A Methodology for Bus Network Design

by

Omar Baba

B.E., Civil Engineering
American University of Beirut, 1992

Submitted to the Department of Civil and Environmental Engineering
in Partial Fulfillment of the Requirements for the Degree of

Master of Science in Transportation

at the

Massachusetts Institute of Technology

September 1995

Signature of Author..
Department of Civil and Environmental Engineering
August 11, 1995

Certified by ....
Nigel H. M. Wilson
Professor of Civil and Environmental Engineering
Thesis Supervisor

Accepted by ....
Joseph Sussman
Chairman, Departmental Committee on Graduate Studies

A Methodology for Bus Network Design

by

Omar Baba

Submitted to the Department of Civil and Environmental Engineering
in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Transportation

Abstract

This thesis develops a methodology for solving the bus network design problem (BNDP) in which one seeks to determine the best configuration of bus routes and their frequencies, given information on the bus transit demand, the street network, the available resources and the operational constraints. The proposed methodology follows a heuristic approach and consists of two main components - route generation (RG) and vehicle allocation (VA) - that are implemented in a single automated design procedure.

The proposed methodology addresses the principal shortcomings of previous heuristic approaches to the BNDP and is characterized by its ability to (1) incorporate the fleet size constraint within the route generation process, (2) identify the major trip patterns in the demand matrix for guidance in the route configuration, and (3) solve the BNDP using a general network design or a transit center network design.

The RG procedure is based on the concept of route skeletons which consist of three nodes - two termini and one intermediate node - that are expanded into full routes by means of nodal insertions. RG creates one route at a time and estimates the number of buses required on the network after each route is generated. If the vehicle estimate is less than the available fleet size, an additional route is generated and the network is re-evaluated; otherwise, VA is initiated. VA calculates the minimum number of buses required on each route so as to satisfy the loading feasibility and passenger assignment constraints. Surplus buses remaining after the minimum vehicle requirement is determined may be allocated on the network to improve its overall performance, or used by RG to create additional routes.

The proposed methodology and automated design procedure are applied to a case study in San Juan, Puerto Rico where a major restructuring of the bus system using a transit center concept has been proposed. Solutions produced by the proposed methodology are compared with the proposed transit center route plan.

Thesis Supervisor: Dr. Nigel H. M. Wilson
Title: Professor of Civil and Environmental Engineering

Acknowledgments

I would like to express my profound appreciation to my advisor, Dr. Nigel Wilson, for his guidance and support throughout my studies at MIT. His encouragement, patience and friendship has made working with him a privilege.

I also would like to express my deep gratitude to Dr. Isam Kaysi at the American University of Beirut for introducing me to transportation planning and suggesting the topic of bus network design.

I am also grateful to Jim Wensley, Larry Englisher and John Attanucci at Multisystems, Inc. for contributing much of the information for the case study. Various students and faculty members of the University of Puerto Rico have also provided me with significant insight and suggestions, and their help is greatly appreciated.

The encouragement of my family and their tremendous support throughout my stay in Boston and at MIT has been a blessing. The company of Jeff Sriver, Bill Cowart, Nicola Shaw, Daniel Rodriguez , John Decker and Alan Hoffman as we traveled the road together has been a pleasure and a great support as well.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Buses are the dominant form of public transport in many metropolitan areas and large cities due to their inherent advantages of low fixed costs and operating flexibility. In 1993, buses carried almost 65% of the 8.3 billion annual passenger trips in the U.S. and accounted for more than 50% of the 39.6 billion annual passenger miles (Transit Fact Book, 1994-1995). In most large cities, however, transit systems (including bus systems) lose some ridership each year owing to increasing car ownership and lower density suburban development. In the last few decades, this has resulted in an automobile-based lifestyle in most metropolitan areas with dramatic redistributions of population, employment, retail centers and other trip generators. During the same period, total urban person trips have increased significantly and have become more dispersed due to the redistribution of land use, whereas transit trip making has grown at a much slower rate or even experienced a decline. One reason for this is that the bus route networks have generally not been improved through restructuring and, thus, could not adapt to the changing demand for transit trips. As a result, buses have become less productive and less cost-effective.

Although bus networks have changed in this period in some of the rapidly growing cities, in many cases the basic network and service structures have remained essentially the same. In fact, bus authorities have failed to take advantage of the flexibility of bus transit in dealing with changing demand patterns, partly because of the absence of systematic procedures to develop and evaluate alternative route and service structures. As a result, the overall redesign of a bus network involves a higher risk of a poorer performance. Bus authorities - recognizing that increasing fares to counterbalance the decrease in cost-effectiveness caused by a decline in passengers would result in accelerating the deterioration in the bus transit market - have resorted to other means of countering the losses. Their main planning focus is often placed on ways for cutting costs, particularly by tackling the most poorly patronized routes, rather than considering major network restructuring. They have also focused on operational planning activities including bus and driver scheduling, since the largest single cost component of providing bus service is the driver's wage and fringe benefits.

This thesis addresses a more strategic aspect of bus service planning in which one seeks to determine a configuration of bus routes and their frequencies that achieves some desired objective, subject to operational and budget constraints. This strategic problem, which will be referred to hereafter as the bus network design problem (BNDP) is at the heart of bus service planning and is a key to the success or failure of an overall service restructuring. In the next section, the BNDP is placed in the context of the overall bus service planning process.

1.1 The Bus Service Planning Process

The overall bus service planning process shown in table 1.1 is a systematic decision sequence consisting of the following elements: route and network design, frequency determination, timetable development, bus scheduling and driver scheduling. The BNDP, as defined in the previous section and for the rest of this thesis, is concerned only with the first two components. Although the activities are shown sequentially, there are strong interrelations between the elements of the problem, and characteristics of the elements down the sequence may well have some effect on higher level decisions. If one considers the overall planning process as an optimization problem, a single global mathematical program formulation could be hypothesized to encompass all the elements of the process. However, because of various complexities, it is infeasible to include all these elements in a single model. Instead, the overall problem is divided into computationally tractable subproblems, often corresponding to the internal organization of the public transit authorities. As such, bus service planning problems are often classified at the strategic, tactical and operational levels. Strategic planning is concerned with the long- term development of the bus route network, a task which requires comprehensive studies and that must stay in place for some time to achieve its full potential impact. Tactical planning

Independent Inputs	**Planning Activity**	**Output**
Demand data Supply data Route performance indices	Route and Network Design	Route changes New routes Operating strategies
Subsidy available Buses available Service policies Current patronage	Frequency Determination	Service frequencies
Demand by time of day Times to first and last trips Running times	Timetable Development	Trip departure times Trip arrival times
Deadhead times Recovery times Schedule constraints Cost structure	Bus Scheduling	Bus schedules
Driver work rules Run cost structure	Driver Scheduling	Driver schedules

Table 1.1: The Bus Service Planning Process

deals with the service frequency determination and is considered to be a medium-term development, whereas operational planning encompasses the remaining schedule-related activities and is conceived of as a frequently performed short-term planning effort.

Among all the activities mentioned above, the ones at the operational planning level have received the most attention by both bus authorities and researchers. From the perspective of bus properties, the scheduling components involve potential tangible cost savings that help improve the financial position of the system. On the other hand, the relatively high frequency of operational planning activities within the planning cycle has motivated researchers to focus on those activities and to develop numerous computer programs to automate them, at least partially.

The BNDP has received less attention than the operational planning subproblems because of its complexity. Approached from a constrained optimization perspective, the BNDP is difficult to formulate as a well-behaved model. A bus network is comprised of links that form a subset of the full street network. In an optimization framework, the provision of bus service on a particular street can be viewed as a 0/1 discrete choice of adding a link to the bus network to enable the flow of bus passengers. On the other hand, frequencies are continuous decision variables that have to be considered in conjunction with the route link discrete variables. Thus, the modeling of the BNDP as an optimization problem would involve mixed integer programming which is a highly complex class of optimization problems. (Difficulties related to the modeling of the BNDP as an optimization problem are discussed further in the next chapter.)

Non-optimization approaches to the BNDP attempt to address the two main facets of the problem by dividing the BNDP into two components: the generation of the network of routes and the determination of service frequencies. However, even in this approach, the BNDP is still a difficult problem for two reasons. First, the route design component is closely related to the frequency determination component, since the vehicle requirement on each route - on which both the cost of the design and its effectiveness critically depend - cannot be known unless frequencies are determined. At the same time, the BNDP is almost always constrained by the size of the available fleet of buses. Consequently, isolating the process of route and network design from frequency determination prevents a reliable assessment of the required number of buses to operate on a particular network of routes, possibly resulting in a design that is not within the operator's resource constraints. Second, bus transit systems are characterized by the fact that it is the users - not the planner - who determine their own routing. Consequently, the frequency determination model within the BNDP must be capable of predicting the choice behavior of passengers with respect to transit paths. In itself, the transit path choice problem is inherently stochastic and time-dependent in nature, and its inclusion in the BNDP adds yet another burden.

1.2 Motivation

It should be obvious that the overall restructuring of a bus network can have significant impacts on the system's performance. In the short term, potential benefits from network restructuring include the improvement in the total travel time and in the level of demand satisfaction. In the longer run, ridership is likely to increase as a result of the improvement in the level of service. Moreover, benefits from optimally, or near-optimally, solving the BNDP are likely to outweigh those resulting from other improvements within the rest of the bus planning activities such as scheduling improvements at the operational level. If the overall bus planning process is viewed as a single optimization problem, the establishment of routes and frequencies would greatly reduce the solution space of the problem. Even when the overall problem is separated into its various components, the marginal benefits obtained from operational improvements would be small if the route network and service frequencies are treated as fixed and given. Consequently, a good solution for the BNDP is necessary for a successful overall restructuring of a bus system.

Prior optimization approaches to the BNDP can be grouped into those predicated on idealization of the network and those dealing with actual routes. The majority of approaches in both categories has not received extensive application. Methods based on idealizing the network are best suited for screening or policy analysis, rather than final design, and thus have limited practical applicability. The second group of methods is inevitably heuristic and is generally complex, non-user oriented and expensive, both in terms of the data needed and the direct cost and staff time required. Moreover, many of the prior heuristic approaches suffer from the following principal shortcomings:

1. The operator's fleet size constraint is not explicitly addressed at the route design level. Certain approaches (Silman et al., 1974; Israeli and Ceder, 1989) start by creating a large pool of "possible" routes. Then, smaller route subsets are selected and the demand is assigned to each one, calculating the corresponding vehicle requirement and specific performance measures. The subset which satisfies the resource constraint and produces

the best overall performance is then selected as the final solution. In another approach (Baaj and Mahmassani, 1993), the vehicle requirement is not calculated until a pre-specified number N of routes is generated. Routes are then either added to or dropped from the set of N routes in order to match the vehicle requirement with the available bus fleet size.

2. Most heuristic approaches do not rely properly on the demand matrix for guidance in the route layout. Approaches which determine all possible routes use constraints on round tip time and circuity ratio to generate the routes. Route subsets are then formed using several strategies which only partially consider the amount of demand satisfied by each subset. Other approaches generally do not take full advantage of the trip patterns contained in the demand matrix.

3. Only a few approaches are capable of incorporating a pre-defined network concept (radial, grid, etc.) in the route generation process. In many large cities, the network restructuring process may be more beneficial if a concept that is different from the one implemented in the existing system is adopted, especially if the new concept is better at serving new trip patterns that have developed since the existing system was conceived.

The approach to the BNDP adopted in this thesis attempts to address the deficiencies described above. This will allow for increasing the chances of acceptance of this approach within transit properties. In particular, the proposed approach will attempt to incorporate the operator's fleet size constraint within the route generation process in order to create a more realistic bus planning tool that addresses one of the operator's main concerns. In addition, the approach will attempt to permit solving the BNDP on the basis of a network concept different from that of the current system. This would also appeal to planning authorities who are usually more reluctant in undertaking service restructuring whenever a new service concept needs to be investigated, because of the additional effort and cost involved.

1.3 Research Objectives

The discussion in the previous sections has demonstrated the importance of the BNDP within the process of bus service planning. The fundamental question addressed in this thesis is the following:

> Given information on the bus transit demand, the street network, the available resources and the operational constraints, how to design the best possible network of bus routes and frequencies?

To this end, the thesis has two major objectives. The first objective is to develop a methodology for solving the BNDP and to implement it as an automated design procedure that can be used as a strategic planning tool in the restructuring of bus systems. Because the optimal solution to the BNDP is too complex to be feasible, a heuristic approach which produces a "good" but not necessarily optimal solution will be adopted. Besides designing networks with the guidance of the existing demand matrix, the methodology will also be capable of generating solutions based on a concept that is different from the one used in the existing system. The development of the methodology will be preceded by a review of previous literature on the BNDP in which the existing approaches for solving this problem are analyzed. The automated design procedure based on the proposed methodology is intended for use as a planning tool which does not necessarily deliver a readily-implementable solution. The focus is on producing a basic description of the network, along with a set of global performance measures that facilitates comparison of the solution to the existing system. The basic network can then be refined in terms of the route layout and service frequencies in order to reach an implementable solution.

The second objective of this thesis is to apply the methodology to a case study in which an existing bus network is to be restructured. The bus system of the metropolitan city of San Juan, Puerto Rico is an appropriate case study for this research because a plan for a major restructuring of the existing system, involving a basic change in the network concept, has been approved and is currently under development. The restructuring is part of a comprehensive study to improve the overall public transportation in the San Juan Metropolitan Area, which has sustained a considerable decline in transit ridership over the last 20 to 25 years. The current bus system is characterized by adequate coverage and connectivity and a relatively low fare, but suffers from low frequencies and a lack of speed, directness and schedule adherence on a large number of routes. Plans for restructuring the network are based on a transit center concept which requires establishing several transit centers located at the major destinations in the region and connecting them with high-frequency trunk routes and medium-to-high frequency local routes. The San Juan case study provides an opportunity to experiment with the design procedure on a real problem, and to compare its performance to the proposed design

1.4 Thesis Organization

Chapter 2 starts by reviewing the literature on the BNDP, describing the various solution approaches and highlighting their strengths, weaknesses and practicality. Next, a more detailed review of the heuristic approach for solving the BNDP is presented in which the two main components of this approach - route generation and frequency determination - are described. In chapter 3, a methodology for solving the BNDP is presented. The main features and assumptions of the methodology are first discussed, and then the two main components are briefly described and placed in the context of the design procedure. Chapter 4 focuses on route generation and starts by discussing the various approaches that were considered for that process. The major steps of route generation in the case of a general network concept and of a transit center concept are then described in detail. Chapter 5 is concerned with vehicle allocation and starts with a review of the theory and practice of vehicle allocation methods. After the proposed vehicle allocation process is described, a numerical example illustrating the performance of route generation and vehicle allocation is presented. In chapter 6, the San Juan case study is used to investigate the performance of the methodology and design procedure and to draw the final conclusions, which are reported in chapter 7.

Chapter 2

Literature Review

This chapter reviews the literature on the various approaches used for solving the BNDP in order to place the research in its appropriate context. The review starts by classifying the different approaches used, briefly describing their main characteristics, strengths, weaknesses and uses. After that, the focus is placed on the mathematical heuristic approach which provides a suitable framework for the automated bus network design procedure sought in this research. This approach is particularly useful in the short-range planning of bus service and can be used to develop an efficient screening tool for investigating the restructuring of existing networks.

2.1 Approaches for Solving the BNDP

The approaches for solving the BNDP used by transit planners or proposed by researchers are reviewed in this section. Because the planning of bus routes and frequencies is understood to include both their development and evaluation, the various approaches for solving the BNDP are classified according to two attributes: (1) the method of generating a solution and (2) the method of evaluation. Based on these two attributes, the approaches for solving the BNDP may be classified into five categories: (1) manual; (2) market analysis; (3) systems analysis; (4) systems analysis with interactive graphics; and (5) mathematical. Table 2.1 shows the classification of these approaches, and the following paragraphs briefly describe the characteristics of each.

2.1.1 Manual

The manual approach relies on qualitative, non-quantifiable guidelines and ad hoc procedures which reflect the professional judgment and practical experience of service

Approach	**Method of Network Generation**	**Method of Network Evaluation**
Manual	manual	manual
Market Analysis Project	manual	partially computerized
Systems Analysis	manual	computerized
Systems Analysis with Interactive Graphics	manual	computerized, with interactive graphics
Mathematical	computerized	computerized

Table 2.1: Classification of Approaches for Solving the BNDP

planners in the transit industry. In this approach, bus networks are manually restructured by planners who often have considerable knowledge of the service area which allows them to propose generally minor improvements to the existing system. The modified bus network is also evaluated manually, rather than by commercially available computer-based analysis tools, and ridership is estimated qualitatively based on comparable routes, without performing passenger assignment. Because of the limited amount of analysis performed, the data required by this approach is usually simple and relatively easy to obtain.

An example of the guidelines used in the manual approach is the NCHRP Report 69 published by the Transportation Research Board (1980) which suggests service planning guidelines in the form of rules of thumb that cover issues such as service area and route coverage, route structure and spacing, route directness, route length, and route duplication. Guidelines with regard to service levels include desirable minimum service frequencies and loading standards. These guidelines are based on interviews with transit planners over a broad spectrum of U.S. and Canadian cities and emphasize practice rather than theory.

In general, the manual approach is most useful for short-range planning, minor changes in the network (such as the extension of a bus route to a new suburb) and fine-tuning of bus networks. However, the manual approach is not suitable for comprehensive network revisions or medium or long-range planning, because the evaluation of the modified network is unlikely to be either consistent or comprehensive. Another weakness of the manual approach is that many value judgments are involved in the restructuring of the network.

2.1.2 Market Analysis Project (MAP)

The MAP approach as described by Bursey et al. (1979) consists of three main components: (1) the description of the existing system; (2) the analysis of the characteristics of the existing system; and (3) the design cycle. The description of the existing system is used, along with other studies, to develop an understanding of its demand and cost characteristics. An alternative network with its characterizing frequencies and fare structure is developed at the beginning of the design cycle, and the demand and cost characteristics of the existing network are then used manually to predict the operator's costs and passenger flows on the proposed network and to appraise its performance. The evaluation of the network is used as the basis for changing its structure or characteristics in the case where the network does not perform as desired, and the design cycle is repeated until a network with satisfactory performance is obtained.

The MAP approach is only partially computerized, with the use of the computer restricted to the examination of passenger origin-destination and household survey data. However, the evaluation of the modified network in terms of estimating ridership and calculating performance measures is performed in a simple way without requiring network analysis tools. As such, this approach is most suitable for small-scale networks in rural and/ or inter-urban applications.

2.1.3 System Analysis (SA)

The major characteristic that distinguishes the SA approach from the other approaches is the use of computer-based models for the evaluation of bus networks which are capable of estimating the change in ridership on the restructured network and assessing its performance. As with the previous two approaches, bus routes and frequencies are established in the SA approach based on the planner's expertise and local knowledge of the area, the operator's capabilities, and the network concepts that might be successfully implemented. The network is then analyzed using computer-based evaluation models, producing a set of performance measures at the route level that form the basis for modifying the proposed network and going through another design cycle, if necessary. The evaluation models which have been used in the SA approach include TRANSEPT (Last and Leak, 1976) which uses the multi-path assignment model of Dial (1971) and TRANSCOM (Chapleau and Chriqui, 1975). More recent computer-based evaluation models use interactive graphics and are discussed in the next section.

The SA approach is far superior to the manual approach because it is more systematic and comprehensive. One of its main advantages is that it allows for variable demand. Without a computer-based evaluation model, the prediction of link volumes is practically infeasible, even for small networks, because of the possibility of having multiple-paths serving certain origin-destination flows. In the case of large and complex networks, the SA approach becomes necessary because of the increased occurrence of route overlap which requires a formal path choice model. On the other hand, the SA approach typically allows only a limited number of alternative networks to be considered because of the effort required to design a network at the level of detail required by the evaluation process, as well as the large amounts of data required in that process.

2.1.4 System Analysis with Interactive Graphics (SAIG)

This approach is essentially the same as the previous one in terms of the generation and evaluation of alternative networks. However, in this approach, a slightly different type of network evaluation is permitted through the use of interactive graphics which provide a computer-assisted, graphically-based operating environment for the planner to interface and interact with the formal procedures for network evaluation. These graphical capabilities support the evaluation process by allowing the planner to see results such as passenger flows on the spatial layout of the network, taking the network evaluation process to a higher level than allowed by the other approaches.

Several computer models with interactive graphics capabilities have been developed for transit network evaluation. Among these are VIPS (Andreasson, 1976) and its successor VIPS-II (Jansson, 1987) developed by the Volvo Corporation, TRANSPLAN (Osleeb et al., 1976), MADITUC (Chapleau et al., 1982) and EMME/2 (Florian et al., 1986). Of these, VIPS-II and EMME/2 provide the most extensive graphical displays, as well as sophisticated evaluation models.

2.1.5 Mathematical

The mathematical approach attempts to employ mathematical techniques to produce optimal bus networks. Mathematical formulations of the BNDP have been concerned primarily with the minimization of an overall cost measure, generally a combination of user costs and operator costs, subject to feasibility constraints. Two types of mathematical approaches can be identified: the analytical type and the heuristic type.

2.1.5.1 Analytical

In this approach, bus networks and the urban areas they serve are idealized by using *simple* bus network geometries and travel demand functions. Extensive work of this type has been based on constrained optimization methods in which design parameters such as route spacing, route length, stop spacing or headways are selected so as to minimize an objective reflecting benefits to the passenger and cost to the operator. Byrne (1976), Hurdle (1973) and Newell (1979) have developed analytical optimization models based on an assumption of fixed demand, limited design parameters and the objective of minimizing the sum of passenger and operator costs. Kocur and Hendrickson (1982) extended this approach to the case of variable demand, a broad range of design parameters and a choice of objectives reflecting user and/ or operator benefits. Nevertheless, all of these methods result in an over-simplification of the actual problem, precluding their use in actual network design. However, this approach may still be suited for screening or policy analysis in which approximate design parameters are to be determined, rather than a detailed final design.

2.1.5.2 Heuristic

The search for heuristic approaches to the BNDP is motivated by five main sources of complexity which preclude the use of formal optimization methods for solving the problem. The first difficulty relates to the mathematical formulation of the problem, namely the difficulty of defining the decision variables. While frequencies can be related to the passengers or operator's costs and expressed in the objective function, the number of routes and their configuration (which are also decision variables for the problem) can not. The second source of complexity results from the non-linearities and non-convexities exhibited by the BNDP formulations. Non-convexities are illustrated by the fact that more buses can be deployed in the network, thereby increasing the operator's costs, and still produce a higher total travel time (worse user costs). As pointed out by Newell (1979), concavity is induced by the waiting time which occurs at the access points to the system, not on the links of the network, and which cannot be easily incorporated in the model. Third, the complexity of the problem grows exponentially with the size of the transit network. The solution of the BNDP whose formulation includes discrete variables becomes extremely difficult with large networks. Fourth, the BNDP has a multi-objective nature. Most of the existing approaches consider reducing user costs and/or operator costs as their sole objective. In practice, however, important trade-offs among other conflicting objectives may need to be addressed in what is inherently a multi-objective problem. An example would be the trade-off between attempting to serve the largest demand possible by creating routes and deploying resources even in low-density areas, or concentrating resources on the more productive routes. A similar trade-off that may be considered is between directness and coverage. Such considerations, which are important in practice, have not typically been included in mathematical programming formulations. The last source of complexity relates to the spatial layout of routes. It is difficult to characterize and incorporate in a formal procedure what constitutes a "good" spatial layout of routes. This aspect has been to a certain extent addressed through design criteria and constraints such as route coverage, route duplication, route length, and directness of service (circuity).

All of the above complexities have shifted the focus on solving the BNDP to the use of heuristic approaches. The heuristic mathematical approach attempts to produce a "good" solution, but does not guarantee optimality. Most of these heuristic approaches decompose the BNDP into its two main components: route generation which creates a network of "good" routes, and frequency determination which allocates buses across the routes. Passenger assignment is also an important element of the heuristic approach and is addressed in the frequency determination subproblem. One advantage of this approach is that it is capable of generating bus network alternatives from scratch. Also, because the generation process may be automated, the overall design process may be performed more rapidly. The disadvantages of this approach, however, are mainly that it is fairly complex and that it does not incorporate the professional judgment and practical experience of the planner. However, the latter problem can be overcome by allowing the planner to interact with the design procedure and incorporate his or her knowledge in the final solution.

2.1.6 Summary of Approaches to the BNDP

Table 2.2 summarizes the strengths and weaknesses of the different approaches to the BNDP and shows the types of urban areas and planning horizons for which each approach is most applicable. Among all the approaches, the heuristic approach is the only one which may be used to accomplish the research objectives set in the previous chapter, particularly the development of an automated network design procedure. The heuristic approach shares similar benefits with the SA approach in terms of being systematic and comprehensive. Furthermore, if successfully implemented in an automated design procedure, the heuristic approach may complement the other manual approaches for network design by providing an alternative solution to the one proposed by the planners. This solution can then be compared to the manual solution, possibly suggesting

Solution Approach	**Strengths**	**Weaknesses**	**Uses**
Manual	• simple • inexpensive • requires less expertise • short time of implementation	• maybe ad hoc • usually not comprehensive • requires value judgments	• small urban areas with few bus routes • simple bus networks • short-range planning • fine-tuning of networks
Market Analysis Project	• systematic • comprehensive • not too expensive	• similar to the manual approach	• small urban areas • operation planning • more suitable for small/urban networks
Systems Analysis	• systematic • comprehensive • allows multiple objectives • not too expensive • widely used in practice	• few options tested • long time of implementation • bias towards current network • large data requirement	• medium/large urban areas • short range (operational) planning • complex networks with many routes • long range (conceptual) planning, but with less details
Systems Analysis with Interactive Graphics	• systematic • more options may be tested • shorter implementation period	• bias towards current network • quite expensive	• depending on models used, can be used for small-large areas, short-long range • moderately complex networks
Analytical	• simple geometric networks and demand functions	• over-simplification of real problem • not suitable for use in practice	• not applicable in real situations • best suited for screening analysis
Heuristic	• more comprehensive than most other approaches (less than SA) • systematic • many options tested • no bias towards current system • short implementation time	• may not produce optimal solution • approach still under development	• short-medium range planning • small-medium areas • new bus networks

Table 2.2: Summary of Approaches to the BNDP

beneficial modifications in the latter. Therefore, the heuristic approach for solving the BNDP will be adopted, and in the following sections, this approach is considered in greater detail.

2.2 The Heuristic Approach for Solving the BNDP

The restructuring of a bus network is likely to have significant impacts on ridership. In fact, one of the long term objectives and potential benefits of network restructuring is the possibility of generating additional ridership by attracting travelers from other modes. The heuristic approach for solving the BNDP may be broadly categorized according to the treatment of demand. Although the assumption of variable demand is more appealing in terms of the objectives of network restructuring, only a few of these heuristics (Dubois et al., 1979; Hasselstrom, 1981) allow variable demand. Difficulties involved in the variable demand assumption include the increased complexity in the formulation and solution of the model. Moreover, there is no strong evidence to suggest that the existing demand models are reliable for route network changes in public transport systems (Multisystems, 1982). Also, bus operators in most metropolitan areas are more concerned about the impact of changes on the current ridership than about the potential for generating new ridership. For these reasons and because of the reduced complexity, models which assume demand to be independent of the service quality offered are more often used at the design phase, combined with constraints to guarantee minimum service levels. In these models, ridership can be estimated separately (outside the model) after the restructuring if desired and, if it is significantly different from the initial one, the new demand matrix can be used to perform a second iteration through the design heuristic.

The heuristic approach includes two major components, route generation and frequency determination, which will be briefly reviewed in the following sections. Passenger assignment is an important element in the approach and is also described.

2.2.1 Route Generation

As mentioned previously, route generation (RG) suffers from a lack of criteria which characterize a "good" network of routes. In the absence of service frequencies, a particular layout of routes holds little information on the level of service provided to the passengers and the total vehicle requirement for the operator. In practice, this problem is solved by adopting a cyclic design process in which a bus network is manually designed by the planner, and then evaluated to provide information to help in modifying the design.

Heuristic procedures devised to solve the BNDP have tried to mimic this iterative process by replacing the manual RG by an automated procedure. Although the automated RG cannot incorporate the judgment of the planner in laying out the routes, it is far superior to exhaustively searching all possible combinations and efficiently eliminating the ones that do not meet certain requirements. One of the earliest examples of heuristic procedures for RG is by Lampkin and Saalmans (1967) who were the first to introduce the concept of skeleton routes - consisting of four nodes, with the end nodes acting as termini of the routes. Based on this concept, a heuristic procedure inserts feasible nodes iteratively into the skeleton route until it is completely developed. The nodal insertion process is based on "desirable properties" that are expected to be present in a "good" network, and ensures that the resulting network satisfies them. These properties reflect basic elements of the level of service provided - such as demand satisfaction and connectivity - and are expressed quantitatively in an objective function used to select feasible insertion nodes during the process of skeleton route expansion. However, one of the shortcomings of this heuristic procedure is that it takes no account of the available fleet size while developing the network of routes, and there is no clear indication of how the process should end.

Silman et al. (1974) applied a RG heuristic that is similar to the one developed by Lampkin and Saalmans to the city of Haifa, Israel. In this procedure, bus routes are generated by a process of insertion and deletion. The process starts with a non-empty initial set of routes which is then augmented by addition of routes selected from a set of candidate four-node skeletons. Once a skeleton is chosen, it is expanded in a manner similar to that in Lampkin and Saalmans' heuristic. The criteria for selecting a route from the candidate set is based on the marginal decrease in the total travel time caused by the addition of the route under consideration. Since frequencies (and thus waiting times) are not known at this stage, travel time between any two nodes is approximated as the minimum of in-vehicle travel time or walk time, in addition to a transfer penalty time in case of transfers (only one transfer is allowed).

Hasselstrom (1981) suggested a heuristic procedure which is among a very few ones to generate bus routes and determine their frequencies simultaneously. The procedure consists of three stages. In the first stage, it is assumed that there is a direct bus link between each pair of nodes in the street network and the frequency of service on this extended network is calculated using a simple model. Next, passengers are allocated on the alternative paths between each origin-destination pair so as to minimize the total time, as in a highway assignment. Little-used links are then eliminated and passenger flows are concentrated on the remaining routes, recalculating the frequencies of service until a satisfactory solution is reached. In the second stage, thousands of possible routes are generated, satisfying constraints on minimum and maximum lengths. In the third and final stage, optimal frequencies are estimated for all the generated routes (in the second stage) by choosing the routes that follow as closely as possible the passenger flows and the frequencies from stage one. The final set of bus routes and frequencies may be improved interactively with input from the planner. The procedure of Hasselstrom used a direct demand model both to estimate a desire matrix based on providing high-quality service throughout the area, and to reduce the demand as the actual design is developed providing less than ideal service between some origin-destination pairs.

Israeli and Ceder (1989) developed a model to create all routes (and transfers) that connect every node in the network and that satisfy round trip time and circuity constraints. Out of this huge pool of feasible routes, the model then generates smaller subsets, with routes in each subset maintaining network connectivity and deviating the least from their corresponding shortest paths. Each set is determined by heuristically solving the set covering problem (a non-linear integer programming problem), which determines the minimum number of routes (out of the pool of feasible routes) that are needed to cover all the nodes in the network. The alternative sets of routes are obtained by imposing incremental changes to an existing set by deleting the "worst" route and re-solving the set covering problem. For each of the subsets, the model assigns the demand and calculates a set of performance measures, including the fleet size requirement. At this point, the planner is able to choose the most suitable subset.

More recently, Baaj and Mahmassani (1993) presented a route generation heuristic as a single component in an artificial intelligence-based design procedure for solving the BNDP. Their route generation algorithm shares several aspects with previous route generation heuristics, notably the idea of skeleton routes (although the definition of skeletons is quite different than that in previous work). In the first step, the algorithm selects the terminal node pair from a ranked list of the M highest-demand terminal node pairs and uses it as a seed for the skeleton to be expanded. The skeleton route is either the shortest path connecting the seed nodes, or an alternate short path (slightly longer) with significantly different nodal composition than the former. The expansion of the M skeletons into routes may be performed according to one of four different node selection and insertion strategies, each one reflecting a certain trade-off between the level of passenger service and operator costs. The various strategies allow the generation of different sets of routes in the final solution. The process of route expansion is continued until a pre-specified level of demand satisfaction is achieved. The resulting set of routes may be improved in a separate procedure after the frequencies of service are determined, also in a separate procedure. Improvements to the solution are attempted at the system coverage level (discontinuation of service on routes that suffer from low ridership) and at the route structure level (joining low ridership routes with other medium to high ridership routes and splitting long routes).

The solution created by Baaj and Mahmassani is heavily guided by the demand matrix of the existing system and thus resulted in a network that is biased towards the existing one. Shih and Mahmassani (1994) extended that work and presented a design methodology that would be capable of designing a bus network using the transit center concept, with the possibility of providing coordinated service at the transit centers. In the first cycle of the overall design process, Shih and Mahmassani utilized the RG procedure of Baaj to create a set of M routes. Flows in the demand matrix are then assigned to the generated routes using a network analysis procedure and several node-level descriptors are computed. The transit center selection procedure is then applied to the given route configuration and the corresponding descriptors so as to identify suitable candidate centers that would offer "good" transfer opportunities. Next, the second cycle of the design process is initiated by the re-application of the RG procedure, this time with the transit centers ranked highest in the set of M skeleton seeds (this set also contains high-demand terminal node pairs). The expansion of the M skeletons is performed in a similar way as in the RG procedure of Baaj, except for minor changes that are introduced to account for the difference in route structure in the transit center concept.

In reviewing RG heuristics, two issues emerge. First, the concept of route skeletons has been used repeatedly by several authors for initiating the process. However, these skeletons are defined differently, either as an ordered sequence of nodes (with the first and the last acting as termini) or as a short(est) path between two terminal nodes. The advantages of each skeleton type have not been previously addressed, although the definition of skeletons might have a large impact on the quality of the solution produced. Therefore, the appropriateness of the skeleton concept to initiate RG will be considered in this thesis. Moreover, if a skeleton route is to be adopted, a comparison between the two types of skeletons needs to be made.

The other more fundamental issue is the need to investigate the possibility of efficiently assessing the vehicle requirement during RG in order to terminate the process once the available bus fleet is exhausted. This issue is important to consider in the design methodology if it is expected to be used in practice, since in real situations, the fleet size constraint is an essential component of any BNDP. Most of the previous heuristics do not address this issue. Some of them allow the generation of a pool of all feasible routes, and then successively select subsets of this pool, determining the frequencies and the vehicle requirement on each subset, and then replacing routes from the pool until a network satisfying the fleet size constraint is found. Other heuristics ignore the fleet size constraint completely, pre-specifying a fixed number of routes to be generated. One guaranteed, yet highly inefficient solution is to implement a vehicle allocation procedure within the route generation algorithm that would update the total vehicle requirement after the generation of each route. This issue will be discussed further in subsequent chapters.

2.2.2 Frequency Determination

The optimization approach to the problem of frequency determination on a given bus network may be viewed as finding the optimal allocation of the resources (bus-hours) on all the routes and across time periods so as to maximize (or minimize) some objective function subject to a set of feasibility constraints. Typically, this problem has been formulated with the objective of minimizing a cost function combining the passenger and operator costs. The former is often expressed in terms of the total travel time incurred by passengers in the network. Operator costs are difficult to quantify in terms of the decision variables (frequencies), so the total number of buses available to the operator is generally used instead, in the form of a fleet size constraint. Other constraints include the minimum operating frequencies and the allowable passenger loading on all routes.

Lampkin and Saalmans (1967) formulate the problem of frequency determination with the objective of minimizing the total passenger travel time, expressed as a function of service frequencies, subject to a fleet size constraint. The total travel time is the sum of individual travel times between each origin-destination pair and consists of wait time, in-vehicle travel time and transfer time. Travel time between each node pair is determined by considering three mutually exclusive cases. In Case A, at least one bus route connects the node pair and it is less beneficial to walk between the two nodes. Case B also considers at least one bus route between the node pair but allows for the possibility of walking instead of waiting for the bus. In Case C, no single bus route exists between the node pair and passengers are forced either to walk or to make a transfer trip. The overall matrix of travel time is calculated first by using walking time in all the cells. Next, node pairs corresponding to cases A and B are considered and their travel time values are updated. Finally, the matrix is improved by comparing for each node pair the direct travel time and the travel time through an intermediate node. The solution technique used is a modified random search procedure in which an initial (guessed) solution of the vector of frequencies is used to start the procedure. After that, new values of the frequencies are produced by random perturbation from the best values of frequencies found to date.

Silman et al. (1974) extended the frequency determination procedure of Lampkin and Saalmans and expressed the objective function in terms of the sum of total travel time and a discomfort penalty. The only feasibility constraint used in this problem is the fleet size constraint and, as in the procedure of Lampkin and Saalmans, the total travel time is calculated by assuming simple passenger path choice strategy, thereby eliminating the need for a formal passenger assignment procedure. The discomfort penalty is included in the objective function as a way of dealing with the capacity constraint which was neglected by Lampkin and Saalmans. The discomfort penalty consists of a cost based on the number of standing passengers per bus and, by minimizing it, the bus capacity constraint is implicitly considered.

Han and Wilson (1982) address the problem of allocating buses in the case of a heavily utilized network with overlapping routes. In such systems, many origin-destination pairs are characterized by multiple paths, so the passenger assignment problem was addressed explicitly. The problem is formulated with the objective function of minimizing the occupancy level at the most heavily loaded point on each route in the system, subject to three constraints: (1) loading feasibility, which stipulates that passengers not be prevented from boarding a bus on their preferred route because of insufficient capacity; (2) the passenger flow assignment, which guarantees that the bus passenger flow on any link is dictated by the passenger assignment submodel; and (3) the fleet size constraint, which ensures that the available fleet size is not exceeded. The passenger assignment subproblem uses a simple path choice criteria which divides passenger flows into captive (served by a unique preferred path) and variable trips (otherwise). The solution to the problem is performed in a two-stage heuristic which first finds the minimum number of buses required, and then allocates surplus buses (if any).

Furth and Wilson (1981) treat the problem of frequency determination as a constrained resource allocation problem with the objective of maximizing the net social benefit (consumer surplus), subject to constraints on subsidy, fleet size and maximum headways. The ridership on each route is assumed to vary as a function of the frequency of service on that route, and is used in the objective function to express the social benefit. The formulation of Furth and Wilson does not require passenger assignment, and ignores the effect of route competition and complementarity effects.

In summary, the frequency determination (or vehicle allocation) problem is well-behaved when formulated as an optimization problem. The solution of the problem, however, may not be computationally tractable unless certain simplifying assumptions about the passenger's path choice process are made. A key issue in this regard is the assumption of route independence, which greatly simplifies the problem so that a passenger assignment submodel would not be necessary. What needs to be addressed in the proposed methodology, however, is the situations in which this assumption may be valid and the impacts of using it on the overall network design process.

2.2.3 Passenger Assignment

Passenger assignment becomes an important part of the frequency determination problem in complex networks with common route segments. From a behavioral standpoint, the objective of the bus passenger is to minimize his or her "inconvenience" which is a weighted sum of access time, wait time, travel time, transfer inconvenience and travel cost. The transit passenger assignment problem has been studied by many authors. Following Dial (1967), many passenger assignment algorithms were based on the assumptions of deterministic running times and exponentially distributed headways on all routes, in which case the market share of passengers for a particular route is simply its frequency share.

Dial's model assigns passengers to a single optimal path in the network with minimum expected travel time. In the case of common routes (or paths) with equal travel times, the assignment is proportional to route frequencies (frequency share rule). Le Clercq (72) presented an improved algorithm for solving Dial's model; however, in most other respects, his approach is similar to that of Dial. Andreasson (1977) expanded on the models of Dial and Le Clercq by performing path assignment based on a simple heuristic to include paths in the passenger's optimal choice set. If the travel time conditional upon choosing a given path is less than the minimum over all paths of the headway plus the travel time after boarding, that route is included in the path assignment. Once the optimal path choice set is determined, path assignment is again based on the relative route frequencies. The procedure discussed by Andreasson was later incorporated as part of the Volvo transit planning package VIPS.

Jansson and Ridderstolpe (1992) show that when headways are deterministic rather than exponential, the frequency share model does not always hold, with passenger shares depending on the degree of schedule coordination. Their model, which is used in the VIPS-II software, assumes the existence of a vehicle timetable, to which vehicle movements adhere perfectly. As a result, depending on the level of coordination of routes, different path assignments may be inferred. For coordinated routes, assignment is again based on the frequency share rule, whereas for uncoordinated routes, a heuristic is used based on the presumption that vehicle departures on any route are uniformly distributed between departures on any other route.

Spiess and Florian (1989) present a different model for transit assignment whereby the traveler does not choose a path, but rather selects a strategy. Spiess' approach allows the transit rider to select any subset of paths leading to the destination with the first vehicle to arrive determining which of the alternative routes is actually taken on an individual trip. This optimal strategy concept allows more realistic modeling of the traveler's behavior and is used in the transit network evaluation model EMME/2.

Hickman (1993) developed a path choice model which takes into account the impacts of real-time information on transit passenger behavior. This dynamic model assumes that the passenger decides his/ her boarding strategy as vehicles arrive at the terminal. The dynamic model is also useful in describing adaptive path choice decisions made during the passenger's trip.

To summarize, transit assignment algorithms may be very complicated to implement and the marginal benefit of using a more realistic assignment algorithm may not be justified, considering the computational cost involved. At the same time, the focus of the proposed methodology will be primarily on the route generation component. The role of frequency determination within the methodology would be to assess the vehicle requirements on each route and to calculate global performance measures in order to guide the process of route generation, rather than to provide a detailed calculation of the frequencies. Moreover, the solution obtained from the proposed methodology may be analyzed in depth by the use of a readily available transit evaluation model such as EMME/2.

Chapter 3

Proposed Methodology

In this chapter, the proposed methodology for solving the BNDP is presented. Section 3.1 discusses the basic features of the solution framework, namely its capability to combine individual components of the network design problem adapted from the literature into a single coherent process. Section 3.1 also describes the development of the network within this framework and the fact that the design process does not yield a single final solution. Section 3.2 discusses the treatment of demand within the proposed methodology and section 3.3 describes the two service concepts provided by the solution framework: general and transit-center. Section 3.4 presents an overview of the two major components of the proposed methodology: (1) the route generation process which creates a single route at a time and incorporates it in the set of routes produced thus far; and (2) the vehicle allocation process which determines the frequency of service on each route of the network by allocating the available bus fleet so as to satisfy a certain service objective and a set of operational and policy constraints. Section 3.5 describes how the planner can introduce improvements to the network solutions created by the design procedure. Finally, section 3.6 outlines the overall design process and provides a flow chart describing the major components.

3.1 Solution Framework

The solution framework combines previous work on the bus network design problem, or specific components of that problem, into a single automated procedure that takes into account the passenger's service requirements and the limited resources of the operator. Although it shares some aspects of modularity with other heuristic design procedures, the proposed procedure is distinguished by the integration of vehicle allocation into the route generation process, thereby addressing the important relationship between those two major components, as well as the tradeoffs of creating networks with different number of routes. In all the previous approaches reviewed in chapter 2, with the exception of the one proposed by Hasselstrom, the generation of a pre-specified and often large number of routes has been performed in isolation from the vehicle allocation process, without any consideration of the operator's resource constraints. Consequently, the evaluation of the final solution may well reveal a network that is not consistent with the available resources.

To achieve a balance between the size of the network and the available resources of the operator, the proposed design procedure is conceived to evaluate the network (partially) every time a new route is created. The purpose of the evaluation is to estimate the total number of buses needed in order to decide whether to create more routes or to terminate route generation. Two methods for evaluating the network are envisioned. The first method calculates the route frequencies using simple rules of thumb, such as the peak load factor method or the square root rule, and uses them to determine the vehicle requirement on each route. These simple methods do not require passenger assignment and impose little computational cost on the overall process. The second method determines a measure of the network's level of demand satisfaction - which is a global performance measure that depends only on the nodal composition of the network (not a function of the service frequencies) - and uses it as an indication of the network's development level and its vehicle requirement. Although it does not explicitly calculate a value of the total vehicle requirement, the second method is still fairly reliable in assessing the amount of resources required and is more computationally efficient than the first method.

If the evaluation suggests the network is likely to consume all the available resources, a more elaborate vehicle allocation process is called upon to determine more accurately the total number of buses required and to calculate a range of network performance measures. On the other hand, if the network requires less buses than the available fleet, vehicle allocation is bypassed, another route is generated and the new network is re-assessed. The route generation process is terminated when vehicle allocation exhausts all the available resources.

The ability to perform network evaluation at close intervals in the route generation process adds another advantage to the proposed solution framework that was not fully explored in this thesis, but may be considered as an extension to this work. The information obtained from the intermediate network evaluation may be used in a variety of ways to guide the generation of routes, thereby transforming network development into a dynamic process with the expectation of achieving a better final solution. However, such actions may eventually reduce the overall efficiency of the design process and, hence, hinder the main feature of the proposed methodology.

Another important feature of the proposed framework is related to the development of the final solution. Because of the nature of the network design problem, it may well be that a better overall solution can be obtained if the route generation process is terminated a few iterations *before* all the resources are exhausted and the remaining buses are allocated to the most heavily patronized routes. This uncertainty is an inherent characteristic of the problem and motivates the investigation for a measure of the marginal benefit of adding a bus to an existing route to improve its wait and transfer times, as opposed to using the same bus to serve additional trips on a newly developed route. In fact, the proposed solution framework builds on that aspect by creating a file that records the performance measures and basic characteristics of several solution networks that satisfy minimum requirements of directness and coverage. As will be described in chapter 5, this file contains two cases of vehicle allocation for each network configuration recorded, the first one requiring a minimum number of buses and the second one utilizing all the available vehicles.

Consequently, when the overall design process is terminated due to the unavailability of additional resources, it yields more than the solution corresponding to the last network configuration (which consumes all the available resources). Instead, several network solutions are contained in the file, each one with a slightly different number of routes, vehicle requirements and performance measures. The planner may then analyze the data at hand and decide on the overall "best" solution.

3.2 Treatment of Demand

The proposed methodology adopts a fixed demand approach to the BNDP, treating the given demand matrix as fixed and independent of the service quality offered between any origin-destination pair. As discussed in the previous chapter, this assumption simplifies the formulation of an already complex problem, and does not necessarily preclude the possibility of conducting a more comprehensive analysis of the system. The impacts of the network restructuring on ridership can still be considered, although not within the design procedure itself, by feeding the service characteristics of the modified network into a separate demand model which can estimate the demand matrix that is consistent with the service provided on that network. If this matrix is significantly different from the one input to the design process, a second iteration through the design procedure can be performed with the new demand matrix.

By assuming a fixed demand matrix, the focus of the design process is placed on increasing the benefits to the current system users rather than on the potential of generating additional ridership. However, bus (and other transit) systems are generally recognized to be an important social service, with the social benefits - including reductions in traffic congestion, pollution and energy use - being proportional to the number of riders. These benefits are likely to be more significant than the ones experienced by the current users, especially when the long-term development of the system is considered. Since the proposed methodology for BNDP addresses the strategic planning of bus systems, an external feedback loop for estimating additional ridership should be used to complement the design procedure. However, within the procedure itself, the variation in demand is not explicitly considered.

3.3 Network Concepts

A demand matrix normally reflects the type of the network used to generate it. For example, if the bus network is radial, the demand matrix is expected to exhibit a radial structure as well, with one row (column) dominating all other rows (columns). If the same matrix is then used to restructure the network, it is very likely that the modified route configuration will also follow a radial pattern. This is particularly true if the design process used attempts to serve all the existing nodes in the current network. On the other hand, the input demand matrix could be different from the existing transit demand, reflecting "desired" trip patterns that correspond to a presumed network concept.

Most of the previous approaches to network design do not specify a particular network concept - such as radial or multi-centered - and tend to be guided only by the existing demand matrix in determining the route configuration. Consequently, the type of network generated is likely to reflect the existing travel patterns, resulting in a network similar in concept to the one that already exists and serving the same type of trips. This limitation is addressed in the proposed methodology by allowing the planner to choose between specifying a transit-center network concept prior to the initiation of the design process, or not specifying a particular concept and letting the process be guided exclusively by the demand matrix. These two options are described in the following sections.

3.3.1 General Network Design

If no network concept is specified, route generation is guided exclusively by the existing demand matrix. The selection of the terminals and intermediate nodes for each route is dictated primarily by the number of trips made between each pair of those nodes. The network is developed by sequentially selecting feasible route "skeletons", sorted in decreasing order by the number of trips, and expanding them into full routes by means of nodal insertions. No attempt is made to serve all the nodes in the existing system, although the design process aims to satisfy the largest amount of demand possible. Consequently, the final solution obtained may be similar to the existing bus network; however it is also expected to deliver a better overall performance because of the tradeoffs of service expansion and concentration that are investigated in the design process.

3.3.2 Transit Center Network Concept

In many major metropolitan areas, many activities such as employment, shopping, medical and educational that were once concentrated in the downtown area have gradually decentralized as a result of extensive suburban development. These fundamental changes in land use have made it difficult for the conventional bus network, which typically follows a radial pattern from the central business district into the suburbs, to provide adequate service for many desired trips within the suburbs. Consequently, some cities have proposed different network concepts designed to provide better service to multi-nucleated metropolises with significant suburban development. The transit center concept revolves around major community retail and/ or employment centers functioning as effective hubs around which transit operations are focused. These centers are served by trunk or main lines that connect the various centers, operating at high frequencies and along the most direct routes. Feeder buses or demand responsive service also serve the transit centers and provide a local collection-distribution function.

When the transit center concept is specified as the basis of the design process, route generation needs to be modified in a number of ways. First, the demand matrix needs to be transformed in order to reflect the redistribution of origin-destination flows in the transit center network. This transformation investigates the possibility of directing any origin-destination flow via an appropriate transit center by attempting to assign each flow to one or two transit center nodes. As a result of the transformation, the demand among transit center nodes, as well as the demand between each of these centers and the other nodes, is increased, reinforcing the role of the centers as major hubs in the network and influencing the process of route generation to consider creating routes to serve them. Because transit center nodes now capture a large portion of the total demand, trunk lines are likely to be created at an early stage of route generation, which seeks the most direct path for this type of route. Feeder routes which increase the levels of network coverage and directness by providing access to the transit centers are also encouraged by the demand matrix transformation. Flows which are not affected by the transformation would still be served by the transit center network, either by being incorporated in feeder routes or by requiring the route generation process to create direct routes to serve them.

The performance of a transit center solution is evaluated based on the original (non-transformed) demand matrix and may well be inferior to that obtained if no particular concept is specified. The reason is that the demand matrix reflects a well-established current service that is ineffective at serving the many newly-developed and dispersed destinations within its service area. Consequently, the demand between those destinations may have been kept artificially low and the routes created to serve them in the solution network may seem to be unjustified. Assessing the performance of the solution network in that case is better performed with a demand matrix estimated based on the proposed service changes.

3.4 The Network Design Procedure

The network design procedure consists of two major components - route generation (RG) and vehicle allocation (VA) - which are integrated into a single automated process. These two components are described in the following sections.

3.4.1 Route Generation

For a given demand matrix, street network information and set of design parameters, the RG process constructs a single bus route at a time and adds it to the set of routes already created. The process of single route generation is preceded by the determination of all feasible initial skeleton routes, each consisting of a sequence of three nodes (two terminal nodes plus one intermediate node) satisfying operational constraints on round trip time and circuity. At each iteration of the single route generation process, the feasible initial skeleton with the highest number of trips is selected and expanded into a connected path by means of nodal insertion. Feasible insertion nodes are selected according to an objective function which combines measures of route directness, and network coverage and connectivity into a single expression. The connected path obtained is referred to as a base route and is investigated for additional insertions that might increase its share of demand without causing lengthy deviations. After all feasible additional insertions have been examined, the route is considered fully developed and is added to the set of previously generated routes.

3.4.2 Vehicle Allocation

VA aims at assigning the available fleet of buses to the network of routes developed thus far so as (1) to ensure enough capacity on each route to accommodate all passengers choosing it and (2) to allocate any surplus buses, if they exist. VA achieves the first objective by assigning the demand matrix onto the network of routes according to a certain model of passenger behavior that depicts the preference of users towards choosing a particular travel path among a set of options. VA includes a path choice model which assigns passengers on each route as a function of its frequency , as well as the frequencies on the competing routes offering similar service. In turn, the frequency of service on each route is determined by assigning buses on the route according to its maximum link flow (which is governed by the passenger assignment). Therefore, the VA process attempts to reach an equilibrium in which the vehicle and passenger assignments are consistent with each other. The allocation obtained when this equilibrium state is reached is referred to as the base allocation and requires the minimum number of buses. The VA process also calculates performance indicators corresponding to the system after base allocation and includes measures of the total in-vehicle travel, wait and transfer time.

If buses remain unused after the base allocation is performed, these buses are assigned using a surplus allocation procedure which also calculates the new performance measures corresponding to the surplus allocation. Both sets of performance measures (corresponding to the base and surplus allocation) are used at the end of the design process to facilitate the selection of the overall "best" solution by the planner.

3.5 Network Improvements

Solutions obtained from the design procedure may benefit from certain modifications to improve their overall performance. Because they involve a high degree of judgment and interaction with the planner, network improvements were not developed as part of the automated design procedure. Instead, the emphasis was placed on creating a route generation process that would reduce the need for further improvement. In particular, RG includes provisions to discourage the formation of overlapping routes. This would improve the network's demand coverage level by allowing the generation of fairly independent routes serving a larger number of distinct nodes. Moreover, RG prevents the formation of excessively circuitous routes and limits the round trip times in order to promote directness of service. These constraints are discussed further in chapter 4.

The solution produced by the RG procedure may require improvements that target short and/ or low-ridership routes. These routes may be improved in more than one way, each with significantly different implications. Low-ridership routes may be joined with medium to high frequency routes, thereby reducing the total vehicle requirement and improving the network's directness. Alternatively, low-ridership routes may be totally eliminated from the network and the buses operating on them used to generate new routes. The proposed methodology was conceived to allow the planner to modify the solution obtained from the design procedure and to feed back the modified network into VA which re-evaluates it and calculates new performance measures. Also, if the modified network requires less buses, RG is resumed in order to create additional routes.

3.6 The Overall Design Process

A flow chart displaying the overall bus network design process is shown in figure 3.1. RG starts with the (optional) demand matrix transformation procedure and proceeds towards the main loop where single routes are created. In the first few iterations of the RG process, individual routes are created without allocating buses to them, since the number

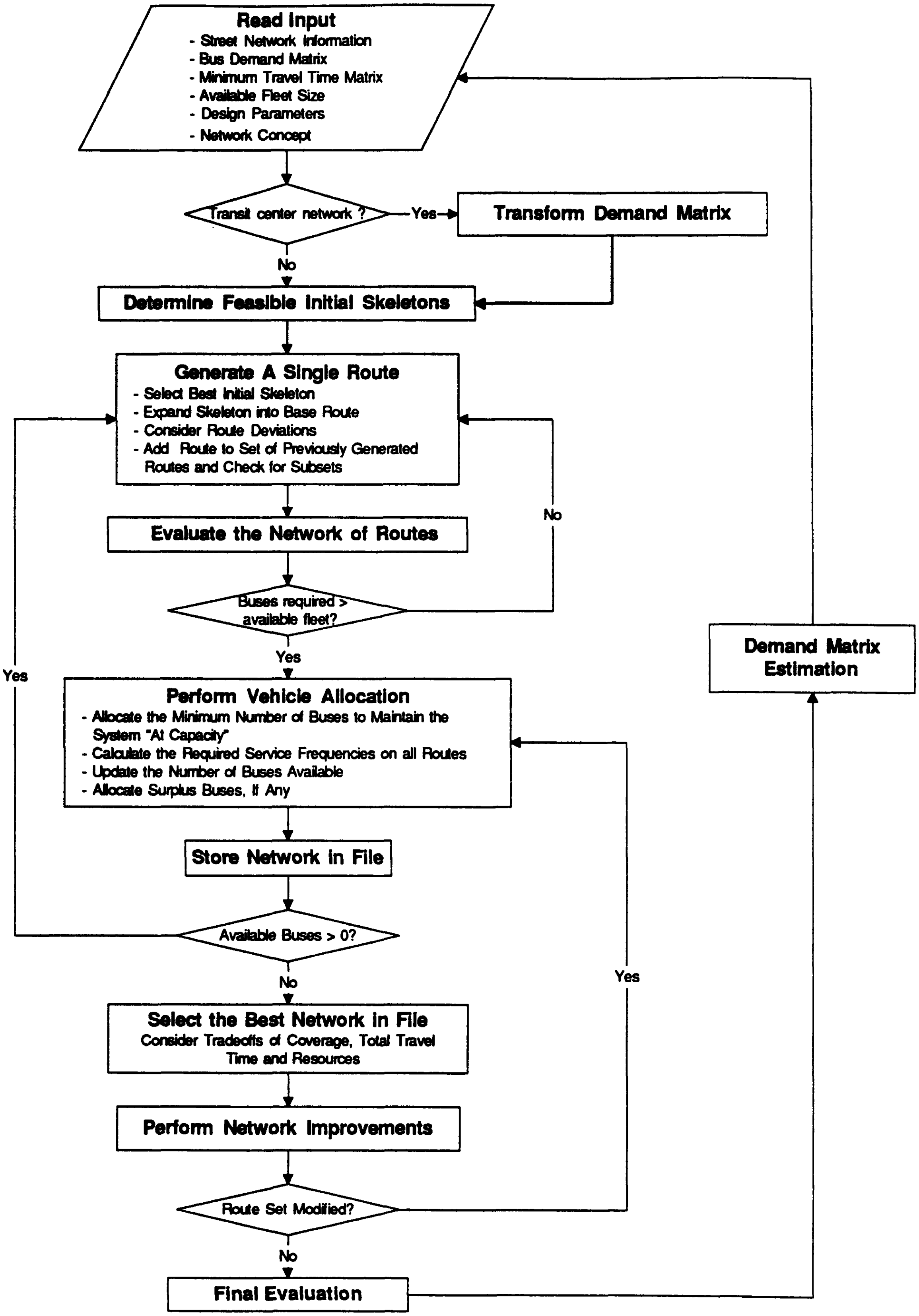


Figure 3.1: The Bus Network Design Process

of buses required by the current network is likely to be below the available fleet size (the vehicle requirement is estimated within the main loop by the network evaluation procedure). As long as this condition prevails, VA is bypassed in order to reduce the total running time. However, when the network reaches a critical size, this condition ceases to be valid and the VA process becomes necessary to estimate the total vehicle requirement more reliably. After it is performed for the first time, the VA process is performed repeatedly after each new route is created until all the resources are exhausted. VA calculates the frequency of service on each route, the total vehicle requirement and the performance measures, and then stores the current solution in a file. When no more buses remain, this file contains network solutions with different sizes and performance measures, but all with vehicle requirements in the vicinity of the available fleet. These solutions are then analyzed by the planner, taking into account tradeoffs among travel time, demand coverage and buses required. The "best" network is then selected and, if necessary, subjected to improvements. If the network is modified after improvements, VA is performed again and the vehicle requirement is recalculated; otherwise, the design process is terminated. A new demand matrix may then be estimated (outside the design procedure) and a new iteration of the overall procedure may be performed, if necessary. The design process is discussed further in chapter 5, after the RG and VA processes are presented.

Chapter 4

Route Generation

This chapter focuses on the generation of bus routes from a given bus demand matrix and a description of the street network that includes for each node its neighboring nodes and the in-vehicle travel times on all possible connecting links. The main component in the network development is the single route generation (SRG) process, which creates individual routes between two terminals by specifying intermediate nodes along a path in the street network.

Two general approaches for developing a single route exist and are discussed in the next sections. The approach for SRG is based partly on the route generation process originally proposed by Lampkin and Saalmans (1967) and shares some aspects with the route generation algorithm implemented by Baaj and Mahmassani (1993). The approach starts by selecting an initial route skeleton consisting of a sequence of three nodes and successively inserts intermediate nodes until a connected base route is found. Additional insertions that increase the base route's share of the total demand without subjecting the passengers to significant time delays are then performed.

The route generation (RG) process creates bus networks that provide fixed-schedule uncoordinated service among the various routes. The process can also construct networks around the concept of transit centers, incorporating better transfer opportunities at these centers, as well as faster and more direct service between two centers in the route design. A similar feature has been implemented in Shih and Mahmassani (1994) with additional considerations for demand responsive services.

Section 4.1 presents an overview of the RG process for the case of a general network design, including a flow chart and a summary of the major steps involved. Section 4.2 describes the proposed approach for route generation and discusses the options that were considered in selecting the approach. The input and the design parameters required for the process are described in section 4.3. Sections 4.4 and 4.5

describe the formation of feasible initial skeletons and the selection of the best initial skeleton before each iteration of the SRG process. The expansion of skeletons into base routes is described is section 4.6 and deviations from the base route are considered in section 4.7. Finally, section 4.8 discusses the modifications implemented in the SRG process for designing networks based on the transit center concept.

4.1 Overview of the Route Generation Process

In the review of previous methods of bus network design presented in chapter 2, two major approaches for route generation were identified. The first approach is attributed to Lampkin and Saalmans and is based on selecting the best four-node initial skeleton, starting and ending with permissible termini, and then expanding the skeleton into a complete route by means of nodal insertions. The second approach was developed by Baaj and Mahmassani and consists of identifying the shortest path (or an alternate short path) between two selected termini, and then expanding this base route by inserting neighboring nodes. In the latter approach, the process is initiated by selecting the seed nodes of the route from a list of terminal node pairs, ranked in decreasing order of the number of trips.

The RG process proposed in this methodology combines some of the basic elements of the previous approaches and is comprised of four main components: (1) pre-processing and the determination of all feasible initial route skeletons, (2) selection of the best initial skeleton, (3) expansion of an initial skeleton into a base route, and (4) consideration of route deviations. The last three components form the SRG process which is performed iteratively until a satisfactory network is developed. The first component is executed only once at the beginning and is used to initiate the SRG process. In the case of a transit center network design, this component is preceded by a demand matrix transformation procedure which examines the possibility of routing each origin-destination flow via one or two designated transit center nodes.

The RG process starts by finding all combinations of initial route skeletons - consisting of a terminal node pair and a major intermediate node. Major nodes are either specified in the input, or determined from the demand matrix as the set of nodes with a large number of associated trips. Each skeleton combination undergoes a filtering process which discards skeletons that violate constraints on round trip time and circuity, and the remaining ones form the set of feasible initial route skeletons.

The SRG process is initiated by approximating the demand served by each feasible initial skeleton and selecting the one with the largest number of trips. The demand associated with each feasible initial skeleton is calculated by aggregating the number of trips between each node pair in the skeleton, ***only if the node pair is not already served by any of the routes that have already been developed.*** This rule is used to discourage the selection of initial skeletons whose nodes are already served by existing routes, thereby reducing the amount of overlap in the network.

After the best initial skeleton is selected, it is then expanded into a base route using an iterative process of nodal insertion. A base route is a connected and relatively direct path (although not necessarily the shortest) between the route's termini and is created as a first-stage expansion of the skeleton route. At this stage, nodal insertion is performed in "gaps" in the skeleton route, with a gap defined as any situation where two consecutive nodes in the skeleton are non-adjacent (i.e. they are not directly joined by a link). Feasible insertion nodes for filling a gap are determined according to two heuristic rules which reject nodes causing (1) a large deviation of the skeleton route from its corresponding shortest path, or (2) a skeleton route round trip time exceeding a maximum allowable value. Feasible insertion nodes are evaluated using an objective function expressing weighted measures of route directness, coverage and connectivity to previously generated routes, and the node which maximizes the objective function is selected and inserted in the skeleton. The process of nodal selection and insertion is repeated for all the gaps in the skeleton until the base route is fully developed or until a gap cannot be filled. In the latter case, the route under expansion is considered infeasible and is discarded. The second stage of route expansion investigates detours from the base route caused by nodes that are one-link away between two consecutive (and adjacent) nodes on the base route. Feasible insertion nodes are determined as in the first stage expansion process, except for a more restrictive condition on route circuity in order to prevent large deviations of the route from its current path. Also, a feasible node is not inserted unless the additional gain in demand outweighs the increase in travel time caused by the detour, calculated for all the passengers affected by it. After the last two consecutive nodes on the route have been examined for possible detours, the route is considered fully developed and is compared to all the routes generated thus far to determine whether it is a subset of any of them. If that is the case, the route is deleted to prevent unnecessary service duplication.

The flow chart of the RG process is shown in figures 4.1a and 4.1b. After a route is fully developed (one iteration of SRG is completed), the process may be either repeated or terminated, depending on the performance of the network of routes obtained thus far. This issue will be addressed in the next chapter after the vehicle allocation process is presented. The major steps involved in the RG process may be summarized as follows:

Step 0: Pre-processing and determination of feasible initial skeletons

- Read the input and design parameters. If the transit center concept is specified as the basis of the design process, transform the demand matrix by routing origin-destination flows via one or two transit centers, where appropriate.
- Create all three-node initial skeleton combinations, consisting of a terminal node pair and an intermediate destination node.
- Determine the set of feasible initial skeletons by eliminating the ones with unacceptable round trip time or circuity.
- Calculate estimates of the demand served by each feasible skeleton and list all feasible skeletons in decreasing order of demand.

Step 1: Start the SRG process (steps 1 through 4) to create the k^{th} route

- Modify the list of feasible initial skeletons by updating the value of demand served by each skeleton. Demand between two nodes in the skeleton that are already served by a route does not count towards the demand associated with that skeleton.

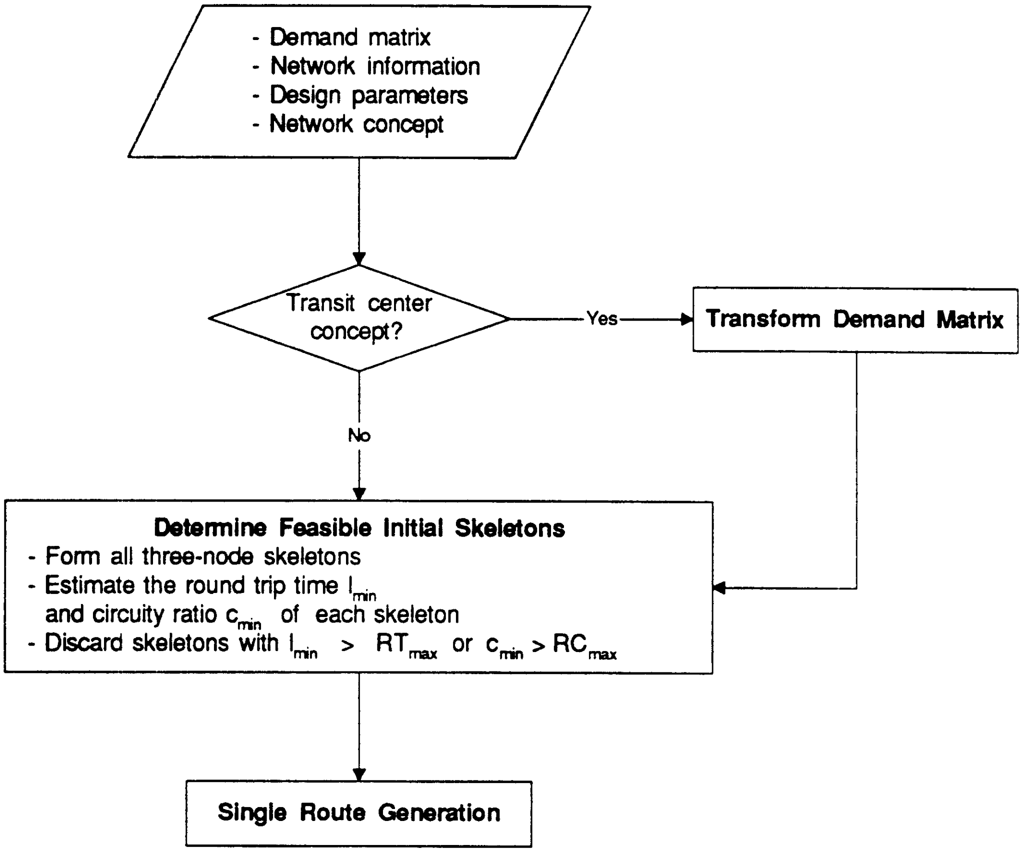


Figure 4.1a: Pre-Processing for the Route Generation Process (Step 0)

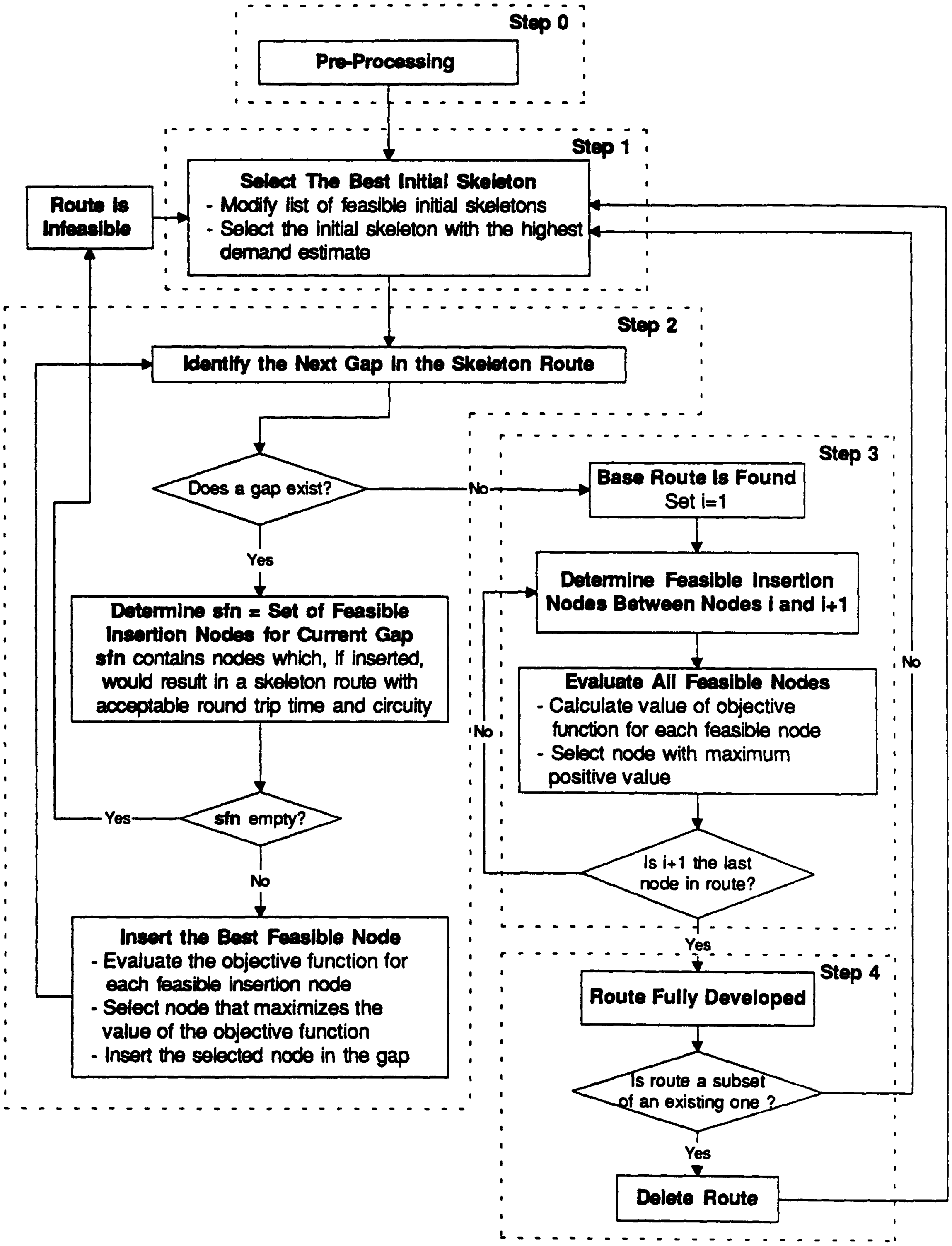


Figure 4.1b: The Single Route Generation Process

- Select the initial skeleton (*r*,*d*,*s*) with the highest demand value.

Step 2: Expand the initial skeleton (*r*,*d*,*s*) into a base route

- For each gap in (*r*,*d*,*s*), examine all feasible insertion nodes and insert the one which maximizes the objective function.
- Repeat step 2 until a base route consisting of a sequence of adjacent nodes is found, or until a gap cannot be filled.
- If a base route is found, go to step 3; otherwise, quit the expansion of the current skeleton and return to step 1.

Step 3: Consider deviations from the base route. Set $n = 1$

- Locate a feasible insertion node that is one link away from nodes n and $n+1$ on the base route. If the additional demand satisfied by inserting that node outweighs the delay caused by the detour then insert the node.
- Set $n = n+1$ and repeat step 3 until node $n+1$ is the last node on the base route.

Step 4: Route *k* is fully developed

- If route *k* is a subset of a previously generated route, then delete it. Otherwise, add route *k* to set of routes.
- Go to step 1 if more routes can be generated.

4.2 The Proposed Approach for Route Generation

In the review in chapter 2, most of the heuristic approaches to route generation consisted of a repetitive and identical process of developing a single route until some termination criterion is met (Lampkin and Saalmans, 1967; Silman and al., 1974; Baaj and Mahmassani, 1993). Furthermore, the generation of routes was divided into two major parts: the creation of route nuclei, and the expansion of these nuclei into full routes. The route nucleus varied in structure among the several approaches, but generally consisted of a terminal node pair and a few intermediate nodes, enough to form an adequate representation of the route it describes and to assess the benefits of including that route in the final design. Since a fully developed route may contain many nodes, the number of possible route combinations is often very large, even in relatively small networks. Moreover, many routes share similar features with other routes and may differ only in a small number of nodes or segments. Therefore, these similar route combinations may be substituted by a smaller number of route nuclei which can be formed and evaluated more efficiently in the first part of route generation. As a result, a priority list for developing route nuclei is established and, in the second part, nuclei are successively selected from that list and expanded into full routes.

The proposed RG process follows the same partitioning mentioned above and combines basic elements from previous approaches. Two distinct approaches for creating route nuclei were presented in chapter 2: (1) the route skeleton approach originally proposed by Lampkin and Saalmans and (2) the base path approach implemented by Baaj and Mahmassani. In the first approach, a route nucleus between two terminal nodes is a *sequence* of four (not necessarily connected) nodes including two intermediate nodes between the termini, whereas in the second approach, it is the shortest *path* between the two terminal nodes.

If these approaches are compared, it can be argued that a node-based route skeleton is capable of exploring more alternative routes for a given terminal node pair than an initial path. By taking advantage of a broad choice of intermediate nodes, the skeleton route approach is in effect evaluating several trajectories of the final route, each one corresponding to a particular selection of intermediate destinations. In contrast, the base path approach determines only the single shortest path between each two terminal nodes, imposing a limit on the trajectory of the route and the destinations it ultimately serves. The objective of investigating multiple routing alternatives for a terminal node pair and selecting the "best" routing is thus defeated. Instead, the initial path approach is overly concerned with finding the most direct route between a certain terminal node pair, overlooking the benefits from considering paths, other than the shortest one, that might be more useful in serving important destinations.

To better understand the advantage of the initial skeleton approach in this regard, consider the situation in figure 4.2 which shows a simple example of a typical radial bus transit network. The downtown district is the only major destination in the area and is assumed to be concentrated at node *n*. Suppose that nodes t_1, t_2, t_3 and t_4 are existing terminals for the two routes in the system, R_1 serving the north-south corridor and R_2 serving the east-west corridor. Assume further that for some reason (traffic congestion, lack of space, etc.), node *n* cannot be a terminal. Because the central core is the only major destination in this purely radial system, bus trips are made only between node *n* and the nodes on each leg of routes R_1 and R_2. To compare the two approaches to route generation, the demand matrix corresponding to the system of figure 4.2 will be used to re-design the bus network serving the same area. Note that in this example, there are six possible combinations of terminal node pairs.

In the case where the base path approach is used, the highest-demand terminal node pair is selected at each iteration. In this example, however, the number of trips between any two terminal nodes is zero, and no clear choice of terminals exists. However, for the sake of argument, suppose that (t_1,t_2) is chosen (arbitrarily) to generate the first route. This route, $R_1^{'}$, shown in figure 4.3a, does not include node *n*. In fact, it is highly unlikely that node *n* would be part of $R_1^{'}$ - even after detours from the initial shortest path are considered and despite the fact that an appreciable demand exists between *n* and both t_1 and t_2. The reason is that node *n* is not on the shortest (or near shortest) path between the route's termini.

On the other hand, if a skeleton approach is used for the same example, all skeleton combinations would be formed first and then ranked in decreasing order of the number of trips. Using only one major destination as an intermediate node, six possible initial skeletons can be formed, each combination consisting of node *n* in between two terminal nodes. Note that in general, more major destination nodes are present in the network and skeleton combinations outnumber terminal node pair combinations. Each

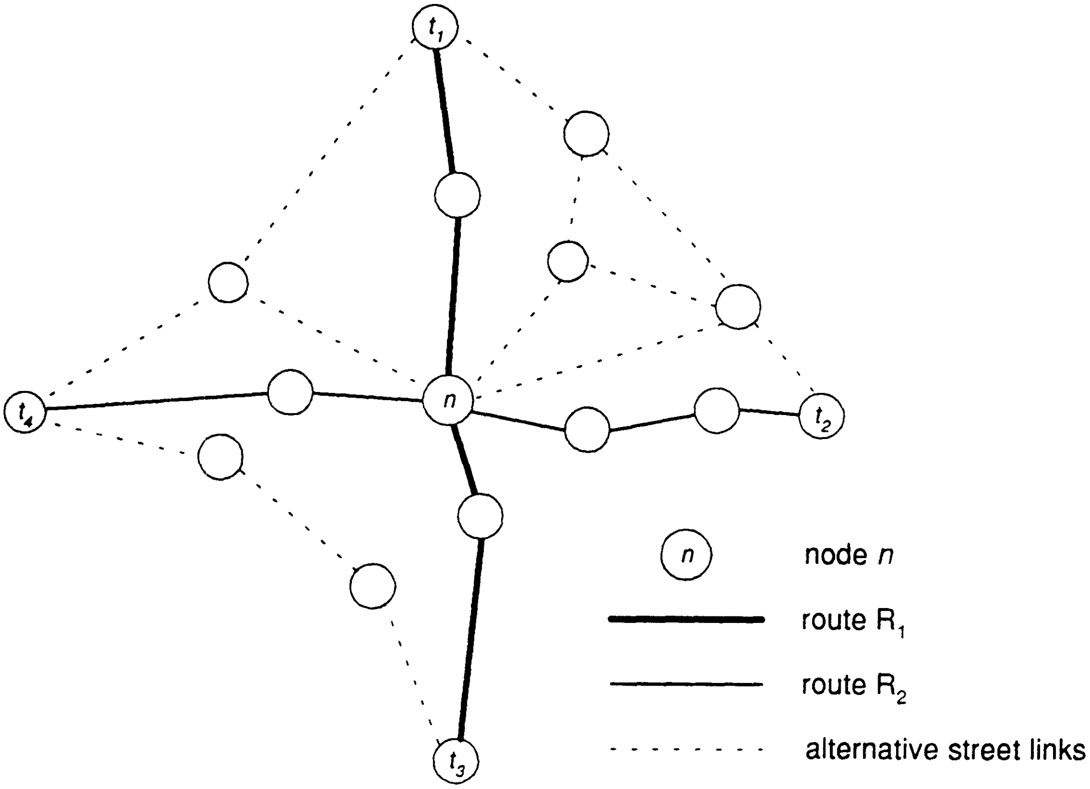


Figure 4.2: Existing Radial Bus Network

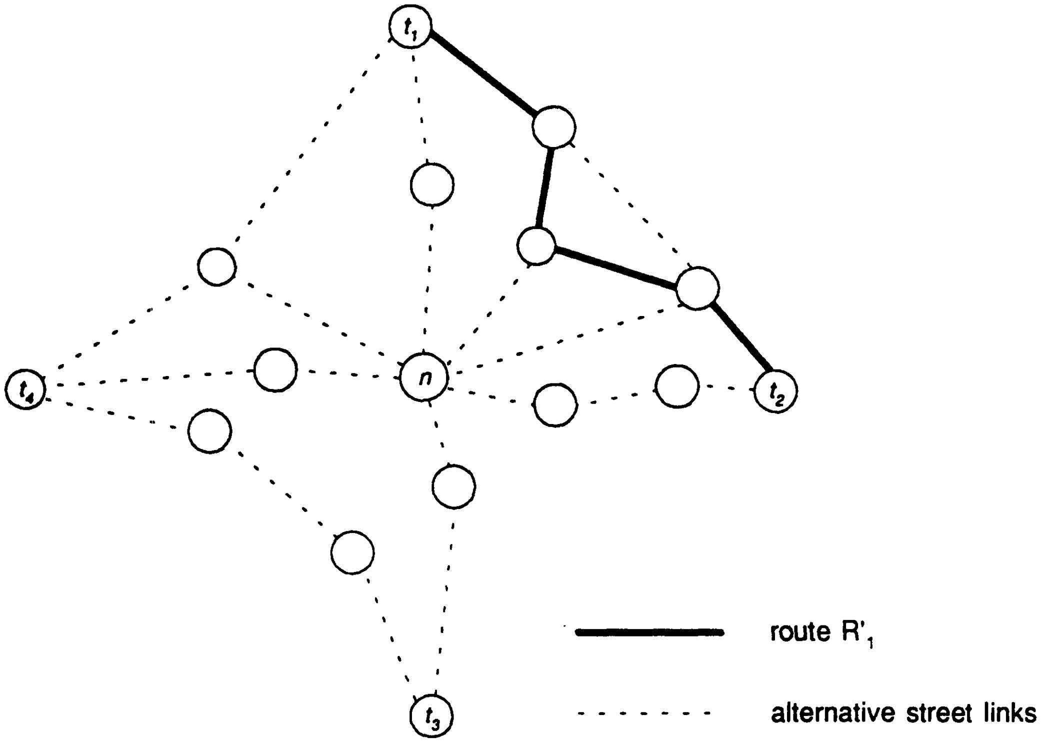


Figure 4.3a: Route R'_1 Created by Base Path Approach

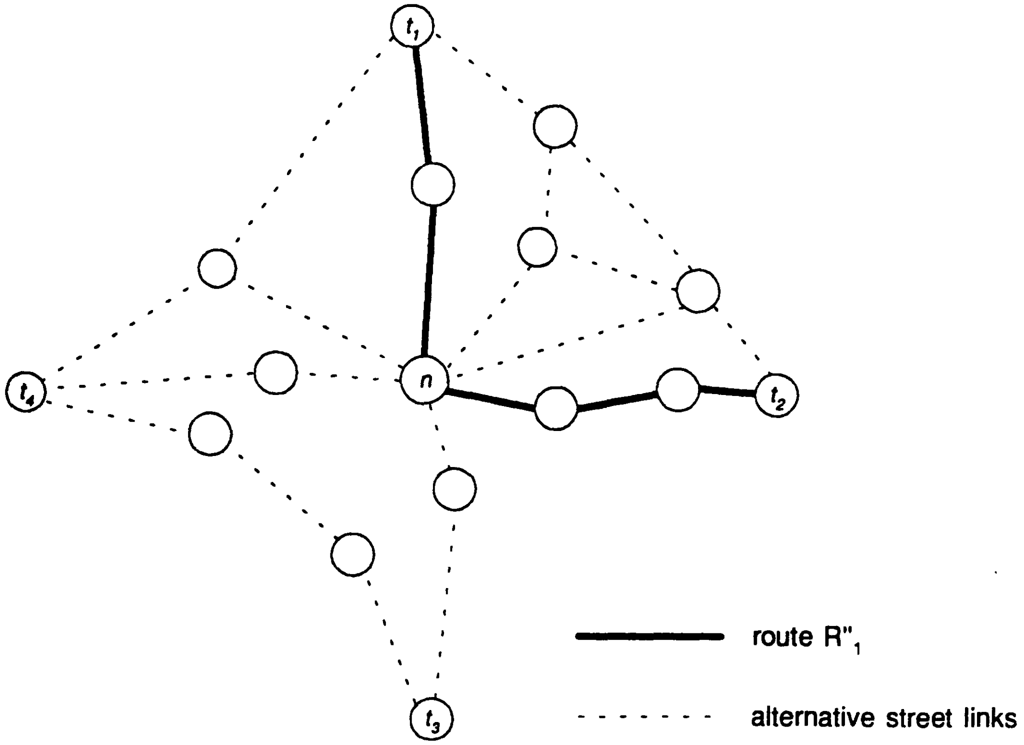


Figure 4.3b: Route R''_1 Created by Skeleton Approach

skeleton has a different number of trips associated with it. For the sake of comparison, suppose that the skeleton (t_1,n,t_2) is selected as having the largest number of trips. The final route R_1'' obtained after (t_1,n,t_2) is developed is shown in figure 4.3b and although it is less direct than R_1', R_1'' clearly serves more trips. In this example, any skeleton other than (t_1,n,t_2) would also yield a better route than R_1', since all final routes would have to go through the major destination node *n*.

The significance of this simple example is that it demonstrates a fundamental problem with the base path approach, namely its limitation on the number of routing options for a given set of terminal nodes. The limitation may cause the neglect of better routing options and, as in the last example, could lead to the formation of routes such as R_1', with no major destination to serve and with a demand that may not warrant its existence in the first place. The impact of this problem extends to the vehicle allocation stage of the network design, in which route R_1' will require a certain (minimum) number of buses that would have otherwise been put to better use on the other routes.

Having established the advantage of the initial skeleton approach in creating route nuclei, the second part of route generation which expands the route nuclei into complete routes is now addressed. An issue that is related to the initial skeletons and that was not discussed in the previous section is the necessity to limit the number of skeleton route combinations. In fact, if no restrictions are placed on the choice of the intermediate node(s), the initial skeleton approach may result in excessively long or circuitous routes after expansion, because no explicit measures to balance the route's directness with the amount of trips serviced are imposed. Long or circuitous routes are undesirable because they penalize the passenger's travel time, and are also difficult to operate because they might create service reliability problems.

For these reasons, the initial skeleton approach should include provisions to ensure that these skeletons are operationally feasible *before* they are considered for expansion. Referring back to the last example, the feasibility of the initial skeleton (t_1,n,t_2) would have to be checked during the formation of skeleton combinations and, if it is within allowable limits of length and circuity, it is allowed to be a candidate skeleton for expansion. (t_1,*n*,t_3) and (t_2,*n*,t_4) both have a circuity ratio of unity and are thus guaranteed to be feasible.

Because more than one alternative of expanding a given initial skeleton may exist, the route length and circuity should also be addressed *during* expansion. Consider the case of figure 4.4 where the skeleton route consists of the terminal nodes *r* and *s* and the intermediate destination node *d*. In this example, the expansion of the route segments [*r*,*d*] and [*d*,*s*] may be performed in two ways each, resulting in different round trip times and circuity levels of the final route. If no restrictions on the process of expansion of the initial skeleton (*r*,*d*,*s*) are imposed, the final route may follow path P - which is the longest and most circuitous possible path - if the benefits from generating such a path more than offset the disbenefits from the additional travel time incurred.

To overcome these problems, the proposed RG process starts with the creation of *feasible* initial skeletons whose corresponding estimates of route length and circuity satisfy maximum allowable values. Furthermore, after it is selected, an initial skeleton is expanded in two stages. In the first stage, the best feasible nodes are inserted in the skeleton until a connected path between the route's terminals is found. To ensure a relatively direct and short path, the same feasibility conditions imposed on the initial skeletons can be evaluated before each insertion, allowing only the feasible nodes which, if inserted, would not result in the skeleton route violating requirements of length and circuity.

The connected path obtained at the end of the first stage is referred to as the base route and is analogous to the base path created in the other approach, except that it is not necessarily the shortest one, since it has to go through pre-specified intermediate nodes. In the second stage, minor detours from the base route are considered incrementally, whereby neighboring feasible nodes that improve the route's share of total demand without excessively penalizing passengers with increased travel time are inserted. The merit of the two-stage approach is that it initially creates a feasible base route which satisfies the operational constraints, leaving the door open to detours in a later stage. Although it may not necessarily capture all the demand it can potentially serve, this route guarantees a basic service that might be extended at the second stage of route expansion.

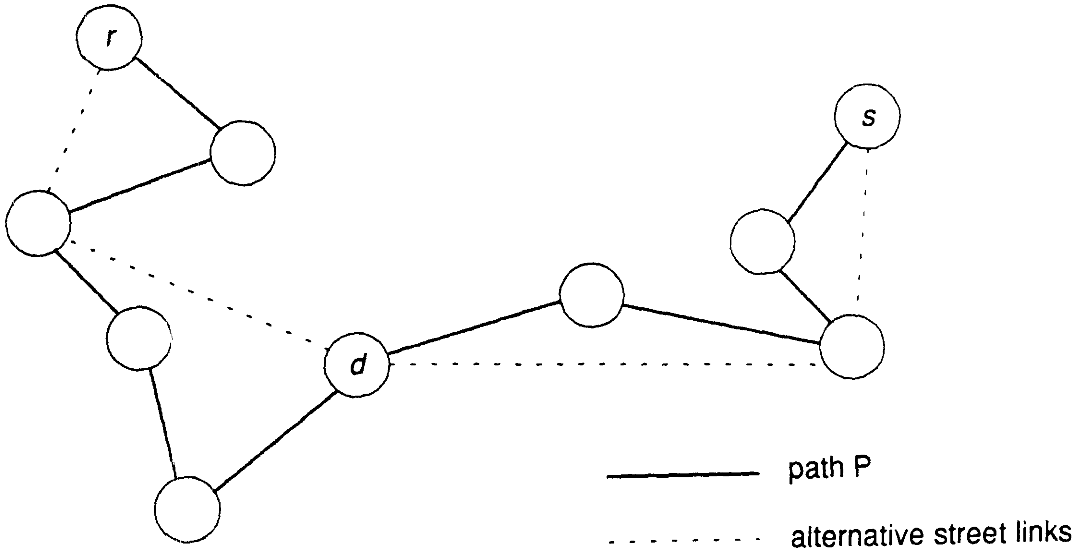


Figure 4.4: Skeleton Route (*r*,*d*,*s*)

On the other hand, a single-stage expansion approach might consider many detours at an early stage of the expansion process, precluding its acceptance as an operationally feasible fully developed route.

Based on all the discussion above, a route generation process which uses an initial skeleton approach is proposed. Furthermore, it is suggested that the expansion of the initial skeletons be performed in two stages, with a base route obtained at the end of the first stage and possible detours from the base route examined in the second stage. The RG process is thus composed of four main components: (1) the determination of the set of feasible initial route skeletons, (2) the selection of an initial skeleton, (3) the expansion of the initial skeleton into a base route, and (4) the consideration of route deviations. Moreover, a procedure for transforming the demand matrix is required in the case of a transit center network concept, in order for the methodology to yield effective results. After a description of the input required by the RG process, each of the four components is described in more detail. The demand transformation procedure is described in section 4.8.1.

4.3 Input and Design Parameters

The input required by the RG process can be grouped into the following categories:

Network:	The street network nodes (with terminal nodes identified separately) and links, the node connectivity information, and the shortest travel time matrix.
Demand:	A symmetric bus demand matrix.
Parameters:	The maximum allowable route round trip time RT_{max}, the maximum allowable route circuity RC_{max}, and the weights w_d, w_l and w_n for the components of the objective function.

The network information represents the basic elements of the street network on which buses can operate. Networks are usually coded from street maps into an abstract, computer-usable format consisting of nodes and links. Links are used to designate the individual road segments (streets, roads, freeways, etc.), whereas nodes represent the intersection of these segments and may also be used as points to shape the topology of the highway system. In the case where it is used to perform demand analysis, the network also contains special nodes called centroids, in addition to regular nodes, that represent the traffic analysis zones considered in the study.

For the purpose of bus network design, a less detailed network than the one used for demand analysis is recommended. Highway networks used in demand analysis serve to model travel by auto, which generally has a larger and more dense access area than transit, thereby requiring a network with a high level of detail. Besides, the proposed network design methodology is intended to be a screening tool that generates preliminary bus routes and frequencies that will require further refinements in terms of the exact alignment and the amount of service provided; thus, no additional benefits are gained by using a highly detailed network. Finally, the execution time of the RG process is highly sensitive to the size of the network used because it includes several operations that are proportional to the number of nodes and links in the network. For all these reasons, it is suggested that the size of the network used for route generation be kept to a minimum, keeping all valuable information such as the major bus transit corridors, destinations and access points to other modes in the network.

To establish the connectivity of the network, a link identification number is used to designate the nodes at the ends of each link. At the link level, the only attribute required is the estimated travel time, which is used to calculate the minimum travel time matrix. Bus demand, expressed as the number of trips per hour, is stored at the node level and is associated with the nodes that replace zone centroids.

The parameters used in the RG process can be classified into two types. The first type consists of parameters pertaining to the service pattern and service levels of the system and may be inferred from service planning guidelines. These guidelines are based on the experience and professional judgment of transit planners rather than on theoretical considerations. NCHRP Report 69 (1980) suggests several criteria for network design, and those specifically related to route generation are shown in table 4.1. Route length, for

1.3 Route Directness/ Simplicity

a. Routes should be direct and avoid circuitous routings. Routes should not be more than 20 percent longer in distance than comparative trips by car.

b. Route deviation shall not exceed 8 minutes per round trip, based on at least 10 customers per round trip.

c. Generally, there should not be more than two branches per round trip.

1.4 Route Length

a. Routes should be as short as possible to serve their markets; excessively long routes should be avoided. Route length generally should not exceed 25 miles round trip or 2 hours.

b. Two routes with a common terminal may become a through route if they have more than 20 percent transfers and similar service requirements, subject to (a).

Table 4.1: Service Planning Guidelines, Selected from NCHRP, Report 69, 1980

example, is suggested to be less that 25 miles or 2 hours round trip (1.4a) and is easily incorporated in the design process (sections 4.4, 4.6.1 and 4.7) with the use of a maximum allowable round trip time parameter RT_{max}. The directness criteria, however, are more difficult to incorporate as stated in table 4.1, because they are either too descriptive (1.3c), or they make reference to information that route generation is not concerned with (1.3a and 1.3b). In that case, an interpretation of the guidelines into a more explicit measure that would fit within the structure of the RG process is necessary. Table 4.2 shows two parameters used in the RG process - the maximum allowable round trip time RT_{max} and the maximum allowable route circuity RC_{max}. A route's circuity in this context is defined as the ratio of its one-way trip length (in time units) to the minimum travel time between its termini. Because no comparable measure is provided in the practical service guidelines,

Design Parameter	Use	Value
Maximum allowable round trip time, RT_{max}	To limit the round trip time estimate on a skeleton route	2 hours (NCHRP, Report 69)
Maximum allowable route circuity, RC_{max}	To limit the circuity of a skeleton route, defined as the ratio of its one-way trip length (in time units) to the minimum travel time between its termini.	1.0 - 1.5

Table 4.2: Design Parameters for the SRG

the value of RC_{max} is specified on a judgmental basis and should be validated after the experimental stage is conducted.

The other parameters used are specific to the proposed RG process and, consequently, are entirely empirical in nature. These parameters consist of the weights w_d, w_l and w_n used in the objective function to combine three measures of the passenger's level of service into a single measure (section 4.6.2). These weights depend on the magnitude of the measures they are associated with and are calibrated accordingly using a heuristic procedure described in appendix A.

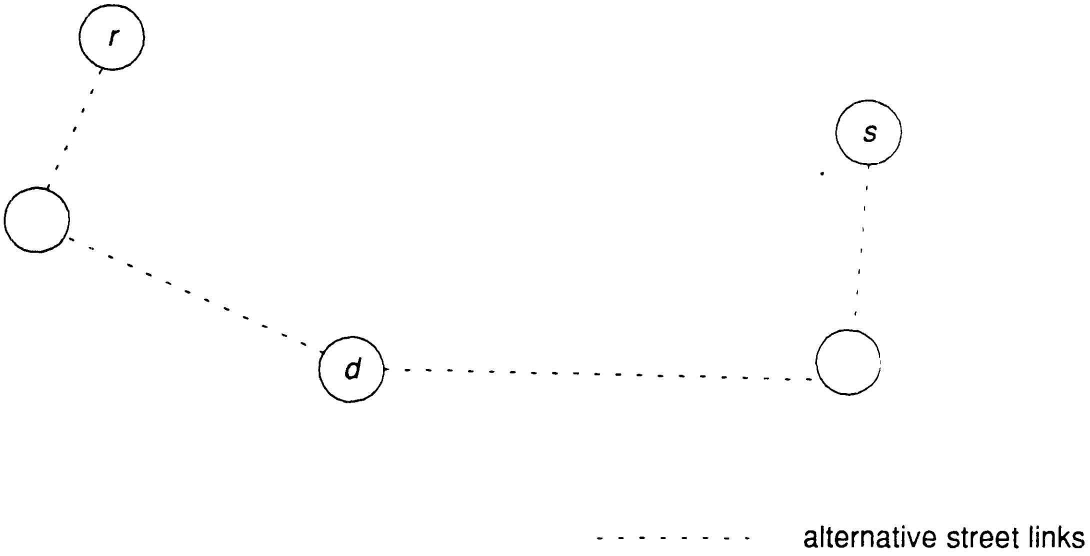
4.4 Determination of Feasible Initial Skeleton Routes

An initial skeleton consists of a sequence of three nodes: the first and last nodes are the termini of the route under development, and the intermediate node is a major generator in the service area. In general, the trajectory of the final route may be known with more certainty if the number of intermediate nodes in the initial skeleton is increased, and thus, a more accurate evaluation of the potential of the skeletons may be achieved. However, additional information drawn from adding one more intermediate node is expected to be small. In addition, the number of initial skeleton node combinations grows rapidly with the increase in the number of intermediate nodes, requiring additional comparisons to select the best initial skeleton. Therefore, only one intermediate node is used. If the demand matrix reflects a radial travel pattern, a single node would be sufficient to influence the formation of skeletons that go through a downtown location (if the deviation is feasible). If a transit center network is to be designed, a single intermediate node would also be sufficient, since the demand matrix would be transformed, prior to the determination of feasible skeletons, in a way to increase the chance of a transit center being chosen as an intermediate node.

Each pair of terminal nodes is matched with every major node to form all possible three-node skeletons. The set of feasible initial skeletons is determined by imposing constraints on the round trip time and circuity of the route corresponding to each skeleton combination. As shown in figure 4.5, a lower bound l_{min} on the round trip time of a route between r and s can be calculated by summing the shortest path travel times $SP_{r,d}$ and $SP_{d,s}$ between the terminal nodes and the intermediate node on its corresponding skeleton (r,d,s). Similarly, the circuity of that route (defined in the previous section) is bounded from below by the circuity c_{min} of its corresponding route skeleton, calculated as the ratio of the sum of $SP_{r,d}$ and $SP_{d,s}$ to $SP_{r,s}$. Therefore, skeletons with values of l_{min} or c_{min} exceeding RT_{max} or RC_{max} respectively are excluded from the set of feasible initial skeletons.

4.5 Selection of the Best Initial Skeleton

For each feasible initial skeleton, the total number of trips that will be served by the route after its full development is approximated by the sum of trips made between each pair of nodes in the skeleton. This approximation is obviously a lower bound, since it ignores additional trips between nodes that have not yet been inserted, as well as any transferring trips from other routes. However, at this early stage of development, this estimate is the only possible indication of the potential trips which may be served by the route, and is used at the beginning of each iteration of the SRG process to select the skeleton with the highest number of trips.



$$l_{min} = SP_{r,d} + SP_{d,s}$$

$$c_{min} = (SP_{r,d} + SP_{d,s}) \,/\, SP_{r,s}$$

Figure 4.5: Minimum Round Trip Time and Circuity of Skeleton (*r*,*d*,*s*)

After a route is fully developed, the demand associated with every initial feasible skeleton is recalculated, ignoring *all* the trips that are satisfied by the set of *all* routes already generated. The selection of the best initial skeleton for the next route to be developed is then made based on the modified values of demand. This calculation is performed for two reasons. First, it prevents previously selected skeletons from being selected in subsequent iterations of the SRG process. Although this could also be prevented by simply discarding these skeletons from the set of feasible initial skeletons, the latter is likely to contain others which overlap with these skeletons. If any one of these overlapping skeletons is then selected at a later stage of RG, the resulting full route is likely to be overlapping with other routes in the network. This leads to the second reason for modifying the demand associated with each feasible skeleton. By doing so, feasible skeletons would be selected for their potential of serving trips that are *not* already (directly) served by the current network. Consequently, the degree of route overlap in the final network is reduced, and its demand coverage is improved.

4.6 Expansion of a Skeleton into a Base Route

The expansion of an initial skeleton into a base route is the first stage of a two-level expansion process required to complete route development. The expansion into a base route is performed using an iterative process of locating "gaps" between nodes already in their position in the skeleton route and filling them with the "best" feasible insertion nodes. A gap is defined as two consecutive nodes i and $i+1$ in a skeleton, where i and $i+1$ are not connected by a single link (non-adjacent). Additional insertions between adjacent nodes in the skeleton are considered in the second stage of expansion (section 4.7), after the base route has been obtained.

To understand how this stage of expansion works, consider the initial skeleton route shown in figure 4.6a which consists of the terminal nodes r and s and intermediate node d. The first gap is located between nodes r and d, which are the first two consecutive and non-adjacent nodes in the skeleton. The first iteration of the expansion process attempts to close the gap between r and d by inserting a feasible neighboring node

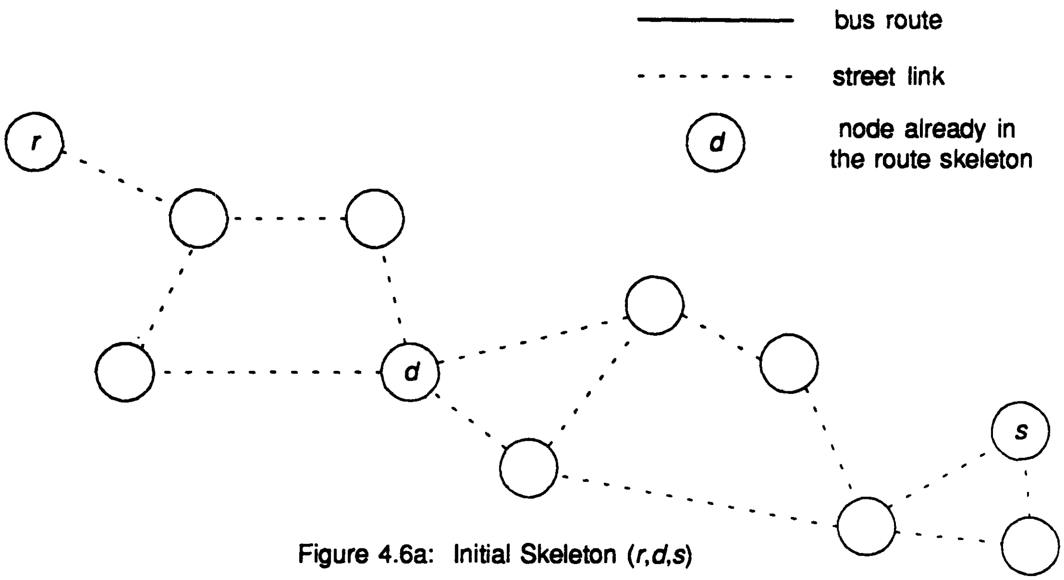


Figure 4.6a: Initial Skeleton (*r*,*d*,*s*)

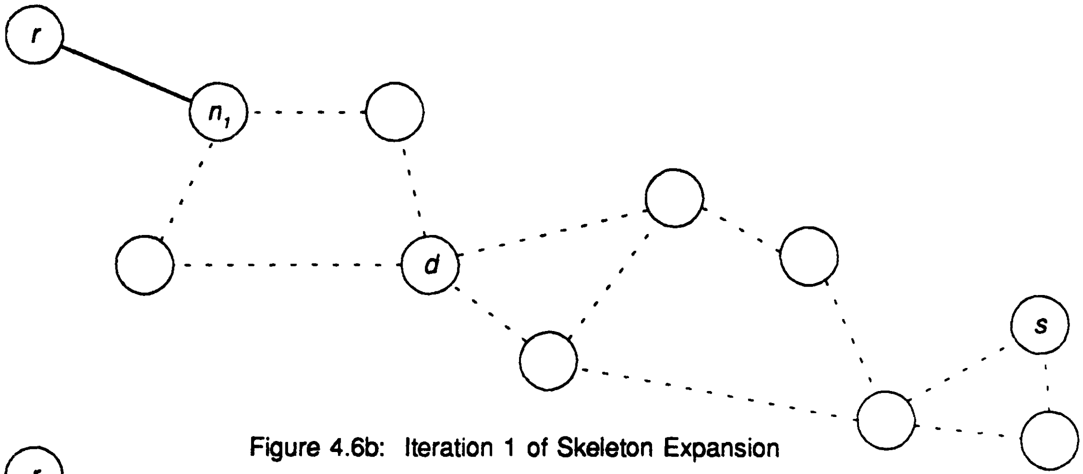


Figure 4.6b: Iteration 1 of Skeleton Expansion

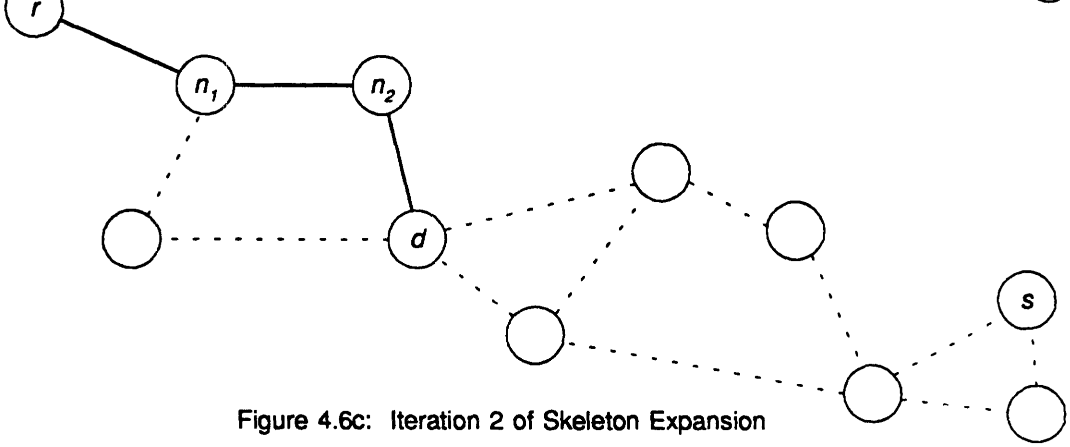


Figure 4.6c: Iteration 2 of Skeleton Expansion

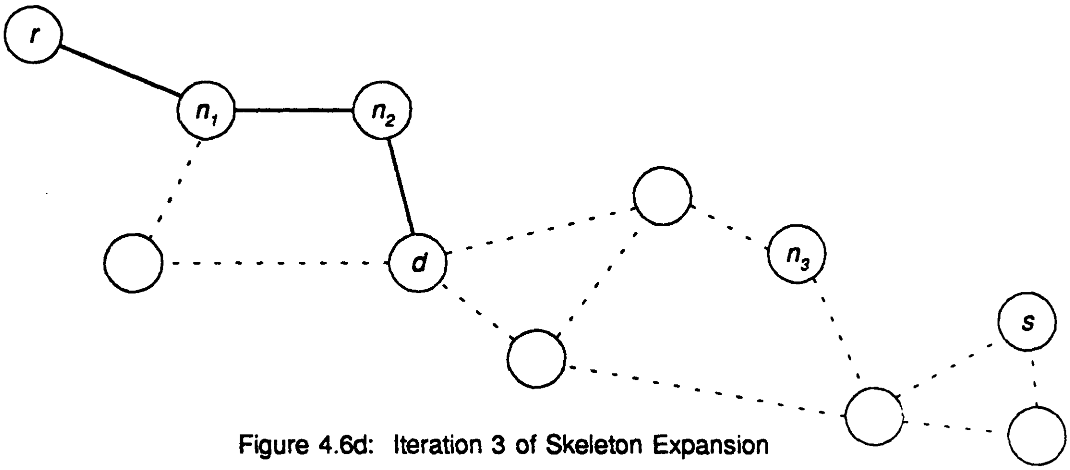


Figure 4.6d: Iteration 3 of Skeleton Expansion

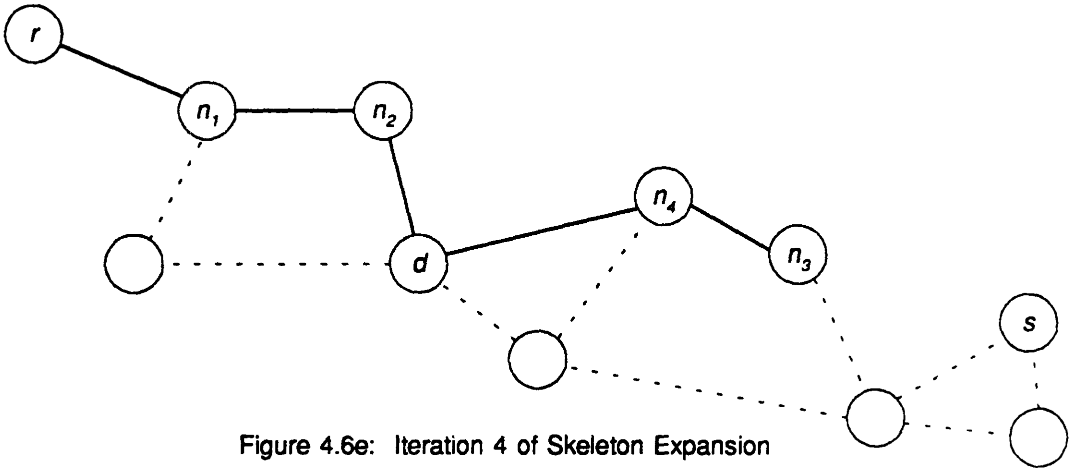


Figure 4.6e: Iteration 4 of Skeleton Expansion

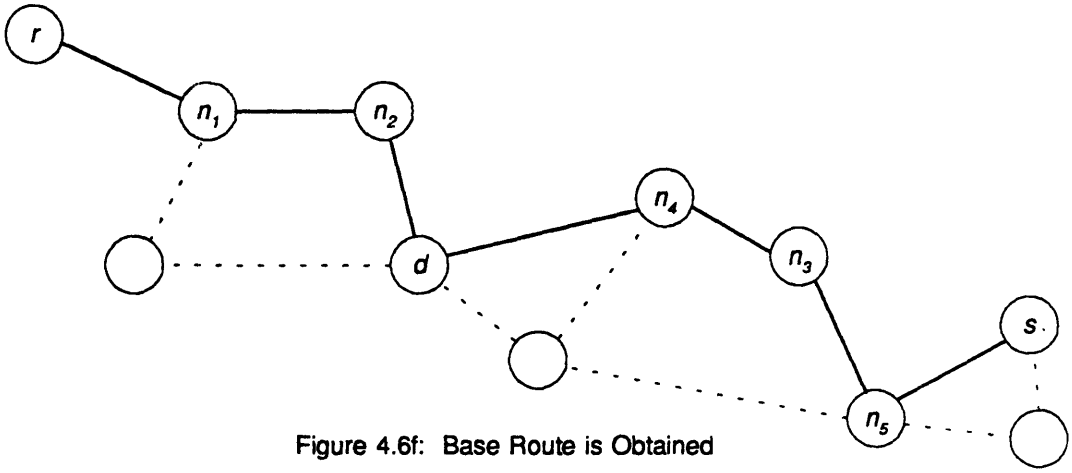


Figure 4.6f: Base Route is Obtained

in between. All the feasible nodes are evaluated and the one which maximizes an objective function is inserted. This situation is illustrated in figure 4.6b in which the best feasible node n_1 is inserted between *r* and *d*. The next gap in the skeleton is now located between nodes n_1 and *d* and the search for feasible insertion nodes for that gap is performed. In figure 4.6c, n_2 is the only feasible insertion node for that gap and is therefore automatically added to the skeleton. The expansion process is continued in the same manner (figure 4.6d through figure 4.6f) until all the gaps in the skeleton are closed or until a gap with no feasible insertion nodes is encountered. In the latter case, the current route is abandoned and a new iteration of the SRG process is initiated with a different skeleton. If all the gaps could be filled, a connected path such as $r\text{-}n_1\text{-}n_2\text{-}d\text{-}n_4\text{-}n_3\text{-}n_5\text{-}s$ of figure 4.6f is obtained. This path is the base route which is examined for deviations in the second stage of route expansion.

4.6.1 Feasible Insertion Nodes

The purpose of creating a set of feasible insertion nodes prior to the selection of the best node to insert in a certain gap is to reduce the number of nodes that need to be evaluated. Given a gap between nodes *i* and *i*+1 in the skeleton route (*r*,*d*,*s*), the set of feasible insertion nodes **sfn** is calculated relative to nodes *i* and *i*+1 and contains all nodes *n* which satisfy both the following conditions:

1. If l_{min} (the lower bound on the round trip time for the route between *r* and *s*), estimated from the skeleton after the insertion of *n* in between nodes *i* and *i*+1, is less than RT_{max} then include node *n* in **sfn**.
2. If c_{min} (the lower bound on the circuity of the route between *r* and *s*), estimated from the skeleton after the insertion of *n* in between nodes *i* and *i*+1, is less than RC_{max} then include node *n* in **sfn**.

Note that these conditions are the same as the initial skeleton feasibility criteria presented in section 4.4 and that values of l_{min} and c_{min} are calculated similarly. The only difference is that the path adopted by the route under expansion is more certain at this stage (with more nodes and links in their final position on the route) than in the initial skeleton stage. Therefore, the calculation of l_{min} and c_{min} produces more accurate estimates of the final route's length and circuity.

4.6.2 Selection of the Best Feasible Node

In order to determine the "best" node for insertion in a certain gap among all the feasible nodes, an objective function which combines several measures of the passenger's level of service is utilized. For each node *n* belonging to **sfn**, the objective function *objctv(n)* is evaluated and the node which maximizes the objective function is selected for insertion. The objective function is expressed as:

$$objctv(n) = w_d \cdot dem(n) - w_l \cdot dev(n) + w_n \cdot rts(n)$$

where *dem(n)* = the additional number of currently *unserved* trips between node *n* and each of the nodes already in the route,

dev(n) = the additional deviation from the shortest path (in passenger-minutes) imposed by the insertion of node *n* in the route,

rts(n) = the number of already generated routes passing through node *n*, and

w_d, w_l and w_n are the weights of the three components.

The first component, *dem(n)*, reflects the goal of increasing network directness by satisfying the largest demand possible without transfers, as well as the objective of having high ridership routes in the final design. In order to reduce route overlap in the final network, the demand between node *n* and node *m* which belongs to the route under expansion is added to *dem(n)* only if *n* and *m* are not both included in a route that has already been developed. The second component, *dev(n)*, is obtained by summing the additional travel time incurred over all the passengers affected by the diversion of the route from its path due to the insertion of *n*. It is used to favor nodes that would result in reasonably direct routes. The last component, *rts(n)*, gives preference to nodes that have already been included in other routes, thereby achieving a well-connected final route network with a high level of demand coverage.

The objective function is not based on any theoretical considerations and is simply an attempt to combine the above three measures and to address the tradeoffs among them in a quantitative manner. The weights w_d, w_l and w_n may also be used to emphasize one or other of the components. The weights are determined empirically and a procedure for calculating their numerical values is presented in appendix A.

4.6.3 Backtracking

During the process of skeleton route expansion, it is possible that the route does a U-turn and retraces part of its path. This undesirable situation is called backtracking and is caused by the situation where the best insertion node is already on the skeleton. Consider the situation in figure 4.7a which shows a gap between nodes *a* and *c* in a skeleton route (*a*,*c*,*e*) under expansion. In searching for a node to insert between *a* and *c*, it may well be that node *d* is the best feasible insertion node according to the value of the objective function. After *d* is inserted (figure 4.7b), the next gap is between *a* and *d* and the only feasible insertion node for that gap is node *c*. Because node *c* already exists in the skeleton, inserting it again between *a* and *d* would cause the skeleton to become *a-c-d-c-e* and the final route to trace segment [*c*,*d*] at least twice. Although this situation cannot be easily anticipated and avoided, backtracking is detected after it has occurred and the node responsible for it is removed from the skeleton route and is prevented from being added back ***in the same gap***. The node to be removed is selected between the two nodes that specify the current gap as the one that was last inserted in the route. In this case, node *d* is removed (being the latter of *a* and *d*) and is not allowed to be a feasible insertion node in the gap between *a* and *c* (figure 4.7c). In the next iteration, node *b* is selected instead as the best node to insert in that gap (figure 4.7d) and the final expansion path would be *a-b-c-d-e*.

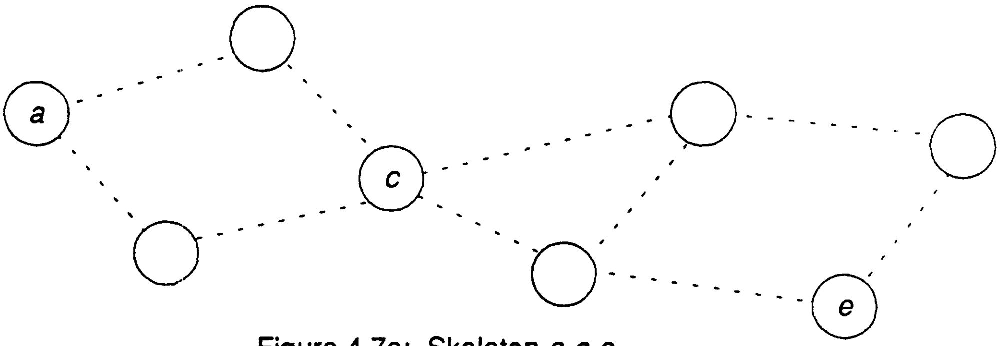


Figure 4.7a: Skeleton *a-c-e*

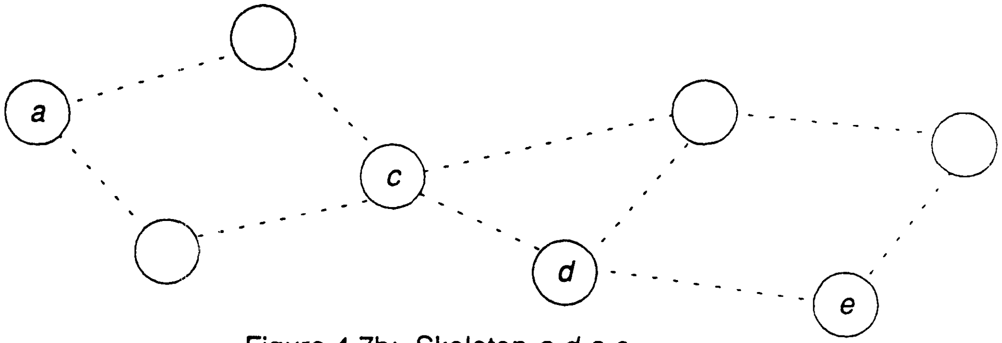


Figure 4.7b: Skeleton *a-d-c-e*

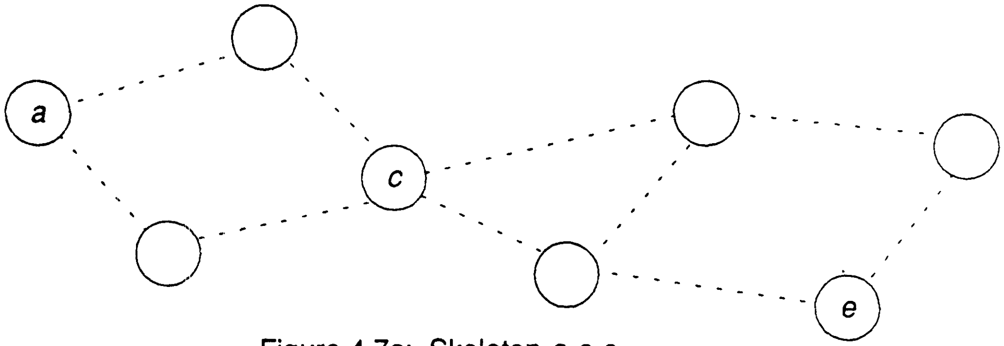


Figure 4.7c: Skeleton *a-c-e*

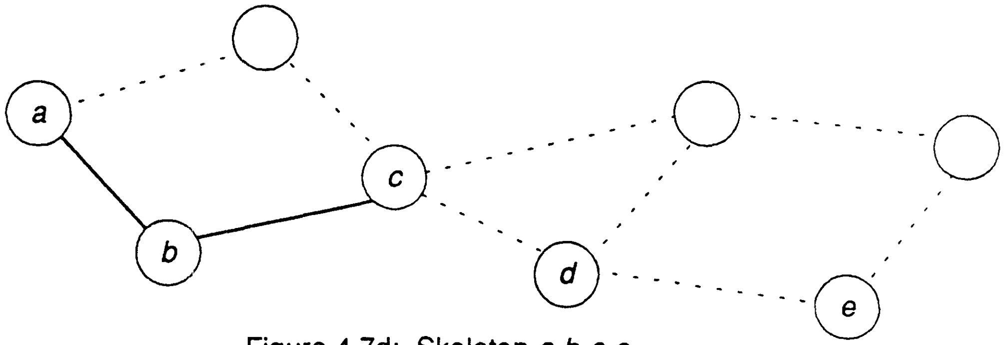


Figure 4.7d: Skeleton *a-b-c-e*

4.7 Deviations from the Base Route

After a base route is found, the second stage of route expansion is initiated by considering additional insertions that would cause deviations from the base route. Specifically, the SRG process examines feasible insertion nodes that are one-link away from two *adjacent* nodes on the base route, and that, if inserted, would significantly increase the direct demand satisfied by the route with only a relatively small increase in travel time. Insertion between adjacent nodes in the base route is intended to complement the first-stage insertion in gaps and to allow the possibility of rerouting a segment of a route by replacing it with another one parallel to it. For each pair of consecutive nodes i and $i+1$ on the base route, feasible insertion nodes are identified relative to i and $i+1$. The feasibility criteria is similar to the one used in the skeleton expansion process (section 4.6.1), except that a more restrictive route circuity condition is used in order to prevent large deviations of the route from its current path. Specifically, a "local" circuity parameter $c_{i,i+1}$ is calculated and compared to the maximum allowable circuity ratio RC_{max}. $c_{i,i+1}$ is determined relative to two consecutive nodes i and $i+1$ on the route under consideration, and measures the deviation between these two nodes due to the insertion of node n. If the local deviation is permissible ($c_{i,i+1} < RC_{max}$) and the round trip time of the route after node n is inserted is less than the maximum allowable, n is considered feasible.

Because nodes only one-link away are considered for insertion, it is unlikely that more than one feasible node will be found. In order to decide whether a feasible node n is worth inserting or not, the same objective function is evaluated for n. A positive value of the objective function indicates that the effects of additional demand directly satisfied by the modified route and the increased opportunity to serve additional destinations through transfer to other routes do indeed offset the increase in passenger-minutes imposed by the detour. If the insertion of node n is deemed to be beneficial, the deviation involving node n is accepted. In the unlikely event that more than one feasible insertion node exists and all of them yield a positive value of the objective function, the one which maximizes the objective function is selected. Finally, the process of additional insertions is terminated when the last two consecutive nodes on the route have been examined.

In order to illustrate the process of additional insertion (stage two) and to compare it with the skeleton expansion process (stage one), the example from section 4.6 is used. Figure 4.8a shows the base route r-n_2-n_3-n_1-n_5-n_4-n_6-s obtained after the skeleton expansion process is terminated. In this example, the only case where an insertion of a node one-link away is possible is between n_1 and n_5. If n_7 is feasible, and if the objective function evaluated at n_7 is positive, then n_7 is inserted between n_1 and n_5, yielding the modified route r-n_2-n_3-n_1-n_7-n_5-n_4-n_6-s shown in figure 4.8b. To appreciate the advantage of a two-stage skeleton expansion approach, note that n_5 was feasible for insertion between n_1 and n_4 during the skeleton expansion (otherwise it would not have become part of the base route) and that node n_7 is very likely to have also been feasible for insertion at that stage, since it satisfies the feasibility condition between n_1 and n_5. Because the objective function value of n_5 exceeded that of n_7, the former was selected to fill the gap between nodes n_1 and n_4, deferring the insertion of n_7 till after securing a feasible base route. If, on the other hand, both n_5 and n_7 were inserted between n_1 and n_5 at the stage of skeleton expansion, the chances of rejecting the base route would have been increased because of the additional length and circuity introduced by n_7.

It is worth mentioning that the process of additional insertions ignores the case of nodes that are farther away than one link from two consecutive nodes on the base route. Figure 4.9 shows an example where nodes m and n would not be considered for insertion between i and j. However, it is assumed that in such cases, the chances of finding a deviation such as (m,n) that is feasible are quite small. Therefore, neglecting deviations more than one-link away may be justified, especially if the computational effort required to detect such cases is also taken into account. This is not to say that every deviation from the base path consists necessarily of only one node. The example in figures 4.10a, 4.10b and 4.10c shows the case where node m, which is one-link away from i and j, is first inserted in to the route and then node n, which is one-link away from m and j, is also inserted in the route. The result is that segment $[i,j]$ has been replaced by the parallel segment $[m,n]$.

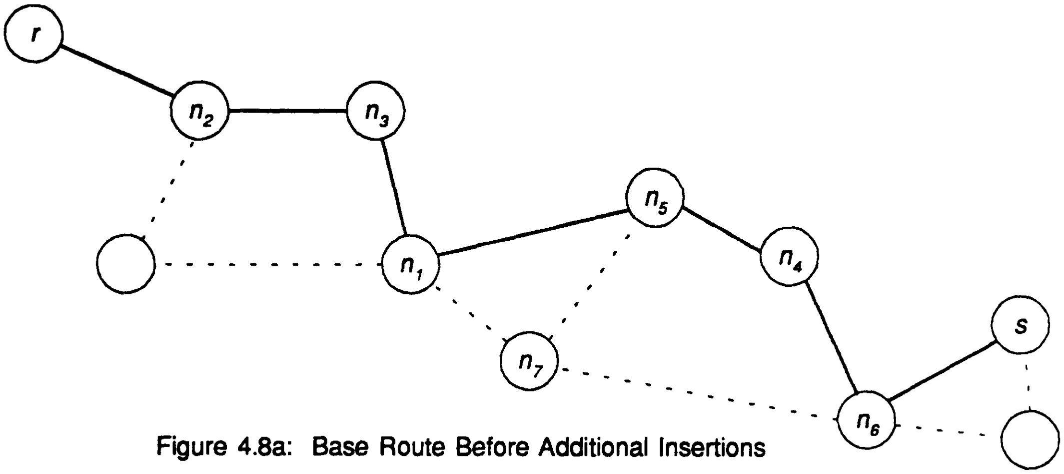


Figure 4.8a: Base Route Before Additional Insertions

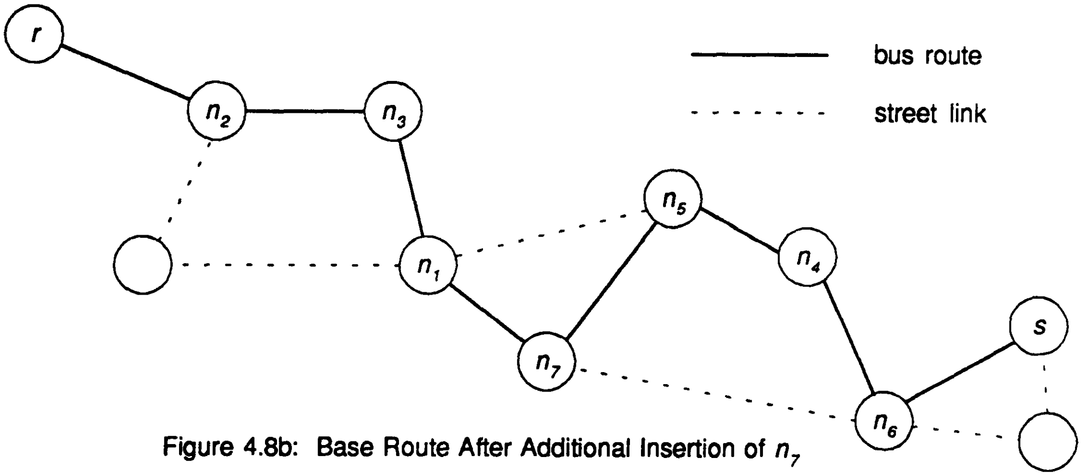


Figure 4.8b: Base Route After Additional Insertion of n_7

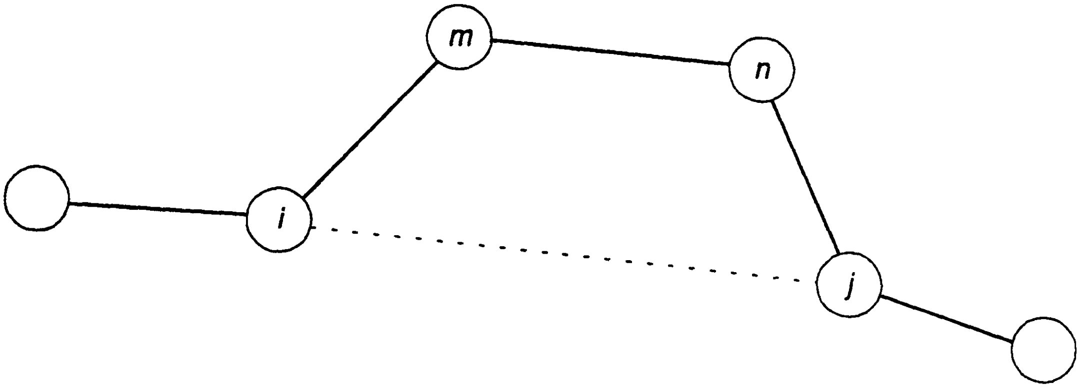


Figure 4.9: Two-Node Deviation from Base Route

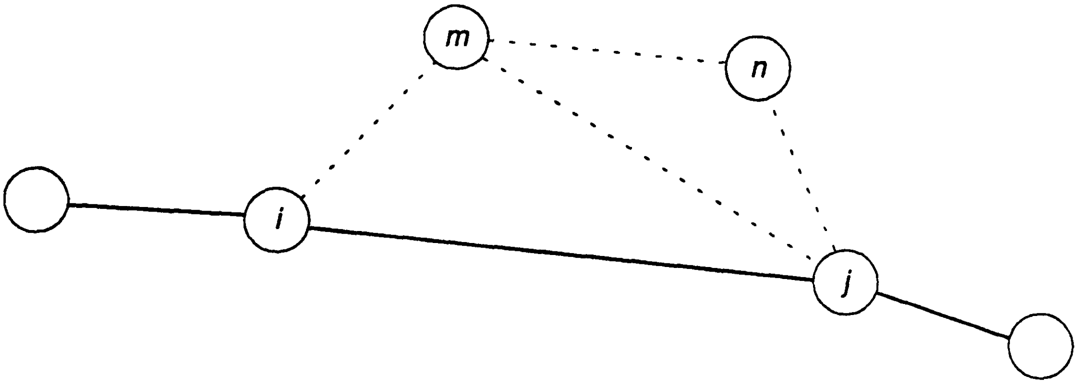


Figure 4.10a: Segment (*i,j*) of Base Route

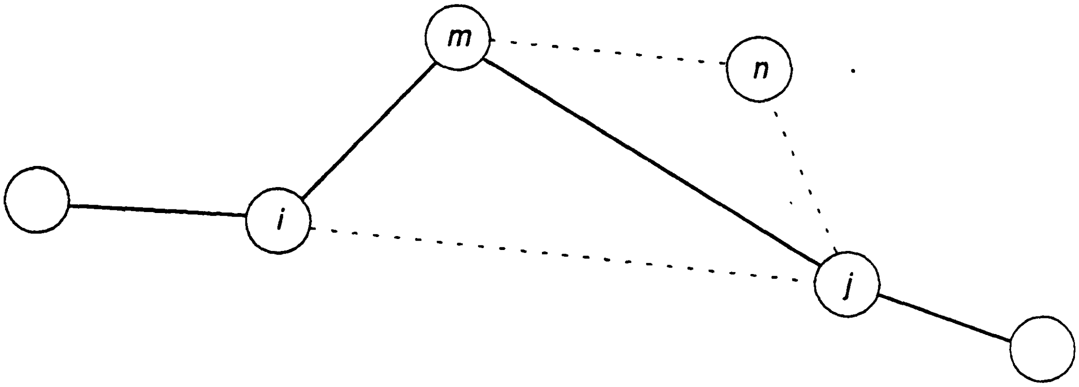


Figure 4.10b: Segment (*i,j*) after First Iteration of Additional Insertions

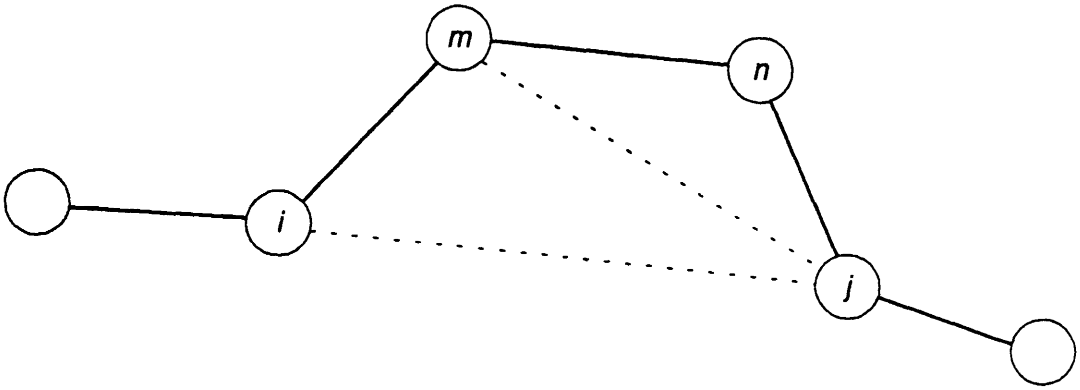


Figure 4.10c: Segment (*i,j*) after Second Iteration of Additional Insertions

4.8 Route Generation with the Transit Center Concept

The process of route generation in the case of a transit center network follows the same basic structure used for general network design presented in the previous sections, except for several modifications that are necessitated by the difference in characteristics among the various route types involved in the transit center concept. As described in the previous chapter, transit centers are served by two types of routes - trunks and feeders. Trunk routes connect a transit center to one or more other transit centers along a direct path and are characterized by high frequency service. Feeder routes operate in a mini-radial pattern with service oriented towards the transit center. These routes operate at a lower frequency and serve a collection-distribution function.

The proposed methodology for bus network design allows the design of a transit center network which requires the specification of a set of transit center nodes in addition to the input used in the general network design. As a screening tool, the proposed methodology can therefore assist transit properties in performing a preliminary analysis of the alternative network scenario without requiring an extensive effort or cost. By performing simple manipulations on the matrix of existing bus demand, the RG process is capable of appraising, at least in a preliminary manner, the redistribution of the trips in the transit center network and then to use the modified matrix as the starting point of the transit center design. This is not to say that the process is capable of completely assessing the change in bus demand or in trip-making patterns due to the new concept; such evaluation requires a full-scale demand analysis and is not within the capabilities of the proposed methodology. The transformations performed on the demand matrix serve only to influence the RG process into acknowledging that transit center nodes would be the major transfer nodes in the new network.

4.8.1 Demand Matrix Transformation

The main feature of a transit center network is that many trips made in the system go through one or more of the designated transit centers. If a demand matrix which corresponds to a radial bus network is used as the basis for the design of such a system, a large portion of the origin-destination flows that is satisfied directly by existing routes might be redirected via one or more transit center nodes in the new network. To illustrate, consider the example of two nodes *a* and *b* which are served by route R in the existing bus system. If the network is to be re-designed around the transit center concept and, if in that design, node *tc*, which is slightly off the route between *a* and *b*, is specified as a transit center, the direct flow $D(a,b)$ would be considered for reassignment via *tc*. This is not to say that the flow $D(a,b)$ would necessarily be reassigned, since certain requirements concerning the feasibility of the potential new route R_{tc} that would serve *a* and *b* via *tc* also have to be met. Even if the transformation is made, *a* and *b* would not necessarily belong to the same route R_{tc}, since one of them may be infeasible at a certain stage of route R_{tc}'s expansion. A flow that is served in the new system via two transit centers would be transformed similarly, except that it would also be assigned between the two transit centers involved. Flows that require more than two transit centers to be satisfied are not considered for transformation, since such trips would require more than two transfers. Routes which serve these flows directly (without going through any transit centers) may be created later if the flows are large enough to justify the direct connection.

The major question in the demand matrix transformation procedure is to determine which flows would be affected by the transit center system. Since no exact answer can be provided unless the network is completely redesigned, heuristic rules are applied to each flow to determine whether, or not, it is likely to be affected and to determine the necessary demand transformation. For a flow $D(o,d)$ between nodes *o* and *d* and a set of transit center nodes tc_1, tc_2, ... tc_m, the following heuristic rules are suggested:

Step 1. If the path $o\text{-}tc_1\text{-}tc_2\text{-}d$ is feasible, assign $D(o,d)$ to $D(o,tc_1)$, $D(tc_1,tc_2)$ and $D(tc_2,d)$, delete $D(o,d)$, and repeat step 1 for another flow. Otherwise, go to step 2.

Step 2. If the path $o\text{-}tc_k\text{-}d$ is feasible, assign $D(o,d)$ to $D(o,tc_k)$ and $D(tc_k,d)$, delete $D(o,d)$, and use step 1 with another flow. Otherwise, go to step 3.

Step 3. If all paths $o\text{-}tc_i\text{-}d$ are infeasible for all $i = 1 \dots m$, do not perform any transformation on $D(o,d)$. Select another flow and go to step 1.

For the sake of consistency, the same skeleton feasibility conditions are used to determine whether the paths involving the transit center nodes are feasible or not. Recalling from section 4.4, these conditions consist of meeting requirements on the maximum allowable length (round trip time) and circuity.

Transformations involving two transit center nodes are given priority over single transit center node transformations, since routes with two transit centers provide better connection and transfer opportunities to other routes in the system. Also, this transformation builds up the demand between the transit centers and promotes the generation of high frequency trunk routes between the centers. Flows that are reassigned via one transit center reinforce the creation of feeder routes connecting to these centers. Finally, flows which cannot be reassigned via any of the transit centers are not modified and may be served in the final design by direct routes, if a skeleton with a non-transit center intermediate node and sufficient demand is found.

It is important to note that there may be other - and probably better - heuristic procedures for transforming the demand matrix than the one proposed above. The main advantage of the proposed three-step transformation procedure is its simplicity and low computational cost relative to its effectiveness in emphasizing the role of transit centers prior to the generation of routes. From a network evaluation perspective, however, the transformation procedure may not achieve the best reassignment of all origin-destination flows to their appropriate transit centers. To understand the limitations of the proposed transformation procedure in that regard, consider figure 4.11a in which the flow $D(o,d)$ between node pair (o,d) is to be reassigned via one or two transit centers. Assume for simplicity that tc_1 and tc_2 are the only two transit centers in the vicinity of nodes o and d. Because of the position of d relative to nodes o, tc_1 and tc_2, the path $o\text{-}tc_1\text{-}tc_2\text{-}d$ is likely to be feasible. Consequently, $D(o,d)$ is reassigned to $D(o,tc_1)$, $D(tc_1,tc_2)$ and $D(tc_2,d)$ and

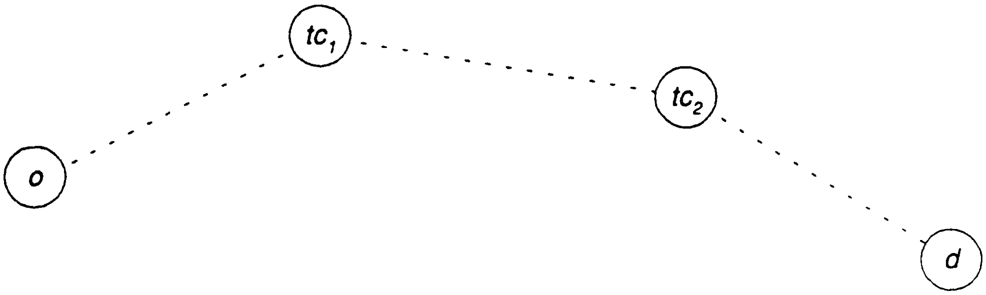


Figure 4.11a: Reassignment of Flow *D*(*o*,*d*) - Case 1

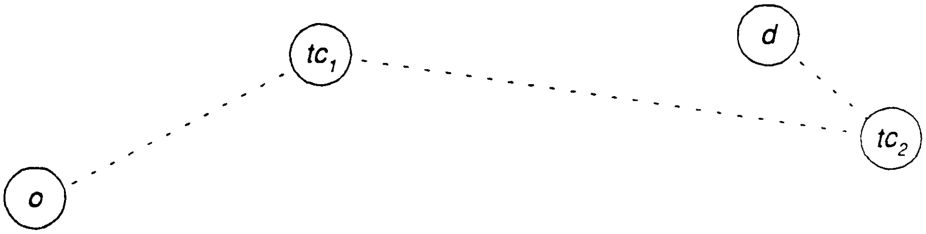


Figure 4.11b: Reassignment of Flow *D*(*o*,*d*) - Case 2

then deleted - according to step 1 of the proposed transformation procedure. In this case, this particular reassignment of $D(o,d)$ is the most desirable one for RG because it increases the demand at both nodes tc_1 and tc_2 and improves the chance of creating a trunk route between them. At the same time, the reassignment of $D(o,d)$ is also adequate from a passenger assignment perspective since the path o-tc_1-tc_2-d is relatively short and direct. (A slightly shorter in-vehicle travel time path between o and d involving only one transit center may exist, although this path would require a larger total travel time because it would not be served by a high frequency trunk route). On the other hand, if node d is positioned as shown in figure 4.11b, it is likely that the path o-tc_1-tc_2-d would also be feasible, resulting in the reassignment of $D(o,d)$ via tc_1 and tc_2 as in the previous case. Although this assignment is beneficial for RG, it is not the most appropriate one in terms of passenger assignment. Instead, a better reassignment of $D(o,d)$ in this case would be via tc_1 only which involves a much shorter path (o-tc_1-d) and eliminates the unnecessary transfer at tc_2.

This simple example demonstrates that the proposed demand matrix transformation procedure is not appropriate for route generation and network evaluation at the same time. Such a procedure would be more sophisticated than the one proposed in this thesis because it would have to compare the benefits (and disbenefits) of the various reassignment alternatives that may be possible for each flow in terms of the number of transfers involved and total travel time. As mentioned earlier, the proposed demand matrix transformation procedure is intended to influence RG in the case of a transit center design and, therefore, does not consider the effects of the transformation beyond the RG process. If the transformed demand matrix is utilized in the VA procedure, the inflated demand at the transit center nodes would cause higher frequencies on the routes containing them than what is actually required. Moreover, the total in-vehicle and transfer time calculated based on the transformed matrix would be significantly greater than their actual values because the proposed transformation procedure sometimes forces the reassignment of flows via more centers than is needed. For these reasons, the transformed demand matrix is used only in the RG procedure, and then replaced by the original (unmodified) demand matrix in the VA procedure for calculating the service frequencies, the number of buses required and the performance measures.

4.8.2 The Generation of Trunk Routes

In general, skeletons containing transit center nodes are expected to rank highest among all feasible skeletons because of the demand transformations which increase the demand for travel to the transit centers, reinforcing their role as the hubs in the network. In particular, skeletons with transit centers at both terminals have the highest demand satisfaction values because the matrix transformation procedure attempts to re-direct flows via two transit centers and, consequently, require high frequencies of service. These skeletons form the trunk routes which provide a direct, high-frequency service between the two transit centers. The feasibility criteria of trunk route skeletons uses a lower circuity ratio (typically 1.2) in order to generate faster and more direct routes between transit centers. Moreover, detours from the base path are not considered for trunk routes. This does not necessarily reduce the demand coverage of the final network, since a large portion of the demand captured by trunk routes is realized via transfers from feeder services rather than from the route itself.

4.8.3 The Generation of Feeder and Other Routes

Feeder routes are expanded from initial skeletons formed between a terminal node and one of the transit centers. These skeletons are also characterized by a large level of demand satisfaction and thus are selected at the early stages of RG. The expansion of feeder routes is performed as in the case of a general network design, including the consideration of additional insertions beyond the base route level.

4.8.4 Transit Center Nodes

In order to produce a network with the transit center concept, a set of transit centers have to be specified in the input. Guidelines for selecting transit center ***nodes within the design process*** were not considered, because it is believed that the choice of transit centers can only be performed by the planner, based on his or her knowledge of the major trip generators in the service area (such as retail and/or employment centers). In addition, the location of transit centers is often dictated by external factors such as the availability of space and access to the street network, which are obviously beyond the scope of the proposed methodology. However, the design procedure could be used to compare different sets of transit centers by modifying the specified transit center nodes in the input.

Chapter 5

Vehicle Allocation

This chapter describes the process of vehicle allocation (VA) within the proposed methodology of bus network design. Several approaches to vehicle allocation are considered, ranging from the simple rules of thumb used in the transit industry to the more sophisticated optimization-based methods. The proposed VA process was selected based on the character of the networks produced by the route generation (RG) process, mainly the fact that the routes in these networks are characterized by some degree of overlap. The method of VA adopted was originally proposed by Han and Wilson (1982) and is capable of analyzing networks with overlapping routes because it incorporates a passenger path choice model. The method aims at allocating the available fleet of buses to the network so as to (1) ensure enough capacity on each route to accommodate all the passengers who would select it, and (2) minimize the occupancy level at the most heavily loaded point on any route in the system. The heuristic solution to this method decomposes the problem into two components. The base vehicle allocation identifies the minimum number of buses required for a feasible solution, satisfying the capacity and other constraints. The surplus vehicle allocation assigns the remaining buses in order to improve the performance of the system in terms of waiting and transfer time.

Section 5.1 discusses several approaches to incorporating vehicle allocation within route generation. Methods of vehicle allocation are reviewed in section 5.2 and the following two sections describe in more detail two approaches for optimal vehicle allocation. Section 5.5 describes the characteristics of the networks created by the RG process and identifies the features sought in the VA process to address them. In sections 5.6 through 5.8, the VA process adopted is described in detail, and in section 5.9, the overall design methodology is discussed in light of the VA process. Two ways of integrating vehicle allocation and route generation are presented in section 5.10 and the next section uses a simple numerical example to illustrate their performance.

5.1 Vehicle Allocation within the Proposed Design Methodology

Vehicle allocation is the second major component after route generation in the proposed network design methodology and is used to assign the limited number of buses available to the network and to estimate the frequency of service on each of the routes. As explained in chapter 3, previous heuristic approaches to the network design problem have almost exclusively separated route generation from vehicle allocation by assigning the available buses after all possible routes, or a desirable number of routes, have been created. The negative implications of this approach were also discussed in that chapter and it was decided that the proposed methodology will address this weakness by investigating ways of integrating route generation and vehicle allocation so as to achieve consistency between the size of the network created and the resources available.

Vehicle allocation and route generation may be integrated at more than one level. The highest degree of integration is achieved by approach 1, shown in figure 5.1, in which the required number of buses is determined every time a new route is added to the network. This approach guarantees that the number of buses required by the final network will not exceed the available fleet size because the vehicle requirement is monitored throughout the network development. However this approach is computationally demanding, since the VA process is likely to require a large amount of calculations. Alternatively, figure 5.2 shows a second approach that might be used, whereby the VA process is called upon less frequently, only after a certain additional number *N* of routes have been created since the last vehicle allocation. In order to reduce the risk of the required buses exceeding the fleet size, this approach may be taken one step further to create approach 3 which includes a procedure to evaluate the current network, positioned within the route generation loop as shown in figure 5.3. This procedure would be used to estimate the total number of buses required after each route is generated, using a simple and computationally efficient method of estimation. When this estimate approaches the available fleet size, the VA procedure which is capable of performing a more accurate and elaborate allocation, is initiated. If any buses remain unused, an

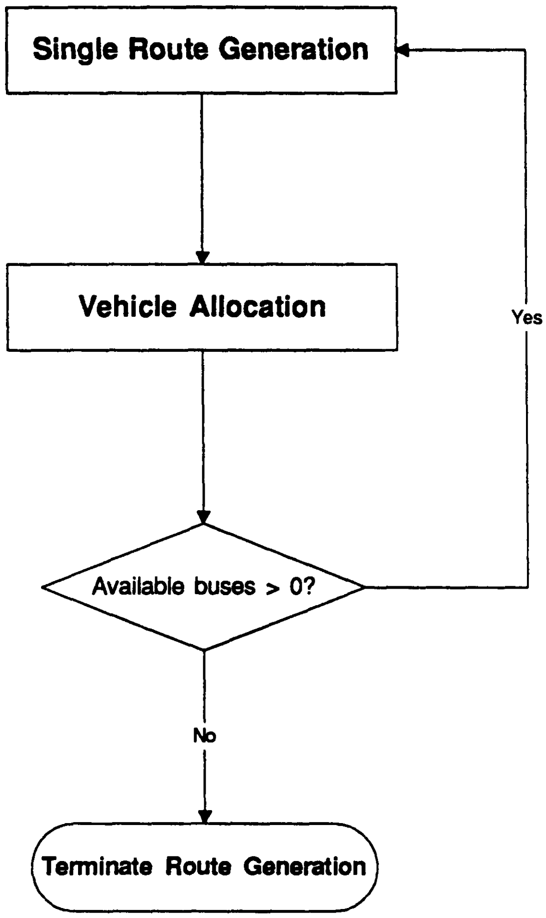


Figure 5.1: Approach 1 for Integrating Vehicle Allocation and Route Generation

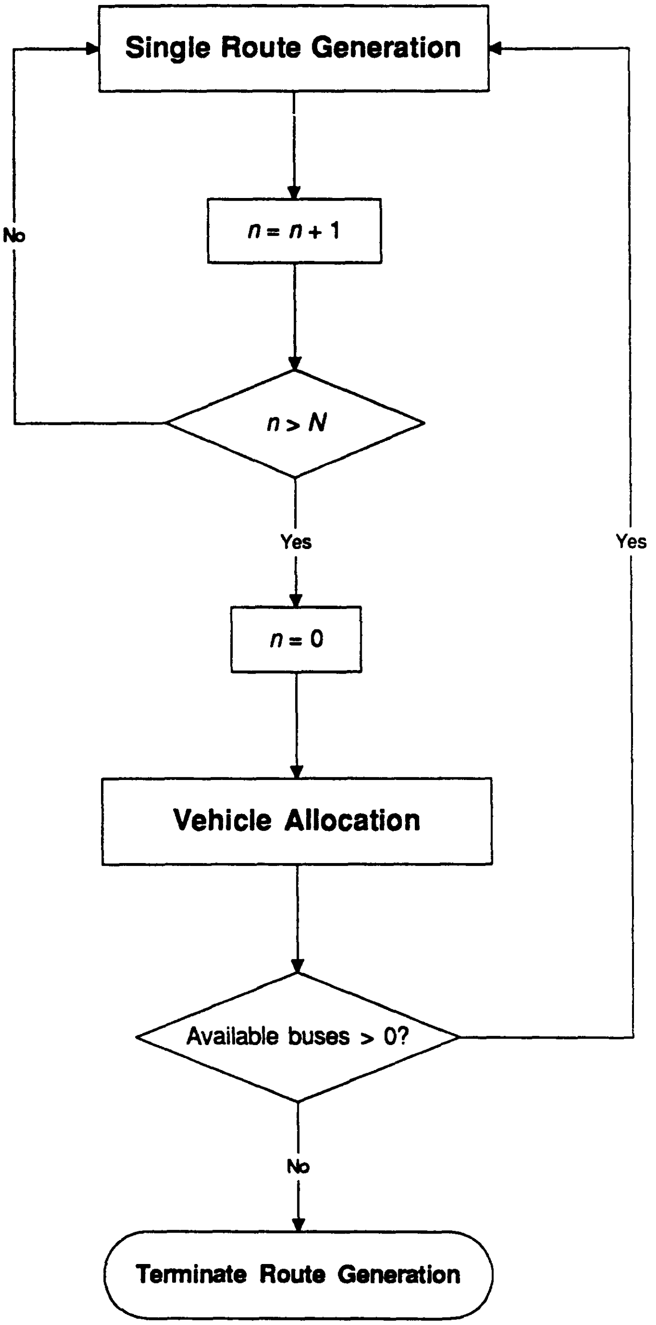


Figure 5.2: Approach 2 for Integrating Vehicle Allocation and Route Generation

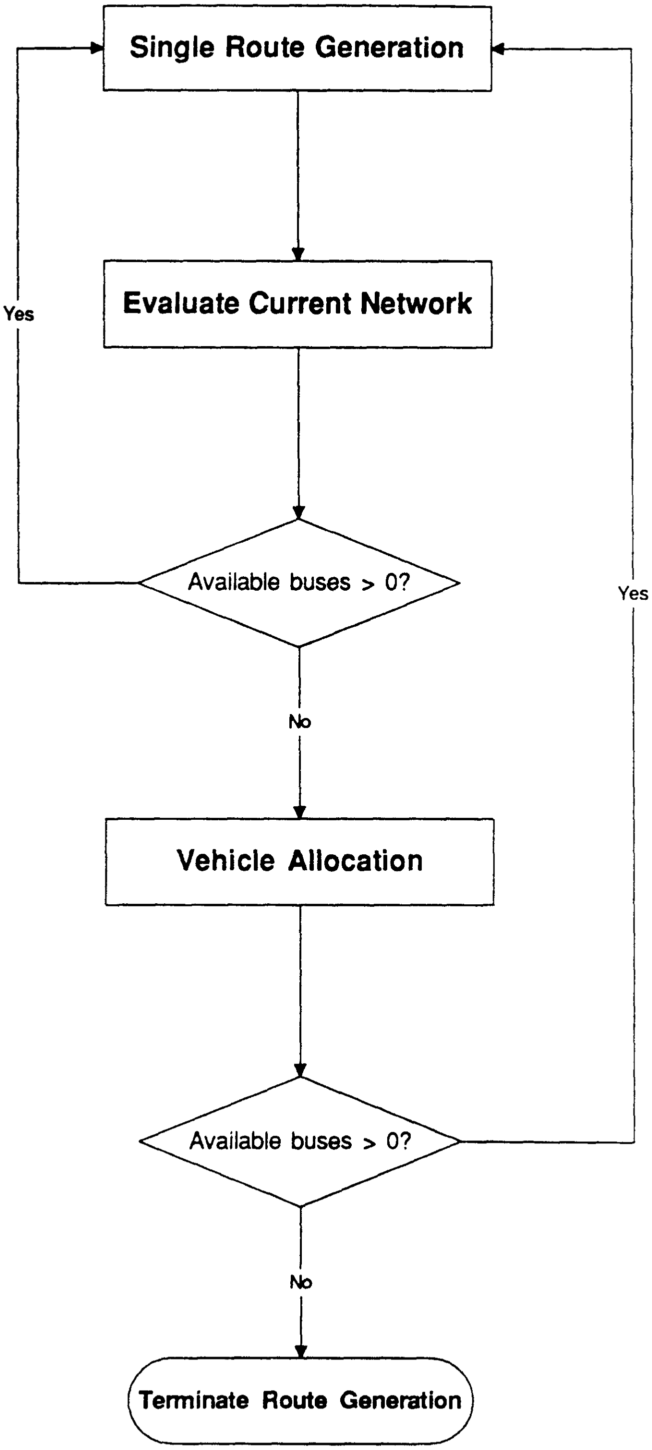


Figure 5.3: Approach 3 for Integrating Vehicle Allocation and Route Generation

additional route is created and the new network is re-evaluated; otherwise, route generation is terminated. Approach 1 is unlikely to be feasible because of its considerable computational cost. Furthermore, since the second approach may be viewed as a derivative of the more general third approach, only the latter will be considered in the rest of this chapter. The procedure for evaluating the network within the route generation loop will be discussed in section 5.10 which describes the overall design process after the integration of the RG and VA processes. In the following sections, several methods for vehicle allocation are reviewed and the advantages and limitations of each presented.

5.2 Methods for Vehicle Allocation

Methods for vehicle allocation vary considerably between theory and practice. In the transit industry, vehicles are allocated according to service standards which provide operators with practical guidelines for setting frequencies. These standards are a result of both codification of existing rules of thumb and a statement of policy, and focus mainly on lower and upper bounds for setting frequencies. However, these methods fall short of ensuring that transit resources are allocated most efficiently because they do not achieve any well-defined objective.

In theory, optimal vehicle allocation may be determined by using mathematical programming methods to solve the following problem: How can the limited resources be allocated to maximize the benefit from operating the system? If the benefits and costs that arise from operating the system can be determined for every bus route as a function of the service frequencies provided, the optimal allocation of resources can be shown to occur when the ratio of marginal benefit to marginal cost (marginal rate of return) on each route is the same and is sufficient to exhaust the available resources. At the optimum, some routes may be operating at a profit and others at a loss; yet, the solution is most efficient since the total benefit cannot be increased by shifting resources from one route to another. However, achieving economic efficiency by balancing the marginal costs and benefits is not a simple task because of the difficulty in making a plausible quantification of the costs and, particularly, the benefits on each route in terms of service frequencies. The fundamental issue in that regard is the assumption about route dependence. If ridership on a route is assumed to be affected only by the service frequency on that route, a relatively simple expression of the benefits can be obtained and the optimality conditions can be easily derived. On the other hand, if the ridership on a route is assumed to depend on the service provided by the competing routes, as well as the service on the route under consideration, the formulation of the resource allocation problem becomes much harder. Section 5.2.3 discusses further the effect of the route dependence assumption on the formulation of the general problem, as well as the implications of this assumption on the applicability of vehicle allocation to various network types. In the following sections, a number of methods for vehicle allocation are discussed.

5.2.1 The Peak-Load Factor Method

In this method, a lower bound on the frequency of service on each route is calculated as the total number of passengers at the peak-load point on the route divided by the product of the vehicle seating capacity and a specified peak-load factor. Because of its simplicity, this method is widely used in the transit industry for setting frequencies, especially during peak periods. The main weakness of this method, however, is that it is not efficient with respect to the optimization of passenger services, because it does not maximizes the user benefits of operating the system. Also, by making the average maximum passenger load factor equal on all routes, this method causes relatively long waiting times on low-ridership routes, compared with an allocation which minimizes the total passenger waiting times. Finally, this method assumes that ridership on each route is fixed, irrespective of the amount of service provided. Consequently, when equal peak-load factors are used to set frequencies, shifts in ridership that occur on competing routes in the system due to the new frequencies are not accounted for, resulting in actual passenger loads that are different from the ones that were originally sought.

5.2.2 The Square Root Rule

The best-known theory for setting frequencies on bus routes is the square root rule, which is based on the minimization of the sum of total passenger wait time costs and total operator costs. The service frequency f_k provided on route k is expressed as:

$$f_k = \sqrt{\frac{b \cdot r_k}{2 \cdot c \cdot t_k}}$$

where r_k = total ridership on route k (passengers per hour),
t_k = round trip time on route k (hours),
b = value of wait time ($ per hour), and
c = bus operating cost ($ per bus per hour).

The advantage of this method is that it realizes a more efficient allocation with respect to the system user than the peak-load factor method because it incorporates the passengers wait time cost in the objective function. However, because the ridership on a route is assumed to be fixed (not a function of the service frequencies), the full social (user and non-user) benefits are not accounted for. Still, the major weakness of the square root rule is that it does not consider bus capacity constraints. In the general case where routes of different lengths exist, the square root rule states that the passenger load should be proportional to the square root of the product of route length and ridership on that route, unlike the peak load factor rule which assumes the same passenger load on all routes. Consequently, on longer and/ or heavier routes, not enough capacity might be provided and the solution would be infeasible, which explains the lack of acceptance of this method by the industry. Nevertheless, this rule can be combined with the peak-load factor method in the following manner: buses on routes which are on the lower end of ridership or length are allocated using the more efficient square root rule, whereas buses on high-ridership or long routes are determined according to the peak load factor method which guarantees enough capacity on these routes.

5.2.3 Mathematical Programming Methods

The mathematical programming approach formulates vehicle allocation as an optimization problem in which an objective function expressing benefits from operating a transit system is maximized subject to a set of constraints. In the general case, benefits arising from operating a transit system include the direct benefits that accrue to the users of the system, as well as the positive externalities that accrue to non-users. Direct benefits are generally represented by the total consumer surplus, calculated using a ridership function on each route expressing the number of passengers as a function of service frequencies. Equivalently, the direct benefits may be stated as the savings in the total passenger wait time, determined for each route as a function of the route's frequency of service and ridership. Positive externalities - such as reduction in congestion and pollution, increased mobility to those with no automobiles, and energy savings - are assumed to be largely collinear with ridership.

Describing ridership in terms of service frequencies (or headways) is the main component in the objective function and needs to be obtained in order to solve the optimization problem. The main distinction among the various optimization models of vehicle allocation is the assumption about route dependence which, in fact, affects the expression of ridership. In general, ridership on one route depends not only on the headway of that route, but also on the headways of competing and complementary routes. In other words, an improvement in service on one route in the system will divert riders from other competing routes because of the passenger route choice behavior. An improvement in service on one route can also raise ridership on another complementary route when there is a large transfer volume between the two. The degree of route dependence, however, may vary among systems, depending on the amount of route competition and complementarity. Many transit systems in large cities outside North America, for example, are characterized by networks with extensively overlapping routes, requiring a vehicle allocation model which recognizes route dependence. On the other hand, if the routes in the network are fairly independent, the problem formulation may be greatly simplified by assuming that ridership on a route depends exclusively on the headway of that route. Table 5.1 summarizes the possible cases of route dependence and displays the various ways of expressing route ridership in terms of frequency. Note that in the first case, ridership is assumed to be unaffected by the service provided and, thus, the social benefits (proportional to ridership) cannot be expressed in the objective function of the allocation problem. This assumption is used in the peak-load factor method and square root rule, which explains why they are less efficient in allocating vehicles than the formal optimization models, which use either of the other two expressions for ridership in their objective functions.

The following sections describe two different approaches for optimal vehicle allocation. The first approach is by Furth and Wilson (1981) and is based on the assumption of independent routes. The second approach by Han and Wilson (1982) recognizes route interdependence and is applicable in the case of networks with overlapping routes.

5.3 Allocation of Buses on Independent Routes (Furth and Wilson, 1981)

The problem is formulated with the objective of maximizing the net social benefit, composed of the weighted summation of consumer surplus and ridership over all the routes in the network, subject to subsidy, fleet size, maximum load factor, and maximum (policy) headway constraints. The assumption of route independence allows a relatively straightforward formulation, resulting in a computationally tractable model. Also, since the ridership on each route is assumed to be a function only of the headway on that route, passenger assignment is not required. The problem is solved by deriving the optimality conditions relating the headways to the other variables in the model and then applying these conditions in an iterative algorithm for obtaining the optimal set of headways.

The advantage of this approach is that it recognizes transit as an important social service, addressing both user and public benefits. One of the key findings reported by the authors while experimenting with this approach is that results obtained when demand is assumed fixed were very similar to those based on variable demand. This was attributed to the fact that providing the best service for current passengers is usually a good way to

Fixed Ridership	Ridership is a Function of the Level of Service	
• ridership for each route k is a constant, r_k	• $r_k = \mathrm{r}(h_k)$, where h_k is the headway on route k	• $r_k = \mathrm{r}(h_1, h_2, ... h_n)$, where h_1, h_2, ... h_n are headways on routes i through n (including k)
• Social benefits cannot be evaluated	• No passenger path choice model required	• Passenger path choice model may be required
• Used in peak-load factor method and square-root rule	• Simple expression of total benefits (social and passenger) in objective function	• More complex expression of total benefits in objective function.

r_k = total ridership on route k

Table 5.1: Ridership Assumptions

attract new ones. The most important limitation of this model, however, is the assumption of route independence. Therefore, care must be taken in applying this model and in interpreting its results in situations where strong route competition or complementarity exists.

5.4 The Allocation of Buses in Heavily Utilized Networks with Overlapping Routes (Han and Wilson, 1982)

This model addresses networks with extensively overlapping routes and with buses frequently operating at, or close to, capacity. Ridership on a route is assumed to be a function of the headway on that route, as well as the headways on the competing routes overlapping with the route under consideration. For that reason, a passenger assignment component is incorporated in the model in the form of passenger flow constraints which relate the volume on each route segment to the origin-destination flows and service frequencies on that route and competing routes offering similar service. A functional form of this constraint cannot be specified, however, because the passenger assignment process depends on path choice criteria that need to be examined for every origin-destination pair in the network. In terms of the objective function, two assumptions are made. First, the social benefits are neglected because ridership cannot be expressed as a closed-form function of headways; thus, the objective function is limited to the direct benefits, expressed as the savings in wait time and crowding levels for all passengers. Second, it is argued that the specification of accurate wait time and crowding level in the objective function is extremely difficult since many buses will be operating under severe capacity constraints. As such, the simplified objective of minimizing the occupancy level at the most heavily loaded point on any route of the system is used instead. This objective may be viewed as equivalent to reducing passenger wait time because of the presence of the loading feasibility constraint, which ensures that passengers are not prevented from boarding a bus on their preferred route (and thus incurring longer waiting and travel times) because inadequate capacity has been allocated to that route.

The loading feasibility constraint stipulates that the passenger load on any route segment - expressed as the ratio of the link flow on that segment to the route's frequency of service - should not be greater than the passenger capacity for each bus. Since the link flow is also a function of frequency, this constraint is endogenous to the model and thus the final solution must be an equilibrium in which the passenger assignment is consistent with the frequency allocation. A two-stage heuristic solution is proposed for this model. In the first stage, base vehicle allocation is performed iteratively by (1) determining frequencies and assigning vehicles to meet the peak loads on each route and by (2) assigning passengers to find the peak loads on each route, until the equilibrium point is reached where passenger and vehicle assignments are consistent with each other. If, at such an equilibrium, the required fleet size is infeasible, then the given demand cannot be satisfied with the existing fleet operating on the given set of routes without violating the loading feasibility constraint and the underlying premise that all passengers are served by their preferred routes. If, on the other hand, the fleet size required is feasible, a second stage surplus vehicle allocation is performed to achieve the objective of reducing the level of crowding on the most heavily loaded points in the system.

The advantage of this method is that it identifies in the base allocation procedure a feasible solution which satisfies all constraints, including the loading feasibility constraint, using a minimum number of buses which provides enough capacity on each route to carry all passengers selecting it. If buses remain after this stage, wait and transfer times are improved by assigning the surplus buses so as to achieve an explicit objective.

5.5 Characteristics of Networks Created by the Route Generation Process

Prior to the selection of a method of vehicle allocation to be used in the proposed methodology, the main characteristics of the networks produced by the RG process are examined in order to adopt a method which is consistent with those characteristics. Route dependence is the key factor in deciding on a particular method, since the routes created by the RG process are expected to exhibit a certain degree of overlap. The reason for this is the nature of the process which relies on systematic rules for inserting nodes in a route under expansion. In the extreme case, if a skeleton route that has already been fully developed is re-selected again for expansion, the same route obtained earlier would be generated, since the criteria of node insertion would be identical in this case. Although this situation is prevented from arising in the RG process by modifying the demand associated with each feasible initial skeleton after each route is completely developed (section 4.5), it is still possible to encounter a situation in which routes in the network are overlapping. Consider, for example, figure 5.4a showing route k which has been created at some point of the RG process from expanding the initial skeleton (r,d,s). Suppose that in the current iteration of RG, (t,d,u) is selected as the best initial skeleton route for expansion. Note that the intermediate (major destination) node d is in common with the skeleton of route k, because the terminal nodes t and u happen to be close to nodes r and s. Consequently, the number of feasible intermediate nodes for the terminal node pair (t,u)

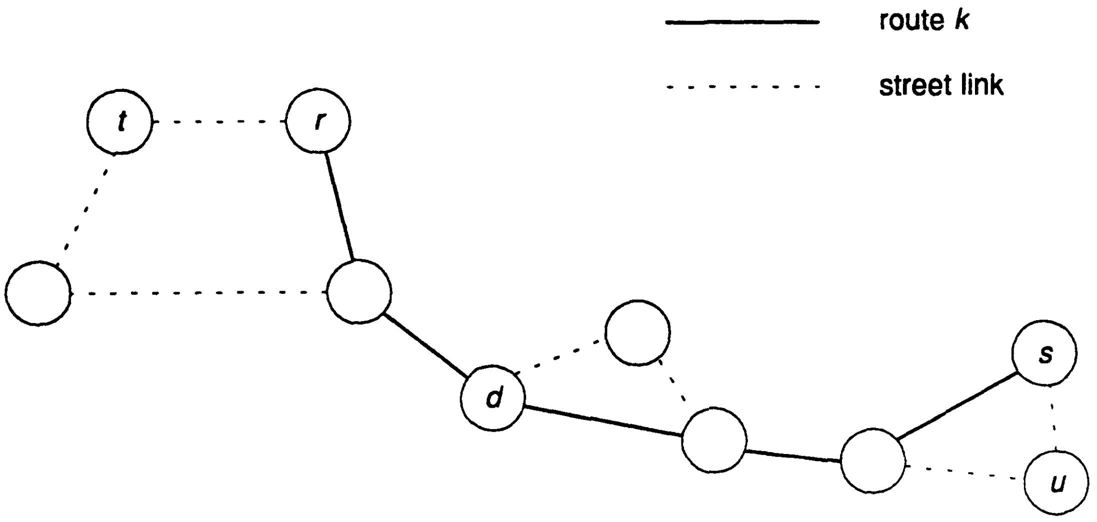


Figure 5.4a: Route *k* between *r and s*

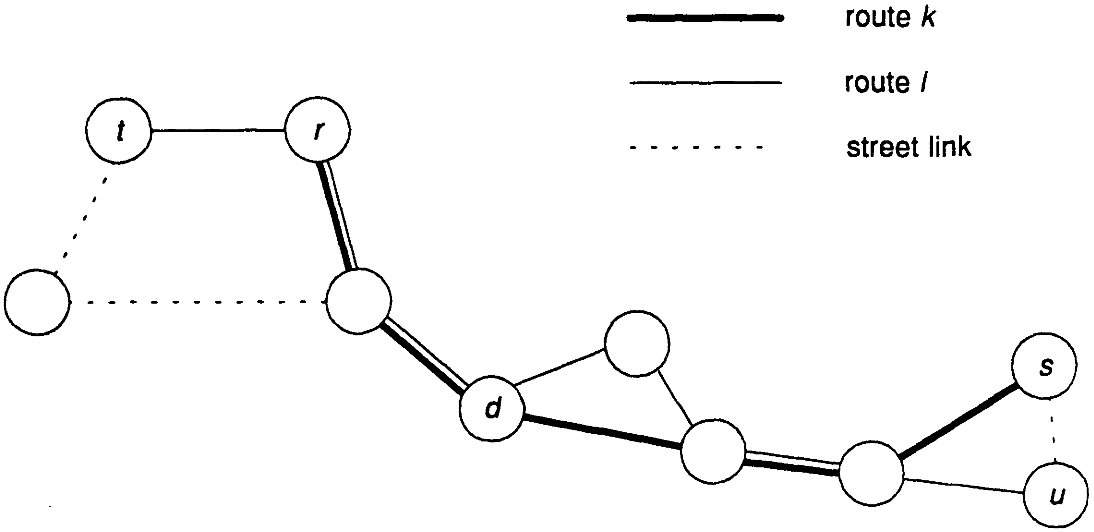


Figure 5.4b: Route *l* between *t* and *u*, overlapping with route *k*

is reduced to *d* and a few other nodes, and if node *d* has a large originating demand, it is likely that it would also be associated with terminals *t* and *u*. Note also that the RG process does not reduce the demand associated with the skeleton route (*t*,*d*,*u*), although it overlaps considerably with (the existing) route *k*, since the node pairs *td*, *du* and *tu* are not served by route *k*. As a result, the final route *l* between *t* and *u* is expected to overlap with route *k* as shown in figure 5.4b, because the initial skeletons of both routes are almost identical.

The possibility of generating networks characterized by overlapping routes is not exclusively caused by the characteristics of the RG process. The street network may also be a contributing factor by restricting access between certain districts in the service area to a small number of roads on which buses could operate. Hence, if more than one route is created between those regions, these routes would be forced to overlap along the same corridor. Overlapping routes may be considered later at the stage of network improvement by attempting to combine pairs of overlapping routes into a single operationally acceptable route. However, these improvements will not be considered until a network which exhausts the available fleet size is generated. Because overlapping routes are likely to be present in that network, the vehicle allocation procedure used in the proposed methodology should be capable of analyzing networks characterized by overlapping routes. This characteristic is essential in order to produce realistic estimates of the vehicle requirements on each route. Therefore, the vehicle allocation method of Han and Wilson is selected to allocate buses in the proposed network design methodology. In the following sections, the base and surplus vehicle allocation procedures are described in more detail.

5.6 The Base Vehicle Allocation Procedure

The base vehicle allocation procedure is performed on a given set of routes to determine the minimum number of buses (and service frequencies) on each route so as to satisfy the loading feasibility constraint. The base vehicle allocation procedure starts by classifying all the origin-destination flows into two mutually exclusive sets. Captive flows are passenger trips with a *unique* preferred path or route, whereas variable flows are defined as the complement of captive flows and consist of passenger trips served by two or more competing paths or routes. The definition of preferred and competing paths depends on the qualities of service (as valued by the riders) on all possible paths and, in the case of the model of Han and Wilson, is based on the lexicographic path choice strategy which considers two criteria: (1) the number of transfers involved in the trip and (2) the in-vehicle travel time incurred on the alternative paths. Hence, a preferred path for node pair *ij* would be one (not necessarily a unique one) with a minimum number of transfers and minimum in-vehicle travel time among all possible paths. Moreover, competing paths for node pair *ij* consist of all the preferred paths between *i* and *j*. To perform the classification of flows, all node pairs are first divided among the following three sets:

- NO_TRANS, containing node pairs which are directly satisfied by at least one route in the network.
- ONE_TRANS, which contains node pairs that are not directly linked by any route in the network, but whose demand may be served by a single transfer.
- UNSAT, which contains all node pairs unsatisfied by 0/ 1 transfers.

Trips that involve more than one transfer are not considered by the base allocation procedure, thereby avoiding a considerable amount of computation and keeping the execution time of the procedure tolerable. It is assumed that a passenger will simply not consider boarding a bus to accomplish a trip requiring two or more transfers, or alternatively that this demand is a negligible portion of the total demand. Trips belonging to the NO_TRANS set and that have a unique minimum in-vehicle travel time route serving them are considered captive. Similarly in the ONE_TRANS set, trips with a unique minimum in-vehicle travel time path are considered captive. The rest of the flows in NO_TRANS and ONE_TRANS are classified as variable. The captive flow on link *(m,l)* of bus route *k* can be determined as follows:

$$CF_{ml}^{k} = \sum_{ij \in D} V^{ij}\, \delta_{ml}^{ij}$$

where D = set of captive flow node pairs,

V^{ij} = flow between the node pair ij,

δ_{ml}^{ij} = 1, if m and l are the nodes on route k used for boarding and exiting on the unique path from i to j, 0 otherwise.

Similarly, the variable flow between nodes m and l which are served by a set of competing routes (not paths) can be determined as follows:

$$VF_{ml} = \sum_{ij \in O} \sum_{p \in P^{ij}} \delta_{ml}^{ijp}\, h_p^{ij} \quad .$$

where O = set of variable flow node pairs,

P^{ij} = set of minimum in-vehicle travel time (preferred) paths between the node pair ij,

h_p^{ij} = flow on path p between the node pair ij,

δ_{ml}^{ijp} = 1, if m and l are the nodes used for boarding and exiting on path p between the node pair ij, 0 otherwise.

The calculation of the path flows h_p^{ij} is performed by assuming that trips between i and j are divided among the competing routes in proportion to the frequency shares on these routes. This rule is based on the assumption that passengers will always board the first arriving bus of any competing route and that vehicle arrivals on each bus route follow a Poisson distribution (Chriqui and Robillard, 1975). In the case of a node pair ij belonging to the NO_TRANS set and served (directly) by a set of competing routes, the path flow on each competing route k is expressed as:

$$h_k^{ij} = \frac{q_k}{\sum_{p \in P^{ij}} q_p} V^{ij}$$

where q_k = service frequency on route k.

In the case of a node pair *ij* belonging to the ONE_TRANS set and served by a set of competing paths each with a single transfer node, the same frequency share rule is applied among the competing paths. However, in this case, each path is designated by its transfer node and consists of a set of different routes with different frequencies. For this reason, the various paths are first classified, with each route originating from node *i* and belonging to one of the competing paths defining a distinct class, and then the frequency share rule is applied to find the "class-flows". After these flows are determined, flows on paths within each class are assumed to be divided according to the relative weights of the transfer nodes involved in the class of paths.

The total passenger flow on link *ij* of route *k* can now be determined as follows:

$$f_{ij}^k = \sum_{m \in A_i^k} \sum_{l \in B_i^k} CF_{ml}^k + \sum_{m \in A_i^k} \sum_{l \in B_i^k} \frac{q_k}{\sum_{r \in X_{ml}} q_r} VF_{ml}$$

where A_i^k = set of nodes on route *k* which are not posterior to node *i*,

B_i^k = set of nodes on route *k* which are posterior to node *i*,

X_{ml} = set of competing routes between nodes *m* and *l*.

To initiate the iterative process of passenger and vehicle assignments, the following procedure is followed:

(1) Determine CF_{ij}^k for every node pair *ij* and for every bus route *k*.

(2) Using only the captive flows, determine an initial set of frequencies $q_k^{(0)}$ for all routes *k* from:

$$C \cdot q_k^{(0)} = \max_{ij} \left(\sum_{m \in A_i^k} \sum_{l \in B_i^k} CF_{ml}^k \right)$$

where C = passenger capacity on each bus.

Set $n = 0$.

(3) Compute the total passenger flow f_{ij}^k on each link ij of each route k, using $q_k = q_k^{(n)}$.

(4) Set $n = n + 1$ and calculate $q_k^{(n)}$ as the solution to the following system of equations:

$$Cq_k^{(n)} = \max_{ij} \left(f_{ij}^k \right)$$

for all links ij on each route k.

(5) If $| \, q_k^{(n)} - q_k^{(n-1)} \, |$ < allowable, then STOP. Otherwise, go to step 4.

5.7 The Surplus Vehicle Allocation Procedure

The aim of this procedure is to allocate buses remaining after the base allocation is performed to the route network so as to improve the total travel time experienced by passengers. Although all of the methods that were considered earlier in section 5.2 for base allocation are still applicable, it should be kept in mind that the solution obtained from base allocation corresponds to an equilibrium between vehicle and passenger assignments. Consequently, care must be taken in selecting a surplus allocation method so as to tamper only minimally with this equilibrium. One simple (almost trivial) way of preserving this equilibrium is suggested by Han and Wilson and consists of simply increasing the service frequencies in proportion to their existing (base allocation) values until all the available buses are consumed. In other words, the surplus frequency on each route k is given by:

$$surp_k = M \frac{q_k}{\sum_{i \in R} q_i . lrt_i}$$

where $surp_k$ = frequency of service on route k, using all of the bus fleet,

M = fleet size,

q_k = frequency of service on route k, determined in the base allocation procedure,

lrt_k = round trip time on route k, and

R = set of all routes.

Since the peak passenger load on each route is proportional to the route's frequency *share*, passenger assignment is unchanged. This method can be viewed as aiming to reduce the crowding level on the most heavily loaded point on each route in the system. In terms of minimizing the total passenger travel time, however, this method is less efficient than the square root rule and is only considered for surplus allocation for the following reasons. The first and most obvious reason is that it preserves the equilibrium solution created by the base allocation. Second, this method is computationally easier to implement than the square root rule. Finally, surplus buses that are subject to allocation are often small in number. As will be discussed further in section 5.9, when a large number of surplus buses remain after base allocation, it is generally more beneficial to use them to create additional routes than to provide better service on the current network. As such, the saving in the passenger travel time obtained from utilizing surplus buses to increase frequencies in proportion to their existing are likely to differ little from the one obtained using the square root rule. Therefore, the former method is adopted for surplus allocation.

5.8 Calculation of Performance Measures

Besides calculating the frequencies of service and the required number of buses on each route of the network, the base and surplus allocation procedures also compute several global performance measures. In terms of the amount of demand served, the vehicle allocation procedures calculate (1) the number of passenger-trips satisfied directly without transfer; (2) the number of passenger-trips served within a maximum of one transfer; and (3) the number of passenger-trips that are not served by the network. Measures which reflect the passenger travel time consist of the total in-vehicle, waiting and transfer time. These performance measures assist the planner in analyzing the solutions produced by the design process and allow the investigation of tradeoffs among the various measures of the service provided.

Passenger in-vehicle travel time is calculated by multiplying route or path flows by the trip time on the corresponding link. For waiting and transfer times, it is assumed that the average waiting time of a passenger at a node is equal to half the combined headway on the competing routes serving that node, calculated as the inverse of the sum of the frequencies on these competing routes. This calculation ignores stochasticity in bus headways.

5.9 Solution Approach

The solution approach has thus far been guided by the goal of ensuring that the number of buses required by the generated network does not exceed the available resources. An inherent feature in the development of the solution that has not yet been addressed, however, is the fact that the network with the best overall performance may not be the one exhausting all the available buses.

To illustrate this situation, consider the case where *N* routes have already been generated using the proposed design methodology. If the base allocation procedure is performed on this network, the total minimum number of buses required would be $B_{min}^{(N)}$. Suppose that the following performance measures corresponding to this network are calculated:

- $\mathrm{COV}^{(N)}$, the percentage of total demand satisfied with 0/ 1 transfer
- $\mathrm{DIR}^{(N)}$, the percentage of total demand satisfied directly (without transfer)
- $\mathrm{IVTT}^{(N)}$, the total in-vehicle travel time
- $\mathrm{WAIT}^{(N)}$, the total wait time
- $\mathrm{XFER}^{(N)}$, the total transfer time

Suppose further that the available fleet size is *FS*. Ignoring for the moment the case where $B_{min}^{(N)}$ is greater than *FS*, two cases are considered. In the first case, $B_{min}^{(N)}$ is equal to (or slightly less than) *FS* and route generation is terminated since creating an additional route would cause the total number of buses to exceed *FS*. The other steps that follow will be described in section 5.10 where the flow chart of the overall methodology is also presented.

In the second case, $B_{min}^{(N)}$ is less than *FS* and two options for the next step to be taken are possible.

Option 1: Continue the RG Process

Another route is created and added to the network, raising the number of routes in the network to *N*+1. The minimum number of vehicles required (obtained from base allocation) is now $B_{min}^{(N+1)}$, which is greater than $B_{min}^{(N)}$. If the performance measures corresponding to this network (of size *N*+1 routes) are compared to the ones corresponding to the previous network (of size *N* routes), the following inferences could be made:

a. $\mathrm{COV}^{(N+1)} \geq \mathrm{COV}^{(N)}$ and $\mathrm{DIR}^{(N+1)} \geq \mathrm{DIR}^{(N)}$. The levels of network coverage and directness cannot decline by the creation of additional routes. In the worst case, coverage and directness may stay the same if the newly generated route *k* does not serve at least one origin-destination flow that was not served in the previous network, either directly or with one transfer.

b. $IVTT^{(N+1)}$ > or < $IVTT^{(N)}$. If route *k* serves origin-destination flows that were not previously served, the total in-vehicle travel time is increased. Although the total travel time is increased in this case, the directness and coverage levels of the network would also be improved. On the other hand, if route *k* competes with other routes and offers a more frequent and direct (shorter trip time) service than the others, the total in-vehicle travel time may be reduced. The reason is that a portion of the flows served by the routes competing with *k* would be re-assigned on *k* and would benefit from its shorter trip time.

c. $WAIT^{(N+1)}$ > or < $WAIT^{(N)}$. The change in total wait time follows the change in in-vehicle travel time in most cases. Thus, similar arguments may be used to show that the changes here are also unpredictable.

d. $XFER^{(N+1)}$ > or < $XFER^{(N)}$. If route *k* creates more opportunities to satisfy trips via one transfer, the total transfer time is increased, along with the network coverage. However, if *k* changes at least a single one-transfer trip into a direct trip, the total transfer time might be improved.

In general, the direction of change in the total travel time (IVTT + WAIT + XFER) cannot be predicted. Although the levels of coverage and directness cannot decrease by creating an additional route, there is a possibility that the overall performance of the augmented network (taking into account all the measures) might be reduced.

Option 2. Terminate the RG Process and Perform Surplus Allocation

Surplus allocation would assign the (FS - $B_{min}^{(N)}$) buses onto the network in order to improve the overall level of service. The levels of network coverage and directness would remain unchanged, since the route configuration is not affected by surplus allocation.

Also, the total in-vehicle travel time would still be the same, since no additional routes are created and the frequencies of service are increased in proportion to their base-allocation values, keeping the same frequency shares among competing routes. The total transfer and wait times, on the other hand, would be improved because of the increased frequency of service on all routes. Therefore, the total travel time is improved.

In the first few route generation iterations, the improvement in the network's coverage and directness clearly outweighs the possible increase in the total travel time and the minimum number of buses required, and the decision to generate more routes is thus justified. However, as the size of the network increases, marginal benefits from demand satisfaction decrease rapidly. As such, it becomes more difficult to decide whether more routes should be created with the surplus buses or whether to terminate route generation and allocate these buses on the network to improve the total passenger travel time. Ideally, a decision rule based on the marginal benefits of assigning one surplus bus to the existing network or of operating it on a newly-created route would have been sufficient. Unfortunately, the performance measures, which are needed to express those marginal benefits, do not behave in a predictable manner and thus cannot be used.

The absence of a clear decision rule implies that the design procedure is not capable of reaching the "best" network solution automatically, without external assistance. This deficiency is not uncharacteristic of heuristic approaches to designing bus networks and is caused by the multiple-objective nature of the problem. To compensate for this deficiency, the planner may intervene in the decision-making process *at the later stages of route generation when surplus buses are about to be completely consumed*, in order to guide the process. However, any external interaction would considerably hinder the running-time efficiency of the route generation process, defeating the purpose of using it as an automated design tool. Instead, it is suggested that the design procedure would automatically investigate several near-best network solutions at the later stages of route generation and keep a record of the main characteristics of each one. These solutions may be compared at the end of route generation and the overall best network then be retrieved.

To implement this solution approach successfully, an efficient and reliable method is needed for estimating the number of surplus buses after each route is generated (within the route generation loop). This is necessary for identifying the k^{th} iteration of the route generation process beyond which the network has entered its final stages. Once the k^{th} iteration is identified, the multiple near-best solutions are determined by assigning any surplus buses and storing the network and its corresponding performance measures in a file *before* generating an additional route. Although the base vehicle allocation is reliable for estimating the minimum vehicle requirement (for identifying the k^{th} iteration), it is not computationally efficient enough to be incorporated within the route generation cycle. Instead, a much simpler procedure is needed. The next section describes in more detail the process of producing several network solutions and the procedure to estimate the number of surplus buses within the route generation loop.

5.10 Integrating Vehicle Allocation with Route Generation

The integration of vehicle allocation and route generation that was addressed earlier in section 5.1 is now re-examined in light of the proposed solution approach outlined in the previous section. The flow chart in figure 5.5 shows the integrated RG and VA processes within the overall design procedure. The number of buses B_{est} required is estimated by the network evaluation procedure incorporated within the RG loop. In the first few iterations of the RG process, B_{est} would still be considerably less than the fleet size FS, and additional routes are created without allocating buses to the network. As the RG process continues, the number of surplus buses $(FS - B_{est})$ diminishes and when this value becomes non-positive, the RG process is interrupted and the base vehicle allocation procedure is initiated. In this procedure, the total minimum number of buses B_{min} required by the current network is calculated. (If the network evaluation procedure within the RG loop is effective, B_{min} will be similar to B_{est}.) If surplus buses exist ($B_{min} < FS$), these buses are allocated on the network in order to improve the level of service. Then, the routes in the network, along with the performance measures corresponding to the base *and* surplus vehicle allocation frequencies, are stored in a file which will be used at the end of the RG

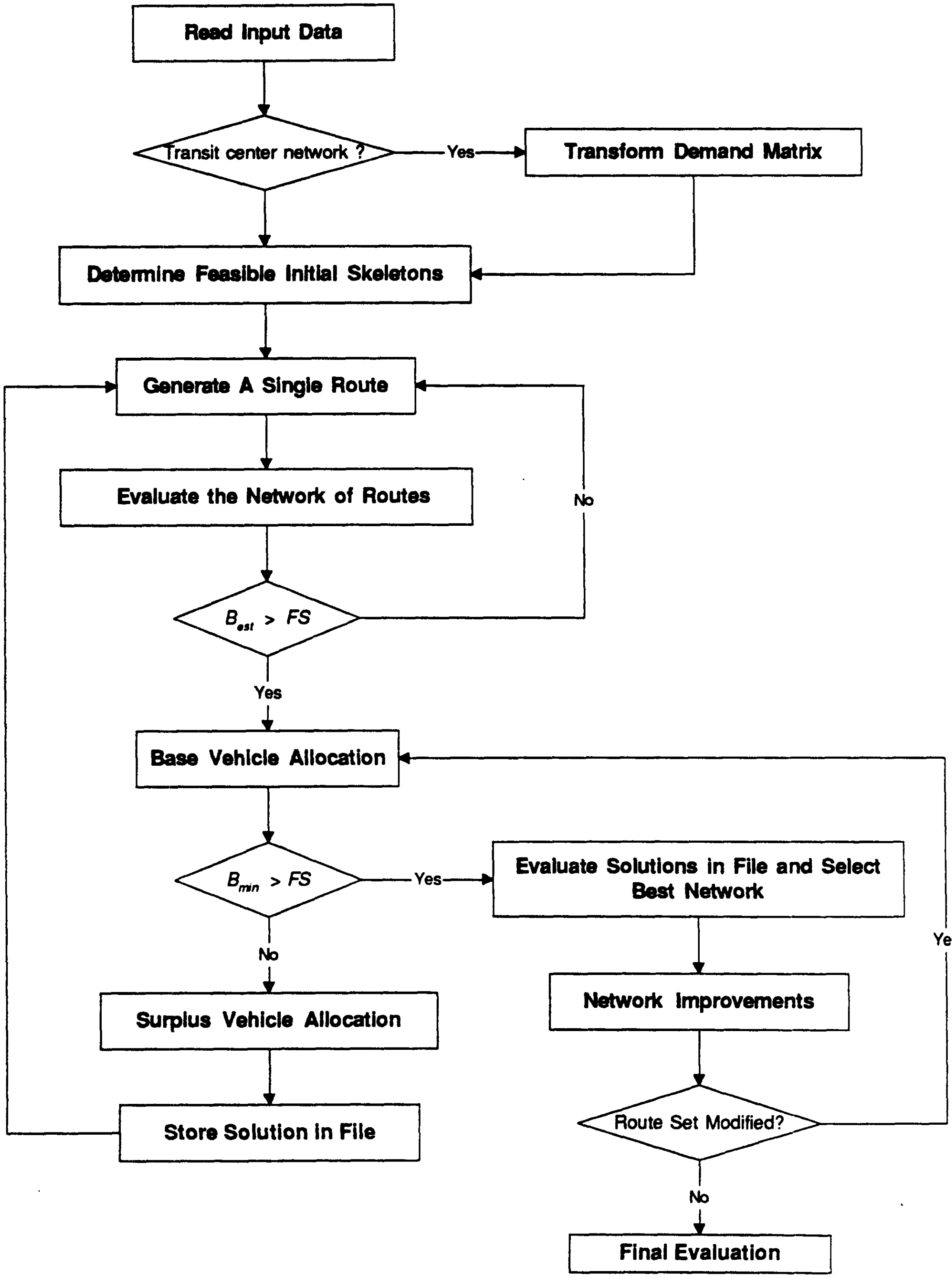


Figure 5.5: The Bus Network Design Process

process to select the "best" overall network solution. Next, the RG process is resumed, and the RG-VA cycle is repeated until all surplus buses are assigned ($B_{min} > FS$) at the end of the base vehicle allocation procedure. At this point, the file used to store previous networks is retrieved and the best overall network (taking into account the total demand satisfied, total passenger travel time, and total buses required) is selected and examined for possible network improvements.

The only component of the RG-VA cycle that has not been addressed yet is the network evaluation procedure which estimates B_{est}. Two methods for estimating this value are suggested:

1. The Direct Method

This method starts by calculating the frequencies of service on all the routes in the current network using a combination of the square root rule and peak-load factor method. On long and/ or high-ridership routes, frequency is calculated using the peak load-factor method, based on the route's maximum link-flow. On the other routes, the square root rule is used instead based on the total ridership values. The maximum link-flow and total ridership on a route are calculated from the origin-destination flows served directly by this route and may be increased by a certain fixed percentage to account for transferring volumes. Also, to prevent over-counting of trips in networks characterized with route overlap, flows on a common route segment are assigned completely to a *single* route, selected (arbitrarily) among all the overlapping routes along the segment. Although this assignment method will not be realistic in predicting the passenger volumes and vehicle requirements on *individual* routes, it should result in a reasonable estimate of the *total* number of buses required, B_{est}.

2. The Indirect Method

This method estimates B_{est} indirectly by examining other, more easily calculated network development indicators which do not depend on the frequencies of service required on the routes. The level of demand satisfied directly (without transfer) by the network, DIR, can be easily calculated after each route is created and may be used for that purpose. When this level exceeds a certain minimum value, DIR_{min}, the route generation loop is interrupted and the base allocation process is initiated. The value of DIR_{min} may be chosen based, for example, on the size and route structure of a comparable (or the existing) bus network. Alternatively, a value that is arbitrarily close to the 100% mark may be used for the first time, causing the RG process to terminate before VA is ever executed, because the minimum level of demand served directly could not be satisfied. In the next run, a lower value is used, and this process is repeated until a solution is found.

These two methods will be compared in the next section where a simple numerical example is presented.

5.11 Numerical Example

In this section, a numerical example is used to illustrate the proposed bus network design methodology. The objectives of this computational experiment are to:

- Examine the performance of the solution produced by the RG procedure for the cases of general and transit center network concepts and compare it with other solutions to the same problem.
- Investigate the benefits of integrating RG with VA and generating multiple near-best solutions.

To achieve these objectives, a benchmark problem, known in the literature as Mandl's network, is used. This problem is based on a case study in Switzerland and was first utilized by Mandl (1979) to demonstrate his transit network design algorithm. (The solution proposed by Mandl suffers from several deficiencies and will not be considered here.) Other authors such as Baaj (1993) and Shih (1994) also used the same problem to investigate the performance of their approaches. Baaj solved Mandl's problem by generating a network which closely follows the demand matrix. Shih generated two solutions - the first with a timed-transfer coordinated service and the second with uncoordinated service. These two networks are identical in terms of route structure and nodal composition; service coordination in the second network was achieved by changing the allocation of buses so as to reduce the total transfer time.

The design algorithms of Baaj and Shih (reviewed in chapter 2) are nearly identical in terms of the general approach. Neither approach is constrained by a fixed fleet size, aiming instead at generating networks with the level of demand satisfaction above a minimum pre-specified value. In this problem, Baaj and Shih determine the required fleet size as part of the solution, considering it as a performance measure rather than a constraint. For the sake of comparison, the proposed design methodology will be constrained by the fleet size to a value that is comparable to the one determined by Baaj and Shih. The solution obtained will then be compared to that of Baaj and Shih in terms of the nodal composition and the performance measures of demand satisfaction and total passenger travel time.

Mandl's network consists of 15 nodes and is illustrated in figure 5.6, with the in-vehicle travel time (in minutes) shown on each link. The demand matrix in table 5.2 contains the peak hourly number of passenger trips between each node pair. Terminals were not originally specified for this network and Baaj and Shih considered all nodes to be feasible termini. Moreover, since the network is small, all the nodes were considered in the determination of feasible initial skeletons, as opposed to limiting these nodes only to major trip generators. Therefore, the number of feasible initial skeletons is maximized.

Baaj and Shih generated solutions with three different sets of design parameters, each set differing from the other in terms of the minimum percentage of the total demand satisfied directly, the node insertion strategy, and the skeleton formation method. The other parameters were kept the same in all three design parameter sets. Because the three solutions produced by Baaj are similar in terms of their overall performance, only one solution will be presented. For the same reason, only one solution by Shih is considered.

The values of the design parameters used in RG and VA are displayed in table 5.3, along with their analogous values in Baaj's and Shih's algorithms. The weights of the objective function are calculated as described in appendix A. The bus peak load factor l_f, and operating cost c, and the passenger's value of time b are used in the network

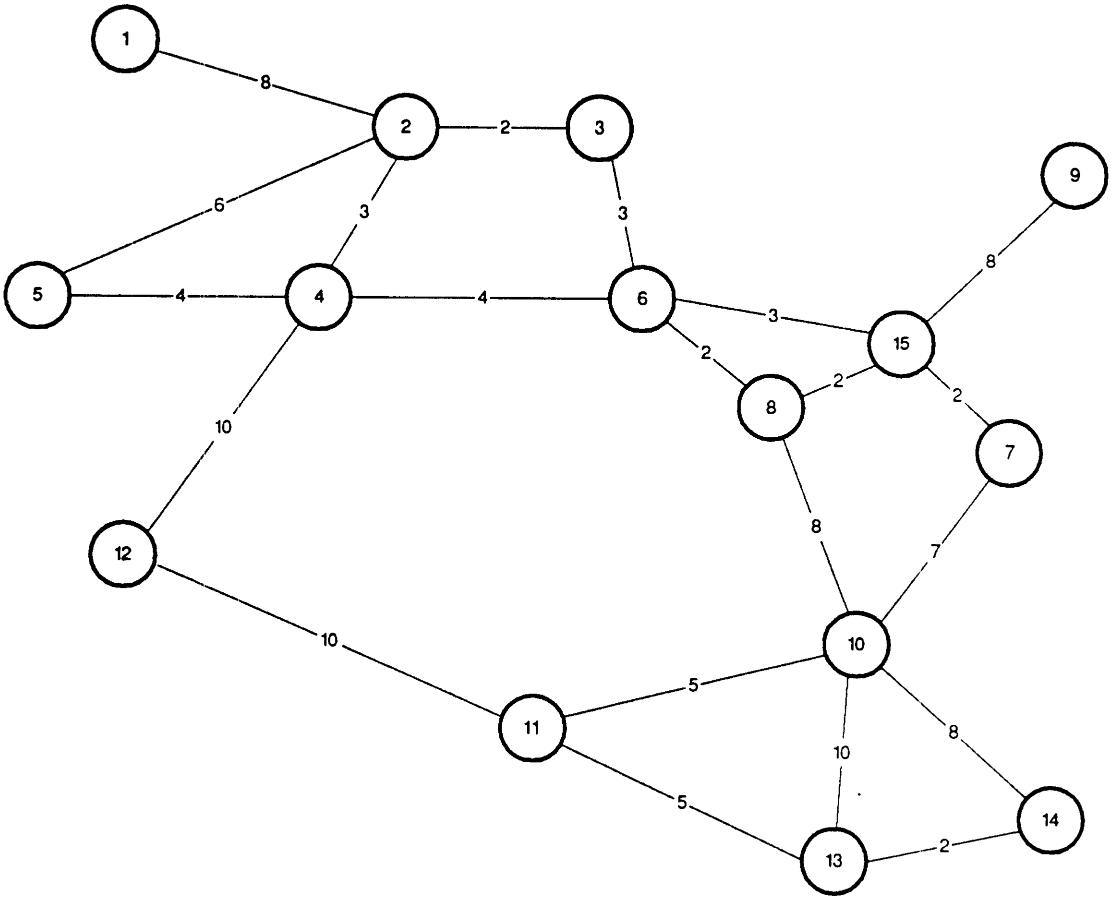


Figure 5.6: Mandl's Network

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0	400	200	60	80	150	75	75	30	160	30	25	35	0	0
2	400	0	50	120	20	180	90	90	15	130	20	10	10	5	0
3	200	50	0	40	60	180	90	90	15	45	20	10	10	5	0
4	60	120	40	0	50	100	50	50	15	240	40	25	10	5	0
5	80	20	60	50	0	50	25	25	10	120	20	15	5	0	0
6	150	180	180	100	50	0	100	100	30	880	60	15	15	10	0
7	75	90	90	50	25	100	0	50	15	440	35	10	10	5	0
8	75	90	90	50	25	100	50	0	15	440	35	10	10	5	0
9	30	15	15	15	10	30	15	15	0	140	20	5	0	0	0
10	160	130	45	240	120	880	440	440	140	0	600	250	500	200	0
11	30	20	20	40	20	60	35	35	20	600	0	75	95	15	0
12	25	10	10	25	15	15	10	10	5	250	75	0	70	0	0
13	35	10	10	10	5	15	10	10	0	500	95	70	0	45	0
14	0	5	5	5	0	10	5	5	0	200	15	0	45	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 5.2: Bus Demand Matrix (passenger trips per hour)

Design Parameters	**Value used in RG, VA**	**Value used by Baaj, Shih**
Maximum round trip time, RT_{max}	120 min.	120 min.
Maximum circuity ratio, RC_{max}	1.5*	1.5
Weight of demand in the objective function, w_d	0.00103	N/A
Weight of circuity in objective function, w_l	0.00019	N/A
Weight of connectivity in objective function, w_n	1.00	N/A
Bus capacity, *cap*	40 pass.	40 pass.
Bus peak load factor, l_f	1.25	1.25
Passenger's value of wait time, *b*	$10 per hour	N/A
Bus operating cost, *c*	$50 per bus	N/A
Minimum % of demand satisfied directly, DIR_{min}	80%	50, 70%
Fleet size, *FS*	90 buses	N/A

Table 5.3: Design Parameters

---

* For trunk routes in the transit center design, a circuity ratio of 1.2 is used.

evaluation procedure to calculate an estimate of the total vehicle requirement, B_{est} (l_f is utilized with the peak load factor method and b and c are used in the square root rule). The minimum percentage of demand satisfied directly by the network DIR_{min} is also used as the alternative (indirect) method of evaluating the network within the RG loop.

In the first run, a network which closely follows the travel patterns in the demand matrix was designed (general network design). The estimate of the number of buses required, B_{est}, was calculated within RG based on the network's level of demand served without transfer (DIR), and the VA process was initiated whenever that measure exceeded 80% (DIR_{min}). The first solution (network A) which satisfied the latter condition had a directly-served demand level of 84%, contained 7 routes and required a minimum of 86 buses (B_{min}) as calculated by the base vehicle allocation procedure. Since the fleet size (90 buses) was not completely consumed, the 4 surplus buses were allocated on the network and route generation was resumed. The next solution (network B) had an 89% level of demand served without transfer, contained 8 routes and required 91 buses. The RG process was terminated at this point, since no surplus buses remained.

In the second run, the two major destination nodes were specified as transit centers and a transit center network was generated for the same problem. The solution consisted of a high frequency trunk route operating between the two transit centers, 8 feeder routes that connect to either one of the two centers, and two other routes serving other major destinations.

The results of both runs are shown at the end of this chapter in tables 5.4 through 5.11, along with the solutions created by Baaj and by Shih. Based on these results, several observations can be made and are presented in the following paragraphs.

Case 1: General Network Design

Demand Satisfaction

The solution created by RG (tables 5.5 and 5.6) was successful in capturing the dominant trip patterns reflected in the demand matrix. Mandl's network is characterized by a radial trip pattern centered around the two highest demand nodes - 6 and 10 - which define the central core. The majority of trips destined to this core are initiated at nodes 1, 2, 6, 8 and 10 which account for 61% of the total passenger trips generated. Both networks A and B created by RG contained a route which serves all of the above major trip generators (route 3). Moreover, the solutions contained routes that either go through the core (routes 1, 3, 5 and 7) or connect to at least one of the nodes 6 or 10 (routes 2, 4 and 6), increasing the level of demand satisfied by the network solution. Network A served all passengers with at most one transfer and achieved an 84% level of direct (no transfer) demand satisfaction. In network B, all passenger-trips were served with at most one transfer, with 89% of the these trips served directly. In comparison, the level of direct demand satisfaction in Baaj's and Shih's networks is 81% and 88%, respectively. On average, routes in networks A and B are longer (36 and 37 minutes of round trip time, respectively) and slightly more circuitous (1.1 circuity ratio) than routes in Baaj's network (30 minute round trip time and 1.0 circuity ratio), which explains the higher level of direct demand attained by networks A and B. Routes in network B are similar to those in Shih's network in terms of average round trip time (37 minutes) and circuity (1.1) and, consequently, network B has a demand satisfaction level that is comparable to that of Shih's solution.

Travel Time

With the same number of passengers served by all network solutions (100%), networks A and B outperformed Baaj and Shih's solutions in terms of in-vehicle and wait times (table 5.4). Among all solutions, network A has the lowest sum of total in-vehicle and wait time which is 12% and 8% less than that on Baaj and Shih's networks respectively. The wait time on networks A and B is better because these networks contain more than one route serving nodes 1, 2, 6, 8 and 10 which are the highest-demand nodes. In network B for example, nodes 1, 2 and 6 are served by routes 3, 6 and 8 (table 5.6), and the combined frequency on the route segment containing these nodes is higher than that on route 4 in Baaj's network (table 5.8) or route 8 in Shih's network (table 5.9). Baaj and Shih calculated the total transfer time as the sum of all 5-minute penalties for each transfer (assuming that each transfer is equivalent to five minutes of in-vehicle travel time). As such, the transfer time reflects the number of transfers made rather than the actual waiting time at transfer points. In contrast, the total transfer time determined in the VA procedure is calculated according to the frequencies of routes that passengers transfer to. Consequently, the total passenger travel time on networks A and B cannot be compared to that of Shih's and Baaj's networks.

Vehicle Requirements

Because they serve more passengers, networks A and B required more vehicles (86 and 90 buses, respectively) than Baaj's network (82 buses). The number of buses required by Shih's network (68 buses[*]) is remarkably low - even when compared to Baaj's network - although the route configuration and demand satisfaction levels are quite similar to networks A and B. While Shih compares his solution to that of Baaj, he does not provide an explanation for this considerable difference in vehicle requirements (20%). A possible reason for this difference could be that Shih's vehicle allocation procedure does not provide enough capacity on each route to satisfy all passengers choosing it.

Estimation of the Number of Buses

The network evaluation procedure within the RG loop was based on the level of demand served (indirect method) in order to compare this method with the one based on a combination of the square root rule and the peak load factor method (direct method). For that purpose, VA was initiated whenever the network's level of direct demand satisfaction, DIR, exceeded 80% ($\mathrm{DIR}_{\mathrm{min}}$). Then, B_{est} was calculated using the square root rule and peak-load factor method and compared to B_{min}, the vehicle requirement produced by the base vehicle allocation process. Results of these calculations are shown in table 5.7 for network A which required 86 buses as determined by VA (base vehicle allocation). For this network, the direct method underestimated B_{est} (69 buses) by a factor of 1.2, the reason being that transfer volumes are not accounted for when using the square root rule or the peak-load factor method. (The 20% difference in vehicle requirement is comparable with the 16% of trips served using transfers in network A.) However, it is still

---

[*] Two other solutions produced by Shih using slightly different sets of design parameters required 84 buses, which is comparable with the vehicle requirement on Baaj's network and networks A and B.

reasonable to use the direct method with a (constant) factor f_{trans} which accounts for transfer volumes. Because these volumes may change during the network development process (depending on the degree of complementarity in the network) f_{trans} would first be estimated and then updated each time after the VA procedure is performed.

Generation of Multiple Solutions

The two solutions created by RG provide an example of the possible tradeoff that may exist among the various service characteristics of alternative networks. The surplus buses (4) remaining after performing base vehicle allocation on network A were assigned to the same network in order to improve its service performance. If network A *after* surplus allocation (table 5.4) is compared to network B, it can be seen that each of the individual components of the total passenger travel time in network A is better, due to the concentration of service on a lesser number of routes, with a 4% improvement in the total passenger waiting time. On the other hand, the percent of demand satisfied without transfer, DIR, increased by 5% from network A (84%) to network B (89%), because of the generation of an additional route (route 8). Although the tradeoffs involved in this simple example may not be significant, they demonstrate the capability of the design methodology to present several near-best network alternatives that can be analyzed in order to reach a better final solution.

Case 2: Transit-Center Network Design

In the second run, nodes 6 and 10 were used as the only two transit centers around which the network is to be designed. The solution obtained (table 5.10) consists of a high-frequency trunk route operating on the segment 6-8-10, as well as 8 medium-to-high frequency feeder routes which collect passengers from the rest of the network and connects them to either one of the centers. Two other routes (8 and 9) which do not connect to either of the transit centers were also generated. Shih does not provide a transit center solution to the Mandl network so a comparison with other transit center solutions is not possible. Instead, the transit center solution will be compared to the general design produced in the first run.

Demand Satisfaction

The transit center network serves less direct demand (78%) than the one produced by the general network design (89%). Compared to network B, the transit center solution does not contain a single route which connects all of the highest-demand nodes (1, 2, 6, 8 and 10), resulting in the difference in the level of direct demand satisfaction. This difference is not caused by an inadequate choice of routes but rather because of the nature of the transit center concept which relies on transferring at the transit centers. In terms of the total demand satisfied, the transit center solution performs nearly as well as the general network design, serving 98% of all passengers within 1 transfer and all of the passengers within two transfers. On average, routes in the transit center solution are shorter (33 minutes of round trip time) than routes generated in the case of a general network design (36-37 minutes). This could also be attributed to the nature of the transit center concept which favors the generation of shorter feeder routes rather than longer (and more circuitous) routes providing direct service between removed destinations.

Travel Time

The total travel time is worst in the transit center design where the minimum number of buses is allocated on the network (base vehicle allocation), especially in terms of wait time. The concentration of buses on the trunk route (route 1) and on a few major feeder routes (routes 5, 7 and 10), as well as the large number of routes generated, result in fewer buses allocated on each route. Consequently, service is *on average* less frequent in the case of a transit center design and wait time is longer. The total transfer time is also higher because of the increased number of transfers made. However, if the surplus buses (21) are allocated on the transit center network (table 5.11), its wait ant transfer times are significantly improved and its total passenger travel time becomes slightly better than that of networks A and B (5 and 6%, respectively).

Vehicle Requirements

The transit center solution consumes less vehicles (69 buses) than network B (90 buses) because it satisfies a smaller amount of total demand. The reason is that routes in the transit center network are generally shorter and less circuitous than in network B.

In summary, the solutions produced by the proposed network design methodology are comparable with the other solutions to Mandl's problem. In terms of demand satisfaction, networks A and B are slightly better than Baaj's solution and in turn, require a few more buses. Although Shih's network is quite similar to network B in various aspects, it requires a significantly lower number of buses. Shih does not comment on the vehicle requirement of his network and it is not clear from the data available how such a saving in buses could be achieved. Networks A and B are clearly better in terms of in-vehicle and wait times than the other solutions, whereas the transfer time was not available for comparison in the other solutions. The transit center solution to Mandl's problem is outperformed by the general network designs in terms of demand satisfaction, although its overall performance is still acceptable in many respects. Moreover, the allocation of surplus buses on the transit center network may considerably enhance its travel time, thereby compensating for its lower demand satisfaction level by providing a more frequent service. A comparison of the transit center network with other transit center solutions could not be made, however, because Shih does not provide a transit center solution to Mandl's problem.

Performance Measures	**Network A (Surplus Allocation)**	**Network B**	**Baaj's Solution**	**Shih's Solution**
Number of routes	7	8	7	8
% Demand served directly	84	89	81	88
% Demand served within 1 transfer	100	100	N/A	N/A
% Demand served within 2 transfers	100	100	100	100
% Demand unserved	0	0	0	0
Number of buses	90	90	82	68
IVTT (pass - min.)	159,464	159,883	180,350	168,023
WAIT (pass - min.)	18,932	19,920	22,804	26,455
XFER (pass - min.)	38,052	38,607	14,800*	9,550*
Total travel time	216,448	218,410	217,954	204,028

Table 5.4: Summary of Solutions to Mandl's Problem - General Network Design

---

* Based on a 5 minute transfer penalty for each trip.

Route Number	Nodal Composition	Frequency (bus/ hour)	Round Trip Time (min.)	Buses Required	Circuity Ratio
1	6-8-15-7-10	34	26	15	1.3
2	10-11-13	27	20	9	1.0
3	1-2-3-6-8-10	23	46	18	1.1
4	12-11-10-14	11	46	8	1.0
5	5-4-6-8-10	26	36	16	1.2
6	1-2-3-6-8-15-7	20	38	13	1.0
7	9-15-6-8-10	10	42	7	1.2
	Average / Total	22	36 / 254	12 / 86	1.1

Table 5.5: RG - VA Solution (Network A - Base Allocation)

Route Number	Nodal Composition	Frequency (bus/ hour)	Round Trip Time (min.)	Buses Required	Circuity Ratio
1	6-8-15-7-10	37	26	16	1.3
2	10-11-13	27	20	9	1.0
3	1-2-3-6-8-10	21	46	16	1.1
4	12-11-10-14	11	46	8	1.0
5	5-4-6-8-10	22	36	13	1.2
6	1-2-3-6-8-15-7	19	38	12	1.0
7	9-15-6-8-10	10	42	7	1.2
8	1-2-3-6-4-5	13	42	9	1.2
	Average / Total	20	37 / 296	11 / 90	1.1

Table 5.6: RG-VA Solution (Network B)

Route Number	Frequency (VA)	Frequency (Square Root/ Peak-Load)	Number of Buses (VA)	Number of Buses (Square Root/ Peak-Load)
1	34	28	15	12
2	27	26	9	9
3	23	18	18	14
4	11	11	8	8
5	26	20	16	12
6	20	13	13	8
7	10	8	7	6
		Total	86	69

Table 5.7: Frequency and Base Vehicle Requirements on Network A

Route Number	Route Composition	Round Trip Time (min.)	Circuity Ratio
1	10-13	20	1.0
2	10-11-12	30	1.0
3	10-14	16	1.0
4	1-2-3-6-8-10	46	1.0
5	9-15-7-10	34	1.0
6	5-4-6-8-10	36	1.0
7	1-2-4-5	30	1.1
	Total/ Average	212/ 30	1.0

Table 5.8: Baaj's Solution

Route Number	**Route Composition**	**Round Trip Time (min.)**	**Circuity Ratio**
1	3-6-15-7-10-11	40	1.1
2	2-3-6-8-15-7-10-11	46	1.1
3	10-14-13	20	1.0
4	1-2-4-6	30	1.2
5	10-11-12	30	1.0
6	9-15-7-10	34	1.0
7	5-4-6-8-10	36	1.0
8	1-2-3-6-8-10-13	66	1.0
	Total/ Average	302/ 37	1.1

Table 5.9: Shih's Solution

Route Number	**Nodal Composition**	**Frequency (bus/ hour)**	**Round Trip Time (min.)**	**Buses Required**	**Circuity Ratio**
1	6-8-10	32	20	11	1.0
2	1-2-3-6	12	26	5	1.0
3	4-6-8-15-7	6	20	2	1.1
4	6-4-12	3	28	1	1.0
5	7-10-13-14	13	38	8	1.3
6	5-4-6-8-15-9	4	40	2	1.1
7	10-11-12	21	30	10	1.0
8	1-2-4-5	4	30	2	1.1
9	5-4-12-11-13	5	58	5	1.0
10	1-2-4-6-8-15-9	23	54	21	1.1
11	10-11-13-14	4	24	2	1.5
	Average / Total	12	33 / 368	8 / 69	1.1

Table 5.10: Transit Center Network Solution -- Base Allocation

Performance Measures	**Transit Center Solution - Base Allocation**	**Transit Center Solution - Surplus Allocation**	**General Design - Network B**
Number of routes	11	11	8
% Demand served directly	78	78	89
% Demand served within 1 transfer	98	98	100
% Demand served within 2 transfers	100	100	100
% Demand unserved	0	0	0
Number of buses	69	90	90
IVTT (pass - min.)	149,170	149,170	159,883
WAIT (pass - min.)	33,984	26,054	19,920
XFER (pass - min.)	39,071	29,954	38,607
Total travel time	222,225	214,585	218,410

Table 5.11: RG - VA Solutions - General and Transit Center Designs

Chapter 6

Case Study

In this chapter, the proposed bus network design methodology and automated design procedure are applied to a real world transit network problem. Data from the bus system of San Juan, Puerto Rico is used in a case study to illustrate the performance of the design methodology. San Juan provides an appropriate case study for this research because a plan to restructure its existing bus system, involving a basic change in the network concept, has been proposed and will be shortly implemented. Consequently, solutions produced by the design methodology can be compared to the proposed network changes.

6.1 Introduction

Ridership on all forms of public transportation has declined in the San Juan Region over the last 20 to 25 years. The combination of increasing auto ownership and deterioration of public bus service reduced the latter's share of person trips within the region to 2.4% in 1990, compared to 19.6% in 1964 (Barton - Aschman Associates, 1993). Currently, the bus network in the San Juan region consists of 42 routes operating with a maximum of 183 buses. The system is generally characterized by adequate coverage and connectivity and a relatively low fare. However, the service suffers from low frequencies - with headways on some routes well over an hour - and a lack of speed, directness and schedule adherence on many routes. Modifications to the system have been limited, except for a significant improvement in bus operations made in October 1990 when service on a major corridor was contracted out to a private operator. Buses on that corridor (referred to as the Metrobus corridor) which connects the highest population and employment density areas in the San Juan region operates on a contraflow exclusive lane and provides a high frequency service.

Although minor realignments of AMA routes to feed the Metrobus corridor have been implemented, network restructuring was still indispensable for improving the overall performance of the system. In 1994, introduction of a transit center network concept was proposed (Multisystems, 1994) as a way of restructuring the system and providing better service. Improvements in service frequency, speed, directness and schedule adherence were specifically targeted in the proposed route structure change. Components of the transit center concept include 13 transit centers located at the major destinations in the region, 7 trunk routes with 4-10 minute headways, and 20 local routes with 20-30 minute headways.

6.2 Objectives of the Case Study

The principle aim of this case study is to examine the performance of the bus network design methodology using data from the San Juan Region and to compare the results to the proposed transit center network plan. Since the design procedure does not produce a complete or detailed solution, the comparison of results will focus primarily on the general layout of routes and the allocation of the available buses on them. Therefore, the case study is not an attempt to design a bus network for San Juan, but rather to experiment with the methodology and to examine its potential as a planning tool. To this end, the case study has the following objectives:

1. To produce a general network design (network G) and a transit center network design (network TC) for the San Juan region. Network G will be compared to the existing bus network with respect to the general route configuration and service frequencies. Network G will be compared to network TC in a similar manner, taking into account the level of demand satisfied by each network and the difference in the total passenger travel time between the two.

2. To compare network TC with the proposed transit center route plan in terms of the different route types involved (trunk and feeder routes) and to comment on the results.

3. To test the performance of the automated design procedure on a large network and to report the required running time of the RG and VA components.

6.3 Data for the Case Study

Figure 6.1 shows the major destinations in the San Juan Region which are located mainly along the central corridor (the Metrobus corridor) which extends between the old part of San Juan and Rio Piedras in the middle of the metropolitan area. The Old San Juan - Rio Piedras spine comprises the highest population density areas in the region, as well the principal employment concentrations in Old San Juan (governmental, retail), Santurce (governmental) and Hato Rey (financial). The other major trip generators in the region include the centers of the Bayamon (west) and Carolina (east) municipalities at the edges of the metropolitan area, which have the second and third highest population densities in the region, respectively.

The demand and travel time information used in the case study were obtained from the San Juan Regional Transportation Plan (Barton - Aschman Associates, 1993) and corresponds to the year 1990. This information is coded on a large network of the San Juan region which consists of more than 6,000 nodes, including the centroids of 765 traffic analysis zones (TAZ's).

For the case study, the demand and travel time information in the San Juan Regional Transportation Plan had to be aggregated to a size that would match with a much smaller street network, preferably not exceeding 200 nodes (for reasons discussed earlier in section 4.3). The street network was created so as to include only the area in which the (desired) bus system is to be operated. This roughly corresponds to the service area of the current bus system and consists of the central core of the San Juan metropolitan region which contains the major destinations (Old San Juan, Santurce, Hato Rey and Rio Piedras) and which extends west to Bayamon and east to Carolina. The street network was then developed manually by selecting all the major bus corridors and streets within this area on which buses can operate. Finally, this network was divided into

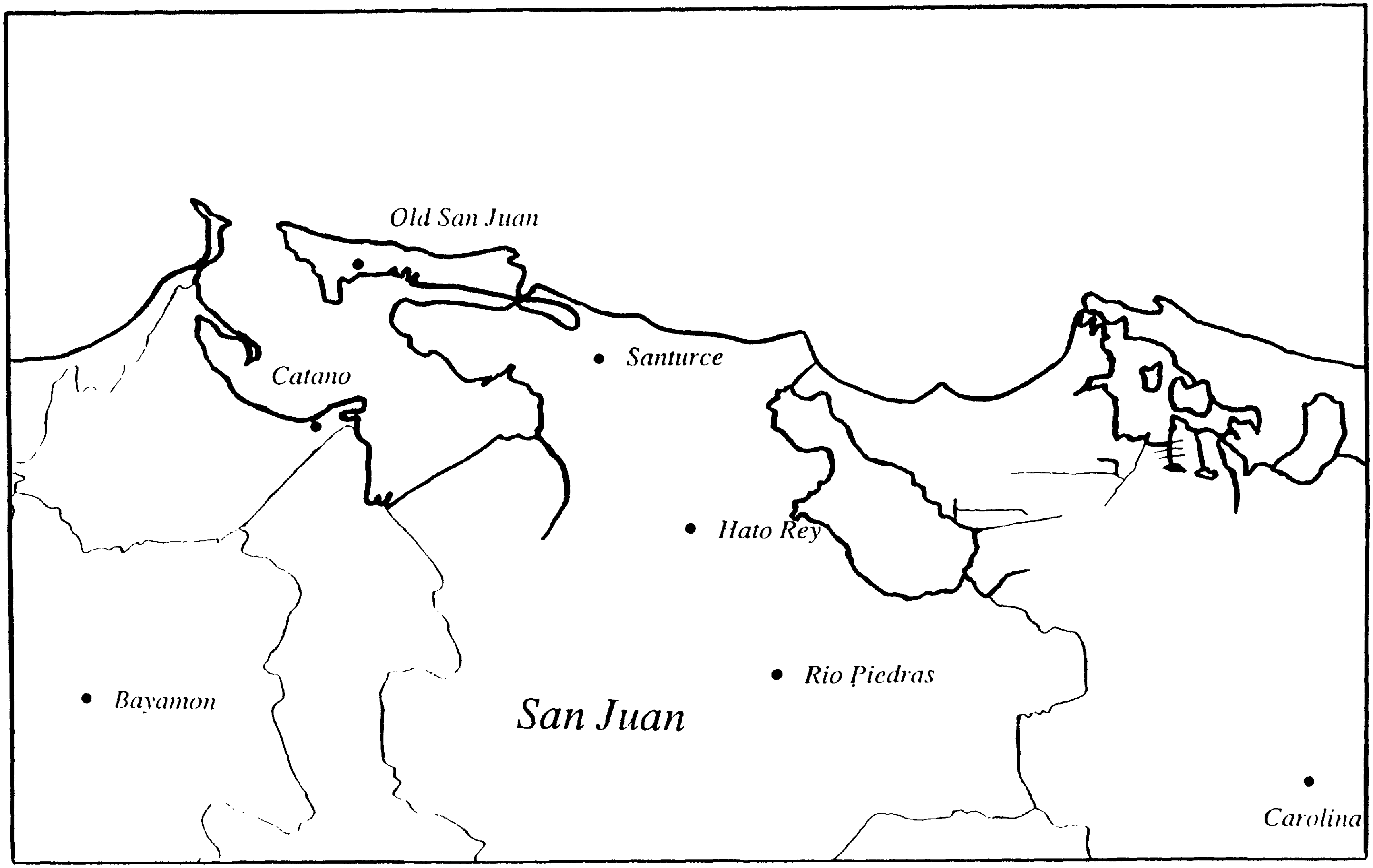


Figure 6.1: Major Destinations in the San Juan Region

238 street segments and 163 nodes based on the locations of existing termini, major destinations and access points to other modes. The connectivity information corresponding to this network was also manually established.

The demand and travel time information in the San Juan Regional Transportation Plan was available at the TAZ level (between each pair of the 765 centroids) and, thus, needed to be aggregated so as to fit the 163-node street network. For the demand matrix, the aggregation was performed first by identifying for each node in the street network its closest TAZ centroid(s)*. After all the centroids were matched, a new demand matrix was created in which column (row) *i* was obtained by summing all the columns (rows) in the existing demand matrix that match with node *i* in the street network into a single column (row). For the travel time matrix, each node in the street network was matched with its ***single*** closest TAZ centroid. A new travel time matrix was then created by assigning the travel time between each node pair in the street network the value of travel time between its matching centroid pair.

The data and design parameters used in the case study are summarized in table 6.1. The number of buses used in the case of a general network design is 183 buses, which is the currently available fleet size (including buses operating on the Metrobus corridor). The existing 8 termini which are located along the central Old San Juan - Rio Piedras corridor and around Bayamon and Carolina centers were used in the design of the network (figure 6.2). Where a transit center concept is used, these terminals were augmented by 7 transit centers located at other principal trip generators in the region, including one at the largest medical complex in the island (Centro Medico) and one near the University of Puerto Rico (UPR) campus in Rio Piedras.

The circuity and round trip time parameters used in the case study are based on the service planning guidelines presented earlier in chapter 4. The maximum limit on the round trip time (120 minutes) is consistent with the San Juan travel time conditions, allowing the formation of routes between the pertinent terminal pairs. A transfer volume

---

* This was achieved using the TransCAD software by overlaying a map of the reduced street network and a map of the TAZ centroids.

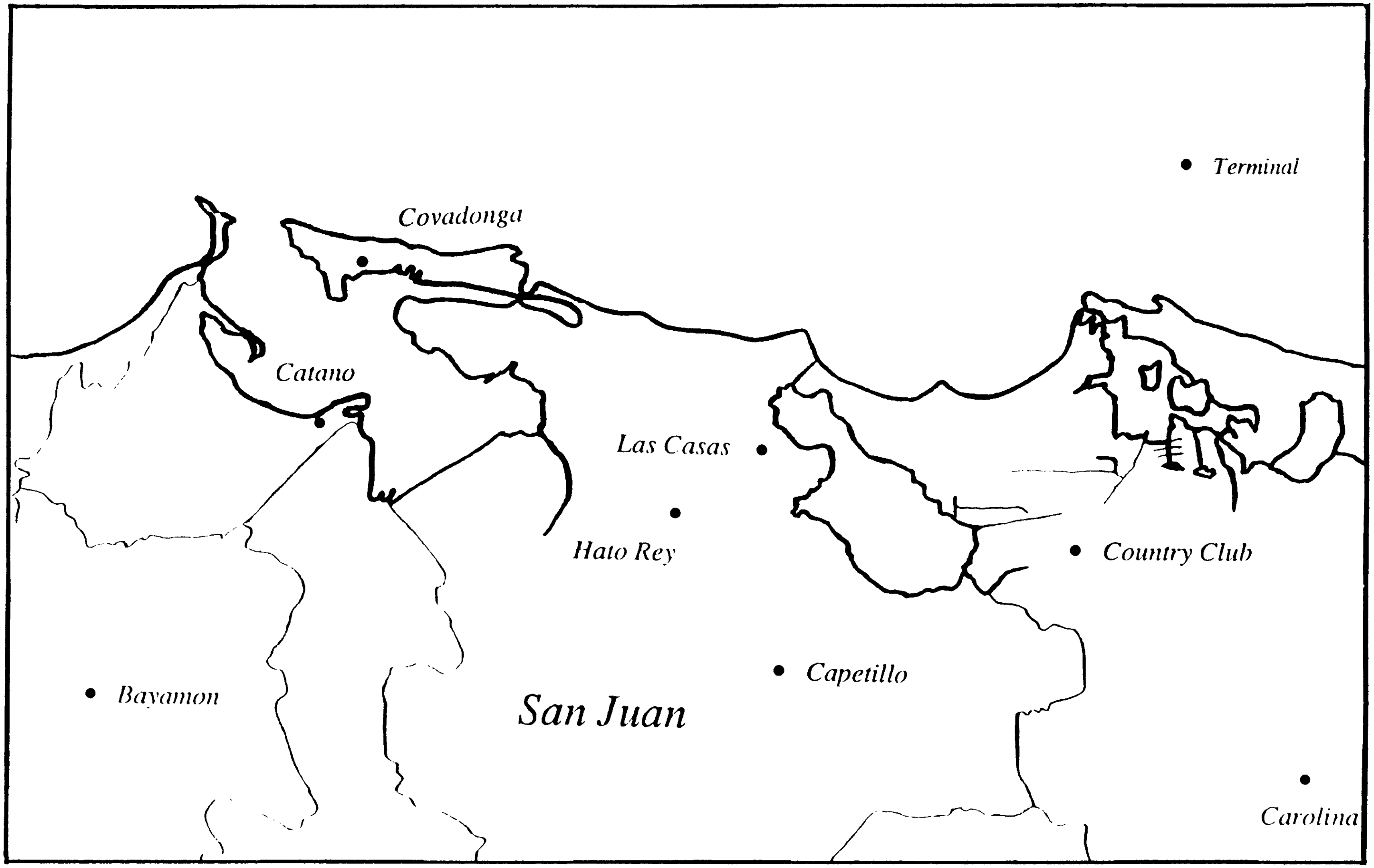


Figure 6.2: Current Termini in the San Juan Region

Input Data

Total number of nodes = 163

Number of links = 238

Number of terminals = 8 (existing)

Number of transit centers = 7 (proposed)

Total number of passenger trips per hour = 21,972

Average link travel time = 4.9 minutes

Fleet size = 183 buses

Design Parameters

Maximum round trip time = 120 minutes

Maximum allowable circuity ratio = 1.5[*]

Bus capacity = 40 passengers

Bus peak load factor = 1.25

Transfer volume factor f_{trans} = 1.3 (initial value)

Table 6.1: Data and Design Parameters for the San Juan Case Study

factor f_{trans} of 1.3[**] is used in the network evaluation procedure (within the RG loop) to account for transfer volumes when estimating the total vehicle requirement B_{est}. B_{est} is calculated using a combination of the square root rule and peak-load factor method for determining frequencies (direct method) *without* considering transfer volumes, and VA is initiated whenever $f_{trans} \cdot B_{est}$ exceeds the available fleet size, *FS*. After VA is performed, a more accurate estimate of the total vehicle requirement is obtained, B_{min}, and the value of f_{trans} is recalculated as B_{min} / B_{est}.

---

[*] For trunk routes, a circuity ratio of 1.2 is used to guarantee more direct service.

[**] A lower initial value (1.1-1.2) may also be used without affecting the network development process. This relatively high value is selected in order to perform the first vehicle allocation at an early stage of route generation.

6.4 Results of the General Network Design

In the first run of the design procedure, a general network was designed and the results are shown in tables B.1 through B.3 in appendix B. For the given round trip time and circuity parameters, 2386 feasible initial skeletons were created at the beginning of RG. The first network generated, G1 (table B.1), consisted of 20 routes and required a minimum of 155 buses. Since network G1 did not completely consume the bus fleet (183 buses), route generation was resumed. Six additional routes were generated, resulting in network G2 (tables B.2 and B.3) which comprised 26 routes and required 186 buses. The RG process was terminated at this point. Performance measures of networks G1 and G2 are summarized in table 6.2.

Individual routes in network G2 will not be compared to routes in the current bus system in San Juan, since many of these routes suffer from insufficient service and require extensive restructuring. Instead, global characteristics of network G2 pertaining to the level of demand satisfied, general route layout, service frequencies and total travel time distribution will be examined and, whenever possible, compared to their corresponding values in the current network.

Route Layout and Demand Satisfaction

Network G2 is effective in serving major destinations and trip generators in the San Juan region. Route 16 (table B.3) follows the current Metrobus corridor - operating between the Covadonga terminal (old San Juan) and the Capetillo terminal (Rio Piedras) - and serves the employment concentrations in old San Juan, Santurce and Hato Rey. Highly populated areas and other employment centers outside this central corridor are served by several routes which feed into route 16. These routes (1, 3, 4, 14, 15, 17, 18, 21 and 26) connect Bayamon, Carolina, Catano, Centro Medico, Country Club, Isla Verde and San Patricio to the central corridor, mainly through Hato Rey. (Route 3 turns north after going through Hato Rey and ends in Santurce, whereas route 1 turns south towards Capetillo). All the other feeder routes (8, 11, 12, 19 and 20) terminate at Capetillo.

Performance Measures	**Network G1**	**Network G2**
Number of routes	20	26
% Demand served directly (0 transfer)	40	44
% Demand served within 1 transfer	74	81
% Demand unserved	13	9
Number of buses required	155	186
IVTT (pass - min.)	1,782,046	1,810,514
WAIT (pass - min.)	93,206	95,933
XFER (pass - min.)	181,899	179,653
Total travel time	2,057,151	2,086,100

Table 6.2: Performance Measures of Networks G1 and G2

Routes which start outside the central core and do not feed into route 16 are few (routes 24 and 25), which explains the relatively low level of direct demand (44%) satisfied by network G2. Most of the trips made between major destinations outside the central core require at least one transfer, either at Hato Rey or Capetillo. In fact, when one-transfer trips are considered, the level of demand satisfaction is increased to 81%. Trips which are unserved by network G2 amount to 9% of the total demand and mainly are on the southern edge of the metropolitan region.

Service Frequencies

Frequencies in network G2 vary between 2 and 7 buses per hour (30 and 9 minute headways, respectively), except on route 16 which operates at a frequency of 15 buses per hour (4 minute headway). Although, the headway on route 16 is the same as the current one on the Metrobus route, the other routes in network G2 have a higher frequency on average than the rest of the routes in the existing network. If route 16 is excluded, the average headway in network G2 is 20 minutes. On the other hand, the average headway in the current network, excluding Metrobus, is 34 minutes (Multisystems, 1994). The higher level of service in network G2 is achieved by concentrating the available fleet size on a smaller number of routes (26, as opposed to 42 in the current network).

Travel Time

The total wait time on network G2 amounts to only 5% of the total passenger travel time, which is consistent with the relatively high level of service provided on most of the routes. The total transfer time contributes 9% of the total travel time, with 86% of the latter incurring as in-vehicle travel time.

6.5 Results of the Transit Center Network Design

In the second run, a transit center concept was used with the introduction of the proposed transit centers and the demand matrix transformation procedure. Results of the transit center design are displayed in tables B.4 through B.12 in appendix B. Two networks were also generated in this case. Network TC1 (table B.4) contained 19 routes and required 125 buses. The second network, TC2 (tables B.5 and B.6), required 145 buses and consisted of 24 routes. Additional routes generated with the surplus buses did not significantly improve the network's demand satisfaction level and were allocated instead on network TC2 (tables B.7 and B.8) to improve its total passenger travel time. A summary of the main characteristics of networks TC1 and TC2 are displayed in table 6.3.

The number of initial feasible skeletons in this case was 7059, which is significantly higher than the one produced in the case of the general network design, although the same parameters of round trip time and circuity were used. The reason is the addition of the 7 transit centers to the existing termini, raising the number of termini to 15 and increasing the number of possible terminal combinations. In the following paragraphs, the route layout, demand satisfaction, service frequencies and travel time on network TC2 will be examined and compared to those of network G2. Also, the results of the demand matrix transformation procedure which preceded the generation of networks TC1 and TC2 will

Performance Measures	**Network TC1**	**Network TC2 - Base Allocation**	**Network TC2 - Surplus Allocation**
Number of routes	19	24	24
% Demand served directly (0 transfer)	35	37	37
% Demand served within 1 transfer	72	78	78
% Demand unserved	11	9	9
Number of buses required	125	145	183
IVTT (pass - min.)	1,445,917	1,665,673	1,665,673
WAIT (pass - min.)	123,254	127,592	101,097
XFER (pass - min.)	215,948	209,186	165,748
Total travel time	1,785,119	2,022,451	1,932,518

Table 6.3: Performance Measures of Networks TC1 and TC2

be presented. A comparison between the routes of network TC2 and the proposed transit center routes will be made later in section 6.6.

Demand Matrix Transformation

The demand matrix transformation procedure re-directed a large portion of the origin-destination flows towards their corresponding nearest one or two transit centers. Table 6.4 shows the total demand originating at the transit center nodes before and after the demand matrix transformation. San Patricio sustained the largest increase in originating demand (255%), since all flows initiating in the vicinity of the Bayamon and Catano areas and directed towards the central core terminate either in Hato Rey or Capetillo. At the same time, San Patricio is the nearest transit center for most of these origin-destination flows. Similarly, UPR falls on a relatively direct path for most of the flows between the outlying areas in the western sector of the San Juan region and the Capetillo terminal.

Transit Center	Demand Originating at Transit Center		Percent Increase in Originating Demand
	Before Demand Matrix Transformation	After Demand Matrix Transformation	
Santurce	1,527	3,145	106
University of Puerto Rico (UPR)	790	2,236	183
Centro Medico	126	210	67
San Patricio	809	2,872	255
Rio Hondo	75	92	23
Isla Verde	588	840	43
Plaza Carolina	328	439	34

Table 6.4: Demand Originating at Transit Centers

Consequently, the demand originating at UPR was also increased considerably (183%). The demand originating at Santurce was also significantly increased by the transformation (106%), since a large number of trips are made between old San Juan and the rest of the region, most of which were re-directed through Santurce. Centro Medico and Isla Verde were less affected than the other transit centers by the demand matrix transformation. Isla Verde (43%) is relatively isolated from the rest of the network and flows re-assigned to it are unlikely to follow a path which satisfies round trip time and circuity constraints. Centro Medico (67%) is not as isolated from the network as Isla Verde, but it competes with the better-located San Patricio and UPR transit centers in capturing re-directed flows, especially those initiating at Bayamon or Catano and directed towards the central core.

Route Layout and Demand Satisfaction

The demand matrix transformation procedure had a significant impact on altering the existing trip-making patterns from being dominantly oriented towards the central corridor to being directed towards the transit centers dispersed throughout the region. The number of routes that do not feed into the central corridor have increased from 2 in network G2 (routes 24 and 25) to 5 in network TC2 (routes 8, 14, 17, 18 and 22). All of these routes contain at least one transit center terminal and were affected by the build up of demand at their corresponding transit center(s). Routes in TC2 which remained oriented towards the central corridor did not use Hato Rey or Capetillo exclusively as the connection point to the central area, as in network G2. For example, route 7 in TC2 connects Catano and San Patricio to Santurce instead of Hato Rey (route 21 in network G2, table B.3). Similarly, route 4 connects Bayamon to three transit centers - San Patricio, Centro Medico and UPR - instead of going through San Patricio and ending in Hato Rey (route 26 in network G2, table B.3).

Despite having more routes connecting destinations outside the central district without the need of transfers, network TC2 did not achieve a higher level of direct demand satisfaction than network G2 (table 6.5). The reason is related to the nature of the transit center concept and the fact that routes in TC2 are, on average, shorter and less circuitous than routes in G2 because of the trunk routes included in the solution. (The average route round trip time in network TC2 is 96 minutes, whereas the corresponding value in G2 is 108 minutes.) The percentage of trips served within one transfer in network TC2 is also lower (78%) than that in G2 (81%), but the level of trips unserved is the same in both networks (9%). All of this indicates that more transfers are involved in network TC2 than in G2.

Performance Measures	Network G2	Network TC2 - Base Allocation	Network TC2 - Surplus Allocation
Number of routes	26	24	24
% Demand served directly (0 transfer)	44	37	37
% Demand served within 1 transfer	81	78	78
% Demand unserved	9	9	9
Number of buses required	186	145	183
IVTT (pass - min.)	1,810,514	1,665,673	1,665,673
WAIT (pass - min.)	95,933	127,592	101,097
XFER (pass - min.)	179,653	209,186	165,748
Total travel time	2,086,100	2,022,451	1,932,518

Table 6.5: Comparison Between Networks G2 and TC2

Service Frequencies

Frequencies on network TC2 in the case of base allocation (tables B.5 and B.6) and network G2 are similar, both averaging 4 buses per hour on a route. In TC2, 12 routes (1 through 8, 20 and 22 through 24) contain more than one transit center and can potentially serve as trunk or feeder routes. The average frequency on these routes is 4 buses per hour (15 minute headway), and is increased to 5 buses per hour (12 minute headway) in the case where surplus buses are allocated. Routes containing only one transit center may serve as feeders and generally have lower frequencies. These routes (9 through 14, 16 through 18, and 21) have an average frequency of 3 buses per hour (4 buses per hour in the case of surplus allocation).

Travel Time

With networks TC2 and G2 roughly serving a similar number of trips, the total travel time on both networks can be compared. Table 6.5 shows that the total wait time on TC2 (base allocation) increased by about 33% from that on G2, and the total transfer time by 16%. The increase in wait time is due to the concentration of service on the potential trunk routes with fewer buses operating on the feeder and other routes (average of 6 buses) than in network G2 (average of 7 buses). The increase in the total transfer time is caused by TC2 providing fewer one-seat rides than G2 in return for a more direct service to major destinations. This is further reflected in the 8% reduction of in-vehicle travel time in TC2 (as compared to G2) which is also caused by the lower level of demand satisfaction.

When the surplus buses are allocated on TC2, the increase of wait time over that of G2 is reduced from 33% to only 5%. Furthermore, transfer time on TC2 actually becomes less than that on G2 by 8%, mainly because most surplus buses are allocated to the potential trunk routes.

If a comparison between networks TC2 and G2 is made based exclusively on the performance measures presented in the previous paragraphs (and summarized in table 6.5), it can be argued that the two networks represent a tradeoff between achieving a higher level of demand satisfaction (G2), and providing a more direct and frequent service (TC2). However, this comparison may undermine the actual demand satisfaction level of network TC2, implying that network G2 is clearly better than TC2 in that respect. In fact, the demand satisfaction level of network TC2 cannot be realistically appraised unless the demand matrix is re-estimated taking into account the changes in service introduced by the transit center concept. Demand matrix re-estimation (which is part of the proposed methodology) would determine a distribution of trips that is different from the existing one, one which is in accordance with the new service provided. If the re-estimated matrix is used to re-evaluate network TC2, better overall performance would be obtained, including a possibly higher level of demand satisfaction. Another important issue that has to be considered in the comparison of G2 and TC2 is that the latter has a potential for attracting additional ridership because it provides more frequent service to major trip generators than G2. Even if the evaluation of TC2 after demand matrix re-estimation reveals an increase in wait or transfer time, the benefits of attracting additional passengers are likely to be large enough to compensate for them.

6.6 Comparison with the Proposed Transit Center Route Plan

In this section, the trunk and feeder routes in network TC2 (surplus allocation, tables B.7 and B.8) are compared with the proposed transit center route plan. In TC2, 9 routes (1 through 6, 8, 20 and 21) were assigned relatively high frequencies (above 4 buses per hour), had a circuity ratio below 1.2, and connected two or more transit centers. These routes are shown in table 6.6 and will be compared to the proposed trunk routes. The remaining 15 routes were considered as feeders, connecting certain termini to their corresponding closest transit center(s), or local routes operating between two termini.

Trunk Routes

All of the 7 trunk routes in the proposed route plan had similar counterparts in TC2. These routes, summarized in table 6.7, roughly follow the same path as the proposed trunk routes (figure 6.3) and go through exactly the same major destinations. Generally, frequencies on the trunk routes in TC2 are slightly lower than the ones proposed, ranging between 4 and 6 buses per hour (15 and 10 minute headways, respectively), except for trunk T6 which provides a much more frequent service, with 15 buses per hour (4 minute headway). The higher frequencies on the proposed trunk routes are caused by the implicit assumption made in the transit center route plan about a build up in ridership as a result of the more frequent service on the major routes. In fact, the proposed route plan requires the deployment of additional buses (around 30 buses) in order to provide 10 minute headways (6 buses per hour) on all trunk routes and 20 minute headways (3 buses per

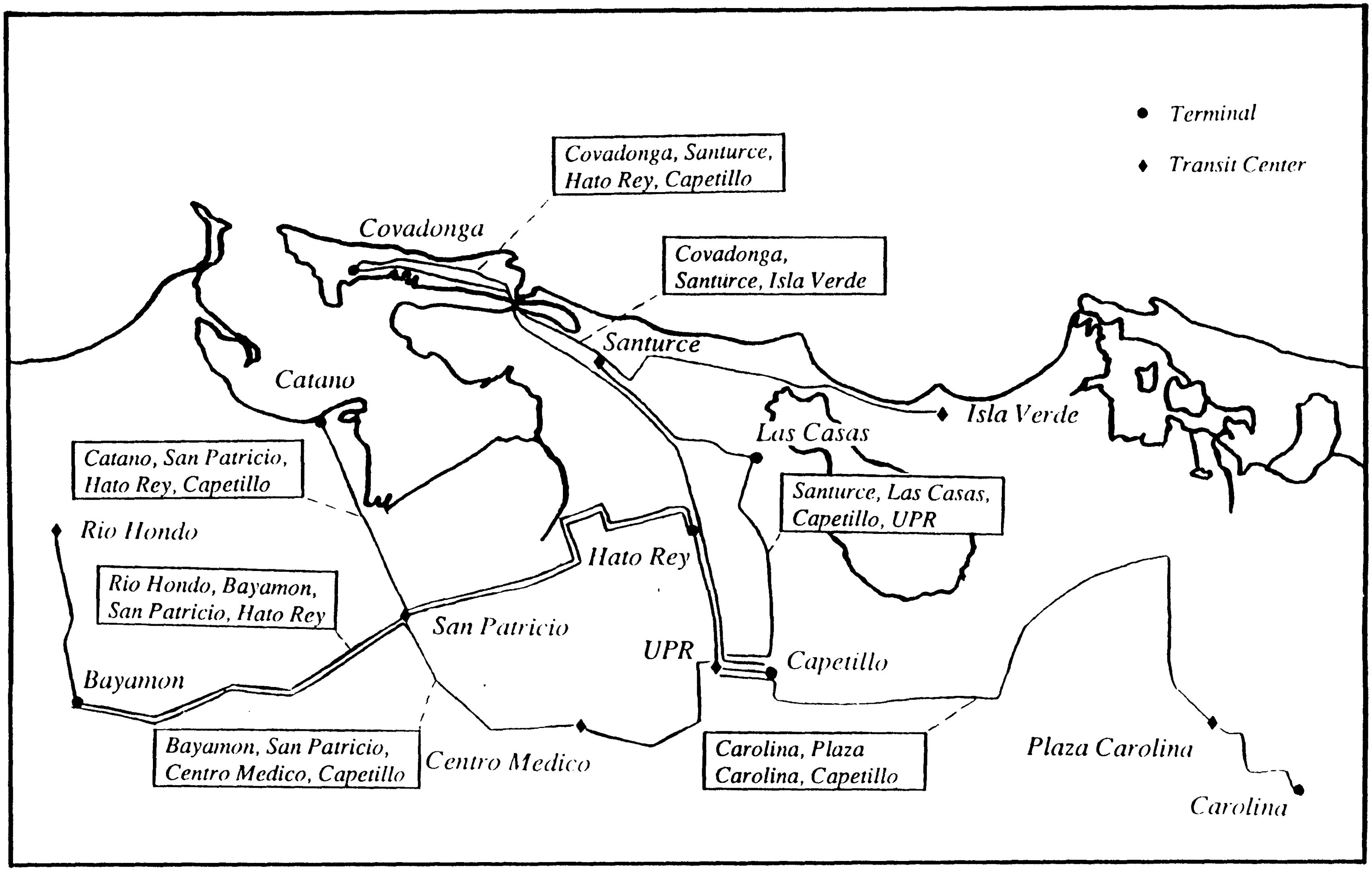


Figure 6.3: Trunk Routes in the Proposed Transit Center Route Plan

Itinerary[*]	**Circuity Ratio**	**Frequencies (bus/hour)**
1: *UPR* - Capetillo - *P. Carolina* - Carolina	1.0	6
2: *Rio Hondo* - Bayamon - *San Patricio* - Hato Rey	1.2	5
3: Covadonga - *Santurce* - Hato Rey - *UPR* - Capetillo	1.2	15
4: *UPR - Centro Medico - San Patricio* - Bayamon	1.2	4
5: Catano - *San Patricio* - Hato Rey - *UPR* - Capetillo	1.2	6
6: *Santurce* - Las Casas - Capetillo - *UPR*	1.2	4
8: *Isla Verde - P. Carolina*	1.2	4
20: *Isla Verde - Santurce* - Covadonga	1.1	5
21: *P. Carolina* - Capetillo	1.1	3

Table 6.6: Potential Trunk Routes in Network TC2

Itinerary of Trunk Routes[*]	**Proposed Frequencies (bus/hour)**	**Frequencies in TC2 (bus/hour)**
T1: Catano - *San Patricio* - Hato Rey - *UPR* - Capetillo	6	6
T2: *UPR* - Capetillo - *P. Carolina* - Carolina	6	6
T3: *UPR - Centro Medico - San Patricio* - Bayamon	6	4
T4: *Rio Hondo* - Bayamon - *San Patricio* - Hato Rey	6	5
T5: *Isla Verde - Santurce* - Covadonga	6	5
T6: Covadonga - *Santurce* - Hato Rey - *UPR* - Capetillo	12 - 15	15
T7: *Santurce* - Las Casas - Capetillo - *UPR*	6	4

Table 6.7: Trunk Routes in the Proposed Route Plan and Network TC2

---

[*] Transit centers shown in italic

hour) on most feeder routes. Therefore, if additional buses are used in the design procedure, all trunk routes in TC2 would have frequencies that match more with the ones proposed.

Feeder and Other Routes

Table 6.8 shows the feeder routes that are similar to the ones in the proposed route plan. Among the 17 major feeder routes in the proposed route plan, 11 shared strong similarities with feeder routes in TC2. Moreover, frequencies on the feeder routes in TC2 are comparable with their counterparts in the proposed route plan, although they are slightly higher on some routes. TC2 also includes one route - route 19 running between the Capetillo and Country Club termini - which does not contain any transit center nodes.

6.7 The Automated Design Procedure

The program implementing the automated design procedure is written in Pascal and was run on a DEC 5000 workstation, although it can be easily converted to run on the more commercially available PC-compatible machines. The program is designed to accommodate networks with up to 200 nodes and can generate and allocate vehicles on up to 50 routes. In order to work with larger networks, a more efficient data structure has to be created and used within the program for storing the larger amount of input data and keeping the running time within tolerable limits.

In the San Juan case study, the bulk of the running time of the design procedure was utilized by the VA process. In the first run of the procedure in which a general network design was sought, the first 20 routes (network G1) were generated in about 2.5 minutes. On the other hand, the VA procedure which followed the generation of these routes required around 37 minutes to determine the frequencies and to calculate the various performance measures, raising the total running time to about 40 minutes. The design procedure similarly required a total of 42 minutes to generate and evaluate network TC2 (24 routes) in the second run, including the time used by the demand matrix

Itinerary of Proposed Feeder / Local Routes (Multisystems, 1994)	**Frequency (bus/ hour)**	**Itinerary of Similar Feeder / Local Routes in Network TC2**	**Frequency (bus/ hour)**
F1: Covadonga - *Santurce*	3	10: Covadonga - *Santurce*	5
F2: Hato Rey - *Santurce*	3	11: Hato Rey - Las Casas - *Santurce*	5
F3: Hato Rey - *Centro Medico*	3	9: Hato Rey - *Centro Medico*	4
F4/ F5: Hato Rey - *P. Carolina*	3	12: Hato Rey - *P. Carolina*	4
F6: *P. Carolina - Isla Verde*	3	8: *P. Carolina - Isla Verde*	4
F7: *P. Carolina* - Carolina	2	17: *P. Carolina* - Carolina	3
F8: Carolina - Country Club	2	18: *P. Carolina* - Country Club	4
F10/ F11: *Centro Medico - San Patricio*	3	22: *Centro Medico - San Patricio*	3
F13: *Centro Medico - UPR* - Capetillo	3	23: *Centro Medico - UPR* - Capetillo	3
F14: Capetillo - *UPR - San Patricio*	3	24: Capetillo - *UPR - San Patricio*	3
F16: *San Patricio - Santurce*	3	7: Catano - *San Patricio - Santurce* -	3

Table 6.8: Comparison of the Proposed Feeder/ Local Routes with Network TC2

transformation procedure which did not cause any observable increase in the total running time. (The additional 2 minutes in the second run are caused mainly by the generation of more routes.)

The most time-consuming component of VA is the calculation of the variable and captive flows, which is related to the passenger assignment model. The VA also uses an iterative solution algorithm to solve for the base frequencies which utilizes a maximum allowable error term. Thus, the running time of VA may be slightly reduced by using a larger allowable error term, although the accuracy of the solution would be affected.

6.8 Summary

The computational experiment using the San Juan case study has demonstrated the performance of the proposed network design methodology and automated design procedure on a large problem. In the case of a general network design, the methodology produced a solution which achieved an acceptable level of demand satisfaction and contained routes that are consistent with the trip-making patterns in the demand matrix. In the case of the transit center design, the methodology generated a considerably different route configuration without suffering from a large reduction in the total demand satisfied and with better overall travel time. The total running time of the automated design procedure was within tolerable limits, although it could be made considerably more efficient by improving the time required by the VA component.

Chapter 7

Conclusions

In this thesis, a methodology for solving the bus network design problem (BNDP) has been developed which incorporates route generation and vehicle allocation in a single automated heuristic procedure. The thesis is motivated by the potential benefits of network restructuring and the lack of heuristic approaches that have received extensive practical application. A literature review revealed that the existing heuristic approaches to the BNDP lack three main elements that have limited their practical applicability. First, the operator's fleet size constraint has not been explicitly addressed at the route design level. Second, the demand matrix has not been adequately used to guide the design process. Finally, only a few approaches have incorporated a pre-defined network concept in the route generation process.

The methodology for solving the BNDP developed in thesis has focused primarily on these important elements and presented a bus planning tool that has the potential to receive practical application from planning authorities. The main components of the methodology consist of route generation (RG), vehicle allocation (VA), network improvements, and demand matrix re-estimation. In this thesis, the RG and VA procedures were developed and incorporated together in order to address, more explicitly, the operator's fleet size constraint. The RG procedure is heavily guided by the demand matrix so as to produce routes that serve the largest possible number of trips. RG can also be used with the same input demand matrix to design networks around the transit center concept by transforming the demand matrix prior to the initiation of the route development process. The VA procedure, which is based on a simple passenger path choice model, allocates buses on the generated routes so as to provide enough capacity to allow each passenger to board the bus on his or her preferred route.

Computational testing of the proposed methodology on Mandl's benchmark problem generated solutions that, in certain aspects, outperformed other solutions to the problem. Computational testing of the methodology on a real-world problem using data from the bus system of San Juan, Puerto Rico, produced results that are comparable with a proposed plan for network restructuring.

7.1 Route Generation

Although it is based on the basic concept of initial route skeletons which is used by other approaches, the proposed RG process introduced several modifications to that concept which lead to better performance in certain aspects of the final solution. The first modification is related to the definition of an initial route skeleton as a sequence of 3 nodes - two termini and one intermediate node - as opposed to the shortest path between the route's termini. This definition does not focus primarily on the directness of the route under development, but rather on the potential of the route to serve a major intermediate destination. As a result, networks created by the RG procedure are more likely to satisfy a large number of trips directly, without transfers. The second modification is concerned with the selection of the best initial skeleton (at each iteration of the RG process) according to the demand between all its node pairs that are unserved by the routes already developed, as opposed to the demand between its terminal nodes. Consequently, routes that have the potential to serve more trips are selected and developed prior to the other routes, increasing the number of trips satisfied in the final solution.

The proposed RG process also includes several measures to reduce the occurrence of route overlap in the final solution. The problem of route overlap can have a significant negative impact on the quality of solution because overlapping routes require a minimum amount of service, therefore consuming valuable resources and potentially preventing additional routes from being generated. At the route skeleton level, a reduction of potential overlap was achieved by ranking skeletons according to demand between skeleton node pairs that are *not* included in routes that have already been developed. This measure does not guarantee avoiding this problem, since a skeleton that overlaps with an existing route *k* and extends further beyond at least one of *k*'s termini is not affected by it. For this reason, another measure to reduce route overlap was imposed on the skeleton expansion process: when measuring the additional demand caused by inserting a node in a skeleton (first component of objective function), only the *directly unserved* demand was used. As a result, the insertion of the same high-demand nodes is discouraged in skeletons that were not affected by the measure to reduce overlap at the skeleton level.

These features that characterize the proposed RG process resulted in satisfactory performance when tested on Mandl's benchmark problem and the San Juan case study. The solution created by RG for Mandl's problem was better in capturing the dominant trip patterns in the demand matrix than the other solutions, resulting in a larger number of trips satisfied. In the San Juan case study, this observation was further reinforced in the case of general network design which produced a network (G2) satisfying 81% of the total trips within only one transfer. Furthermore, the routes in G2 were fairly distinct.

In the case of a transit center design, RG performs a demand matrix transformation which is used to modify the distribution of trips in a way that is consistent (at least in a preliminary manner) with a transit center network. In the San Juan case study, the transformation was effective in replacing routes oriented towards the central corridor by routes that are clustered around the various transit centers scattered throughout the region.

7.2 Vehicle Allocation

Developing a VA procedure was not the main objective of this thesis. As explained in the first chapter, the problem of optimal frequency determination (which can also be viewed as a problem of vehicle allocation) has received much more attention by bus planning authorities and researchers than the BNDP. Moreover, numerous computer programs for network evaluation (such as EMME/2) have been recently developed that are capable of performing passenger assignment in complex networks - such as the ones characterized by common route segments - based on sophisticated passenger behavioral models. These programs may be used with the solution obtained from the RG procedure to assign the passengers on the network, determine the required frequencies and perform a detailed analysis of the performance of the service provided. The role of the VA procedure within the methodology, however, is to estimate the total vehicle requirement on a particular network in order to guide the RG process.

In that regard, the proposed methodology relied on two separate procedures for vehicle allocation/ frequency determination: (1) the network evaluation procedure within the RG loop which uses a combination of the square root rule and the peak-load factor method for determining frequencies; and (2) the more extensive VA procedure which is based on the work of Han and Wilson (1982). Frequencies calculated using a combination of the square root rule and the peak-load factor method may be more efficient with respect to the passenger's travel time than the base vehicle allocation procedure in the case of routes with a low degree of competition and complementarity. However, as the network's complexity increases, these methods lose their advantage because they cannot ensure adequate capacity on each route. On the other hand, the base vehicle allocation procedure used in the VA procedure is capable of finding the equilibrium between passenger and vehicle assignments and determine the minimum number of buses required on each route to maintain a desired peak load factor. Computational testing on Mandl's problem revealed that the combination of the square root rule and the peak-load factor method underestimated the vehicle requirement in the case of large transfer volumes. However, these methods are still used in the proposed methodology in order to estimate roughly the number of buses required and to initiate the VA procedure which produces a more reliable estimate of the vehicle requirement.

7.3 Further Research

The bus network design problem has many facets that were not considered in this thesis and that can become topics of further research. For example, the problem can be extended to encompass more than one vehicle type, each with a different capacity and/ or cost structure, thereby requiring a different model to determine the optimal (or near optimal) allocation of service on the network of routes. Another possible extension of the problem would consider the vehicle size as a variable that needs to be determined as part of the solution. Shih and Mahmassani (1994) have already presented a transit network design procedure which includes a model for vehicle sizing.

Even the aspects of bus network design that were dealt with in this thesis may be further extended. Route generation can be improved by adopting more sophisticated node selection and insertion strategies. For example, an improved route generation procedure would be capable of monitoring a wide range of performance measures ***while developing the network*** and can proceed by creating routes that enhance specific aspects of the current network. In that sense, the route generation process would become more dynamic.

In the case of a transit center design, a procedure to identify the best (or near best) transit centers may be incorporated in the methodology. Such procedure could start by proposing a certain number of feasible transit centers, based on criteria of location and/ or travel time, and then use information obtained from the VA procedure (such as the number of transfers at each one of these candidate nodes) to update the choice of transit centers and perform another evaluation.

The demand matrix transformation procedure used in the case of a transit center design can also be improved. Besides emphasizing the role of transit centers during the process of route generation, a better procedure would be capable of appraising - in a relatively accurate and yet efficient method - the potential increase in ridership caused by the change in network concept. The modified demand matrix would then be used to influence the network evaluation process in order to allocate adequate frequencies on the trunk and feeder routes.

More efficient vehicle allocation procedures could significantly improve the overall performance of the methodology by allowing higher levels of integration with route generation. Also, efficient vehicle allocation procedures would enable the evaluation of a larger number of networks in a relatively shorter time.

Another direction for further research would focus on network improvements which is a major component of the proposed methodology that was not developed in this thesis. Although the need for network improvement actions depends greatly on the performance of the RG procedure, solutions produced by the design procedure would still benefit from actions at the route structure level. These actions include reducing low ridership/ frequency routes, route splitting, and devising new combinations of routes through branch exchange.

Finally, further testing of the methodology on different networks and their corresponding demand matrices is needed in order to form a better judgment of the methodology and to investigate its practical application.

Appendix A

Determination of the Weights w_d, w_l and w_n of the Objective Function

The following is a procedure for generating values of the weight w_d, w_l and w_n of the objective function. Recalling from chapter 4, the objective function is expressed as:

$$objctv(n) = w_d \cdot dem(n) - w_l \cdot dev(n) + w_n \cdot rts(n)$$

where *dem(n)* = the additional number of unserved trips between the node *n* and each of the nodes already in the route,

dev(n) = the additional deviation from the shortest path (in passenger-minutes) imposed by the insertion of node *n* in the route, and

rts(n) = the number of already generated routes passing through node *n*

The weights are determined so as to (1) reflect the relative trade-offs among the terms of the objective function, and (2) to reject a node *n* for insertion in a base route if *objtv(n)* is negative (second stage insertion).

The following relationships are used to determine the weights:

1. $w_d = a \cdot \frac{number\ of\ routes}{total\ demand}$; $w_l = b \cdot \frac{number\ of\ routes}{total\ demand} \cdot \frac{1}{average\ link\ length} \cdot \frac{1}{15}$; $w_n = c$
2. $a - b + c = 0$
3. a, b and c are strictly positive.

The relationships are based on the idea of selecting the weights so as an "average" node n would make the objective function value go to zero. $\frac{number\ of\ routes}{total\ demand}$, for example, can be viewed as the average additional demand incurred by the addition of a typical node.

In the example used in chapter 5, total demand = 15570 trips

number of routes = 8 (expected)

average link length (in-vehicle time) = 5.33 minutes

if we use $a = 2$, $b = 3$ and $c = 1$, the values for the weights are:

$w_d = 0.001028$

$w_l = 0.000193$

$w_n = 1$

Appendix B

Results of the San Juan Case Study

Route Number	Itinerary	Round Trip Time (min.)	Frequency (bus/ hr)	Number of Buses
1	Capetillo - Hato Rey - San Patricio - Catano	120	10	20
2	Las Casas - Hato Rey	42	2	2
3	Capetillo - Santurce - Hato Rey - Covadonga	120	13	26
4	Hato Rey - Santurce - Catano	120	5	10
5	Hato Rey - Country Club	120	3	6
6	Las Casas - Santurce - Covadonga	86	5	7
7	Capetillo - UPR - Hato Rey	62	3	3
8	Las Casas - Hato Rey - Capetillo	88	2	3
9	Country Club - Capetillo	94	2	3
10	Hato Rey - Las Casas - Covadonga	102	3	4
11	Covadonga - Catano	100	3	5
12	Capetillo - UPR - Centro Medico - San Patricio - Bayamon	120	9	18
13	Country Club - Capetillo - UPR - Hato Rey	120	4	8
14	Covadonga - Las Casas - Capetillo	120	3	6
15	Bayamon - Hato Rey	92	3	5
16	Catano - UPR - Capetillo	120	6	12
17	Hato Rey - Santurce - Covadonga	82	5	7
18	Las Casas - Catano	108	2	3
19	Country Club - Isla Verde - Hato Rey	108	2	3
20	Carolina - Isla Verde - Hato Rey	120	2	4
	Total / Average	2044 / 102	4	155 / 8

Table B.1: Network G1

Route Number	Itinerary	Round Trip Time (min.)	Frequency (bus/ hr)	Number of Buses
1	Capetillo - Hato Rey - San Patricio - Catano	120	7	14
2	Capetillo - Hato Rey - Covadonga	120	7	14
3	Hato Rey - Santurce - Catano	120	3	6
4	Hato Rey - Country Club	120	3	4
5	Las Casas - Santurce - Covadonga	86	3	3
6	Capetillo - UPR - Hato Rey	62	3	3
7	Las Casas - Hato Rey - Capetillo	88	2	3
8	Country Club - Capetillo	94	2	3
9	Hato Rey - Las Casas - Covadonga	102	3	4
10	Covadonga - Catano	100	3	5
11	Capetillo - UPR - Centro Medico - San Patricio - Bayamon	120	7	14
12	Country Club - Capetillo - UPR - Hato Rey	120	4	8
13	Hato Rey - Santurce - Covadonga	108	5	7

Table B.2: Network G2

Route Number	**Itinerary**	**Round Trip Time (min.)**	**Frequency (bus/ hr)**	**Number of Buses**
14	Country Club - Hato Rey	108	2	4
15	Carolina - Isla Verde - Hato Rey	120	3	6
16	Covadonga - Santurce - Hato Rey - UPR - Capetillo	120	15	30
17	Bayamon - Hato Rey	92	3	5
18	Las Casas - Hato Rey - San Patricio - Catano	108	5	10
19	Catano - San Patricio - UPR - Capetillo	120	6	12
20	Carolina - Capetillo	104	2	3
21	Catano - San Patricio - Hato Rey	120	2	4
22	Covadonga - Las Casas - Capetillo	120	2	4
23	Covadonga - Santurce - Las Casas	92	4	6
24	Bayamon - San Patricio - Las Casas	120	2	4
25	Catano - Las Casas	118	2	3
26	Hato Rey - San Patricio - Bayamon	120	2	4
	Total / Average	2822 / 108	4	186 / 7

Table B.3: Network G2 (Continued)

Route Number	**Itinerary**[*]	**Circuity Ratio**	**Round Trip Time (min.)**	**Frequency (bus/ hr)**	**Number of Buses**
1	Carolina - *P. Carolina* - Capetillo	1.0	104	5	9
2	Bayamon - *San Patricio* - Hato Rey	1.2	114	4	8
3	Covadonga - *Santurce* - Hato Rey - *UPR* - Capetillo	1.2	120	12	24
4	*UPR - Centro Medico - San Patricio* - Bayamon	1.2	104	3	5
5	Catano - *San Patricio* - Hato Rey - *UPR* - Capetillo	1.2	120	5	10
6	*Santurce* - Las Casas - Capetillo - *UPR*	1.2	120	3	6
7	*Santurce - San Patricio* - Catano	1.2	86	2	3
8	*Isla Verde - P. Carolina*	1.2	86	3	5
9	Hato Rey - *Centro Medico*	1.4	104	3	5
10	Covadonga - *Santurce*	1.5	100	4	6
11	*Santurce* - Las Casas - Hato Rey	1.4	120	4	8
12	*P. Carolina* - Hato Rey	1.0	112	3	7
13	*Santurce* - Las Casas	1.5	78	4	5
14	*San Patricio* - Las Casas	1.0	80	3	4
15	Covadonga - Catano	1.0	100	4	7
16	Hato Rey - *San Patricio*	1.2	60	2	2
17	*P. Carolina* - Carolina	1.4	36	2	1
18	*P. Carolina* - Country Club	1.3	98	3	5
19	Country Club - Capetillo	1.1	72	4	5
	Total / Average	1.2	1814 / 95	4	125 / 7

Table B.4: Network TC1

---

[*] Transit centers are shown in italic

Route Number	**Itinerary**[*]	**Circuity Ratio**	**Round Trip Time (min.)**	**Frequency (bus/ hr)**	**Number of Buses**
1	Carolina - *P. Carolina* - Capetillo - *UPR*	1.0	110	5	9
2	*Rio Hondo* - Bayamon - *San Patricio* - Hato Rey	1.2	120	4	8
3	Covadonga - *Santurce* - Hato Rey - *UPR* - Capetillo	1.2	120	12	24
4	*UPR* - *Centro Medico* - *San Patricio* - Bayamon	1.2	104	3	5
5	Catano - *San Patricio* - Hato Rey - *UPR* - Capetillo	1.2	120	5	10
6	*Santurce* - Las Casas - Capetillo - *UPR*	1.2	120	3	6
7	*Santurce* - *San Patricio* - Catano	1.2	86	2	3
8	*Isla Verde* - *P. Carolina*	1.2	86	3	5
9	Hato Rey - *Centro Medico*	1.4	104	3	5
10	Covadonga - *Santurce*	1.5	100	4	5

Table B.5: Network TC2 - Base Allocation

---

[*] Transit centers are shown in italic

Route Number	Itinerary[*]	Circuity Ratio	Round Trip Time (min.)	Frequency (bus/ hr)	Number of Buses
11	*Santurce* - Las Casas - Hato Rey	1.4	120	4	8
12	*P. Carolina* - Hato Rey	1.0	112	3	7
13	*Santurce* - Las Casas	1.5	78	4	5
14	*San Patricio* - Las Casas	1.0	80	3	4
15	Covadonga - Catano	1.0	100	4	7
16	Hato Rey - *San Patricio*	1.2	60	2	2
17	*P. Carolina* - Carolina	1.4	36	2	1
18	*P. Carolina* - Country Club	1.3	98	3	5
19	Country Club - Capetillo	1.1	72	4	5
20	*Isla Verde - Santurce -* Covadonga	1.1	120	4	8
21	*P. Carolina* - Capetillo	1.1	88	2	3
22	*Centro Medico - San Patricio*	1.3	72	2	3
23	*Centro Medico - UPR* - Capetillo	1.4	110	2	4
24	Capetillo - *UPR - San Patricio*	1.3	78	2	3
Total / Average		1.2	2296 / 96	4	145 / 6

Table B.6: Network TC2 - Base Allocation (Continued)

---

[*] Transit centers are shown in italic

Route Number	**Itinerary***	**Circuity Ratio**	**Round Trip Time (min.)**	**Frequency (bus/ hr)**	**Number of Buses**
1	Carolina - *P. Carolina* - Capetillo - *UPR*	1.0	110	6	11
2	*Rio Hondo* - Bayamon - *San Patricio* - Hato Rey	1.2	120	5	10
3	Covadonga - *Santurce* - Hato Rey - *UPR* - Capetillo	1.2	120	15	30
4	*UPR* - *Centro Medico* - *San Patricio* - Bayamon	1.2	104	4	6
5	Catano - *San Patricio* - Hato Rey - *UPR* - Capetillo	1.2	120	6	13
6	*Santurce* - Las Casas - Capetillo - *UPR*	1.2	120	4	8
7	*Santurce* - *San Patricio* - Catano	1.2	86	3	4
8	*Isla Verde* - *P. Carolina*	1.2	86	4	6
9	Hato Rey - *Centro Medico*	1.4	104	4	6
10	Covadonga - *Santurce*	1.5	100	5	6

Table B.7: Network TC2 - Surplus Allocation

---

* Transit centers are shown in italic

Route Number	**Itinerary**[*]	**Circuity Ratio**	**Round Trip Time (min.)**	**Frequency (bus/ hr)**	**Number of Buses**
11	*Santurce* - Las Casas - Hato Rey	1.4	120	5	10
12	*P. Carolina* - Hato Rey	1.0	112	4	8
13	*Santurce* - Las Casas	1.5	78	5	7
14	*San Patricio* - Las Casas	1.0	80	4	4
15	Covadonga - Catano	1.0	100	5	7
16	Hato Rey - *San Patricio*	1.2	60	3	2
17	*P. Carolina* - Carolina	1.4	36	3	1
18	*P. Carolina* - Country Club	1.3	98	4	5
19	Country Club - Capetillo	1.1	72	5	5
20	*Isla Verde - Santurce -* Covadonga	1.1	120	5	18
21	*P. Carolina* - Capetillo	1.1	88	3	9
22	*Centro Medico - San Patricio*	1.3	72	3	3
23	*Centro Medico - UPR* - Capetillo	1.4	110	3	4
24	Capetillo - *UPR - San Patricio*	1.3	78	3	3
Total / Average		1.2	2296 / 96	5	183 / 8

Table B.8: Network TC2 - Surplus Allocation (Continued)

---

[*] Transit centers are shown in italic

Route Number	Itinerary*	Frequency (bus/ hr)
T1	Catano - *San Patricio* - Hato Rey - *UPR* - Capetillo	6
T2	*UPR* - Capetillo - *P. Carolina* - Carolina	6
T3	*UPR - Centro Medico - San Patricio* - Bayamon	6
T4	*Rio Hondo* - Bayamon - *San Patricio* - Hato Rey	6
T5	*Isla Verde - Santurce* - Covadonga	6
T6	Covadonga - *Santurce* - Hato Rey - *UPR* - Capetillo	12 - 15
T7	*Santurce* - Las Casas - Capetillo - *UPR*	6

Table B.9: Trunk Routes in Proposed Transit Center Route Plan (Multisystems, 1994)

Route Number	Itinerary*	Circuity Ratio	Frequency (bus/ hr)
5	Catano - *San Patricio* - Hato Rey - *UPR* - Capetillo	1.2	10
1	*UPR* - Capetillo - *P. Carolina* - Carolina	1.0	8
4	*UPR - Centro Medico - San Patricio* - Bayamon	1.2	7
2	*Rio Hondo* - Bayamon - *San Patricio* - Hato Rey	1.2	9
20	*Isla Verde - Santurce* - Covadonga	1.1	9
3	Covadonga - *Santurce* - Hato Rey - *UPR* - Capetillo	1.2	15
6	*Santurce* - Las Casas - Capetillo - *UPR*	1.2	10

Table B.10: Trunk Routes in Network TC2

---

[*] Transit centers are shown in italic

Route Number	Itinerary[*]	Frequency (bus/ hr)
F1	Covadonga - *Santurce*	3
F2	Hato Rey - *Santurce*	3
F3	Hato Rey - *Centro Medico*	3
F4	Hato Rey - *P. Carolina* (1)	3
F5	Hato Rey - *P. Carolina* (2)	3
F6	*P. Carolina - Isla Verde*	3
F7	*P. Carolina* - Carolina	2
F8	Carolina - Country Club	2
F9	*P. Carolina* - Capetillo - *UPR*	3
F10	*Centro Medico - San Patricio* (1)	3
F11	*Centro Medico - San Patricio* (2)	3
F12	*Centro Medico* - Bayamon	2
F13	*Centro Medico - UPR* - Capetillo	2
F14	Capetillo - *UPR - San Patricio*	3
F15	Catano - *Rio Hondo*	2
F16	*San Patricio - Santurce*	3
F17	*Isla Verde* - Capetillo	3

Table B.11: Major Feeders and Local Routes in Proposed Transit Center Route Plan (Multisystems, 1994)

---

[*] Transit centers are shown in italic

Route Number	Itinerary[*]	Frequency (bus/ hr)
7	*Santurce* - *San Patricio* - Catano	2
8	*Isla Verde* - *P. Carolina*	4
9	Hato Rey - *Centro Medico*	3
10	Covadonga - *Santurce*	3
11	*Santurce* - Las Casas - *Hato Rey*	7
12	*P. Carolina* - Hato Rey	4
13	*Santurce* - Las Casas	5
14	*San Patricio* - Las Casas	3
15	Covadonga - Catano	4
16	Hato Rey - *San Patricio*	2
17	*P. Carolina* - Carolina	2
18	*P. Carolina* - Country Club	3
19	Country Club - Capetillo	4
21	Capetillo - *P. Carolina*	6
22	*Centro Medico* - *San Patricio*	2
23	*Centro Medico* - *UPR* - Capetillo	2
24	Capetillo - *UPR* - *San Patricio*	2

Table B.12: Feeders and Local Routes in Network TC2

---

[*] Transit centers are shown in italic

Bibliography

1. Andreasson, I., "A Method for the Analysis of Transit Networks", Advances in Operations Research: Proceedings of the Second European Congress on Operations Research, p.1-8, 1977.
2. Baaj, M. H. and Mahmassani, H. S., "An AI-Based Approach for Transit Route System Planning and Design", Journal of Advanced Transportation, Vol. 25, no. 2, pp.187-210, 1993
3. Barton-Aschman Associates Inc., "San Juan Regional Transportation Plan", Department of Transportation and Public Works, Commonwealth of Puerto Rico, 1993
4. Bursey, N., McCallum, D., Macbriar, I. and Mills, K., "The National Bus Company 'MAP' market analysis project", Traffic Engineering and Control, 20, p.575-582, 1979
5. Byrne, B., "Cost Minimizing Positions, Lengths and Headways for Parallel Public Transit Lines Having Different Speeds", Transportation Research, 10, p.209-214, 1976
6. Chapleau, R. and Chriqui, C., "Public Transport Assignment Calculation with TRANSCOM Suite', Proceedings of PTRC Summer Annual Meeting, University of Warwick, July 75
7. Chapleau, R., Allard, B. and Canova, M., "MADITUC: Un Model de Planification Operationnelle Adapte aux Entreprises de Transport en Commun de Taille Moyenne", publication no. 265, Centre de Recherche sur les Transports, Universite do Montreal, 1982
8. Chriqui, C. and Robillard, P., "Common Bus Lines", Transportation Science, 9, pp.115-121, 1975
9. Chua, T. A., "The Planning of Urban Bus Routes and Frequencies: A Survey", Transportation, vol. 12, pp.147-172, 1984

10. Dial, R. B., "Transit Pathfinder Algorithm", Highway Research Record, 205, pp.67-85, 1967
11. Dubois D., Bell G. and Llibre M., "A Set of Methods in Transportation Network Synthesis and Analysis", Journal of Operations Research Society, 30, pp.797-808, 1979.
12. Florian, M. and Spiess, H., " Models and Techniques for the Planning of Urban and Regional Transportation Networks", publication no. 463, Centre de Recherche sur les Transports, Universite do Montreal, 1986
13. Furth, G. and Wilson, N. H. M., "Setting Frequencies on Bus Routes: Theory and Practice", Transportation Research Record, vol. 818, pp.1-7, 1981
14. Han, A. F. and Wilson, N. H. M., "The Allocation of Buses in Heavily Utilized Networks with Overlapping Routes", Transportation Research B, vol. 16B, no. 3, pp.221-232, 1982
15. Hasselstrom, D., "Public Transportation Planning", Sweden, BAS edition, 1981
16. Hickman, M., "Assessing the Impact of Real-Time Information on Transit Passenger Behavior", MIT, 1993
17. Hurdle V., "Minimum Cost Locations for Parallel Public Transportation Lines", Transportation Science, 7, p.97-102, 1973
18. Jansson, K., "VIPS-II: New Computer System for Public Transit Planning", paper presented at the American Public Transit Association Annual Meeting, 1987
19. Jansson, K. and B. Ridderstolpe, "A Method for the Route-Choice Problem in Public Transport Systems", Transportation science, 26(3), pp.246-251, 1992
20. Kocur, G. and Hendrickson, C., "Design of Local Bus Service with Demand Equilibrium", Transportation Science, 16, 1982
21. Lampkin, W. and Saalmans, P. D., "The Design of Routes, Service Frequencies, and Schedules for a Municipal Bus Undertaking: A Case Study", Operations Research Quarterly, vol. 18, no. 4, pp.375-397, 1967
22. Last, A. and Leak, S. E., "TRANSEPT: A Bus Model', Traffic Engineering and Control, 17,p.14-17, 1976

23. Le Clercq, F., "A Public Transport Assignment Method", Traffic Engineering and Control, 14, p.91-96, 1972

24. Morlok, E. K., "Supply Functions for Public Transport: Initial Concepts and Models", International Symposium on Traffic Equilibrium Methods, Montreal 1974

25. Multisystems Inc., "Route Level Demand Models: A Review", U.S. Department of Transportation Report DOT 1-82-6, Washington, D.C., 1982.

26. Multisystems Inc., "Comprehensive Bus and Publico Plan", 1994

27. NCHRP report 69, "Bus Route and Schedule Planning Guidelines", Transportation Research Board, National Research Council, Washington, D. C., 1980

28. Newell, G., "Some Issues Relating to the Optimal Design of Bus Routes", Transportation Science, 13, p.20-35, 1979

29. O' Brien, W. B., "An Example of the Functional Design and Costing of Transit Routes", Urban Bus Transit - A Planning Guide, University of Waterloo, Canada, 1974

30. Osleeb, J. P. and Moellering, H., "TRANSPLAN: An Interactive Geographic Information System for the Design and Analysis of Urban Transit Systems, URISA, 1976

31. Pratt, R. H. and Schultz, G. W., "A Systems Approach to Sub-Area Transit Service Design", Transportation Research Record, 419, p.37-47, 1972

32. Shih, M. C. and Mahmassani, H., "A Planning and Design Model for Transit Route Networks with Coordinated Operations", prepared for presentation at the Triennal Symposium on Transportation Analysis, Capri, Italy, July 1994

33. Silman, L. A., Barzily, Z. and Passy, U., "Planning the Route System for Urban Buses", Computers and Operations Research, vol. 1, pp.201-211, 1974

34. Spiess, H. and M. Florian, "Optimal Strategies: A New Assignment Model for Transit Networks", Transportation Research B, 23B (2), pp.83-102, 1989