is accepted.

At this point all the terminal node pairs have been considered. Node 4 is disconnected from the route network because none of the shortest routes between terminal nodes provides service to this node.

5.3.2 Part 2 - Connecting disconnected nodes

This part is connecting the disconnected nodes to the route network in a way that minimizes the increase in the value of the objective function. This is done by accepting the connection with which minimizes the increase in value of the route evaluation functions, for the routes
involved in the connection. There is only one disconnected node so this part is performed only once.

Step 4: Adding a disconnected node to an existing route. The disconnected node (node 4) can be added to either route 1 or route 2. The demand between node 4 and the nodes on route 2 is higher, but node 4 can not be added to route 2 because the riding time on route 2 with node 4 is 60 minutes for 1-3-2-4-5, and 55 minutes for 1-3-4-2-5. Both riding times are above the upper bound for route riding time. Node 4 can be added to route 1 (3-2-5) in two ways, 3-4-2-5 with riding time of 45 minutes, and 3-2-4-5 with riding time of 50 minutes. The shorter connection (denoted route 3) is considered, and the direct demand is assigned to the route (None of the demand between nodes 3 and 2 is assigned to route 3 because the riding time is above the maximum excess riding time). Now the value of the route evaluation function is calculated:

Route 3 (figure 5.4)

\[ \Delta PH_3 = \frac{1}{60} \times 15 \times (30 - 20) = 2.5 \quad \text{[passenger-hours]} \]

\[ WH_3 = \frac{65 \times 50}{2 \times 40} = 40.62 \quad \text{[passenger-hours]} \]

\[ UH_3 = \frac{1}{60} \times [(40 \times 45 - (15 \times 15 + 20 \times 40 + 10 \times 25)] = 8.75 \quad \text{[space-hours]} \]

\[ Z_3 = \frac{(4.4 \times 2.5 + 1.0 \times 40.62 + 4.0 \times 8.75)}{65} = 1.33 \]

Since all the demand between nodes 3 and 2, and between nodes 3 and 5, is satisfied by route 2, the route has changed and the route evaluation function is calculated again:
Figure 5.4  ROUTE 3  (riding time in minutes)

THE 0-D DEMAND MATRIX
AFTER ACTIVATING ROUTES 2 & 3
(PASSENGERS/HOUR)

3 15 - demand between nodes 3 and 4 satisfied by route 3.

P = 65 passengers
Z = 1.33
Figure 5.5: ROUTE 2 (riding time in minutes)

The 0-D Demand Matrix
After Activating Routes 3 & 2
(Passengers/Hour)

1 15 - demand between nodes 2 and 3 satisfied by route 1.

\[ P = 125 \text{ passengers} \]
\[ \bar{z} = 0.717 \]
Route 2 when route 3 is active (figure 5.5)

$\Delta PH_2 = 0$ [passenger-hours]

$WH_2 = \frac{125 \times 50}{2 \times 95} = 32.89$ [passenger-hours]

$UH_2 = \frac{1}{60} \times [95 \times 30 - 10 \times (60 + 95 + 45)] = 14.17$ [space-hours]

$Z_2 = \frac{1.0 \times 32.89 + 4.0 \times 14.17}{125} = 0.717$

Step 5: Replacing a node on an existing route with a disconnected node. Node 4 can only be replaced with node 2 on either routes 1 or 2. The demand between node 4 and the nodes on route 2 is higher than the demand between node 4 and the nodes on route 1. Therefore node 2 on route 4 is replaced with node 4 yielding the route 1-3-4-5 (denoted route 4). The direct demand is assigned to the route, and the value of the route evaluation function is calculated:

Route 4 when route 1 is active (figure 5.6)

$\Delta PH_4 = \frac{1}{60} \times [10 \times (45 - 30) + 13 \times (35 - 20)] = 5.75$ [passenger-hours]

$WH_4 = \frac{103 \times 50}{2 \times 68} = 37.87$ [passenger-hours]

$UH_4 = \frac{1}{60} \times [68 \times 45 - (10 \times 60 + 15 \times 68 + 20 \times 38)] = 11.33$ [space-hours]

$Z_4 = \frac{4.4 \times 5.75 + 1.0 \times 37.87 + 4.0 \times 11.33}{103} = 1.05$

Since node 2 is not on route 4, the demand to and from node 2 is now satisfied by route 1 only. The demand between nodes 3
Figure 5.6 ROUTE 4 (riding time in minutes)

THE 0-D DEMAND MATRIX
AFTER ACTIVATING ROUTES 1 & 4
(PASSENGERS/HOUR)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>30</td>
<td>420</td>
<td>430</td>
<td>410</td>
<td></td>
<td>30</td>
</tr>
<tr>
<td>2</td>
<td>30</td>
<td>130</td>
<td>25</td>
<td>120</td>
<td></td>
<td>55</td>
</tr>
<tr>
<td>3</td>
<td>420</td>
<td>130</td>
<td>415</td>
<td>112</td>
<td>413</td>
<td>35</td>
</tr>
<tr>
<td>4</td>
<td>330</td>
<td>25</td>
<td>415</td>
<td>415</td>
<td></td>
<td>25</td>
</tr>
<tr>
<td>5</td>
<td>410</td>
<td>120</td>
<td>112</td>
<td>415</td>
<td>413</td>
<td>0</td>
</tr>
</tbody>
</table>

1. Demand between nodes 2 and 3 satisfied by route 1.

P = 103 passengers
Z = 1.05
figure 5.7 ROUTE 1 (riding time in minutes)

THE 0-D DEMAND MATRIX
AFTER ACTIVATING ROUTE 1 & 4
(PASSENGERS/HOUR)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>30</td>
<td></td>
<td>30</td>
</tr>
<tr>
<td>2</td>
<td>30</td>
<td></td>
<td>1</td>
<td>30</td>
<td>25</td>
<td>55</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>20</td>
<td>1</td>
<td>30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>30</td>
<td>4</td>
<td>15</td>
<td>4</td>
<td>25</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>10</td>
<td>1</td>
<td>20</td>
<td>4</td>
<td>15</td>
</tr>
</tbody>
</table>

1. Demand between nodes 2 and 3 satisfied by route 1.

\[ P = 62 \text{ passengers} \]
\[ Z = 0.70 \]
and 5 is now divided between routes 1 and 4. Since route 1 changed its route evaluation function is calculated again:

Route 1 when route 4 is active (figure 5.7)

\[ \Delta PH_1 = 0 \quad \text{[passenger-hours]} \]

\[ WH_1 = \frac{62 \times 50}{2 \times 42} = 36.9 \quad \text{[passenger-hours]} \]

\[ UH_1 = \frac{1}{60} \times [42 \times 20 - 10 \times (42 + 32)] = 1.66 \quad \text{[space-hours]} \]

\[ Z_1 = \frac{(1.0 \times 36.9 + 4.0 \times 1.66)}{62} = 0.70 \]

The connection is evaluated according to the sum of the route evaluation functions of all the routes involved. The connection that minimizes this sum is activated. In this case there are two possibilities:

1. Adding node 4 to route 1. The values for the route evaluation functions are:

\[ Z_{add} = Z_2 + Z_3 = 0.717 + 1.33 = 2.047 \]

2. Replacing node 4 with node 2 on route 2. In this case the demand on the segment 3-2-5 is divided between routes 1 and 4. The values for the route evaluation functions are:

\[ Z_{rep} = Z_1 + Z_4 = 0.70 + 1.05 = 1.75 \]

Thus, the connection chosen is replacing node 4 with node 2 on route 2.
The route network includes two routes:

route 1: 3-2-5 with riding time of 20 minutes.
route 4: 1-3-4-5 with riding time of 45 minutes.

5.3.3 Part 3 - Trip assignment

In this part the transfer demand is assigned to the route network, and the frequency share model is used to divide the demand more realistically among routes and transfer paths.

Step 6: Dividing the flow of passengers on common segments among routes. No common segments on the route network, but the demand is still divided between routes that provide service between the same nodes. In this case the demand between nodes 3 and 5 (d35 = 25 passengers) is divided between routes 1 and 4 according to the frequency on the routes (see appendix A for a detailed discussion of this procedure):

**Iteration 0**

\[ d^0_{351} = 12 \]

\[ f^0_1 = \frac{42}{50} = 0.84 \]

\[ d^0_{354} = 13 \]

\[ f^0_4 = \frac{68}{50} = 1.36 \]

**Iteration 1**

\[ d^1_{351} = \left[ \frac{0.84}{(0.84 + 1.36)} \right] \times 25 = 10 \]
\[ f^1_1 = \frac{40}{50} = 0.8 \]
\[ d^{1\times 4} = \left[ \frac{1.36}{(0.84+1.36)} \right] \times 25 = 15 \]
\[ f^1_4 = \frac{70}{50} = 1.4 \]

**Iteration 2**

\[ d^{2\times 1} = \left[ \frac{0.8}{(0.8+1.4)} \right] \times 25 = 9 \]
\[ f^2_1 = \frac{39}{50} = 0.78 \]
\[ d^{2\times 4} = \left[ \frac{1.4}{(0.8+1.4)} \right] \times 25 = 16 \]
\[ f^2_4 = \frac{71}{50} = 1.42 \]

**Iteration 3**

\[ d^{3\times 1} = \left[ \frac{0.78}{(0.78+1.42)} \right] \times 25 = 9 \]
\[ f^3_1 = \frac{39}{50} = 0.78 \quad \text{(No change).} \]
\[ d^{3\times 4} = \left[ \frac{1.42}{(0.78+1.42)} \right] \times 25 = 16 \]
\[ f^3_4 = \frac{71}{50} = 1.42 \]

9 passengers are assigned to route 1 and 16 to route 4. The routes with this new demand assignment are shown in figures 5.8 and 5.9. Route evaluation functions are not recalculated.

**Step 7:** Transfer trip assignment. There are two pairs of transfer
THE 0-D DEMAND MATRIX
AFTER ACTIVATING ROUTE 1 & 4
AND DIVIDING THE DEMAND BETWEEN NODES 3 & 5
(PASSENGERS/HOUR)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>30</td>
<td>4</td>
<td>4</td>
<td>30</td>
<td>4</td>
<td>55</td>
</tr>
<tr>
<td>2</td>
<td>30</td>
<td>1</td>
<td>1</td>
<td>30</td>
<td>4</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>30</td>
<td>4</td>
<td>15</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>15</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>16</td>
<td>4</td>
<td>15</td>
</tr>
</tbody>
</table>

1 30 - demand between nodes 2 and 3 satisfied by route 1.

\[ P = 59 \text{ passengers} \]
Figure 5.9  ROUTE 4  (riding time in minutes)

The 0-D demand matrix after activating routes 1 & 4 and dividing the demand between nodes 3 & 5 (passengers/hour)

P = 106 passengers
demand nodes (1,2) and (2,4). The demand between nodes 1 and 2 is higher so this pair is considered first.

Node pair (1,2)

There are two transfer paths between nodes 1 and 2:

1. Transfer path 1: 1-3 on route 2 and 3-2 on route 1, total riding time of 20 minutes.

2. Transfer path 2: 1-3-4-5 on route 2 and 5-2 on route 1, total riding time of 55 minutes. This riding time is above the upper bound on riding time.

Thus all the demand between nodes 1 and 2 is appropriately assigned to routes 1 and 2.

Node pair (2,4)

There are two transfer paths between nodes 2 and 4:

1. Transfer path 3: 2-3 on route 1 and 3-4 on route 2, total riding time of 25 minutes.

2. Transfer path 4: 2-5 on route 1 and 5-4 on route 2, total riding time of 30 minutes.

Both transfer paths are acceptable. Since the first route on both of them is route 1 the demand is divided equally between them. Note that even though the transfer paths include the same routes, the travel on them is in opposite directions. A passenger waiting at node 2 will take the first vehicle to arrive in either direction. The routes with all the demand
The 0-D Demand Matrix
After Activating Route 1 & 4
And Assigning Transfers 1 & 2
(Passengers/Hour)

1 - 30 demand between nodes 2 and 3 satisfied by route 1.

\[
P = 114 \text{ passengers} \\
\text{(55 on transfers)}
\]

2-4 = 13/5 - transfer via node 5
Figure 5.11 ROUTE 4 (riding time in minutes)

THE 0-D DEMAND MATRIX
AFTER ACTIVATING ROUTES 1 & 4
AND ASSIGNING TRANSFERS 1 & 2
(PASSENGERS/HOUR)

\[
\begin{array}{cccccc}
1 & 2 & 3 & 4 & 5 & \text{total} \\
1 & 130 & 430 & 130 & 130 & 130 \\
2 & 130 & 430 & 130 & 130 & 130 \\
3 & 420 & 130 & 415 & 19 & 416 \\
4 & 430 & 125 & 415 & 415 & 415 \\
5 & 410 & 120 & 19 & 416 & 415 \\
\end{array}
\]

130 - demand between nodes 2 and 3 satisfied by route 1.

\[P = 161 \text{ passengers} \]
\[(55 \text{ on transfers})\]

1-2 = 30/3 - transfer via node 3

\[1-3 = 20 \]
\[1-4 = 30 \]
\[1-5 = 10 \]
\[2-4 = 12/3 \]
\[3-4 = 16 \]
\[3-5 = 16 \]
\[4-5 = 15 \]
\[54 \]
assigned to them are shown in figures 5.10 and 5.11. After the transfer trips are assigned to the appropriate routes the frequencies change so the demand on between nodes that are serviced by more than one route is divided between the routes again according to the new frequencies. In this example nodes 3 and 5 are served by both routes 1 and 2:

**Iteration 0**

\[ d_{351}^0 = 9 \]

\[ f_{1}^0 = \frac{81}{50} = 1.62 \]

\[ d_{354}^0 = 16 \]

\[ f_{4}^0 = \frac{90}{50} = 1.8 \]

**Iteration 1**

\[ d_{351}^1 = \left[\frac{1.62}{(1.62+1.8)}\right] \times 25 = 12 \]

\[ f_{1}^1 = \frac{84}{50} = 1.68 \]

\[ d_{354}^1 = \left[\frac{1.8}{(1.62+1.8)}\right] \times 25 = 13 \]

\[ f_{4}^1 = \frac{90}{50} = 1.8 \quad \text{(the maximum arc load is not effected).} \]

**Iteration 2**

\[ d_{351}^2 = \left[\frac{1.68}{(1.68+1.8)}\right] \times 25 = 12 \quad \text{(No change).} \]

Since all the demand is satisfied, the objective function for the network can be calculated:
figure 5.12 ROUTE 1  (riding time in minutes)

THE 0-D DEMAND MATRIX
AFTER ACTIVATING ROUTE 1 & 4
AND ASSIGNING TRANSFERS 1 & 2
(PASSENGERS/HOUR)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>30</td>
<td>4</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>0</td>
</tr>
</tbody>
</table>

130 - demand between nodes 2 and 3 satisfied by route 1.

\[ \mathbb{P} = 117 \text{ passengers} \]
(55 on transfers)

2-4 = 13/5 - transfer via node 5
**THE 0-D DEMAND MATRIX**

**AFTER ACTIVATING ROUTES 1 & 4**

**AND ASSIGNING TRANSFERS 1 & 2**

(PASSengers/HOUR)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>total</strong></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

- 1 - demand between nodes 2 and 3 satisfied by route 1.

**P = 158 passengers**

(55 on transfers)

- 1-2 = 30/3 - transfer via node 3
route 1 with transfer trips (figure 5.12)

$$PH_1 = (1/60) \times 10 \times (84+45) = 21.5 \quad \text{[passenger-hours]}$$

$$WH_1 = 117 \times 50 / (2 \times 84) = 34.82 \quad \text{[passenger-hours]}$$

$$UH_1 = (1/60) \times [84 \times 20 - 10 \times (84+45)] = 6.5 \quad \text{[space-hours]}$$

route 4 with transfer trips (figure 5.13)

$$PH_2 = (1/60) \times (10 \times 90 + 15 \times 80 + 20 \times 51) = 52.0 \quad \text{[passenger-hours]}$$

$$WH_2 = 158 \times 50 / (2 \times 90) = 43.89 \quad \text{[passenger-hours]}$$

$$UH_2 = (1/60) \times [90 \times 45 - (10 \times 90 + 15 \times 80 + 20 \times 51)] = 15.5 \quad \text{[space-hours]}$$

$$Z = 4.4 \times (21.5 + 52.0) + 1.0 \times (34.82 + 43.89) + 4.0 \times (6.5 + 15.5) = 490.11$$

5.3.4 Part 3 - Route network improvement

In this part attempts are made to improve the route network.

Step 8: Replacing transfer trips with direct service. First the transfer paths are ranked according to the transfer path evaluation function, which is based on the total travel time on the transfer including both waiting and riding times. In this case all three transfer paths include both routes 1 and 4:

$$w_1 = (60/2) \times 50 / 84 = 18 \quad \text{[minutes]}$$

$$w_4 = (60/2) \times 50 / 90 = 17 \quad \text{[minutes]}$$
\[ Z^1 = (1/60) \times 30 \times (55 - 20) = 17.5 \quad [\text{passenger-hours}] \]

\[ Z^2 = (1/60) \times 25 \times (60 - 20) = 16.67 \quad [\text{passenger-hours}] \]

\[ Z^3 = (1/60) \times 25 \times (65 - 20) = 18.75 \quad [\text{passenger-hours}] \]

Transfer trips 2 and 3 are serving the same node pair, so in the improvement procedure only one will be considered. This is only in cases where the transfer path includes only two routes.

Transfer path 3 is considered first. Since nodes 2 and 4 are not terminal nodes they have to be connected to terminals. The connection that yields the shortest route is 3-4-2-5. This route (denoted route 5) is added to the route network, and the demand is assigned to the appropriate routes. The objective function for the network is re-calculated (figure 5.14):

route 1

\[ PH_1 = (1/60) \times 10 \times (72+22) = 15.67 \quad [\text{passenger-hours}] \]

\[ WH_1 = 82 \times 50/(2 \times 72) = 28.47 \quad [\text{passenger-hours}] \]

\[ UH_1 = (1/60) \times [72 \times 20 - 10 \times (72+22)] = 8.33 \quad [\text{space-hours}] \]

route 4

\[ PH_2 = (1/60) \times (10 \times 90 + 15 \times 61 + 20 \times 31) = 40.58 \quad [\text{passenger-hours}] \]

\[ WH_2 = 119 \times 50/(2 \times 90) = 33.06 \quad [\text{passenger-hours}] \]

\[ UH_2 = (1/60) \times [90 \times 45 - (10 \times 90 + 15 \times 61 + 20 \times 31)] = 26.92 \]
figure 5.14 ADDING ROUTE 5 TO THE NETWORK

**ROUTE 1**

\[ P = 82 \text{ passengers} \]

(30 on transfers)

**ROUTE 4**

\[ P = 119 \text{ passengers} \]

(30 on transfers)

**ROUTE 5**

\[ P = 49 \text{ passengers} \]
route 5

\[ \text{PH}_S = (1/60) \times (15 \times 7 + 20 \times 32 + 10 \times 17) = 15.25 \quad \text{[passenger-hours]} \]

\[ \text{WH}_S = 49 \times 50 / (2 \times 32) = 38.28 \quad \text{[passenger-hours]} \]

\[ \text{UH}_S = (1/60) \times [32 \times 45 \times (15 \times 7 + 20 \times 25 + 10 \times 10)] = 8.75 \quad \text{[space-hours]} \]

\[ Z = 4.4 \times (15.67 + 40.58 + 15.25) + 1.0 \times (28.47 + 33.06 + 38.28) + 4.0 \times (8.33 + 26.92 + 8.75) = 590.41 \]

Thus, replacing transfers 2 and 3 with direct service does not improve the route network. This is mainly because the route proposed to provide the direct service has very low service frequency.

Transfer path 1 has no excess riding time (has the same riding time as the shortest path) so providing a direct service is going to save mainly the waiting time for the passengers. Introducing a new route, that will have low service frequency, is not likely to improve the route network.

Step 9: Adding a node to a route. The values of the objective function for both routes is below the threshold for accepting a route so no improvement is attempted on the individual routes.

The final solution according to the heuristic algorithm consists of two routes (see figures 5.12 and 5.13):
route 1: 3-2-5
route 4: 1-3-4-5

There are two active transfer paths, from 1 to 2 on routes 1 and 2 via node 3 and from 2 to 4 on routes 1 and 2 via nodes 5 and 3. The value of the objective function is \( Z = 490.11 \).

5.4 The optimal solution according to the MINOS package

This section presents the optimal solution obtained from the MINOS package. The detailed formulation with all the variables and constraints is presented in appendix C.

The initial set of feasible routes for the MINOS formulation included the following nine routes (riding time between 20 and 50 minutes):

Between terminal nodes 1 and 5:
1) 1-2-5
2) 1-3-2-5
3) 1-3-4-5

Between terminal nodes 1 and 3:
4) 1-2-3

Between terminal nodes 3 and 5:
5) 3-2-5
6) 3-2-4-5
7) 3-4-5
8) 3-4-2-5
9) 3-1-2-5

The optimal solution according to the MINOS package includes two routes:
figure 5.15 THE MINOS SOLUTION

ROUTE 1

P = 180 passengers
(55 on transfers)

ROUTE 2

P = 95 passengers
(55 on transfers)
route 1: 1-3-2-5
route 2: 3-4-5

There are two active transfer paths, from 1 to 4 on routes 1 and 2 via node 3 and from 2 to 4 on routes 1 and 2 via nodes 3 and 5. The value of the objective function is calculated below (figure 5.15):

route 1

$$PH_1 = \frac{(1/60)\times 10\times (90+97+48)}{2} = 39.17 \quad \text{[passenger-hours]}$$

$$WH_1 = \frac{180\times 50}{2\times 97} = 46.39 \quad \text{[passenger-hours]}$$

$$UH_1 = \frac{(1/60)\times [97\times 30 - 10\times (90+97+48)]}{2} = 9.33 \quad \text{[space-hours]}$$

route 2

$$PH_2 = \frac{(1/60)\times (15\times 67+20\times 38)}{2} = 29.42 \quad \text{[passenger-hours]}$$

$$WH_2 = \frac{95\times 50}{2\times 67} = 35.45 \quad \text{[passenger-hours]}$$

$$UH_2 = \frac{(1/60)\times [67\times 35 - (15\times 67+20\times 38)]}{2} = 9.67$$

$$Z = 4.4\times (39.17+29.42) + 1.0\times (46.39+35.45) + 4.0\times (9.33+9.67) = 459.64$$

The running time for the program was approximately 70 seconds.

5.5 Comparing the solutions

The two solutions are shown in figure 5.16. The routes chosen in both cases are very similar and the difference between the values of the objective function is relatively small (6.6 percent). The reason that the heuristic approach did not produce the optimal solution can be found in
figure 5.16: COMPARING THE SOLUTIONS

THE HEURISTIC APPROACH SOLUTION

\[ Z = 490.11 \]

THE MINOS SOLUTION

\[ Z = 459.64 \]
the procedures for selecting non shortest-path routes, namely connecting
the network by adding or swapping nodes (steps 4 and 5 in the algorithm).
The heuristic approach is choosing the route to connect the disconnected
node to, according to the highest demand between the disconnected node
and the nodes on the route. This is done in an attempt to provide as much
direct service as possible, but in some cases it introduces a certain
bias towards routes with more nodes because they are likely to have more
demand to and from the disconnected node. When a node on an existing
route is replaced with a disconnected node, some of the demand associated
with the replaced node may have to be satisfied by transfer rather than
direct service. So the connection of a disconnected node to the network
is saving transfer trips on the disconnected node side but may add trans-
fer trips to and from the node it replaced.

In the example network the difference between the two solutions is
that in the heuristic approach node 4 was replaced with node 2 on route 1
(which has more demand to and from node 4 because it has more nodes)
rather than on route 4. Furthermore, the second possibility was not
considered at all. One way to overcome this problem is to consider all
possible connections. This method is inefficient and involves high compu-
tation costs which this approach in this work is trying to avoid. Another
approach is to look at the total number of transfer trips saved (or
created) by each connection, and choose the one for which this number is
maximized (or minimized). This may be very much dependent on the network
structure. In the example network both connections (node 4 to route 1 or
route 4) save 30 transfer trips and add 30 transfer trips so this measure
does not help in choosing between the two.

In order to try and better understand the differences between the
two solutions, the following table summarizes the values of the three
components of the objective function (without the weights) for each
network (the lower value is highlighted, difference in parenthesis):
<table>
<thead>
<tr>
<th>MINOS</th>
<th>HEURISTIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>PH</td>
<td>68.59</td>
</tr>
<tr>
<td>WH</td>
<td>81.84</td>
</tr>
<tr>
<td>UH</td>
<td>19.00</td>
</tr>
<tr>
<td>TOTAL</td>
<td>169.43</td>
</tr>
</tbody>
</table>

It is evident from this table that the network produced by MINOS is a better one since it has less riding time and unutilized capacity even though it has higher total waiting time. This comparison is done without the weights in order to capture the effect of the network parameters without the cost coefficients. Both routes have excess riding time on both direct and transfer service. The difference in PH is because the demand between nodes 1 and 5 is satisfied in the heuristic solution by the direct route which is 50 percent longer than the shortest route. As for the waiting times the heuristic solution has less waiting time but the distribution of the demand among the arcs on each route, which causes the relatively high frequencies and low waiting times, also results in poor utilization of equipment which shows in the high UH value.

To sum up this discussion, it is hard to find a criterion that will always indicate the best choice in terms of route selection without enumerating all the possibilities. This choice is dependent, to a large extent, on the network structure. Even though direct service is usually better than transfer service, in some cases replacing a direct service with a transfer path may yield a network with a lower value of the objective function. All in all the solution obtained by the heuristic approach is very close to the optimal solution and involves much less work.
CHAPTER 6

SUMMARY AND CONCLUSIONS

6.1 Summary and conclusions

This work can be divided into three parts. In the first part the Transit Network Design problem was defined and prior work on this subject was reviewed. Hasselstrom's work is the only one so far that has been fully implemented. Since this implementation is expensive to execute, there is a need for an acceptable, easy to implement, inexpensive approach.

In the second part of this work the Transit Network Design problem was formulated as an optimization program. This was done for two reasons. First in order to demonstrate the size of the problem and the fact that it can not be solved optimally for real life network problems. Second in an attempt to learn more about the structure of the problem and the desired solution approach by comparing the optimal solution with a heuristic solution for a small network example.

The third part of the work presented a new heuristic approach for the solution of the Transit Network Design problem, based on generating a relatively small initial set of shortest routes between terminal nodes that are altered later in an attempt to improve the route network. A route generation algorithm was presented and applied to a small example network. This example network problem was also solved optimally by a non-linear optimization computer package.
The formulation of the example network problem, with 5 nodes and 7 arcs, as an optimization program required 107 variables and 144 constraints. This indicates that large, real life Transit Network Design problems can not be solved optimally because of the size of the problem.

The solution obtained by applying the heuristic algorithm is very similar to the optimal solution obtained by using the MINOS optimization package. The difference in the network evaluation function between the solutions is less than seven percent. Even though the example network is small the similarity of the solutions implies that this approach is in the right direction.

The main feature of this algorithm, which is not present in prior work, is the fact that it is generating a small set of routes and therefore the cost involved in applying it, even to large network problems, is likely to be low. Since the number of routes generated is small, and close to the number of actual routes in the final solution, a critical part of the algorithm is the selection of the pairs of terminals between which the routes are generated. Different selection strategies will yield different route networks. Another important part in the algorithm is the procedure for connecting the network. In this procedure non shortest path routes are selected to connect the network so all the demand can be satisfied. Thus, the excess time spent by passengers on the network depends, to large extent, on this part.

The reason that the heuristic algorithm did not produce the optimal solution is the procedure for connecting the route network. In this procedure the algorithm selects the connections based on the demand between the routes and the disconnected node. This approach does not always minimize the value of the objective function. Other criteria for selecting routes were examined briefly in chapter 5, none of which gave clear indication of the best selection method. It is probably dependent,
to large extent, on the structure of the network.

6.2 Future research

Since the weak spot in the algorithm is the procedure for connecting disconnected nodes, more work can be done on trying to improve this procedure. One way of doing this is to look at how the different components of the objective function are effected by different connection selection strategies. The selection of terminal node pairs for the shortest routes generation procedure is also an important part of the algorithm, and has high impact on the outcome. Different strategies for selecting these terminal nodes can be tested and evaluated in order to find the best one.

More work can be done in developing this approach further so that some of the assumptions made here can be relaxed. The demand matrix was treated as fixed and one possible direction for further research is to incorporate variable demand. Another direction may be to try the algorithm on large, real life networks and perform some sensitivity analysis of the algorithm to various parts in it like the selection of terminal nodes in part 1, or the selection of transfer paths and routes to be improved in the network improvement part. The effect of the network structure on the outcome of the algorithm can be tested by applying the algorithm to networks with different structures.

The example network presented in this work is very small. An attempt can be made to optimally solve a larger network example, and compare it with a solution obtained by applying the heuristic algorithm. This may give some more insight on the problem and help to better understand its characteristics.
REFERENCES

pp.331-344.

"A Set of Methods in Transportation Network Synthesis and Analysis".

Friedlander Alex E., (1968).
"A Method of Schedule and Route Planning in Urban Mass Transit".

"Public Transportation Planning - Mathematical Programming
Approach". Phd. thesis, Department of Business Administration,
University of Gothenburg, Gothenburg, Sweden.

"Design of Local Bus Service with Demand Equilibrium". Transpn. Sci.
16.

Lampkin W., Saalmans, P.D. (1967).
"The Design of Routes, Service Frequencies and Schedules for a
No. 4, pp. 375-397.

"Evaluation and Optimization of Urban Public Transportation Net-
works". Presented at EURO III in Amsterdam, The Netherlands, April
9-11, 1979.


"Designing Urban Transit Systems: An Approach to the Route-Technology Selection Problem". PB 204 881, University of Washington, Seattle, WA.


"Application of a Route Structing Algorithm to City Bus Lines". Georgia Institute of Technology, Atlanta, Georgia, (1974).

APPENDIX A

DIVIDING THE DEMAND ON A COMMON PART BETWEEN TWO ROUTES

This appendix details the procedure of dividing the demand between two routes that share a common part, and shows that this procedure converges. For simplicity, but without loss of generality, the example here has two routes, 1 and 2, and a common part consisting of a single arc. If the common part has more than one arc, each pair of nodes on the common part will be treated separately. For example, if the common part has three nodes, o-l-m, the demand will be divided into three parts, o->1, o->m and l->m and each one of them will be treated separately as one arc.

The common arc is denoted i-j. For simplicity the load on this arc is assumed to consist only of passengers traveling between nodes i and j. This is done without any loss of generality because all other passengers, going through arc i-j, can be represented by a constant and do not effect the calculations.

At the beginning all the demand between nodes i and j is assigned to route 1. The load on arc i-j of route 2 is therefore 0. At each iteration the demand is divided between the routes proportionally to the frequencies ratio. As the demand is modified the frequencies change and yield yet another change in the demand assignment which causes another change in the frequencies. This iterative procedure continues until there is no change in the frequencies of either route from one iteration to the next. The demand is rounded, at each iteration to the nearest integer.
Notation

\( d_{ij} \) - The total demand between nodes i and j. [passengers/hour].

\( d^n_{i,j,k} \) - The demand between nodes i and j, satisfied by route k, at iteration n. [passengers/hour].

\( f^n_k \) - The frequency on route k at iteration n. [vehicles/hour].

\( F^n \) - The sum of the frequencies on both routes at iteration n.

\( F^n = f^n_1 + f^n_2 \) [vehicles/hour].

\( C \) - Vehicle capacity. [passengers/vehicle].

\( L_1 \) - The highest arc load on route 1 at iteration 0. [passengers].

\( L_2 \) - The highest arc load on route 2 at iteration 0. [passengers].

\( L^n_k \) - The highest arc load on route k at iteration n, n \( \neq 0 \). [passengers].

The analysis will be divided into two cases depending on the relation between the highest arc load on route 1, \( L_1 \), and the total demand between nodes i and j, \( d_{ij} \):

In case 1 the load on arc i-j is smaller than (or equal to) the highest arc load on route 1, \( L_1 \geq d_{ij} \), so the frequency on route 1 is not determined by arc i-j but by another arc on route 1 (which is not on route 2).

In case 2 the frequency on route 1 is determined, initially, by the load on arc i-j, \( L_1 = d_{ij} \).

Each of the aforementioned cases is further divided into two sub-cases according to the relation between the highest arc load on route 2, \( L_2 \), and the total demand between nodes i and j, \( d_{ij} \):

a. \( L_2 \geq d_{ij} \). The frequency on route 2 will never be determined by the load on arc i-j.
b. \( L_2 < d_{ij} \). It is possible that the load on arc \( i-j \) will exceed \( L_2 \) in which case it will determine the frequency on route 2.

Within the two sub-cases, a and b, it is important weather or not the route frequency is determined by the load on arc \( i-j \).

**Case 1: \( L_1 \geq d_{ij} \), frequency on route 1 is not determined by arc \( i-j \).**

The initial conditions (iteration 0) are the same for all sub-cases.

**Iteration 0**

\[
d^{0}_{ij1} = d_{ij} \quad [A.1]
\]

\[
f^{0}_{1} = (1/C)*L_{1} \text{ (Fixed because } L_{1} \geq d_{ij}). \quad [A.2]
\]

\[
d^{0}_{ij2} = 0 \quad [A.3]
\]

\[
f^{0}_{2} = (1/C)*L_{2} \quad [A.4]
\]

**Iteration 1**

\[
d^{1}_{ij1} = (f^{0}_{1}/(f^{0}_{1} + f^{0}_{2}))*d_{ij} \quad [A.5]
\]

\[
f^{1}_{1} = (1/C)*L_{1} \text{ (No change from iteration 0).} \quad [A.6]
\]

\[
d^{1}_{ij2} = (f^{0}_{2}/(f^{0}_{1} + f^{0}_{2}))*d_{ij} \quad [A.7]
\]

**Case 1.a. \( L_2 \geq d_{ij} \)**

\[
f^{1}_{2} = (1/C)*L_{2} \quad [A.8]
\]

In this case there is no change in the frequencies from iter-
ation 0 to iteration 1 so this is the final distribution of the
demand on arc i-j between routes 1 and 2:

route 1 - \( d_{ij}^1 = \left( f_1^0 / (f_1^0 + f_2^0) \right) * d_{ij} \)
route 2 - \( d_{ij}^2 = \left( f_2^0 / (f_1^0 + f_2^0) \right) * d_{ij} \)

**Case 1.b. \( L_2 \leq d_{ij} \)**

In this case the frequency on route 2 will change only if the
part of the demand between nodes i and j transferred from route
1 to route 2 in iteration 1, \( d_{ij}^1 \), is greater than the
initial highest arc load on route 2, \( L_2 \). The following is a
proof that this can not happen and \( d_{ij}^1 - L_2 \) is always
negative.

From iteration 1:

\[
d_{ij}^1 = \left( f_2^0 / (f_1^0 + f_2^0) \right) * d_{ij} \quad \text{[A.9]}
\]

But from iteration 0:

\[
f_1^0 = (1/C) * L_1 \quad \text{and} \quad f_2^0 = (1/C) * L_2 \quad \text{so:} \quad \text{[A.10]}
\]

\[
d_{ij}^1 = \left( (1/C) * L_2 / [(1/C) * L_1 + (1/C) * L_2] \right) * d_{ij} = \left[ L_2 / (L_1 + L_2) \right] * d_{ij} \quad \text{[A.11]}
\]

The highest value that \( d_{ij}^1 \) can get is when \( d_{ij} = L_1 \). Than:

\[
d_{ij}^1 = \left[ L_2 / (L_1 + L_2) \right] * L_1 = L_2 * L_1 / (L_1 + L_2) \quad \text{[A.12]}
\]

In this case the difference between \( d_{ij}^1 \) and \( L_2 \) is:

\[
d_{ij}^1 - L_2 = L_2 * L_1 / (L_1 + L_2) - L_2 = \left[ L_2 * L_1 - L_2 * (L_1 + L_2) \right] / (L_1 + L_2) = \left[ L_2 * L_1 - L_2 * L_1 - L_2 * L_2 \right] / (L_1 + L_2) = -L_2 * L_2 / (L_1 + L_2) \quad \text{[A.13]}
\]
This means that in case 1 it is always true that $d^1_{ij2} < L_2$.

Since in case 1 $L_1 \geq d_{ij}$ and $L_2 > d^1_{ij2}$ the system will always achieve equilibrium after one iteration. The reason is that the frequencies on both routes will not change from iteration 0 to iteration 1 because they are not determined by the load on arc $i-j$.

**Case 2:** $L_1 = d_{ij}$ frequency on route 1 is determined initially by arc $i-j$.

The initial conditions (iteration 0) are the same for all sub-cases.

### Iteration 0

$$d^0_{ij1} = d_{ij}$$  \[A.14\]

$$f^0_1 = (1/C)*d^0_{ij1} = (1/C)*d_{ij} = (1/C)*L_1$$  \[A.15\]

$$d^0_{ij2} = 0$$  \[A.16\]

$$f^0_2 = (1/C)*L_2$$  \[A.17\]

### Iteration 1

$$d^1_{ij1} = (f^0_1/(f^0_1 + f^0_2))*d_{ij}$$  \[A.18\]

$$d^1_{ij2} = (f^0_2/(f^0_1 + f^0_2))*d_{ij}$$  \[A.19\]

**Case 2.a, $L_2 \geq d_{ij}$**

$$f^1_2 = (1/C) * L_2$$  (No change from iteration 0).  \[A.20\]

In this case the frequency on route 2 will not change because it is determined by an arc other than arc $i-j$ which has an arc
load higher than $d_{ij}$. There will be a shift of demand on arc $i-j$ from route 1 to route 2 in iteration 1. There are two cases regarding the demand between nodes $i$ and $j$ satisfied by route 1.

1) $d_{ij1}$ is not the highest arc load on route 1, and $L_{11}$ is the highest arc load on route 1 (now $L_{11}$ determines the frequency on route 1):

$$f_{1}^{1} = (1/C)*L_{11}^{1} \quad [A.21]$$

and another iteration is performed.

**Iteration 2**

$$d_{ij1}^{2} = [f_{1}^{1}/(f_{1}^{1} + f_{1}^{2})]*d_{ij} \quad [A.22]$$

$$f_{1}^{2} = (1/C)*L_{11}^{1} \quad \text{(No change from iteration 1).} \quad [A.23]$$

$$d_{ij2}^{2} = [f_{2}^{1}/(f_{1}^{1} + f_{1}^{2})]*d_{ij} \quad [A.24]$$

$$f_{2}^{2} = (1/C)*L_{2} \quad \text{(No change from iteration 0).} \quad [A.25]$$

In this case there is no change in the frequencies from iteration 1 to iteration 2 so the final distribution of demand on arc $i-j$ between routes 1 and 2 is as shown at iteration 2 above.

2) $d_{ij1}$ is still the highest arc load on route 1, (it still determines the frequency on route 1). As long as this is true, there will be a shift of demand from route 1 to route 2. This is due to the fact that the frequency on route 2 ($L_{2} \geq d_{ij}$) is higher than the frequency on route 1.
(L₁ = dᵢⱼ) which is decreasing.

When, at iteration n, dⁿᵢⱼ₁ will no longer be the highest arc load on route 1, the frequency on route 1 will be determined by the load on another arc, Lⁿ₁. In this case at iteration n+1 the frequencies on both routes will not change and the system will be at equilibrium (similar to case 2.a.1).

**Case 2.b. L₂ < dᵢⱼ**

Since on both routes the load on arc i-j is smaller than the total demand between nodes i and j, the distribution of demand on arc i-j between routes 1 and 2 dependent on weather or not the frequency on the route, at iteration 1, is determined by the load on arc i-j. There are three possibilities:

1) The route frequency on both routes is not determined by the load on arc i-j. The frequency on routes 1 and 2 is determined by the highest load arc on each route, at iteration 1, L¹₁ and L¹₂ respectively. This means that the frequencies on both routes will not change any more and the distribution of demand on arc i-j between routes 1 and 2 is the one achieved at iteration 1.

2) It was shown before (case 1.b) that the frequency on route 2 will not be determined by the load on arc i-j at iteration 1. As a result, this case is similar to case 2.b, where L₁ = dᵢⱼ and L₂ < dᵢⱼ, which was discussed earlier.

3) The route frequency on route 1 determined by the load on arc i-j. The route frequency on both routes is determined by the load on arc i-j. This will happen when the initial
load on route 2, \( L_2 \), is very small, yielding a very low frequency. In practice, this route will probably not be accepted. In the rare event that this may happen the frequencies are:

\[
\begin{align*}
 f^1_1 &= (1/C)*d^1_{ij1} \quad [A.26] \\
 f^1_2 &= (1/C)*d^1_{ij2} \quad [A.27]
\end{align*}
\]

and another iteration is performed.

**Iteration 2**

\[
\begin{align*}
 d^2_{ij1} &= [f^1_1/(f^1_1 + f^1_2)]*d_{ij} \quad [A.28] \\
 d^2_{ij2} &= [f^1_2/(f^1_1 + f^1_2)]*d_{ij} \quad [A.29]
\end{align*}
\]

from iteration 1:

\[
\begin{align*}
 f^1_1 &= (1/C)*d^1_{ij1} - (1/C)*[f^0_1/(f^0_1 + f^0_2)]*d_{ij} \quad [A.30] \\
 f^1_2 &= (1/C)*d^1_{ij2} - (1/C)*[f^0_2/(f^0_1 + f^0_2)]*d_{ij} \quad [A.31]
\end{align*}
\]

and from iteration 0:

\[
\begin{align*}
 f^0_1 &= (1/C)*d^0_{ij1} - (1/C)*d_{ij} - (1/C)*L_1 \quad [A.32] \\
 f^0_2 &= (1/C)*L_2 \quad [A.33]
\end{align*}
\]

so the expressions for \( f^1_1 \) and \( f^1_2 \) become:

\[
\begin{align*}
 f^1_1 &= (1/C)*\{(1/C)*L_1/((1/C)*L_1+(1/C)*L_2)\}*d_{ij} - (1/C)*[L_1/(L_1 + L_2)]*d_{ij} \quad [A.34]
\end{align*}
\]
\[ f^1_{2} = (1/C) \left( \frac{(1/C) \times L_2/((1/C) \times L_1 + (1/C) \times L_2)}{L_2/(L_1 + L_2)} \right) \times d_{ij} \]  
\[ = (1/C) \times \left[ L_2/(L_1 + L_2) \right] \times d_{ij} \]  
\[ \text{[A.35]} \]

but since \( d_{ij} = L_1 \)

\[ f^1_{1} = (1/C) \times \left[ L_1 \times L_1/(L_1 + L_2) \right] \] and  
\[ \text{[A.36]} \]

\[ f^1_{2} = (1/C) \times \left[ L_2 \times L_1/(L_1 + L_2) \right] \]  
\[ \text{[A.37]} \]

after substituting this in the equations for \( d^2_{ij1} \) and \( d^2_{ij2} \), and since the terms \( (f^1_{1} + f^1_{2}) \) and \( (1/C) \) cancel:

\[ d^2_{ij1} = \left[ f^1_{1}/(f^1_{1} + f^1_{2}) \right] \times d_{ij} = \left[ L_1 \times L_1/(L_1 \times L_1 + L_1 \times L_2) \right] \times d_{ij} = \]  
\[ = \left[ L_1/(L_1 + L_2) \right] \times d_{ij} = d^1_{ij1} \]  
\[ \text{[A.38]} \]

\[ d^2_{ij2} = \left[ f^1_{2}/(f^1_{1} + f^1_{2}) \right] \times d_{ij} = \left[ L_2 \times L_1/(L_1 \times L_1 + L_1 \times L_2) \right] \times d_{ij} = \]  
\[ = \left[ L_2/(L_1 + L_2) \right] \times d_{ij} = d^1_{ij2} \]  
\[ \text{[A.39]} \]

Since the demand distribution does not change from one iteration to the next, the system is at equilibrium and this demand distribution is final.

This concludes the enumeration of all possible cases regarding the distribution of the demand on a common part between two routes. It was shown that the distribution of the demand proportionally to the frequencies ratio, will always converge to an equilibrium state.
APPENDIX B

THE ROUTE EVALUATION FUNCTION

This appendix describes some properties of the route evaluation function and illustrates, with an example, the trade-offs between high level of service (high service frequency) and high route density. Another example is used to show the importance of the number of passengers in the evaluation of a route.

The efficiency of a route can be represented by its density, \( \rho \), defined as the ratio between the total load (in passenger-hours), and the desired maximum capacity (in space-hours) of the route. The desired maximum capacity of route \( k \) is simply the service frequency on the route, \( f_K \) (in vehicles/hour), times the desired maximum vehicle occupancy, \( C \) (in passengers/vehicle), multiplied by the total travel time on the route \( T_k \) (in hours). The density, \( \rho \), can have values between 0 and 1, where \( \rho = 1 \) means the best possible utilization of equipment where all spaces on all vehicles on the route are utilized at all times.

While the passengers would like to minimize the time spent on the network, the operator seeks to maximize the utilization of the available equipment (the value of the route density, \( \rho \)). These objectives are naturally conflicting since maximizing \( \rho \) may result in very low service frequencies which mean long waiting times. The following example illustrates the trade-offs between high frequency, \( f \), and high route density, \( \rho \):

120
Consider route (A) with the load profile shown in figure B.1a. It has a service frequency of 6 vehicles/hour (assuming desired maximum vehicle occupancy of 50 passengers/vehicle) so at least 13 vehicles are needed (assuming that a round trip takes exactly two hours and the drivers rest for 10 minutes after each round trip). The density of this route is $3/4$, meaning that only 75% of the route's desired maximum capacity is utilized. From the passengers point of view the level of service is relatively high with a bus every 10 minutes and mean waiting time, assumed to be half the headway, of 5 minutes per passenger. The total mean waiting time on this route, assuming a total of 300 passengers, is 2.5 hours.

Figures B.1b and B.1c show the load profiles of two separate routes, (B) and (C) respectively, that satisfy the same demand. Both routes have density 1.0 and service frequency of 3 vehicles/hour. From the operator's point of view both (B) and (C) have a better utilization of equipment than (A). All spaces on all vehicles are utilized at all times, and less vehicles are needed to satisfy the demand. Route (B) requires 4 vehicles and route (C) only 7, a total of 11 vehicles. As for the level of service, the mean waiting time per passenger is 10 minutes, twice as high as for the original route (A).

The following table summarizes these two alternatives:

<table>
<thead>
<tr>
<th></th>
<th>rho</th>
<th>vehicles</th>
<th>per pass.</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>route (B):</td>
<td>1.0</td>
<td>4</td>
<td>0.166</td>
<td>2.5</td>
</tr>
<tr>
<td>route (C):</td>
<td>1.0</td>
<td>7</td>
<td>0.166</td>
<td>2.5</td>
</tr>
<tr>
<td>routes (B) + (C):</td>
<td>1.0</td>
<td>11</td>
<td>0.166</td>
<td>5.0</td>
</tr>
<tr>
<td>original route:</td>
<td>3/4</td>
<td>13</td>
<td>0.083</td>
<td>2.5</td>
</tr>
</tbody>
</table>
figure B.1: TRADEOFFS BETWEEN DENSITY AND FREQUENCY

B.1a: original route A, \(\rho = 3/4, \ f = 6, \ T = 60\)

B.1b: route B, \(\rho = 1.0, \ f = 3, \ T = 30\)

B.1c: route C, \(\rho = 1.0, \ f = 3, \ T = 60\)
Clearly routes (B) and (C) are a better solution from the operator's perspective but a much worse one from the passengers point of view.

The mean waiting time for passengers boarding a single route is assumed to be half the mean headway between vehicles on the route. This is due to the assumptions that passengers arrive randomly at bus stops but the buses arrive at regular deterministic intervals:

\[ w_k = \min(C \times \left\{ \frac{60}{2 \times q_k} \right\}, \frac{60}{2 \times F^m_k}) \]  \hspace{1cm} [B.1]

where \( w_k \) is the waiting time for a single passenger on route \( k \) [minutes], \( C \) is the desired maximum vehicle occupancy [passengers/vehicle], \( q_k \) is the highest arc load on route \( k \) [passengers/hour] and \( F^m_k \) is the minimum required frequency on route \( k \) [vehicles/hour] based on policy headways.

The measure of waiting time for the entire route \( k \), \( WH_k \), [passenger-hours] would be:

\[ WH_k = \frac{1}{60} \times P_k \times w_k \]  \hspace{1cm} [B.2]

where \( P_k \) is the total number of passengers on route \( k \), not including transfer passengers, and is defined as:

\[ P_k = \sum_{(i,j) \in k} d_{ijk} \]  \hspace{1cm} [B.3]

where \( d_{ijk} \) is the demand between nodes \( i \) and \( j \) satisfied by route \( k \) [passengers/hour], and the expression \( (i,j) \in k \) means "... over all node pairs \( i,j \) that belong to route \( k \)". For example on the route 1-3-7-4, 1,7 is a node pair but 7,1 is not.

As for riding time, since there is a minimum riding time between each pair of nodes on the network, it is sufficient to take into account
only the excess time, over the shortest possible riding time, spent on the route by all passengers. The total time spent on route $k$ by all passengers (excluding transfer passengers which are considered later), $PH_k$, is [passenger-hours]:

$$PH_k = (1/60) \sum_{(i,j) \in k} d_{ijk} t_{ijk} \tag{B.4}$$

where $t_{ijk}$ is the riding time between nodes $i$ and $j$ on route $k$ [minutes].

The minimum possible riding time for these passengers, $PH^m_k$, is:

$$PH^m_k = (1/60) \sum_{(i,j) \in k} d_{ijk} t^m_{ij} \tag{B.5}$$

where $t^m_{ij}$ is the minimum riding time between nodes $i$ and $j$ [minutes].

The excess riding time over the minimum riding time, for all non-transfer passengers on route $k$ [passenger-hours], is therefore:

$$\Delta PH_k = PH_k - PH^m_k = (1/60) \sum_{(i,j) \in k} d_{ijk} (t_{ijk} - t^m_{ij}) \tag{B.6}$$

The operator’s measure is called unutilized-space-hours, $UH_k$, and is defined as the difference between the route's desired maximum capacity [space-hours] and the route load [passenger-hours]:

$$UH_k = (1/60) q_k T_k - PH_k \tag{B.7}$$

where $T_k$ is the total riding time on route $k$ [minutes].

The density of route $k$, $\rho$, is defined as:
\[ \rho_k = (60 \times PH_k)/(q_k \times T_k) \]  \[ \text{[B.8]} \]

Substituting \( PH_k \) in equation B.7 with \((1/60) \times \rho_k \times q_k \times T_k\) from equation B.8 the unutilized-space-hours term \( UH_k \) becomes:

\[ UH_k = (1/60) \times (q_k \times T_k - \rho_k \times q_k \times T_k) = (1/60) \times (1 - \rho) \times q_k \times T_k \]  \[ \text{[B.9]} \]

Equation B.9 means that usually a route that has a high \( UH \) value will have low \( \rho \) value and vice versa. In case a minimum frequency, \( F^m_k \), has to be maintained on route \( k \), the measure unutilized-space-hours becomes:

\[ UH_k = \max \left\{ (1/60) \times q_k \times T_k - PH_k, (1/60) \times F^m_k \times C \times T_k - PH_k \right\} \]  \[ \text{[B.10]} \]

As mentioned above, the operator wishes to minimize \( UH \) while the passenger wishes to minimize both \( \Delta PH \) and \( WH \). Hence, one possible evaluation function is simply the sum of all the aforementioned measures:

\[ Z_k = \Delta PH_k + WH_k + UH_k \]  \[ \text{[B.11]} \]

A major drawback of this function is the fact that it does not directly incorporate the number of passengers using the route. This may cause misleading results as shown in the following example.

Consider two shortest path routes with the load profiles shown in figures B.2a and B.2b.
figure B.2: COMPARISON OF TWO ROUTES

B.2a: Load profile of route 1, \( P = 60 \text{ pass.}, \rho = 1/3, \ Z = 80 \text{ pass.-hrs.} \)

B.2b: Load profile of route 2, \( P = 125 \text{ pass.}, \rho = 2/3, \ Z = 162 \text{ pass.-hrs.} \)
The parameters of these routes are:

<table>
<thead>
<tr>
<th>route 1</th>
<th>route 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_1 = 20$ passengers</td>
<td>$P_2 = 40$ passengers</td>
</tr>
<tr>
<td>$q_1 = 60$ passengers/hour</td>
<td>$q_2 = 60$ passengers/hour</td>
</tr>
<tr>
<td>$T_1 = 50$ minutes</td>
<td>$T_2 = 110$ minutes</td>
</tr>
<tr>
<td>$P_1 = 60$ passengers</td>
<td>$P_2 = 125$ passengers</td>
</tr>
<tr>
<td>$\rho_1 = 1/3$</td>
<td>$\rho_2 = 2/3$</td>
</tr>
<tr>
<td>$UH_1 = 33.33$ place-hours</td>
<td>$UH_2 = 36.66$ place-hours</td>
</tr>
<tr>
<td>$PH_1 = 21.66$ passenger-hours</td>
<td>$PH_2 = 73.33$ passenger-hours</td>
</tr>
<tr>
<td>$WH_1 = 25$ passenger-hours</td>
<td>$WH_2 = 52.08$ passenger-hours</td>
</tr>
</tbody>
</table>

where $P_k$ is the average passenger load on route $k$, defined as:

$$P_k = \frac{PH_k}{T_k} \quad [B.12]$$

and $P_k$ is the total number of non transfer passengers on route $k$ as defined in equation B.3.

Since both routes are along the shortest path $\Delta PH_1 = \Delta PH_2 = 0$ and:

$$Z_1 = \Delta PH_1 + WH_1 + UH_1 = 0 + 25 + 33.33 = 58.33 \quad [\text{pass.-hours}] \quad [B.13a]$$
$$Z_2 = \Delta PH_2 + WH_2 + UH_2 = 0 + 52.08 + 36.66 = 88.74 \quad [\text{pass.-hours}] \quad [B.13b]$$

In the example above $UH$ fails to indicate which route has higher density $\rho$. Even though $\rho_2 > \rho_1$, $UH_2 > UH_1$ implying that route 1 has a better (higher) $\rho$ value. Furthermore the value of the route evaluation function, $Z$, is much lower for route 1 than for route 2 implying, again, that route 1 is better than route 2. Actually route 2 is much better because it serves more than twice as many passengers (125 as opposed to
60 on route 1) and it utilizes the equipment more efficiently ($\rho_2 = 2/3$ when $\rho_1 = 1/3$ only). Route 2 is "penalized", according to this evaluation function, for serving more passengers.

In order to overcome these problems, some measure of the load on the route must be incorporated into the route evaluation function. This can be done by dividing all the terms in the function by the total number of non-transfer passengers on the route, which will yield the following results:

1. The term $UH/P$ will be the average number of unutilized-space-hours per passenger on the route.

2. The term $WH/P$ will simply be the average waiting time experienced by each passenger on the route.

3. The term $\Delta PH/P$ will be the average excess riding time per passenger on the route.

All these terms will be expressed in time units.

The modified terms of the route evaluation function for the examples in figure B.2 become:

\[
\begin{align*}
\rho_1 &= 1/3 \\
UH_1/P_1 &= 0.555 \text{ hours} \\
WH_1/P_1 &= 0.417 \text{ hours} \\
\Delta PH_1/P_1 &= 0 \text{ hours} \\
Z_1/P_1 &= 0.972 \text{ hours}
\end{align*}
\]

\[
\begin{align*}
\rho_2 &= 2/3 \\
UH_2/P_2 &= 0.293 \text{ hours} \\
WH_2/P_2 &= 0.417 \text{ hours} \\
\Delta PH_2/P_2 &= 0 \text{ hours} \\
Z_2/P_2 &= 0.71 \text{ hours}
\end{align*}
\]

where $UH_1/P_1 > UH_2/P_2$ which indicates correctly the fact that $\rho_1 < \rho_2$. $WH/P$ is the waiting time experienced by every passenger on the route. In
this example both routes have the same waiting time because the maximum arc load, which determines the waiting time, is equal (60 passengers/hour).

After dividing the route evaluation function by the total number of passengers using the route, it succeeds in indicating that route 2 is better than route 1, \( Z_2 < Z_1 \). Now the function gives the average time cost per passenger for the route, for both the user and the operator.

Different components of the route evaluation function may have different significance in the route evaluation process. Waiting time, for example, may be more important to passengers than riding time, so it should receive more weight in the function. Similarly, the operator's measure, \( \text{UH} \), may be of greater importance than both waiting and travel times. In order to capture these trade-offs between the components, weights, to be determined by the operator, may be added to the route evaluation function.

Based on the discussion above, the route evaluation function used in this work is:

\[
Z_k = \left( \theta \Delta PH_k + \mu WH_k + \varphi UH_k \right) / P_k \tag{B.14}
\]

where \( \theta \), \( \mu \) and \( \varphi \) are the weights for \( \Delta PH \), \( WH \) and \( UH \) respectively.
This appendix details the formulation of the network example. It starts with a description of the network with all the feasible routes and all feasible transfer paths. Next the objective function and the constraints are fully stated, followed by the output of the MINOS program.

C.1 Network description

The network has five nodes, three of which are terminal nodes, and seven arcs (see figure 5.1). Travel on the arcs is possible in both directions.

All feasible routes (riding time between 20 and 50 minutes) by nodes:

Between terminal nodes 1 and 5:
1) 1→2→5
2) 1→3→2→5
3) 1→3→4→5

Between terminal nodes 1 and 3:
4) 1→2→3

Between terminal nodes 3 and 5:
5) 3→2→5
6) 3→2→4→5
7) 3→4→5
8) 3→4→2→5
9) 3→1→2→5

All feasible transfer paths, limited to two routes on each transfer path:

<table>
<thead>
<tr>
<th>node pair</th>
<th>first part</th>
<th>possible routes</th>
<th>second part</th>
<th>possible routes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1→2</td>
<td>1-3</td>
<td>2, 3</td>
<td>3-2</td>
<td>2, 4, 5, 6</td>
</tr>
<tr>
<td></td>
<td>1-3-4</td>
<td>3</td>
<td>4-2</td>
<td>6, 8</td>
</tr>
<tr>
<td>1→3</td>
<td>1-2</td>
<td>1, 4</td>
<td>2-3</td>
<td>2, 4, 5, 6</td>
</tr>
<tr>
<td>1→4</td>
<td>1-2</td>
<td>1, 4</td>
<td>2-4</td>
<td>6, 8</td>
</tr>
<tr>
<td></td>
<td>1-3</td>
<td>2, 9</td>
<td>3-4 or 3-2-4</td>
<td>7, 8 or 6</td>
</tr>
<tr>
<td></td>
<td>1-2-5</td>
<td>1, 9</td>
<td>5-4</td>
<td>3, 6, 7</td>
</tr>
<tr>
<td></td>
<td>1-3-2-5</td>
<td>2</td>
<td>5-4</td>
<td>3, 6, 7</td>
</tr>
<tr>
<td>1→5</td>
<td>1-2</td>
<td>4</td>
<td>2-5</td>
<td>5, 8</td>
</tr>
<tr>
<td>2→3</td>
<td>2-5</td>
<td>1</td>
<td>5-4-3</td>
<td>7</td>
</tr>
<tr>
<td>2→4</td>
<td>2-1</td>
<td>1, 4, 9</td>
<td>1-3-4</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>2-3 or 2-1-3</td>
<td>2, 4, 5 or 9</td>
<td>3-4</td>
<td>3, 7</td>
</tr>
<tr>
<td></td>
<td>2-5</td>
<td>1, 2, 5, 9</td>
<td>5-4</td>
<td>3, 7</td>
</tr>
<tr>
<td>2→5</td>
<td>2-3</td>
<td>4</td>
<td>3-4-5</td>
<td>3, 7</td>
</tr>
<tr>
<td>3→4</td>
<td>Ø</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3→5</td>
<td>Ø</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4→5</td>
<td>Ø</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

C.2 The objective function

The objective function to be minimized:

\[
Z = 0.4 \sum_{k \in K} \sum_{(i,j) \in k} (d_{ijk} + d_{ijk}^n) t_{ijk} + \sum_{k \in K} \left( \frac{1}{2^*f_k} \right) \left[ \sum_{k \in K} (d_{ijk} + d_{ijk}^n) \right] + (4/60) \sum_{k \in K} f_k \cdot C \cdot T_k \quad \text{[C.1]}
\]
\[ Z = 0.4\left[ (d_{121} + d_{131}^2 + d_{141}^2 + d_{241}^1)\times20 + (d_{151} + d_{141}^5)\times30 + (d_{251} + d_{231}^5 + d_{241}^5)\times30 \right] + \\
(1/2*f_1)\left[ (d_{121} + d_{151} + d_{141}^2 + d_{131}^2 + d_{141}^2 + d_{141}^5 + d_{231}^5 + d_{241}^1 + d_{241}^5) + f_1*T_1*(200/60) \right] + \\
0.4\left[ (d_{132} + d_{142}^3)\times10 + d_{122}\times20 + (d_{152} + d_{142}^5)\times30 + (d_{222} + d_{242}^3) + d_{352}\times20 + (d_{252} + d_{242}^5)\times10 \right] + \\
(1/2*f_2)\left[ (d_{132} + d_{142} + d_{152} + d_{322} + d_{352} + d_{252} + d_{142}^3 + d_{142}^5 + d_{242}^3 + d_{242}^5) + f_2*T_2*(200/60) \right] + \\
0.4\left[ (d_{133} + d_{123}^3)\times10 + (d_{143} + d_{123}^4 + d_{243}^1)\times25 + (d_{153} + d_{143}^5 + d_{243}^5)\times15 + (d_{353} + d_{253}^3)\times35 + (d_{453} + d_{5243}^5)\times20 \right] + \\
(1/2*f_3)\left[ (d_{133} + d_{143} + d_{153} + d_{343} + d_{353} + d_{453} + d_{123}^3 + d_{4123} + d_{124} + d_{324}^3 + d_{243}^5 + d_{253}) + f_3*T_3*(200/60) \right] + \\
0.4\left[ (d_{124} + d_{124}^2 + d_{154} + d_{144}^1)\times20 + d_{134}\times30 + (d_{234} + d_{244}^3 + d_{254}^5)\times10 \right] + \\
(1/2*f_4)\left[ (d_{124} + d_{413} + d_{234} + d_{144}^2 + d_{154} + d_{144}^1 + d_{244}^3 + d_{244}^5 + d_{252}) + (f_4*T_4*(200/60) \right) + \\
0.4\left[ (d_{325} + d_{135}^2 + d_{245}^3 + d_{125})\times10 + d_{355}\times20 + (d_{255} + d_{155} + d_{245}^5)\times10 \right] + \\
(1/2*f_5)\left[ (d_{325} + d_{355} + d_{255} + d_{152} + d_{135} + d_{215} + d_{245} + d_{5245}) + f_5*T_5*(200/60) \right] + \\
0.4\left[ (d_{326} + d_{126} + d_{218})\times10 + (d_{346} + d_{146})\times30 + d_{356}\times20 + (d_{246} + d_{126} + d_{246})\times20 + d_{256}\times40 + (d_{456} + d_{5146})\times20 \right] + \\
(1/2*f_6)\left[ (d_{326} + d_{346} + d_{356} + d_{246} + d_{256} + d_{456} + d_{3126} + d_{4126} + d_{2136} + d_{2146} + d_{3146} + d_{5146}) + f_6*T_6*(200/60) \right] + \\
0.4\left[ (d_{347} + d_{247} + d_{147})\times15 + (d_{357} + d_{5237} + d_{3257})\times35 + (d_{457} + d_{147} + d_{5247})\times20 \right] + \\
(1/2*f_7)\left[ (d_{347} + d_{357} + d_{457} + d_{147} + d_{5147} + d_{5237} + d_{3247} + d_{3247} + d_{3247} + d_{3247}) \right] + \\
132
\[
d^5_{247} + d^3_{257} + f_7 T_7*(200/60) + \\
0.4*((d_{348} + d^3_{148})*15 + d_{328}*35 + d_{558}*45 + (d_{428} + d^4_{128} + d^2_{148})*20 + d_{458}*30 + (d_{258} + d^2_{158})*10) + \\
(1/2*f_9)*(d_{348} + d_{328} + d_{358} + d_{428} + d_{458} + d_{258} + d^4_{128} + d^2_{148} + d^3_{148} + d^2_{158}) + f_8 T_8*(200/60) + \\
0.4*((d_{319} + d^3_{149})*10 + (d_{329} + d^3_{249})*30 + d_{559}*40 + (d_{129} + d^1_{249})*20 + (d_{159} + d^5_{149})*30 + (d_{259} + d^5_{249})*10) + \\
(1/2*f_9)*(d_{319} + d_{329} + d_{359} + d_{129} + d_{159} + d_{259} + \\
d^3_{149} + d^5_{149} + d^1_{249} + d^3_{249} + d^5_{249}) + f_9 T_9*(200/60)
\]

There are 107 variables in this formulation. The total number of demand variables is 89, 42 representing direct service, \(d_{ijk}\), and 47 represent service along a transfer path, \(d^n_{ijk}\). There are 9 maximum arc load variables, \(f_k\), and 9 binary variables, \(y_k\).

\[C.3\] The constraints:

1) Frequency share model - direct demand.

\[
d_{ijk} = (f_k y_k / \sum_{k' \in (i,j)} f_k y_k) \cdot d_{ij} \quad \text{[C.2]}
\]

\[
d_{121} = (f_1 y_1 / f_1 y_1 + f_4 y_4 + f_9 y_9) \cdot d_{12}
\]

\[
d_{124} = (f_4 y_4 / f_1 y_1 + f_4 y_4 + f_9 y_9) \cdot d_{12}
\]

\[
d_{129} = (f_9 y_9 / f_1 y_1 + f_4 y_4 + f_9 y_9) \cdot d_{12}
\]

\[
d_{151} = (f_1 y_1 / f_1 y_1 + f_9 y_9) \cdot d_{15}
\]

\[
d_{159} = (f_9 y_9 / f_1 y_1 + f_9 y_9) \cdot d_{15}
\]

\[
d_{251} = (f_1 y_1 / f_1 y_1 + f_2 y_2 + f_5 y_5 + f_8 y_8 + f_9 y_9) \cdot d_{25}
\]

\[
d_{252} = (f_2 y_2 / f_1 y_1 + f_2 y_2 + f_5 y_5 + f_8 y_8 + f_9 y_9) \cdot d_{25}
\]

133
$$d_{255} = (f_5*y_5/f_1*y_1+f_2*y_2+f_5*y_5+f_8*y_8+f_9*y_9)*d_{25}$$
$$d_{258} = (f_5*y_5/f_1*y_1+f_2*y_2+f_5*y_5+f_8*y_8+f_9*y_9)*d_{25}$$
$$d_{259} = (f_9*y_9/f_1*y_1+f_2*y_2+f_5*y_5+f_8*y_8+f_9*y_9)*d_{25}$$

$$d_{132} = (f_2*y_2/f_2*y_2+f_3*y_3+f_9*y_9)*d_{13}$$
$$d_{133} = (f_3*y_3/f_2*y_2+f_3*y_3+f_9*y_9)*d_{13}$$
$$d_{139} = (f_9*y_9/f_2*y_2+f_3*y_3+f_9*y_9)*d_{13}$$

$$d_{322} = (f_2*y_2/f_2*y_2+f_4*y_4+f_5*y_5+f_8*y_8)*d_{32}$$
$$d_{324} = (f_4*y_4/f_2*y_2+f_4*y_4+f_5*y_5+f_8*y_8)*d_{32}$$
$$d_{325} = (f_5*y_5/f_2*y_2+f_4*y_4+f_5*y_5+f_8*y_8)*d_{32}$$
$$d_{326} = (f_8*y_8/f_2*y_2+f_4*y_4+f_5*y_5+f_8*y_8)*d_{32}$$

$$d_{352} = (f_2*y_2/f_2*y_2+f_5*y_5)*d_{35}$$
$$d_{355} = (f_5*y_5/f_2*y_2+f_5*y_5)*d_{35}$$

$$d_{343} = (f_3*y_3/f_3*y_3+f_7*y_7+f_8*y_8)*d_{34}$$
$$d_{347} = (f_7*y_7/f_3*y_3+f_7*y_7+f_8*y_8)*d_{34}$$
$$d_{348} = (f_8*y_8/f_3*y_3+f_7*y_7+f_8*y_8)*d_{34}$$

$$d_{353} = (f_3*y_3/f_3*y_3+f_7*y_7)*d_{35}$$
$$d_{357} = (f_7*y_7/f_3*y_3+f_7*y_7)*d_{35}$$

$$d_{453} = (f_3*y_3/f_3*y_3+f_8*y_8+f_7*y_7)*d_{45}$$
$$d_{456} = (f_8*y_8/f_3*y_3+f_8*y_8+f_7*y_7)*d_{45}$$
$$d_{457} = (f_7*y_7/f_3*y_3+f_8*y_8+f_7*y_7)*d_{45}$$

$$d_{246} = (f_8*y_8/f_8*y_8+f_8)*d_{24}$$
$$d_{248} = (f_8*y_8/f_8*y_8+f_8)*d_{24}$$

Total of 29 constraints.
2) Frequency share model - transfer demand. The last index in each variable is the transfer number.

\[ d_{ijr}^N = \left[ f_{ij1}^* \prod_{k \in R} y_k \right] \sum_{k \in R} (f_{ij1}^* \prod_{k \in R} y_k) d_{ij} \]  

\[ d_{121}^3 = (f_3 y_3 y_4 \div f_3 y_3 y_4 + f_3 y_3 y_5 + f_3 y_3 y_6) d_{12} \]
\[ d_{122}^3 = (f_3 y_3 y_5 \div f_3 y_3 y_4 + f_3 y_3 y_5 + f_3 y_3 y_6) d_{12} \]
\[ d_{123}^3 = (f_3 y_3 y_6 \div f_3 y_3 y_4 + f_3 y_3 y_5 + f_3 y_3 y_6) d_{12} \]
\[ d_{124}^4 = (f_3 y_3 y_6 \div f_3 y_3 y_5 + f_3 y_3 y_6) d_{12} \]
\[ d_{125}^4 = (f_3 y_3 y_6 \div f_3 y_3 y_5 + f_3 y_3 y_6) d_{12} \]
\[ d_{136}^2 = (f_1 y_1 y_2 \div f_1 y_1 y_2 + f_1 y_1 y_5 + f_1 y_1 y_6) d_{13} \]
\[ d_{137}^2 = (f_1 y_1 y_6 \div f_1 y_1 y_2 + f_1 y_1 y_5 + f_1 y_1 y_6) d_{13} \]
\[ d_{138}^2 = (f_1 y_1 y_6 \div f_1 y_1 y_2 + f_1 y_1 y_5 + f_1 y_1 y_6) d_{13} \]
\[ d_{149}^2 = (f_1 y_1 y_6 \div f_1 y_1 y_5 + f_1 y_1 y_6 + f_1 y_4 y_5) d_{14} \]
\[ d_{1410}^2 = (f_1 y_1 y_6 \div f_1 y_1 y_5 + f_1 y_1 y_6 + f_1 y_4 y_5) d_{14} \]
\[ d_{1411}^2 = (f_1 y_4 y_6 \div f_1 y_1 y_5 + f_1 y_1 y_6 + f_1 y_4 y_5) d_{14} \]
\[ d_{1412}^2 = (f_1 y_4 y_6 \div f_1 y_1 y_5 + f_1 y_1 y_6 + f_1 y_4 y_5) d_{14} \]
\[ d_{1413}^3 = (f_2 y_2 y_7 \div f_2 y_2 y_7 + f_2 y_2 y_8 + f_2 y_2 y_6 + f_2 y_2 y_7 + f_9 y_8 y_9 y_9 y_8) d_{14} \]
\[ d_{1414}^3 = (f_2 y_2 y_8 \div f_2 y_2 y_7 + f_2 y_2 y_8 + f_2 y_2 y_6 + f_9 y_8 y_9 y_7 + f_9 y_8 y_9 y_8) d_{14} \]
\[ d_{1415}^3 = (f_2 y_2 y_9 \div f_2 y_2 y_7 + f_2 y_2 y_8 + f_2 y_2 y_6 + f_9 y_8 y_9 y_7 + f_9 y_8 y_9 y_8) d_{14} \]
\[ d_{1416}^3 = (f_9 y_8 y_7 \div f_2 y_2 y_7 + f_2 y_2 y_8 + f_2 y_2 y_6 + f_9 y_8 y_7 + f_9 y_8 y_9 y_8) d_{14} \]
\[ d_{1417}^3 = (f_9 y_8 y_8 \div f_2 y_2 y_7 + f_2 y_2 y_8 + f_2 y_2 y_6 + f_9 y_8 y_7 + f_9 y_8 y_9 y_8) d_{14} \]
\[ d_{1418}^3 = (f_9 y_8 y_9 \div f_2 y_2 y_7 + f_2 y_2 y_8 + f_2 y_2 y_6 + f_9 y_8 y_7 + f_9 y_8 y_9 y_8) d_{14} \]

135
\( f_0 \times y_0 \times y_0 + f_0 \times y_0 \times y_0 \times y_0 \times y_0 \times d_{14} \)

\[ d_{1419}^5 = (f_1 \times y_1 \times y_3 + f_1 \times y_1 \times y_3 + f_1 \times y_1 \times y_3 + f_1 \times y_1 \times y_3 + f_0 \times y_0 \times y_0 + f_0 \times y_0 \times y_0 + f_0 \times y_0 \times y_0 + f_2 \times y_2 \times y_7) \times d_{14} \]

\[ d_{1420}^5 = (f_1 \times y_1 \times y_3 + f_1 \times y_1 \times y_3 + f_1 \times y_1 \times y_3 + f_1 \times y_1 \times y_3 + f_0 \times y_0 \times y_0 + f_0 \times y_0 \times y_0 + f_0 \times y_0 \times y_0 + f_2 \times y_2 \times y_7) \times d_{14} \]

\[ d_{1421}^5 = (f_1 \times y_1 \times y_3 + f_1 \times y_1 \times y_3 + f_1 \times y_1 \times y_3 + f_1 \times y_1 \times y_3 + f_0 \times y_0 \times y_0 + f_0 \times y_0 \times y_0 + f_0 \times y_0 \times y_0 + f_2 \times y_2 \times y_7) \times d_{14} \]

\[ d_{1422}^5 = (f_0 \times y_0 \times y_3 + f_1 \times y_1 \times y_3 + f_1 \times y_1 \times y_3 + f_1 \times y_1 \times y_3 + f_0 \times y_0 \times y_0 + f_0 \times y_0 \times y_0 + f_0 \times y_0 \times y_0 + f_2 \times y_2 \times y_7) \times d_{14} \]

\[ d_{1423}^5 = (f_0 \times y_0 \times y_3 + f_1 \times y_1 \times y_3 + f_1 \times y_1 \times y_3 + f_1 \times y_1 \times y_3 + f_0 \times y_0 \times y_0 + f_0 \times y_0 \times y_0 + f_0 \times y_0 \times y_0 + f_2 \times y_2 \times y_7) \times d_{14} \]

\[ d_{1424}^5 = (f_0 \times y_0 \times y_3 + f_1 \times y_1 \times y_3 + f_1 \times y_1 \times y_3 + f_1 \times y_1 \times y_3 + f_0 \times y_0 \times y_0 + f_0 \times y_0 \times y_0 + f_0 \times y_0 \times y_0 + f_2 \times y_2 \times y_7) \times d_{14} \]

\[ d_{1425}^5 = (f_2 \times y_2 \times y_3 + f_1 \times y_1 \times y_3 + f_1 \times y_1 \times y_3 + f_1 \times y_1 \times y_3 + f_0 \times y_0 \times y_0 + f_0 \times y_0 \times y_0 + f_0 \times y_0 \times y_0 + f_2 \times y_2 \times y_7) \times d_{14} \]

\[ d_{1426}^5 = (f_2 \times y_2 \times y_3 + f_1 \times y_1 \times y_3 + f_1 \times y_1 \times y_3 + f_1 \times y_1 \times y_3 + f_0 \times y_0 \times y_0 + f_0 \times y_0 \times y_0 + f_0 \times y_0 \times y_0 + f_2 \times y_2 \times y_7) \times d_{14} \]

\[ d_{1427}^5 = (f_2 \times y_2 \times y_3 + f_1 \times y_1 \times y_3 + f_1 \times y_1 \times y_3 + f_1 \times y_1 \times y_3 + f_0 \times y_0 \times y_0 + f_0 \times y_0 \times y_0 + f_0 \times y_0 \times y_0 + f_2 \times y_2 \times y_7) \times d_{14} \]

\[ d_{1528}^2 = (f_4 \times y_4 \times y_5 + f_4 \times y_4 \times y_5 + f_4 \times y_4 \times y_5) \times d_{15} \]

\[ d_{1529}^2 = (f_4 \times y_4 \times y_5 + f_4 \times y_4 \times y_5 + f_4 \times y_4 \times y_5) \times d_{15} \]

\[ d_{2430}^1 = (f_1 \times y_1 \times y_3 + f_1 \times y_1 \times y_3 + f_4 \times y_4 \times y_3 + f_9 \times y_9 \times y_9) \times d_{24} \]

\[ d_{2431}^1 = (f_4 \times y_4 \times y_3 + f_1 \times y_1 \times y_3 + f_4 \times y_4 \times y_3 + f_9 \times y_9 \times y_9) \times d_{24} \]

\[ d_{2432}^1 = (f_0 \times y_0 \times y_3 + f_1 \times y_1 \times y_3 + f_4 \times y_4 \times y_3 + f_9 \times y_9 \times y_9) \times d_{24} \]

\[ d_{2433}^3 = (f_2 \times y_2 \times y_3 + f_2 \times y_2 \times y_3 + f_2 \times y_2 \times y_7 + f_4 \times y_4 \times y_7 + f_5 \times y_5 \times y_3 + f_5 \times y_5 \times y_7 + f_9 \times y_9 \times y_9 + f_9 \times y_9 \times y_9) \times d_{24} \]

\[ d_{2434}^3 = (f_2 \times y_2 \times y_3 + f_2 \times y_2 \times y_3 + f_2 \times y_2 \times y_3 + f_4 \times y_4 \times y_7 + f_5 \times y_5 \times y_3 + f_5 \times y_5 \times y_7 + f_9 \times y_9 \times y_3 + f_9 \times y_9 \times y_7) \times d_{24} \]

\[ d_{2435}^3 = (f_4 \times y_4 \times y_3 + f_2 \times y_2 \times y_3 + f_2 \times y_2 \times y_7 + f_4 \times y_4 \times y_7 + f_5 \times y_5 \times y_3 + f_5 \times y_5 \times y_7 + f_9 \times y_9 \times y_3 + f_9 \times y_9 \times y_7) \times d_{24} \]
\[ d_{2436} = (f_s y_5 y_7 + f_s y_5 y_3 + f_s y_6 y_7) * d_{24} \]
\[ d_{2437} = (f_s y_5 y_5 y_3 + f_s y_5 y_2 y_3 + f_s y_5 y_2 y_7 + f_s y_4 y_4 y_7 + f_s y_5 y_5 y_3 + f_s y_5 y_5 y_7 + f_s y_5 y_9 y_5 y_9 y_9 y_7) * d_{24} \]
\[ d_{2438} = (f_s y_5 y_5 y_3 + f_s y_5 y_2 y_3 + f_s y_5 y_2 y_7 + f_s y_4 y_4 y_7 + f_s y_5 y_5 y_3 + f_s y_5 y_5 y_7 + f_s y_5 y_9 y_5 y_9 y_7) * d_{24} \]
\[ d_{2439} = (f_s y_5 y_5 y_3 + f_s y_5 y_2 y_3 + f_s y_5 y_2 y_7 + f_s y_4 y_4 y_7 + f_s y_5 y_5 y_3 + f_s y_5 y_5 y_7 + f_s y_5 y_9 y_5 y_9 y_7) * d_{24} \]
\[ d_{2440} = (f_s y_5 y_5 y_3 + f_s y_5 y_2 y_3 + f_s y_5 y_2 y_7 + f_s y_4 y_4 y_7 + f_s y_5 y_5 y_3 + f_s y_5 y_5 y_7 + f_s y_5 y_9 y_5 y_9 y_7) * d_{24} \]
\[ d_{2441} = (f_s y_5 y_5 y_3 + f_s y_5 y_2 y_3 + f_s y_5 y_2 y_7 + f_s y_5 y_1 y_3 + f_s y_5 y_1 y_7 + f_s y_5 y_5 y_3 + f_s y_5 y_5 y_7 + f_s y_5 y_9 y_5 y_9 y_7) * d_{24} \]
\[ d_{2442} = (f_s y_5 y_5 y_3 + f_s y_5 y_2 y_3 + f_s y_5 y_2 y_7 + f_s y_5 y_1 y_3 + f_s y_5 y_1 y_7 + f_s y_5 y_5 y_3 + f_s y_5 y_5 y_7 + f_s y_5 y_9 y_5 y_9 y_7) * d_{24} \]
\[ d_{2443} = (f_s y_5 y_5 y_3 + f_s y_5 y_2 y_3 + f_s y_5 y_2 y_7 + f_s y_5 y_1 y_3 + f_s y_5 y_1 y_7 + f_s y_5 y_5 y_3 + f_s y_5 y_5 y_7 + f_s y_5 y_9 y_5 y_9 y_7) * d_{24} \]
\[ d_{2444} = (f_s y_5 y_5 y_3 + f_s y_5 y_2 y_3 + f_s y_5 y_2 y_7 + f_s y_5 y_1 y_3 + f_s y_5 y_1 y_7 + f_s y_5 y_5 y_3 + f_s y_5 y_5 y_7 + f_s y_5 y_9 y_5 y_9 y_7) * d_{24} \]
\[ d_{2445} = (f_s y_5 y_5 y_3 + f_s y_5 y_2 y_3 + f_s y_5 y_2 y_7 + f_s y_5 y_1 y_3 + f_s y_5 y_1 y_7 + f_s y_5 y_5 y_3 + f_s y_5 y_5 y_7 + f_s y_5 y_9 y_5 y_9 y_7) * d_{24} \]
\[ d_{2446} = (f_s y_5 y_5 y_3 + f_s y_5 y_2 y_3 + f_s y_5 y_2 y_7 + f_s y_5 y_1 y_3 + f_s y_5 y_1 y_7 + f_s y_5 y_5 y_3 + f_s y_5 y_5 y_7 + f_s y_5 y_9 y_5 y_9 y_7) * d_{24} \]
\[ d_{2447} = (f_s y_5 y_5 y_3 + f_s y_5 y_2 y_3 + f_s y_5 y_2 y_7 + f_s y_5 y_1 y_3 + f_s y_5 y_1 y_7 + f_s y_5 y_5 y_3 + f_s y_5 y_5 y_7 + f_s y_5 y_9 y_5 y_9 y_7) * d_{24} \]
\[ d_{2448} = (f_s y_5 y_5 y_3 + f_s y_5 y_2 y_3 + f_s y_5 y_2 y_7 + f_s y_5 y_1 y_3 + f_s y_5 y_1 y_7 + f_s y_5 y_5 y_3 + f_s y_5 y_5 y_7 + f_s y_5 y_9 y_5 y_9 y_7) * d_{24} \]
\[ d_{2549} = (f_s y_5 y_5 y_3 + f_s y_5 y_2 y_3 + f_s y_5 y_2 y_7 + f_s y_5 y_1 y_3 + f_s y_5 y_1 y_7 + f_s y_5 y_5 y_3 + f_s y_5 y_5 y_7 + f_s y_5 y_9 y_5 y_9 y_7) * d_{16} \]
\[ d_{2550} = (f_s y_5 y_5 y_3 + f_s y_5 y_2 y_3 + f_s y_5 y_2 y_7 + f_s y_5 y_1 y_3 + f_s y_5 y_1 y_7 + f_s y_5 y_5 y_3 + f_s y_5 y_5 y_7 + f_s y_5 y_9 y_5 y_9 y_7) * d_{15} \]

A total of 50 constraints.
3) The demand constraints:

\[ \sum_{k \in K} (d_{ijk} + d_{ijk}^n)y_k = d_{ij} \quad \forall (i,j) \in N : i \neq j \text{ and } \forall k \in K \quad [C.4] \]

The notation DMN0102 means the demand constraint for node pair (1,2).

DMN0102: \( d_{121}y_1 + d_{122}y_2 + d_{133}y_3 + d_{124}y_4 + d_{125}y_5 + d_{126}y_6 + d_{128}y_8 + d_{128}y_8 + d_{129}y_9 = 30 \)

DMN0103: \( d_{213}y_1 + d_{132}y_2 + d_{133}y_3 + d_{124}y_4 + d_{213}y_5 + d_{233}y_6 + d_{312}y_8 = 20 \)

DMN0104: \( d_{214}y_1 + d_{141}y_1 + d_{312}y_2 + d_{142}y_2 + d_{143}y_3 + d_{144}y_4 + d_{145}y_5 + d_{146}y_6 + d_{147}y_7 + d_{214}y_8 + d_{214}y_8 + d_{214}y_9 + 30 \)

DMN0105: \( d_{151}y_1 + d_{152}y_2 + d_{153}y_3 + d_{215}y_4 + d_{215}y_5 + d_{215}y_6 + d_{159}y_9 = 10 \)

DMN0203: \( d_{523}y_1 + d_{322}y_2 + d_{326}y_6 + d_{527}y_7 + d_{328}y_8 + d_{329}y_9 = 30 \)

DMN0204: \( d_{1241}y_1 + d_{5242}y_1 + d_{324}y_2 + 2d_{5242}y_2 + d_{5}y_3 + d_{243}y_3 + d_{244}y_4 + d_{244}y_4 + d_{245}y_5 + d_{245}y_5 + d_{246}y_6 + d_{247}y_7 + d_{5247}y_7 + d_{248}y_8 + d_{3249}y_9 + d_{3249}y_9 + d_{249}y_9 = 25 \)

DMN0205: \( d_{251}y_1 + d_{252}y_2 + d_{253}y_3 + d_{3254}y_4 + d_{255}y_5 + d_{256}y_6 + d_{3257}y_7 + d_{258}y_8 + d_{259}y_9 = 20 \)

DMN0304: \( d_{223}y_2 + d_{346}y_6 + d_{347}y_7 + d_{348}y_8 = 15 \)

DMN0305: \( d_{352}y_2 + d_{353}y_3 + d_{355}y_5 + d_{356}y_6 + d_{357}y_7 + d_{358}y_8 + d_{359}y_9 = 25 \)

DMN0405: \( d_{453}y_3 + d_{456}y_6 + d_{457}y_7 + d_{458}y_8 = 15 \)

10 constraints as the number of O-D demand pairs on the network, 5
choose \( 2 = (1/2) \times 5 \times 4 = 10 \).

4) the same set of constraints without the \( y \)'s to ensure that the \( y_k \)'s are exactly 0 or 1.

\[
\sum_{k \in K} (d_{i,j,k} + d^n_{i,j,k}) = d_{i,j} \quad \forall \ (i,j) \in N : i \neq j \text{ and } \forall \ k \in K \quad [C.5]
\]

The notation DMNY0102 means the demand constraint for node pair (1,2), without the \( y_k \)'s.

DMNY0102: \( d_{121} + d_{122} + d^3_{133} + d^4_{123} + d_{124} + d^3_{125} + d^3_{126} + d^4_{126} + d^4_{128} + d_{129} = 30 \)

DMNY0103: \( d^2_{131} + d_{132} + d_{133} + d_{124} + d^2_{135} + d^2_{136} + d_{319} = 20 \)

DMNY0104: \( d^5_{141} + d^5_{141} + d^3_{142} + d^5_{142} + d_{143} + d^2_{146} + d^3_{146} + d^5_{146} + d^5_{147} + d^3_{147} + d^2_{148} + d^3_{148} + d^5_{149} + d^3_{149} = 30 \)

DMNY0105: \( d_{151} + d_{152} + d_{153} + d^2_{215} + 4d^2_{155} + d^3_{158} + d_{159} = 10 \)

DMNY0203: \( d^5_{231} + d_{232} + d_{236} + d^5_{237} + d_{328} + d_{329} = 30 \)

DMNY0204: \( d^1_{241} + d^5_{242} + d_{244} + 2d^5_{244} + d^4_{242} + d^7_{243} + d^5_{243} + d^1_{244} + d^2_{244} + d^2_{245} + d^5_{245} + d_{246} + d^3_{247} + d^5_{247} + d_{428} + d^1_{249} + d^3_{249} + d^5_{249} = 25 \)

DMNY0205: \( d_{251} + d_{252} + d^3_{253} + d^3_{254} + d_{255} + d_{256} + d^3_{257} + d_{258} + d_{259} = 20 \)

DMNY0304: \( d_{232} + d_{346} + d_{347} + d_{348} = 15 \)

DMNY0305: \( d_{352} + d_{353} + d_{355} + d_{356} + d_{357} + d_{358} + d_{359} = 25 \)

DMND0405: \( d_{453} + d_{456} + d_{457} + d_{458} = 15 \)

A total of 10 demand constraints without the \( y_k \)'s.

5) The maximum arc load constraint. There is one maximum arc load constraint for each arc on each route to make sure that the maximum arc load is not smaller than the actual load on any of the arcs:
\[ f_k - (1/C) \sum_{(h,l) \ni (i-j)} (d_{hlk} + d_{hl}^*) k \text{ and } k \in K \]  

where the expression \((h,l) \ni (i-j)\) means "...over all node pairs \((h,l)\) on route \(k\) that include arc \((i-j)\) in the path between them".

The notation QX010201 means the maximum arc load constraint for arc 1-2 on route 1. These constraints are expressed in terms of the frequency.

QX010201: \( f_1 - (1/50) \sum (d_{121} + d_{151} + d^2_{141} + d^2_{141} + d^3_{141} + d_{1241}) \geq 0 \)
QX020501: \( f_1 - (1/50) \sum (d_{151} + d_{251} + d^2_{141} + d^5_{231} + d_{1241}) \geq 0 \)
QX010302: \( f_2 - (1/50) \sum (d_{132} + d_{122} + d^5_{141} + d^5_{231} + d^5_{241}) \geq 0 \)
QX030202: \( f_2 - (1/50) \sum (d_{322} + d_{352} + d^2_{142} + d^2_{3242}) \geq 0 \)
QX020502: \( f_2 - (1/50) \sum (d_{252} + d^5_{142} + d^5_{142} + d_{152} + d_{352}) \geq 0 \)
QX010303: \( f_3 - (1/50) \sum (d_{133} + d_{143} + d_{153} + d^3_{123} + d^4_{123} + d^3_{1243}) \geq 0 \)
QX030403: \( f_3 - (1/50) \sum (d_{343} + d_{353} + d^4_{123} + d^1_{243} + d^3_{243} + d^3_{253} + d_{143} + d_{153}) \geq 0 \)
QX040503: \( f_3 - (1/50) \sum (d_{433} + d^3_{243} + d^3_{253} + d_{153} + d_{353}) \geq 0 \)
QX010204: \( f_4 - (1/50) \sum (d_{124} + d_{134} + d^2_{144} + d^2_{154} + d^1_{244}) \geq 0 \)
QX020304: \( f_4 - (1/50) \sum (d_{134} + d_{234} + d^3_{244} + d^3_{254}) \geq 0 \)
QX030205: \( f_5 - (1/50) \sum (d_{325} + d_{355} + d^2_{125} + d^2_{135} + d^2_{245}) \geq 0 \)
QX020505: \( f_5 - (1/50) \sum (d_{255} + d_{355} + d^2_{155} + d^5_{245}) \geq 0 \)
QX030206: \( f_6 - (1/50) \sum (d_{322} + d_{342} + d_{356} + d^3_{126} + d^2_{136} + d^3_{146}) \geq 0 \)
QX020406: \( f_6 - (1/50) \sum (d_{346} + d_{256} + d_{356} + d_{346} + d^4_{126} + d^2_{146} + d^3_{146}) \geq 0 \)
QX040506: \( f_6 - (1/50) \sum (d_{335} + d_{325} + d_{456} + d^4_{146}) \geq 0 \)
QX030407: \( f_7 - (1/50) \sum (d_{347} + d_{357} + d^3_{147} + d^5_{237} + d^2_{247} + d^3_{257}) \geq 0 \)
QX040507: \( f_7 - (1/50) \sum (d_{357} + d_{457} + d^5_{147} + d^5_{237} + d^3_{247} + d^3_{257}) \geq 0 \)
QX030408: \( f_8 - (1/50) \sum (d_{348} + d_{328} + d_{358} + d^3_{148}) \geq 0 \)
QX040208: \( f_8 - (1/50) \sum (d_{428} + d_{358} + d_{458} + d_{328} + d^4_{128} + d^2_{148} + d^3_{148}) \geq 0 \)
QX020508: \( f_8 - (1/50) \sum (d_{358} + d_{458} + d_{258} + d^2_{158}) \geq 0 \)
QX030109: \( f_9 - (1/50) \sum (d_{319} + d_{329} + d_{359} + d^3_{149} + d^3_{249}) \geq 0 \)
QX010209: \( f_9 - (1/50) \sum (d_{129} + d_{159} + d_{329} + d_{359} + d^5_{149} + d^1_{249} + d^3_{249}) \geq 0 \)
QX020509: \( f_0 - (1/50)*(d_{ss0} + d_{1s0} + d_{2s0} + d_{140} + d_{240}) \geq 0 \)

There are 23 maximum arc load constraints, one for each arc on each route on the network.

6) The minimum frequency constraints:

\[ f_k - F^m_k y_k \geq 0 \quad \forall k \in K \]  \[ \text{(C.7)} \]

where \( C \) is 50 passengers/vehicle and the policy headway is chosen here to be 30 minutes so \( F^m_k = 2 \) vehicles/hour.

MINFRQ01: \( f_1 - 2y_1 \geq 0 \)
MINFRQ02: \( f_2 - 2y_2 \geq 0 \)
MINFRQ03: \( f_3 - 2y_3 \geq 0 \)
MINFRQ04: \( f_4 - 2y_4 \geq 0 \)
MINFRQ05: \( f_5 - 2y_5 \geq 0 \)
MINFRQ06: \( f_6 - 2y_6 \geq 0 \)
MINFRQ07: \( f_7 - 2y_7 \geq 0 \)
MINFRQ08: \( f_8 - 2y_8 \geq 0 \)
MINFRQ09: \( f_9 - 2y_9 \geq 0 \)

Total of 9 constraints as the number of routes.

7) Transfer sides equality constraints. This constraints are to make sure that when the demand between a pair of nodes, \((i,j)\), is satisfied by a transfer path, the number of passengers leaving node \( i \) is equal to the number of passengers arriving at node \( j \).

The notation TY010203 means the demand between nodes 1 and 2 satisfied by a transfer path with transfer at node 3.

TY010203: \( d_{123}^3 y_3 - (d_{125}^3 y_5 + d_{126}^3 y_6) = 0 \)
TY010204: $d_{123}^8 y_3 - (d_{126}^8 y_6 + d_{128}^8 y_8) = 0$
TY010302: $d_{131}^8 y_1 - (d_{135}^8 y_5 + d_{136}^8 y_6) = 0$
TY010402: $d_{141}^8 y_1 + d_{144}^8 y_4 - (d_{148}^8 y_8 + d_{148}^8 y_8) = 0$
TY010403: $d_{142}^3 y_7 - (d_{149}^3 y_9 + d_{147}^3 y_7 + d_{148}^3 y_8 + d_{148}^3 y_8) = 0$
TY010405: $d_{141}^8 y_1 + d_{149}^8 y_9 - (d_{148}^8 y_8 + d_{147}^8 y_7) = 0$
TY010502: $d_{154}^8 y_4 - (d_{155}^8 y_5 + d_{158}^8 y_8) = 0$
TY020305: $d_{123}^5 y_1 - d_{237}^5 y_7 = 0$
TY020401: $d_{241}^1 y_7 + d_{244}^1 y_4 + d_{249}^1 y_9 - d_{243}^1 y_3 = 0$
TY020403: $d_{242}^1 y_2 + d_{244}^3 y_4 + d_{248}^3 y_8 + d_{249}^3 y_9 - (d_{243}^3 y_3 + d_{247}^3 y_7) = 0$
TY020405: $d_{241}^5 y_1 + d_{242}^5 y_2 + d_{246}^5 y_8 - (d_{249}^5 y_9 + d_{243}^5 y_3 + d_{247}^5 y_7) = 0$
TY020503: $d_{254}^3 y_4 - (d_{253}^3 y_3 + d_{257}^3 y_7) = 0$

A total of 12 constraints for the transfer sides.

8) Fleet size constraint:

$$\frac{(2/60)}{C} \sum_{k \in K} T_k f_k - Q \leq 0 \quad [C.8]$$

where the expression $(2/C)T_k f_k$ is a proxy of the number of vehicles required for route $k$, and $Q$ is the total available fleet.

FLEETSIZ: $(2/50)*(f_1 t_1 + f_2 t_2 + f_3 t_3 + f_4 t_4 + f_5 t_5 + f_6 t_6 + f_7 t_7 + f_8 t_8 + f_9 t_9) -500 \leq 0$

C.4 Summary of the formulation

In the formulation of this small network problem with 5 nodes and 7 arcs there are 107 variables and 144 constraints. Each run of the program took one minute to achieve optimal solution on a D.E.C micro-vax computer.