Technical Report Documentation Page

1. Report No. SWUTC/04/167244-1	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Optimal Transit Route Network Design Problem: Algorithms, Implementations, and Numerical Results		5. Report Date May 2004 6. Performing Organization Code
7. Author(s) Wei Fan and Randy B. Machemehl		8. Performing Organization Report No. Report 167244-1
9. Performing Organization Name and Address Center for Transportation Research University of Texas at Austin 3208 Red River, Suite 200 Austin, Texas 78705-2650		10. Work Unit No. (TRAIS) 11. Contract or Grant No. 10727
12. Sponsoring Agency Name and Address Southwest Region University Transportation Center Texas Transportation Institute Texas A&M University System College Station, Texas 77843-3135		13. Type of Report and Period Covered 14. Sponsoring Agency Code

15. Supplementary Notes

Supported by General Revenues from the State of Texas.

16. Abstract

Previous approaches used to solve the transit route network design problem (TRNDP) can be classified into three categories: 1) Practical guidelines and ad hoc procedures; 2) Analytical optimization models for idealized situations; and 3) Meta-heuristic approaches for more practical problems. When the TRNDP is solved for a network of realistic size in which many parameters need to be determined, it is a combinatorial and NP-hard problem in nature and several sources of non-linearities and non-convexities involved preclude guaranteed globally optimal solution algorithms. As a result, the meta-heuristic approaches, which are able to pursue reasonably good local (possibly global) optimal solutions and deal with simultaneous design of the transit route network and determination of its associated service frequencies, become necessary.

The objective of this research is to systematically study the optimal TRNDP using hybrid heuristic algorithms at the distribution node level without aggregating the travel demand zones into a single node. A multi-objective nonlinear mixed integer model is formulated for the TRNDP. The proposed solution framework consists of three main components: an Initial Candidate Route Set Generation Procedure (ICRSGP) that generates all feasible routes incorporating practical bus transit industry guidelines; a Network Analysis Procedure (NAP) that determines transit trips for the TRNDP with variable demand, assigns these transit trips, determines service frequencies and computes performance measures; and a Heuristic Search Procedure (HSP) that guides the search techniques. Five heuristic algorithms, including the genetic algorithm, local search, simulated annealing, random search and tabu search, are employed as the solution methods for finding an optimal set of routes from the huge solution space. For the TRNDP with small network, the exhaustive search method is also used as a benchmark to examine the efficiency and measure the quality of the solutions obtained by using these heuristic algorithms.

Several C++ program codes are developed to implement these algorithms for the TRNDP both with fixed and variable transit demand. Comprehensive experimental networks are used and successfully tested. Sensitivity analyses for each algorithm are conducted and model comparisons are performed. Numerical results are presented and the multi-objective decision making nature of the TRNDP is explored. Related characteristics underlying the TRNDP are identified, inherent tradeoffs are described and the redesign of the existing transit network is also discussed.

17. Key Words Transit, Route, Network Design, Multi-objective Decision Making, Heuristic Search, Genetic Algorithm, Local Search, Simulated Annealing, Tabu Search, Random Search, Exhaustive Search, Network Analysis, Headway, Transfer, Long-walk		18. Distribution Statement No restrictions. This document is available to the public through NTIS: National Technical Information Service 5285 Port Royal Road Springfield, Virginia 22161	
19. Security Classif.(of this report) Unclassified	20. Security Classif.(of this page) Unclassified	21. No. of Pages 267	22. Price

Form DOT F 1700.7 **(8-72)** **Reproduction of completed page authorized**

Optimal Transit Route Network Design Problem: Algorithms, Implementations, and Numerical Results

By
Wei Fan
Randy B. Machemehl

Research Report SWUTC/04/167244-1

Southwest Region University Transportation Center
Center for Transportation Research
University of Texas at Austin
Austin, Texas 78712

May 2004

DISCLAIMER

ABSTRACT

Previous approaches used to solve the transit route network design problem (TRNDP) can be classified into three categories: 1) Practical guidelines and ad hoc procedures; 2) Analytical optimization models for idealized situations; and 3) Meta-heuristic approaches for more practical problems. When the TRNDP is solved for a network of realistic size in which many parameters need to be determined, it is a combinatorial and NP-hard problem in nature and several sources of non-linearities and non-convexities involved preclude guaranteed globally optimal solution algorithms. As a result, the meta-heuristic approaches, which are able to pursue reasonably good local (possibly global) optimal solutions and deal with simultaneous design of the transit route network and determination of its associated service frequencies, become necessary.

The objective of this research is to systematically study the optimal TRNDP using hybrid heuristic algorithms at the distribution node level without aggregating the travel demand zones into a single node. A multi-objective nonlinear mixed integer model is formulated for the TRNDP. The proposed solution framework consists of three main components: an Initial Candidate Route Set Generation Procedure (ICRSGP) that generates all feasible routes incorporating practical bus transit industry guidelines; a Network Analysis Procedure (NAP) that determines transit trips for the TRNDP with variable demand, assigns these transit trips, determines service frequencies and computes performance measures; and a Heuristic Search Procedure (HSP) that guides the search techniques. Five heuristic algorithms, including the genetic algorithm, local search, simulated annealing, random search and tabu search, are employed as the solution methods for finding an optimal set of routes from the huge solution space. For the TRNDP with small network, the exhaustive search method is also used as a benchmark to examine the efficiency and measure the quality of the solutions obtained by using these heuristic algorithms.

Several C++ program codes are developed to implement these algorithms for the TRNDP both with fixed and variable transit demand. Comprehensive experimental networks are used and successfully tested. Sensitivity analyses for each algorithm are conducted and model comparisons are performed. Numerical results are presented and the multi-objective decision making nature of the TRNDP is explored. Related characteristics underlying the TRNDP are identified, inherent tradeoffs are described and the redesign of the existing transit network is also discussed.

ACKNOWLEDGMENTS

The authors recognize that support for this research was provided by a grant from the U.S. Department of Transportation, University Transportation Centers Program to the Southwest Region University Transportation Center which is funded 50% with general revenue funds from the State of Texas.

EXECUTIVE SUMMARY

The transportation system is one of the basic components of an urban area's social, economic and physical structure. As a major part of the urban transportation system, public transit has been widely recognized as a potential way of reducing air pollution, lowering energy consumption, improving mobility, lessening traffic congestion, increasing productivity, providing more job opportunities, promoting retail sales, and rationalizing urban development patterns. In addition to providing mobility for transit-captive users (e.g., people with low incomes, disabled or unable to drive, elderly, children, or those who don't own a car), public transportation also offers meaningful travel alternatives for transit-choice users who might choose transit for the sake of cost, speed, safety, convenience, traffic avoidance, or environmental issues.

However, during the past 30 to 40 years, the bus transit share of total travel has been declining. With suburban sprawl and dispersion of employment, automobile use is challenging public transportation systems. Therefore, maintaining bus transit ridership is a big problem that many bus transit agencies have to face today. Of many methods that have been proposed and/or implemented to expand the transit market, improving the transit level-of-service is a key concept. An operationally and economically efficient bus transit network can help meet these requirements, as well as potentially reduce congestion and conserve energy.

It is generally accepted that the Bus Transit Route Network Design Problem (BTRNDP) should be addressed in the context of bus planning process. Ceder and Wilson (1986), for the first time, defined and presented a conceptual model for the whole bus planning process as a systematic decision sequence, which consists of five levels: network design, frequency setting, timetable development, bus scheduling and driver scheduling. Quite a few past research efforts were devoted to the last two stages: bus scheduling and driver scheduling. However, the critical determinants of system performance from both the operator and user standpoints, are the choices of a bus route network pattern and the corresponding service frequencies. These have received less attention due to their inherent complexity.

Generally speaking, the network design related problem involves the minimization (or maximization) of some intended objective subject to a variety of constraints, which reflect the system performance requirements and/or resource limitations. In the past decade, several people began to realize this bus planning process need and several research efforts have examined the bus transit route network design problem (BTRNDP). However, most of the approaches are still largely dependent on the planners' or researcher's intuition, experience and knowledge about the existing transit network. Furthermore, to make the BTRNDP tractable, many assumptions were made and the problems were over-simplified, making their solutions questionable and therefore preclude them as generally accepted applications for practical transportation networks. To design an optimal bus transit route network that can provide the "best" service given a variety of resource constraints, innovative modeling concepts coupled with scientific tools or systematic procedures are urgently needed.

This research is intended to systematically examine the underlying characteristics of the optimal bus transit route network design problem (BTRNDP). A multi-objective nonlinear mixed integer model is built and the inherent complexity and implementation difficulty are described. Several efficient and flexible heuristic algorithms are employed and compared to come up with an optimal transit route network both with fixed and variable transit demands. Numerical results including sensitivity analyses are presented for comprehensive experimental networks and characteristics underlying the BTRNDP are discussed in details. Summary and conclusions are made and further research directions are also given.

The goal of this research is to develop a flexible algorithmic solution framework to implement the computer-aided design of bus transit route networks and provide various good solutions to accommodate different service requirements. The proposed work in this research is intended to fulfill the following objectives:

1) To identify knowledge that can reflect current related practice and existing rules of thumb for bus transit route network design issues;
2) To develop several robust and systematic efficient heuristic algorithms that can incorporate the above knowledge, and to test a set of designed algorithmic procedures to search intelligently for an optimal solution;
3) To explicitly account for the multi-objective nature of the transit route network design problem and to explore the capability to evaluate various performance measures from the points of view of both the operator and transit users for various service options and to develop the ability to ascertain the built-in characteristics of tradeoffs between various conflicting performance-measure variables involving the bus transit route network problem;
4) To systematically assess various service design concepts for the design and/or redesign for the transit route networks under different scenarios, such as both with fixed and variable transit demands, with and without demand aggregations.

Due to the inherent complexity and combinatorial NP-hard nature of the BTRNDP, traditional exact analytical optimization methodology is impracticable. The proposed work in this research is oriented to developing hybrid heuristic approaches to finding an acceptable and operationally implementable route network and associated service plans that can provide alternative design concepts corresponding to different service requirements in a reasonable time domain. Three algorithmic procedures are developed to provide various service options, namely, the initial candidate route set generation procedure, the network analysis procedure and the heuristic search procedure. The solution methodology differs from existing approaches in many aspects and the expected contributions from this research are summarized as follows:

1) Ability to apply a set of designed algorithmic procedures to search intelligently for an optimal solution without the loss of applicable service planning guidelines and the transit planners' knowledge and expertise;
2) Ability to produce a decent route network reflecting the inherent tradeoffs between conflicting performance-measures. This includes explicit consideration of the multi-objective nature of the bus transit route network design problem and the capability to evaluate performance measures and service options from the points of view of both the operator and transit users;
3) Ability to account for the practical characteristics of real-world transit demand and consider the demand assignment procedure under a microscopic "centroid-connector-link" level with particular concerns for transfer and long-walk related paths;
4) Ability to systematically apply heuristic algorithms to produce quality solutions for the BTRNDP and identify the most appropriate one(s) under certain circumstances;
5) Ability to explore the design and/or redesign for transit route networks with variable transit demand in the context of fixed total travel demand as well as that with fixed demand, and with or without demand aggregation.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE
INTRODUCTION

1.1 Problem Statement and Motivation

The transportation system is one of the basic components of an urban area's social, economic and physical structure. As a major part of the urban transportation system, public transit has been widely recognized as a potential way of reducing air pollution, lowering energy consumption, improving mobility, lessening traffic congestion, increasing productivity, providing more job opportunities, promoting retail sales, and rationalizing urban development patterns. In addition to providing mobility for transit-captive users (e.g., people with low incomes, disabled or unable to drive, elderly, children, or those who don't own a car), public transportation also offers meaningful travel alternatives for transit-choice users who might choose transit for the sake of cost, speed, safety, convenience, traffic avoidance, or environmental issues.

As the most dominant form among all public transportation modes in American cities, bus transit is significant in several aspects. According to the unpublished Transit Fact Book (2002), buses accounted for almost 61% of the 9.4 billion annual U.S. transit trips and about 45% of the 47.7 billion annual transit passenger miles in 2000. They provide service for cities of all sizes, making it an essentially indispensable part of the urban transportation system.

However, during the past 30 to 40 years, the bus transit share of total travel has been declining. With suburban sprawl and dispersion of employment, automobile use is challenging public transportation systems. Therefore, maintaining bus transit ridership is a big problem that many bus transit agencies have to face today. Of many methods that have been proposed and/or implemented to expand the transit market, improving the transit level-of-service is a key concept. An operationally and economically efficient bus transit network can help meet these requirements, as well as potentially reduce congestion and conserve energy.

The bulk of the transportation network research mainly focuses on the automobile (e.g., traffic assignment procedures). However, most of this work is not applicable to the transit industry. The basic difference between private and public transportation can be illustrated by Figure 1.1. Assuming that an acceptable level of service is always maintained and that the supply of the public transit capacity is adequate (i.e., the route frequency is determined by the transit demand on any route), it is expected that as demand increases, the level of service provided by a transit system might improve because lower headways might be provided and therefore, the possibility of the more efficient usage of the transit might be higher. Conversely, the level of service offered to auto users declines as the demand increases due to traffic congestion. Such characteristic of public transit distinguish it from the auto. Therefore it stands out as an urban travel solution that deserves more attention and more research effort.

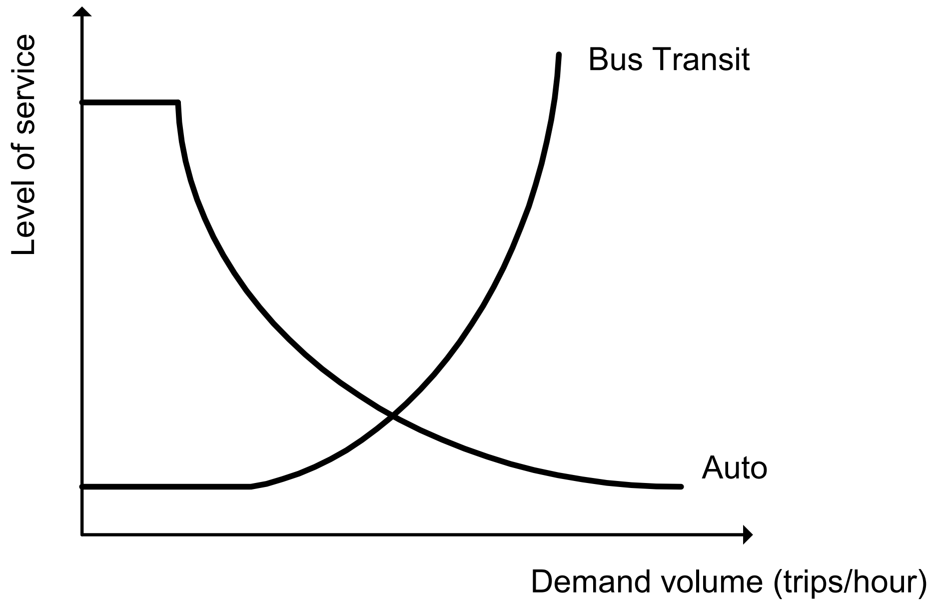


Figure 1.1 Relationship between Level of Service and Demand Volume for Auto and Transit

It is generally accepted that the Bus Transit Route Network Design Problem (BTRNDP) should be addressed in the context of bus planning process. Ceder and Wilson (1986), for the first time, defined and presented a conceptual model for the whole bus planning process as a systematic decision sequence, which consists of five levels: network design, frequency setting, timetable development, bus scheduling and driver scheduling, as shown in Figure 1.2, where the left to right order marks the transition from the highest to the lowest level in the bus planning process. Namely, as illustrated, the output of each level positioned in the left in the sequence becomes an input into lower level decisions on the right. Because the decisions made further down the sequence will have some effects on higher level ones, these levels are not independent and actually interactive, making the feedback in the sequence a repeated process. Furthermore, quite a few past research efforts were devoted to the last two stages: bus scheduling and driver scheduling. This concentration is understandable because in addition to the automation necessity of the scheduling process, these two activities largely affect the operator cost, which includes the drivers' wages, vehicle running and maintenance costs. However, the critical determinants of system performance from both the operator and user standpoints, are the choices of a bus route network pattern and the corresponding service frequencies. These have received less attention due to their inherent complexity.

Targeted to serve centralized core-oriented land user patterns, most traditional bus route networks are either radial or grid-like, providing fixed-route, fixed schedule, uncoordinated service. However, during the past several years, significant spatial redistribution and demographic changes have been taking place in most U.S. cities, making the land-use patterns of cities increasingly decentralized. The changes of population growth and suburbanization have transformed the associated trip distribution patterns from a traditional multiple origin, single destination (CBD) pattern to a multiple origin, multiple destination one. As a result, traditional bus route networks are no longer appropriate for cities with multi-centered and spatially dispersed trip patterns, making the reevaluation and possible redesign of the entire transit route network justified.

Transit authorities have recognized the emerging problems and have made incremental modifications to the traditional transit network. However, due to the absence of systematic procedures, most of these improvements are confined to extensions of old routes to new developing areas and/or discontinuation of service to other areas. Such changes are highly dependent on the transit planners' experience, judgment and knowledge of the existing demand patterns, land use patterns and resource constraints. Furthermore, in most cases, the overall layout and basic structure of the transit route network in most U.S cities remain radial or grid-like, making the service provided neither effective nor efficient. Consequently, user frustration precludes the transit system as a competitive alternative to private automobiles. Furthermore, reliance on the automobile has contributed to a series of problems, including traffic congestion, more fuel consumption, and intensified air pollution. The need for scientific tools or systematic procedures to reevaluation and/or redesign bus transit route networks is thus apparent.

Generally speaking, the network design related problem involves the minimization (or maximization) of some intended objective subject to a variety of constraints, which reflect the system performance requirements and/or resource limitations. In the past decade, several people began to realize this bus planning process need and several research efforts have examined the bus transit route network design problem (BTRNDP). However, most of the approaches are still largely dependent on the planners' or researcher's intuition, experience and knowledge about the existing transit network. Furthermore, to make the BTRNDP tractable, many assumptions were made and the problems were over-simplified, making their solutions questionable and therefore preclude them as generally accepted applications for practical transportation networks. To design an optimal bus transit route network that can provide the "best" service given a variety of resource constraints, innovative modeling concepts coupled with scientific tools or systematic procedures are urgently needed.

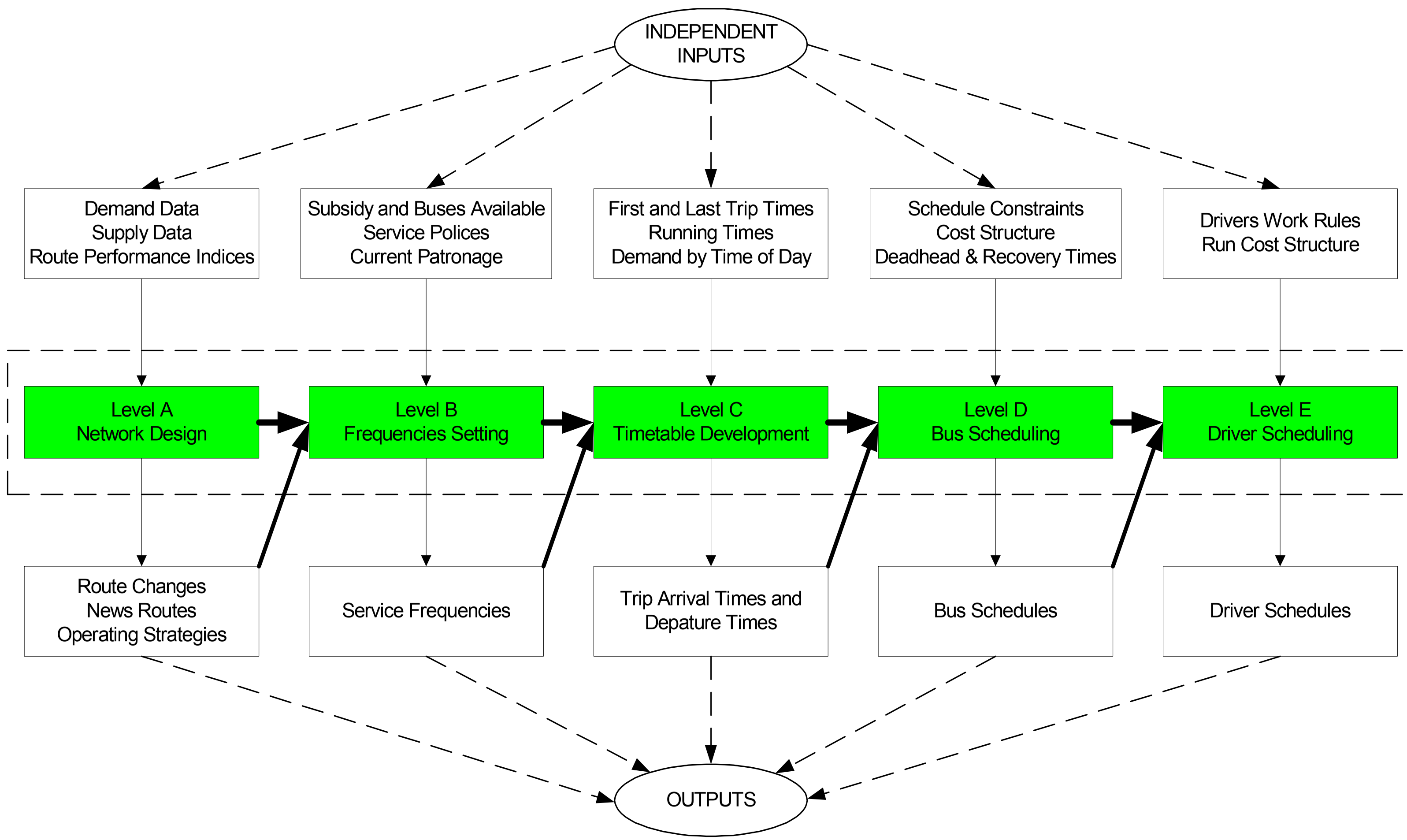


Figure 1.2 Bus Planning Process (adapted and modified from Ceder and Wilson, 1986)

This research is intended to systematically examine the underlying characteristics of the optimal bus transit route network design problem (BTRNDP). A multi-objective nonlinear mixed integer model is built and the inherent complexity and implementation difficulty are described. Several efficient and flexible heuristic algorithms are employed and compared to come up with an optimal transit route network both with fixed and variable transit demands. Numerical results including sensitivity analyses are presented for comprehensive experimental networks and characteristics underlying the BTRNDP are discussed in details. Summary and conclusions are made and further research directions are also given.

1.2 Study Objectives

The goal of this research is to develop a flexible algorithmic solution framework to implement the computer-aided design of bus transit route networks and provide various good solutions to accommodate different service requirements. The proposed work in this research is intended to fulfill the following objectives:

5) To identify knowledge that can reflect current related practice and existing rules of thumb for bus transit route network design issues;
6) To develop several robust and systematic efficient heuristic algorithms that can incorporate the above knowledge, and to test a set of designed algorithmic procedures to search intelligently for an optimal solution;
7) To explicitly account for the multi-objective nature of the transit route network design problem and to explore the capability to evaluate various performance measures from the points of view of both the operator and transit users for various service options and to develop the ability to ascertain the built-in characteristics of tradeoffs between various conflicting performance-measure variables involving the bus transit route network problem;
8) To systematically assess various service design concepts for the design and/or redesign for the transit route networks under different scenarios, such as both with fixed and variable transit demands, with and without demand aggregations.

1.3 Expected Contributions

Due to the inherent complexity and combinatorial NP-hard nature of the BTRNDP, traditional exact analytical optimization methodology is impracticable. The proposed work in this research is oriented to developing hybrid heuristic approaches to finding an acceptable and operationally implementable route network and associated service plans that can provide alternative design concepts corresponding to different service requirements in a reasonable time domain. Three algorithmic procedures are developed to provide various service options, namely, the initial candidate route set generation procedure, the network analysis procedure and the heuristic search procedure.

The solution methodology differs from existing approaches in many aspects and the expected contributions from this research are summarized as follows:

9) Ability to apply a set of designed algorithmic procedures to search intelligently for an optimal solution without the loss of applicable service planning guidelines and the transit planners' knowledge and expertise;
10) Ability to produce a decent route network reflecting the inherent tradeoffs between conflicting performance-measures. This includes explicit consideration of the multi-objective nature of the bus transit route network design problem and the capability to evaluate performance measures and service options from the points of view of both the operator and transit users;
11) Ability to account for the practical characteristics of real-world transit demand and consider the demand assignment procedure under a microscopic "centroid-connector-link" level with particular concerns for transfer and long-walk related paths;
12) Ability to systematically apply heuristic algorithms to produce quality solutions for the BTRNDP and identify the most appropriate one(s) under certain circumstances;
13) Ability to explore the design and/or redesign for transit route networks with variable transit demand in the context of fixed total travel demand as well as that with fixed demand, and with or without demand aggregation.

1.4 Research Overview

The research is structured as shown in Figure 1.3. In this chapter, the significance and the motivation of the optimal transit route network design problem (BTRNDP) has been discussed in the context of bus transit planning activities, followed by descriptions of study objectives and expected contributions.

Chapter 2 presents a comprehensive literature review of previous solution approaches to the BTRNDP primarily in chronological order. Previous approaches that were used to solve the BTRNDP can be classified into three categories: 1) Practical guidelines and ad hoc procedures; 2) Analytical optimization models for idealized situations; 3) Meta-heuristic approaches for more practical problems. In addition, from another perspective, the literature is also summarized according to six distinguishing features: objective function, demand, constraints, decision variables, passenger behavior and solution techniques. Finally, the difficulties in solving the BTRNDP are presented.

Chapter 3 introduces background terminology in the BTRNDP and mathematical notations to be used in the model formulation. A mathematical nonlinear mixed integer programming model for the BTRNDP is formulated in this chapter. Associated constraints and characteristics of the user cost, operator cost and unsatisfied demand cost are also presented. This chapter concludes with discussions of the shortcomings of previous approaches to solve the BTRNDP.

Chapter 4 presents the proposed solution framework for the BTRNDP in this research, which consists of three main components: an Initial Candidate Route Set Generation Procedure (ICRSGP) that generates all feasible routes incorporating some practical guidelines that are commonly used in the bus transit industry; a Network Analysis Procedure (NAP) that computes many performance measures; and a Heuristic Search Procedure (HSP) that guides the search techniques. Different heuristic algorithms including genetic algorithm (GA), local search (LS), simulated annealing (SA), random search (RS) and tabu search (TS) algorithms are proposed to find an optimum set of routes from the huge solution space. For small network, an exhaustive search method (ESM) is also used as a benchmark to examine the efficiency and measure the solution quality obtained from these heuristic algorithms. Two scenarios, namely the BTRNDP with fixed transit demand and the BTRNDP with variable transit demand, are considered. The solution framework and its distinct features are presented, along with discussions of the rationale for choosing the employed heuristic algorithms as the solution techniques. A brief literature review on each of these algorithms is also presented.

Chapter 5 presents the details of the Initial Candidate Route Set Generation Procedure (ICRSGP). A literature review of solution approaches to shortest path algorithm (including label-setting and label-correcting algorithms) and the K-shortest path algorithm is conducted. The chosen Dijkstra's Algorithm and Yen's K-shortest Path for the BTRNDP are then described and the two route feasibility constraints are also discussed. In the end, a small example is introduced to illustrate these two algorithms.

Chapter 6 contains details of the network analysis procedure (NAP) primarily for the BTRNDP with fixed transit demand, which is used to analyze and evaluate the alternative network structures and determine their associated service frequency. Two major components of the NAP, namely, the transit trip assignment model and the frequency setting procedure are presented. The algorithm skeleton and details of its solution methodologies are discussed. Characteristics associated with each component are also described. This chapter concludes with an illustrative application to a transit network example.

Chapter 7 presents details of the BTRNDP with fixed transit demand. Five heuristic algorithms, including genetic algorithm, local search, simulated annealing, random search and tabu search algorithm, as well as the exhaustive search algorithm as benchmark for the BTRNDP with small network, are used to solve the BTRNDP with fixed transit demand. Solution frameworks based on each of these algorithms for the BTRNDP are presented. A small example network using a genetic algorithm as the representative heuristic solution algorithm is introduced for illustrating the proposed methodology. Finally, a summary concludes this chapter.

Chapter 8 presents details of the BTRNDP with variable transit demand. The concepts of variable total demand and variable transit demand are first presented in the urban planning processes. The characteristics underlying the determinants of discrete choice such as the utility and disutility functions are discussed. Multinomial logit model (MNL) and nested logit model (NLM) are compared and their respective pros and cons are given. An innovative two-staged model, consisting of binary logit model-inversely proportional model (BLM-IPM), is proposed to assign the total demand to car and transit mode choice. The solution framework for the BTRNDP with variable transit demand is therefore introduced and the differences between this approach and that in Chapter 7 are discussed. Details of its two components, namely, the ICRSGP and the NAP are presented. Concepts with regard to transit demand equilibrium procedure and headway convergence process are included. Details of the implementation process are also discussed. This chapter concludes with a summary.

Chapter 9 provides details of the algorithm implementation issues. The details of the input data formats, the network representation and the data structure for organizing all network related data using C++ are presented. Three comprehensive experiments (i.e., the BTRNDP for small size, medium size and large scale network) are conducted, where all five heuristic algorithms and the exhaustive search method as a benchmark are employed to solve the BTRNDP in two scenarios, namely the BTRNDP both with fixed and variable demand. Sensitivity analyses are conducted for each heuristic algorithm and related numerical results including the computation of a variety of performance measures and objective functions are presented and compared. A variety of characteristics underlying the multi-objective BTRNDP under different scenarios are therefore described. Effects of the route set size, network size and demand aggregations are presented in detail. Issues of the redesign of the existing transit network are also discussed. Summary and conclusions are given and general guidelines for the TRNDP are also presented.

Chapter 10 concludes with summaries of the proposed algorithms, solution approaches and research results. Suggestions for future research are also provided.

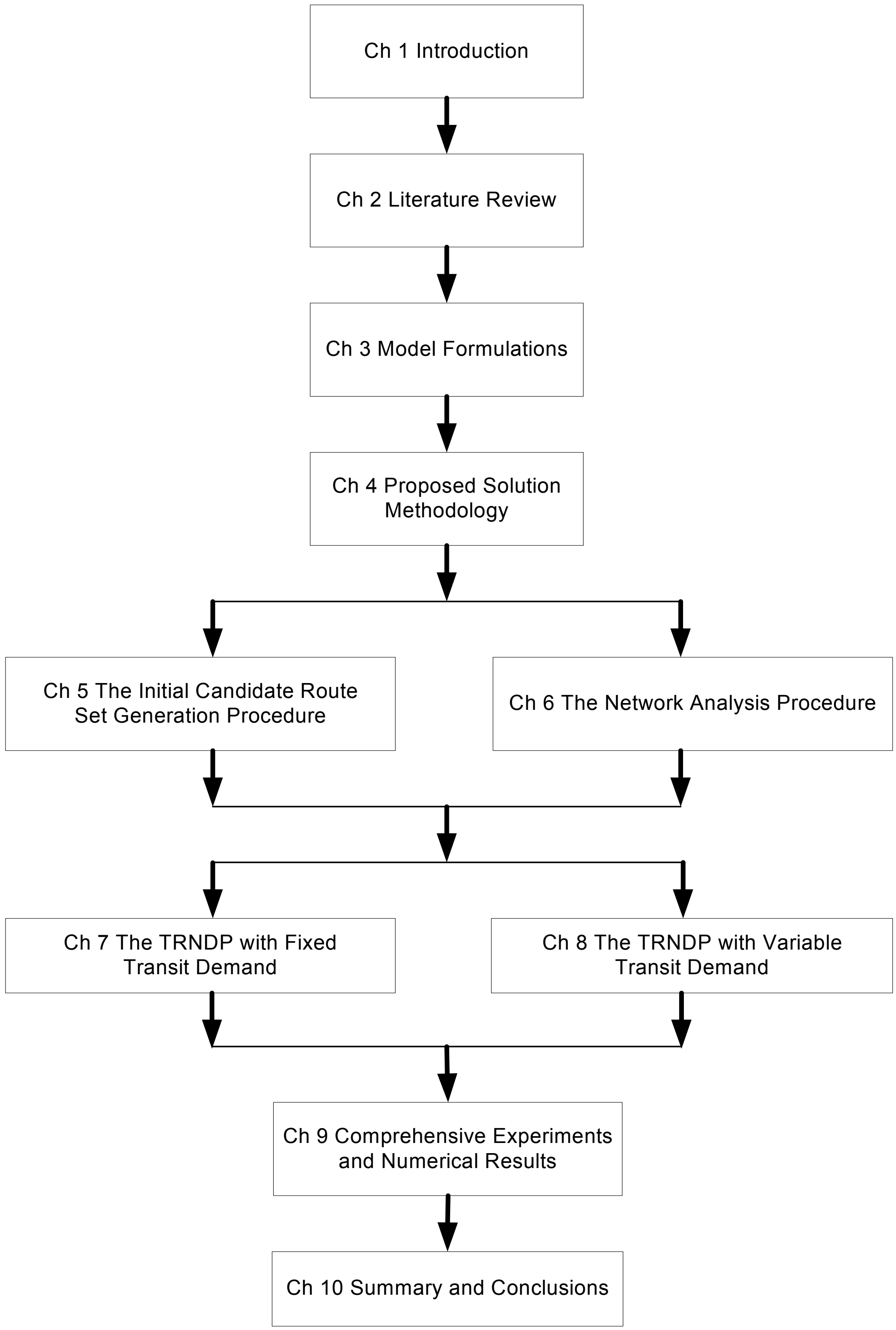


Figure 1.3 Research Structure

CHAPTER TWO
LITERATURE REVIEW

2.1 Introduction

As described before, generally speaking, the bus transit route network design problem involves finding a bus transit route network configuration and associated service frequencies that achieve a desired objective with a variety of given constraints. Many research efforts have examined the BTRNDP. As can be seen, previous approaches that were used to solve the BTRNDP can be classified into three categories: 1) Practical guidelines and ad hoc procedures; 2) Analytical optimization models for idealized situations; 3) Meta-heuristic approaches for more practical problems.

This chapter focuses on the BTRNDP literature review. Several studies deserve particular attention because they provide a solid basis for this research. A comprehensive review of transit network design literature is presented in chronological order. The following sections are organized as follows. Practical guidelines and ad hoc procedures are introduced first in section 2.2, followed by discussions of application of traditional operations research analytical optimization models and meta-heuristics approaches in section 2.3. Section 2.4 presents the summary of the literature review and section 2.5 discusses the difficulties in solving the BTRNDP. Finally, summary and conclusions in section 2.6 complete this chapter.

2.2 Practical Guidelines and Ad hoc Procedures

When dealing with the BTRNDP, researchers have to refer to practical guidelines for the development of operational feasibility constraints. NCHRP Synthesis of Highway Practice 69 (1980) provides suggested rules-of-thumb service-planning guidelines. Adapted from NCHRP 69 for particular use for the transit route network design problem, Table 2.1 shows selected service-planning guidelines with regard to service patterns and service levels. Regarding service patterns, important practical guidelines include service area and route coverage, route structure and spacing, route directness–simplicity, route length, and route duplication. In terms of service levels, guidelines consist of service period, policy headways desirable and minimum service frequency, loading standards, and road speeds. Based on the professional judgment and extensive practical experience of operations practitioners in the transit agencies in many Canadian and U.S. cities, these guidelines emphasize practice and short-term transit planning rather than theory and long-range planning.

Table 2.1 provides a summary of practical service planning guidelines that the operations planner can apply to the BTRNDP. Although obeying these guidelines is not sufficient to provide an optimal solution to the transit route network problem, violations of them may result in ineffective or even infeasible operations. On the other hand, however, it is also widely accepted that one cannot ensure that this approach can always yield a satisfactorily operationally and economically efficient bus transit route network only with the help of scientific tools or systematic procedures. In recent years, many researchers and transit planners began to realize this problem.

Recently, many research efforts have been focused on the applications of operations research and/or mathematics to the BTRNDP. Generally speaking, considering whether route network configurations and their associated frequencies (or other related design parameters) are determined in a sequential or in a joint manner, these approaches can be categorized as analytical optimization models and meta-heuristic approaches. The following chapters will focus on the discussion of these two methods.

2.3 Analytical Optimization Models and Meta-heuristic Approaches

Tradition Operations Research (OR) analytical optimization models were used in the early stages of the research on transit route network design problems. Rather than determine both the route structure and design parameters simultaneously, these analytical optimization models were primarily applied to determine one or several design parameters (e.g., stop spacing, route spacing, route length, bus size, frequency of service) on a predetermined transit route network structure. The Meta-heuristic approaches primarily dealt with simultaneous design of the transit route network and determination of its associated service frequencies.

Examples of this traditional operations research analytical optimization model can be seen in the work of Newell (1979), Oldfield and Bly (1988) and Leblanc (1987). Generally speaking, these analytical optimization models are very effective in solving optimization-related problems for networks of small size or with one or two decision variables. However, when it comes to the transit route design problem for a network of realistic size in which many parameters need to be determined, this approach does not work very well. Due to the inherent complexity involved in the BTRNDP and to deal with it from a systematic point of view, the meta-heuristic approaches, which do not guarantee to find the global optimal solution, were therefore proposed. The following sections present the literature on these two methods in chronological order.

1. SERVICE PATTERN
   1.1. Service Area and Route Coverage
      a. Service area is defined by operating authority or agency.
      b. Provide 1/4-mile coverage where population density exceeds 4,000 persons per sq mile or 3 dwelling units per acre. Serve at least 90 percent of residents.
      c. Provide 1/2-mile coverage where population density ranges from 2,000 to 4,000 persons per mile (less than 3 dwelling units per acre). Serve 50 to 75 percent of the population.
      d. Serve major employment concentrations, schools, and hospitals.
      e. Serve area within two-mile radius of park-and-ride lot.
   1.2. Route Structure and Spacing
      a. Fit routes to major streets and land use patterns; provide basic grid system where streets form grid; provide radial or radial-circumferential system where irregular or radial street patterns exists.
      b. Space routes at about 1/2-mile in urban areas, 1 mile in low-density suburban areas, and closer when terrain inhibits walking.
   1.3. Route Directness --- Simplicity
      a. Routes should be direct and avoid circuitous routings. Routes should be not 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 be not more than two branches per trunk-line route.
   1.4. Route Length
      a. Routes should be as short as possible to serve their markets; excessively long routes should be avoided. Long routes require more liberal travel times because of the difficulty in maintaining reliable schedules.
      b. Route length generally shall not exceed 25 miles round trip or 2 hours.
      c. Two routes with a common terminal may become a through route if they have more than 20 percent transfers and similar service requirement, subject to (b). This usually results in substantial cost savings and reduces bus movements in the central business district.
   1.5. Route Duplication
      a. There should be one route per arterial except on approaches to the CBD or a major transit terminal. A maximum two routes per street (or two branches per route) is desirable.
      b. Express service should utilize freeways or expressways to the maximum extent possible.
      c. Express and local services should be provided on separate roadways, except where frequent local service is provided.

2. SERVICE LEVELS
   2.1. Service Period
      a. Regular service: 6 a.m. to 11 p.m./midnight, Mon.-Fri.
         Priorities: weekday commuter, 6-10 a.m. and 3-7 p.m.; weekday, 6 a.m. -7 p.m.;
         Saturday, 7 a.m. -7 p.m.; Evenings, 7 p.m.-midnight; Sundays, 9:00 a.m. –7:00 p.m.
      b. Owl service: selected routes, large cities-24hr.
      c. Suburban feeder service: weekdays, 6-9a.m., 4-7 p.m.
         (Some services 6 a.m. to 7 p.m.)
      d. Provide Saturday and Sunday service over principle routes except in smaller communities, where Sunday service is optional.
   2.2. Policy Headways Desirable --- Minimum Service Frequency
      a. Peak: 20 minutes - urban; 20-30 minutes - suburban.
      b. Midday: 20 minutes - urban; 30 minutes – suburban.
      c. Evening: 30 minutes - urban; 60 minutes – suburban.
      d. Owl: 60 minutes.
   2.3. Loading Standards
      a. Peak 30 minutes: 150 percent.
      b. Peak hour: 125-150 percent.
      c. Transition period: 100-125 percent.
      d. Midday/evening: 75-100 percent.
      e. Express: 100-125 percent.
      f. Suburban: 100 percent.
      Note: Policy headway may result in considerably lower load factors.
   2.4. Road Speeds
      a. Central area: 6-8 mph.
      b. Urban: 10-12 mph.
      c. Suburban: 14-20 mph.

Table 2.1 Suggested Service Planning Guidelines (Selected from NCHRP 69, 1980)

2.3.1 Lampkin and Saalmans's research work

Lampkin and Saalmans (1967) uncoupled the design of transit routes and the setting of frequencies and tackled them separately. In this paper, they proposed an optimization model to determine the routes of the networks first and then to assign frequencies to the generated set of routes in a second stage. In the first phase, a heuristic algorithm was developed to design the transit network in attempt to transport a maximum number of passengers given a fixed OD-matrix while considering trips without transfers (i.e., to optimize the passenger-kilometer criteria). The objective of the second stage is to allocate service frequencies to the already-generated routes so that the total travel time was minimized given the available number of vehicles. Employing a sequential rather than a simultaneous approach, some issues regarding construction of timetables and the assignment of individual buses to journeys were also discussed using a conventional method and a linear programming model respectively.

2.3.2 Rea's research work

Different than any of previous or later approaches, Rea (1971) didn't specify an objective function, but sought a "satisfactory" solution that meets certain operator specified performance levels. A service specific model was developed to determine which links should be used to construct routes for a public transport network. In this model, Rea employed single path assignment, assuming that all passengers traveling between a specific origin-destination pair use the single least weighted cost path under a fixed transit demand context.

2.3.3 Silman, Barzily, and Passy's research work

Silman, Barzily, and Passy (1974) presented a planning method for urban bus route systems, trying to minimize the sum of journey time (included allowance for transfer times) and discomfort penalties proportional to the number of passengers who cannot find seats. A two-staged approach was employed. First, the candidate routes set was constructed through several repetitions of a route addition and deletion process. Second, the optimal frequencies for a set of already-generated routes were determined by a gradient method under the constraints of a given number of available buses. It was also pointed out that the optimal value of the objective function in the second phase served as a more accurate evaluator of a generated set of routes.

2.3.4 Mandl's research work

Mandl (1979) employed a single path assignment method, assuming that all passengers traveling between a specific origin-destination pair use the single least weighted cost path. In this work, Mandl presented a heuristic algorithm to find the optimal routes. Also, Mandl assumed a constant frequency on all bus routes (policy headway), which made the BTRNDP a much simpler problem. The optimization process was described as follows. A set of feasible routes was examined and possible reductions in the average cost using this set were attempted. The new set was compared with the older one on the basis of performance and if found better, it was accepted and the search procedure repeated until no new improvements could be found.

2.3.5 Dubois, Bell, and Llibre's research work

Dubois, Bell, and Llibre (1979) categorized the network generation problem into three sub-problems, i.e., to choose a set of streets, to choose a set of bus lines, and to determine optimal frequencies. Firstly, they proposed a heuristic model to find an optimal subset of streets, intending to minimize the total travel time under an investment constraint. Then, a model was presented to find the optimal bus lines given the chosen street subset. In the last stage, optimal frequencies for the lines of the chosen network were determined in a variable demand formulation context in which the total trip matrix was estimated first, and then a diversion curve based on expected travel times was used to estimate the public transit share from the already predicted total trip matrix. Namely, the transit demand between any origin-destination pair was treated as a variable that responded to the network design solution. The work included transit trips that required transfers.

2.3.6 Newell's research work

Newell (1979) discussed some issues relating to the optimal design of bus routes. He pointed out the nonconvexity of an objective function designed to minimize the total cost. He noted that the higher the demand for trips on a route, the larger is the quantity and quality of service that one can provide and this distinguishes the bus transit demand assignment process from that of the automobile demand assignment process. Theoretical analyses of the bus route design were presented but no real-world applications were provided.

2.3.7 Hasselstrom's research work

Hasselstrom (1981) sought to determine a set of optimal bus routes and frequencies simultaneously with maximization of the consumer's surplus in the context of variable demand formulations. To obtain this goal, Hasselstrom employed a complex two-level optimization model to generate routes by initially assigning desired trips onto a network of all possible transit links, then forming routes using normal practice criteria. A direct model based on the multiple path assignment method was used to estimate a demand matrix that could provide service of high quality throughout the area although it offered less than ideal service between some origin-destination pairs. Note that the disadvantage of the models presented in this work was that although the bus routes and frequencies were determined simultaneously, two different optimization problems had to be formulated.

2.3.8 Ceder and Wilson's research work

As mentioned in Chapter 1, Ceder and Wilson (1986) placed the bus network design activity in the context of other bus service functions including setting frequencies, timetable development, bus scheduling and driver scheduling. A two level methodological approach was presented for the design of the bus route network, in which the first level considered only the passenger viewpoint and the second level accounted for both the passenger and operator viewpoint. Two constraints that were considered included minimum frequency and fleet size. The first level was handled by an optimization program while the second level relied on heuristic techniques.

Very similar later papers include those from Israeli and Wilson (1991) and Ceder and Israeli (1998). In the second paper, they proposed a nonlinear mixed integer programming model and formulated the minimization of generalized cost and fleet size as their two-level objective functions. An algorithmic method consisting of seven components was presented to solve this two-staged model. Assessment of the performance of the existing transit network was achieved using a multi-objective programming approach from the aspects of operator efficiency and passenger level of service. Finally, a small example was suggested for the proposed algorithm, but was not actually applied.

2.3.9 Leblanc's research work

Leblanc (1987) formulated a transit network design model for determining the frequencies of existing transit routes. A refinement to the conventional mode-split assignment model was introduced, in which transit frequencies for each distinct transit line were used to calculate transit access time and transfer times. The author pointed out that preliminary computational testing of this model showed that it can predict mode choices and link flows more accurately.

2.3.10 Van Nes, Hamerslag and Immers's research work

Van Nes, Hamerslag and Immers (1988) reviewed existing design methods and existing optimization methods and proposed a model to design the public transport network while trying to maximize the number of direct trips given a certain fleet size. Furthermore, the relation between supply and demand for public transport services was examined and a direct demand model was used based on the simultaneous distribution-modal split model for a variable transit demand formulation. A heuristic solution method was presented so that the simultaneous selection of routes, assignment of frequencies and the determination of the number of passengers were achieved in this research work.

2.3.11 Baaj, Shin and Mahmassani's research work

Baaj and Mahmassani (1990, 1991 and 1995) proposed an AI-based solution approach that consists of three major components. The first part was a route generation algorithm (RGA) that generated different sets of routes corresponding to different tradeoffs between user cost and operator cost. The second part was a transit route network analysis procedure (TRUST) to evaluate a given transit route network as well as to set its associated route frequencies. The third part was a route improvement algorithm (RIA) to improve the initially generated sets of routes so as to obtain feasible and implementable route networks. The code language used for development was LISP.

2.3.12 Shin and Mahmassani's research work

Shin and Mahmassani (1994 and 1998) proposed essentially the same approach as that of Baaj, in which an artificial intelligence (AI)-based search approach guided by expert knowledge was used to solve the bus transit route network design problem. Realizing the importance of addressing the multi-objective nature of the problem, Baaj and Shin selected different weights to examine the total travel time (including walk time, wait time, in-vehicle travel time and transfer-related time), total demand satisfied and the required fleet size to operate the transit system. The approaches employed in these two works consisted of three major procedural components: a route generation procedure, a network evaluation procedure, a transit center selection procedure and a network improvement procedure. The difference between Shin's work and Baaj's work was primarily that the later tackled the bus transit network design problem mainly on conventional service concepts (namely, providing fixed route, fixed schedule and uncoordinated systems, with the same vehicle size on all routes) while the former incorporated three additional service design concepts including route coordination, variable vehicle size, and demand responsive service. Shin's two works provided some insights to the bus transit route network design problem. However, these works are heavily dependent on the experience and judgment of the transit planners and knowledge of the existing demand patterns, land use patterns and resource constraints, preclude these works becoming viable in real-world applications.

2.3.13 Constantin and Florian's research work

Constantin and Florian (1995) considered the problem of optimizing frequencies of transit lines in a transit network with a goal to minimize the total expected travel and waiting time on the network while satisfying fleet size constraints. A nonlinear nonconvex mixed integer model was formulated first and was then converted into a bi-level Min-Min nonconvex optimization problem and a projected (sub)gradient algorithm was used to solve this problem. Some computational results were obtained from an example city transit network.

2.3.14 Pattnaik, Mohan and Tom's research work

Pattnaik, Mohan and Tom (1998) formulated the urban bus route network design problem with fixed transit demand as an optimization problem of minimizing the overall cost (i.e., the sum of user total travel time cost and operator cost). Constraints included frequency and load factor feasibility. In this paper, a genetic algorithm based model was presented to solve the route network design problem so that a route configuration with a set of transit routes and associated frequencies could be determined simultaneously. The solution methodology consisted of two phases. First, a set of candidate routes competing for the optimum solution was generated. Second, the optimum set was evaluated and the best solution was chosen using a GA.

2.3.15 Yang, Chien and Hou's research work

Yang, Chien and Hou (1999) considered the problem of determining an optimal feeder bus route in attempt to minimize the sum of user and supplier costs subject to geographic, capacity, and budget constraints. It was pointed out that as the number of the links increases, the number of feasible bus routes increases dramatically, making the transit route network design problem computationally intractable for realistic urban networks. Therefore, a genetic algorithm was chosen as the solution methodology to determine optimal bus route locations and headways. The presented examples demonstrated that the genetic algorithm could provide a solution of good quality.

2.3.16 Lee and Vuchic's research work

Lee and Vuchic (2000) stated that transit demand should depend on the network configuration and associated route frequencies. Therefore, they presented an iterative approach to tackle the dynamic characteristics of the transit route network design problem. The objective of this approach was to minimize user total travel time subject to constraints of frequency on each route. To simultaneously estimate the transit demand and generate the optimal transit network, a modal split modeling procedure was proposed. Furthermore, adapted from the solution approach used by Rea (1971), an AI-based heuristic method was presented to solve the transit network design problem along with considerations of variable transit demand under a fixed total travel demand. Sensitivity analyses examining the relationship between the optimal transit network and the design inputs were also presented.

2.3.17 Ngamchai and Lovell's research work

Ngamchai and Lovell (2001) studied an optimal timed-transfer bus transit route network design problem. A model was formulated as the minimization of total cost (including total user cost and operator cost) to optimize the bus transit route configuration and associated service frequencies. The procedure consisted of three components including a route generation algorithm, route evaluation algorithm and route improvement algorithm. Several genetic operators were developed for the genetic algorithm, which was the proposed solution approach. Results were presented by implementing the GA for a small example network.

2.4 Summary of the literature review

One can see from the above literature review that essentially speaking, there are six distinguishing features that characterize the transit route network design problems: objective function, demand, constraints, decision variables, passenger behavior, and solution techniques. The following provides another perspective of the previous literature: 1) <u>Objective Functions</u>: The most widely used objective function is minimization of generalized cost (or time) or maximization of consumer surplus; 2) <u>Demand</u>: For simplicity, most researchers pursue the optimal transit network with fixed demand. However, since transit demand is largely dependent on the transit route network structure and its associated frequencies, the BTRNDP with variable demand is preferred and considered in this research although computational complexity is added; 3) <u>Constraints</u>: Feasibility constraints often include: minimum operating frequencies; a maximum load factor; a maximum allowable bus fleet size; maximum and minimum limits on route lengths; maximum number of routes; minimum route ridership volumes; restriction to headways (policy headway); and/or constraints of the directness of routes as measured by circuity factors; 4) <u>Decision Variables</u>: Network route configuration and service frequency are the decision variables; 5) <u>Passenger Behavior</u>: Previous transit trip assignment models can be divided into two groups, namely, single path assignment and multiple path assignment models. Another aspect of passenger behavior that is also of interest is the passenger willingness to make transfer trips. In addition to the extra passenger waiting time incurred as a result of transfers, a transfer penalty is also often incorporated; 6) <u>Solution Methodology</u>: Most approaches rely on practical guidelines and are guided by genetic algorithm or artificial intelligence (AI)-based search procedures.

In conclusion, Table 2.2 presents a summary of the existing transit route network design models. Note that the following notations are introduced for convenience of the descriptions.

*: No explicit objective function, but solutions that meet certain operator-specified performance measure requirements are generated
**: Final bus routes are determined manually
***: Problem formulation only
****: Multi-objective approach is considered and generated solutions can reflect different tradeoffs among conflicting objectives
AI: artificial intelligence approach
GA: genetic algorithm approach

Year	Author	Objectives	Fixed vs. Variable Demand	Single vs. Multiple path assignment	Decision Variables	Solution Techniques
1967	Lampkin and Saalmans	Generalized time	Fixed	Multiple	Route and Frequency	Sequential
1972	Rea	*	Fixed	Single	Route and Frequency**	Sequential
1974	Silman, Barzily, and Passy	Generalized cost	Fixed	Multiple	Route and Frequency	Sequential
1979	Mandl	Generalized time	Fixed	Single	Route	Sequential
1979	Dubois, Bell, and Llibre	Generalized time	Fixed	Multiple	Route and Frequency	Sequential
1979	Newell	Generalized cost	Fixed	Single	Route	Sequential
1981	Hasselstrom	Consumer surplus	Variable	Multiple	Route and Frequency	Simultaneous + AI
1986	Ceder and Wilson***	Generalized time	Fixed	Multiple	Route and Frequency	Sequential
1987	Leblanc	Min. operator's cost & Max. transit usage	Fixed	Single	Frequency	Sequential
1988	Van Nes, Immers, and Hamerslag	Number of direct trips	Variable	Multiple	Route and Frequency	Simultaneous + AI
1990	Baaj and Mahmassani	****	Fixed	Multiple	Route and Frequency	Sequential+AI
1991	Israeli and Wilson	Generalized time, and fleet size	Variable	Multiple	Route and Frequency	Sequential
1994	Shin and Mahmassani	****	Fixed	Multiple	Route and Frequency	Sequential+AI
1995	Constantin and Florian	Minimize the total travel and waiting time	Fixed	Single	Frequency	Sequential
1998	Pattnaik, Mohan and Tom	Generalized cost	Fixed	Multiple	Route and Frequency	Simultaneous + GA
1999	Yang, Chien and Hou	Generalized cost	Fixed	Single	Route and Frequency	Simultaneous + GA
2000	Lee and Vuchic	Generalized cost	Variable	Multiple	Route and Frequency	Simultaneous + AI
2001	Ngamchai and Lovell	Generalized cost	Fixed	Multiple	Route and Frequency	Simultaneous + GA

Table 2.2 Summary of Transit Network Design Models

2.5 Difficulties in solving the BTRNDP

The literature describing previous solution approaches to the BTRNDP has been reviewed. As partly mentioned by several researchers (e.g., Baaj, 1990) and noticed by the authors, six main sources of complexity often preclude finding a unique optimal solution for the Bus Transit Route Network Design Problem (BTRNDP):

(1) Great difficulty in defining the decision variables and expressing the objective function accordingly;
(2) Non-convexities and Non-linearities are involved in the cost associated with the transit network configuration;
(3) Combinatorial complexity arises from the discrete nature of the route design problem, making the BTRNDP an NP-hard problem;
(4) Many important tradeoffs among conflicting objectives need to be addressed, making the BTRNDP an inherently multi-objective problem;
(5) Spatial layout of routes makes it very hard to design an acceptable and operationally feasible set of routes with the need to address many important design criteria;
(6) The nature of variable transit demand even with a given total travel demand makes the already-difficult BTRNDP more complex.

These sources of complexity render the solution search space computationally intractable and the computational burden of the problem grows exponentially with the size of the bus transit network. Figure 2.1 presents a graphical representation of global and local optimum for the bus transit route network design problem. It can be seen that as the solution space for bus transit route network increases, the number of local optima grows. The inherent nonlinear mixed integer and discrete properties make the BTRNDP a really complex combinatorial optimization and an NP-Hard problem. Therefore, traditional optimization approaches are unable to solve such problems, necessitating more intelligent strategies and scientific tools. Heuristic algorithms that are designed by many researchers are a necessary approach for solving the BTRNDP. In this context, a hybrid heuristic algorithm-based solution methodology for the BTRNDP is proposed in the next chapter.

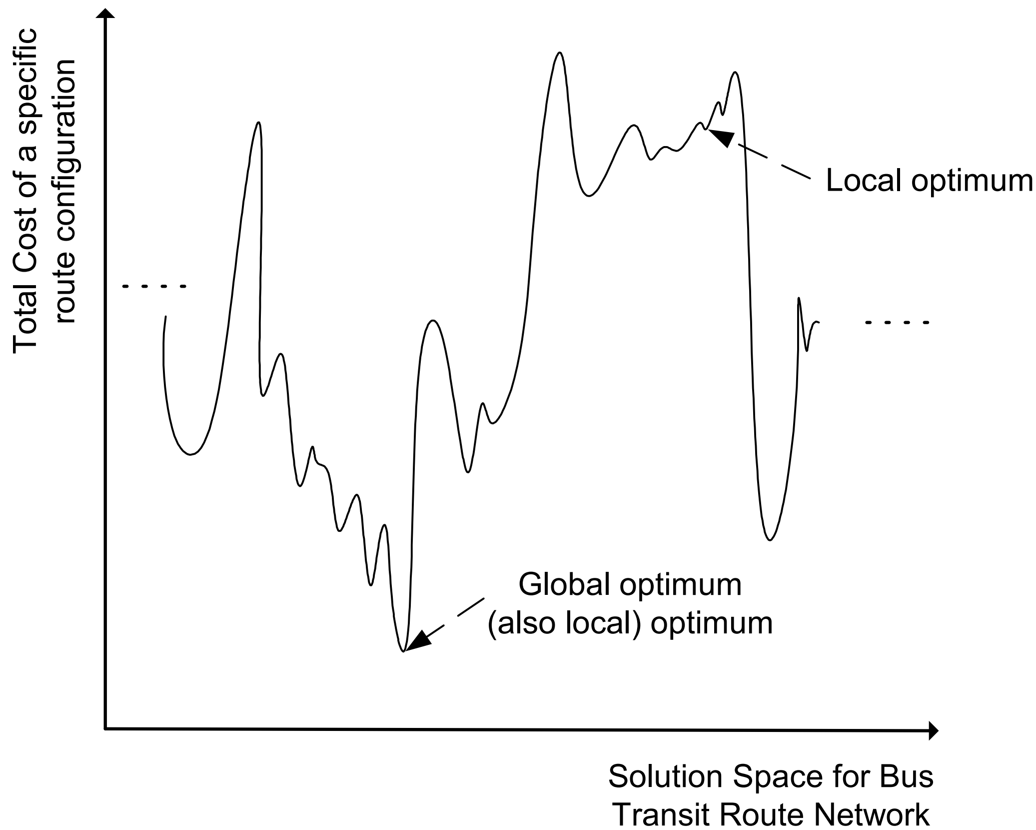


Figure 2.1 Global and Local Optimum for the Transit Route Network Design Problem

2.6 Summary and Conclusions

Previous approaches to the transit route network design problem can be classified into three categories, practical guidelines and ad hoc procedures, analytical optimization models for idealized situations, and meta-heuristic approaches for more practical problems. Analytical optimization models are primarily applied to determine certain design parameters on a predetermined transit route network structure. To determine both the route structure and design parameters simultaneously, meta-heuristic approaches for practical problems are utilized. A comprehensive literature review of previous analytical optimization models and meta-heuristic approaches is conducted in chronological order and categorized into six distinguishing features that characterize the optimal transit route network design problem: objective function, demand, constraints, decision variables, passenger behavior, and solution techniques.

The difficulties in solving the BTRNDP are presented. Six significant sources of complexity often preclude finding a unique global optimal solution for the BTRNDP and make this problem an NP-Hard one. Due to its inherent complexity, the heuristic algorithms are identified as the necessary approaches for solving the BTRNDP.

The research is intended to model the transit route network design problem and propose an appropriate heuristic algorithm-based solution framework to overcome the above difficulties. The next chapter presents the model formulations for the BTRNDP, followed by introduction of the solution methodology to be used in this research.

CHAPTER THREE
MODEL FORMULATIONS

3.1.Introduction

Generally speaking, the optimal transit route network design involves determining a network configuration with a set of transit routes and associated service frequencies that achieve a desired objective with a variety of given constraints.

Basically, the problem to be addressed can be defined in general terms as follows. For this initial discussion, the transit demand matrix is assumed to be fixed, that is, not dependent on the route structure. Therefore, given the transit demand matrix, the highway network and spatially represented zone system in a certain city, one must find a set of routes that correspond to chosen tradeoffs between user cost, operator cost and unsatisfied demand cost. Operator cost refers to the cost of operating the number of buses used in the transit network systems and user cost consists of four components, including walking cost (the cost of transit users' access from and to the bus stops), waiting time (cost of transit users' waiting for a bus), transfer cost, and in-vehicle time (time needed for transit users' riding a bus from the origin bus stop to the bus stop near their intended destination). The unsatisfied demand cost is the penalty set for those passengers who cannot be served by the proposed transit route network. As a result, solution to the problem requires design of a system of bus routes and selection of a set of bus frequencies on each route.

This chapter is focused on the model formulation of the BTRNDP and is organized as follows. First, definitions of terms and notations are introduced in section 3.2. Second, assumptions of the BTRNDP in this research are presented in section 3.3. Third, mathematical notations, objective function and constraints are proposed and characteristics underlying the model formulation are described respectively in section 3.4, 3.5 and 3.6. Fourth, shortcomings of previous approaches are then pointed out in section 3.7 and a summary in section 3.8 concludes this chapter.

3.2.Definitions of Terms and Notations

Essentially speaking, the public transportation system is described in terms of "nodes", "links" and "routes". In this section, terms and notations are defined and explained in order to provide a basis for the model formulation of the BTRNDP and the subsequent chapters.

A <u>node</u> is used to represent a specific point for loading, unloading and/or transfer in a transportation network. Generally speaking, there are three kinds of nodes in a bus transit network system: (a) Nodes representing centroids of specific zones; (b) Nodes representing road intersections; and (c) Nodes with which zone centroid nodes are connected to the network through centroid connectors (called "distribution nodes"). Note that nodes could be real identifiable on the ground or fictitious. For example, nodes of the second kind are real while the first and the third are fictitious (these two are also user-defined). Furthermore, the term "distribution nodes" is introduced especially for the third kind of node in this research. A link joins a pair of nodes and represents a particular mode of transportation between these nodes, which means that if different modes of transportation are involved with the same link, these are represented as two links, say walk mode and transit mode. This is natural since the travel time associated with every mode specific link is different. A route is a sequence of nodes. Every consecutive pair of the sequence must be connected by a link of the relevant mode. The bus line headway on any particular route is the inter-arrival time of buses running on that route. A graph (network) refers to an entity G = {N, A} consisting of a finite set of N nodes and a finite set of A links (arcs) which connect pairs of nodes. A transfer path is a progressive path that uses more than one route.

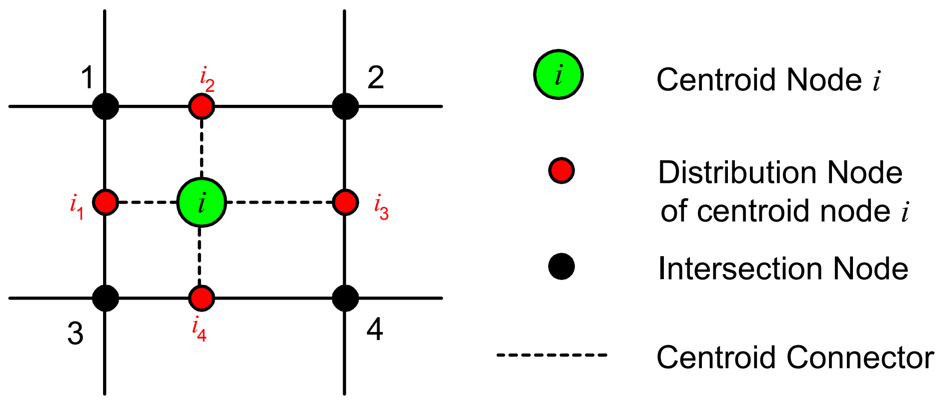


Figure 3.1 Graphical Representations for Nodes, Links and Zones

Figure 3.1 gives a graphical representation for nodes, links and zones. In this graph, links refer to any pair of nodes except the centroid nodes, (e.g., section from 1 to i_1 is a link). The intersection nodes always refer to the road intersections. The region that is encompassed by nodes 1, 2, 3 and 4 is called a traffic "zone". Note that a typical geographical zone system may be based upon census boundaries to furnish zone boundaries on the arterial street system and all land areas are encompassed by streets or major physical barriers. The zone centroids are located somewhere near the centers of the zones and zone connectors are used to connect these centroids to the modeled network. Generally, the Centroid node represents the "demand" center (origin and/or destination) of a specific traffic zone. Distribution nodes are the junctions of centroid connectors and road links (e.g., i_1, i_2, i_3 and i_4). It should be pointed out that centroid connectors are usually fictitious and are used only for the sake of convenience for network modeling. In addition, it is the distribution nodes that are used as the origins and/or destinations for implementation of the shortest path algorithm and k shortest path algorithm. Distribution nodes might physically represent bus stops. Furthermore, an important characteristic of these centroid connectors is the distances that transit users have to walk to get to the routes that provide service to their intended destinations. Note that the terms, "arc" and "link" are used interchangeable.

3.3.Assumptions of the BTRNDP

For simplicity, the following assumptions are made for the BTRNDP in this research:

1. For the BTRNDP with fixed demand, the bus transit demand pattern (i.e., transit O-D table) is static and constant for all periods of study, while for the BTRNDP with variable demand, the total travel demand is fixed but transit demand is variable. In any case, bus transit demand and/or total travel demand is symmetric (namely, O-D matrices are square and symmetric) in the studied system.
2. Every bus in the transportation system travels with constant speed on all routes. This means that passenger in-vehicle travel time is not subject to the traffic conditions on the roads composing the bus transit network.
3. Passenger access times to bus stops are independent of the transit network configurations.
4. All buses have the same capacity and the same load factor.
5. Transfers can take place at any distribution node and any intersection node.

3.4.Model Formulation

Consider a connected network composed of a directed graph G = {N, A} with a finite number of nodes, N connected by A arcs. The following notations are used.

<u>Sets/Indices:</u>

$i, j \in N$	Centroid nodes
$r_k \in R$	Routes
$i_t \in N$	t-th distribution node of centroid node i
$tr \subset R$	transfer paths that use more than one route from R

<u>Data:</u>

R_{max} = maximum allowed number of routes for the route network;

N = number of centroid nodes in the route network;

D_{max} = maximum length of any route in the transit network;

D_{min} = minimum length of any route in the transit network;

d_{ij} =bus transit travel demand between centroid nodes i and j;

h_{max} =maximum headway required for any route;

h_{min} =minimum headway required for any route;

L_{max} =maximum load factor for any route;

P =seating capacity of buses operating on the network;
W =maximum bus fleet size available for operations on the route network;
V_b = travel speed of buses on the route network;
V_w = passenger walking speed;
C_v = per-hour operating cost of a bus; ($/vehicle/hour)
O_v = operating hours for the bus running on any route; (hours)
C_1, C_2, C_3 = weights reflecting the relative importance of three components including the user costs, operator costs and unsatisfied demand costs respectively;

Decision Variables:
M = the number of routes of the current proposed bus transit network solution;
r_m = the *m*-th route of the proposed solution, $m = 1,2,\cdots,M$;
D_{r_m} = the overall length of route r_m;
$d_{ij}^{r_m}$ =the bus transit travel demand between centroid nodes *i* and *j* on route r_m;
d_{ij}^{tr} =the bus transit travel demand between centroid nodes *i* and *j* along transfer path *tr*;
DR_{ij} = the set of direct routes used to serve the demand between centroid nodes *i* and *j*;
TR_{ij} = the set of transfer paths used to serve the demand between centroid nodes *i* and *j*;
$t_{ij}^{r_m}$ = the total travel time between centroid node *i* and *j* on route r_m;
t_{ij}^{tr} = the total travel time between centroid node *i* and *j* along transfer path *tr*;
h_{r_m} = the bus headway operating on route r_m; (hours/vehicle)
L_{r_m} = load factor on route r_m;
T_{r_m} = the round trip time of route r_m; $T_{r_m} = \dfrac{2D_{r_m}}{V_b}$;
N_{r_m} = the number of operating buses required on route r_m; $N_{r_m} = \dfrac{T_{r_m}}{h_{r_m}}$;
Ω_{r_m} = set of links (a, b) assigned to route r_m;
$Q_{r_m}^{\max}$ = the maximum flow occurring on the route r_m;
$$\beta_{ijab}^{r_m} = \begin{cases} 1, \text{if } d_{ij}^{r_m}, \text{O-D flow on route } r_m, \text{ will travel over link (a,b)} \\ 0, \text{otherwise} \end{cases}$$

Objective Function:

The objective is to minimize the sum of user cost, operator cost and unsatisfied demand costs for the studied bus transit network. The objective function is as follows:

$$\min z = \mathrm{C}_1 * (\sum_{i\in N}\sum_{j\in N}\sum_{r_m\in DR_{ij}} d_{ij}^{r_m} t_{ij}^{r_m} + \sum_{i\in N}\sum_{j\in N}\sum_{tr\in TR_{ij}} d_{ij}^{tr} t_{ij}^{tr}) + \mathrm{C}_2 * \mathrm{C}_\mathrm{v} * \mathrm{O}_\mathrm{v} * (\sum_{m=1}^{\mathrm{M}} \frac{T_{r_m}}{h_{r_m}})$$

$$+ \mathrm{C}_3 * (\sum_{i\in N}\sum_{j\in N} d_{ij} - \sum_{i\in N}\sum_{j\in N}\sum_{r_m\in DR_{ij}} d_{ij}^{r_m} - \sum_{i\in N}\sum_{j\in N}\sum_{tr\in TR_{ij}} d_{ij}^{tr})$$

s.t.

$\mathrm{h}_{\min} \le h_{r_m} \le \mathrm{h}_{\max}$	$r_m \in \mathrm{R}$	(headway feasibility constraint)
$L_{r_m} = \frac{Q_{r_m}^{\max} * h_{r_m}}{\mathrm{P}} \le \mathrm{L}_{\max}$	$r_m \in \mathrm{R}$	(load factor constraint)
$\sum_{m=1}^{\mathrm{M}} N_{r_m} = \sum_{m=1}^{\mathrm{M}} \frac{T_{r_m}}{h_{r_m}} \le \mathrm{W}$	$r_m \in \mathrm{R}$	(fleet size constraint)
$\mathrm{D}_{\min} \le D_{r_m} \le \mathrm{D}_{\max}$	$r_m \in \mathrm{R}$	(trip length constraint)
$M \le \mathrm{R}_{\max}$		(maximum numbers of routes constraint)

3.5. Constraints of the BTRNDP

There are five constraints that were considered and included for the BTRNDP in this research, namely headway feasibility, load factor, fleet size, trip length, and maximum number of routes. Essentially, all these constraints are operator constraints and the subsequent sections discuss them.

3.5.1 Headway Feasibility Constraint

$$\mathrm{h}_{\min} \le h_{r_m} \le \mathrm{h}_{\max} \qquad r_m \in \mathrm{R}$$

Generally speaking, the most commonly used service frequencies in the transit industry can be grouped into three categories: supply frequency, policy frequency, and demand frequency:

1) Supply frequency is dependent on the operator's resources including limited fleet size. It is the maximum frequency that the operator can provide under current resource and economic constraints.
2) Demand frequency is determined by transit demand. This frequency is the minimum frequency that provides just enough capacity to meet the demand on the maximum link flow so that on the other links of this route, the demand is always less than the capacity.
3) Policy frequency can serve as a lower bound and an upper bound for service frequency and is usually used by transit operators when the supply frequency is much greater than the demand frequency or vice versa. Policy headways are most effective in systems that provide service for low-demand areas. However, when demand is high, especially during peak hours in large cities, policy headway is much less efficient. In this case, the demand frequency should be used.

In the real world, as well as in the bus transit route network design process, the demand frequency approach is preferred because it reflects the purpose of transit operations, which is to provide customer-oriented service. Furthermore, it is also expected that if the transit service that is proposed using the demand frequency approach falls below a minimum threshold, the policy frequency might be used instead. However, the supply frequency seems to be very rarely used. Based on these descriptions, demand frequency and the policy frequency are jointly employed.

The headway feasibility constraints shown in this section reflect the necessary usage of policy headways on extreme situations. Furthermore, generally speaking, the maximum bus line headway and the minimum bus line headway have been chosen as user-specified with 60-minutes and 5-minutes being the most commonly used maximum and minimum. Demand frequency is described in the following load factor constraint.

3.5.2 Load Factor Constraint

$$L_{r_m} = \frac{Q_{r_m}^{\max} * h_{r_m}}{\mathrm{P}} \le \mathrm{L}_{\max} \qquad r_m \in \mathrm{R}$$

This constraint reflects demand frequency. Generally speaking, headway is influenced by the demand and the vehicle capacity, which is determined by the seating and standing capacity (i.e., standee rule). Note that the maximum flow on the critical link of route r_m can be determined as: $Q_{r_m}^{\max} = \max\limits_{(a,b)\in\Omega_{r_m}} \sum_{i=1}^{N}\sum_{j=1}^{N} \beta_{ijab}^{r_m} d_{ij}^{r_m}$. The load factor is often used to represent vehicle capacity. A load factor of 1.0 means every seat is occupied and higher load factor reflects the user-specified choice to include standing passenger spaces as part of the vehicle capacity.

Actually, this constraint usually can be transformed to the headway determination, where this minimum frequency can be computed as: $f_{r_m} = \dfrac{\mathrm{Q}_{r_m}^{\max}}{\mathrm{PL}_{r_m}}$. The maximum headway is therefore computed as: $h_{r_m} = \dfrac{\mathrm{PL}_{r_m}}{\mathrm{Q}_{r_m}^{\max}}$.

3.5.3 Fleet Size Constraint

$$\sum_{m=1}^{\mathrm{M}} N_{r_m} = \sum_{m=1}^{\mathrm{M}} \frac{T_{r_m}}{h_{r_m}} \le \mathrm{W} \qquad r_m \in \mathrm{R}$$

This constraint represents the resource limits of the operating organization. As mentioned before, the supply frequency is usually highly dependent on the operator's resources. Namely, the limited fleet size is expected to have a significant impact on the level of service that can be provided by the BTRNDP solution network. This constraint guarantees that the optimal network pattern never uses more vehicles than the currently available ones.

3.5.4 Trip Length Constraint

$$\mathrm{D}_{\min} \le D_{r_m} \le \mathrm{D}_{\max} \qquad r_m \in \mathrm{R}$$

This constraint avoids routes that are too long because bus schedules on very long routes are too difficult to maintain. Meanwhile, to guarantee the efficiency of the network, the length of routes should not be too small. Furthermore, the thresholds of the maximum or minimum lengths of the bus routes are usually user-defined values.

3.5.5 Maximum Number of Routes Constraint

$$M \le R_{\max}$$

In solving the BTRNDP, transit planners often set a maximum number of routes, which is based on the fleet size and this has a great impact on the later driver scheduling work. This constraint is introduced to add realism to the optimal solution network.

The above sections present brief descriptions for the five constraints that are considered. There are other possible constraints, which may include: directness of route feasibility constraint and minimum route ridership volume constraint. Since these constraints are partially correlated with the above constraints, they are not considered in this research.

3.6.Objective Function of the BTRNDP

Our objective is to minimize the sum of user cost, operator cost and unsatisfied demand cost for the whole network under study as follows:

$$\min z = \mathrm{C}_1 * (\sum_{i \in N} \sum_{j \in N} \sum_{r_m \in DR_{ij}} d_{ij}^{r_m} t_{ij}^{r_m} + \sum_{i \in N} \sum_{j \in N} \sum_{tr \in TR_{ij}} d_{ij}^{tr} t_{ij}^{tr}) + \mathrm{C}_2 * \mathrm{C}_{\mathrm{v}} * \mathrm{O}_{\mathrm{v}} * (\sum_{m=1}^{\mathrm{M}} \frac{T_{r_m}}{h_{r_m}})$$
$$+ \mathrm{C}_3 * (\sum_{i \in N} \sum_{j \in N} d_{ij} - \sum_{i \in N} \sum_{j \in N} \sum_{r_m \in DR_{ij}} d_{ij}^{r_m} - \sum_{i \in N} \sum_{j \in N} \sum_{tr \in TR_{ij}} d_{ij}^{tr})$$

where the first term is the total user cost (including the user cost on direct routes and that on transfer paths), the second part is the total operator cost, and the third component is the cost resulting from total transit demand excluding those satisfied by a specific network configuration. Note that C_1, C_2 and C_3 are introduced to reflect the tradeoffs between the user cost, the operator cost and unsatisfied transit ridership, making the BTRNDP a multi-objective optimization problem. Generally, operator cost refers to the cost of operating the required buses. User cost usually consists of four components, including walking cost, waiting cost, transfer cost, and in-vehicle travel cost. Note that there are also tradeoffs among the user cost that transit users always consider when it comes to the mode choice and route choice if transit is chosen (see Chapter 6). It should also be noted that the model here is formulated for the BTRNDP with fixed transit demand. If one tackles the BTRNDP with variable transit demand, only very minor modifications are needed. Put another way, given the total travel demand rather than the transit trip demand, one needs to determine the transit trip matrix besides the transit route configuration and route frequencies. In this context, the first two parts of this model remain unchanged. However, the third component should be modified to the total travel demand excluding the total transit demand that can be served by the current proposed transit route network. Details for the BTRNDP with variable transit demand are presented in Chapter 8. The following sections describe these components.

3.6.1 Transit User Costs

3.6.1.1 Passenger Access Time

Transit users usually have to walk from their origin (e.g., home address) to get to the routes that are close to their origin and provide service for them to reach their intended destination or vice versa. These distances usually vary among different persons. A good proxy for these distances is the lengths of the centroid connectors (namely, the distance from the centroid zones to the distribution node where the transit users wait for the buses to get to their destination. Obviously, the actual "shortest" distance to their intended distribution node depends upon the specific transit network configuration and road structure. However, once the network structure is given and the transit network configuration is proposed, the passengers' walking cost can be computed.

3.6.1.2 Transit Users' Waiting time

Waiting time is route-based and is defined as the amount of time that passengers spend at the transit stops. Generally, there are two scenarios describing transit users' waiting time: 1) Non-transfers: Assume that passengers arrive randomly at the bus transit stops, the expected waiting time for transit users on route r_m is half of h_{r_m}, the transit vehicle headway operating on route r_m. This assumption is reasonable especially when the vehicle headway is short (say, less than 10 minutes). However, according to results from previous research (e.g., Fan and Machemehl, 2002), as the headway increases, passengers tend to coordinate their arrivals with bus arrivals according to the published bus schedule, making the expected waiting time less than the half headway. For simplicity, the half headway model for predicting passenger waiting times is employed. 2) Transfers: It is expected that when transfers occur, the passenger waiting time will greatly depend on the schedules of the two connecting routes. It is expected that headway coordination can have impacts on waiting times for transfer passengers, namely, the absence of headway coordination imposes additional passenger waiting time at that transfer node. Perfect coordination, on the other hand, mediates this penalty. Clearly, if all routes are coordinated, then transfer penalties are not incurred, so waiting cost consists only of delays at origin stops. For simplicity, uncoordinated transfers are considered in this research.

3.6.1.3 In-Vehicle Travel Time

The user in-vehicle travel time is defined as the total passenger riding time in transit vehicles. Essentially, this travel time is still route-based and this component can be computed as the sum of the travel time on each link along the route. Assume that the travel speed of buses on the route network is a user-defined constant. Then, the total travel time can be computed as the total distance of this route divided by the bus traveling speed.

3.6.1.4 Transit Transfer-Related Time:

Usually, there are two time components when transfers are involved: 1) The user transfer walking time, which is defined as the amount of time that passengers spend walking from the initial stop to the transfer transit stop; and 2) The user waiting time at the bus stop where the second bus is boarded. Note that the first part has been mentioned and discussed in 3.6.1.1 and the second part has described in 3.6.1.2. Obviously, the transfer-related time is not route-based but it is actually network-based.

It is noted that transit users always want to walk less and avert as many transfers as possible. This passenger behavior enforces the need for the transfer penalty in the objective function. For simplicity, it is assumed that no more than one transfer can occur in this research. A simple equation is introduced here as an example to show how the transfer-related total travel time when transit users take a route that needs one transfer is computed: <u>Transfer-related-total-travel-time</u> = <u>walking time to 1st bus</u> + <u>waiting time for 1st bus</u> + <u>travel time on 1st bus</u> + <u>walking time from 1st to 2nd bus</u> + <u>waiting time for 2nd bus</u> + <u>transfer penalty</u> + <u>travel time on 2nd bus</u> + <u>walking time to the destination</u>. The details of how to compute the user travel time are presented in Chapter 6.

3.6.2 Transit Operator Costs:

As described before, operator cost is directly related to the fleet size, which is needed to provide all vehicle trips along the chosen set of routes. The operator cost is route-based and is therefore:

$$\text{Operator costs} = \mathrm{C_v} * \mathrm{O_v} * \sum_{m=1}^{\mathrm{M}} \frac{T_{r_m}}{h_{r_m}}$$

where the notations are as introduced before.

3.6.3 Unsatisfied Demand Costs

One of the important objectives of the BTRNDP is to provide bus transit service to as many transit users as possible. However, in most cases, it is almost impossible to provide service for all the travelers who want to use transit. Namely, some traveler demands may not be satisfied due to resource constraints. Therefore, minimizing the number of unsatisfied demands is included as part of the objective for the BTRNDP.

In the above BTRNDP model formulation, it should be noted that the third component, $(\sum_{i\in N}\sum_{j\in N} d_{ij} - \sum_{i\in N}\sum_{j\in N}\sum_{r_m\in DR_{ij}} d_{ij}^{r_m} - \sum_{i\in N}\sum_{j\in N}\sum_{tr\in TR_{ij}} d_{ij}^{tr})$ refers to unsatisfied demand for a specific network configuration. Note that the first part represents the total transit demand for the current transit network. The second is the total transit demand that can be satisfied through the direct route service and the third part is demand that can be met through one or more transfers. Therefore, the unsatisfied transit demand can be formulated as shown.

However, to the author's knowledge, almost all the previous research on the BTRNDP did not include the cost of unsatisfied demand. This point is quite obvious because the optimal solution to the BTRNDP without considering unsatisfied demand costs should be "no service provided at all" and "zero" cost could always be the best solution that one can achieve for the minimization problem in this case. Therefore, inclusion of unsatisfied demand cost in the objective function is justified.

3.6.4 Multi-Objective Decision Making Problems

Note that the coefficients included in each component of the objective function are introduced to reflect different tradeoffs between the inherently conflicting natures of the user cost, the operator cost and the unsatisfied demand cost. The tradeoff can be described as follows: if one wants to design an optimal transit route network with an objective to solely minimize the unsatisfied demand cost, then the operator cost could be very high. Conversely, if one wants to minimize the operator cost for a decent network, the sum of the user cost and unsatisfied demand cost would be very high because little or no transit service is provided. Figure 3.2 gives a graphical representation for these tradeoffs.

Based on these considerations, the coefficients are explicitly introduced to account for the above-mentioned tradeoffs. Also note that these coefficients are user-defined since different transit planners many have different design requirements and may wish to set different values for these coefficients. Certainly, different user-defined coefficients could result in different optimal BTRNDP solutions. Sensitivity analyses are conducted and the multi-objective decision making nature is explored in Chapter 9 to show tradeoffs inherent in the BTRNDP.

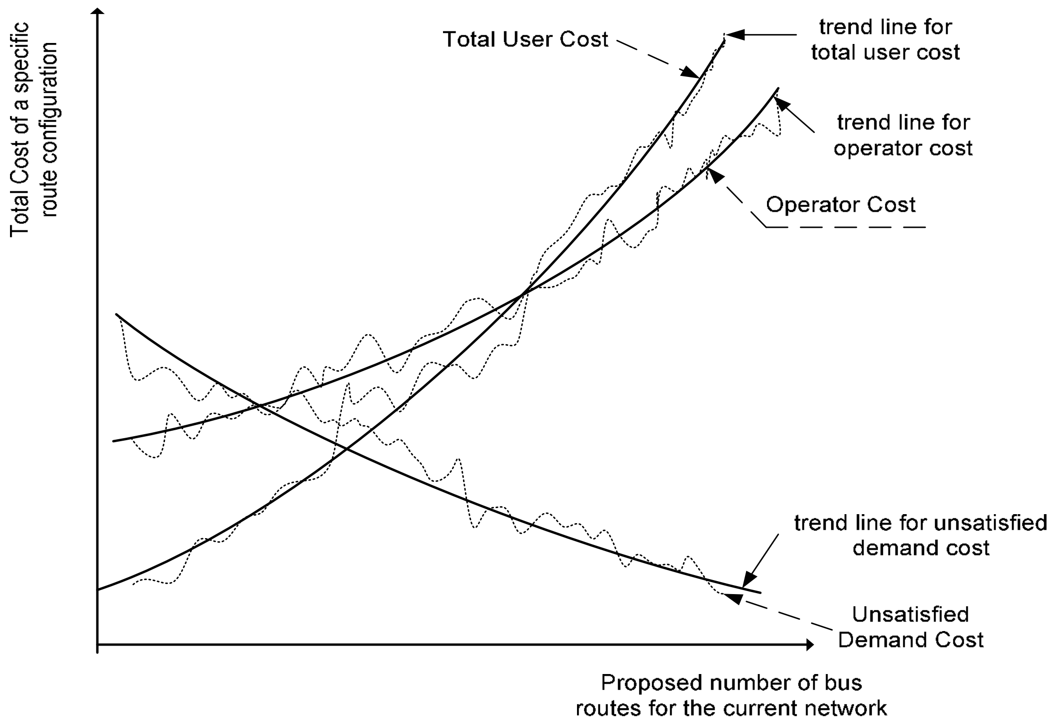


Figure 3.2 Graphical Representations for the Tradeoffs Inherent in the User Cost, Operator Cost and Unsatisfied Demand Cost

3.7.Shortcomings of Previous Approaches

The literature review in Chapter 2 indicates that, there are several shortcomings in the previous research work. The principal shortcomings include:

1. All previous approaches, for simplicity, tackle the trip assignment model by implicitly aggregating the transit demand associated with a zone into a single point without considering it as being distributed among bus stops (represented here as distribution nodes). It is expected that results obtained under the single point assumption reduce the solution network to an approximation (if not incorrect) for real-world applications.

2. Most approaches ignore the inherent multiple objective nature of the transit route network design problem and only consider generalized cost (time) as their single objective. In addition to constructing bus routes to ensure the connectivity of all (or most) O-D demand pairs, the penalties should be set for the total unsatisfied demand and this cost should be incorporated into the multi-objective function for the transit network design problem.

3. Most approaches consider the BTRNDP in the context of assuming total transit demand being fixed. The solution obtained under this assumption is problematic since transit demand is largely dependent on the combinations of transit route network structure and its associated frequencies. There is a dynamic relation between variable demand and the optimal transit route network. To ensure the quality of the solution to the BTRNDP, variable demand should be considered.

4. Most approaches fail to incorporate well the practical service panning guidelines in the process of solving the BTRNDP. As a result, the route network design and its service frequency solutions are sometimes either poor, uneconomical, or even operationally infeasible. To ensure a solution of better quality, ad hoc professional judgment and practical experience from the practitioners in the transit industry need to be incorporated.

5. Most approaches ignore some essential aspects of the problem and therefore rely on some shaky procedures. For example, when computing the total travel time, focus is primarily on the total in-vehicle travel time without proper consideration of the walking time, waiting time, and transfer-related times involved. Transfer penalties are usually not carefully considered. As a result, the transit route choice assignment model based on the relative travel time of auto and transit is not mathematically precise. Careful attention should be given to this problem.

3.8.Summary

This chapter focuses on the model formulations for the BTRNDP. The objective of the BTRNDP is formulated as an optimization problem of minimizing the overall cost including the user cost, operator cost and unsatisfied demand cost. Since the user cost are associated with time and the operator cost are related to money, one must use some empirical knowledge to combine user cost and operator cost into a single metric. Some of the feasibility constraints include minimum operating frequencies on selected routes, a maximum load factor on every bus route, a maximum allowable bus fleet size, and a maximum and minimum limit on the route length. Different components of the user cost, operator cost as well as the unsatisfied demand cost are discussed.

Principal shortcomings of previous approaches include: 1) failure to consider the BTRNDP where the demand should be distributed among the bus stops serving an origin (destination) rather than to the centroid (single point); 2) failure to address the inherent multiple objective nature of the transit route network design problem; 3) failure to consider the transit route network design problem in the context of variable demand; 4) failure to incorporate practical service guidelines; and 5) failure to consider essential aspects of the problem in the solution process. The proposed solution methodology for the BTRNDP is presented in the next chapter.

CHAPTER FOUR
PROPOSED SOLUTION METHODOLOGY

4.1 Introduction

Previous approaches to solve the BTRNDP were discussed and their shortcomings were pointed out in the preceding chapter. Due to the inherent complexity in the BTRNDP and its NP-hard nature that were presented in Chapter 2, using traditional analytical optimization models to solve this problem is difficult. In this chapter, several solution methodologies that account for these characteristics of the BTRNDP are presented. It explicitly incorporates several practical guidelines and industry rules of thumb and uses a heuristic-based mathematical tool to search the solution space. Building on, adapting and modifying several previous approaches, this approach offers a more comprehensive transit route network design procedure.

The proposed solution framework consists of three main components: an Initial Candidate Route Set Generation Procedure (ICRSGP) that generates all feasible routes incorporating practical guidelines that are commonly used in the bus transit industry; a Network Analysis Procedure (NAP) that assigns the transit trips, determines the service frequencies on each route and computes many performance measures; and a Heuristic Search Procedure (HSP) that guides the search techniques. Not that five heuristic algorithms, including the Genetic Algorithm (GA), Local Search (LS), Simulated Annealing (SA), Random Search (RS) and Tabu Search (TS) algorithms, as well as the Exhaustive Search Method (ESA) as a benchmark for the BTRNDP with small network, are proposed to select an optimum set of routes from the huge solution space. This chapter is organized as follows. First, section 4.2 presents the solution framework and its distinct features. Section 4.3 offers an overview of the ICRSGP, followed by discussions of the NAP in section 4.4. Section 4.5 presents the rationale for choosing these six algorithms as solution techniques and a brief literature summary of each algorithm is provided. Finally, this chapter concludes with a summary in section 4.6.

4.2 Proposed Solution Framework and its Distinct Features

Three main features of the solution methodologies proposed in this research distinguished this work from the previous ones: 1) this solution approach explicitly incorporates several practical guidelines and industry rules of thumb. However, unlike previous approaches, this approach is not heavily dependent on user experience but is scientifically guided by the heuristic-based mathematical tool to search the optimal solution space intelligently in a reasonable time domain; 2) this approach, for the first time, employs a multiple path trip assignment model in the NAP and explicitly considers transfer and long-walk related characteristics among routes under a much more real situation (i.e., at a microscopic "centroid-connector-link" level); 3) this approach explicitly considers the network analysis procedure for the BTRNDP under two different scenarios, namely fixed and variable transit demand and uses an iterative procedure to obtain route frequencies and total transit trip demand. Note that this chapter discusses the solution methodology solely for the BTRNDP with fixed demand and that for the BTRNDP with variable transit demand is presented in details in Chapter 8.

The solution framework employed in this research is presented in Figure 4.1.

4.3 The Initial Candidate Route Set Generation Procedure

The initial candidate route set generation procedure (ICRSGP) is a design algorithm that configures all candidate routes for the current transit network. It requires the user to define the minimum and maximum route lengths. The knowledge of the transit planners has a significant impact on the initial route set skeletons, that is, different user requirements result in different route set solution space. The ICRSGP relies mainly on algorithmic procedures including the shortest path and k-shortest path algorithms and is guided by user-defined values that incorporate the knowledge and expertise of the transit planners. Furthermore, it should be noted that the ICRSGP remains the same for both fixed and variable demand, namely, given the user-defined minimum and maximum route length constraints, the same set of candidate feasible routes is generated for the BTRNDP with either fixed or variable demand. The details of the ICRSGP are discussed in Chapter 5 and an illustrative application to a small network is also presented.

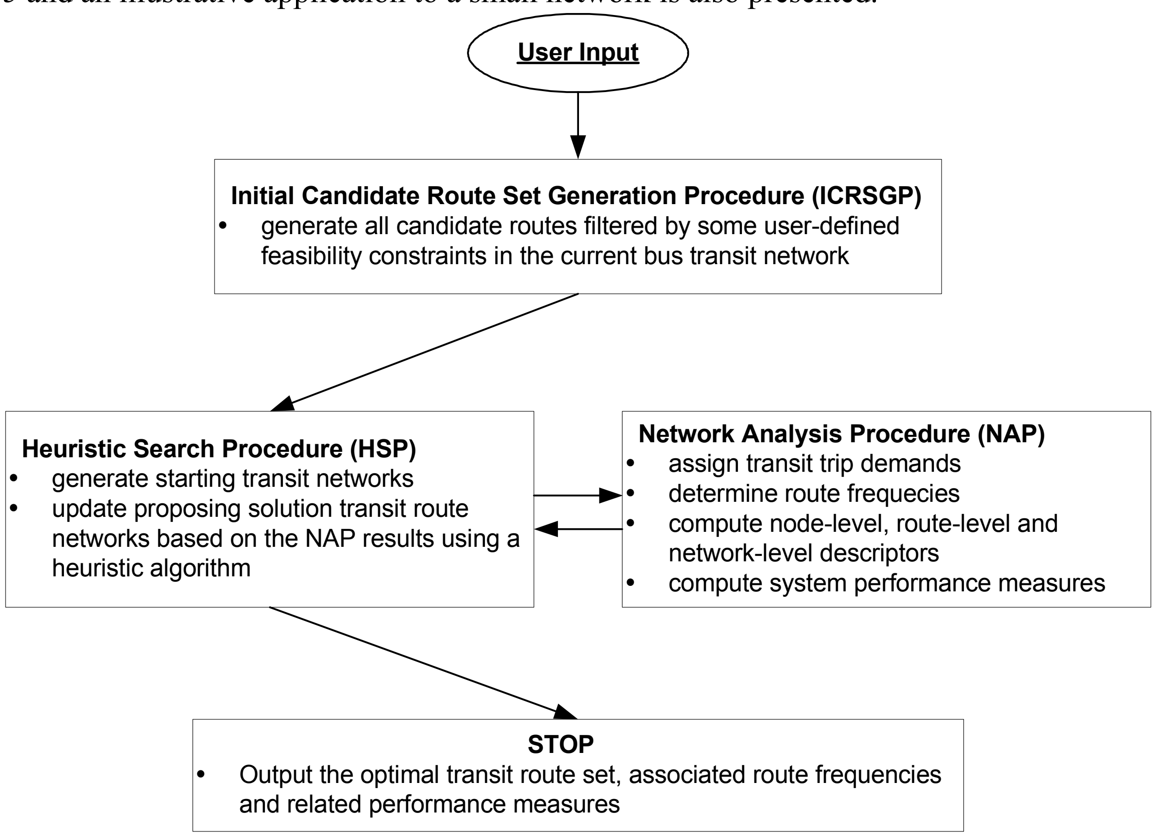


Figure 4.1 Flow Chart of the Proposed Solution Methodology

4.4 The Network Analysis Procedure

Basically speaking, the network analysis procedure (NAP) is a bus transit network evaluation tool with the ability to determine route frequencies. To accomplish these tasks for the BTRNDP with fixed demand, NAP employs an iterative procedure that seeks to achieve internal consistency of service frequencies on each route in the whole proposed solution transit network. Furthermore, the iterative procedure in the NAP contains two major components, namely, a transit trip assignment model and a frequency setting procedure. For the BTRNDP with variable demand, an additional procedure, namely a transit demand equilibration procedure is added and described in Chapter 8.

Once each possible candidate route is generated by the ICRSGP, the solution space is formed. Then, as a specific set of routes are proposed as the solution transit route network using any of the employed HSP, the NAP is called to evaluate the alternative network structure and to determine associated route service frequencies. The whole process in the NAP can be described as follows. First, an initial set of route frequencies are specified because it is necessary before the beginning of the trip assignment process. Then, the transit trip assignment model is utilized to assign the passenger transit demand matrix to a given set of routes associated with a particular network configuration. The service frequency on each route in this transit route network system is then computed, which is used further as the input frequency for the next iteration in the frequency setting procedure if these frequencies are considered to be quite different from the previous input frequencies by a user-defined parameter. The whole process won't be terminated until the internal consistency of route frequencies is achieved. As a result, the frequency on a given set of routes associated with a particular network configuration is determined. For any given route network configuration and its associated service frequency, many system performance measures that can reflect the quality of service are computed. Also, the user cost (i.e., the cost experienced by the transit users), the operator cost (i.e., the fleet size) and the transit demand unsatisfied by this proposed transit network are determined. As a result, the optimal route configuration for a particular network and its route frequencies are then obtained by choosing the one that achieves the minimum objective function value.

Furthermore, as mentioned, two different network analysis procedures (NAP) are designed for the BTRNDP, with fixed demand and that with variable demand, respectively. The main difference between these two procedures is that the transit demand equilibration procedure is added to the second one. The NAP in this research differs from the previous approaches in three main features. First, for the first time, the NAP considers the transit demand at the distribution node level rather than at an aggregate single node level. Second, the NAP employs a multiple path transit assignment model that explicitly considers transfer and long-walk related characteristics among routes and uses an iterative procedure to obtain route frequencies at a microscopic "centroid-connector-link" level. Third, the NAP can explicitly consider the variable transit demand characteristics for the BTRNDP.

Details of the NAP are presented in Chapter 6 and these include the algorithmic skeleton of the NAP, the transit trip assignment model and the frequency setting procedure. Also included is the computation of many system performance measures that reflect the quality of service, determination of the transit user cost, the operator cost (i.e., the fleet size) and the unsatisfied demand cost.

4.5 Solution Techniques

Due to the inherent complexity and NP-hard nature of the BTRNDP, traditional optimization methods for capturing the global optimum solution are not applicable for realistic, large-scale networks. As a result, several heuristic methods including Genetic Algorithms (GA), Local Search (LS), Simulated Annealing (SA), Random Search (RS) and Tabu Search Algorithms (TS) are employed to solve the BTRNDP. Among many available methods, these algorithms are particularly well suited to this type of the problem because they allow efficient reformulation of the problem, making the solution process computationally more efficient than other approaches. In addition, to examine the quality of the solution produced by GA, LS, SA, RS and TS, an Exhaustive Search Method (ESM) is also used for the BTRNDP with small network. These algorithms are implemented and compared using comprehensive experimental networks, and related issues and characteristics also are identified, in Chapter 9.

As mentioned, the first five heuristic algorithms develop good local (possibly global) solutions. Figure 4.2 presents an illustration of the solution approaches. For a specific route set size (i.e., the number of routes), these algorithms find a local optimum for the current transit network and this local optimal solution is appended to a local optimum set. As the route set size changes, another local optimum is found using these algorithms and the local optimum set is updated. The same process repeats until a user-defined limit (e.g., the number of local optima found reaches the maximum value). The best solution in the local optimum set is selected as the optimal solution to the BTRNDP. Although different from the exhaustive search method, which guarantees the global optimal solution (this process might take a huge amount of time), these heuristic algorithms can find a very good solution (a decent transit route network) in a reasonable time domain.

The subsequent sections present a brief literature review of the GA, LS, SA, RS, TS and ESA algorithms step by step. Details of the applications of these six algorithms to the BTRNDP, both with fixed and variable transit demand, are discussed in Chapter 7 and Chapter 8, respectively. Comprehensive experiments and numerical results are presented in Chapter 9.

4.5.1 Genetic Algorithm

The first papers in the first literature dealing with GAs are those of Holland (1975) and Schwefel (1981). More recently, Goldberg (1989), Chambers (1995) and

Michalewicz (1999) have discussed several applications of GAs in optimization problems. Generally speaking, a GA is a local search algorithm, which starts from an initial collection of strings (a population) representing possible problem solutions. Each string of the population is called a *chromosome*, and has an associated *fitness function* that contributes to the generation of new populations by means of genetic operators (called *reproduction*, *crossover* and *mutation*, respectively). Every position in a chromosome is called a *gene* and its value is an *allelic* value. This value may vary according to an assigned *allelic alphabet*; the most common allelic alphabet is {0,1}. At each generation, the algorithm uses the fitness function values to evaluate the survival capacity of each string *i* of the population using simple operators in order to create a new set of artificial creatures (a new population) which tries to improve on the current fitness function values by using pieces of the oldest ones. Figure 4.3 shows a simple layout of a GA implementation.

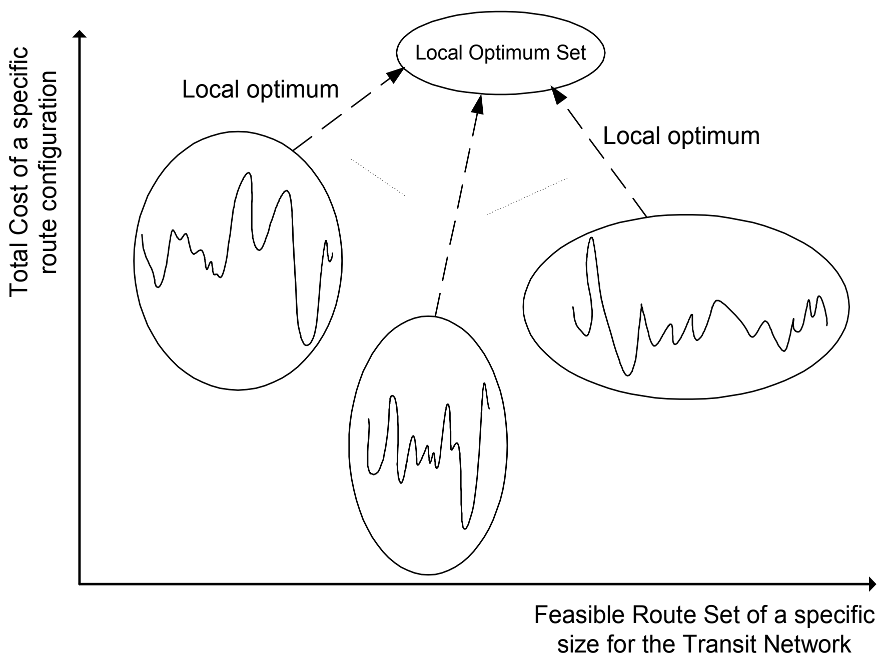


Figure 4.2 Approaches for Finding the Local Optimum for the BTRNDP

Before describing the generic operators in detail, a number of differences between this method and other local search techniques can be listed as follows:

- GA operates with codes of the parameter set and not with the parameters themselves;
- GA searches for a population of points and not a single point;
- GA uses objective function information and not derived or auxiliary knowledge;
- GA uses probabilistic not deterministic transition rules.

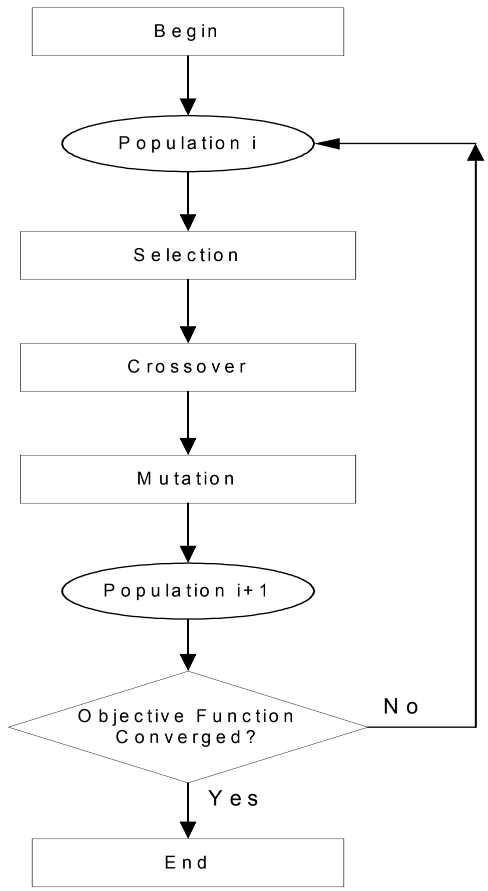


Figure 4.3 Algorithm Skeleton of a GA Implementation

These particular aspects make this method applicable in a very general way, without the limitations imposed by other local search methods (i.e., continuity, derivative existence, and unimodality). Moreover, it makes possible the exploitation of consequent information from more points in the domain of the solutions, reducing the probability of finding false peaks, i.e., traps or local optima.

The working method of genetic algorithms is very simple and involves nothing more than copying strings or swapping partial strings. The simplicity of the operations and the ability to find good solutions are two characteristics that make this method very attractive. In the next subsections, the GA representation is first presented and the *selection*, *crossover* and *mutation* operators are then discussed.

4.5.1.1 Representation

Without loss of generality, it is assumed that we wish to minimize a function of *k* variables, $f(x_1, x_2, ..., x_k):\ R^k \rightarrow R$. And suppose further that each decision variable x_i

can take values within a domain $D_i = [a_i, b_i] \subseteq R$ and $f(x_1, x_2, ..., x_k) > 0$ for all $x_i \in D_i$.

In the process of implementing a GA, the first important thing to do is to set up an encoding method to represent these variables. To optimize the function f, one should encode decision variables into binary strings meeting some desirable required precision. For example, if the required precision is five places after the decimal point, one should cut the domain D_i into $(b_i - a_i) * 10^5$ equal size ranges. The minimum required bits m_i for the variable x_i could be calculated by the following equations:

$$2^{m_i - 1} - 1 < (b_i - a_i)10^5 \leq 2^{m_i} - 1$$

Then, a representation of any variable x_i coded as a binary string of length m_i can satisfy the precision requirements. Mapping from a binary string to a real decimal number for any variable x_i can be interpreted as follows:

$$x_i = a_i + (decimal\ value\ of\ m_i)\frac{(b_i - a_i)}{2^{m_i} - 1} \qquad (1)$$

where m_i is the binary string representation of x_i.

For each chromosome (as a potential solution), we would have a k-dimensional vector $S_i = (x_1, x_2, ..., x_k)$. Therefore, it is expected that the representation of a single solution would have a binary string of length $m = \sum_{i=1}^{k} m_i$. The first m_1 bits map into a value within the range $[a_1, b_1]$, the next m_2 bits map into a value within the range $[a_2, b_2]$ and so on; the last m_k bits map into a value from the range $[a_k, b_k]$.

To initialize a population, one can set a pre-specified number (*pop_size*) of chromosomes randomly or use his/her available knowledge to arrange these sets of initial solutions. Then, the rest of the algorithm is straightforward: we can evaluate each chromosome using the objective function f based on the decoded sequences of variables in each generation, select a new population with respect to the probability distribution based on the fitness values, and alter the chromosomes in the new population by mutation and crossover operators. After a certain number of generations, if no further improvement is observed, the current kept best chromosome represents an optimal (possibly the global) solution. Or, one can also stop running the algorithm after a fixed number of iterations depending on speed and resource criteria.

4.5.1.2 Selection

In most practices, a roulette wheel, a fitness-proportional approach in essence, is adopted for the selection process, which, generally speaking, begins by spinning the roulette wheel *pop_size* times and each time a single chromosome is selected for a new population. The specific procedure can be described as follows:

- Calculate the fitness value $obj(S_i)$ $(S_i = (x_1, x_2, ..., x_k))$ for each chromosome S_i $(i = 1, 2, \cdots, pop\_size)$;
- Get the total summation of fitness value of the population $T = \sum_{i=1}^{pop\_size} obj(S_i)$;
- Calculate the probability of a selection p_i for each chromosome S_i $(i = 1, 2, \cdots, pop\_size)$ $p_i = obj(S_i)/T$;
- Calculate the cumulative probability of q_i for each chromosome S_i $(i = 1, 2, \cdots, pop\_size)$ $q_i = \sum_{i=1}^{pop\_size} p_i$;
- Generate a random number r within the range [0,1);
- If $r \leq q_1$, then select the first chromosome S_1; else select the i-th chromosome S_i $(2 < i \leq pop\_size)$ such that $q_{i-1} < r \leq q_i$.

4.5.1.3 Crossover

Since the principles involved in the one-cut-point method are essentially the same as that in the two-cut-point method, the one-cut-point method is chosen to illustrate the crossover process here. Basically, crossover with the one-cut-point method selects randomly the cut-point and then the offspring is generated by exchanging the right parts of two parents. Given the probability of crossover p_c as one of the parameters of a genetic system, the number $p_c \cdot pop\_size$ of chromosomes is expected to undergo the crossover. And the crossover procedure can be described as follows:

- Set $i = 1$;
- Generate a random number r within the range [0,1);
- If $r \leq P_c$, then select chromosome S_i for crossover;
- $i = i + 1$;
- Repeat the above steps until $i > pop\_size$.

If the number of selected chromosomes obtained above is even, then they can be easily and randomly mated. If the number of selected chromosomes is odd, one can either randomly add one extra chromosome or randomly remove one selected chromosome. For each pair of these selected chromosomes, we generate a random integer number *pos* within the range $[1, 2, \cdots, m-1]$, where m is the total length of the bits in a chromosome as mentioned before. Obviously, the *pos* indicates the position of the crossing point and the crossover procedure can be illustrated as follows:

$$\left.\begin{array}{l} B_1 = \left(b_1 b_2 \cdots \bar{b}_{pos} \middle| \bar{b}_{pos+1} \cdots b_m\right) = (u, v) \\ B_2 = \left(c_1 c_2 \cdots \bar{c}_{pos} \middle| \bar{c}_{pos+1} \cdots c_m\right) = (x, y) \end{array}\right\} \Rightarrow \left\{\begin{array}{l} B_1^{'} = \left(b_1 b_2 \cdots \bar{b}_{pos} \middle| \bar{c}_{pos+1} \cdots c_m\right) = (u, y) \\ B_2^{'} = \left(c_1 c_2 \cdots \bar{c}_{pos} \middle| \bar{b}_{pos+1} \cdots b_m\right) = (x, v) \end{array}\right.$$

4.5.1.4 Mutation

Mutation is performed on a bit-by-bit basis. That is to say, every bit in the whole population has a chance to undergo mutation, i.e., change from 0 to 1 or vice versa, with the probability being equal to the mutation rate p_m. And it is expected that $p_m \cdot m \cdot pop\_size$ bits will be mutated. The whole mutation procedure can be described as follows:

- Set $i = 1$;
- Generate a random number $r \in [0, 1)$ for each bit;
- If $r \le p_m$, mutate the bit;
- $i = i + 1$;
- Repeat the above steps until $i > m \cdot pop\_size$.

If p_m is chosen to be small, many bits (genes) that might be usefully chosen for further improvements will be rarely examined. On the other hand, if p_m is chosen to be too big, the offspring will lose their resemblance to their parents due to random perturbations, and as a result, the GA will lose the ability to learn from the search history (see Lee and Machemehl, 1998). Therefore, this probability parameter needs to be carefully chosen.

4.5.2 Local Search

Local search has a long tradition in combinatorial optimization. An instance of a combinatorial optimization problem typically consists of a set of feasible solutions and a cost function over the solutions. The problem consists in finding a solution with the optimal cost among all feasible solutions. As mentioned, generally the problems addressed are computationally intractable, thus approximation algorithms have to be used. One class of approximation algorithms that has been successful in spite of their simplicity are local search methods. Local improvement ideas have found application in many domains, including constraint satisfaction, as well as routing and scheduling.

Local search is based on the concept of a neighborhood. A neighborhood of a solution *p* is a set of solutions that are in some sense close to *p*, for example because they can be easily computed from *p* or because they share a significant amount of structure with *p*. It is also known that the neighborhood generating function may, or may not, be able to generate the optimal solution.

A typical local search method, which is widely used to solve combinatorial optimization algorithms, can be instantiated in four steps: 1) Selection of an initial feasible solution; 2) Generation of a local neighborhood; 3) Cost function to minimize; and 4) Selection of the next point. During the general implementation process, the local search method generally starts with a complete, but presumedly sub-optimal solution, and then checks the "neighborhood" of that solution to see if any of them is better. If the answer is positive, then this improved solution is adopted as the current best choice and the same process repeats. If not, then the local search can either give up, assuming the current solution is good enough (local optimum). Or, it can pick a totally random solution to start again with this newly-chosen starting point. Figure 4.4 shows the skeleton for the local search method.

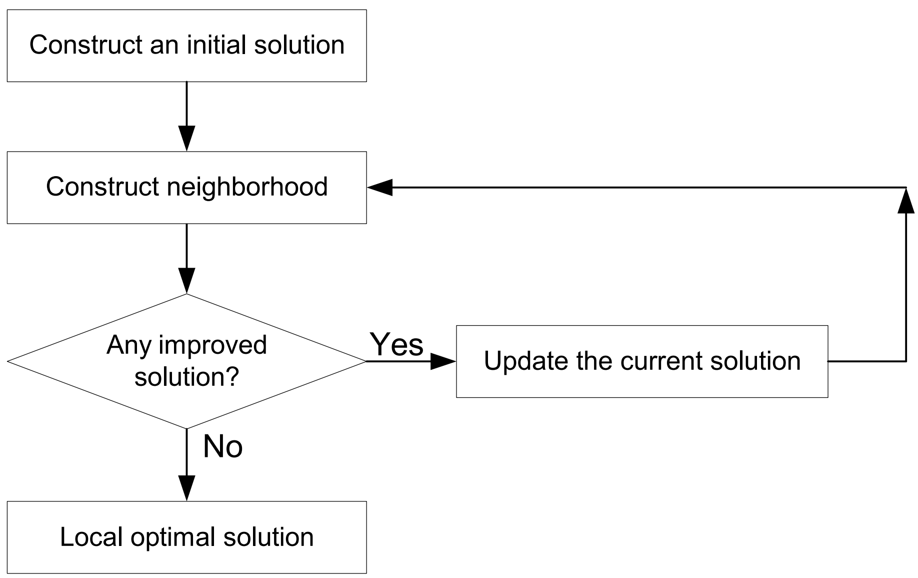


Figure 4.4 Skeleton for the Local Search Method

As can be seen, the basic idea of local search is the iterative improvement process, which starts with an initial solution and searches a neighborhood of the solution for a lower cost solution. If such a solution is found, it replaces the current solution and the search continues. Otherwise, the algorithm returns a locally optimal solution. Figure 4.5 gives a graphical depiction for the search process in the local search method.

Note that there are several variations of this basic algorithm. Undoutedly, the most dominant factor that might affect the quality of the solution is how to define the "neighborhood", i.e., the nearby solutions. Obviously, a different definition rule could result in a different solution of different quality. This research comes up with an efficient approach to define the neighborhood and all the details are presented in Chapter 7 and Chapter 8.

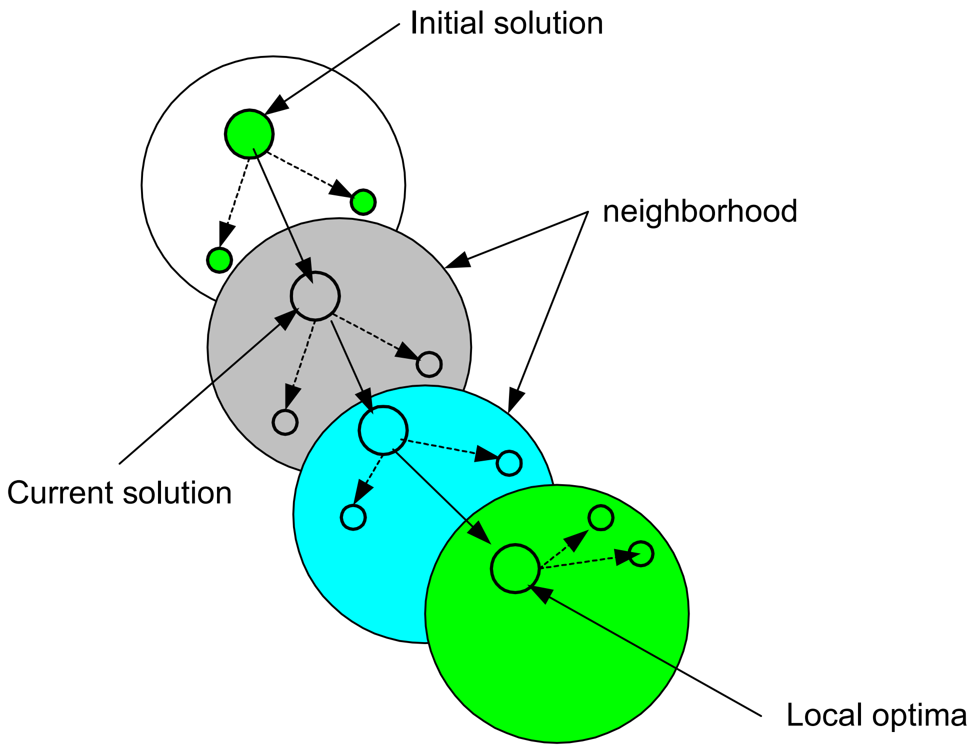


Figure 4.5 Graphical Representations for the Search Process in the Local Search Method

Note that the local search method obtains local optima, not necessarily the global optimum. Also, it can be seen that the local search method might get stuck after a certain number of steps and no local improving move can be achieved. One way to fix this problem and to improve the quality of the solution is to restart the search from several initial points and choose the best of the local optima reached from them. Certainly, the best solution should be remembered for each random restart.

In all, it is stressed that the appeal of the local search method to solve combinatorial optimization relies on its simplicity and good average-case behavior. Therefore, it is employed as one of the solution approaches for the BTRNDP and its implementation details are presented later.

4.5.3 Simulated Annealing

As one of the widely used heuristic approaches (including genetic algorithm and local search) to solve combinatorial problems, simulated annealing (SA) can produce a good though not necessarily global optimal solution within a reasonable computing time. Essentially speaking, simulated annealing can be regarded as a “Randomized Variation” of the local search method although more advanced than the local search method because it attempts to minimize the probability of being stuck in a low-quality local optimum.

As mentioned in the previous section, a simple form of local search (a descent algorithm) starts with an initial solution perhaps chosen at random. A neighbor of this solution is then generated by some suitable mechanism and the change in cost is calculated. If a reduction in cost is found, the current solution is replaced by the generated neighbor; otherwise the current solution is retained. The process is repeated until no further improvement can be found in the neighborhood of the current solution, so the descent algorithm terminates at a local minimum.

Although the local search method is simple and quick to execute, the disadvantage of the method is that the local minimum found may be far from any global minimum. In this sense, simulated annealing is introduced to alleviate the shortcomings of the local search. Essentially speaking, simulated annealing is a Monte Carlo simulation based search algorithm. The term "simulated annealing" is derived from the process of heating and then cooling a substance slowly to finally arrive at the solid state. In simulation, a minimum of the cost function corresponds to this ground state of the substance. The whole search algorithm simply mimics the physical process as follows. In the early stages of the execution, the temperature is high, which results in a higher probability for jumping to occur more frequently. In this case, the frequent jumping, which occurs as a way of avoiding local minima, may produce a higher probability of a poor solution. Put another way, simulated annealing selects the next point randomly. If a lower cost solution is found, it is selected. If a higher cost solution is found, it has a non-zero selection probability. The function that governs the behavior of the acceptance probability is called the cooling schedule. As the execution time elapses, the temperature decreases and the cooling schedule reduces the frequency of jumping. The simulation process terminates after a number of successive executions with no improvements, and returns the best solution found. The following figure provides an illustration of the SA algorithm in pseudo-code:

Basically, as can be seen, the annealing schedule consists of 1) the initial value of T; 2) a cooling function; 3) the number of iterations $N(t)$ to be performed at each temperature and 4) a stopping criteria to terminate the algorithm. In SA, the algorithm attempts to avoid entrapment in a local optimum by sometimes accepting a neighborhood move that increases the value of the objective function. The acceptance or rejection of an uphill move is determined by a sequence of random numbers, but with a controlled probability. The probability of accepting a move which causes an increase δ in f is called the acceptance function and is normally set to $\exp(-\delta / T)$ where T is a control parameter which is analogous to temperature in physical annealing.

Simulated Annealing algorithm;
Select an initial state $i \in S$;
Select an initial temperature $T > 0$;
Set temperature change counter $t = 0$;
Repeat
Set repetition counter $n = 0$ (number of iterations to be performed at each temperature)
    Repeat
    Generate state *j*, a neighbour of *i*;
    Calculate $\delta = f(j) - f(i)$;
    If $\delta < 0$ *then* $i = j$
        else if random (0,1) < exp($-\delta / T$) *then* $i = j$;
    n=*n*+1;
    until $n = N(t)$;
t=*t*+1;
$T = T(t)$;
until stopping criteria is true.

Figure 4.6 Simulated Annealing Algorithm in Pseudo-code (Adapted from Eglese, 1990)

Note that it is the probabilistic nature of the simulated annealing algorithm that guarantees the exploration of other solution spaces instead of terminating at the first local optimum. It can be proven that simulated annealing converges asymptotically to the optimal solution. However, such convergence requires exponential time, which means that the only drawback of simulated annealing is the long execution time to obtain high-quality solutions. That is to say, to achieve the global optimum solution using simulated annealing, the price is a slower cooling procedure and more iterations at each temperature level. However, it is widely accepted that simulated annealing can provide a near-optimal or at least local optimal solution within a reasonable time domain. Therefore, simulated annealing is employed as one of the solution approaches for the BTRNDP. The associated details using SA to solve the BTRNDP both with fixed and variable transit demand are presented later in this research. The performance comparisons between all these algorithms are presented in Chapter 9.

4.5.4 Random Search

Essentially speaking, random search is a Monte Carlo based simulation optimization method. When it is used for the BTRNDP, at each step for a specified route set size, the method randomly chooses a solution set from the whole solution space and evaluates the generated solution set by comparing the objective function value to previous steps. The optimal solution is achieved when the solution with the least objective function value (the least sum of total user cost, operator cost and unsatisfied demand cost) is found. Since this method is simply and easy to implement, random search is employed as one of the solution approaches in this research. Figure 4.7 presents the skeleton of this method in pseudo-code.

Random Search Algorithm;
Randomly generate an initial solution $i \in S$;
Set the number of iterations $t = 0$;
Evaluate the solution *i* and set the optimal objective solution $O(i)$;
Repeat
Generate a diffrent $j \in S$;
If the $O(j) < O(i)$, *then* $i = j$;
t=t+1;
until stopping criteria is true.

Figure 4.7 Random Search Algorithm in Pseudo-code

4.5.5 Tabu Search

The Tabu Search has traditionally been used on combinatorial optimization problems and frequently has been applied to many integer programming problems, routing and scheduling, traveling salesman and related problems. The basic concept of Tabu Search is presented by Glover (1977) who described it as a meta-heuristic superimposed on another heuristic. The overall approach is to avoid entrainment in cycles by forbidding or penalizing moves which take the solution, in the next iteration, to points in the solution space previously visited (hence "tabu"). The Tabu Search begins by marching to a local minima. To avoid retracing the steps used, the method records recent moves in one or more Tabu lists. The original intent of the list was not to prevent a previous move from being repeated, but rather to insure it was not reversed. The Tabu lists are historical in nature and form the Tabu search memory. The role of the memory can change as the algorithm proceeds. For initializations at each iteration, the objective is to make a coarse examination of the solution space, known as "diversification", but as locations of the candidate solutions are identified, the search is more focused to produce local optimal solutions in a process of "intensification". In many cases, various implementation models of the Tabu Search method can be achieved by changing the size, variability, and adaptability of the Tabu memory to a particular problem domain.

The chief limitation of a local search method (i.e., the hill climbing procedure) is that it might get stuck at a local optimal point that might be far from the global optimum. As one of the heuristic approaches to overcome this shortcoming, Tabu Search (TS) algorithm emulates an intelligent attitude by using an adaptive memory and can therefore avoid being entrapped at the local optimum with the aid of a memory function. It is an intelligent search technique that hierarchically explores one or more local search procedures in order to search quickly for the global optimum.

Tabu search explores the solution space by moving from a solution to the solution with the best objective function value in its neighborhood at each iteration even in the case that this might cause the deterioration of the objective. (In this sense, moves are defined as the sequences that lead from one trial solution to another.) As mentioned, to avoid cycling, solutions that were recently examined are declared forbidden or "tabu" for a certain number of iterations and associated attributes with the tabu solutions are also stored. The tabu status of a solution might be overridden if it corresponds to a new best solution, which is called "Aspiration". There are groups of Tabu Search methods that use either short term memory or intermediate and long term memory strategies. The recency-based memory functions require specifying the tabu tenure *m* and the frequency-based memory generally adds long term memory.

Several features inherent in the TS also might include the diversification and intensification procedure. The diversification strategy use counts to diversify search and drive to new regions by penalizing moves with greater frequency counts, preserving the aggressiveness of the search. That is to say, it undertakes to generate solutions that embody compositions of attributes significantly different from those encountered previously during the search process. Conversely, keeping track of high quality local optima gives rise to complementary strategy of intensification. Put another way, the intensification strategy intends to create solutions by aggressively encouraging the incorporation of "good attributes". Intensification and diversification are fundamental cornerstones of longer term memory in tabu search.

Let the set $S(x)$ define a "neighborhood function" that consists of those moves from the current solution x to a next trial solution. Let T denote a subset of S that contain elements that are called "tabu moves" and "OPTIMUM" as the objective evaluation function. A basic version of the Tabu Search Algorithm without "aspiration" can be presented as follows:

Tabu Search Algorithm;

Step1. Select an initial $x \in X$ and let $x^* = x$.
Set the iteration counter $k = 0$.
Set the Tabu set $T = \phi$.

Step2. If $S(x) - T = \phi$, go to Step 4.
Otherwise, set $k = k + 1$.
Select $s_k = S(x) - T$ such that $s_k(x) = \text{OPTIMUM}(s(x) \mid s \in S(x) - T)$.

Step3. Let $x = s_k(x)$.
If $c(x) < c(x^*)$, where x^* denotes the best solution currently found, let $x^* = x$.

Step4. If the number of iterations has reached the maximum user-defined iterations either in total or since x^* was last improved, or if $S(x) - T = \phi$ upon reaching this step directly from Step 2, stop.
Otherwise, update Tabu set T and associated attributes and return to Step 2.

Figure 4.8 Basic Tabu Search Algorithm in Pseudo-code (Adapted from Glover, 1989)

As mentioned by Glover (1989), by the preceding form of OPTIMUM, each execution in Step 2 moves from the current solution x to an $s(x)$ that yields the greatest improvements, or if not improved, the least disimprovement in the objective function, subject to the restriction that only non-tabu moves are allowed. Put another way, tabu search algorithm makes a "best available move" at each step (like the greedy algorithms). Especially, when the "aspiration" is considered in a more advanced tabu search algorithm, if no improvements can be found in the current non-tabu lists but improvements can be made in the tabu moves lists, then one can allow tabu moves and let it override the rules. Since the tabu search algorithm is problem-specific, for simplicity, probably one shouldn't take too much time describing the general concepts. However, details of the application of this algorithm for the BTRNDP are presented in Chapter 7 and Chapter 8.

The difference between SA and TS algorithms is that TS exploits memory, which is absent from the simulated annealing algorithm. That is to say, TS emphasizes guiding the search by reference to multiple thresholds including *tenures* for tabu-active attributes and conditional stipulations of *aspiration criteria*. Conversely, SA relies on several parameters as mentioned before such as *temperature*, *stopping criteria* and randomly rather than deterministically tackles the neighbor solutions.

In addition, the difference between GA and TS algorithms is that GA tracks history over subpopulations and selects the parents in each generation based on the objective functions and alters the new solutions randomly through *crossover* and *mutation* procedures. On the contrary, TS introduces *memory* and *aspiration* strategy and employs the *intensification* and *diversification* procedure to intelligently search the neighbor solutions. As one of the advanced heuristic methods, Tabu Search is generally regarded as a method that can provide a near-optimal or at least local optimal solution within a reasonable time domain for the BTRNDP.

4.5.6 Exhaustive Search

The Exhaustive Search Method (ESM) is an approach to search for the global optimal solution over the whole solution space. By simply enumerating and comparing the objective function for all possible solutions, the global optimal solution, which has the least objective function (the least sum of total user cost, operator cost and unsatisfied demand cost) is found. Figure 4.9 presents the skeleton of this method in pseudo-code.

Exhaustive Search Algorithm;
Generate an initial solution $i \in S$;
Evaluate the solution *i* and set the optimal objective solution $O(i)$;
Repeat
    Generate a different solution $j \in S$;
    If the $O(j) < O(i)$, *then* $i = j$;
until all the solutions in the solution space are traversed.

Figure 4.9 Exhaustive Search Algorithm in Pseudo-code

Note that the ESM can always be used, but evaluating all possible solutions can become prohibitively time consuming if the solution space is large. As a result, it is expected that ESM can only be applied for the BTRNDP when the network size is small, which can be illustrated as follows. Suppose that all feasible routes (the number of which is represented by FR_{max}) are found by the ICRSGP. As noted before, R is used to represent the number of routes proposed in the current solution network and R_{max} is used to represent the user-defined maximum allowed number of routes for the solution transit network. The solution space can therefore be represented by the sum of a specific combination: $\sum_{R=1}^{R_{max}} C_{FR_{max}}^{R}$. For example, suppose $FR_{max} = 100$ and $R_{max} = 10$. The solution space is then $\sum_{R=1}^{10} C_{100}^{R} = 1.94159 \times 10^{13}$. As a result, it can be seen that the exhaustive search method might work well for a very small network. However, when the network becomes larger, the ESM becomes an unrealistic tool to solve the BTRNDP. Note that in real-world applications, the number of feasible routes can be at least 1000 and the solution space therefore can increase to infinity, precluding the ESM as a viable approach for solving the BTRNDP. In other words, the ESM can be successfully applied to find the global optimal bus route set for a very small size transit network. When the transit network is large, the solution space grows exponentially so the computation time using ESM to solve the BTRNDP grows exponentially. Note that in this research, the sole purpose of employing ESM to solve the BTRNDP is to use its solution as the benchmark to examine the efficiency and measure the quality of the solutions obtained by using heuristic algorithms including the GA, LS, SA, RS and TS. It is noted that GA, LS, SA, RS and TS are probably more reasonable approaches to solving the large-scale transit route network design problem in a reasonable time domain.

4.6 Summary

This chapter focuses on the proposed BTRNDP solution methodology. The solution methodology must account for the NP-hard nature and other related characteristics of the BTRNDP and explicitly incorporates practical guidelines and industry guidelines and rules of thumb.

The proposed solution framework consists of three main components: an Initial Candidate Route Set Generation Procedure (ICRSGP) that generates all feasible routes incorporating practical guidelines that are commonly used in the bus transit industry; a Network Analysis Procedure (NAP) that assigns transit trips, determines the service frequencies on each route and computes performance measures and a Heuristic Search Procedure (HSP) that guides the search techniques. Five heuristic algorithms, including the Genetic Algorithm, Local Search, Simulated Annealing, Random Search and Tabu Search Algorithms, as well as the Exhaustive Search Method (ESM) as a benchmark for the small network BTRNDP, are proposed to select an optimum route set from the huge solution space. The proposed solution framework and its distinct features as well as the overview of the ICRSGP, and the components of the NAP are discussed. The rationale for choosing these five heuristic algorithms as solution techniques is presented and a brief literature review of these six algorithms is also sequentially discussed. The details of the ICRSGP and NAP are described in Chapter 5 and 6 respectively. Implementation details of all proposed algorithms for the BTRNDP both with fixed and variable transit demand are presented in Chapters 7 and 8 respectively.

CHAPTER FIVE
THE INITIAL CANDIDATE ROUTE SET GENERATION PROCEDURE

5.1. Introduction

Chapter 4 presented the solution framework for the BTRNDP. The first component of this framework is the initial candidate route set generation procedure (ICRSGP), which is a design algorithm that configures all candidate routes for the current transit network. Given a user-defined minimum and maximum route length, ICRSGP finds all feasible routes for the current transit network.

As pointed out, the ICRSGP relies mainly on algorithmic procedures including the shortest path and k-shortest path algorithms. This chapter focuses on the details of the ICRSGP and is organized as follows. Section 5.2 presents an overview of the ICRSGP. Section 5.3 reviews the shortest path algorithm and k shortest path algorithm from a systematic approach. Label-setting and label-correcting algorithms are presented and compared. To illustrate these two algorithms, a case study for Dijkstra's algorithm and Yen's k-shortest path algorithm is also presented. Section 5.4 discusses the route feasibility constraints that are applied to the initial candidate route generation process. Section 5.5 contains applications of the Dijkstra's shortest path algorithm and Yen's k-shortest path algorithm to a small network. Section 5.6 concludes this chapter with a summary.

5.2. Overview of the Initial Candidate Route Set Generation Procedure

As mentioned, the initial candidate route set generation procedure (ICRSGP) is a design algorithm that: 1) is guided by user-defined parameters that incorporate the knowledge and expertise of the transit planners; 2) generates all candidate routes inside the current transit network using the algorithmic procedures including the shortest path and k-shortest path algorithms; 3) provides a solution space that can be used for the BTRNDP both with fixed and variable demand.

The ICRSGP requires a limited amount of input data, which can be simply grouped into two categories:

1) <u>Network</u>: The number of zones, the location of the centroid nodes, the user-defined specification of the locations of the distribution nodes on the road links for each centroid node, the specification of each intersection node, the link connectivity list specifying for each node its accessible neighboring nodes as well as the distance (on the road network) to each of neighboring nodes.
2) <u>Parameters</u>: The average traveling speed of each bus, the minimum route length and the maximum route length.

The overall skeleton of the ICRSGP is presented in Figure 5.1, where the structure of the ICRSGP consists of the following steps:

Step 1. Generate routes by finding the shortest path between each centroid node pair in the studied bus transit network;
Step 2. Check all the routes that were generated in Step 1 for the minimum and maximum route length constraints. If any route satisfies the constraints, then the route is accepted as a candidate route;
Step 3. Generate alternative routes by finding the k shortest path between the same centroid node pair as that in step 1;
Step 4. Check for feasibility constraints for each alternate route. If the route satisfies these constraints, then the alternate route is accepted as a candidate route. Otherwise, it is removed from the solution space;
Step 5. Stop the ICRSGP, store all the kept routes with their respective labels as the candidate route set.

The details of each step in the ICRSGP can be described as follows.

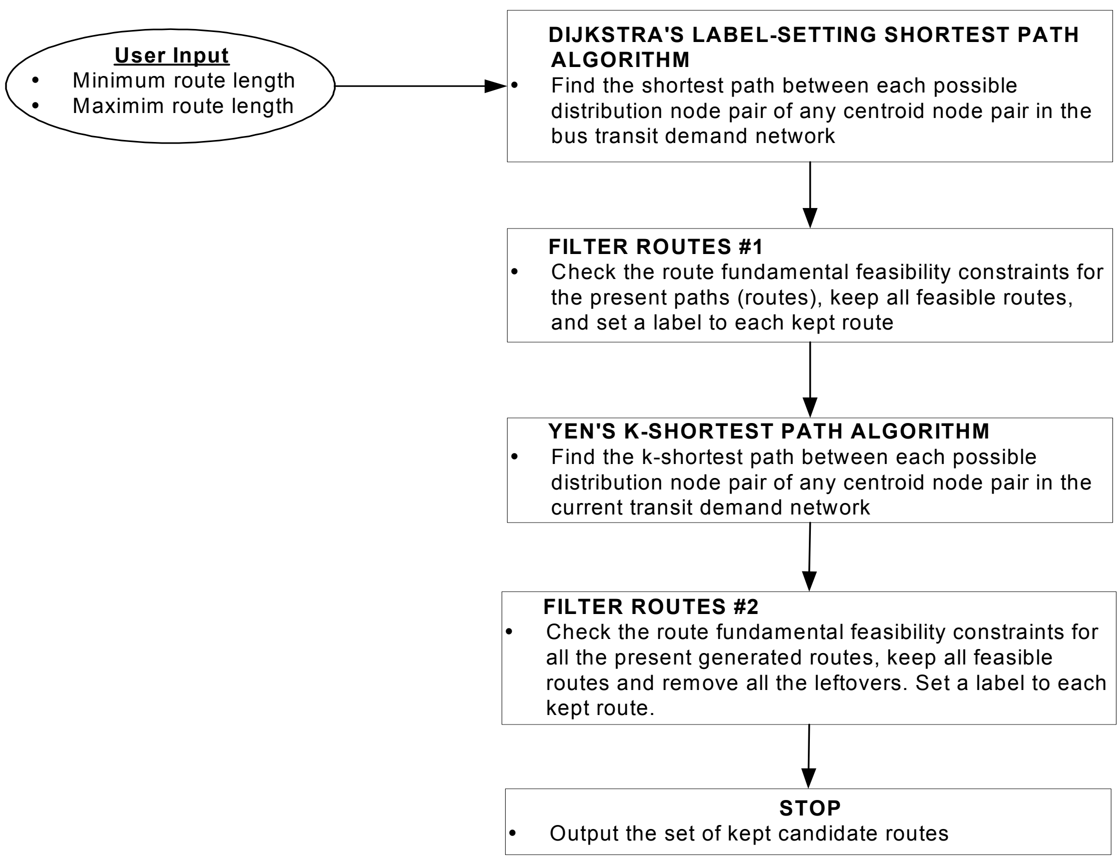


Figure 5.1 Skeleton of the Initial Candidate Route Set Generation Procedure

In Step 1, the ICRSGP starts by using Dijkstra's Label-Setting Shortest Path Algorithm to find the shortest path P(i, j) between each centroid node pair (i, j) in the bus transit demand network, each of these paths may form a feasible route element in the optimal bus transit route network solution space.

In Step 2, the initial set of all routes generated in Step 1 is checked by two feasibility filter tests: 1) the minimum route length constraint; and 2) the maximum route length constraint. Those routes that pass these two tests will be kept. The distance of these respective routes will be recorded and a label will be set to keep track of each one.

In Step 3, all the alternate routes are generated using modified Yen's k-th Shortest Path Algorithm between the same origin and destination as that generated in Step 1. At each step, the generated path is checked for Step 4 until all feasible routes have been generated.

In Step 4, the ICRSGA checks the route fundamental feasibility for each present route generated in Step 3. The feasibility constraints that are used here still include two components: 1) the minimum route length constraint; and 2) the maximum route length constraint. All feasible routes that satisfy these two constraints are kept and labeled, and all the leftovers are removed.

In step 5, the ICRSGP stores all the routes with their respective labels as elements of the overall candidate solution route set, and this set becomes input data to be evaluated in the network analysis procedure.

Note that the data structure of the ICRSGP, the network representation and the C++ algorithm implementation of the ICRSGP (and the NAP in Chapter 6) are discussed in Chapter 9. As can be easily seen, the ICRSGP relies mainly on algorithmic procedures including a shortest path algorithm and a k-shortest path algorithm. The following section presents a systematic literature review for these algorithms.

5.3. Overview of Shortest Path and K-Shortest Path Algorithms

5.3.1 Shortest Path Algorithm

The Shortest Path problem lies in the core of many transportation and logistics problems. Approaches for solving shortest path algorithms have been extensively researched in the past decades. The following section provides a brief literature review of the shortest path algorithm.

5.3.1.1 Notations

Consider a given a network $G = (N, A)$, where N is the set of all the nodes and A is the set of all the arcs in this network. Let C_{ij} be the distances (costs) associated with arc or link $(i, j) \in A$. Let node s be the source (origin) node and node i be any node other than s, i.e., $i \in N, i \neq s$ in the network. Define the length of a directed path as the sum of the distances of the arcs in the path. The Shortest Path Problem is essentially this kind of problem: for every node $i \neq s$, find a directed path of shortest length from s to i. To put it another way, the shortest path problem can be formulated as follows: *Given* 1) Network $G=(N, A)$; 2) $C_{ij,} (i, j) \in A$; and 3) Source node s and any node $i \neq s$, *the shortest path problem* is how to find the shortest path from the origin node s to node i. Note that the costs in this research are the travel times on specific links, which are always positive and therefore means that the network does not contain a negative cycle. Note that when handling any shortest path problem, this requirement must be satisfied. Otherwise, it will be substantially harder to solve the shortest path problem because with a negative cycle, it is an NP-complete problem, for which there exists no algorithm that can be solve this kind of problem in a polynomial time.

The shortest path problems that researchers have studied include three different types: 1) Single-source shortest path problems, which find the shortest path from one node to all others; 2) All-pairs shortest path problems, which find the shortest path from every node to every other node; and 3) Other generalizations of the shortest path problem. These different types of shortest path problems have very similar characteristics in solution methodology. It is expected that with some minor modifications, the solution algorithm for one type of shortest path problem can be applied to solve other types. Essentially speaking, the shortest path problem in this research belongs to the second type, namely, all-pairs shortest path problem. The following will discuss two distinct solution algorithms that are used to solve shortest path problems.

5.3.1.2 Label-Setting and Label-Correcting Algorithms

According to the network flow literature, the algorithmic approaches that can be applied to solve shortest path problems can be classified into two groups: label setting and label correcting. Both groups of algorithms are iterative and both employ the labeling method in computing one-to-all (all to all) shortest paths. The two groups of algorithms differ, however, in the ways in which they update the estimate (i.e., upper bound) of the shortest path associated with each node from step to step and how they "converge" to the optimal shortest path distances. In label-setting algorithms, one label will be designated as permanent (optimal) at each iteration. However, in label-correcting algorithms, all labels will be considered as temporary until they all become permanent in the final step. According to Ahuja, Magnanti and Orlin (1993), the relationship between Label-Setting and Label-Correcting Algorithms can be summarized in Table 5.1. A literature review of these two algorithms is conducted in the following chapters.

5.3.1.3 Label-Setting Algorithm: Dijkstra's Algorithm

The most basic label-setting algorithm is Dijkstra's Algorithm, which finds the shortest path from the source node s to all other nodes in a directed network $G = (N, A)$

with nonnegative arc lengths. Dijkstra's algorithm creates labels associated with nodes. These labels represent the distance (cost) from the source node to each particular node. Within the network, there exist two kinds of labels: temporary and permanent. Temporary labels are given to nodes that have not been reached. The value given to these temporary labels can vary. The main idea of Dijkstra's algorithm is to change the temporary labels into permanent ones as the shortest path tree adds them. The permanent label of a node denotes the shortest path distance from the source node to the node. For any given node, there must be a permanent label or a temporary label, but not both.

For a mathematical description of Dijkstra's Algorithm, some notations are introduced. For node i, let $A(i)$ represent the arc adjacency list of node *i*. Let P denote the set containing all the nodes with permanent labels, and $\overline{\mathrm{P}}$ be the set containing all the nodes with temporary labels. Initially, every node has a temporary label and the distance from the source node to all other nodes are initialized to ∞. At each step, Dijkstra's algorithm chooses the node $i \in \overline{P}$ with the least temporary label distance, and makes it permanent, records its predecessor index, and updates the temporary values of all nodes $j \in A(i)$. Repeat this procedure until all nodes become permanent ones. Then, output the distance and the shortest path consisting of a series of the predecessor nodes. A formal description of Dijkstra's algorithm is given in Figure 5.2.

Dijkstra's Shortest Path Algorithm
begin
    d(i) = ∞ and p(i) = -1 for each node $i \in N$;
    d(s) = 0 and p(s) = 0;
    $\mathrm{P} = \Phi$ and $\overline{\mathrm{P}} = \mathrm{N}$;
    while |P|<N **do**
    begin
        Select a node i of smallest d(i) = min{d(j): $\mathrm{j} \in \overline{\mathrm{P}}$ }
        Make node i permanent and delete i from $\overline{\mathrm{P}}$, $\mathrm{P} = \mathrm{P} \cup \{\mathrm{i}\}$ $\overline{\mathrm{P}} = \overline{\mathrm{P}} - \{\mathrm{i}\}$
        for each (i, j) $\in$ A(i) **do**
            if d(j) > d(i) + C_{ij} **then**
            begin
                Do distance update
                d(j) : = d(i) + C_{ij} and pred(j) : = i;
            end;
    end;
end;

Figure 5.2 Dijkstra's Algorithm

Due to the easy-to-implement characteristics of Dijkstra's Algorithm, this algorithm is employed in this research to find all feasible shortest paths between each centroid node pair.

Table 5.1 Label-Setting and Label-Correcting Algorithm Comparison

	Shortest Path Algorithms	
	Label-Setting Algorithm	***Label-Correcting Algorithm***
Typical Algorithms	• Dijkstra’s	• Bellman-Ford • Floyd-Warshall
Similar Characteristics	• iterative; • assign tentative distance labels to nodes at each step; • label setting can be viewed as a special case of label-correcting	
Different Characteristics	updating and converging	
	• designate one label as permanent at each iteration	• consider all labels as temporary until the final step, when they all become permanent
	Applications	
	• acyclic networks with arbitrary arc lengths (could be negative) • non-negative arc lengths • much more efficient (better worst-case complexity bounds) but less flexibility	• label-correcting can be applied to all classes of problems • more algorithmic flexibility but less efficient

5.3.2 K Shortest Path Algorithm

Compared to the shortest path problem, the k shortest path problem receives much less attention. However, several researchers, such as Yen (1971) and Lawler (1977), have made k shortest path a full-fledged field. In the literature, the algorithmic approaches that can be applied to solve the k shortest path problems can also be classified into two groups: label setting and label correcting. Due to the easy-to-implement characteristics of Yen's k shortest path algorithm (1971), this algorithm is employed in this research to find all feasible k shortest paths between node pairs. However, some modifications are performed to accommodate the requirements of the BTRNDP. A small example is also given to show how these algorithms work step by step. The following presents the modified k shortest path algorithm that was presented by Yen (1971) for finding the K loopless paths that have the shortest lengths from one node *s* to another node *i* in a network.

5.3.2.1 Notations:

In this subsection, notations are introduced for describing this algorithm. (Note that these notations are only confined to this section.) Suppose that a network has N-nodes, and let

i ------ node in the network, $\mathrm{i} = 1, 2, \cdots, \mathrm{N}$, where 1 is the origin node (source) and N is the destination node (sink);
$1 - i - \cdots - j,\ i \neq j \neq \cdots \neq 1$ ------ one loopless path from 1 to *j*, passing through $i, \cdots$;
N_k ------ the number of the nodes included in the kth shortest path;
$Q_i^{\ k}$ ------ the *i*-th node of the k-th shortest path.

Since the concept of computing k-shortest paths between all node pairs is essentially the same as that of the single source-sink node pair, it is reasonable to only consider the algorithm for computing single source-sink k-shortest path. Therefore, for the convenience of the description, one can set $Q_1^{\ k} = 1,\ Q_{N_k}^{\ \ k} = N$. Further notation regarding this algorithm is given as follows:

$A^k = 1 - Q_2^{\ k} - Q_3^{\ k} - \cdots - Q_{(N_k - 1)}^{\ \ k} - N, \quad k = 1, 2, \cdots, K$ ------ the *k*th shortest path from 1 to N where $Q_2^{\ k}, Q_3^{\ k}, \cdots, Q_{(N_k-1)}^{\ \ k}$ are respectively the 2nd, 3rd, $\cdots$, (N_k - 1)st nodes of the *k*-th shortest path;

$A_i^{\ k} = A_{Q_t}^{\ \ k},\ t = 1,\ 2,\ 3,\ \cdots,\ (N_k - 1)$ ------ a set of paths that deviat from A^{k-1} at *i*, (i.e., Q_t), the t-th node of the k-th shortest path. Put it in another way, $A_i^{\ k}$ is the shortest of the paths that coincide with A^{k-1} from the source node 1 to the t-th node i (*i.e.* Q_t)

on the path and then deviate to a node that is different from any of the (t+1)st nodes of those $A^j, j=1,2,\cdots,k-1,$ which have the same subpaths from 1 to the t-th nodes as does A^{k-1}; and finally it gets to the sink node N by a shortest subpath without passing any node that has already been included in the first part of this path. At this point, it should be noted that $A_{Q_t}{}^k$ is a loopless path, which means that all nodes can only appear once in each path;

The candidate k-th shortest path $A_i{}^k$ can be partitioned into two parts, with the first part being $R_i{}^k$ and the second part (also the last part) being $S_i{}^k$. $R_i{}^k$ ------ the first part of $A_i{}^k$ coincides with A^{k-1} from the first node 1 to the t-th node i, i.e., $1-Q_2{}^k-Q_3{}^k-\cdots-Q_t{}^k$ in $A_i{}^k$; and $S_i{}^k$ ------ the second part of $A_i{}^k$ consisting of the last part of $A_i{}^k$ has only one node coinciding with A^{k-1}, which is i, the t-th node of the shortest path. i.e., the second part of $A_i{}^k$ is $(Q_t{}^k)-Q_{t+1}{}^k\cdots Q_{(N_k-1)}{}^k-N$ in $A_i{}^k$.

5.3.2.2 Modified Yen's K Shortest Path Algorithm

Figure 5.3 presents the modified Yen's K shortest path algorithm for computing single source-sink k-shortest paths.

5.3.2.3 Remarks for Yen's K-Shortest Path Algorithm

Yen is one of the first researchers to propose an algorithm to solve the k-shortest path problem. The most significant part of Yen's k-shortest path algorithm lies in its straightforward description and easy-to-implement characteristics. The whole procedure of this algorithm can be organized as follows: First, use a standard shortest path algorithm, (e.g. Dijkstra's algorithm) to find the shortest path from the source node to the sink node and store it in the results SET A. Then every node in the shortest path except the destination node will be selected once and for each such node, the path from the source node to the current node is defined as the first part of the path. The second part (also the last part) of the path is obtained from the calculation result of another shortest path from each selected node to the destination node. It should be emphasized that for the second part of the path, two restrictions are placed:

algorithm K-shortest path;

begin:

Find A^1 ------ the shortest path using Dijkstra's algorithm. Store

$$A^1 = 1 - Q_2{}^1 - Q_3{}^1 - \cdots - Q_{(N_1-1)}{}^1 - N \text{ in the SET A.}$$

$k = 1$;

Num = N_k;

while $k \leq$ K **do**

begin

$k = k+1$;

for each $t = 1,\ 2,\ 3,\ \cdots,\ (N_{k-1} - 1)$ **do**

- Check if the subpath consisting of the first t nodes of A^{K-1} in sequence coincide with the subpath consisting of the first t nodes of A^j in sequence for $j = 1, 2, \cdots, k-1$. If so, set $d_{tq} = \infty$ --- where (q) is the (t+1)st node of A^j; otherwise, make no changes.
- Apply Dijkstra's algorithm to find the shortest path from Q_t, the t-th node to N, allowing it to pass through those nodes that are not yet included in the path. Note that the subpath from node 1 to Q_t is $R_i{}^K$, the first part of $A_i{}^K$; and the subpath from Q_t to N is $S_i{}^K$, the last part of $A_i{}^K$. Note also if there is more than one subpath from (i) to (N), that have the minimum length, take any arbitrary one of them and denote it by $S_i{}^K$.
- Find $A_i{}^K$ by joining $R_i{}^K$ and $S_i{}^K$. Then add $A_i{}^K$ to SET B.

Find from SET B the path(s) that have the minimum length. Denote this path (or an arbitrary one, if there are more than one such paths by A^K and move it from SET B to SET A and leave the rest of the paths in SET B.

end;

end;

Figure 5.3 Modified Yen's K Shortest Path Algorithm

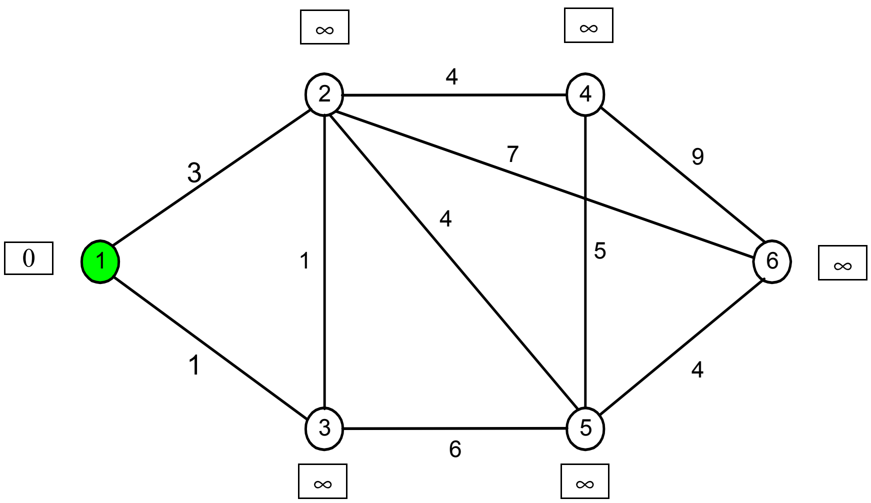
1) It is a loopless subpath, i.e., it cannot pass through any node that has already been included in the path;
2) It cannot branch from the current node on any edge used by a previously found (k-1) shortest path. The node and edge marking procedure that is used in this algorithm can prevent the second part of the path from looping or simply following the route of a previous (k-1) shortest path. If a new second part of the path is found, it will be appended to the first part of the path to form a complete path from source to sink node, with the current node being the connecting node. This path forms a candidate for the next KSP. All such paths that were found this way will be stored in SET B, in which the path that has the shortest distance will be selected as the next k-th shortest path, and will be transferred to the results set

in SET A. The same process will be repeated, until the required number of k shortest paths or some other requirements has been met.

Based on the above literature review, a case study is performed to illustrate these two algorithms and the details are presented step by step.

5.3.3 Case Study: Dijkstra's Algorithm and Yen's K-Shortest Path Algorithm

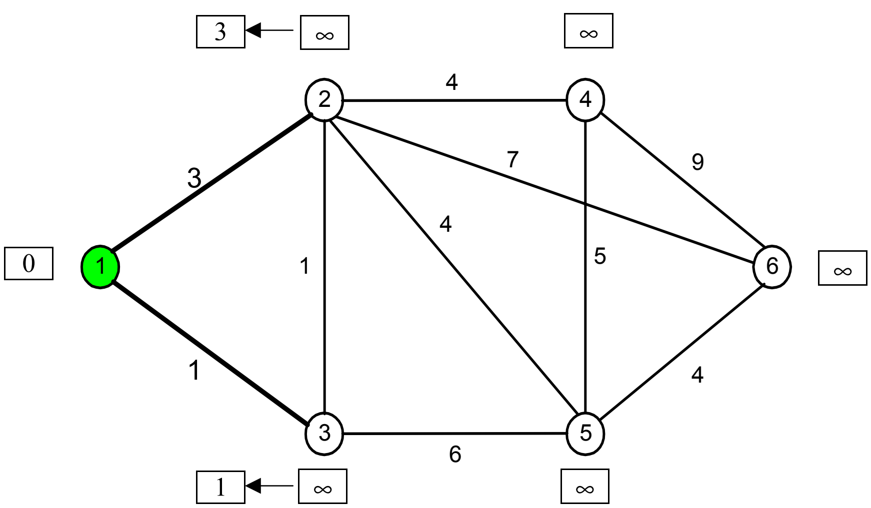
A worked example for Label-Setting---Dijkstra's algorithm is as follows:



Step 1:

$P = \{1\}, \bar{P} = \{2,3,4,5,6\}$
$|P| = 1$

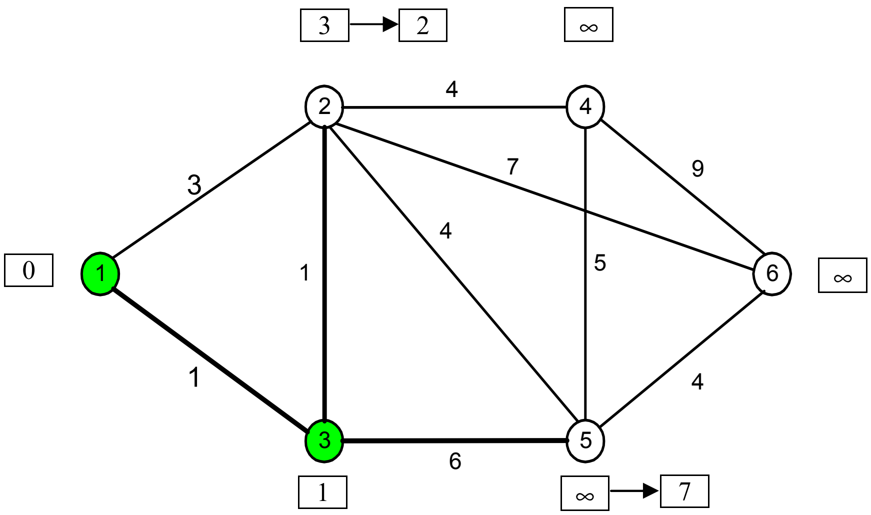
i	1	2	3	4	5	6
d(i)	0	∞	∞	∞	∞	∞
p(i)	0	-1	-1	-1	-1	-1



Step 2:

$P = \{1\}, \bar{P} = \{2,3,4,5,6\}$
$|P| = 1$

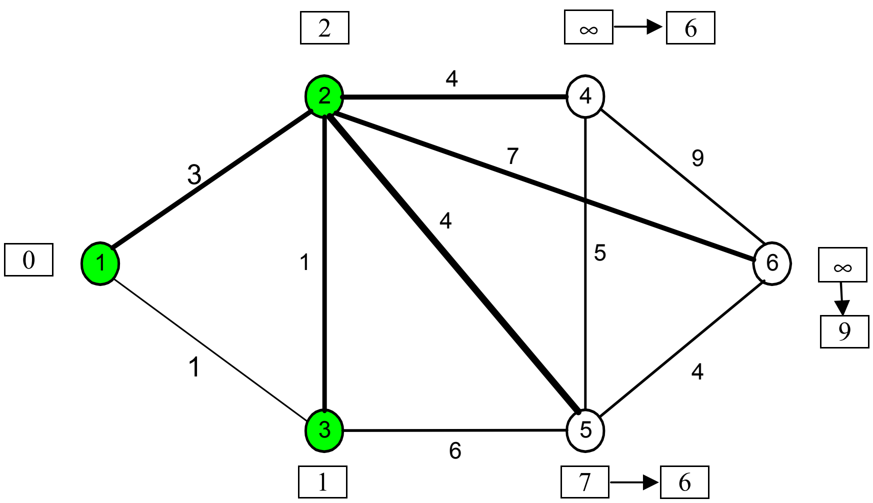
i	1	2	3	4	5	6
d(i)	0	3	1	∞	∞	∞
p(i)	0	1	1	-1	-1	-1



Step 3:

$P = \{1,3\}, \overline{P} = \{2,4,5,6\}$
$|P| = 2$

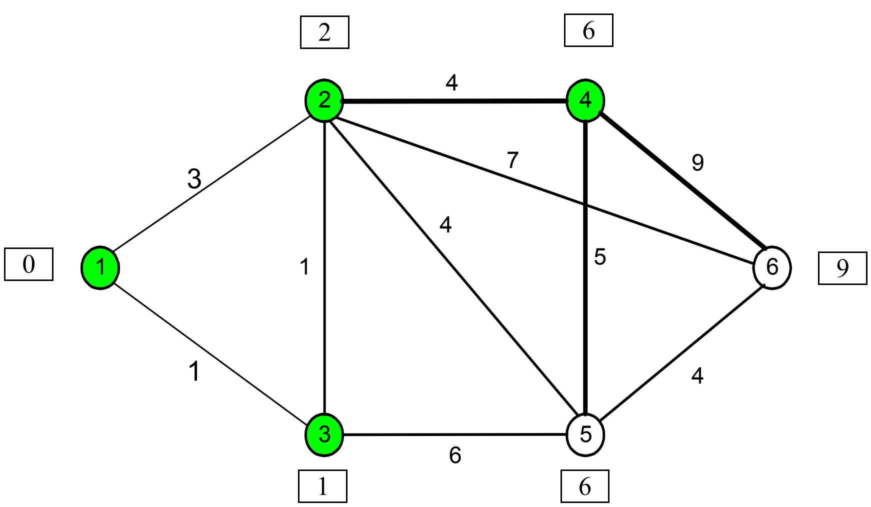
i	1	2	3	4	5	6
d(i)	0	2	1	∞	7	∞
p(i)	0	3	1	-1	3	-1



Step 4:

$P = \{1,2,3\}, \overline{P} = \{4,5,6\}$
$|P| = 3$

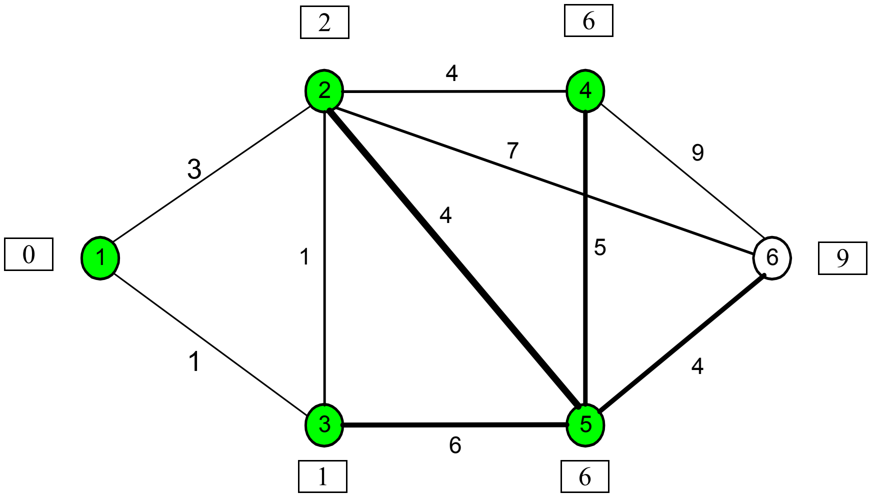
i	1	2	3	4	5	6
d(i)	0	2	1	6	6	9
p(i)	0	3	1	2	2	2



Step 5:

$P = \{1,2,3,4\}, \overline{P} = \{5,6\}$
$|P| = 4$

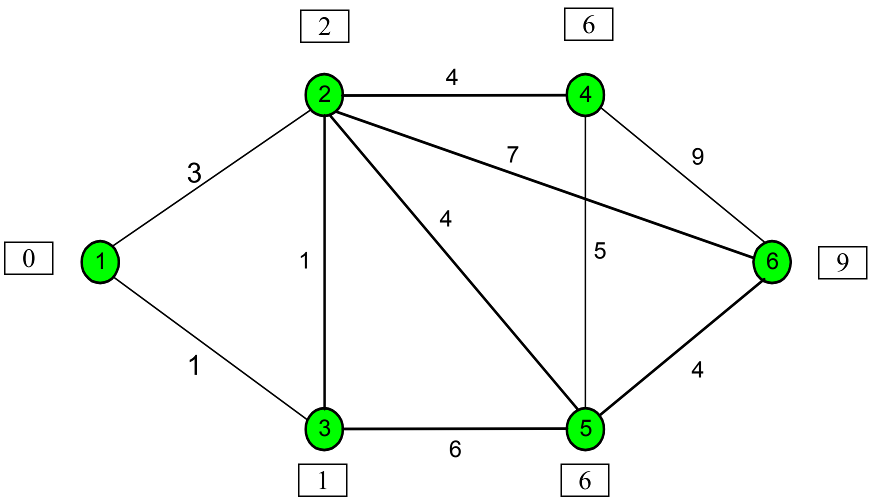
i	1	2	3	4	5	6
d(i)	0	2	1	6	6	9
p(i)	0	3	1	2	2	2



Step 6:

$P = \{1,2,3,4,5\}, \overline{P} = \{6\}$
$|P| = 5$

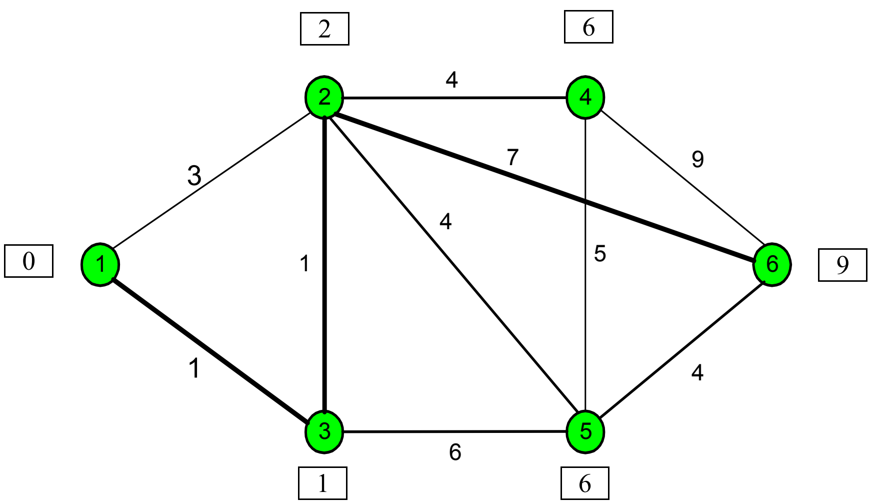
i	1	2	3	4	5	6
d(i)	0	2	1	6	6	9
p(i)	0	3	1	2	2	2



Step 7:

$P = \{1,2,3,4,5,6\}, \overline{P} = \Phi$
$|P| = 6$

i	1	2	3	4	5	6
d(i)	0	2	1	6	6	9
p(i)	0	3	1	2	2	2

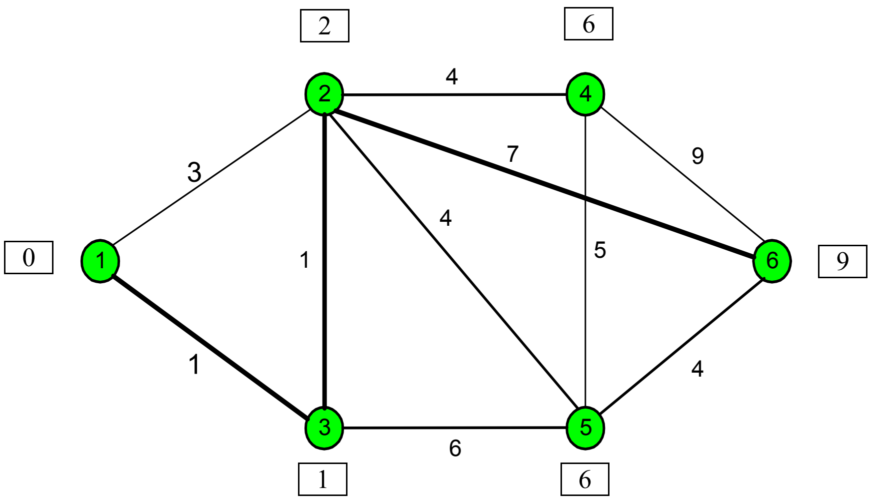


Step 8:

$P = \{1,2,3,4,5,6\}, \overline{P} = \Phi$
$|P| = 6$

Output the shortest path

Given the shortest path results, Yen proposed an algorithm to solve the k-shortest path problem. The skeleton of this algorithm has been described before. For convenience and consistency, the same example network that was used in Dijkstra's algorithm is used to show how this algorithm works.



Step 1:

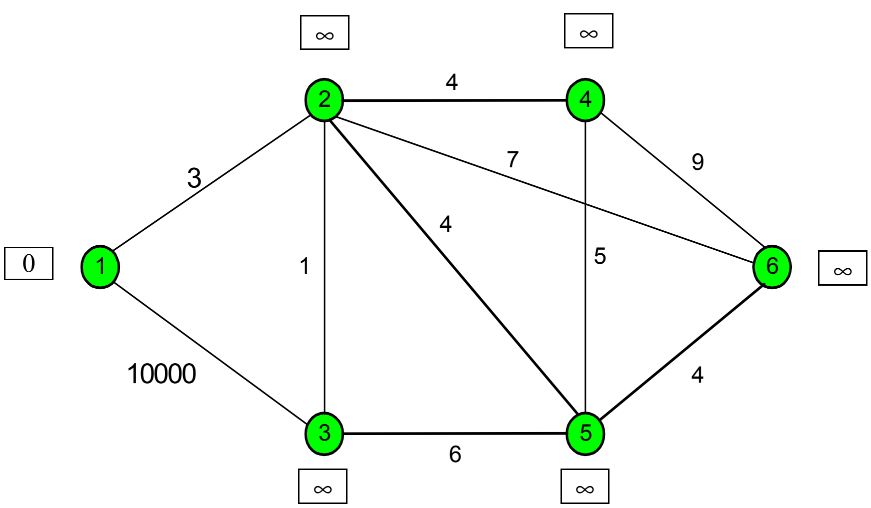
SET A: A^1 = 1-3-2-6

$N_1 = 4$

$A^2 = \Phi$

SET B: $B = \Phi$

1. Set $C_{13} = \infty$ Using Dijkstra's algorithm to find the shortest path from source node 1 to sink node 6.



Step 2:

SET A: A^1 = 1-3-2-6

$N_1 = 4$

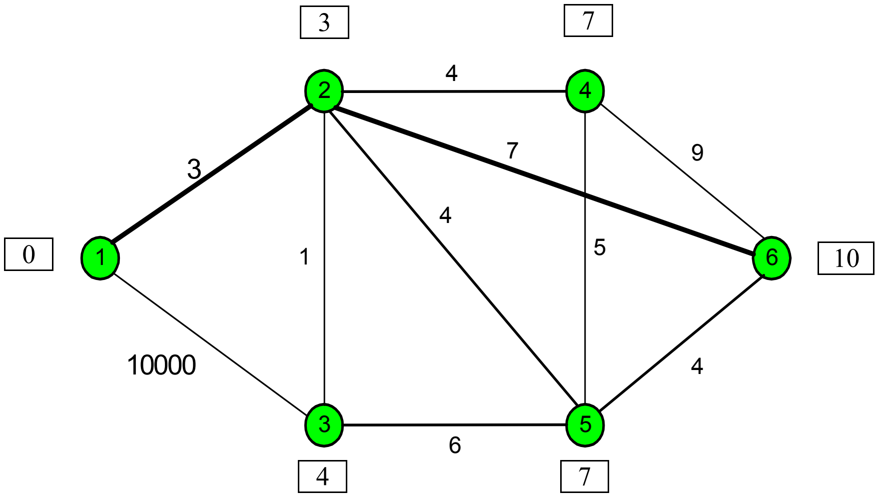
$A^2 = \Phi$

SET B: $B = \Phi$

Select node 1.

$R_1^{\,2} = 1; S_1^{\,2} = \Phi$;

Set arc $C_{13} = \infty$



Step 3:

SET A: A^1 = 1-3-2-6

$N_1 = 4$

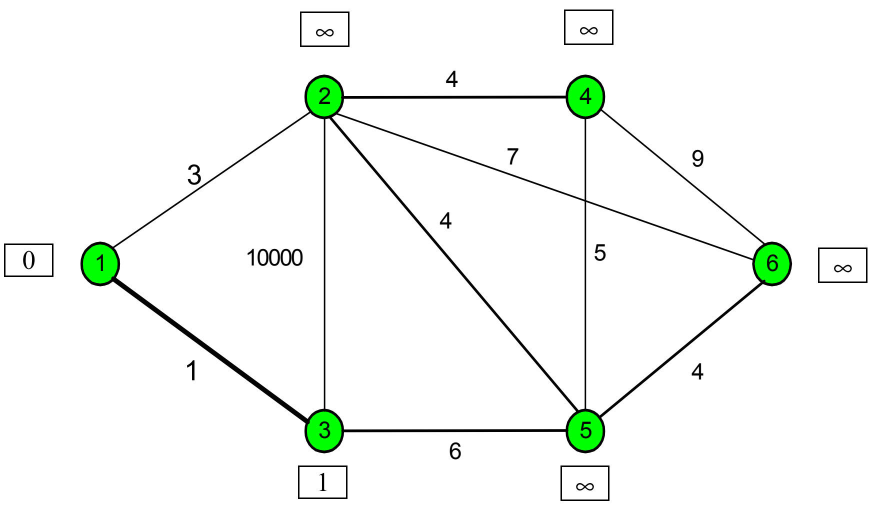
$A^2 = \Phi$

SET B: $R_1^{\ 2}$ =1; $S_1^{\ 2}$ =1-2-6;

$A_1^{\ 2}$ =1-2-6;

$d_1^{\ 2}$ =10;

2. The first arc 1→3 is recorded as the root path. Since 3→2 is used in the previous shortest path, it cannot be used again in the 2-nd shortest path. Therefore, set $C_{32} = \infty$ Using Dijkstra's algorithm to find the shortest path from node 3 to sink node 6.



Step 4:

SET A: A^1 = 1-3-2-6

$N_1 = 4$

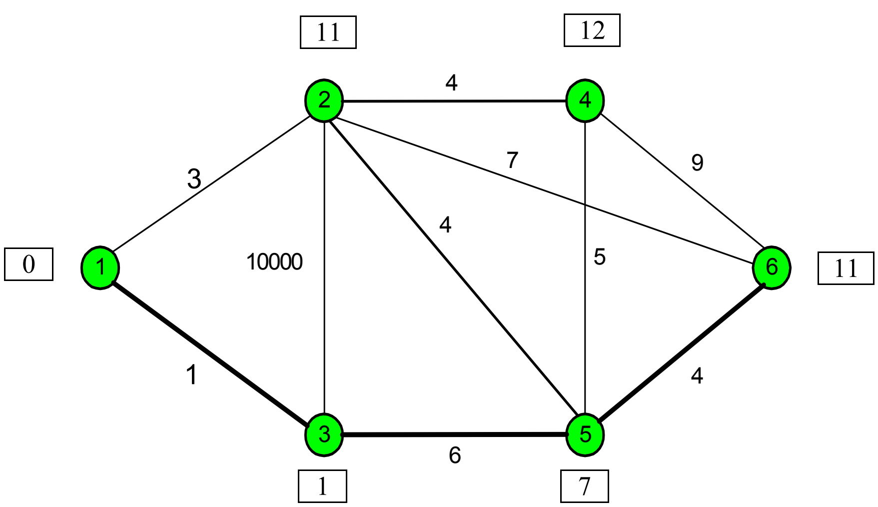
$A^2 = \Phi$

SET B: $A_1^{\ 2}$ =1-2-6;

$d_1^{\ 2}$ =10;

$R_2^{\ 2}$ =1-3; $S_2^{\ 2} = \Phi$;

Set arc $C_{23} = \infty$



Step 5:

SET A: A^1 = 1-3-2-6

$N_1 = 4$

$A^2 = \Phi$

SET B: $A_1^{\ 2}$ =1-2-6;

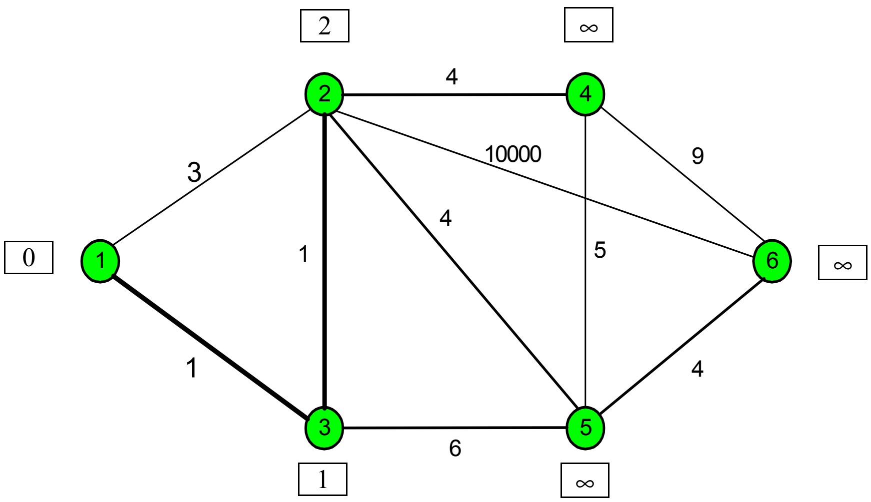
$d_1^{\ 2}$ =10;

$R_2^{\ 2}$ =1-3; $S_2^{\ 2}$ =3-5-6**;**

$A_2^{\ 2}$ =1-3-5-6;

$d_1^{\ 2}$ =11;

3. The first two arcs 1→3→2 are recorded as the root path. Since 2→6 is used in the previous shortest path, it cannot be used again in the 2-nd shortest path. Therefore, set $d_{26} = \infty$ Using Dijkstra's algorithm to find the shortest path from node 2 to sink node 6.



Step 6:

SET A: A^1 = 1-3-2-6

$N_1 = 4$

$A^2 = \Phi$

SET B: $A_1^{\ 2}$=1-2-6;

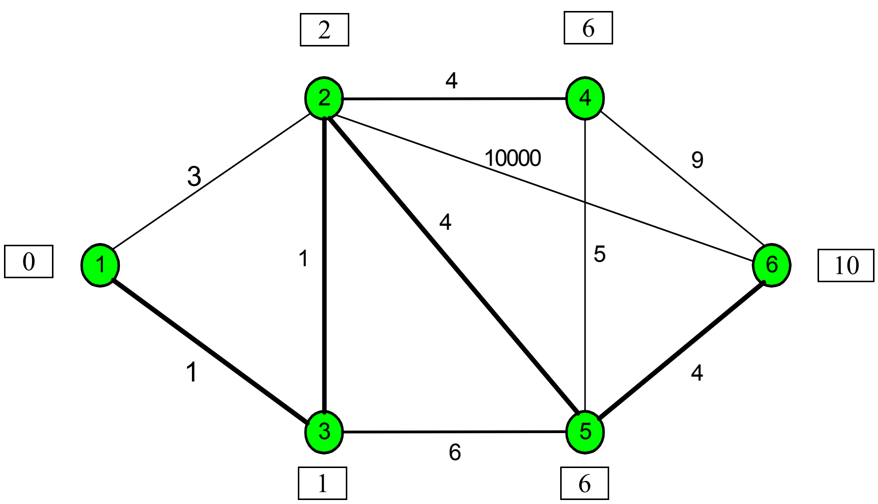
$d_1^{\ 2}$=10;

$A_2^{\ 2}$=1-3-5-6;

$d_1^{\ 2}$=11;

$R_3^{\ 2}$=1-3-2; $S_3^{\ 2} = \Phi$;

Set arc $C_{26} = \infty$



Step 7:

SET A: A^1 = 1-3-2-6

$N_1 = 4$

$A^2 = \Phi$

SET B: $A_1^{\ 2}$=1-2-6;

$d_1^{\ 2}$=10;

$A_2^{\ 2}$=1-3-5-6;

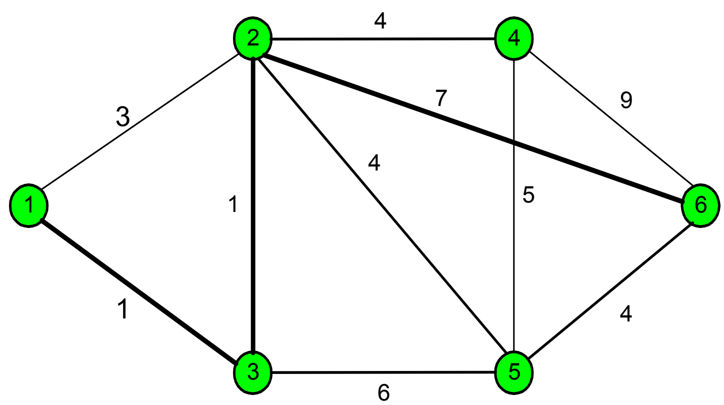
$d_1^{\ 2}$=11;

$R_3^{\ 2}$=1-3-2; $S_3^{\ 2}$ = 2-5-6;

$A_3^{\ 2}$=1-3-2-5-6;

$d_3^{\ 2}$=11;

Therefore, three paths will be stored in LIST B. (10, 11, 10). Since there are two paths whose shortest path value are the smallest in these paths, one can pick any path out of LIST B and transfer it to A. Meanwhile, delete it and leave all other paths in LIST B.



Step 8:

SET A: A^1 = 1-3-2-6

$N_1 = 4$

A^2 =1-2-6

$N_2 = 3$

SET B: A_2^2 =1-3-5-6;

A_3^2 =1-3-2-5-6

The same procedure is repeated until the required number of k shortest paths or some requirement is met (For the BTRNDP as shown later, this requirement is that the maximum route length cannot be exceeded). As a result, the k-shortest paths that can meet the user's requirement can be found and applied for future use.

5.4. Route Feasibility Constraints

Different from many previous approaches, the route feasibility constraints in this research only include two components: a minimum route length constraint and a maximum route length constraint. Furthermore, these two components are user-defined and the details of these two constraints have been described in Chapter 2.

5.5. Applications for a Small Network

To illustrate the ICRSGP, a small example network with 7 zones and 15 road intersections is designed, and is shown in Figure 5.4.

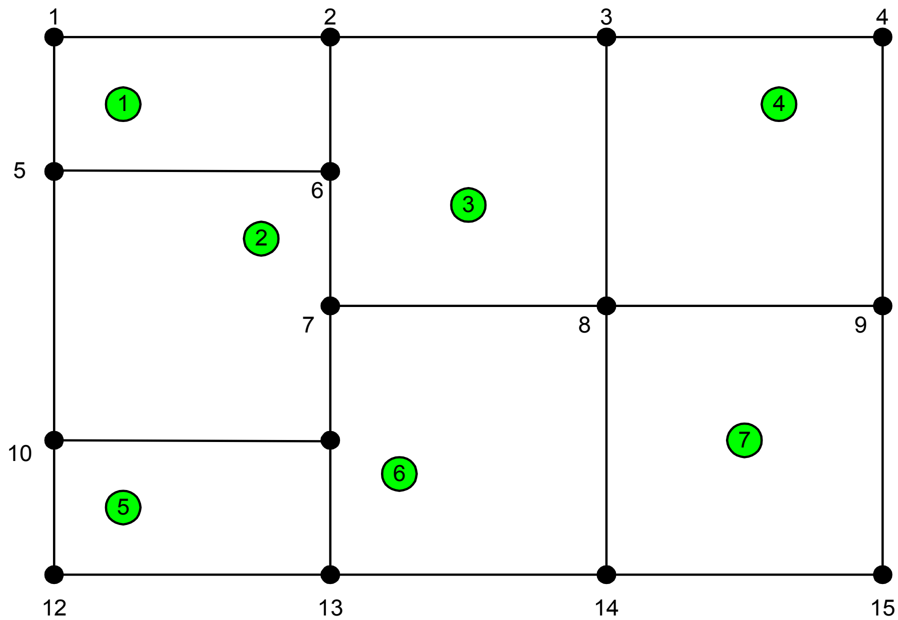


Figure 5.4 A Small Network with Road Structures and the Centroid Node of Several Zones

As noted before, the ICRSGP discussed in this chapter first considers the BTRNDP under the "centroid" level. The centroid location of the each traffic zone and the locations of the distribution nodes on the road links for each zone are specified by the users (commonly transit planners). The location specification of each intersection node, the connectivity list specifying for each node and its accessible neighboring nodes, as well as the distance (on the road network) to each of neighboring nodes, the average bus speed, the minimum and maximum route length are specified by the users. Figure 5.5 shows an intermediate process for distribution nodes for this example network and Figure 5.6 presents the final chosen distribution node specifications of this small network.

After the network "transformation" processing, 20 centroid distribution nodes, 35 nodes, and 82 arcs are obtained in this example network. Excluding the possible routes whose origins and/or destinations are the same zone, the number of overall routes (undirected) between "meaningful" pairs is 147. Furthermore, if the minimum and maximum route length is defined as 800 and 1600 meters respectively, then there are 100 feasible routes (out of 147 total routes) whose distances satisfy these two route length constraints and there are 47 nonfeasible routes that are generated from this example network. As shown, after obtaining all feasible shortest paths, the k-th shortest path algorithm comes into play. Namely, for the same origin/destination node pair in a certain shortest path, the second, third and k-th shortest paths are found until the distance of these routes violated the user-defined route length constraints. In this case, the k-shortest path algorithm stops temporarily and the current generated k-th shortest paths are appended to the feasible route set and kept as possible members of the solution route set. The whole process is repeated until all the current 147 meaningful origin/destination pairs in the network are completed. Note that for this example network, 186 additional paths of the k-th shortest path category are generated. As a result, the candidate route set for this example network consists of 286 paths, any of which is used as part of the solution route set. Table 5.2 presents a detailed representation of all the solution route space for the example network. Note that the detail of the network representations using C++ are presented in Chapter 9 and details of finding an optimal transit route network are presented in the next chapters.

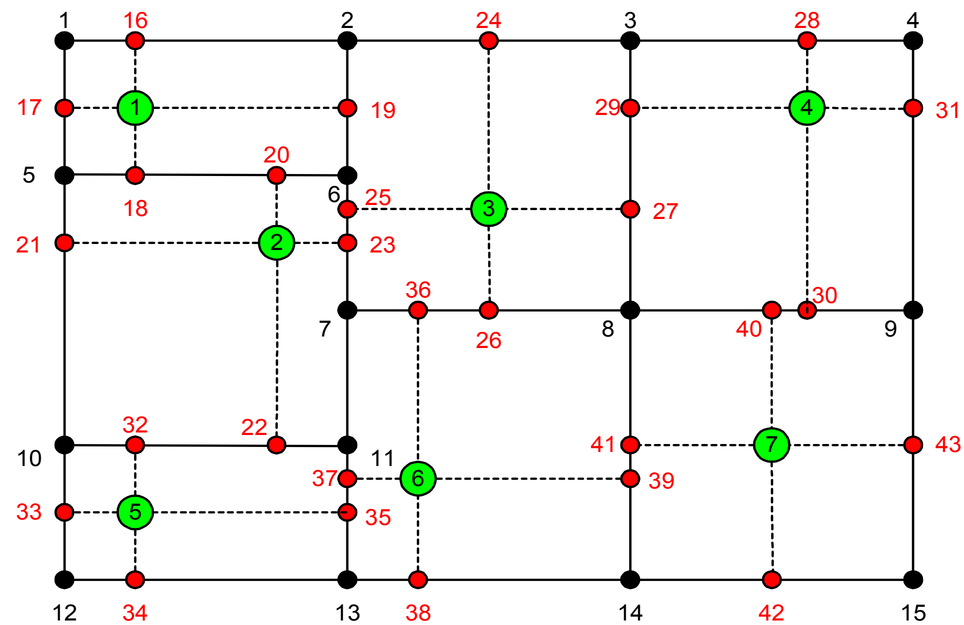


Figure 5.5 Intermediate Processes for Distribution Nodes

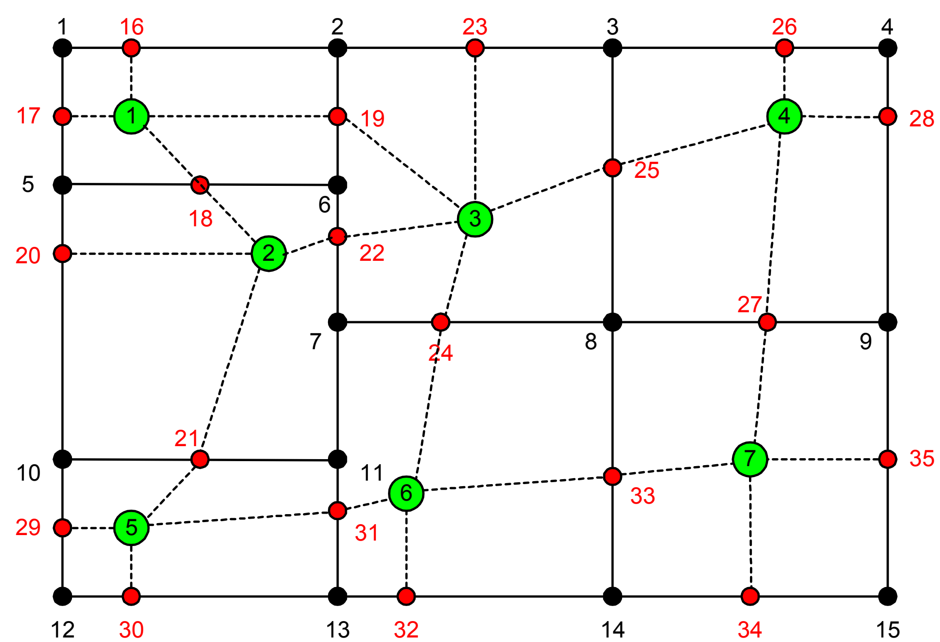


Figure 5.6 A Small Network with Distribution Nodes for the ICRSGP Illustration

Table 5.2 Representation of the Solution Route Space for the Example Network

Route Number	Orig.	Dest.	Shortest Path or k-th Shortest Path Representations of Nodes and Links	Dist.	Number of Nodes	Traversed Zones	Number of Zones Traversed
1	16	21	21-10-20-5-17-1-16	900	7	1,2,5	3
2	16	24	24-7-22-6-19-2-16	850	7	1,2,3,6	4
3	16	25	25-3-23-2-16	875	5	1,3,4	3
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
98	30	35	35-15-34-14-32-13-30	1300	7	5,6,7	3
99	31	35	35-15-34-14-32-13-31	1125	7	5,6,7	3
100	32	35	35-15-34-14-32	900	5	6,7	2
101	16	21	21-11-7-22-6-19-2-16	1100	8	1,2,3,5	4
102	16	21	21-11-7-22-6-18-5-17-1-16	1300	10	1,2,3,5	4
103	16	21	21-10-20-5-18-6-19-2-16	1500	9	1,2,3,5	4
104	16	24	24-7-22-6-18-5-17-1-16	1050	9	1,2,3,6	4
105	16	24	24-8-25-3-23-2-16	1350	7	1,3,4,6	4
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
284	31	35	35-9-27-8-33-14-32-13-31	1525	9	4,5,6,7	4
285	32	35	35-9-27-8-33-14-32	1300	7	4,6,7	3
286	32	35	35-9-27-8-24-7-11-31-13-32	1500	10	3,4,5,6,7	5

5.6. Summary

This chapter focuses on the details of the initial candidate route set generation procedure (ICRSGP). The overall solution framework for the ICRSGP and its distinct features are presented. The shortest path and k-shortest path algorithms, the two algorithmic procedures that the ICRSGP heavily relies on, are reviewed systematically. Label correcting and label setting algorithms are discussed and compared. The two chosen algorithms, Dijkstra's shortest path algorithm and Yen's k-shortest path algorithm are presented in detail. The route feasibility constraints are also discussed. A case study on applications of these two algorithms is illustrated. The next chapter is mainly oriented to the details of the network analysis procedures (NAP) for the BTRNDP with fixed transit demand.

CHAPTER SIX
THE NETWORK ANALYSIS PROCEDURE

6.1 Introduction

As mentioned before, there could be many feasible (and local optimal) solutions to the BTRNDP. To measure the quality of each proposed solution, the objective function components that are expressed in mathematical formulation must be evaluated. Similarly, a variety of other performance measures or service quality that are of concern to both the transit users and operator should be considered.

The Network Analysis Procedure (NAP) is a procedure that can analyze and evaluate the alternative network structures and determine their associated route service frequencies. For any given route network configuration in the solution space that was generated by the ICRSGP and proposed by a heuristic search procedure (HSP), the NAP determines route frequencies and evaluates the resulting transit network system by computing many performance measures that can reflect the quality of service, which include the user cost (i.e., the cost experienced by transit users), the operator cost (i.e., the fleet size) and the unsatisfied demand cost.

The NAP in this research differs from the previous approaches in three main features. First, for the first time, the NAP employs a multiple path transit trip assignment model that explicitly considers the transfer and long-walk related characteristics among routes under a much more real situation (namely, in the context of "centroid" node level instead of aggregating the traffic zone to a single node). Second, the NAP can explicitly consider the transit trip assignment model for the BTRNDP under two distinct scenarios, i.e., fixed transit demand and variable transit demands. Third, the NAP uses discrete choice modeling techniques and an iterative procedure to obtain the transit demand between each traffic zone pair, assign transit demand and determine the route frequencies when their internal convergences are achieved.

This chapter centers on the details of the NAP for the BTRNDP with fixed demand. Details of the NAP designed for the BTRNDP with variable transit demand are discussed in Chapter 8. This chapter is organized as follows. Section 6.2 presents an overview of the NAP, where the input data, output data, the general description, the algorithm skeleton and flow chart for the NAP are discussed. Section 6.3 describes the assumptions made for the NAP. Section 6.4 reviews the literature on the transit trip assignment model and then presents the detail of the transit trip assignment model for the NAP. In section 6.5, the frequency setting procedure (FSP), including the demand and policy frequency, the solution approach and its preset parameters are discussed. Section 6.6 focuses on the example network illustrations for the NAP, especially with the transit trip assignment procedure. Finally, section 6.7 concludes this chapter with a summary.

6.2 Overview of the NAP with Fixed Transit Demand

Basically speaking, the NAP is primarily used as a bus transit network evaluation tool with the incorporation of the ability to determine the route frequencies. After each possible set of candidate routes that define the network configuration is generated by the ICRSGP, a variety of data inherent in the bus transit route network system are obtained as its output data. Most of these data act further as input data and can be evaluated and analyzed via the NAP. As a result, the optimal route configuration for a particular network and its route frequencies are then obtained in the output data from the NAP. Figure 6.1 illustrates the relations between the ICRSGP and the NAP.

The following two subsections present the input data that are required for the NAP and the output data that are generated via the NAP.

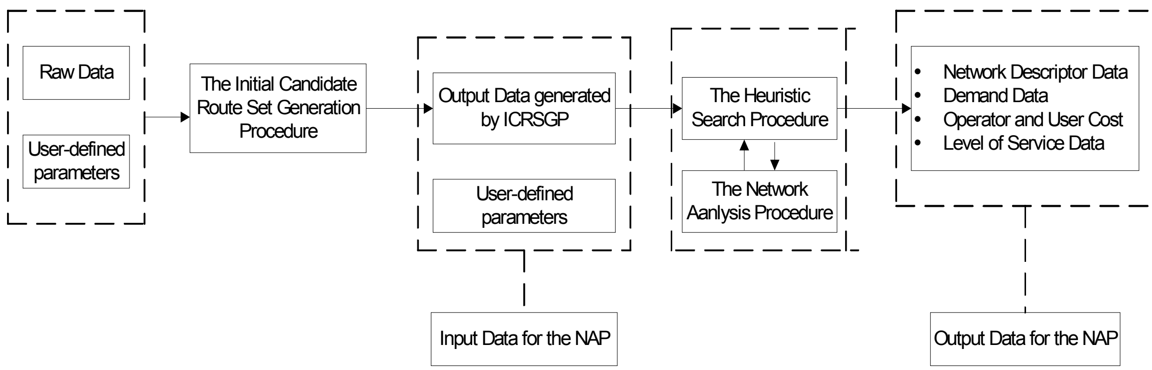


Figure 6.1 Relations between the ICRSGP and the NAP

6.2.1 Input Data for the NAP

As shown in Figure 6.1, the input data needed for the NAP includes the output data generated from ICRSGP, some user-defined parameters and the solution route set proposed by the HSP, which can be summarized as follows:

1) Network Information, which contains the number and the location (coordinates) of the centroid nodes where the demand originates and/or head, the number and the location of demand distribution nodes on the road links that are connected with each centroid node through associated centroid connectors and all the feasible routes generated from the ICRSGP and their associated node connectivity lists;
2) Demand Data, which includes a symmetric (although not necessarily) demand matrix representing the number of passenger trips using bus transit between each pair of traffic zones;
3) Design parameters, which refers to some parameters that are set by the network designers: the initial service frequencies on each route; the transfer penalty per transfer (expressed in equivalent minutes of bus transit in-vehicle travel time); the bus

seating capacity; the maximum load factor on each route; and the different weights for operator cost, users cost and the unsatisfied demand cost. If the policy headway approach is employed, one also needs to specify the corresponding service headway for each bus route. In addition, as mentioned later in this chapter, a possible maximum number of bus routes and a possible maximum number of iterations in the algorithm implementation (namely, GA, SA, LS, RS, TS and ESM) for each specific set of bus routes may be included.

4) Proposed solution route set, which is generated and guided by the heuristic search procedure (HSP).

6.2.2 Output Data for the NAP

Once the above input data for the NAP has been generated by the ICRSGP and/or inputted by the network designers, these data are analyzed via the NAP. The proposed optimal transit route network configuration, the route frequencies and associated transit demand, along with a wide variety of performance measures, are then obtained in the output data from the NAP, which can be categorized as follows:

1. Network Description Data:
    1) Node information, which includes flows from and terminated at each centroid node, as well as the transfer flow at each link node.
    2) Link information, which contains link flows along each route. The maximum link flow on each route, along with maximum load factor, is used to determine the route frequency on each route. (The details are discussed in Section 6.5.)
    3) Route information, which includes round trip time, total number of passengers, and fleet size required on each route.)
2. Demand Data, which includes total number of demand trips of the following seven categories: unsatisfied, 0-transfer-0-longwalk, 1-transfer-0-longwalk, and 0-transfer-1-long-walk, 1-transfer-1-longwalk, 2-transfer-0-longwalk, 0-transfer-2-longwalk.
3. User Cost, Operator Cost and Unsatisfied Demand Cost, which contains the total cost experienced by the transit users (including four parts: access time, waiting time, in-vehicle travel time and transfer time if any), the operator cost (i.e., the required number of buses run on each route in the solution transit network configuration) and the penalty for the transit demand not satisfied by the current transit network.
4. Level of Service, which contains the service frequency and the load factor run on each route in the solution network.

6.2.3 Description of the NAP

As mentioned before and illustrated in Figure 6.1, the NAP is an analytical tool that can evaluate and analyze the input bus transit network and possesses the ability to determine the route frequencies. To accomplish these tasks, the NAP employs an iterative procedure that seeks to achieve internal consistency of route frequencies (and transit demand equilibration for the BTRNDP with variable demand). Furthermore, the iterative procedure in the NAP contains two major components, namely, the transit trip assignment model and the frequency setting procedure.

Once a specific set of routes is proposed in the overall candidate solution route set generated by the ICRSGP, the NAP is called to evaluate the alternative network structure and determine route frequencies. The whole NAP process can be described as follows. First, an initial set of route frequencies are specified because they are necessary before the beginning of the trip assignment process. Then, hybrid transit trip assignment models are utilized to assign the passenger trip demand matrix to a given set of routes associated with the proposed network configuration. Assuming that the transit tripmaker always attempts to complete his/her trip with an intention to avoid transfers and have the least walking distance possible, the trip assignment model provides transit user route service in the following order: 0-transfer-0-longwalk (direct service), 1-transfer-0-longwalk and/or 0-transfer-1-long-walk, 1-transfer-1-longwalk and/or 2-transfer-0-longwalk and/or 0-transfer-2-longwalk. If none of the above three categories of paths is available, then the transit demand involved are unsatisfied. The service frequency for each route is then computed and used as the input frequency for the next iteration in the transit trip assignment and frequency setting procedure. If these route frequencies are considered to be different from previous input frequencies by a user-defined parameter (for example, they differ by more than 10%), the process iterates until the internal consistency of route frequencies is achieved. Once this convergence is achieved, route frequencies and several system performance measures (such as the user cost, the fleet size and the (un-)satisfied transit demands) are thus obtained.

Figure 6.2 presents the flow chart for the NAP. The details of the transit trip assignment model and frequency setting procedure are discussed in Section 6.4 and 6.5 respectively. The algorithm skeleton of the NAP is presented in the following subsection.

6.2.4 Algorithm Skeleton

In summary, the solution algorithm skeleton of the NAP can be described by the following steps:

Step 1. Set an initial frequency for each bus route that defines the whole bus transit network configuration.

Step 2. Assign the transit demand to the proposed solution route network.

a. Set $i = 1, j = 1$.

b. Assign the transit demand d_{ij} into paths of four possible categories: direct route service at the first level (i.e., 0-transfer-0-longwalk), 1-transfer-0-longwalk and/or 0-transfer-1-long-walk paths at the second level, 1-transfer-1-longwalk and/or 2-transfer-0-longwalk and/or 0-transfer-2-longwalk paths at the third level and unsatisfied demand at the fourth level according to the rules associated with each category, and update related performance measures and the network descriptors.

c. If $j \le \mathrm{N}$ (namely, the total number of traffic zones), then set $j = j + 1$ and repeat 2.2.

d. If $i \le \mathrm{N}$, then set $i = i + 1$ and $j = 1$, repeat 2.2.

Step 3. Compute the service frequencies associated with each route.

Step 4. Check whether the service frequencies for each route converge respectively during two consecutive steps. If not, then replace the input route frequencies for the next iteration with the current computed service frequencies, goto Step 2.

Step 5. Output the service frequencies on each route, compute the user cost, operator cost and the unsatisfied demand cost, the objective function value and all the other system performance measure descriptors.

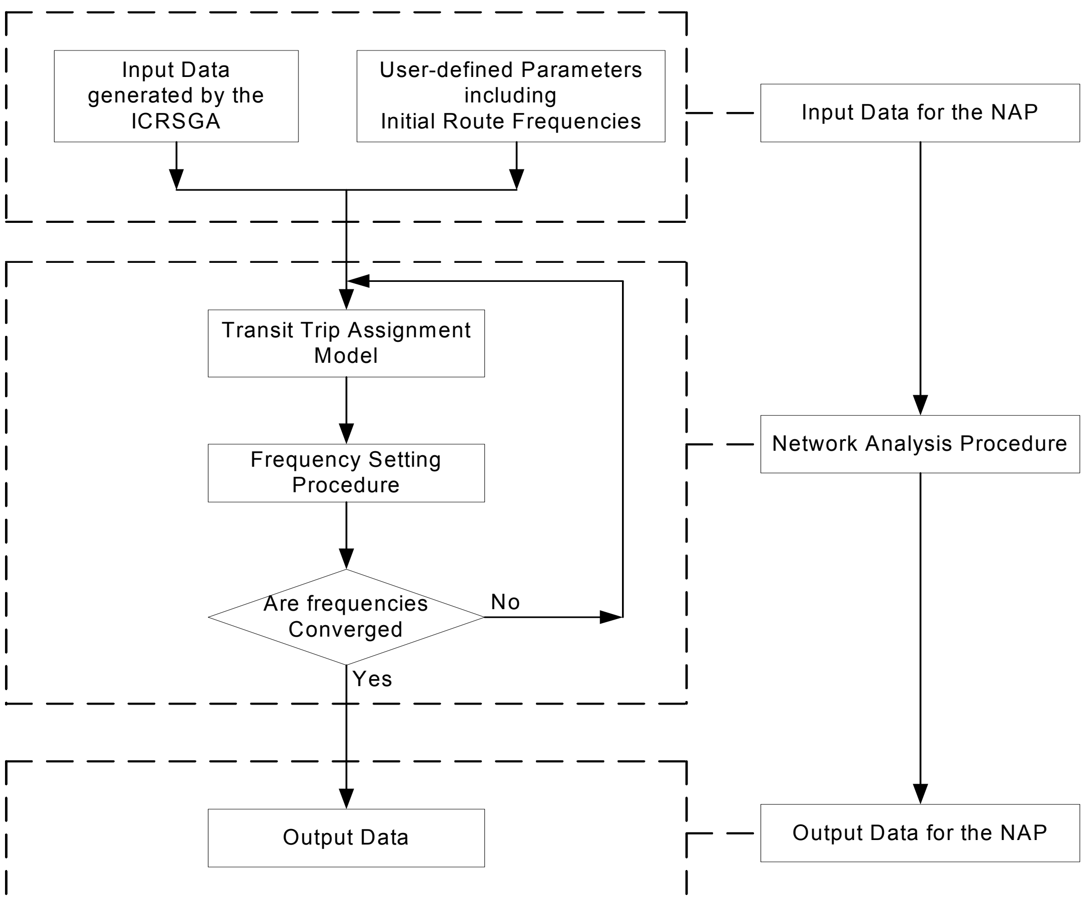


Figure 6.2 Flow Chart for the Network Analysis Procedure

6.3 Assumptions of the NAP

The Network Analysis Procedure is implemented under the following assumptions:

1. All components relevant to the passenger travel time from the origins to the destinations are not subject to congestion effects;
2. There is an industry-wide standard for a service area of one-quarter mile to either side of each route and a half mile is assumed to be the maximum walking distance that a transit user can tolerate.

6.4 Transit Trip Assignment Model for the NAP

As mentioned before, the NAP for the BTRNDP with fixed demand consists of two major components: the transit trip assignment model and the frequency setting procedure. This section focuses on the transit trip assignment model, which is the core of the whole Network Analysis Procedure.

6.4.1 Overview of the Transit Trip Assignment Model

Generally speaking, as an essential part of the BTRNDP, the transit trip assignment process refers to the assignment of the passenger trip demand matrix to a given set of routes associated with a particular network configuration. As indicated by Shih and Mahmassani (1996), the significance of the transit trip assignment for the BTRNDP can be demonstrated by two aspects: (a) the number of vehicles and the frequencies run on each route is largely dependent on the number of trips assigned to the transit network routes; and (b) the evaluation of various cost and performance measure for the transit system needs accurate passenger flow information on each node and each link along the routes. A better transit trip assignment model could undoubtedly help analyze and evaluate the existing transit network more accurately, redesign a transit route network more efficiently and possibly yield a better solution to the transit route network design problem.

Similarly as categorized by Speiss and Florian (1989), the transit assignment problem has been investigated by several researchers during past decades, either as a separate problem (see Dial, 1967 and Rapp etc, 1976) or as a subproblem of more complex models, such as transit network design (see Lampkin and Saalmans, 1967; Hasselstrom, 1981; Baaj and Mahmassani, 1990 and Shih and Mahmassani 1996), or multimodal network equilibrium problem (see Florian and Spiess, 1983).

Due to the phenomenon of passengers' waiting for the arrival of a transit vehicle at the transit stops, the transit trip assignment problem differs from the auto trip assignment problem. For transit users to get to the intended destination, the different availability of the buses on different routes at different time periods, along with possibly different total travel time, may greatly affect the decision of the transit users. Therefore, passenger route choices in a transit network, especially in large urban areas with highly overlapping routes, are much more complex.

Having recognized these characteristics, many researchers have proposed several transit trip assignment models, such as single path assignment and multiple path assignment models as a result of the modifications to the auto traffic assignment process. Typical efforts include the following. Lampkin and Saalmans (1967) assigned passenger transit trips to several competing paths according to a "frequency-share" rule, where the faction of total transit passengers on a specific path is proportional to the probability that vehicles serving that path arrive earlier than other competing paths. As indicated by many researchers, the "frequency-share" rule is the most commonly used rule for route assignment at uncoordinated operations terminals in the node aggregation level. In this rule, it is assumed that transit passengers will always board the first arriving vehicle of any competing route. As a result, this rule stipulates that a route carries a proportion of the flow equal to the ratio of its frequency to the sum of the frequencies of all competing paths. For example, suppose that d_{ij} is the demand from origin i to destination j. If there are three acceptable competing routes $r_{1,} r_{2,}$ and r_3 whose frequencies are $f_{1,} f_{2,}$ and f_3 respectively, then, the demand assigned on all the links of the route r_1 between node i and j is determined as follows by the rule: $(d_{ij})_{r_1} = f_1 \Big/ (\sum_{i=1}^{3} f_i) = [f_1/(f_1+f_2+f_3)] * d_{ij}$, the demand that route r_2 carries is $\{[f_2/(f_1+f_2+f_3)] * d_{ij}\}$, and the demand that route r_3 carries is $\{[f_3/(f_1+f_2+f_3)] * d_{ij}\}$. Furthermore, if the stochastic characteristic of the waiting time is ignored, the average passenger waiting time (in minutes) incurred in this case is as follows: $w_{ij} = 60.0 \Big/ (2 * \sum_{i=1}^{3} f_i) = 30.0/(f_1+f_2+f_3)$ (minutes).

Dial (1967) discussed the necessity and methods for inclusions of transfer penalty in the computation costs and a minimum weighted time path assignment was proposed in his paper. Spiess and Florian (1989) presented an optimal strategy and solved it using a label-correcting algorithm to allow the transit patrons to reach their destinations at minimum generalized cost. For the first time to consider trip assignment to the overlapping routes, Han and Wilson (1982) proposed a lexicographic strategy where transfer avoidance and/or minimization were considered as the primary criterion for transit user route choice. The main feature in the context of this model is that the number of transfers is considered as the most important criterion and preemptive priority is given to this consideration. It starts by searching all the transit paths between node i and j with no transfers. Only if none of these types of paths is found will paths with only one transfer be considered. Different from the procedure introduced before, this research is intended to account for more details involved in the trip assignment model. For example, for any given demand node pair, if more than one path has the same number of (or no) transfers, an additional rule is introduced for checking the travel cost on each competing route. Only the paths whose travel cost is within a particular range are recorded. Trips are then assigned into the network using an analytical allocation model that reflects the relative level of service on these competing paths. Note that the lexicographic decision structure of Han and Wilson (1982) is adapted in this research due to its behavior realism and plausibility. The details of this procedure in this research are described in the next section.

Horowitz (1987) pointed out two obvious facts that always affect multipath trip assignment for the transit network. The first one is that passengers dislike transferring and will always try to avoid as many transfers as possible and the second is that transit users also dislike long walks at either the origin or the destination of their trips. Therefore, alternative paths with particularly long walks and/or excessive (two or more) transfers through the networks will never be considered in normal conditions. If any of these cases happens, transit users might resort to a better path, look for a different destination, choose a different mode, or possibly completely forego their trips. Furthermore, it is noted that transit users usually only have a reasonable choice of alternative paths. A multipath transit trip assignment algorithm must first determine the acceptable choice of alternative paths and must next split passengers among a small choice set of these reasonable paths based on their relative merits. Note that the multipath trip assignment model is implemented given a specific route set network and their associated service frequencies on each route.

Baaj and Mahmassani (1990) adopted Han and Wilson's lexicographic strategy and the "frequency-share" rule from Lampkin and Saalmans in their TRUST (Transit Route Analyst). Meanwhile with some modifications, they proposed a filtering process to apply a threshold check to eliminate the undesirable paths with a trip time exceeding the shortest trip time among all candidate paths by a specified percentage. Mainly based on the work of Baaj and Mahmassani (1990), Shih and Mahmassani (1996) extended the TRUST for timed-transfer transit system with some considerations on the coordinated operations terminals, while the lexicographic strategy and the "frequency-share" rule were still employed.

These previously developed models provide a solid basis for further research on the transit trip assignment problem. However, there are three major shortcomings that exist in previous models: 1) These models are proposed for the transit network under the assumption that each zonal transit demand is aggregated as a single node, ignoring the fact that multiple routes can pass by the same zone connected to different distribution nodes on different links. As a result, some commonly used rules that were proposed in the context of this assumption, such as "frequency-share" rule, may not work properly if these assumptions are violated; 2) These models did not consider the phenomenon of the long-walk-involved paths where transit users might walk a tolerable distance to their neighboring zones to take the bus. This situation might commonly exist when there is no direct service provided for transit users at a specific zone but provide such service at their neighbors; and 3) These models are uniformly limited to transit networks with fixed transit demands. However, it is generally accepted that a variable relationship exists between the transit demand and the transit network route configuration where they are actually dependent upon each other.

These shortcomings of the existing transit trip assignment models greatly preclude their practical applications (such as to the bus transit route network design problem). It is expected that if considered in the context of a transit network with variable transit demands and at the distribution node level, trip assignment becomes much more complicated. For example, the "frequency-share" rule might work well in the context of the zone aggregation level. However, the widely used "frequency-share" rule could be problematic in the real world situation, where the transfer characteristics between related routes need to be considered in the context of the "centroid" node level since the routes that pass by a centroid node are not necessarily on the same link on the same road. In this case, a multiple path assignment model that can explicitly consider the transfer characteristics between related routes at the "centroid" node level should be considered. To improve its behavior realism and model practicability, one needs to consider the transit trip assignment in the context of the transit network with variable transit demands and at the "centroid level". The related details of the BTRNDP with variable demand at "centroid" node level are presented in chapter 7. In this section, a more general and practical trip assignment model for the BTRNDP with fixed demand at "centroid" node level is presented.

The following section proposes the details of an innovative transit trip assignment procedure in the NAP that can accommodate this necessity.

6.4.2 Transit Trip Assignment Model

The trip assignment models presented here build on the lexicographic strategy (see Han and Wilson, 1982), Horowitz (1987) and several further research works (such as Baaj, 1990, and Shih and Mahmassani, 1996). However, remarkable modifications have been made to accommodate more complex considerations for real world applications.

It is assumed that transit users usually consider the following criteria in their transit route choice: the number of transfers required to reach the destination, the number of long walks needed and how long the long walks take, and the trip times for different choices. Furthermore, it is assumed that trip-makers always attempt to choose the path that has the lowest number of transfers and/or least number of long walks to get to the destination. If more than one path has the same number of transfers, and/or long walks, then a decision is made based on the total travel time for those competing paths whose trip times do not exceed the minimum travel time by a specified threshold. Based on these criteria, the proposed hybrid transit trip assignment models in this paper can be seen from an "hierarchical" point of view, which consists of four levels that can be described as follows. The first level is 0-transfer-0-longwalk paths (i.e., routes than can provide direct service); the second level is 0-transfer-1-longwalk path and/or 1-transfer-0-longwalk paths; the third level is 2-transfer-0-long-walk paths, 1-transfer-1-long-walk and/or 0-transfer-2-longwalk paths and the fourth level is the "no service available".

Figure 6.3 shows the flow chart of the proposed transit trip assignment model that is designated for the NAP with fixed transit demand. Details of the transit trip assignment models for the BTRNDP with variable transit demands are discussed in Chapter 8.

It should be noted that the trip assignment process considers each zone (centroid node) pair separately. For a specific centroid node pair, as shown in Figure 6.3, the trip assignment model first checks the existence of the 0-transfer-0-longwalk paths. If any path of this category is found, then the transit demand between this centroid node pair can be provided with direct route service and the demand is therefore distributed to these routes. If not, the existence of paths of the second category, i.e., 0-transfer-1-longwalk path and 1-transfer-0-longwalk paths are checked. If none of these paths is found, the proposed procedure will continue to search for paths of the third category, i.e., paths with 2-transfer-0-long-walk, 1-transfer-1-long-walk and/or 0-transfer-2-longwalk. Only if no paths that belong to these three categories exist, there would be no paths in the current transit route system that can provide service for this specific centroid node pair (i.e., these demands are unsatisfied). Note that at any level of the above three steps, if more than one path exists, a "travel time filter" is introduced for checking the travel time on the set of competing paths obtained at that level. If one or more alternative paths whose travel time is within a particular range pass the screen process, an analytical nonlinear model that reflects the relative utility on these competing paths proposed at that level is used to assign the transit trips between that centoid node pair to the network. The whole process is repeated until all the travel demand pairs in the studied network are traversed. Details of this trip assignment procedure are presented in the following subsections.

To facilitate the description, four notations are introduced for a given centroid node pair (i, j).

- RS_i denotes the set of routes that pass by centroid node *i*
- NS_i denotes the set of distribution nodes on the link for centroid node *i*
- NB_i denotes the set of distribution nodes of the zones that are neighbors to centroid node *i* and to which a long walk distance is within a pre-specified range that a transit user can tolerate
- $\mathrm{SNR}_k, k \in \mathrm{NB}_i$ denotes the set of routes that pass by *k*, one of the neighbor distribution nodes of centroid node *i*

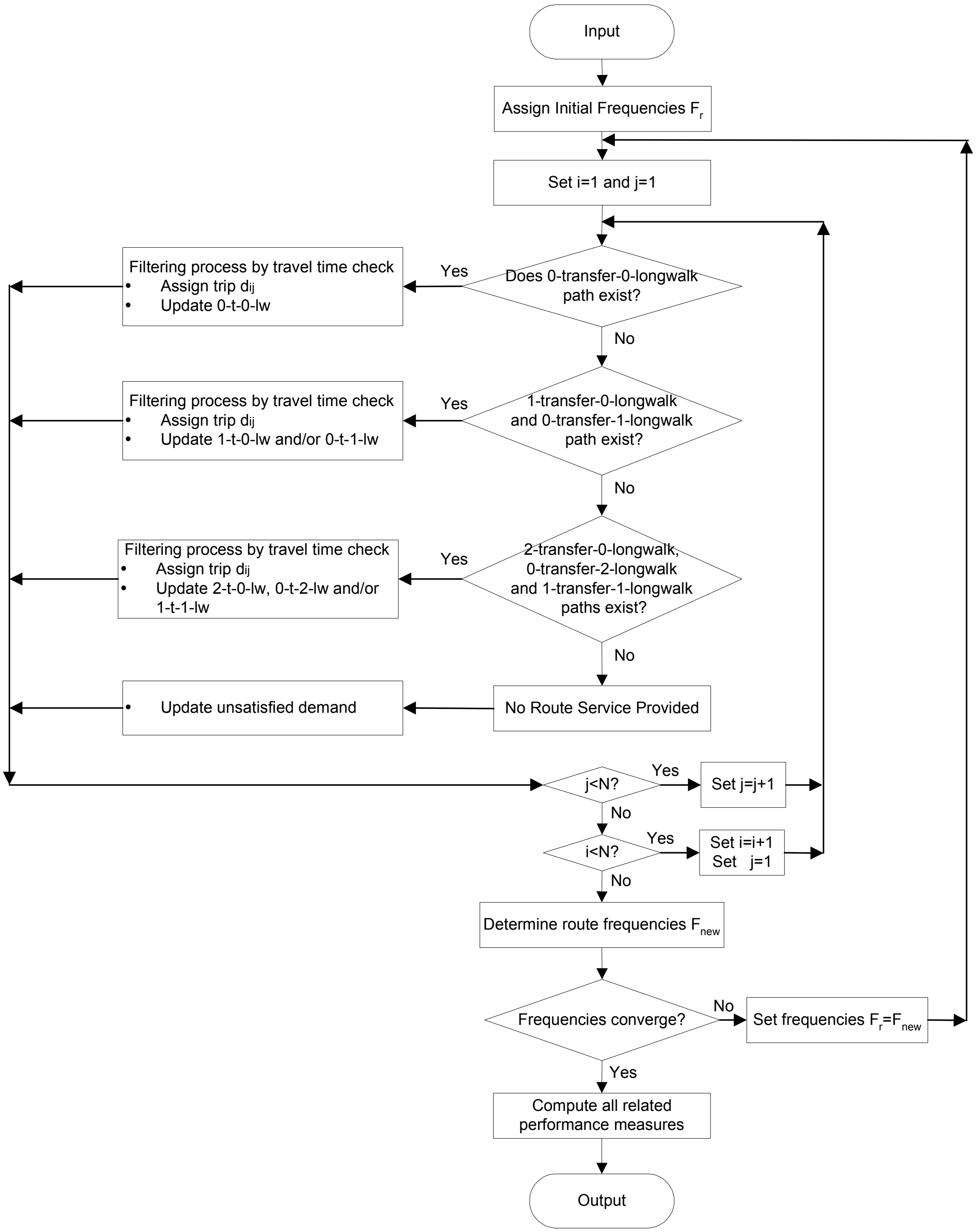


Figure 6.3 The Network Analysis Procedure (NAP) for the BTRNDP

6.4.2.1 First Level (0-transfer-0-long-walk paths)

For a given centroid node pair (i, j), the trip assignment model checks the intersection of two sets of routes RS_i and RS_j, which pass by the origin i and the destination j respectively. If $\mathrm{RS}_i \cap \mathrm{RS}_j \neq \Phi$, then there exist some routes that have both node i and j on their node lists, which means that these routes can provide direct service for those passengers who want to travel from centroid node i to j. The routes in this set are classified as 0-transfer-0-long-walk paths for node pair (i, j). Let R_{kij} denote one such path where passengers board route r_k at zone i (the origin) and travel on this route to zone j (the destination). The demand is distributed among these route sets and the parameter, "0-transfer-0-long-walk paths", is then updated.

If more than one such path exists, then an additional rule is introduced for checking the travel cost (total travel time) on each competing route. First of all, the minimum travel cost is found by comparing the total travel time on all routes. Any route whose travel cost exceeds the minimum value by a specified threshold (say 30 percent) is rejected. Demand is then assigned to the acceptable routes using an analytical nonlinear allocation rule that reflects the relative utility on these competing paths rather than the widely used "frequency share" rule (where the travel demand zone is aggregated as a single node and the flow distributed to a route is proportional to its frequency).

Let TT_{kij} denote the total travel time required for the transit users who want to travel from origin zone i to destination zone j by taking bus transit on route r_k and ik represent the distribution node (i.e., the access point) on the road link of centroid node i. Then TT_{kij} can be computed as follows:

$$TT_{kij} = t_{i,ik}^{access} + t_{ik,k}^{wait} + t_{ik,k,kj}^{invt} + t_{kj,j}^{access}$$

where:

$t_{i,ik}^{access}$ ----- the access (walking) time from centroid node i to the distribution node ik on the link where route r_k locates;

$t_{ik,k}^{wait}$ ----- the waiting time experienced by passengers at node ik on the link for transit run on route r_k;

$t_{ik,k,kj}^{invt}$ ----- the in-vehicle travel time required from distribution node ik on the link to the distribution node kj of destination j on the route r_k;

$t_{kj,j}^{access}$ ----- the access (walking) time from the distribution node kj on the link to the centroid node of destination j;

Suppose that there are several paths such as R_{kij} that satisfy $r_k \in \mathrm{RS}_i \cap \mathrm{RS}_j$. Then the analytical allocation model divides travelers proportional to travel time as follows:

$$(d_{ij})_{r_k} = \frac{1/TT_{kij}}{\sum_{r_m \in \mathrm{SR}_i \cap \mathrm{SR}_j} (1/TT_{mij})} * d_{ij}$$

When the headway on a specific route is less than 10 minutes, the average passenger waiting time can be estimated using half headway model. However, as headway increases, passengers might tend to plan their arrivals to the transit stops. In this case, a more accurate estimation of average passenger waiting times can be obtained using the models developed by Fan and Machemehl (2002 and 2003).

However, if the intersection of SR_i and SR_j is an empty set (i.e. $\mathrm{SR}_i \cap \mathrm{SR}_j = \Phi$) or if at least one of them is empty, then there is no direct route connecting these two nodes *i* and *j*. Since no direct service can be provided, those passengers who want to travel from zone *i* to zone *j* or vice versa may have to make transfers or take long walks to get served by some alternative routes. In this case, the next level of transfer paths (1-transfer-0-long-walk paths) and/or long-walk paths (0- transfer-1-long-walk paths) are then checked. Note that the availability of these two paths is checked in the same priority since it is assumed that the transit tripmaker always attempts to complete his/her trip making as few transfers as possible and as little walking distance as possible. Namely, if no direct routes that connect zone *i* to zone *j* exist, the trip assignment model checks for both 1- transfer paths and 1-long-walk paths at the same time for possible services.

6.4.2.2 Second Level

As mentioned before, if the demand between a specific centroid node pair cannot be satisfied directly at the first level, then paths (i.e., 0-transfer-1-longwalk paths and/or 1-transfer-0-longwalk paths) at the second level are checked. This level is designed to check whether the trip can be completed with one transfer or one long walk and the details are presented as follows. For convenience, the following notations are introduced:

- NLR_{it} denotes the node lists in $r_t \in \mathrm{RS}_i$ (i.e. r_t is one of the routes that pass by centroid node *i*).
- $\mathrm{NLR}_{jt'}$ denotes the node lists in $r_{t'} \in \mathrm{RS}_j$ (i.e. $r_{t'}$ is one of the routes that pass by centroid node *j*).

6.4.2.2.1 1-transfer-0-long-walk paths

By examining the node lists of every possible combination of r_t and $r_{t'}$, the trip assignment model seeks to find the intersection set of nodes contained in both NLR_{it} and

$\mathrm{NLR}_{jt'}$. If the intersection set is not empty, then its contents are possible transfer nodes between route r_t and $r_{t'}$. For example, if the intersection set is $\{\mathrm{n}_1, \mathrm{n}_2, \cdots, \mathrm{n}_\mathrm{k}\}$, then there are k 1-transfer-0-longwalk paths. Let $\mathrm{TR}_{it(n_1)t'j}$ denote a path from zone *i* to zone *j*, where passengers board route r_t at origin node *i*, and stay on route r_t until node n_1, where passengers transfer to route $r_{t'}$ and travel on it until the destination node *j* is reached. Assume that the average waiting time at the transfer station is half of the receiving route headway and assume that the transfer penalty is specified by the users. For each possible path involving one transfer, the total travel time can be estimated as follows:

$$
\begin{aligned}
TT_{it(n_1)t'j} &= t_{i,it}^{access} + t_{it,t}^{wait} + t_{it,t,n_1}^{invt} + t_{n_1,t'}^{wait} + t_{transfer\_penalty} + t_{n_1,t',jt'}^{invt} + t_{jt',j}^{access} \\
&= t_{i,it}^{access} + \frac{60}{2*f_t} + t_{it,t,n_1}^{invt} + \frac{60}{2*f_{t'}} + t_{transfer\_penalty} + t_{n_1,t',jt'}^{invt} + t_{jt',j}^{access}
\end{aligned}
$$

After the total travel time associated with this type of path is computed, note that different from the approach taken in 0-transfer-0-longwalk route service, the filtering process and corresponding analytical allocation model do not come into play until all the possible 1-transfer-0-long-walk paths and 0-transfer-1-long-walk paths have been obtained. However, these 1-transfer-0-long-walk paths are kept until the 0-transfer-1-long-walk paths are searched.

6.4.2.2.2 0-transfer-1-long-walk paths

This section is focused on the procedure to check whether the trip can be completed with one long walk and no transfers. Suppose the procedure has found that origin centroid node *i* cannot be directly reached by any route in the existing transit system. The trip assignment model first identifies the set of zones that are neighbors to centroid node *i* and to which a long walk distance would be within a pre-specified range tolerable by a transit user and records such neighboring nodes in NB_i. Also recorded is $\mathrm{SNR}_k, k \in \mathrm{NB}_i$, the set of routes that pass through *k*, any of the neighbor distribution nodes of centroid node *i*. The model then checks the intersection of these two sets of routes. If $\mathrm{SNR}_{k\in \mathrm{NB}_i} \cap \mathrm{RS}_j \neq \Phi$, then there exist some routes that have both node *k* (neighbors to centroid *i*) and *j* on their node lists, which means that these routes can provide route service (with 0-transfer-1-long-walk involved) for those passengers who want to travel from centroid node *i* to *j*. The routes in this set are classified as 0-transfer-1-long-walk paths for centroid node pair (*i*, *j*). Let $\mathrm{R}_{ki^N j}$ denote such a path that represents passengers boarding route r_k at origin node i^N (one of the neighbor distribution nodes of centroid node *i*) and traveling on this route to destination *j*. Suppose that i^o is the chosen node on the link that transit users want to walk through because it provides the shortest path to i^N compared to other nodes $i^{'}$ ($i^{'} \in \mathrm{NS}_i$) on the link.

Let TT_k denote the travel time required for transit users who want to travel from centroid node *i* to centroid node *j* by taking a bus on route r_k, a 1-longwalk-0-transfer path. Then it can be computed as follows:

$$TT_{ki^N j} = t_{i,i^o}^{access} + t_{i^o,i^N}^{access} + t_{i^N,k}^{wait} + t_{i^N,k,kj}^{invt} + t_{kj,j}^{access}$$

where:

t_{i,i^o}^{access} ----- the access (walking) time from centroid node *i* to its associated distribution node (i.e., access point) i^o on the link;

t_{i^o,i^N}^{access} ----- the access (walking) time from node i^o that belongs to centroid node *i* to i^N, the distribution node neighbored to centroid node i on the link;

$t_{i^N,k}^{wait}$ ----- the waiting time experienced by passengers at distribution node i^N on the link for buses on route r_k;

$t_{i^N,k,kj}^{invt}$ ----- the in-vehicle travel time required from node i^N on the link to the distribution node *kj* of centroid node *j* on the link on the route r_k;

$t_{kj,j}^{access}$ ----- the access (walking) time from the distribution node *kj* on the link to the destination centroid node *j*;

As mentioned, after the total travel time associated with all the possible 1-transfer-0-long-walk paths and/or 0-transfer-1-long-walk paths are obtained, a filtering process that is similar to the rule at first level is used if more than one such path exists. Only the paths whose total travel time is within a particular range are recorded as acceptable paths. The trip assignment procedure assigns the demand d_{ij} according to an analytical allocation model as follows.

For 1-transfer-0-long-walk paths, suppose several paths such as n_k satisfy $n_k \in \mathrm{NLR}_{it} \cap \mathrm{NLR}_{jt^{'}}$. For 0-transfer-1-long-walk paths, suppose that several paths such as r_k satisfy $r_k \in \mathrm{SNR}_{t \in \mathrm{NB}_i} \cap \mathrm{RS}_j$. Then for 1-transfer-0-long-walk paths, the analytical allocation model divides travelers proportional to travel time as follows:

$$(d_{ij})_{r_k} = \frac{1/t_{in_k j}}{\sum\limits_{n_{kk} \in \mathrm{NLR}_{it} \cap \mathrm{NLR}_{jt^{'}}} (1/t_{in_k j}) + \sum\limits_{kk \in \mathrm{SNR}_{t \in \mathrm{NB}_i} \cap \mathrm{RS}_j} (1/TC_{kk})} * d_{ij}$$

While for 0-transfer-1-long-walk paths, the analytical allocation model divides travelers proportional to travel time as follows:

$$(d_{ij})_{r_k} = \frac{1/TC_k}{\sum\limits_{n_{kk}\in \mathrm{NLR}_{it}\cap \mathrm{NLR}_{jt'}} (1/t_{in_k j}) + \sum\limits_{kk\in \mathrm{SNR}_{t\in \mathrm{NB}_i}\cap \mathrm{RS}_j} (1/TC_{kk})} * d_{ij}$$

After the demand is distributed among these route sets, the parameter, "1-transfer-0-long-walk paths" and/or "0-transfer-1-long-walk paths" is therefore updated. However, if none of the paths at this level can be found, the next level of transfer paths and/or long-walk paths are checked and details are presented in the next subsections.

6.4.2.3 Third Level

If none of the 0-transfer-1-longwalk paths and/or 1-transfer-0-longwalk paths can be found, then passengers who want to travel from node i to j might have to resort to paths with two transfers, two long walks, and/or one transfer and one long walk.

6.4.2.3.1 2- transfer-0-long-walk paths

This procedure here is intended to search for all paths with exactly two transfers between node *i* and *j* and try to find a route that passes through neither node *i* or node *j,* but shares a node with a route *t* that passes through node *i* ($r_t \in \mathrm{RS}_i$) and shares another through node *j* with route $r_{t'}$, ($r_{t'} \in \mathrm{RS}_j$). As far as computation is concerned, it can be noted that RS, the set of these routes, is simply the complement of the union of the previous generated SR_i and SR_j (i.e., $\mathrm{SR} = \overline{\mathrm{SR}_i \cup \mathrm{SR}_j}$, see Baaj and Mahmassani, 1990).

Let $(n_k, n_{k'}) \in SET2TRANSFER$ denote one pair of the transfer nodes that are suitable for 2-transfer paths and $\mathrm{TR}_{i,ik_1,k_1,n_k,k_2,n_{k'},k_3,jk_3,j}$ denotes a path from node *i* and *j*, which passengers board route k_1 at origin node ik_1 on the link that belongs to centroid node *i*, and stay on the route k_1 until node n_k, where passengers makes his/her first transfer to route k_2 and travel on it until node $n_{k'}$, where he/she makes second transfers to route k_3 and travel on it until node jk_3, and then walks to the destination centroid node *j*. Again, assume that the average waiting time for the transfer bus at the transfer station is half of the headway run on the transfer route. For this type of path that involves two transfers, the total travel time can be estimated as follows:

$$t_{i,n_k,n_{k'},j} = t^{access}_{i,ik_1} + t^{wait}_{ik_1,k_1} + t^{invt}_{ik_1,k_1,n_k} + t^{wait}_{n_k,k_2} + t_{transfer\_penalty\_1} + t^{invt}_{n_k,k_2,n_{k'}} + t^{wait}_{n_{k'},k_3} + t_{transfer\_penalty\_2}$$
$$+ t^{invt}_{n_{k'},k_3,jk_3} + t^{access}_{jk_3,j}$$
$$= t^{access}_{i,ik_1} + t^{wait}_{ik_1,k_1} + t^{invt}_{ik_1,k_1,n_k} + \frac{60}{2*f_{k_2}} + t_{transfer\_penalty\_1} + t^{invt}_{n_k,k_2,n_{k'}} + \frac{60}{2*f_{k_3}} + t_{transfer\_penalty\_2}$$
$$+ t^{invt}_{n_{k'},k_3,jk_3} + t^{access}_{jk_3,j}$$

6.4.2.3.2 0-transfer-2-long-walk paths

The procedure considers the situations that both centroid nodes in the demand pair d_{ij} cannot be reachable by any route in the existing transit system. Using the same principles as before, this section is focused on the procedure to check whether the trip can be completed with two long walks and no transfers.

The trip assignment model first checks whether the set of centroid nodes that are neighbors to centroid node *i* and are within a long walk distance and records such neighboring nodes in NB_i. Also recorded is $\mathrm{SNR}_k, k \in \mathrm{NB}_i$, the set of routes that pass through *k*, any neighbor centroid of node *i*. The same process is also applied for centroid node *j*. The model then checks the intersection of these two sets of routes. If $\mathrm{SNR}_{k\in \mathrm{NB}_i} \cap \mathrm{SNR}_{kk\in \mathrm{NB}_j} \neq \Phi$, then there exist some routes that have both node *k* (that are neighbored to centroid *i*) and node *kk* (that are neighbored to centroid *j*) on their node lists, which means that these routes can provide route service for those passengers who want to travel from centroid node *i* to *j* or from centroid node *j* to *i* (with 0-transfer-2-long-walk). The routes in this set are classified as 0-transfer-2-long-walk paths for node pair (*i*, *j*). Let $\mathrm{R}_{ki^N j}$ denote a path that represents passengers boarding route r_k at origin node i^N (the neighbor centroid node of *i*) and traveling on this route to destination node j^N (the neighbor centroid node of *j*.) Suppose that i^o is the chosen node on the link that the transit users want to walk through because it has the shortest path to i^N compared to other nodes i' ($i' \in \mathrm{NS}_i$) on the link. (So is j^o).

Let TC_k denote the travel time required for the transit users who want to travel from centroid node *i* to centroid node *j* by taking bus transit on route r_k. Then it can be computed as follows:

$$TC_k = t^{access}_{i,i^o} + t^{access}_{i^o,i^N} + t^{wait}_{i^N,k} + t^{invt}_{i^N,k,kj} + t^{access}_{kj,j^N} + t^{access}_{j^N,j}$$

where:

t^{access}_{i,i^o} ----- the access (walking) time from centroid node *i* to its associated node i^o on the link;

$t^{access}_{i^o,i^N}$ ----- the access (walking) time from node i^o that belongs to centroid *i* to the node i^N on the link that belongs to neighboring centroid of *i*;

$t^{wait}_{i^N,k}$ ----- the waiting time experienced by passengers at node i^N on the link for transit run on route r_k ;

$t^{invt}_{i^N,k,kj}$ ----- the in-vehicle travel time required from node i^N on the link to the node *kj* on the link on the route r_k ;

t^{access}_{kj,j^N} ----- the access (walking) time from the node *kj* on the link to the node j^N that belongs to the destination centroid node *j*;

$t^{access}_{j^N,j}$ ----- the access (walking) time from the node j^N on the link to its associated destination centroid node *j*.

6.4.2.3.3 1- transfer-1-long-walk paths

This procedure here is designed to check whether the trip can be completed with 1-transfer-1-long-walk paths. Note that the long walk can happen at either the origin or destination or while transferring. For convenience, it is assumed that centroid node *i* is the node that cannot be reachable by any route in the existing transit system (i.e., the long walk is at the origin centroid node). The trip assignment model first checks whether the set of centroid nodes that are neighbors to centroid node *i* and to which a long walk distance are within a pre-specified range and records such neighboring nodes in NB_i. Also recorded is $\text{SNR}_k, k \in \text{NB}_i$, the set of routes that pass through *k*, a neighbor centroid of node *i*. For facilitation of the description, the following notations are introduced:

- $\text{NLR}_{i^N t}$ denotes the node lists in $r_t \in \text{SNR}_k, k \in \text{NB}_i$ (i.e. r_t is one route that passes through node i^N on the link that belongs to the neighboring centroid node of *i*).
- $\text{NLR}_{jt'}$ denotes the node lists in $r_{t'} \in \text{SR}_j$ (i.e. $r_{t'}$ is one route that passes through node *j*)

By examining the node lists of every possible combination of r_t and $r_{t'}$, the trip assignment model seeks to find the intersection set of nodes contained in both $\text{NLR}_{i^N t}$ and $\text{NLR}_{jt'}$. If the intersection set is not empty, then its contents are possible transfer nodes between route r_t and $r_{t'}$. For example, if the intersection set is $(\text{n}_1, \text{n}_2, \cdots, \text{n}_\text{k})$, then there are k 1-transfer paths. Let $\text{TR}_{i^N t(n_1) t' j}$ denote a path from node i^N to *j*, and passengers board route *t* at origin node i^N, and stay on route *t* until node n_1, where they transfer to route $t^{'}$ and travel on it until destination node *j* is reached. Suppose that i^{o} is the chosen node on the link that the transit users want to walk through because it has the shortest path to i^{N} compared to other nodes $i^{'}$ ($i^{'} \in \mathrm{NS}_{i}$) on the link. Assume that the average waiting time for the transfer bus at the transfer station is half of the transfer route headway. For each possible path involving one transfer and one long walk, the total travel time can be estimated as follows:

$$
\begin{aligned}
t_{in_1 j} &= t_{i,i^o}^{access} + t_{i^o,i^N}^{access} + t_{i^N,t}^{wait} + t_{i^N,t,n_1}^{invt} + t_{n_1,t^{'}}^{wait} + t_{transfer\_penalty} + t_{n_1,t^{'},jt^{'}}^{invt} + t_{jt^{'},j}^{access} \\
&= t_{i,i^o}^{access} + t_{i^o,i^N}^{access} + \frac{60}{2 * f_t} + t_{i^N,t,n_1}^{invt} + \frac{60}{2 * f_{t^{'}}} + t_{transfer\_penalty} + t_{n_1,t^{'},jt^{'}}^{invt} + t_{jt^{'},j}^{access}
\end{aligned}
$$

After obtaining all the possible 2-transfer-0-long-walk paths, 1-transfer-1-long-walk paths, and/or 0-transfer-2-long-walk paths, the transit trip assignment procedure assigns the demand d_{ij} to these competing paths. As described before, if more than one such path exists, then a "travel time filter" is introduced for checking the travel cost on competing routes. Any route whose travel cost exceeds the minimum value by a specified threshold (say 30 percent) is rejected. Demand is then assigned to the acceptable routes (possibly to paths of these three categories) using an analytical nonlinear proportional allocation rule, which is the same as that in Level 1 and 2.

6.4.2.4 Fourth Level (no service available)

As shown in Figure 6.3, for a given centroid node pair (*i*, *j*), if none of the paths of the first three levels can be found, then there is no route service available for this specific centroid node pair. In this case, demands from zone *i* to zone *j* (or vice versa) cannot be satisfied by transit in the current transit network system. In this case, the parameters regarding unsatisfied demand are updated and the model will increment *j* and repeat all these processes until $\mathrm{j} \geq \mathrm{N}$. Note that all the above processes are repeated until all the transit demand pairs shown in the trip table matrix are checked.

6.5 Frequency Setting Procedure

As the core of the Network Analysis Procedure, the previous section has presented the transit trip assignment model. To determine the service frequency for each route and further to evaluate the objective function including the user cost and operator cost, it is necessary to include a Frequency Setting Procedure as part of the NAP. This section centers on the iterative frequency setting procedure that yields internally consistent service frequencies when it is coupled with the transit trip assignment model.

6.5.1 Demand Frequency and Policy Frequency

As mentioned in the headway feasibility constraints in Chapter 3, the frequencies commonly used in the transit industry can be grouped into three categories: supply frequency, policy frequency, and demand frequency. Demand frequency and policy frequency are chosen in the NAP to set the route service frequencies.

As described, the demand frequency can be computed as follows. The route-based transit demands are obtained by assigning the trips in a transit demand matrix using the trip assignment model. Based on these transit demands, the total transit trips using route r_k and its corresponding Q_k^{max}, the maximum link flow of route r_k are computed. As a result, assuming that transit demand is symmetric, f_k, the route frequency for route r_k can always be computed as follows:

$$f_k = \frac{Q_k^{max}}{PL_k}$$

where:

f_k = minimum route frequency for route r_k (hrs);
Q_k^{max} = maximum hourly link flow occurring on the route r_k (people/hour);
P = the bus seating capacity for passengers operating on the network's route;
L_k = the user-defined maximum load factor allowed on route r_k;

Obviously, different values for maximum load factors can meet different transit operator operational considerations. Furthermore, different load factors may be chosen for different subset of bus routes. However, for simplicity, one load factor is assumed for all routes in the transit network although its value can be changed to perform sensitivity analyses. Furthermore, as described, the policy frequency is used for setting the bus line headway in extreme situations. The following section presents an iterative process that is employed in the FSP to compute the demand frequency for all routes.

6.5.2 Solution Approach for the FSP

A commonly used rule in the transit industry, especially for congested networks, is the computation of route frequencies in order to achieve a preset peak load factor. To make this rule meaningful, the demand assignment must be performed over the peak hour period. However, one would expect that the NAP can be used for different time-of-day periods. For less congested periods, the method using peak load factor may yield frequencies that are much higher than what the riders expect as reasonable. In this case, the minimum policy headway would come into play.

As described before, service frequencies are computed with a preset load factor and an iterative procedure. First, an initial set of frequencies is predefined as inputs for the transit trip assignment model. (For example, the NAP simply assumes identical initial frequencies of 6 buses/hour for all routes associated with a particular network configuration.) Then, based on these frequencies, the NAP calculates new service frequencies on all routes by implementing the trip assignment model. Demands must be reassigned in order to be consistent with the current service frequencies and new route frequencies are recomputed again based on updated transit demand characteristics on all routes. For any specific iteration in the frequency setting procedure, if the output frequencies are considered to be quite different from the input frequencies (say, they differ by more than 10%), then the NAP continues until convergences are achieved between the revised frequencies and the input frequencies. In other words, the NAP develops the internal consistency of route frequencies through an iterative process, which has been shown in Figure 6.2.

6.5.3 Preset Parameters for the FSP

As described before in this chapter, some parameters must be chosen by the users when using the FSP in the NAP. In the NAP, the penalty for each transfer is chosen to be 5 minutes equivalent in-vehicle travel time, the bus seating capacity is selected as 40 passengers, and the maximum load factor on each route is taken as 1.3. Other design-related parameters include different weights that the transit operators might put on different components of the transit user time. However, the traffic delay at each intersection and the mean passenger boarding and deboarding times are not considered. Table 6.1 presents all the NAP input Parameters.

Parameter	Notation	Value
$C_{transfer\_penalty}$	the penalty for each transfer	5 minutes
P	bus seating capacity	40 passengers/bus
L_k	maximum load factor on each route	1.3

Table 6.1 Input Parameters for the NAP

It is expected that different chosen preset network-related parameters can result in different solution quality. To see how the quality of the optimal bus transit network solution changes corresponding to different chosen values for these parameters, sensitivity analysis that can reflect changes in the performance of objective functions can be conducted and related characteristics can be identified.

6.6 Network Example Illustrations for the Trip Assignment Procedure

To show how the nonlinear transit trip assignment model works, a small network example is presented in Figure 6.4. A single demand node pair (1,7) that is served by six routes is illustrated. The shortcomings of previous models can be seen in Figure 6.4: 1) Four nodes connecting to Centroid 7 through the centroid connector cannot be treated as a single node; and 2) the considerations for passenger route choice should be involved with both the numbers of transfers and the numbers of long walks. Therefore, previous trip demand assignment models should be modified to accommodate these changes.

The transit trip assignment procedure gives several input specifications. In this example, the penalty for each transfer is chosen to be 5 minutes of equivalent in-vehicle travel time. Equal weights are put on different components of the transit user time. Table 6.2 shows the Link Travel Time (expressed in minutes) and Route Headways (hours per bus) for this network example. Since there is no direct route service available for passengers traveling between zone 1 and zone 7, the existence of 1-transfer-0-longwalk paths and the 0-transfer-1-longwalk paths is checked. Table 6.3 presents the four candidate paths and their characteristics including all the components of passenger travel costs.

According to the transit trip assignment model presented before, the results of Transit Trip Assignment from demand node pair (1, 7) are given in Table 6.4. From these results, it can be seen that if there are 100 transit users who want to travel from zone 1 to zone 7, then this transit trip assignment model will distribute 56 passengers to P3, the third candidate path and 44 passengers to P4, the fourth candidate path as shown in Table 6.3. These results clearly suggest that the model is more reasonable than that obtained from previous methods because a significant number of transit users will choose the best path on P3 as expected. However, no demands are distributed on this path if previous models are used and obviously this result is not plausible. Therefore, this method is more applicable to the real-world transit route network design problem.

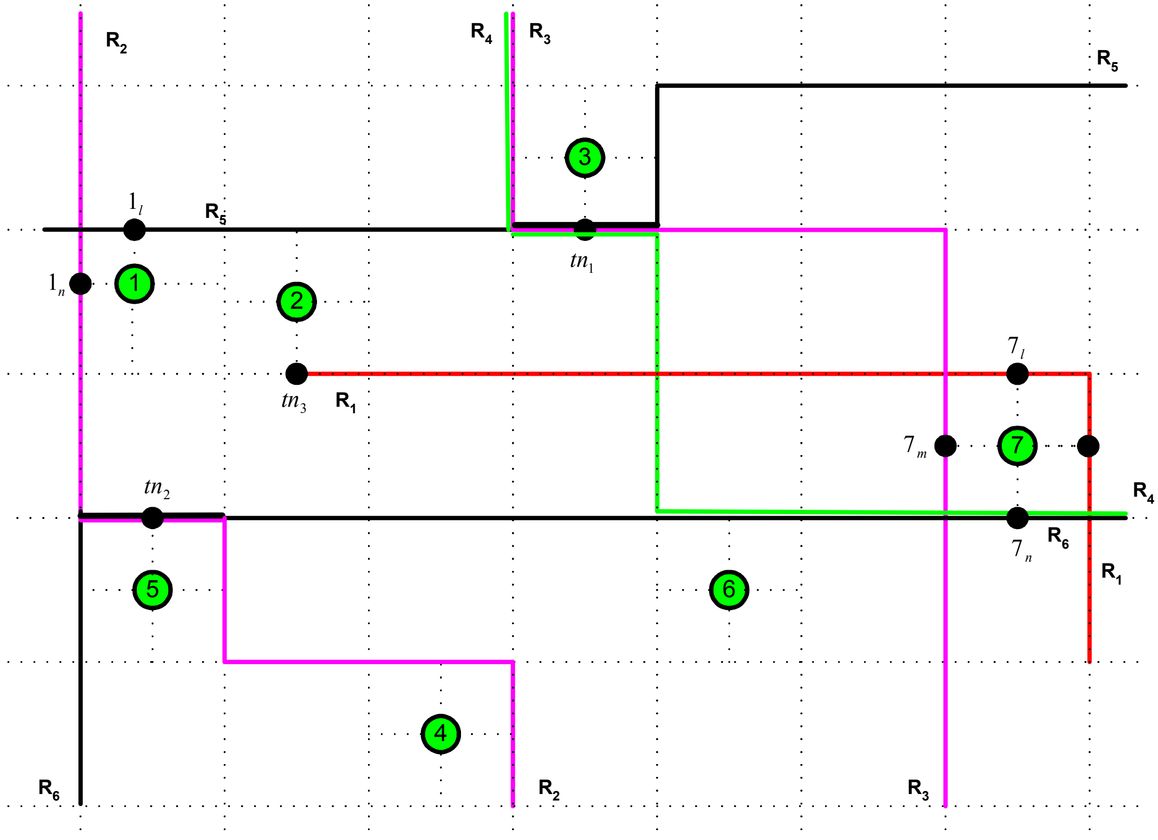


Figure 6.4 Network Examples for the Trip Assignment Model

Table 6.2 Link Travel Time (minutes) and Route Headways (hour per bus)

Links	L1	L2	L3	L4	L5	L6
Link Node Pairs	$1_l \to tn_1$	$tn_1 \to 7_m$	$tn_1 \to 7_n$	$tn_3 \to 7_l$	$1_n \to tn_2$	$tn_2 \to 7_n$
Travel Times	8	10	9	15	4	10
Routes	R1	R2	R3	R4	R5	R6
Route Headways	8	12	12	15	20	10

Table 6.3 Candidate Paths and their Characteristics

	Paths			
	P1	P2	P3	P4
Path Category	1transfer-0longwalk	1transfer-0longwalk	0transfer-1longwalk	1transfer-0longwalk
Path Components	$1 \rightarrow 1_l \xrightarrow{R_5} tn_1 \xrightarrow{R_3} 7_m \rightarrow 7$	$1 \rightarrow 1_l \xrightarrow{R_5} tn_1 \xrightarrow{R_4} 7_n \rightarrow 7$	$1 \rightarrow tn_3 \xrightarrow{R_1} 7_l \rightarrow 7$	$1 \rightarrow 1_n \xrightarrow{R_2} tn_2 \xrightarrow{R_6} 7_n \rightarrow 7$
Walking Cost	1+1	1+1	5+1	1+1
Waiting Cost	10+6	10+7.5	4	6+5
In-vehicle Travel Cost	8+10	8+9	15	4+10
Transfer Penalty	5	5	0	5
Overall Travel Cost	41	41.5	25	32
Filter Results	N	N	F	F

Note: 1) All the cost is uniformly expressed in minutes. Threshold is chosen to be 30%;
2) "N" means the filter result is "non-feasible" and "F" means the filter result is "feasible".

Table 6.4 Results of the Transit Trip Demand Assignment

Paths	Percentage of Transit Demand Assignment	
	Current Method	Previous Method
P1	0	0.31
P2	0	0.30
P3	0.56	0
P4	0.44	0.39

6.7 Summary

The Network Analysis Procedure (NAP) is a procedure that can analyze and evaluate the alternative network structures and determine their associated service frequencies. For any given route network configuration, many system performance measures including associated route service frequencies that can reflect the quality of service are computed. In addition, the transit user cost, operator cost (i.e., the fleet size) and the unsatisfied demand cost are determined.

The NAP discussed in this chapter differs from the previous approaches in four main aspects: 1) the ability to explicitly consider the transfer and long-walk related characteristics among routes under a much more real situation (in the context of centroid nodes); 2) the ability to assign the trip demands under a much more complex situation that considers the numbers of transfers and the numbers of long walks as the most important criteria and to solve the innovative trip demand assignment procedure using a nonlinear analytical allocation model for the first time; and 3) the ability to explicitly consider the transit trip assignment model in the context of variable transit demands.

Two major components of the NAP, namely, the transit trip assignment model and the frequency setting procedure are presented. The algorithm skeleton and the details of its solution methodologies are discussed. Characteristics associated with each component are described, and this provides a solid basis for the solution algorithm implementations for the BTRNDP in the next chapters.

CHAPTER SEVEN

The TRNDP WITH FIXED TRANSIT DEMAND

7.1 Introduction

The previous chapters have covered model formulations and corresponding solution methodology for the BTRNDP. Chapter 3 discussed the model formulations and Chapter 4 proposed solution methodology, which consists of three main procedures, namely the initial candidate route set generation procedure, the network analysis procedure and the heuristic search procedure. The characteristics of six solution approaches, i.e., the genetic algorithm, local search, simulated annealing, random search, tabu search algorithm and the exhaustive search method were also presented. Chapter 5 discussed the details of the ICRSGP and chapter 6 presented a detailed description of the NAP including the transit trip assignment model and frequency setting procedure.

This chapter focuses on the solution framework for the BTRNDP with fixed demand. Previously proposed concepts including the genetic algorithm, local search, simulated annealing, random search, tabu search algorithm and exhaustive search method, which are intended to systemize and organize the ICRSGP and NAP, are discussed in detail. A similar solution framework for the BTRNDP with variable transit demand is presented in Chapter 8. Comprehensive experiments and corresponding numerical results using these six algorithms are discussed in chapter 9.

This chapter is organized as follows. Section 7.2 presents the solution framework of the genetic algorithm implementation model for the BTRNDP with fixed demand. Implementation models of the BTRNDP with fixed demand using local search, simulated annealing, random search, tabu search algorithm and exhaustive search solution methods are discussed in sections 7.3, 7.4, 7.5, 7.6 and 7.7 respectively. Section 7.8 uses the GA model as a representative example for these algorithms to illustrate the network applications for the BTRNDP with fixed demand. Finally, a summary concludes this chapter in section 7.9.

7.2 Genetic Algorithm Implementation Model

As mentioned before, since the GA provides a robust search as well as a near optimal solution in a reasonable time, this approach is employed as one of the candidate heuristic search solution techniques for the BTRNDP. Before implementing the genetic algorithm model, a set of potential routes, consisting of the whole solution space, has been generated by the ICRSGP. The objective of the genetic algorithm model presented here is to scientifically guide the transit route solution set generation process and select an optimum set of routes from the candidate route set solution space with the sum of the total user cost, operator cost and unsatisfied demand cost being minimized.

Input data for the studied network, including the node, link, zone and network data, are required and must be defined by the user. The user-defined data must also include the value of a minimum and maximum number of bus routes in the solution network and a series of GA-related parameters such as the population size (which partially depends on the network size and the users' knowledge), the number of generations, the crossover probability and the mutation probability. Generally speaking, the Genetic Algorithm Implementation Model can be presented as follows.

Step 1. Set n=1 and initialize all the performance measure parameters;
Step 2. Define the n-related dynamic allocated arrays required for GA implementation;
Step 3. Generate the GA population formulation for the current route set size and initialize each chromosome (population) randomly;
Step 4. Set generation=0;
Step 5. Call the network analysis procedure;
Step 6. Evaluate the objective function;
Step 7. Keep the current solution;
Step 8. Generate next population
- Selection
- Crossover
- Mutation

Step 9. If generation<_MAX_GEN, increment generation by 1 and go to step 5; else update the current best solution if improved;
Step 10. Set n=n+1;
Step 11. If n<=_MAX_ROUTES, go to step 2 and repeat the same process;
Step 12. Output the best route set from the best solution found, and several performance measures.

Figure 7.1 presents the flow chart of a genetic algorithm implementation model for the BTRNDP with fixed demand. Details of the whole process can be described as follows:

In Step 1, the route set size (i.e., the number of routes in the solution network) is set to 1. All performance measures such as the parameters used for representing the seven transit demand categories, _direct_route (i.e., 0-transfer-0-longwalk), 1-transfer-0-longwalk, 0-transfer-1-longwalk, 1-transfer-1-longwalk, 2-transfer-0-longwalk, 0-transfer-2-longwalk and unsatisfied demand are initialized;

In Step 2, computer memory to be occupied by dynamically allocated arrays is established for the GA components including the gene, the population, and the updated population. Note that the space is related to the route set size. The larger the route set size, the more the space needed to meet the requirements;

In Step 3, the GA population formulation for the current route set size is generated and the initial population (i.e., starting feasible solution) is generated randomly;

In Step 4, the generation number, which is used to record the number of generations for the current number of route set of the genetic algorithm, is initialized to 0;

In Step 5, the network analysis procedure is called to perform the transit trip assignment and frequency setting procedures. The output of this step is the transit demand between each centroid node pair allocated to a specific route and the frequencies on each route in the current population. Several different transit demand related parameters, including _direct_route, 1-transfer-0-longwalk, 0-transfer-1-longwalk, 1-transfer-1-longwalk, 2-transfer-0-longwalk, 0-transfer-2-longwalk and unsatisfied demand, and information regarding each node, link, route and network are obtained accordingly. Also the objective function that is represented as the sum of the user cost, operator cost and the unsatisfied demand cost (exactly as included in the model formulation is chapter 3) is obtained;

In Step 6, the objective function obtained in Step 5 is evaluated;

In Step 7, the current best solution (namely, the population with least objective function) up to current generation is kept;

In Step 8, the next population is generated through the GA selection, crossover and mutation processes;

In Step 9, the generation number is compared to the user-defined maximum number of generations or other stopping criteria. If the maximum number of generations has not been reached, then increment the generation by 1 and goto step 5 and repeat the same process. Otherwise, update the current best solution if improved;

In Step 10, the route set size (i.e., the number of routes in the proposed transit solution network) is incremented by 1;

In Step 11, the number of routes is checked. If it doesn't exceed the user-defined maximum number of routes, then goto step 2 and repeat the same process until the maximum route set size is reached;

In Step 12, the GA route population with the least objective function value is outputted as the best solution found. Meanwhile, several performance measures are obtained.

The GA implementation model for the BTRNDP has been successfully tested and details of the numerical results are presented in Chapter 9.

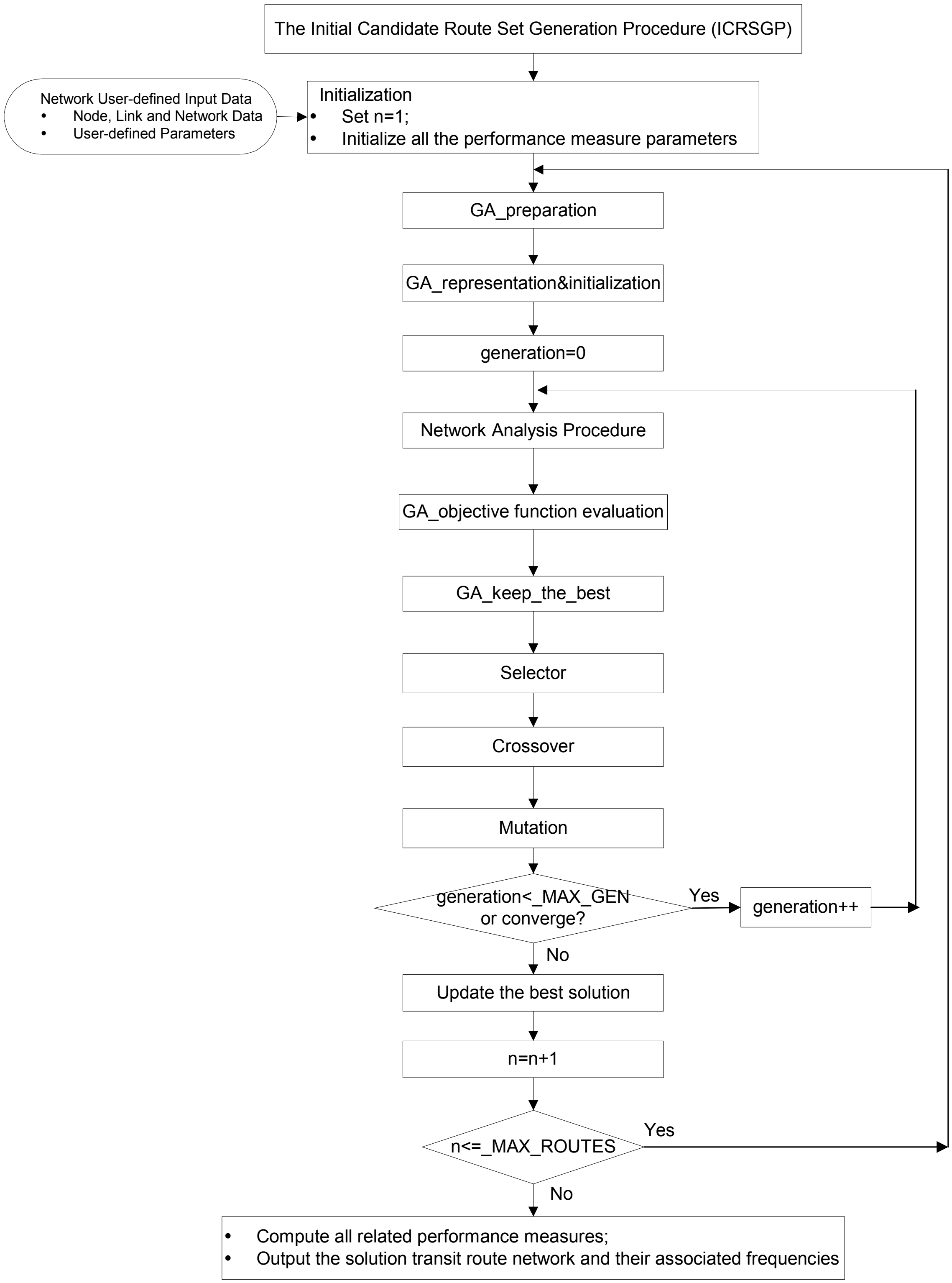


Figure 7.1 Genetic Algorithm Model for the BTRNDP with Fixed Demand

7.3 Local Search Implementation Model

As described in Chapter 4, the local search method is used as one of the solution approaches for the BTRNDP. Generally speaking, there are several variations of this basic algorithm. Different definition rules could result in solutions of different quality. Furthermore, the definition of the "neighborhood", i.e., the nearby solutions of the current solution, might affect the quality of the solution most. Note that like the GA, the local search method is employed to choose an optimal route set from the whole route set solution space that has been generated by the ICRSGP.

In this research, the "neighborhood" for any route *i* is defined as the routes that are next to route *i* in the solution space because the neighbors will share a significant amount of structure (say, the origin and destination node might be the same but are on different k-th shortest paths or even have the same centroid origin and centroid destination node). Furthermore, this definition rule is very simple and easy to implement. As a result, this definition is employed for all heuristic algorithms that are based on the definition of "neighborhood" including LS, SA and TS algorithms (except GA and RA) in this research. Figure 7.2 presents the local search implementation model for the BTRNDP with fixed demand.

7.4 Simulated Annealing Implementation Model

Simulated annealing can be regarded as a "randomized variation" of the local search method. Compared to the local search method, simulated annealing is a more advanced algorithm because it attempts to minimize the probability of being stuck in a low-quality local optimum. As described in chapter 4, randomness can help the search process as follows: 1) Randomly choose a local move. If it is improving, then take it as the current best solution; 2) If it is a tie, randomly decide whether to take it; and 3) If it is only a little worse, randomly decide whether to take it as the updated solution.

Note that the neighborhood definition is the same as that in LS and Figure 7.3 presents the simulated annealing implementation model for the BTRNDP with fixed demand.

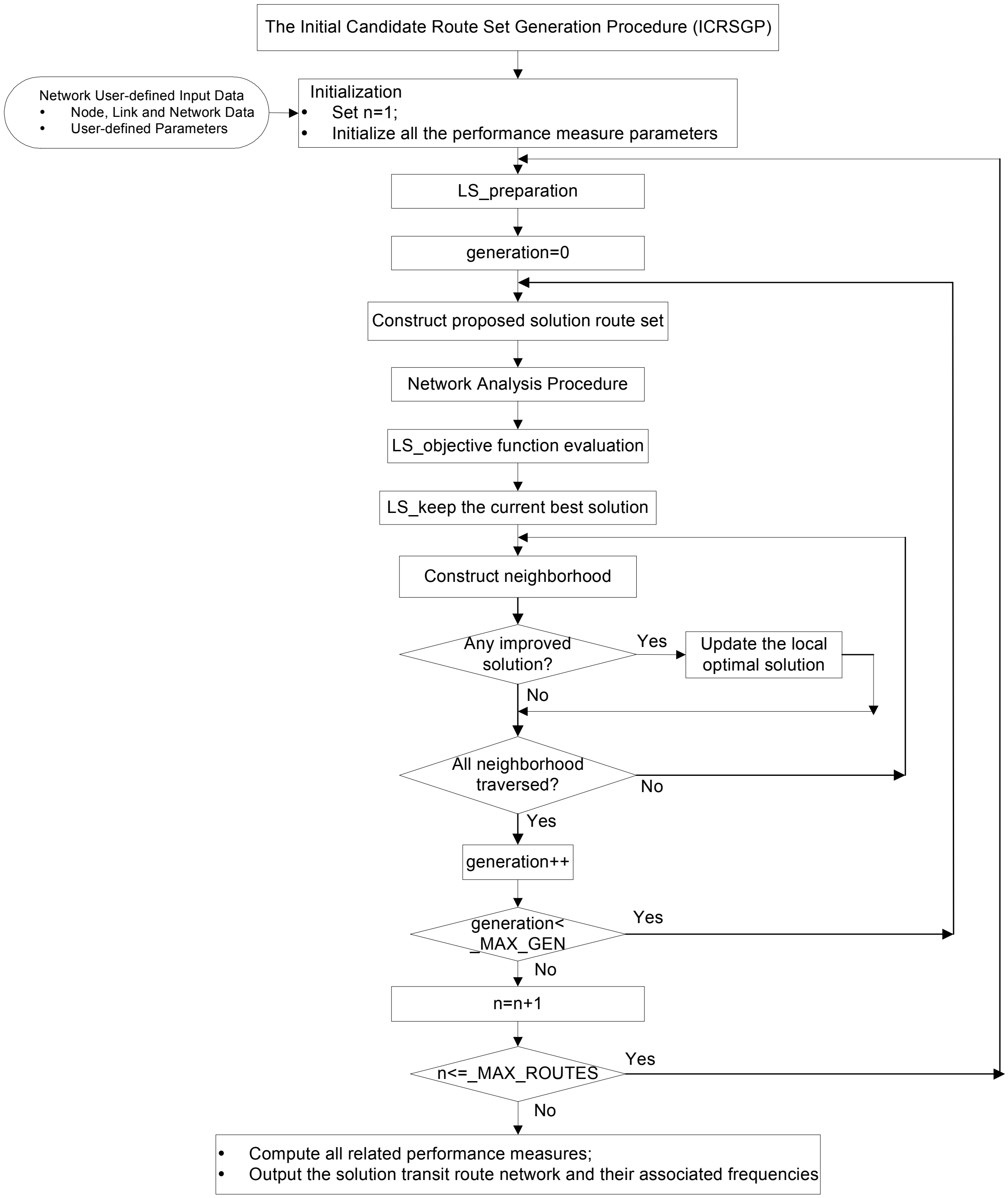


Figure 7.2 Local Search Model for the BTRNDP with Fixed Demand

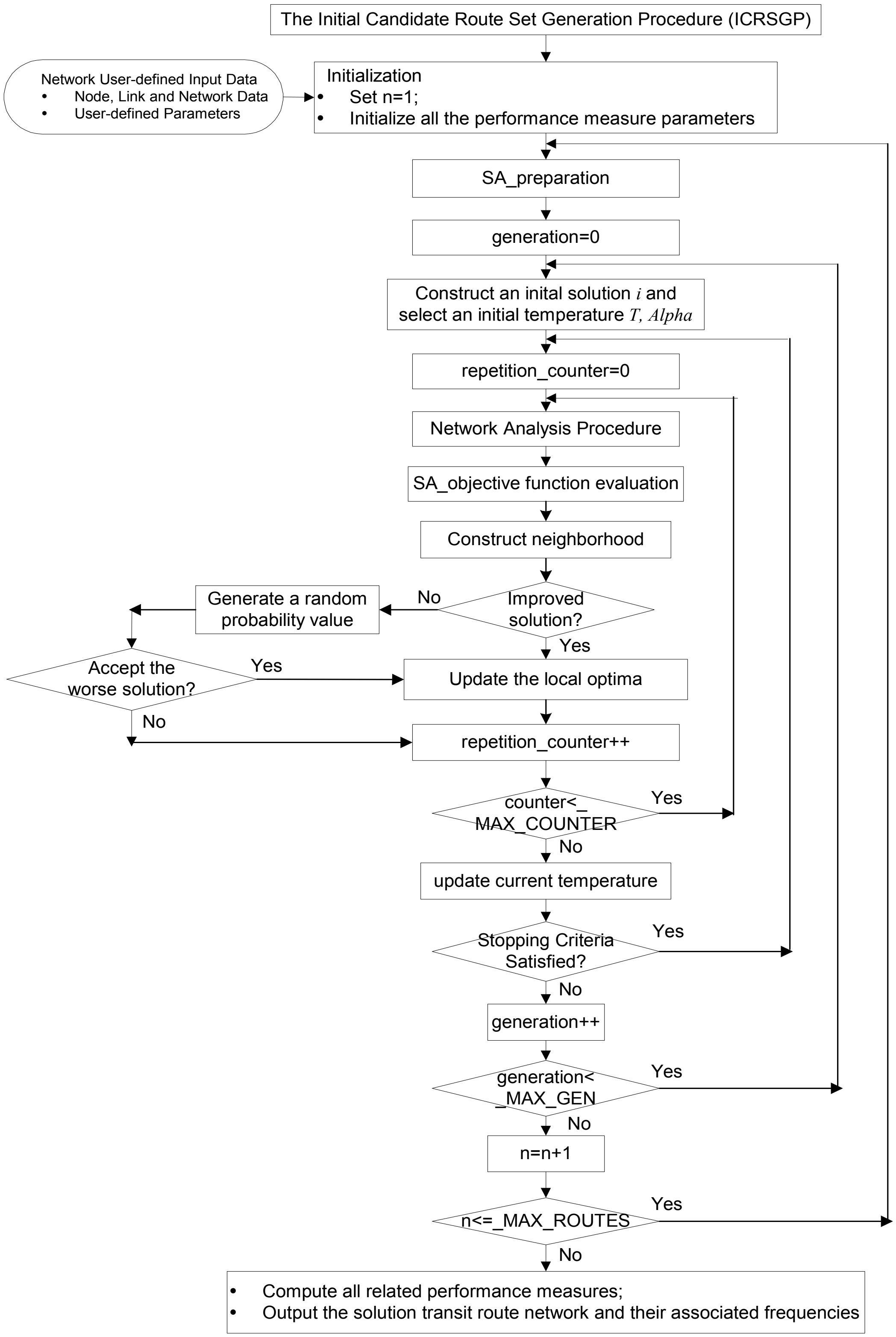


Figure 7.3 Simulated Annealing Model for the BTRNDP with Fixed Demand

7.5 Random Search Implementation Model

As described in chapter 4, random search is essentially a Monte Carlo based simulation optimization method. When it is used for the BTRNDP, this method randomly chooses a solution for each step of a specified route set size from the solution space and evaluates the generated solution sets by comparing their objective function values. The whole process is repeated until the route set size reaches the user-defined maximum route set size. The optimal transit route network solution is the one with least objective function magnitude (i.e., the least sum of total user cost, operator cost and unsatisfied demand cost). Since this method is simple and easy to implement, random search is employed as one of the solution approaches in this research. Figure 7.4 presents the random search implementation model for the BTRNDP with fixed demand.

7.6 Tabu Search Implementation Model

As other heuristic algorithms, applying tabu search (TS) methods require a significant amount of knowledge specific to that problem. To make tabu search a potentially efficient algorithm for the BTRNDP, careful attention is required. Note that one of the significant contributions in this research is using the tabu search algorithm to solve the BTRNDP both with fixed and variable transit demand. Since it is the first time for the tabu search methods to be applied for the BTRNDP, a detailed description of the BTRNDP-specific tabu search is presented.

7.6.1 Solution Representation

At any iteration t of the algorithm, let n represent the proposed solution route set size. A candidate bus transit route solution network can be represented by $X^t = (R_1^t, R_2^t, \cdots R_i^t, \cdots, R_n^t)$, where R_i^t $(i = 1,2,\cdots,n)$ denotes the i-th bus route in the proposed solution set. Let $f(X^t)$ represent the objective function as shown in Chapter 3 for the proposed solution network defined by this n transit route network configuration $X^t = (R_1^t, R_2^t, \cdots, R_n^t)$.

7.6.2 Initial Solution

In this research, all initial solutions for three different versions of the tabu search algorithms are generated randomly the same way as that shown in the genetic algorithm, local search, simulated annealing and random search methods.

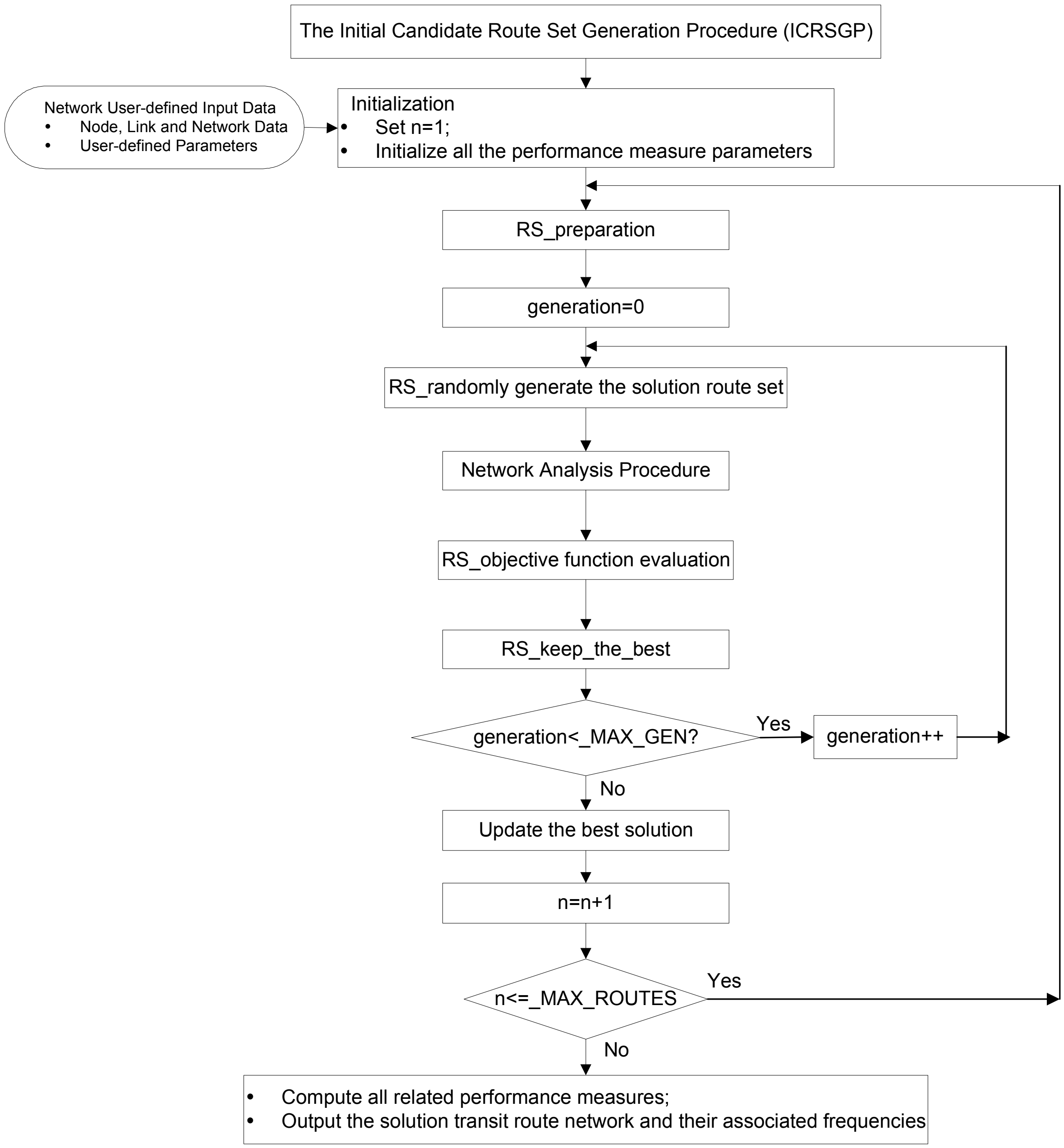


Figure 7.4 Random Search Model for the BTRNDP with Fixed Demand

7.6.3 Neighborhood Structure

The neighborhood of a feasible solution route network set X^t is another feasible solution obtained by moving one of the routes in the current proposed solution set, say the *i*-th route R_i^t to one of the routes that is next to R_i^t (namely R_{i-1}^t or R_{i+1}^t) in the stored solution space. For route 1, the neighborhood can be defined as route 2 and route N, where N is the total number of routes in the stored solution space. For route N, the neighborhood can be defined as route 1 and route (N-1). The neighborhood of any route *i* $(1 < i < \mathrm{N} - 1)$ that lies somewhere in the middle of the solution route space can be defined as the routes that are next to R_i^t. $Z(X_{ij}^t)$, the objective function value of a new solution X^{t+1} that is obtained from X^t by moving R_i^t to one of its neighbors R_j^t at generation *t* can be computed as follows: $Z(X_{ij}^t) = f(X^{t+1})$.

7.6.4 Moves and Tabu Status

At the beginning of this process, no move is tabu (i.e., forbidden). At any iteration, the algorithm executes the best non-tabu move to a feasible neighbor of the current solution. However, if a tabu move yields a better incumbent, it is also implemented. Whenever a move is performed, the reverse move is declared tabu for *m* iterations, where *m* is either a user-defined parameter or a randomly generated one that follows a discrete uniform distribution in an interval $[m_{\min}, m_{\max}]$, where $m_{\min}$ and $m_{\max}$ are the user-defined minimum and maximum parameters of the algorithm. Comparisons of the model performance between these two strategies including the fixed and variable tabu tenure can be performed and sensitivity analyses can be done to get the optimal parameters in either case as will be shown in Chapter 9.

7.6.5 Diversification and Intensification

This part is developed to combine the diversification and intensification procedures to further explore the solution space for a possibly better solution route network. It starts from the best found solution route set and introduces a major perturbation by allowing *q* routes $(1 \le q \le n)$ to move up to *w* more than their current solution location (say *q* = 2 and *w* = 10). This is called “diversification”. To respect the original characteristics of the tabu search, this procedure is never applied more than once during a given operation (called “intensification”). Note that tabu moves are also applied to this situation. If this move is toward one direction (say increasing direction) of the current route, then moves toward to the opposite direction (i.e., decreasing direction) are prevented for a certain number of iterations (say using the same *m*). Model performance comparisons of the tabu search algorithms between using and not using this procedure can be achieved and the better approach will be identified.

7.6.6 Implementation Model Summary

The proposed tabu search algorithms for the BTRNDP in this research include two main procedures and they are described as follows.

1. Neighborhood Search Procedure

At iteration *t*, let $X^t = (R_1^t, R_2^t, \cdots, R_n^t)$ be a feasible solution of value $f(X^t)$. Let $N(X^t)$ be the set of feasible neighbors of X^t, as defined before. The best neighbor of X^t is a solution $X_{i^*j^*}^t \in N(X^t)$ obtained by moving $R_{i^*}^t$ to its best neighbor $R_{j^*}^t$. Similarly define the best feasible non-tabu neighbor of X^t as $X_{\bar{ij}}^t \in N(X^t)$. ($X_{i^*j^*}^t$ and $X_{\bar{ij}}^t$ may coincide). Let X^* be the incumbent (the best known feasible solution) and let $Z(X^*)$ be its value.

If $Z(X_{i^*j^*}^t) < Z(X^*)$, set $X^* = X^{t+1} = X_{i^*j^*}^t$ and $Z(X^*) = Z(X^{t+1}) = Z(X_{i^*j^*}^t)$. Declare the move of a route from $R_{j^*}^t$ to $R_{i^*}^t$ tabu for *m* iterations, where *m* can be a fixed user-defined parameter or is uniformly distributed with $m \in [m_{\min}, m_{\max}]$. If $Z(X_{i^*j^*}^t) > Z(X^*)$ and all moves defining the solutions of $N(X^t)$ are tabu, set $\delta = 1$ and return. Otherwise, set $X^{t+1} = X_{\bar{ij}}^t$ and $Z(X^{t+1}) = Z(X_{\bar{ij}}^t)$. Declare the move of a route from R_j to R_i tabu for *m* iterations, where *m* has the same definition as used before.

2. Diversification and Intensification Procedure

This procedure is the same as that in Neighbor Search but defines $N(X^t)$ differently. It allows *q* routes ($1 \le q \le n$) to move up to *w* more than the current solution location in the solution space (Note that in this research, this procedure is called the "shakeup" procedure. Furthermore, for simplicity, *q* is set to *n* and *w* is set as a user-defined parameter). When a route is moved in one direction (say the increasing direction), moving back in the opposite direction is declared tabu for *m* iterations, where *m* uses the same notation as before.

Tabu Search Algorithm for the BTRNDP

Step 1. Randomly generate an initial feasible solution route network $X^t = (R_1^t, R_2^t, \cdots, R_n^t)$ with route size *n* in the proposed solution set.

Step 2. Set $\delta = 0$, $t = 1$ and $X^* = X^t$;
While ($\delta = 0$ and $t \le \text{MAX\_Iterations}$)
  Apply Neighborhood Search to the solution X^t;
  $t = t + 1$.

Step 3. Apply the "Diversification and Intensification" procedure to X^*.
Apply Neighborhood Search to the solution X^* until $\delta = 1$ or $t > \text{MAX\_Iterations}$).

Step 4. Output the current best solution found.

As mentioned before, since TS provides a robust search as well as a near optimal solution within a reasonable time domain, this algorithm is employed as the solution technique for the BTRNDP. Before implementing the Tabu Search algorithms, a set of potential routes, consisting of the whole solution space, has been generated by the ICRSGP. The objective of the Tabu Search algorithm presented here is to select an optimal set of routes from the candidate route set solution space with the sum of the total user, operator and unsatisfied demand cost being minimized.

A flow chart that provides the Tabu Search algorithm-based solution framework for the BTRNDP can be seen in Figure 7.5. Note that the "neighborhood" for any route *i* is defined as the route left or right of route *i* stored in the solution space. At the beginning of the TS implementation, the initial solution is randomly generated. In the second (and later) generation, the HSP is used to guide the generation of the new transit route solution set and after it is proposed at each generation, the search process is started. The network analysis procedure is then called to assign the transit trips between each centroid node pair and determine the service frequencies on each route and evaluate the objective function for each proposed solution route set. For each iteration, if a solution route set is detected to improve over the current best one, the current best solution is updated. The new proposed solution sets are generated and are evaluated in the same way. If convergence is achieved or the number of generations is satisfied, the iteration for a specific route set size ends. Then, the proposed solution route set size is incremented and same processes are repeated until the maximum route set size is reached. The best solution among all transit route solution sets is adopted as the optimal solution to the BTRNDP for the current studied network.

Moreover, in this research, three versions of TS algorithms are used: 1) Tabu search without shakeup procedure (i.e., without the diversification and intensification procedure as defined before); 2) Tabu search with shakeup procedure and fixed tabu tenure (i.e., the number of restrictions set for the tabu moves are fixed); and 3) Tabu search with shakeup procedure and variable tabu tenure (i.e., the number of restrictions set for the tabu moves are randomly generated). All three different variations of tabu search methods are implemented and algorithm comparisons are presented in Chapter 9.

7.7 Exhaustive Search Implementation Model

As mentioned in Chapter 4, exhaustive search is an approach to search for the global optimal solution by enumerating and comparing the objective functions for all possible solutions. Note that in this research, the sole purpose of employing ESM to solve BTRNDP is to use its global optimal solution as the benchmark to examine the efficiency and measure the quality of solutions obtained from the heuristic algorithms, especially when the network size is small. Figure 7.6 gives the exhaustive search implementation model for the BTRNDP with fixed demand.

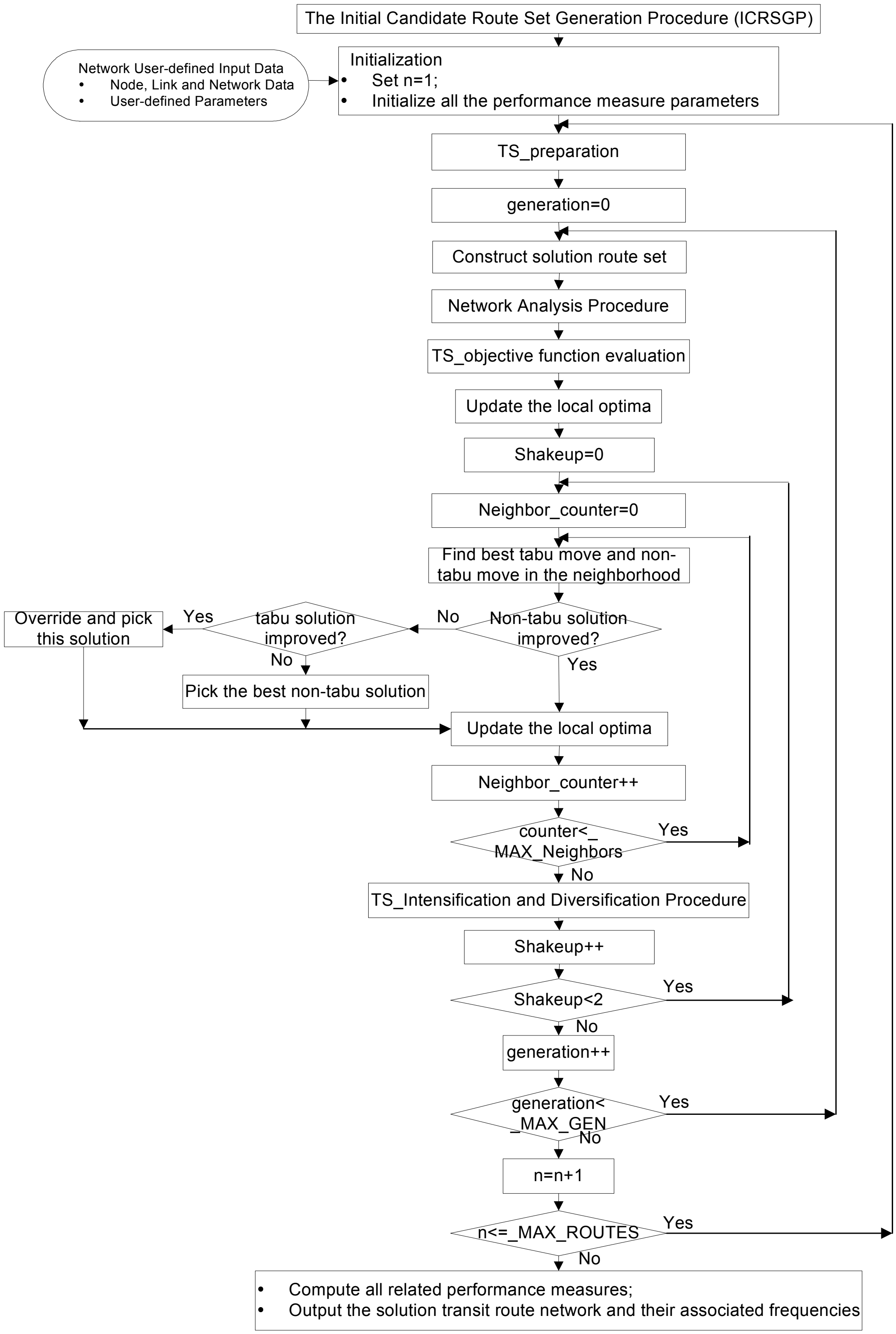


Figure 7.5 Tabu Search Model for the BTRNDP with Fixed Demand

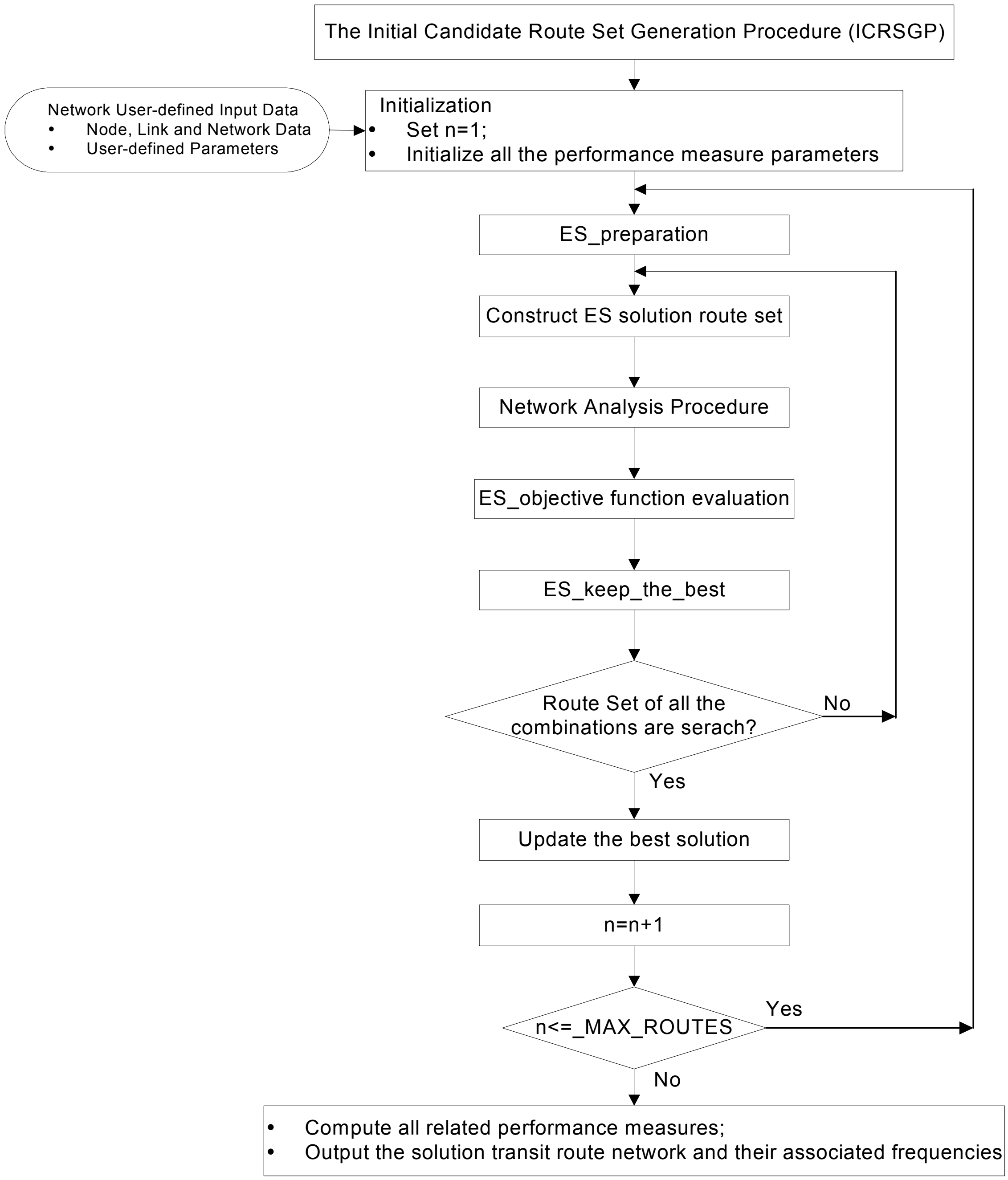


Figure 7.6 Exhaustive Search Model for the BTRNDP with Fixed Demand

7.8 Example Network Illustrations

This section uses the GA model as an example to present the application of these heuristic algorithms to a small network. In order to better explain this method, the same example network used in previous chapters is used. After the transit route solution space is generated by the ICRSGP, the GA model is used to find the optimum set of routes from this space. Fixed string length coding is employed to represent the GA coding-decoding scheme. The previous flow chart can be reformulated as follows. The number of routes will be fixed in each step. Only the solution route set would be considered as variables and an optimal route set will be found for each chosen number of routes. One can find the best number of routes and route set by comparing optimal solutions for the various cases.

It can be seen from Chapter 5 that the solution space contains 286 feasible routes. Let any feasible route be labeled as an integer between [1, 286] and the precision requirement of the route number is p=0.1. Therefore, to represent these 286 routes, one needs the string length to be 12 using equation ($2^{11}-1<(286-1)10^{1}<2^{12}-1$) as shown in Chapter 4. Suppose that one needs to decode a typical sub-string 101100100101. Using the transformation method described in the GA section of Chapter 4, x_i can be computed as follows: $x_i = 1+2853*(286-1)/(2^{12}-1)=199.6$. As a result, the 200th route is adopted as the route corresponding to the substring 101100100101 by simply rounding off this number.

As mentioned, during the GA model implementation process for BTRNDP with fixed demand, the GA algorithm tries to find the optimum set of routes from the solution space consisting of these 286 candidate routes. In implementing GA, the route set varies from each outer loop (say 1 to 4 at most) but is fixed inside the inner loop during each successive generation. To start the iteration, the minimum route set size, which is 2 routes for this example network, is adopted. Suppose one sets the population size to be 3 and the current route set size in the proposed solution is 2. These 286 candidate routes will compete for the optimum set with a feasible solution containing only two routes. Therefore, each population contains two routes that are represented by a string containing two substrings, as shown in Figure 7.7, in which each substring has length 12 and represents a specific route number in binary code. Then, in this case, each population, which consists of two routes with a string length of 24 and concatenated by two substrings, represents a candidate solution route set in the solution space defined by the 286 candidate routes. Once the code and decode scheme are finalized, the GA model can be used to find the optimal route set for the studied network. Note that the initial solution is formed by randomly generating a population of such solution vectors. The coding for two individual candidate routes in a population is shown in Figure 7.7.

Candidate set (1)

0	1	0	0	1	0	1	1	0	0	0	1		1	1	0	0	0	0	1	0	1	0	0	0

(Route no.: 85) (Route no.: 218)

Candidate set (2)

1	0	1	0	1	0	1	1	1	0	0	1		1	1	0	1	0	0	1	0	0	0	0	1

(Route no.: 192) (Route no.: 235)

Candidate set (3)

0	0	0	0	1	0	0	0	0	0	0	0		0	0	0	0	0	0	1	0	0	0	0	1

(Route no.: 10) (Route no.: 3)

Figure 7.7 Graphical Representations for Each Chromosome in GA Model

Table 7.1 presents a simple representative snapshot of the solution space for the example network (Related demand matrix and parameter settings for this network are described in Chapter 9). After the finalization of the coding and decoding scheme and the generation of the initial population, the genetic search can be started. The network analysis procedure is called to get the transit demand distributed and determine the service frequencies associated with each route set. Then, the population is evaluated, i.e., to find the objective function value (as shown in the previous model) of each individual in the population. It should be noted that each individual provides a solution of a set of routes that tries to satisfy the demands, subject to the constraints (The user-defined parameters included in the constraints are presented in Chapter 9). Also, the current best solution is kept. In the next steps, the genetic operators, namely, selection, crossover, and mutation, which have been examined before, are applied to this initial population. New populations are generated using these operators and are evaluated in the same way. For each iteration, the best solution will be updated if the solution improves over the current one. The iteration for a specific route set size completes if convergence is achieved or the number of generations is satisfied. Then, the size of the route set is incremented and successive iterations are made. All the work will be repeated until the maximum route set size is reached. The best solution among all the route sets is adopted as the solution. Details of the numerical results are presented in Chapter 9.

7.9 Summary

This chapter focuses on six different solution frameworks, namely, the genetic algorithm, local search, simulated annealing, random search, tabu search algorithms and exhaustive search model, for the BTRNDP with fixed demand. Related algorithm skeletons for each solution approach are sequentially presented. As a typical illustration example for these six algorithms, a GA model is applied to solve the BTRNDP with fixed demand using a small network. Computer programming indicates the validity, effectiveness and efficiency of the pilot study of the example network. The next chapter presents related characteristics for the BTRNDP with variable demand and its corresponding solution framework. Applications of these solution frameworks to comprehensive experiments are conducted and associated numerical results are presented in Chapter 9.

Table 7.1 Representation of the Route Set Solution Space for the Example Network

Route Set Size	Route Number	Orig.	Dest.	Shortest Path or k-th Shortest Path Representations of Nodes and Links	Dist.	Traversed Zones	Set Headway	Operator Costs	User Costs	Unsatisfied Demand Costs	Fleet Size
1	85	27	29	29-10-21-11-7-24-8-27	1325	2,3,4,5,6,7	35	40000	1334790	3300000	4
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
2	262	27	30	30-12-29-10-21-11-7-24-8-27	1525	2,3,4,5,6,7	60	50000	782498	0	5
	194	21	25	25-3-23-2-19-6-22-7-11-21	1375	1,2,3,4,5	60				
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
3	262	27	30	30-12-29-10-21-11-7-24-8-27	1525	2,3,4,5,6,7	60	80000	1587223	0	8
	193	21	23	23-2-19-6-18-5-20-10-21	1400	1,2,3,5	60				
	229	24	26	26-3-23-2-19-6-22-7-24	1200	1,2,3,4,6	47				
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
_MAX _ROUTES	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

CHAPTER EIGHT

The TRNDP WITH VARIABLE TRANSIT DEMAND

8.1 Introduction

The preceding chapters presented solution frameworks for the BTRNDP under the assumption that the transit demands are fixed. Although this assumption of fixed transit demand makes the BTRNDP much simpler to solve, this approach is also problematic because the transit demand actually depends on a specific combination of the transit network structure and the city highway transportation network. Furthermore, the BTRNDP is always investigated for a city highway network that already exists. Therefore, one can expect that different transit network patterns can and will result in different transit demands, even given the same city highway transportation network. The actual interactions between the transit network patterns and the variable demand makes the optimal solution to the BTRNDP with fixed transit demand questionable in real world situations. Therefore, when considering the BTRNDP, one needs to take the variable demand nature into account.

Generally speaking, variable demand can be classified into two cases: 1) Variable total demand, which may result from the feedback process of the Urban Transportation Planning Process (UTPP); and 2) Variable transit demand, which is due to the variable relationship in modal split between auto and transit mode under given total travel demand. Motivated by this, this chapter focuses on the solution framework for the BTRNDP with variable demand. A systematic literature review on variable demand is performed and a corresponding solution framework for the BTRNDP with variable demand is developed.

This chapter is organized as follows. Section 8.2 describes characteristics of variable total demand. Section 8.3 examines characteristics of variable transit demand. The mode split procedure including the attributes of alternatives and the setting decision rule for mode choice between the auto and transit modes is discussed. Utility and disutility functions are introduced and the commonly used multinomial logit model and nested logit model are also reviewed. Underlying characteristics of the MNL and NL models are discussed and advantages and disadvantages of each model are pointed out. A two-stage BLM-IPM model (binary logit model-inversely proportional model) for determining the mode choice between auto and transit routes is presented for the BTRNDP with variable transit demand. Section 8.4 discusses the NAP for the BTRNDP with variable transit demand. Section 8.5 presents the solution framework for the BTRNDP with variable transit demand. Section 8.6 concludes this chapter with a summary.

8.2 Variable Total Demand

Generally speaking, transportation demand-supply characteristics can be seen from a microeconomics point of view. On one hand, the spatial separation of people and goods creates demand for transportation; on the other hand, supply of transportation is represented by the service of the whole transportation (including the highway and transit) network system. At this point, the objective of travel demand forecasting is to quantify the amount of travel on the whole transportation system.

As can be seen in the literature, the commonly used methodology in travel demand forecasting process is the traditional four-step process. As explained by its name, this process consists of four basic phases (see Khisty and Lall, 1998):

1. *Trip Generation* forecasts the number of trip ends generated in each zone;
2. *Trip Distribution* connects origin-destination trip ends forecasting trips;
3. *Mode Choice* predicts how the trips will be divided among the available modes of travel;
4. *Network Assignment* predicts the routes or paths that the trips will take.

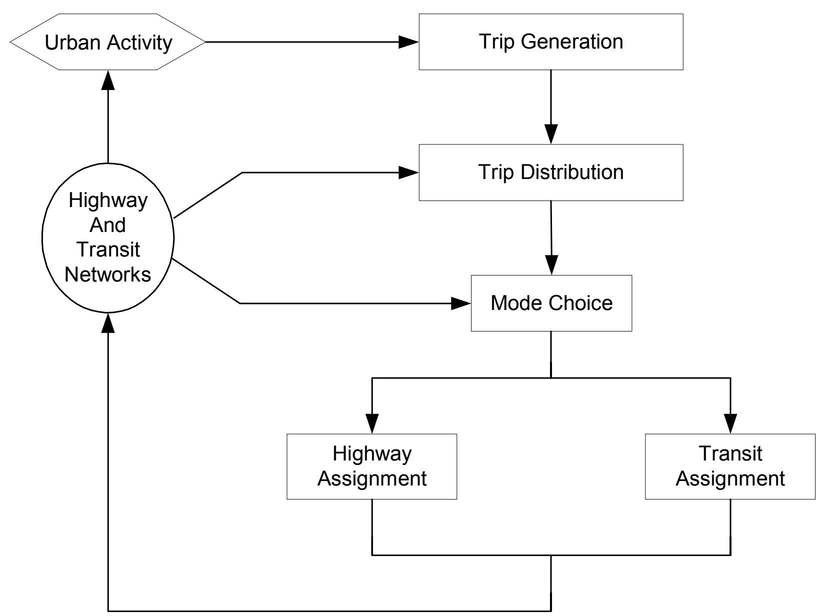


Figure 8.1 Traditional Four-Step Process Used in Transportation Planning (FHWA/UMTA, 1977)

Figure 8.1 shows how these four phases fit together into the travel demand forecasting process that has been used in the Urban Transportation Planning Process. From this figure, one can see that total travel O-D demand can be estimated from the first two steps of the UTPP, which are trip generation and trip distribution. Therefore, the problem comes from the fact that the generation of the total travel O-D demands (including the auto and transit demand) generally results from the services provided by the existing highway and transit networks. On the other hand, the optimal transit route network and the optimal highway network design problems should be solved given the total travel demand between any planning zone pair. The variable relationship between the total travel demand and the solution network suggests that the optimal transit route network design should be part of UTPP as shown in Figure 8.1. Put another way, the design of the optimal transit route network should be solved simultaneously rather than sequentially with the determination of total travel demand. The urban planning process is not finished until equilibrium (convergence) occurs in total travel O-D demands via appropriate iteration processes. In this case, the highway and transit route network are obtained at the same time without any additional cost. However, total work involved this planning process would make it practically intractable. Furthermore, many assumptions must be made and several parameters must be chosen and these would make the results questionable (at least less precise) even if the whole process could be tackled. Therefore, although the variable total travel O-D demand theoretically should be considered in the BTRNDP, for simplicity, total demand is not investigated in this research. Instead, the highway network and the total travel O-D demands between travel zone pairs in the studied transportation network are assumed to be fixed and given throughout this research.

8.3 Variable Transit Demand

Given the city highway and transit network, generally speaking, three factors can influence transit demand at the aggregate level: 1) the total travel demand; 2) the attributes (e.g., travel time and cost) of each alternatives (such as auto, bus transit and share ride, etc.) and the characteristics of the trip makers (e.g., household size, household income, the number of autos available and residential density, etc.) and 3) a setting decision rule for the modal split to determine transit demand from the total travel O-D demand. A simple version of the procedures that can be used to estimate the transit O-D demand is shown in Figure 8.2:

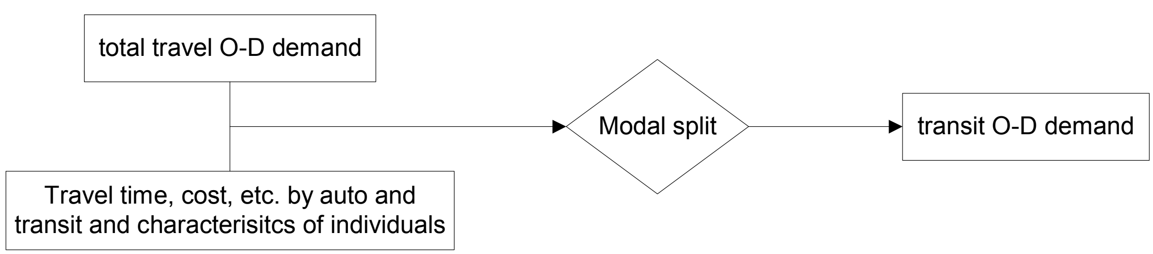


Figure 8.2 Procedures to Estimate Transit O-D Demand

As described in the previous section, total travel demand is assumed as fixed and given. Namely, this research focuses on the BTRNDP under the assumption that the total travel origin-destination demand (i.e., total person trips) is given. Furthermore, characteristics of the travelers are usually not considered in computing the transit demand. The reasons can be seen from two aspects: First, the characteristics of each individual who wants to travel from and/or to a specific zone usually are not known by the forecasters. Second, the current empirical results for mode split usually do not incorporate characteristics of the travelers for the preliminary transportation planning process. However, as can be seen later on, these unobserved traveler characteristics can be captured in the error terms in the discrete choice model structure. Based on these considerations, the following sections only discuss two major components including the attributes of the alternatives and the setting decision rule for modal split.

8.3.1 The Attributes of Alternatives

The attributes of alternatives that are expected to influence traveler preferences/choices among alternatives can include total travel time, in-vehicle travel time, out-of-vehicle travel time, travel cost, the number of transfers (only for transit modes), walk distance, and reliability of on-time arrival etc. Note that the BTRNDP is usually studied as an early stage of the preliminary transportation planning process. Furthermore, even the current discrete mode choice related empirical results usually don't include all the characteristics except two main level-of-service descriptors, namely travel time and cost, of auto and transit mode. For simplicity, only the travel time and travel costs involving auto mode and each candidate transit route are considered in the utility function, which is discussed in section 8.3.2.

Whenever the BTRNDP is discussed, probably one also has to mention the city highway network. This should not seem a surprise since the level of service provided by the transit network is not separated from the highway network. Furthermore, the former depends largely on the latter, as can be seen from the following descriptions.

As mentioned, travel times and travel demand are all related to the feedback processes involved in the urban transportation planning process. It can be expected that bus transit in-vehicle travel time on a specific bus route is dependent on the traffic volumes including passengers, cars and transit on that route at that time. In the meantime, the traffic volumes are determined by the modal split model that is partly dependent on the auto and transit travel time. However, as a major part of the transit travel time, the bus transit in-vehicle travel time largely determines the latter. The following Figure 8.3 clearly shows the cyclic relationships.

Due to the stochastic dynamic nature of traffic flow in the city network, quantifying the bus in-vehicle travel time and auto travel time precisely is a complex process. For simplicity, it is assumed that both auto in-vehicle travel times and transit in-vehicle travel times on the links for a specific combination of bus transit route network and city transportation network are fixed and can be computed. The way to compute the travel time for each mode is discussed in chapter 9. However, transit travel time for the users and the transit demand are still variable, the magnitudes of which depend on the transit route network solution (and route frequencies) to the BTRNDP.

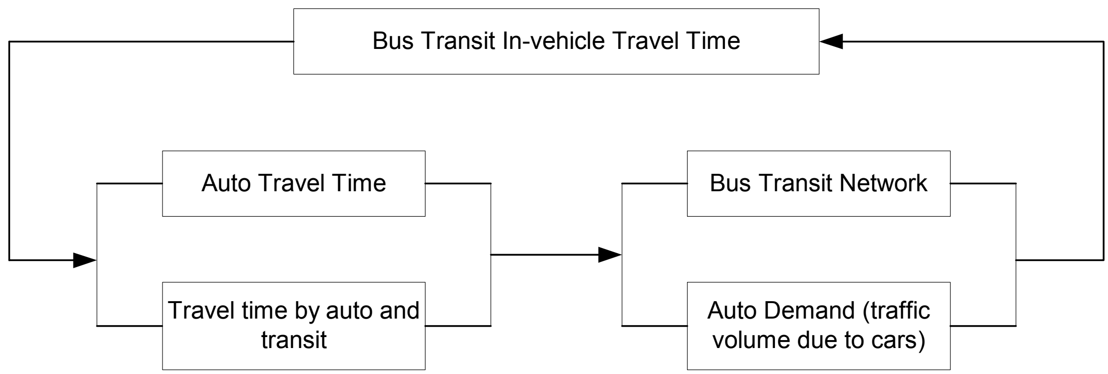


Figure 8.3 Cyclic Relationships regarding Travel Time of Auto and Transit

8.3.2 Setting Decision Rule for the Modal Split

Given the total travel demand, and travel time and cost of each travel mode (auto and transit), a decision rule for determining the modal split is necessary for estimating the distribution of the transit and the auto demand.

In a typical travel situation, trip-makers can select among several travel modes, which may include walking, driving, taking the bus, riding the bicycle, and sharing rides. A mode choice (or mode split) model is concerned with the prediction of how travel will be split among the modes available to the travelers. Mode choice decisions are made according to the individual preference, trip type and relative levels of service associated with the available modes. As described by Papacostas and Prevedouros (2001), it is likely that trip-makers would establish a pattern of mode choice that remains relatively constant as long as the conditions remain the same. When significant changes in these conditions occur, trip-makers respond to varying degrees by shifting from one mode to another.

In this research, it is assumed that trip-makers only have two travel modes, driving the car or taking the bus. This might be a suitable assumption for most cities of medium or small size in the U.S., where there is no other transit mode, such as light rail and subway. For simplicity, as mentioned, only travel times and costs are considered in the mode split model.

As far as discrete choice modeling is concerned, both ordered and unordered response logit models can be used. However, unordered response logit models are regarded as the most popular due to their flexibility and characteristics. Therefore, this research uses this model form. The following section introduces the utility and disutility functions and discusses the model structure and underlying characteristics of the multinomial logit (MNL) and nested logit model (NL).

8.3.2.1 Utility and Disutility Functions

Generally speaking, a utility function measures the degree of satisfaction that people derive from their mode choice. A disutility function represents the generalized cost (i.e., impedance) that is associated with each mode choice. Usually, the characteristics of each choice, the trip purpose, and the socioeconomic status of the individual making that choice determine the magnitude of these two functions. However, to specify a utility function, one needs to select both the relevant variables and the particular functional form relating the selected variables.

As a result, U_i, the utility (and disutility) function derived from choice *i* can be typically expressed as the linear weighted sum of the independent variables (i.e., the attributes specific to choice *i*):

$$U_i = \beta_i^{'} x_i + \varepsilon_i$$

where U_i = the utility derived from mode choice *i*;

x_i = a vector of the exogenous variables that are specific to mode choice *i*;

β_i = a vector of parameters for mode choice *i*;

ε_i = a vector of random terms, which are usually assumed to be identically independently Gumbel distributed with variance normalized to be 1;

Since the independent variables included in this equation typically represent losses, such as travel times and costs, U_i, the utility for mode choice *i* is always negative and is in essence a disutility. Furthermore, different mode choices might have different utility functions, including different variables with different weights. But usually, the independent variables represent the level of service, cost and convenience, etc. associated with that mode. In this sense, this model is always called mode-specific. In practical applications, because it is unlikely to include all the relevant variables except travel time and costs for each mode in the utility function, it is always reasonable to attempt to capture these unobserved characteristics in the random terms.

In this research, it is assumed that the trip-makers only have two available mode choices, which are auto and bus transit. Based on this point, U_i, the utility of the mode choice *i*, can be calibrated by the following equation, which will be used later in the Multinomial Logit (MNL) or Nested Logit Model (NL) for auto and transit demand estimation:

$$U_i = \beta_{0i} + \beta_{1i} X_{1i} + \beta_{2i} X_{2i}$$

Where U_i = utility function of mode *i*;

$\beta_{0i}, \beta_{1i}, \beta_{2i}$ = weighted coefficients of respective attributes of mode *i*;

X_{1i} = total travel time of mode *i*;

X_{2i} = travel cost of mode *i*.

A utility-based mode choice model estimates the share of each mode based on the utility associated with it. The fraction of the travelers that will select a given mode is related to the competing modal utilities. Their relationship has been investigated in various forms called discrete choice models, the most popular of which are the logit models including the MNL and NL. The following sections discuss these two models in detail.

8.3.2.2 Multinomial Logit Model (MNL)

As defined by Khisty and Lall (1998), the logit formulation is a share model that divides travelers between the various modes depending on each mode's relative desirability for any given trip. If one mode is faster, cheaper, or has more favorable features than other competitive modes, this mode would be relatively more desirable. As a result, more people will select it than others. The better a mode, the more utility it has for the potential travelers.

For a specific demand pair, the auto travel time and cost are based upon the shortest path. When only 1 transit route is found for a specific demand pair, the MNL structure introduced here is then collapsed to the binary logit model. However, the essence of the mode choice process remains unchanged so that this equation still can be used except that only two choices are considered. However, when two or more than two transit routes are found for a specific demand pair, the MNL structure might not be appropriate for describing the mode split process. The reason is discussed in detail in the following section. Figure 8.4 presents a multinomial logit model structure for auto and multiple bus transit route choices that can be applied to the BTRNDP.

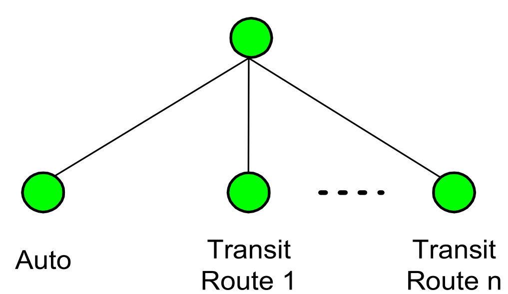


Figure 8.4 Multinomial Logit Model Structure for Auto Use and Transit Route Choices

The utility function for this MNL model structure can be expressed as follows:

$$U_{\text{Auto}} = V_{\text{Auto}} + \varepsilon_{\text{Auto}}$$
$$U_{\text{Transit\_route\_1}} = V_{\text{Transit\_route\_1}} + \varepsilon_{\text{Transit\_route\_1}}$$
$$\vdots$$
$$U_{\text{Transit\_route\_n}} = V_{\text{Transit\_route\_n}} + \varepsilon_{\text{Transit\_route\_n}}$$

As a result, P_i, the probability of travelers using mode choice *i*, is given by:

$$P_i = \frac{e^{U_i}}{\sum_{k=1}^{n} e^{U_k}}$$

where U_i = utility of mode *i*

U_k = utility of mode *k*

n = total number of modes in consideration

If there is only one transit route available, then the MNL collapses to a binary logit model. Figure 8.5 shows the logit curve for the binary choice model under two available choices of mode: transit and auto.

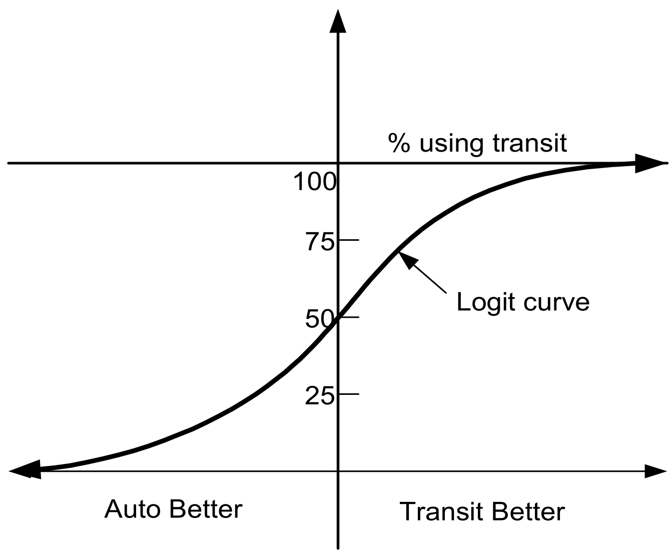


Figure 8.5 Logit Curve (FHWA/UTMA, 1977)

As mentioned before, it is generally accepted that the BTRNDP is usually investigated before the existence of rail transit, however, one might need to improve, redesign or restructure the bus transit network where rail already operates. For simplicity, this research assumes that the BTRNDP is investigated in the former case. Note that only minor modifications for the proposed model structure (as will be described later in Chapter 9) are needed to accommodate the BTRNDP in the second case.

The Multinomial Logit Model (MNL) structure has been widely used in both urban and intercity mode choice models primarily due to its simple mathematical form, ease of estimation and interpretation, and the ability to add or remove choice alternatives. However, the MNL model has also been widely criticized for its IIA (independence of irrelevant alternatives) property. The IIA property of the MNL restricts the ratio of choice probabilities for any pair of alternatives to be independent of the existence and characteristics of other alternatives in the choice set. This restriction implies that introduction of a new mode or improvements to any existing mode will affect all other modes proportionately. That is, the new or improved mode will reduce the probability of existing modes in proportion to their probabilities before the change (see Chapter 7, Bhat, 2003).

The IIA property is a major limitation of the MNL model as it implies equal competition between all pairs of alternatives, an inappropriate assumption in many choice situations. If this model is applied to the BTRNDP among drive alone and several candidate transit routes, the results could be misleading because these bus transit route alternatives are likely to be more similar to each other than they are to the auto mode due to shared attributes which are not included in the measured portion of the utility function. For example, buses may have the same fare structure and operating policies, the same lack of privacy, control of the environment, and so on. Such similarities, if not included in the measured portion of the utility function, lead to correlations between the error terms associated with these alternatives, a violation of the assumptions which underlie the derivation of the MNL.

The way in which this undesirable characteristic of the IIA property manifests itself can be illustrated using this example. Suppose that there are three modes which consist of auto, bus transit route 1 and route 2 and assume that the choice probabilities for these alternatives (for an individual or a homogeneous group of individuals) are 70%, 20% and 10% for auto, bus transit route 1 and bus transit route 2 respectively. If the service on bus transit route 2 were to be improved in such a way as to increase its choice probability to 15%, the MNL model would predict that the shares of the other alternatives would decrease proportionately as shown in Table 8.1, maintaining the probability ratios between the auto mode and bus transit route 1 alternative. As a result, the MNL model predicts that most of the increased ridership in bus transit route 2 mode comes from auto mode. This is inconsistent with the expectation and empirical evidence that most of the new riders in bus transit route 2 will be diverted from those in bus transit route 1. This inconsistency is a direct result of the IIA property of the MNL model. Thus, in these types of choice situations, the MNL model will yield incorrect predictions of diversions from existing modes.

Table 8.1 Illustration of the IIA Property on Predicted Choice Probabilities

Alternatives	**Choice Probability Before Improvements to Bus Transit Route 2**	**Choice Probability After Improvements to Bus Transit Route 2**	**Change in Choice Probabilities**
Auto	0.700	0.661	0.039
Bus Transit Route 1	0.200	0.189	0.011
Bus Transit Route 2	0.100	0.150	0.050

As can be seen, the limitation of the MNL model results from the assumption of the independent distribution of error terms in the utility of the alternatives that is used to derive the MNL model. Different models can be derived through the use of different assumptions concerning the structure of the error distributions of alternative utilities. Among them, the Nested Logit (NL) model is the simplest and most widely used. The NL model represents important deviations from the IIA property but retains most of the computational advantages of the MNL model (see Bhat, 2003). The NL model is characterized by grouping (or nesting) subsets of alternatives that are more similar to each other with respect to excluded characteristics than they are to other alternatives.

Alternatives in a common nest exhibit a higher degree of similarity and competitiveness than alternatives in different nests. This level of competitiveness, represented by cross-elasticities between pairs of alternatives (the impact of a change in one mode on the probability of another mode) is identical for all pairs of alternatives in the nest. Complex tree structures can be developed which offer substantial flexibility in representing differential competitiveness between pairs of alternatives; however, the nesting structure imposes a system of restrictions concerning relationships between pairs of alternatives as will be discussed later in this chapter.

8.3.2.3 Nested Logit Model (NL)

Different from the MNL model, the Nested Logit model is based on the assumption that some of the alternatives share common components in their random error terms. That is, the random term of the nested alternatives can be decomposed into a portion associated with each alternative and a portion associated with groups of alternatives. For example, consider a situation where a traveler has *n* bus transit routes available for making an urban trip. The utility equations for these alternatives are:

$$
\begin{aligned}
&U_{\text{Auto}} = V_{\text{Auto}} + \varepsilon_{\text{Auto}} \\
&U_{\text{Transit\_route\_1}} = V_{\text{Transit}} + V_{\text{Transit\_route\_1}} + \varepsilon_{\text{Transit}} + \varepsilon_{\text{Transit\_route\_1}} \\
&\vdots \\
&U_{\text{Transit\_route\_n}} = V_{\text{Transit}} + V_{\text{Transit\_route\_n}} + \varepsilon_{\text{Transit}} + \varepsilon_{\text{Transit\_route\_n}}
\end{aligned}
$$

The utility terms for transit route 1 to transit route *n* each include a distinct observed component, $V_{\text{Transit\_route\_1}}$ and $V_{\text{Transit\_route\_n}}$, and a common observed component, V_{Transit}; they also include distinct random components, $\varepsilon_{\text{Transit\_route\_1}}$ and $\varepsilon_{\text{Transit\_route\_n}}$, and a common random component, $\varepsilon_{\text{Transit}}$. The common error component creates a covariance relationship between the total error components for transit route 1, $\varepsilon_{\text{Transit}} + \varepsilon_{\text{Transit\_route\_1}}$, and transit route *n*, $\varepsilon_{\text{Transit}} + \varepsilon_{\text{Transit\_route\_n}}$. This covariance, which violates the assumption underlying the MNL model, represents an increased similarity between pairs of nested alternatives (including all the transit routes) and leads to greater similarity and cross-elasticity between these alternatives. Note that the total error for each of these *n* alternatives is assumed to be Gumbel distributed with variance parameter equal to one, as in the MNL model. Figure 8.6 presents a graphical representation for the theoretical nested logit model structure for the auto and bus transit route choices.

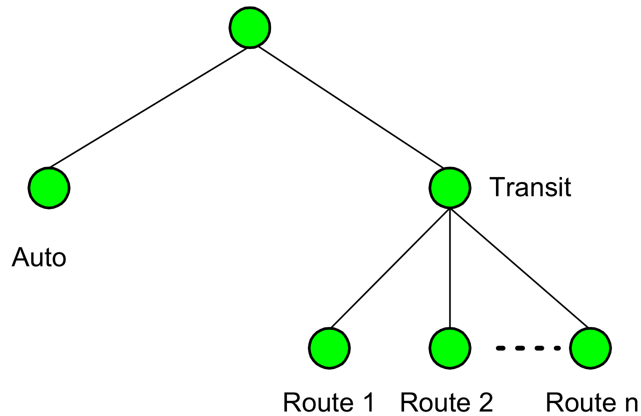


Figure 8.6 Theoretical Nested Logit Model Structure for Auto and Transit Route Choices

The disadvantage of MNL is as follows: Compared to the nested logit model, the multinomial logit model tends to overestimate the transit demand given the fixed total travel demand because it usually draws more auto demand due to the IIA property inherent in MNL model structure. The advantage of MNL is also obvious: it can be easily used to estimate bus transit demand.

The advantage of NL is also clear: It reflects the realistic situation and therefore probably is more reasonable for mode choice estimation. Especially when empirical results are available, the quality of the results using the nested logit model can be guaranteed. The disadvantage of NL is as follows: Practically, it is difficult to get empirical results because usually just one transit route is considered in the case alternatives. Also, it is harder to use and is more complex compared to the multinomial logit model. Table 8.2 presents the characteristics comparisons between the MNL and Nested Logit Model.

Table 8.2 Characteristics Comparisons between the MNL and Nested Logit Model

Features	**Model Structures**	
	Multinomial Logit Model	**Nested Logit Model**
Advantages	Easy to estimate the demand	Reflect the realistic situation
Disadvantages	Tend to overestimate the transit demand	More complex a model Hard to get the empirical results

8.3.2.4 BLM-IPM Model

As discussed in the preceding two sections, both the Multinomial Logit (MNL) and Nested Logit (NL) Model structure can be used when computing the transit demand fraction of total travel demand. Furthermore, when two or more feasible transit routes are available, each transit route can be regarded as one choice alternative competing with other transit routes and the car mode. Examples of MNL applications can be seen in Lee and Vuchic (2000). However, as already presented, the MNL model is not appropriate for the BTRNDP due to its IIA (independence of irrelevant alternatives) property and the NL model cannot work very well for the BTRNDP because of the unavailability of empirical results for model calibration. Especially, as the number of feasible transit routes increases, the quality of results from both models becomes worse and therefore the solution qualities cannot be guaranteed if one uses either of these two models.

In this research, a two-stage BLM-IPM model is proposed to overcome the disadvantages of these two models for the mode choice between auto and transit routes. The first stage uses the binary logit model (BLM) to compute the individual probability of choosing the transit mode using current empirical results and the utility derived from the driving shortest path and the shortest transit path (transfer penalty and travel cost can be added if necessary). The aggregate share of the transit demand can be computed by multiplying individual choice probability by the total demand. The second stage of this model uses an "inversely proportional model" (IPM) to assign transit demand to the competing transit routes that are inversely proportional to the total travel time. Note that this model implicitly assumes that transit users have made the decision to take transit before they decide which transit route to choose if more than one transit route is available. It is expected that this BLM-IPM model will work much better than the MNL and therefore it is used for the mode choice model in this research.

The following sections combine the above conclusions and apply them to the BTRNDP with variable transit demand. The details are presented as follows.

8.4 The NAP for the TRNDP with Variable Transit Demand

Basically speaking, the NAP for the BTRNDP with variable transit demand is a bus transit network evaluation tool with the ability to decide the transit trip demand between each centroid node pair, assign the transit trips to each route on the proposed solution network and determine associated route frequencies. To accomplish these tasks for the BTRNDP with variable transit demand, NAP employs an iterative procedure that seeks to achieve internal consistency of the transit trip demand and the route frequencies. Furthermore, the iterative procedure in the NAP contains three major components, namely, a transit demand equilibration procedure, a transit trip assignment procedure and a frequency setting procedure.

Once the overall candidate solution route set is generated by the ICRSGP and a specific set of routes is proposed by the HSP, the NAP is called to evaluate the alternative network structure and to determine transit demand and route service frequencies. The whole process in the NAP can be described as follows. First, an initial set of route frequencies are necessarily specified at the beginning of the trip assignment process. Based on these data, transit demands are set tentatively according to the mode choice model or one can assign initial transit demand directly by multiplying the total demand by a specific percentage (say 30%) at the same time when one assigns the initial route frequencies. Second, the trip assignment model is utilized to assign the transit trip demand matrix to the set of routes associated with a particular network configuration. Third, the frequency setting procedure is called to determine the service frequencies on each route in the current proposed solution network. Fourth, the transit demand is recomputed using the mode choice model. These are used as the input frequency and input transit demand for the next iteration in the frequency setting procedure and transit demand equilibration procedure respectively. If the bus transit demand matrix and these frequencies are considered to be different from the previous input frequencies by a user-defined parameter (say 5%), the process iterates until the internal consistency of the transit demand matrix and route frequencies is achieved. If convergences are achieved both in the transit demand matrix and route frequencies, the transit demands and the route frequency are then determined. Meanwhile, many other performance measures are also obtained.

Figure 8.7 shows the flow chart of the NAP for the BTRNDP with variable transit demand. It should be noted that the trip assignment process considers each zone (centroid node) pair separately. Also, the transit trip assignment model developed here for the BTRNDP with variable transit demand adapts the same four-level transit assignment process as that presented in Chapter 6 for the BTRNDP with fixed transit demand. As already described, this model considers the number of transfers and/or the number of long walks to the bus station as the most important criterion. (Note that long walk paths refer to those paths that include a walk to a bus stop from a trip origin not directly served by transit.) It first checks the existence of the 0-transfer-0-longwalk paths. If any path of this category is found, then the transit demand between this centroid node pair can be provided with direct route service and the demand is therefore distributed to these routes. If not, the existence of paths of the second category, i.e., 0-transfer-1-longwalk path and 1-transfer-0-longwalk paths are checked. If none of these paths is found, the proposed procedure will continue to search for paths of the third category, i.e., paths with 2-transfer-0-long-walk, 1-transfer-1-long-walk and/or 0-transfer-2-longwalk. Only if no paths that belong to these three categories exist, there would be no paths in the current transit route system that can provide service for this specific centroid node pair (i.e., these demands are unsatisfied). Note that at any level of the above three steps, if more than one path exists, a "travel time filter" is introduced for checking the travel time on the set of competing paths obtained at that level. If one or more alternative paths whose travel time is within a particular range pass the screening process, an analytical nonlinear model (i.e., the inversely proportional model) is used to assign the transit trips to the competing transit routes between that centoid node pair that are inversely proportional to the total travel time. In addition, policy headway and the demand headway are used together to determine the frequencies on each route in the frequency setting procedure. The whole process is repeated until all the travel demand pairs in the studied network are traversed. Related details of the transit trip assignment model can be seen from the work by Fan and Machemehl (2003).

Note that NAP for the BTRNDP with variable demand is essentially the same as that with fixed demand except that the former adds one more procedure, namely, the transit demand determination procedure. Once internal consistency of both the route frequencies and transit demand equilibration are achieved, the optimal transit route network can be developed and a variety of computational performance measures is obtained accordingly.

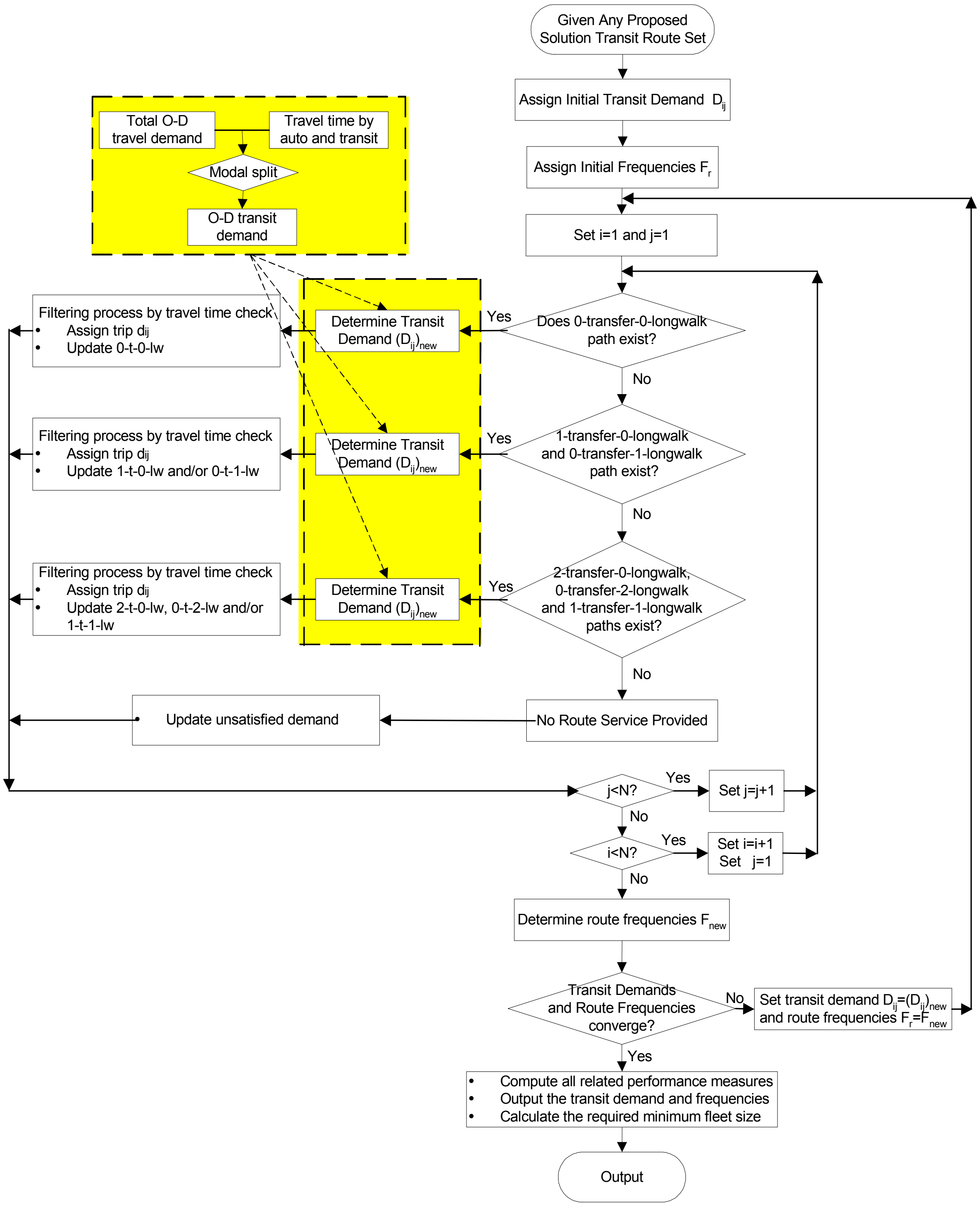


Figure 8.7 Network Analysis Procedure (NAP) for the BTRNDP with Variable Transit Demand

8.5 Solution Framework for the TRNDP with Variable Transit Demand

As mentioned before, the solution methodology for the BTRNDP with variable demand does not differ significantly from that for the BTRNDP with fixed demand. The solution framework for the BTRNDP with variable transit demand consists of three parts: an Initial Candidate Route Set Generation Procedure (ICRSGP) that generates all feasible routes incorporating practical industry guidelines; a Network Analysis Procedure (NAP) that evaluates the proposed route network, computes performance measures and finds the optimal transit network; and a Heuristic Search Procedure (HSP) that guides the search techniques. For the first part, the whole ICRSGP procedure for the BTRNDP with variable demand is the same as that for the BTRNDP with fixed demand as described in chapter 5. The NAP for the BTRNDP with variable demand evaluates and analyzes the input bus transit network and determines the transit demand matrix and route frequencies. It uses the same methodology as that employed in the BTRNDP with fixed demand and adds some additional procedures to tackle the characteristics of variable transit demand. The details have been discussed in previous sections.

The six previously proposed solution techniques, including the genetic algorithm, local search, simulated annealing, random search, tabu search and exhaustive search methods are all used for the BTRNDP with variable transit demand. As pointed out, the difference between the solution frameworks of each of these six algorithms for the BTRNDP with variable transit demand and that with fixed transit demand only lies in the NAP part. But once the NAP with fixed demand is replaced by the NAP with variable transit demand, each of the solution frameworks proposed for the BTRNDP with fixed transit demand can work for the BTRNDP with variable transit demand. As a typical illustration, a genetic algorithm based solution framework for the BTRNDP with variable demand is shown in Figure 8.8. For the other five proposed solution techniques, similar replacement work can be done.

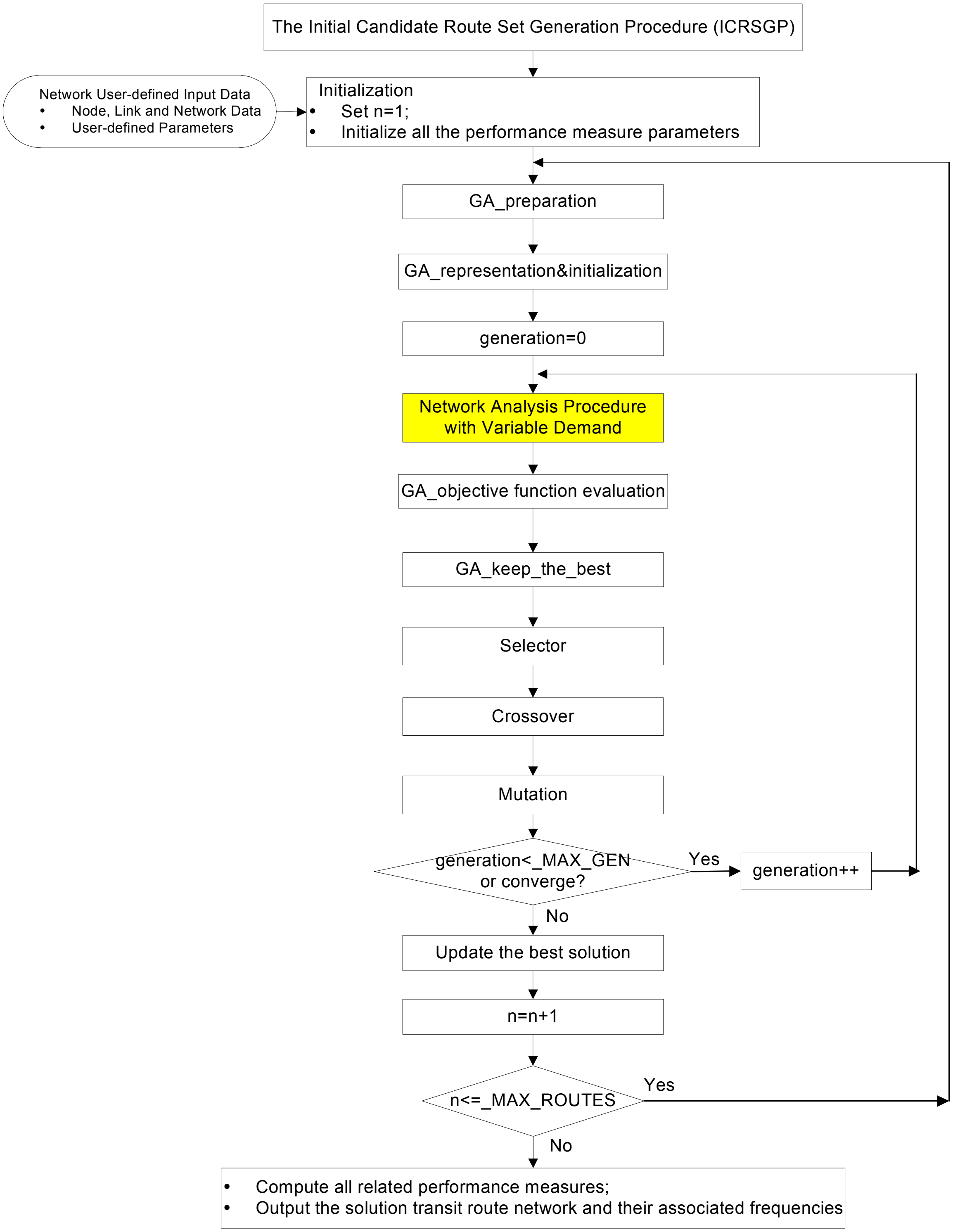


Figure 8.8 A Genetic Algorithm-Based Solution Framework for the BTRNDP with Variable Transit Demand

8.6 Summary

This chapter focuses on the BTRNDP with variable demand. The underlying characteristics of both variable total demand and variable transit demand are presented. The mode split procedure including the attributes of alternatives and the setting decision rule for mode choice between the auto and transit modes are discussed. Utility and disutility functions are introduced and the commonly used MNL and NL model are also reviewed. The underlying characteristics of the MNL and NL are discussed and advantages and disadvantages of each model are pointed out. A two-stage BLM-IPM model (binary logit model-inversely proportional model) for determining the mode choice between auto and transit routes is presented for the BTRNDP with variable transit demand. It is also pointed out that the ICRSGP for the BTRNDP with variable demand is the same as that proposed in Chapter 6 for the BTRNDP with fixed demand. The NAP for the BTRNDP with variable transit demand, which builds on the BTRNDP with fixed transit demand, is described. Three major components of the NAP, namely, the transit demand equilibration procedure, transit trip assignment procedure and the frequency setting procedure are presented. The solution framework for the BTRNDP with variable transit demand is presented. Implementation methods and numerical results based on comprehensive experimental networks are discussed in the next chapter.

CHAPTER NINE
COMPREHENSIVE EXPERIMENTS AND NUMERICAL RESULTS

9.1 Introduction

As described, the objective of the BTRNDP is to develop an optimal transit route network that meets user-defined requirements subject to limited economic and operational resources. Solution methodologies for the BTRNDP both with fixed and variable transit demand are presented in previous chapters. For the BTRNDP with fixed demand, the solution framework contains three major components, namely, an Initial Candidate Route Set Generation Procedure (ICRSGP) that generates all feasible routes incorporating practical bus transit industry guidelines; a Network Analysis Procedure (NAP) that assigns transit trips, determines service frequencies and computes performance measures; and a Heuristic Search Procedure (HSP) that guides the search techniques and updates the proposed solution network. For the BTRNDP with variable transit demand, similar solution frameworks are used except that the transit demand equilibration procedure is added and included in the NAP to accommodate the variable transit demand characteristics. Moreover, for the BTRNDP under both scenarios, five heuristic algorithms, including the genetic algorithm, local search, simulated annealing, random search and tabu search algorithms, along with an exhaustive search algorithm as a benchmark for the BTRNDP with small network, are employed as the solution methods for finding an optimum set of routes from the huge solution space.

This chapter focuses on the algorithm implementation and numerical results for these proposed solution methodologies. Details of the C++ program codes that are developed to implement the above-mentioned six algorithms for the BTRNDP both with fixed and variable transit demand are presented. Comprehensive experimental networks are designed and successfully tested for the BTRNDP. Numerical results are presented and related characteristics underlying the BTRNDP are identified.

This chapter is organized as follows. Section 9.2 describes the example networks that are used for illustrating the proposed solution methodologies. Section 9.3 presents the network representation and algorithm implementation details of the C++ programming codes for the proposed methodology for solving the BTRNDP. Sections 9.4 and 9.5 discuss the comprehensive numerical results of the six proposed solution methods, including the genetic algorithm, local search, simulated annealing, random search, tabu search algorithm and exhaustive search method with fixed and variable transit demand respectively using the designed example networks. Sensitivity analyses for each algorithm are conducted and algorithms are compared based on the multi-objective decision making nature of the BTRNDP. Section 9.6 identifies the characteristics underlying the BTRNDP. Effects of the route set size on the objective function and its components are examined and compared for the BTRNDP both with fixed and variable transit demand. Section 9.7 investigates the large network extensions, in which effects of network size and demand aggregations on the computation speed and solution quality are studied. Section 9.8 describes the redesign of the existing transit network issues and related numerical results are also presented. Finally in section 9.9, a summary concludes this chapter.

9.2 Example Network Configuration

To test the feasibility of the proposed solution methodologies in previous chapters and examine and compare the quality of the solutions derived from these methodologies, three experimental networks are designed, which are shown in Figures 9.1, 9.2 and 9.3 respectively. Although the size of any of these three networks is relatively smaller compared to the real world networks, they can be used as a pilot study to get a "feel" about how these algorithms work for the BTRNDP and hopefully generate or provide some guidelines for future applications of these algorithms on large networks. For simplicity, they are named "small", "medium" and "large" network respectively according to their relative size. In these example networks, the centroid node point of each traffic demand zone, the road intersection nodes and current road network structure are also given. For example, the small network contains 7 travel demand zones and 15 road intersections.

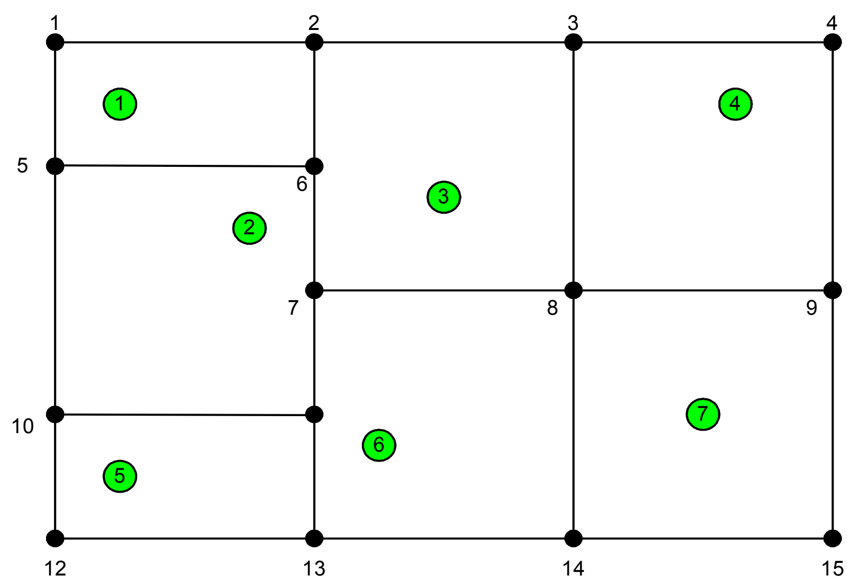


Figure 9.1 A Small Example Network for Case Study

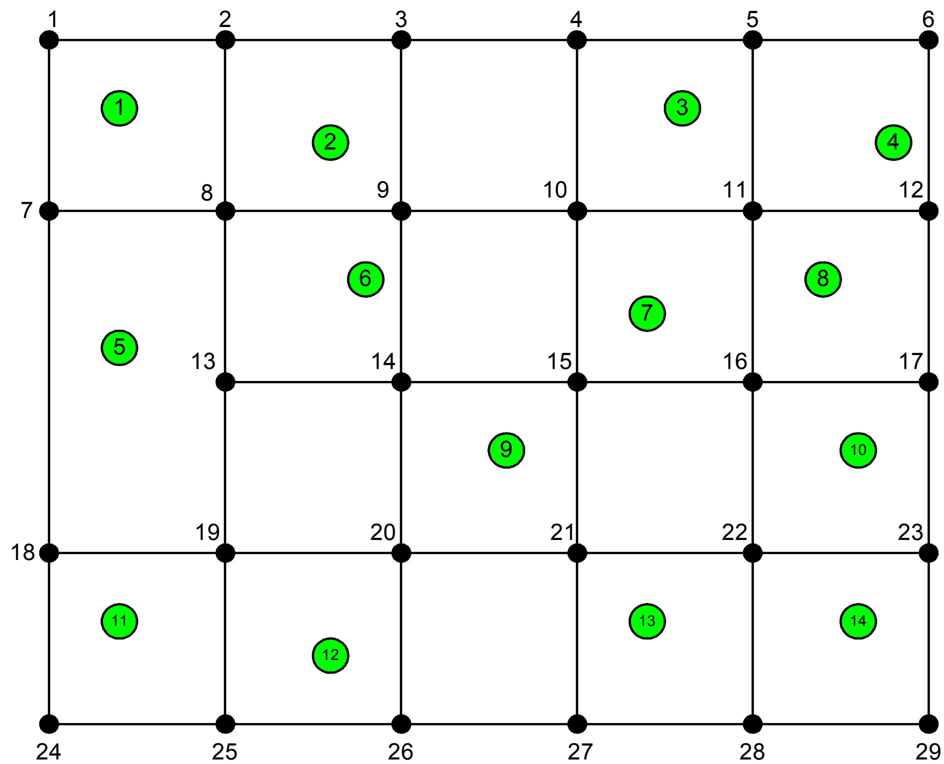


Figure 9.2 A Medium Example Network for Case Study

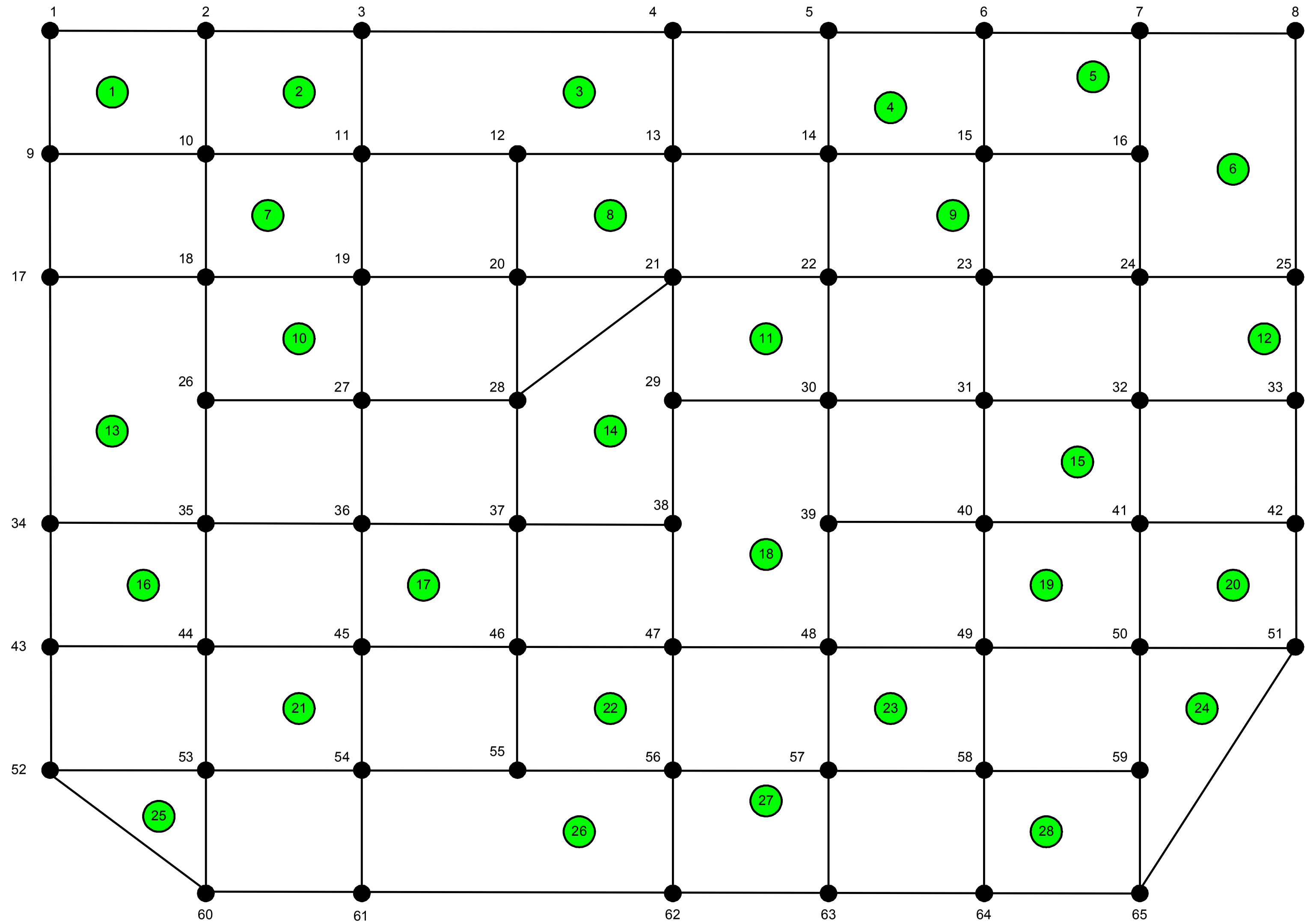


Figure 9.3 A Large Example Network for Case Study

Note that the node, link and network information are usually known beforehand and they can be inputted directly. To implement and compare the proposed solution methodologies, these three example networks are preprocessed so that all network information such as the link connectivity, link distances, node locations, transit origin-destination trip demands (fixed demand) and total origin-destination trip demands (variable transit demand) is obtained. However, the demand distribution nodes are generally user-defined and therefore have to be carefully specified by the users (i.e., transit planners). As mentioned before, different user-defined input files can result in different optimal solution networks. In this research, the network is preprocessed as follows: 1) the zonal demands are distributed the same way as with highway network demand; and 2) if the same road link contains two or more demand distribution nodes from different zones, these distribution nodes are aggregated as one node at their middle point. For example, after this preliminary process, 20 centroid distribution nodes, 35 nodes, and 82 arcs are obtained in the small example network as shown in Figure 9.4. The minimum and maximum route lengths are defined. As mentioned in Chapter 5, in the first example phase, the ICRSGP generates 286 feasible routes whose distances satisfy these two route length constraints. The medium and large network example networks, after preliminary processing, are shown in Figures 9.5 and 9.6.

Required information such as the XY coordinates of each node, zone, network and transit or/total travel demand from each zone to the other zones for these three example networks are presented in the APPENDIX. For a particular city network, relevant and necessary data must be specified in four files: zone.dat, network.dat, xy.dat and demand.dat. The input formats are designed for these three networks (especially using the small network as an input illustration example) and included in the APPENDIX.

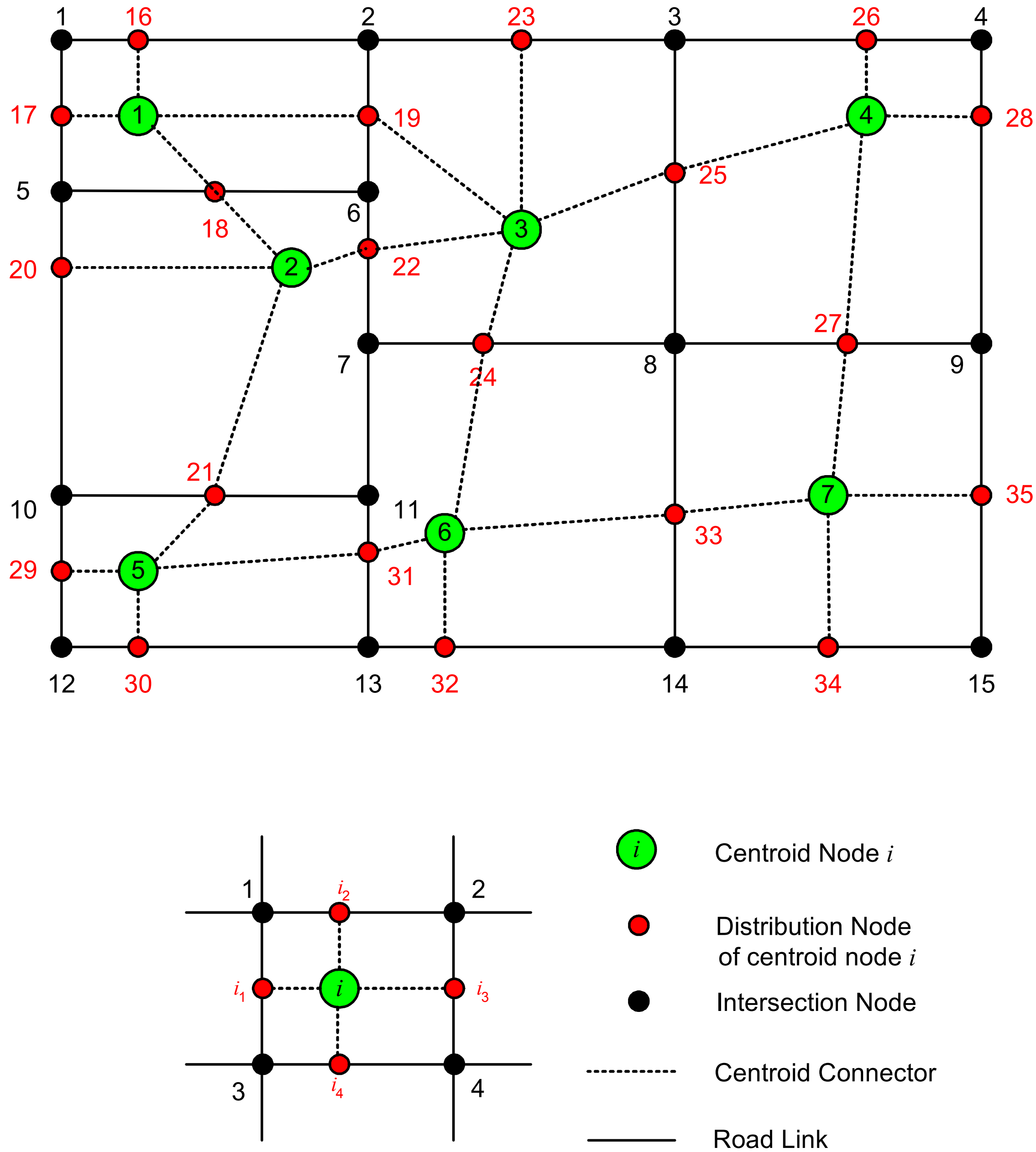


Figure 9.4 The Preprocessed Small Example Network for Case Study

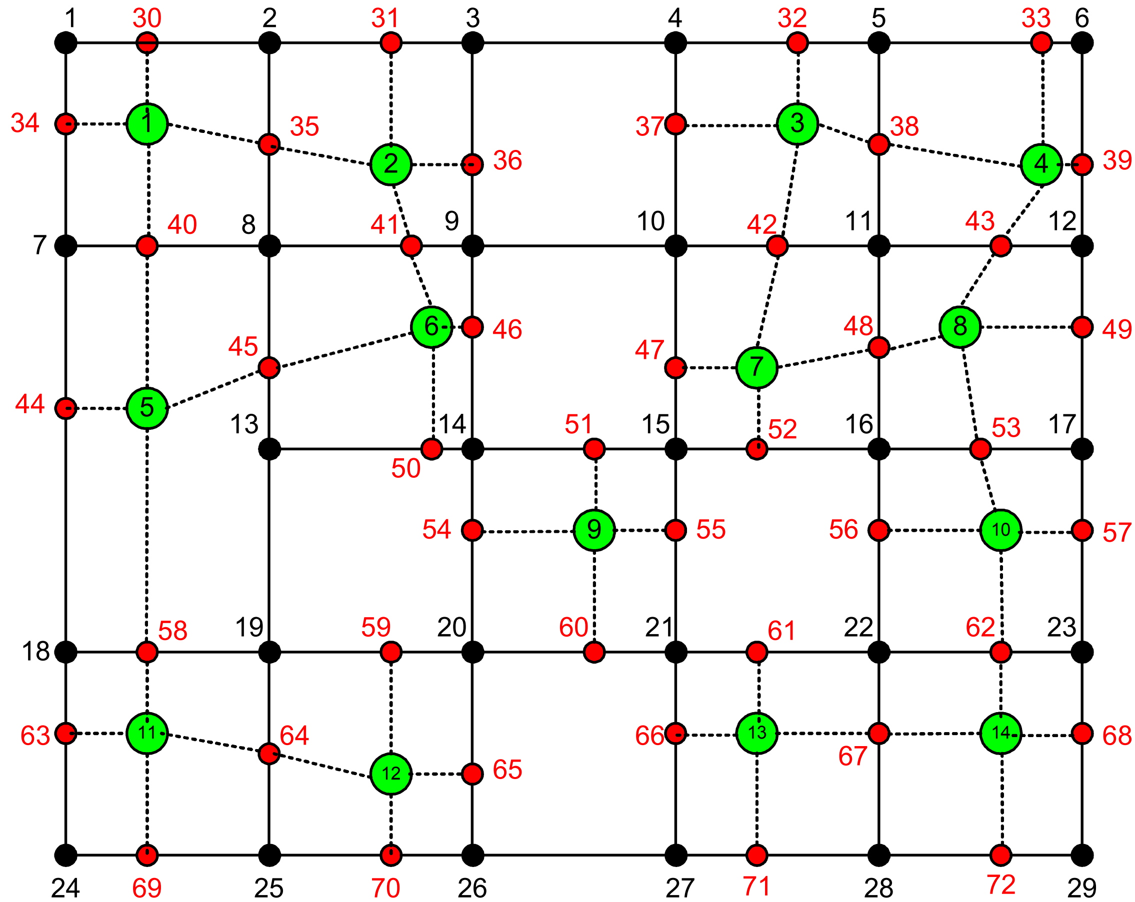


Figure 9.5 The Preprocessed Medium Example Network for Case Study

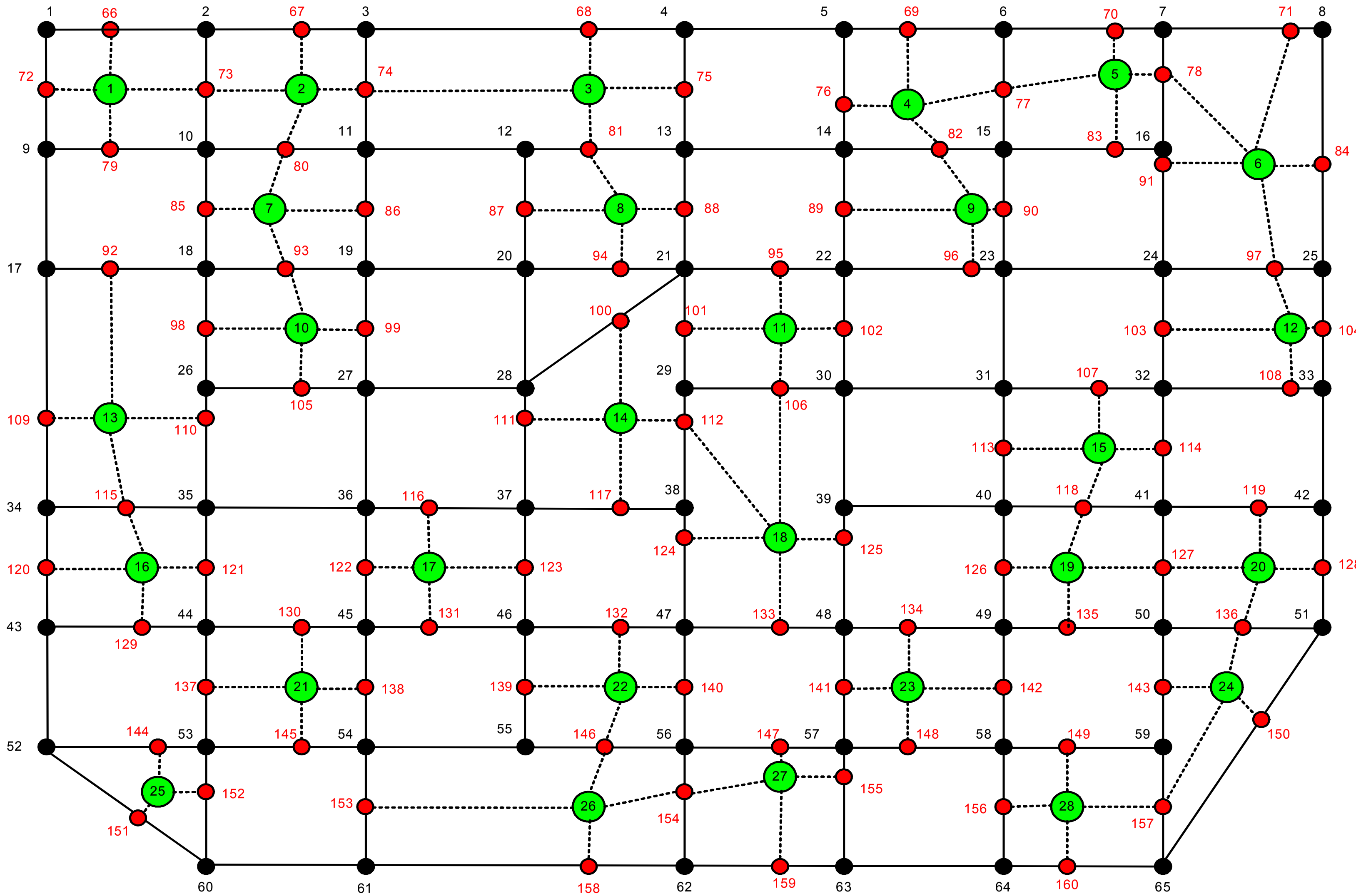


Figure 9.6 The Preprocessed Large Example Network for Case Study

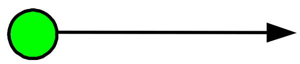
9.3 Computer Implementations

9.3.1 Network Representations and Data Structures:

The fundamental network data provided in four input files, (zone.dat, network.dat, xy.dat, and demand.dat) include all the information for zone, node, link and network data for the current studied network.

Given the above information, the data structures are developed to organize these data so that the C++ program can be used to design the optimal transit route network. To facilitate the implementation, four classes are defined in the C++ programs:

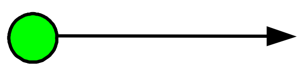
- Node:



NodeID
ZoneID
X,Y Coordinates
Print_Node_Information

A special flag value represented by ZoneID is set to distinguish centroid nodes of zones from intersection and distribution nodes. Another special flag value is used to distinguish intersection nodes from distribution nodes (bus stops). Note that all this information is generally user-defined and for the example is presented in the input file network.dat, as shown in the APPENDIX.

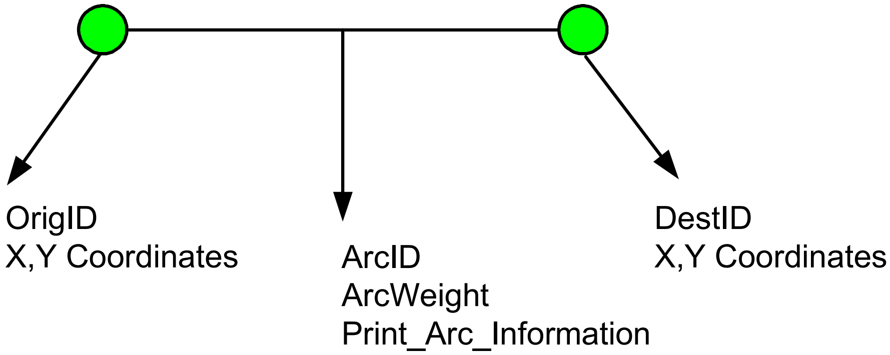
- Zone:



ZoneID
NumDistNodes
NumLongWalkNodes
LongWalkNodes
X,Y Coordinates
Print_Zone_Information

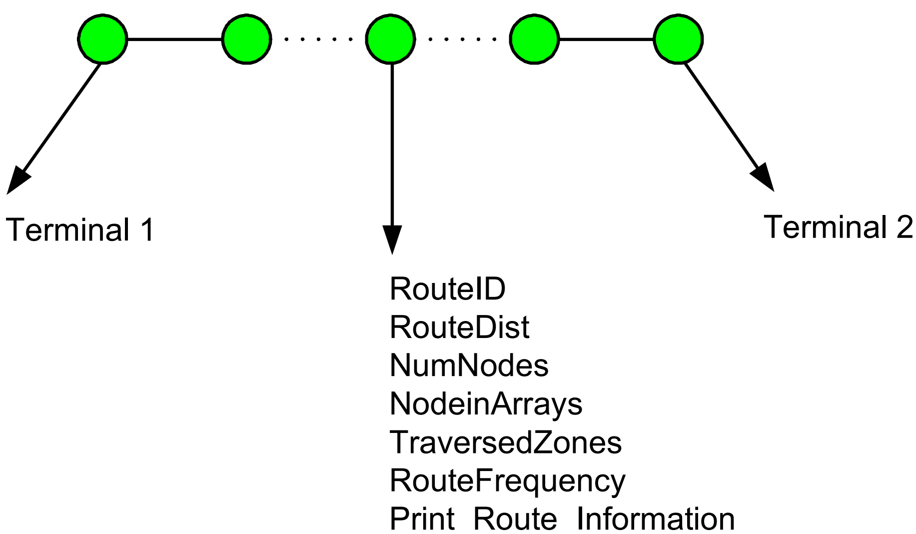
A specific numeric ID, and X and Y coordinates are set for each centroid node. The number of distribution nodes associated with this centroid is also recorded and these distribution node IDs are stored. The total numbers of intersection nodes or distribution nodes in the network that can be reached within a user-defined maximum walking distance from this centroid node are recorded. Related zone information can be printed whenever needed. This user-defined information is presented in the input file zone.dat, as shown in the APPENDIX.

- Arc:



A specific numeric ArcID is set for each arc (link) in the studied network. The Forward Star Representation is the employed data structure in this research to efficiently determine the set of arcs outgoing from any node. The origin node and destination node of this arc are represented by corresponding node IDs respectively and their X and Y coordinates are also read from the node class information and stored for consistency. Based on the stored X and Y coordinates of the origin and destination nodes, the weight (distance) of this arc is therefore computed and stored in ArcWeight. Note that fundamental arc information in the whole network is also user-defined and can be read from the input file zone.dat, as shown in the APPENDIX.

- Route:



As before, a numerical route ID is set for any generated route. At the same time, two terminals are recorded and the route distances are stored. The total number of nodes contained in any route and all nodes along the route are stored in dynamically allocated arrays. The zones that can be accessed by the route are also recorded. Also, the route frequency on this route is computed and stored in RouteFrequency parameter. Note that the route information is generated by the proposed solution methodology using shortest path and k-shortest path procedures rather than obtained from the user-defined input file.

- Shortest Path:

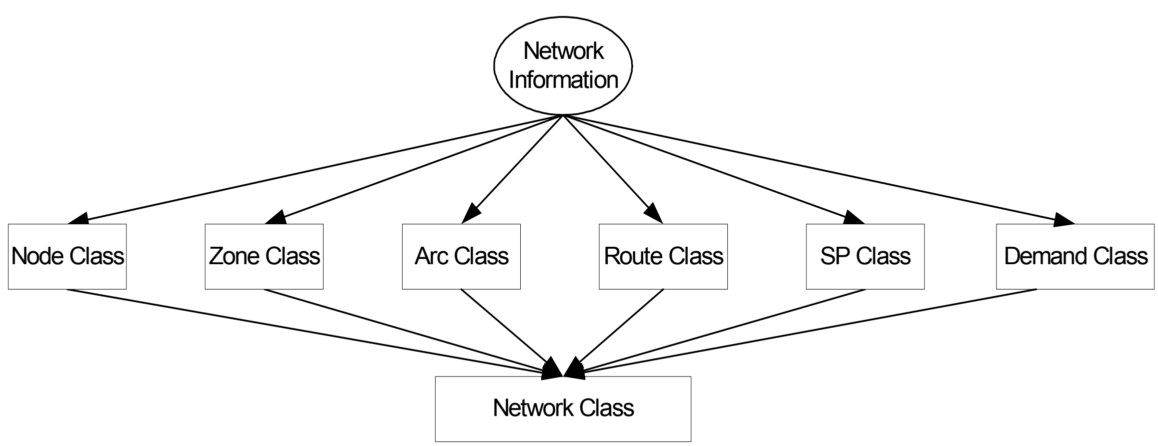
Generally speaking, the data structures used in the Dijkstra's shortest path and modified Yen's k-shortest path algorithm follow the commonly used ones, which contain information such as NodeStatus, Previous_Node, Distance and Print_Sp as shown in Chapter 5. Once these shortest paths that meet the route length requirements are generated, they are stored.

- Demand:

The demand data are stored in an O-D trip demand matrix for the studied network. Put another way, the demand data includes the origin and destination zone information and the trip demands between any O-D pair. Obviously, this information is known before one designs the optimal transit route network and is stored in the user-defined input file demand.dat.

- Network:

Network class files are used to organize the data from the above classes and treat them either as private numbers or as public member functions. The Forward Start Representation data structure is used to organize all network information. For simplicity of description, the relationships among these defined classes can be simply presented as the following graph.



Note that computer memories are dynamically allocated in network class for several sub-classes such as Zone, Node, Link, Route, Shortest Path and Demand classes. Furthermore, the functions that are developed to implement the solution algorithms have been successfully tested and numerical results are presented in this chapter.

9.3.2 Preset User Defined Parameters

As described before, some parameters are required when one designs the optimal transit route network. These data can be defined through the input file or the list of parameters in the network file by the transit network designers. Generally speaking, the parameters for the minimum and maximum route lengths are needed for the Initial Candidate Route Set Generation Procedure (ICRSGP), which uses Dijkstra's label-setting shortest path and modified Yen's K-th shortest path algorithm to generate all feasible paths. The Network Analysis Procedure (NAP), which contains the transit trip assignment procedure and frequency setting procedure, requires several parameters such as the bus travel speed, the load factor, the bus capacity transfer penalty and the minimum and maximum headways.

In this research, the penalty for each transfer is chosen to be 5 minutes of equivalent in-vehicle travel time, the bus seating capacity is 40 passengers, the maximum load factor is taken as 1.3, the minimum headway is set as 5 minutes and the maximum headway is chosen as 60 minutes, the transit vehicle and car speed are chosen as 25 and 40 mile per hour respectively. Other design-related parameters include different weights that the transit operators might put on different components of the transit user time. Note that the traffic delay at each intersection and the passenger boarding and deboarding time are not considered although one should do so for a real-world BTRNDP application. However, only very minor modifications are needed for the computer programs to accommodate them. In addition, as mentioned before, for each heuristic algorithm, different parameters can be chosen that can result in different solution network performance. Sensitivity analyses that reflect changes in the objective function performance corresponding to different values for these parameters are conducted and presented in the following sections.

Using the example networks, several C++ computer programs were made to implement the Initial Candidate Route Set Generation Procedure (ICRSGP). Dijkstra's label-setting shortest path algorithm and modified Yen's k-th shortest path algorithm were successfully implemented and all feasible routes and related numerical results were written to a file called outputReport.dat. For the Network Analysis Procedure (NAP), a C++ program was successfully tested implementing the transit trip assignment model and frequency setting procedure for the BTRNDP with fixed demand as well as the transit demand determination procedure for the BTRNDP with variable transit demand. In conclusion, the proposed solution methodologies were implemented using small, medium and large example networks as shown before and have successfully solved the BTRNDP both with fixed and variable transit demand. The following sections present comprehensive numerical results for each algorithm and associated sensitivity analyses. Algorithm comparisons and characteristics underlying the BTRNDP both with fixed and variable demand are also discussed.

9.4 The TRNDP with Fixed Transit Demand

The following sections first present sensitivity analyses for the heuristic algorithms, including the GA, LS, SA, RS and TS methods. Essentially speaking, each algorithm has continuous parameters that can fall within a very large range. For example, the number of generations in each algorithm can vary from 1 to positive infinity. One cannot try all combinations for all parameters including weight set levels for each objective function component. Furthermore, even if one could find the optimal parameter set at a specific weight set level, this parameter set might not be optimal at another weight set level. Therefore, for simplicity, several discrete values are chosen for each continuous parameter in all algorithms and the optimal parameter set is decided sequentially for each algorithm at a commonly used weight set level (here 0.4, 0.4 and 0.2 is chosen for the weight of user cost, operator cost and unsatisfied demand cost respectively). The following section first presents the sensitivity analyses for each algorithm using the small network as an example and then summarizes the sensitivity analyses for both the small and medium networks.

9.4.1 Genetic Algorithm

9.4.1.1 Implementation Presetting

After all candidate routes are generated by the ICRSGP, the GA model is used to find the optimum set of routes from the set of candidate routes. The route representation, the preset parameters and the algorithm implementation skeleton for the GA model for the BTRNDP have been discussed in previous chapters. The numerical results from the sensitivity analyses are presented as follows.

9.4.1.2 Numerical Results and Sensitivity Analyses

As mentioned before, the performance of the proposed genetic algorithm model greatly depends upon the chosen population size, stopping criteria (i.e., the number of generations), crossover probability and mutation probability. The following sections present sensitivity analyses of these parameters.

9.4.1.2.1 *Effect of Population Size*

The effect of population size is examined by varying this value from 5 to 100 and the result is given in Figure 9.7(1). It can be seen from the figure that as the population size increases, the objective function value tends to decrease. It is also noted that the larger the chosen population size, the more the computation time. When the population size reaches 30, the optimal objective function is achieved, suggesting that 30 should be chosen as the optimal population size for the example network. Similar results can be found in Goldberg (1989), where a population size of 30 to 50 was recommended.

9.4.1.2.2 *Effect of Generations*

The effect of stopping criteria is investigated by choosing the number of generations ranging from 5 to 50 and the result is provided in Figure 9.7(2). As can be seen, the lowest objective function value can be achieved with 20 generations. Therefore, 20 is chosen as the optimal number of generations in terms of solution quality.

9.4.1.2.3 *Effect of Crossover Probability*

The effect of crossover probability is also studied by varying this value from 0.1 to 0.9. The result shown in Figure 9.7(3) indicates that 0.8 might be the optimal value and as a result, it is recommended.

9.4.1.2.4 *Effect of Mutation Probability*

The effect of mutation probability is examined by varying this value from 0.0001 to 0.2 and the result is presented in Figure 9.7(4). Achievement of the lowest objective function value at 0.1 suggests that the optimal mutation probability might be 0.1.

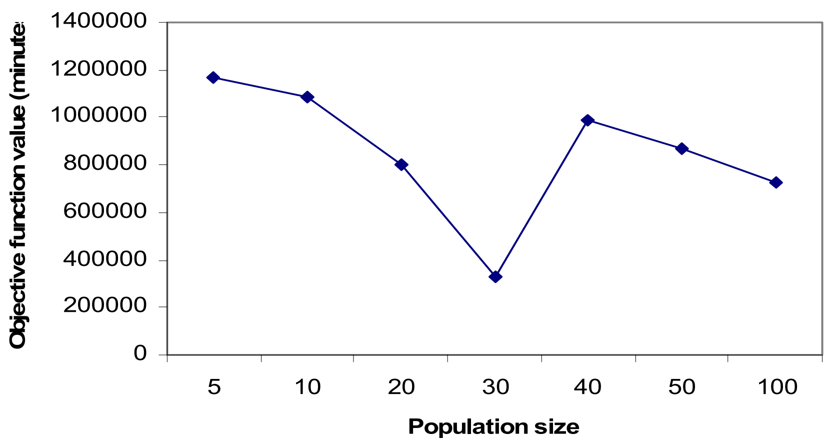


Figure 9.7(1)

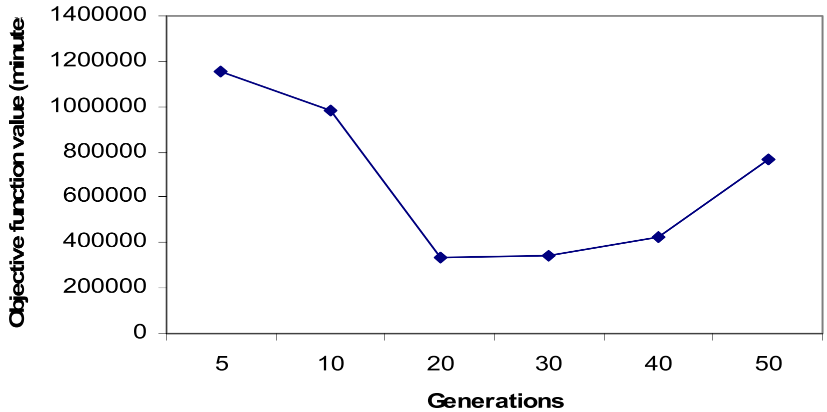


Figure 9.7(2)

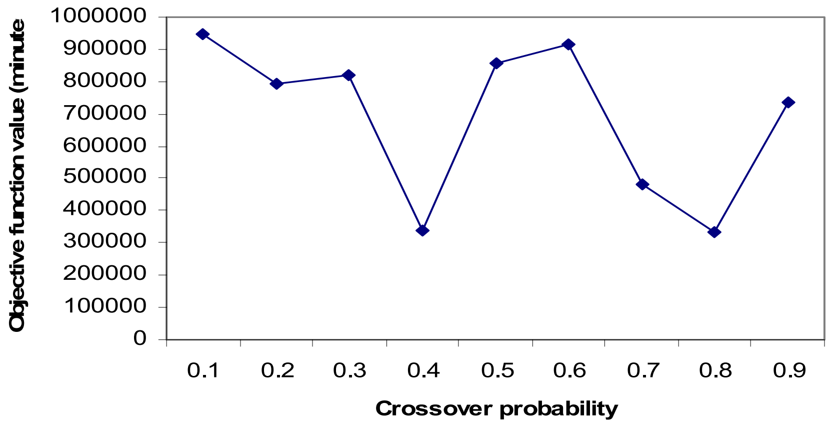


Figure 9.7(3)

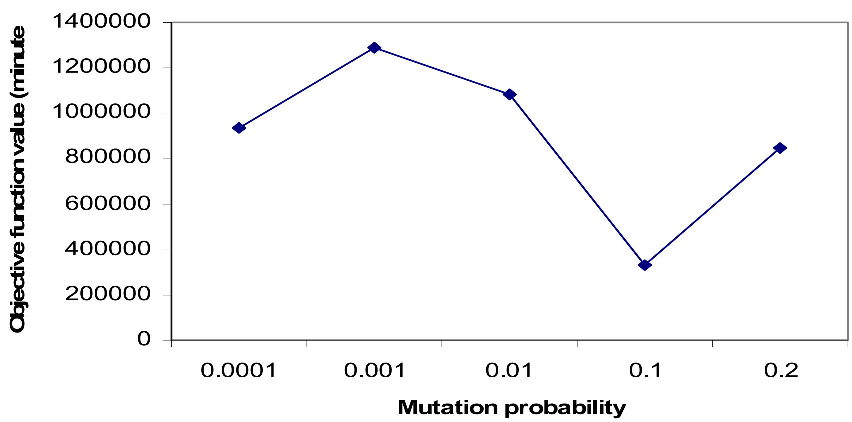


Figure 9.7(4)

Figure 9.7 Sensitivity Analyses for the Genetic Algorithm

9.4.2 Local Search

As can be seen from Chapters 4 and 7, the local search method only has one parameter, which is the number of generations. The effect of the number of generations is studied by choosing the number of generations ranging from 100 to 10000 and the result is provided in Figure 9.8. As can be seen, the lowest objective function value can be achieved with 100 generations. Therefore, 100 is chosen as the optimal number of generations in terms of both solution quality and efficiency.

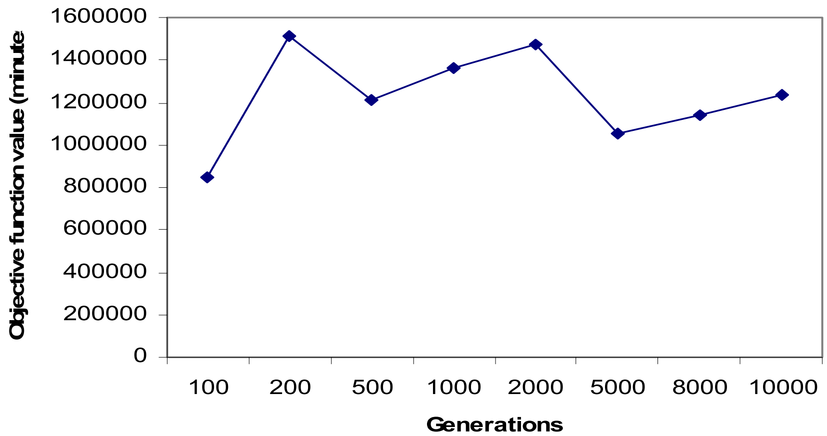


Figure 9.8 Sensitivity Analyses for the Local Search Algorithm

9.4.3 Simulated Annealing

As mentioned before, the performance of the proposed simulated annealing algorithm model greatly depends upon the chosen parameters of the temperature, the stopping criteria (i.e., the number of generations), the alpha value and the repetition counter. The following sections present the sensitivity analyses of these parameters.

9.4.3.1 *Effect of Temperature*

The effect of the initial temperature value is examined by varying this value from 100 to 10000 and the result is given in Figure 9.9(1). It can be seen from the figure that as the initial temperature increases, the objective function value changes unpredictably. It is also noted that the larger the chosen initial temperature, the longer the computation time. When the initial temperature is chosen as 1000, the lowest objective function value is achieved, suggesting that 1000 should be chosen as the optimal initial temperature for the small example network.

9.4.3.2 *Effect of Generations*

The effect of stopping criteria is investigated by choosing the number of generations ranging from 5 to 100 and the result is provided in Figure 9.9(2). As can be seen, the lowest objective function value is achieved with 30 generations. Therefore, 30 is chosen as the optimal number of generations in terms of efficiency.

9.4.3.3 *Effect of Alpha Value*

The effect of the alpha value is also studied by varying this value from 0.1 to 0.9. The result shown in Figure 9.9(3) indicates that 0.6 might be the optimal value and as a result, it is recommended.

9.4.3.4 *Effect of Repetition Counter*

The effect of the repetition counter is examined by varying this value from 5 to 50 and the result is presented in Figure 9.9(4). As can be seen, the lowest objective function value is achieved at 10, suggesting that the optimal repetition counter might be 10.

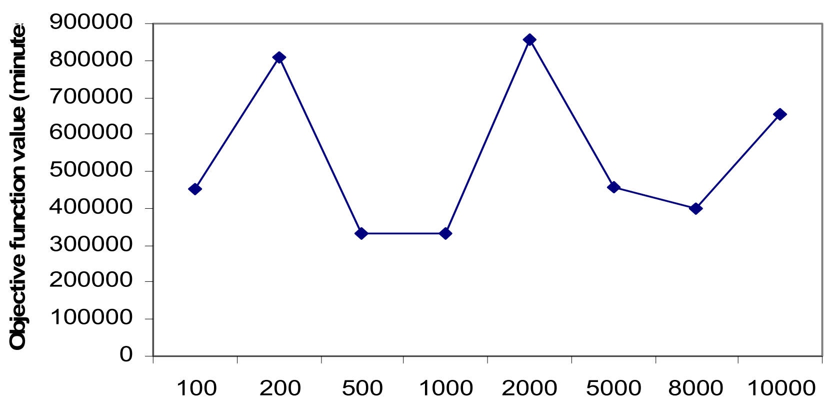


Figure 9.9(1)

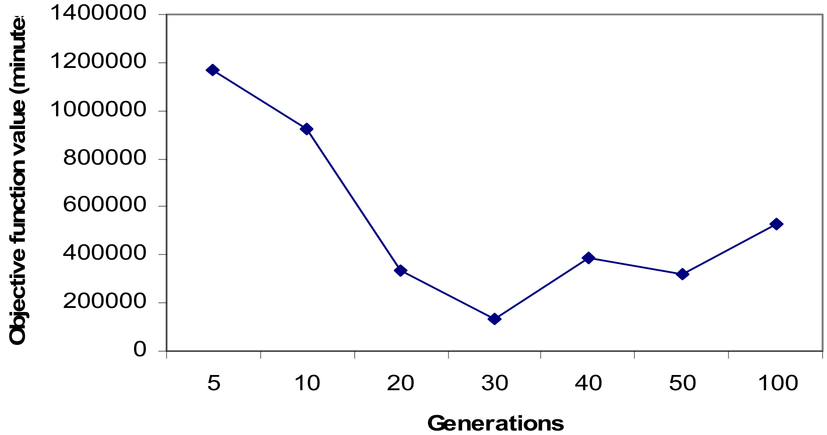


Figure 9.9(2)

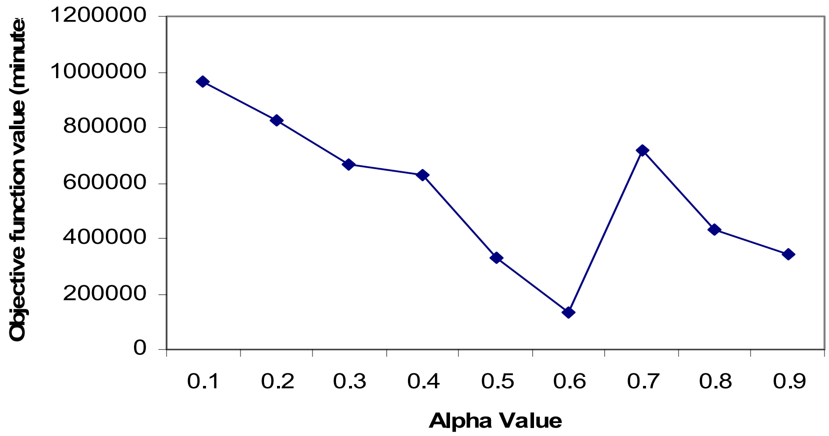


Figure 9.9(3)

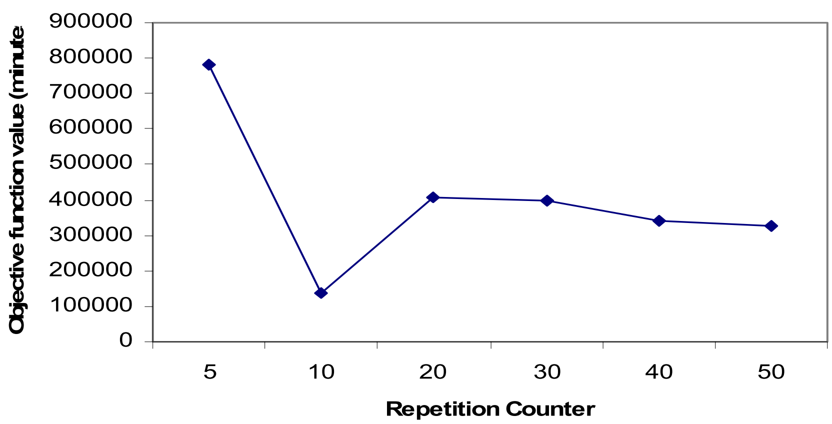


Figure 9.9(4)

Figure 9.9 Sensitivity Analyses for the Simulated Annealing Algorithm

9.4.4 Random Search

The application of the random search algorithm for the BTRNDP has been discussed in Chapters 7 and 8. As described, the random search also has only one parameter, which is the number of generations. The effect of the number of generations is studied by choosing the number of generations ranging from 5 to 10000 and the result is provided in Figure 9.10. As can be seen, the lowest objective function value can be achieved with 5000 generations. Therefore, 5000 is chosen as the optimal number of generations in terms of solution quality.

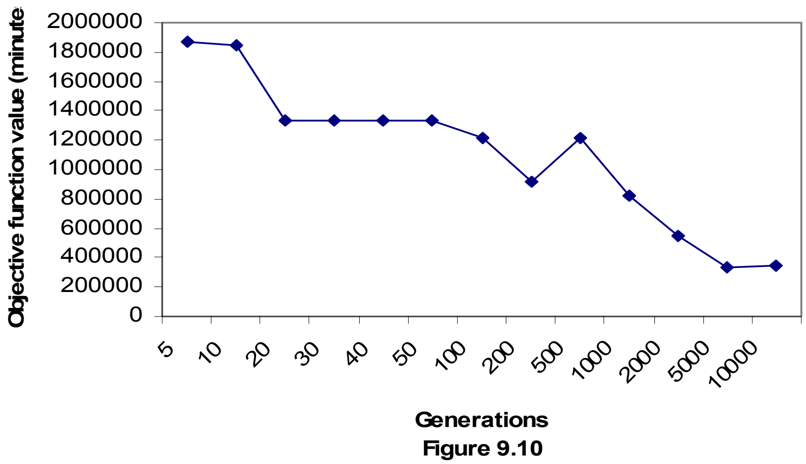


Figure 9.10 Sensitivity Analyses for the Random Search Algorithm

9.4.5 Tabu Search Methods

As mentioned in Chapter 7, three versions of the TS methods are used In this research,: 1) Tabu search without shakeup procedure (i.e., without the diversification and intensification procedure as defined before) and with fixed tabu tenures; 2) Tabu search with shakeup procedure and fixed tabu tenures (i.e., the number of restrictions set for tabu moves are fixed); and 3) Tabu search with shakeup procedure and variable tabu tenures (i.e., the number of restrictions set for tabu moves are randomly generated). The differences underlying each TS algorithm are self-explained by the names. The sensitivity analyses for each version are presented as follows.

9.4.5.1 Tabu without Shakeup and with Fixed Tenures

9.4.5.1.1 *Effect of Generations*

The effect of the number of generations is examined by varying this value from 5 to 100 and the result is given in Figure 9.11(1). It can be seen from the figure that as the number of generations increases, the objective function value tends to decrease. It is also noted that the larger the chosen number of generations, the longer the computation time.

When the number of generations reaches 30, the optimal objective function is achieved, suggesting that 30 should be chosen as the optimal generations for the small network. Therefore, a generation of 30 was recommended.

9.4.5.1.2 *Effect of Tabu Tenures*

The effect of Tabu tenures (i.e., the number of restrictions) is investigated by choosing this number ranging from 5 to 40 and the result is provided in Figure 9.11(2). As can be seen, the lowest objective function value occurred with 10 restrictions. Therefore, 10 is chosen as the optimal number of tabu move tenures.

9.4.5.1.3 *Effect of Search Neighbors*

The effect of search neighbors is also studied by varying this value from 10 to 100. The result shown in Figure 9.11(3) indicates that 20 might be the optimal value and as a result, it is recommended.

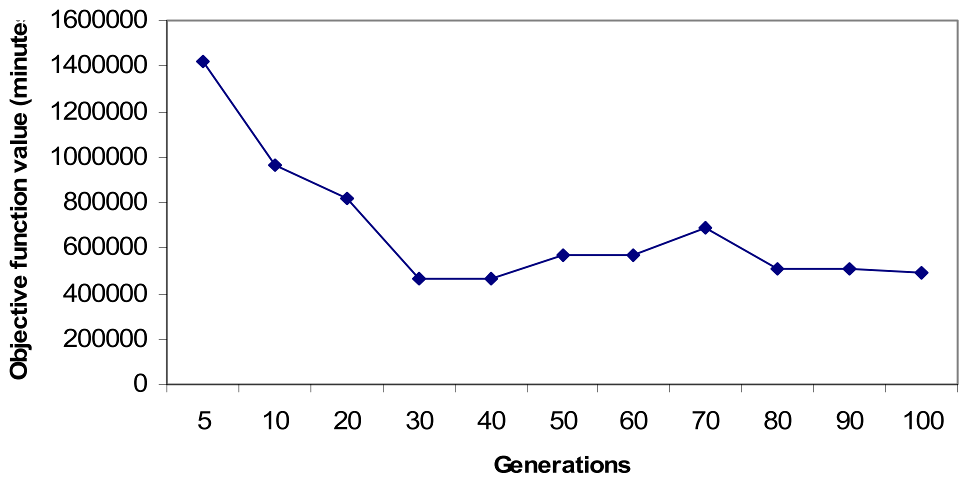


Figure 9.11(1)

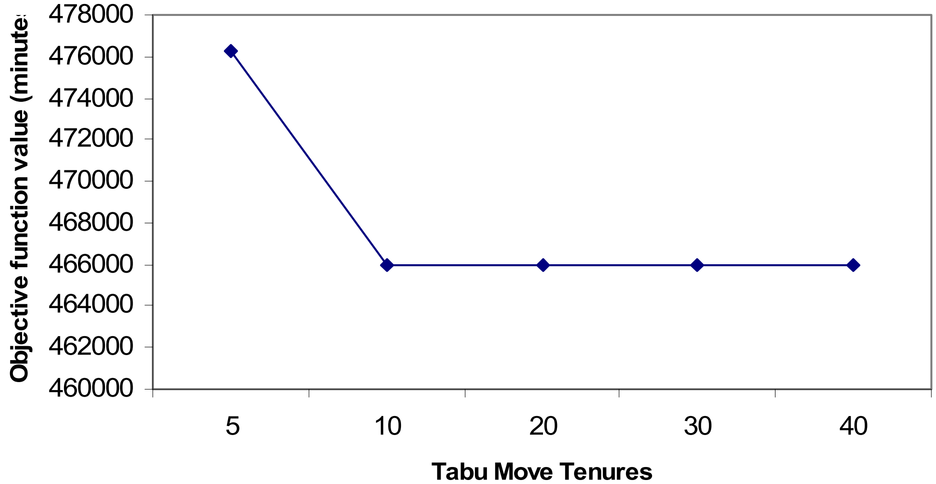


Figure 9.11(2)

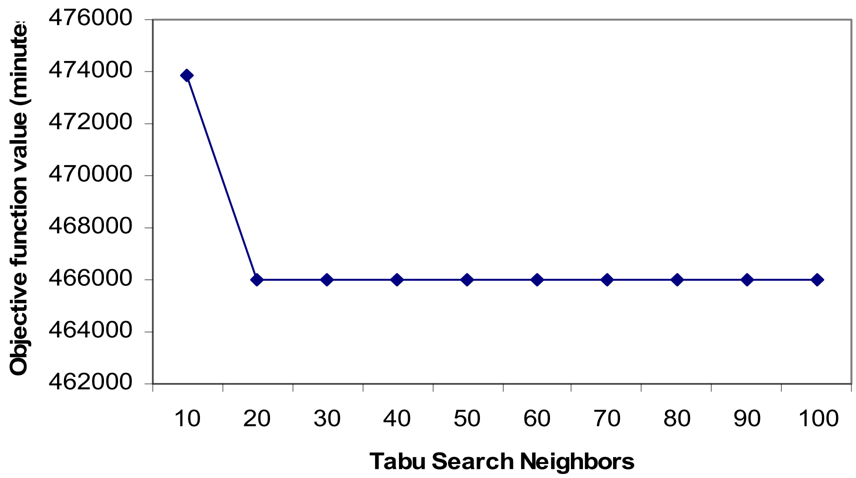


Figure 9.11(3)

Figure 9.11 Sensitivity Analyses for the Tabu Algorithm without Shakeup and with Fixed Tenures

9.4.5.2 Tabu with Shakeup and Fixed Tenures

The effect of each parameter involved with this TS algorithm is examined by varying the parameter value within a specific range and the result is given in Figure 9.12. It can be seen from the figure that 80, 10, 10 and 50 might constitute the optimal parameter set and therefore, they are recommended as the optimal number of generations, the number of search neighbors, the tabu tenures and the shakeup number respectively.

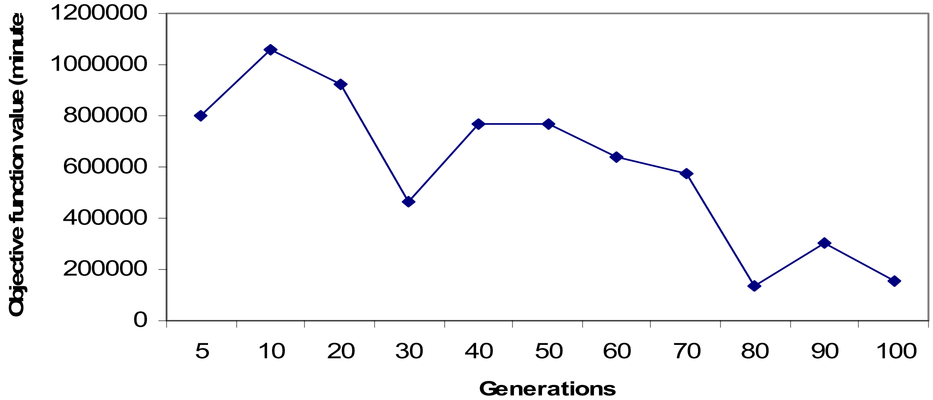


Figure 9.12(1)

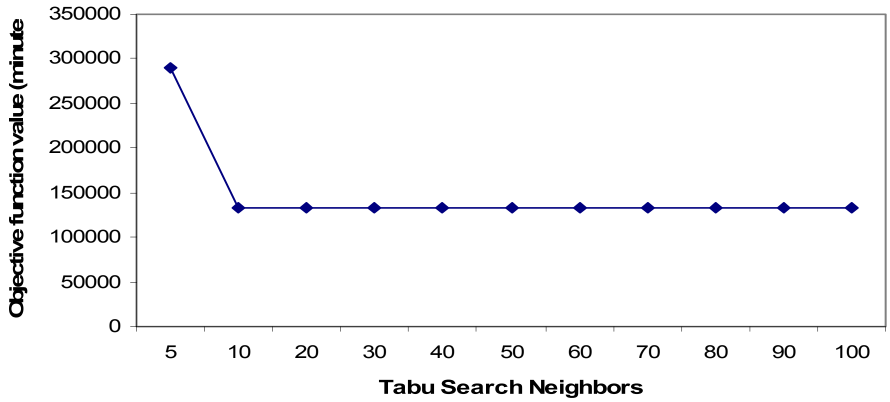


Figure 9.12(2)

Figure 9.12(3)

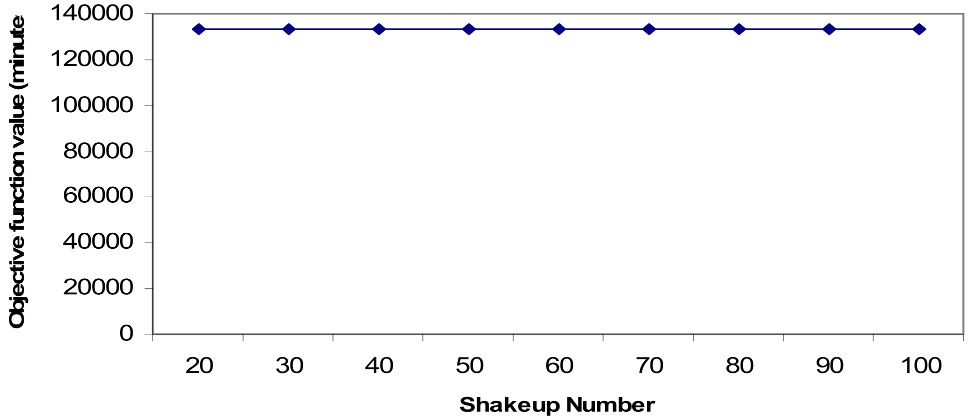


Figure 9.12(4)

Figure 9.12 Sensitivity Analyses for the Tabu Algorithm with Shakeup and Fixed Tenures

9.4.5.3 Tabu with Shakeup and Variable Tenures

Similarly, the effect of each parameter involved with this TS algorithm is also examined by varying the parameter value within a specific range and the result is given in Figure 9.13. It can be seen from the figure that 20, 40 and 50 might constitute the optimal parameter set and is recommended as the optimal number of generations, the number of search neighbors and the shakeup number respectively.

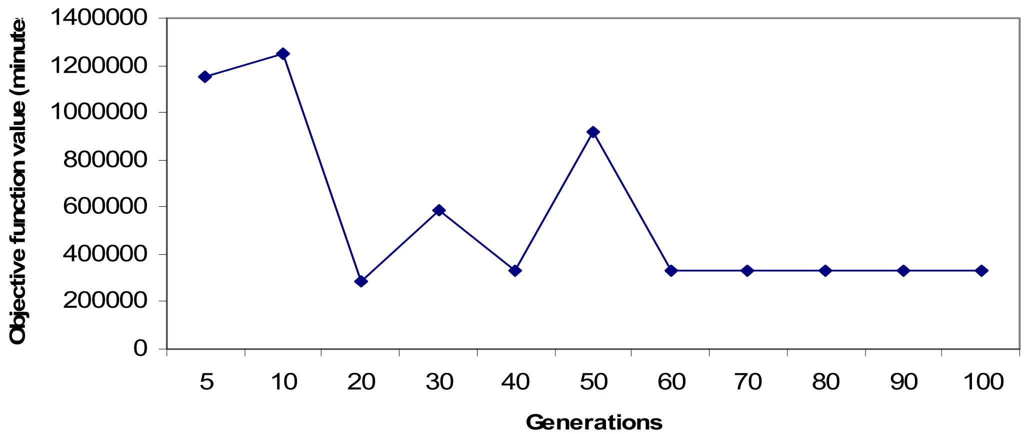


Figure 9.13(1)

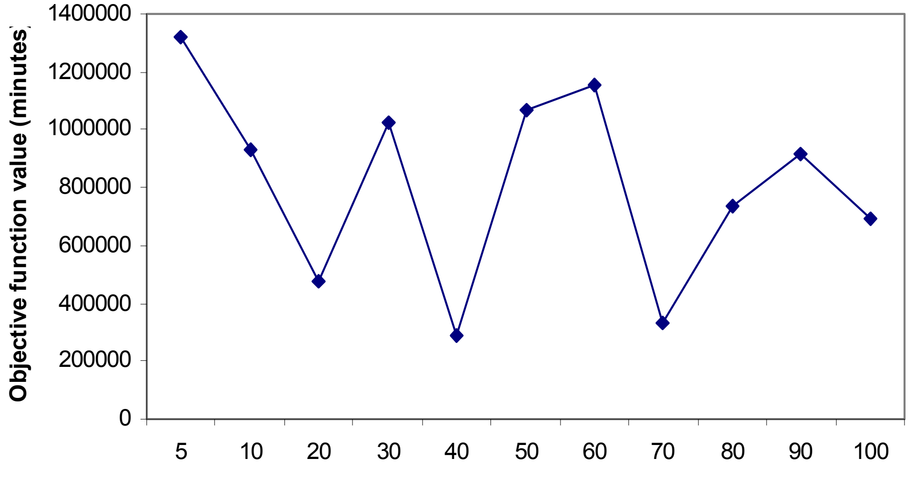


Figure 9.13(2)

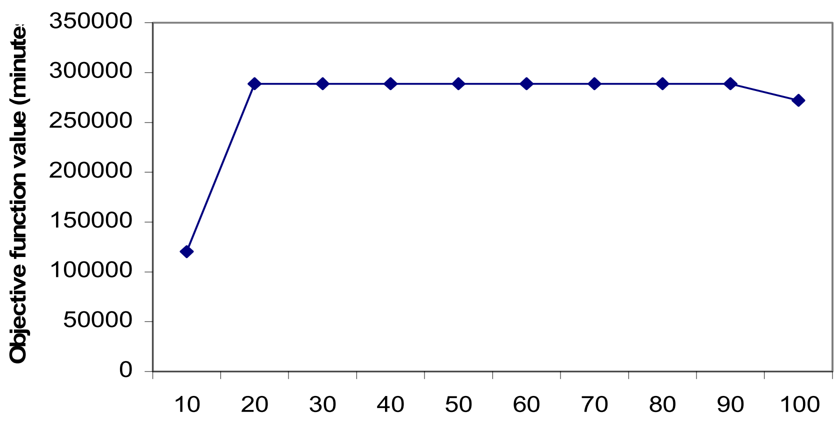


Figure 9.13(3)

Figure 9.13 Sensitivity Analyses for the Tabu Algorithm with Shakeup and Variable Tenures

9.4.6 Summary of Sensitivity Analyses

The above subsections presented the sensitivity analyses for each heuristic algorithm including the GA, LS, SA, RS and TS methods for the BTRNDP with fixed transit demand for the small example network. For sensitivity analyses regarding the BTRNDP for the medium network, the same procedure can be followed and the following table provides a summary of these sensitivity analyses for each algorithm for the BTRNDP with fixed transit demand (including the small and medium networks).

Table 9.1 Summary of Algorithm Sensitivity Analyses for the BTRNDP with Fixed Demand

			Fixed Demand	
			Small Network	**Medium Network**
Genetic Algorithm		Population Size	30	40
		Generations	20	20
		Crossover Probability	0.8	0.8
		Mutation Probability	0.1	0.1
Local Search		Generations	100	2000
Simulated Annealing		Temperature	1000	10000
		Generations	30	20
		Alpha Value	0.6	0.7
		Repetition Counter	10	10
Random Search		Generations	5000	10000
Tabu Search	**Tabu w/o Shakeup and with Fixed Tenures**	Generations	30	50
		Tenures	10	10
		Search Neighbors	20	20
	Tabu w/t Shakeup and Fixed Tenures	Generations	80	50
		Tenures	10	10
		Search Neighbors	10	10
		Shakeup Number	50	50
	Tabu w/t Shakeup and Variable Tenures	Generations	20	100
		Search Neighbors	40	70
		Shakeup Number	50	50

One can see from this table that the values in the optimal parameter set for each algorithm change slightly as the network size increases from the small network to the medium one. This might suggest that the optimal parameter set can be generalized and used for different networks. However, as can be seen, they might also depend on specific network characteristics including the network size and its configuration. Therefore, it is recommended that when one uses this software or wants to develop a solution framework to design an optimal transit route network, it will be better to perform the sensitivity analyses for the employed algorithm for the studied specific network although one can use the above optimal values in the parameter set as a general guideline.

9.4.7 Exhaustive Search

As presented in previous chapters, an exhaustive search method is developed as a benchmark for the BTRNDP with fixed demand using the small network to examine the efficiency and measure the quality of solutions obtained from all the above heuristic algorithms. Figure 9.14 gives the related numerical results. As one can see from Figure 9.14(1) and 9.14(2), when the route set size is very small (such as equal to 1), all heuristic algorithms seem to be quite good because the global optimal can be captured. This is expected because the solution space only contains 286 feasible routes in this case. Put another way, for route set size 1, there are only 286 possibilities (i.e., the 1st route to the 286th route). The sensitivity analyses show that the generations normally would be greater than 286. Therefore, the probability of capturing the global optimum solution is very high. However, as the route set size (i.e., the number of transit routes in the proposed solution network) increases, the solution qualities obtained from these heuristic algorithms seem to decrease dramatically. This is also expected because the combinations of feasible route sets change from 286 to 40755 to 3858140 as route set size increase from 1 to 2 to 3. That is to say, the solution space for a specific route set size increase exponentially as described in Chapter 4 and as shown Figure 9.14 (3). As a result, the possibility of capturing the global optimal solution for the heuristic algorithm will decrease dramatically (if not exponentially). Therefore, the percentage increases compared to the global optimum for all heuristic algorithms increase dramatically as shown in Figure 9.14 (2). However, the price that the exhaustive search method paid for the achievement of the global optimum value is its extremely long computation time compared to any heuristic algorithm, which is clearly shown in Figure 9.14(4).

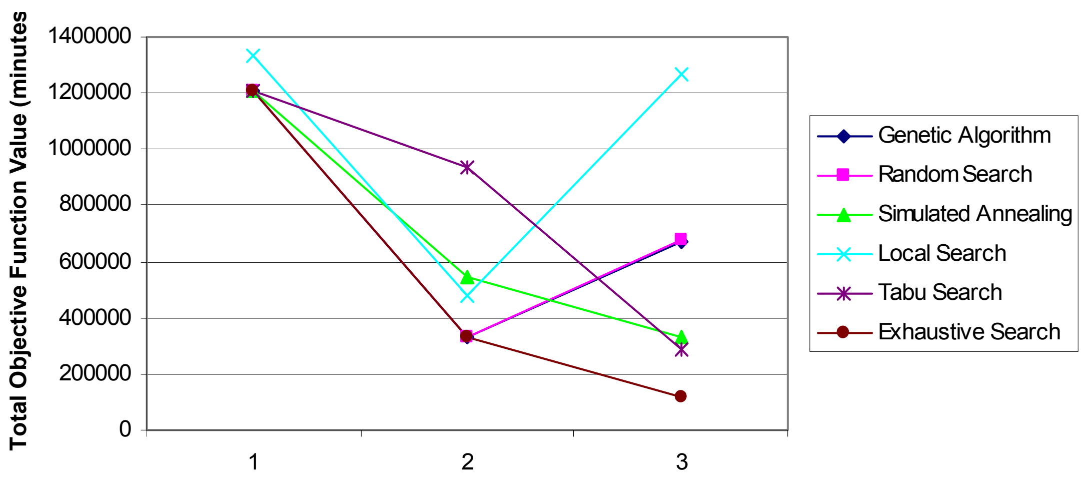


Figure 9.14(1)

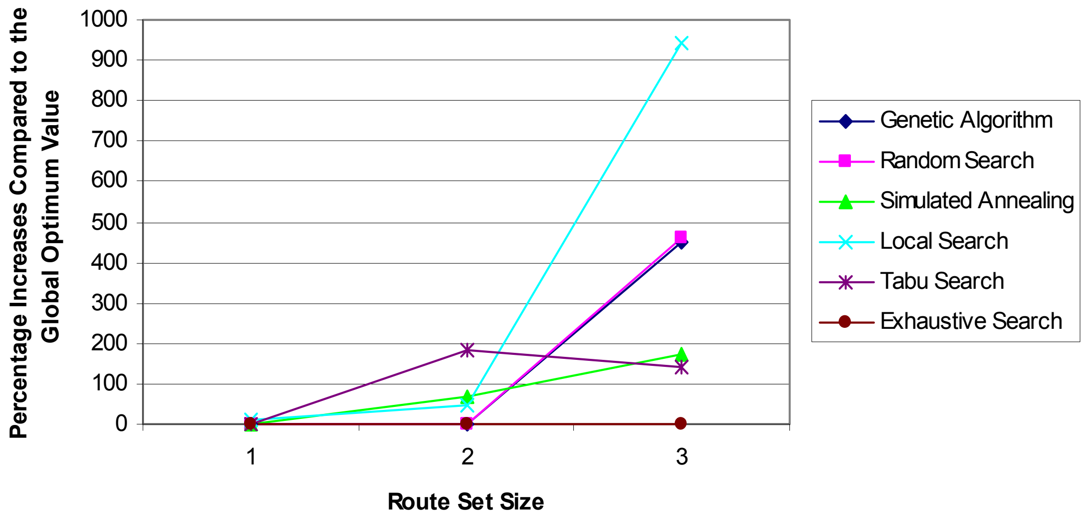


Figure 9.14(2)

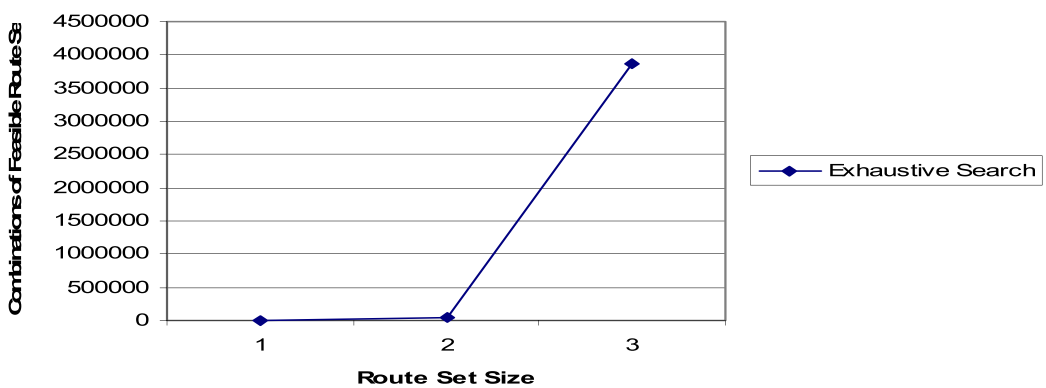


Figure 9.14(3)

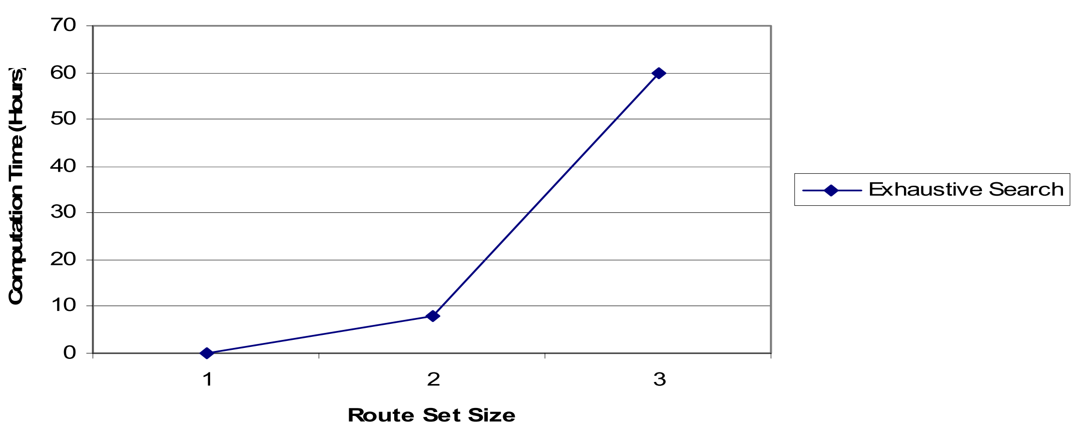


Figure 9.14(4)

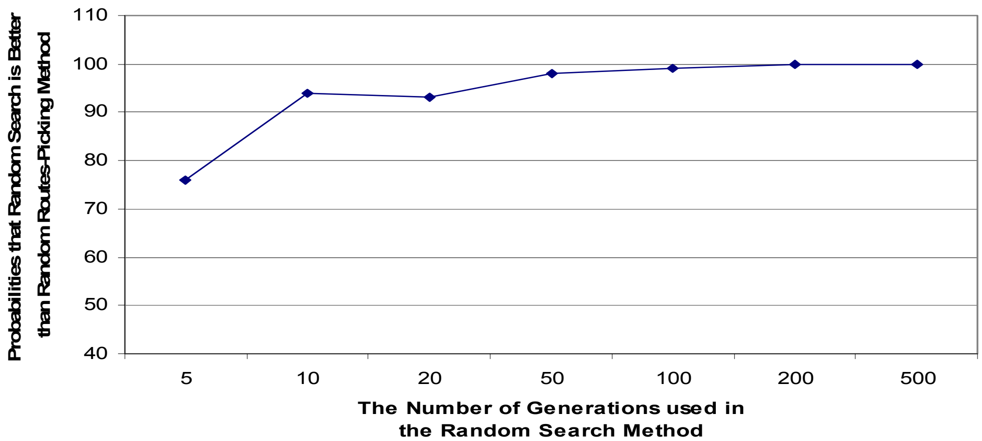


Figure 9.14 (5)

Figure 9.14 Numerical Results Comparisons between the Heuristic Algorithms and the Exhaustive Search Method

As can be seen in Figure 9.14 (1) and (2), as the number of specified routes increases, the quality of the solutions obtained from the heuristic algorithms decreases significantly compared to that from the exhaustive search method. One therefore might reasonably doubt whether the heuristic search methods should be used rather than just randomly picking the transit routes for simplicity in the network. To clarify this issue, the random search method is chosen as the representative heuristic algorithm and compared with random routes-picking method. By running the same programs using 100 consecutive replications, comparison between these two algorithms are preformed and the result is presented in the Figure 9.14 (5). As can be seen, as the number of generations currently used in the Random Search Method increases from 5 to 500, the probabilities that the random search method outperforms the random routes-picking method improved from 76% to 100%. Furthermore, when the number of generations used in the Random Search method exceeds 200, one can guarantee that the random search method performs much better than the random routes-picking method. Since the optimal parameter for the number of generations used in the random search method shown in the previous sensitivity analyses table is 5000 or 10000, which is much greater than 200, one can be very sure that the random search method produces much better solutions than the random routes-picking method in every scenario.

In conclusion, it is recommended that the heuristic search algorithms, rather than the exhaustive search method, should be used for the BTRNDP with the large networks.

9.4.8 Multi-Objective Decision Making and Algorithm Comparisons

As mentioned before, the model performance based on each proposed algorithm greatly depends upon the chosen value of parameters inherent in that algorithm. In previous sections, an optimal set of user-defined parameters associated with each algorithm are found by first assigning a commonly used weight set to each of the three objective function components and then running the developed programming codes based on that algorithm several times. The sensitivity analyses are then performed and the optimal parameter set are found by choosing those resulting in the lowest objective value from that algorithm. For example, the optimal parameter set for the genetic algorithm for the BTRNDP with the small network is 30 for population size, 20 for the number of generations, 0.8 for the crossover probability and 0.1 for the mutation probability value. In this section, these optimal chosen parameters for each algorithm are used and applied to the BTRNDP at different chosen weight set levels. The objective is to see how the quality of these algorithms varies across different weight levels and one might therefore know which algorithm can be used to best solve the BTRNDP. The following sections first compare the three employed tabu search algorithms. Based on the comparison results, the best tabu algorithm is then chosen for each scenario and compared with other heuristic algorithms including the GA, LS, SA and RS method. Related numerical results are included.

9.4.8.1 Tabu Search Algorithm Comparisons

Three versions of TS algorithms have been introduced and their sensitivity analyses have been conducted. To examine which variation of the TS algorithms is most suitable for the BTRNDP with fixed demand, they are compared from a multi-objective decision making perspective using both the small and medium network. Figure 9.15 presents numerical results for the TS algorithm comparisons with fixed demand using the small network.

Weights of Unsatisfied Demand Cost=0.1

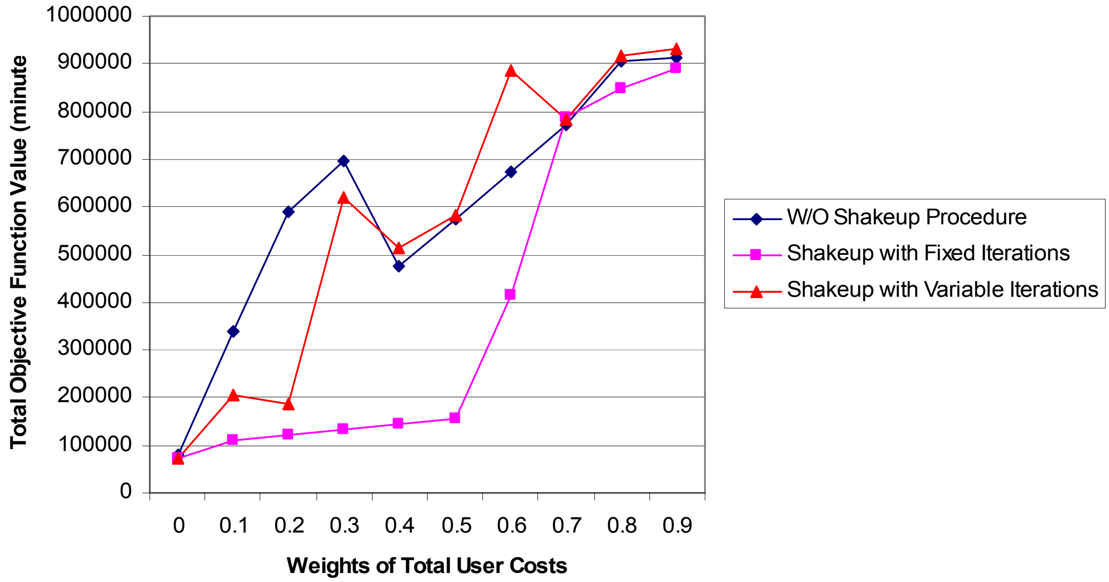


Figure 9.15(1)

Weights of Unsatisfied Demand Cost=0.2

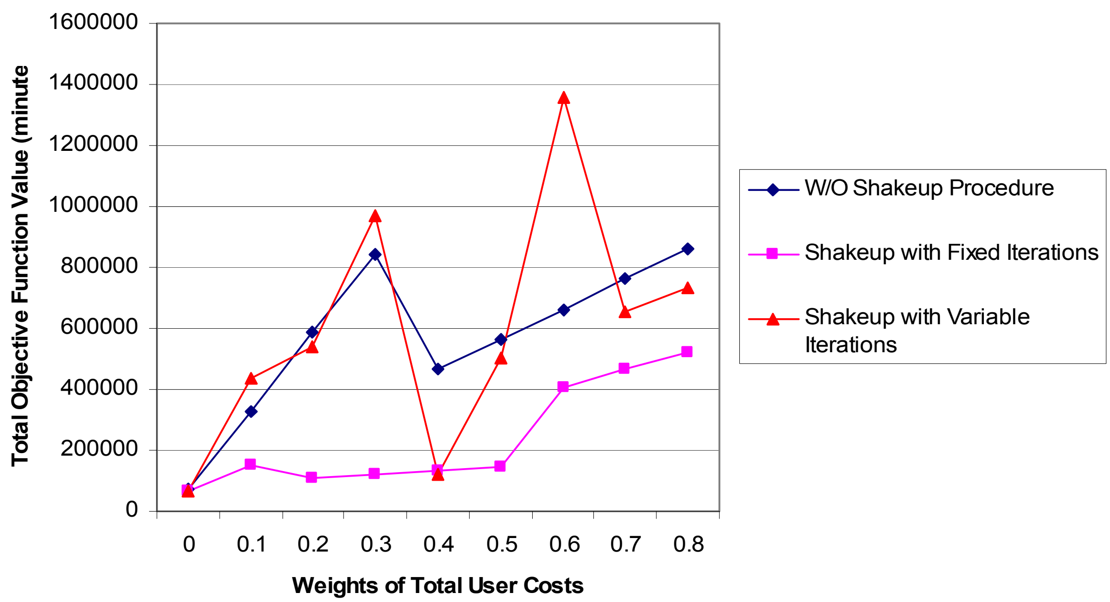


Figure 9.15(2)

Weights of Unsatisfied Demand Cost=0.4

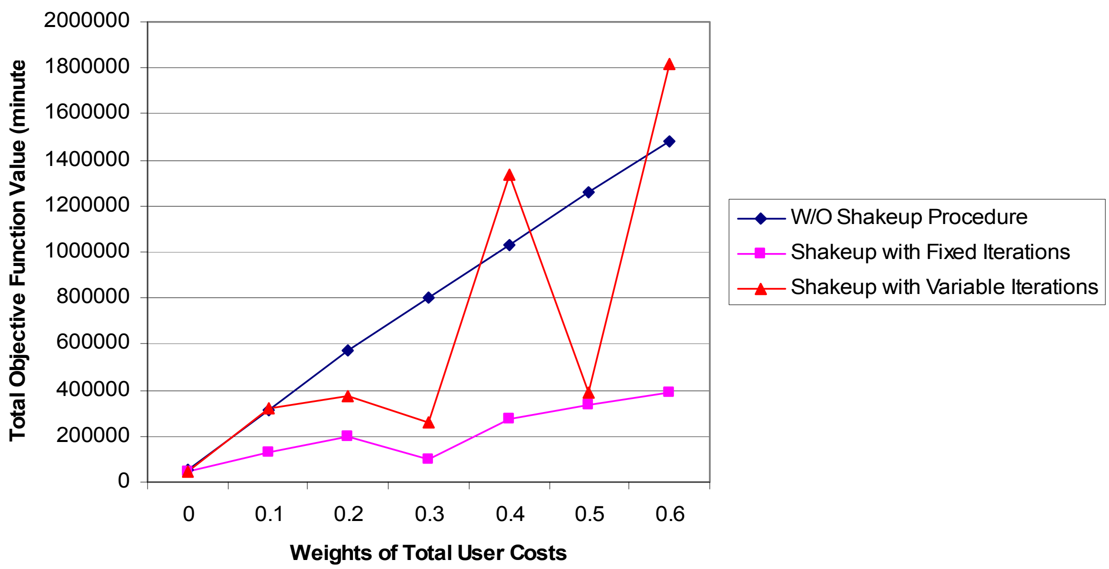


Figure 9.15(3)

Weights of Unsatisfied Demand Cost=0.6

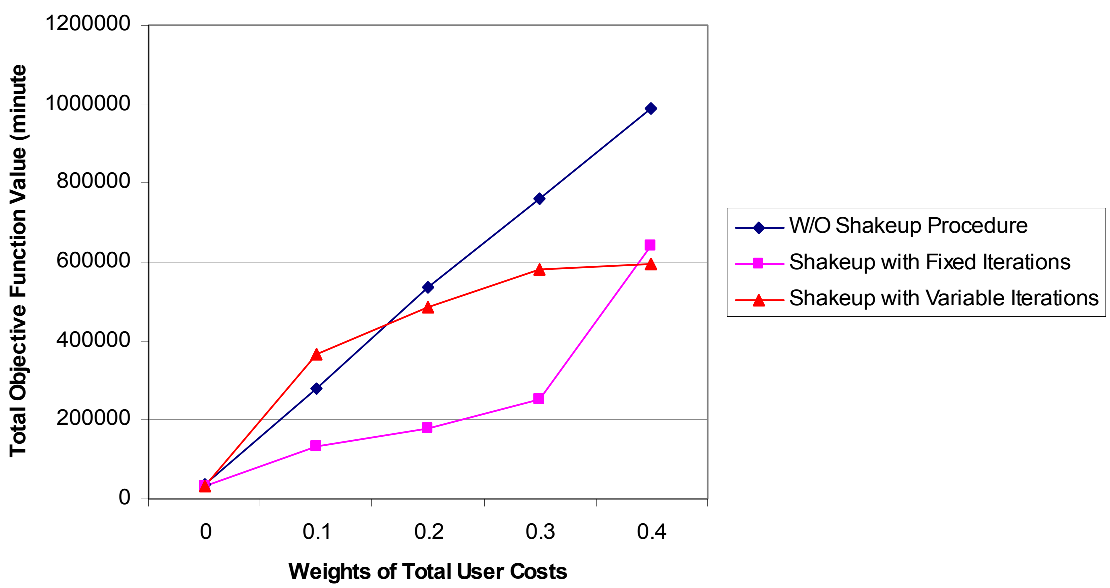


Figure 9.15(4)

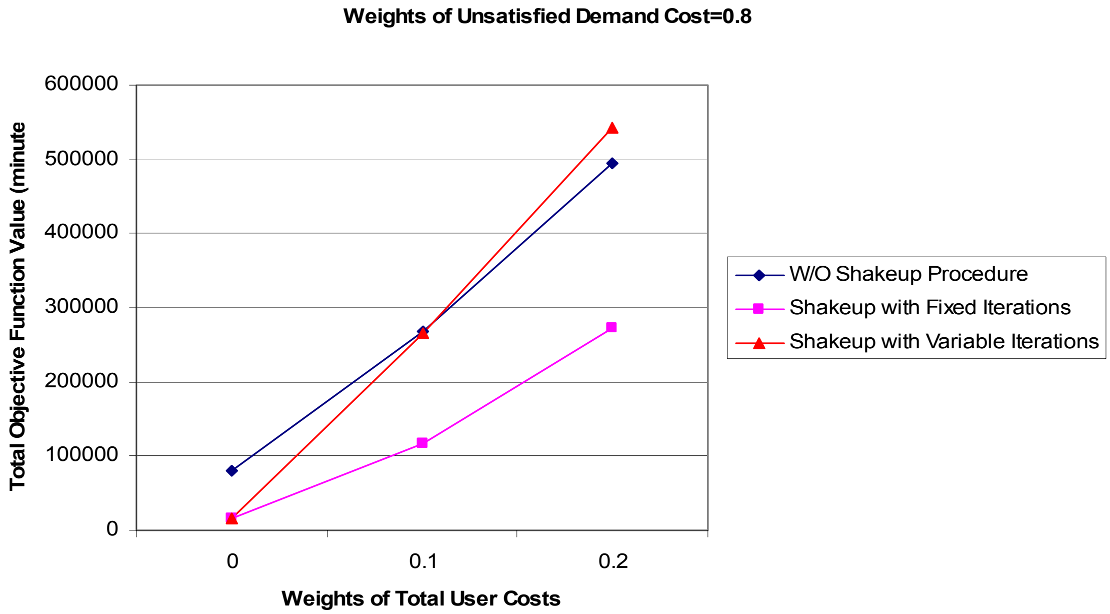


Figure 9.15(5)

Figure 9.15 Tabu Search Algorithm Comparisons using Small Network for the BTRNDP with Fixed Demand

Figure 9.16 presents the numerical results for the TS algorithm comparisons for the BTRNDP with fixed demand using the medium network.

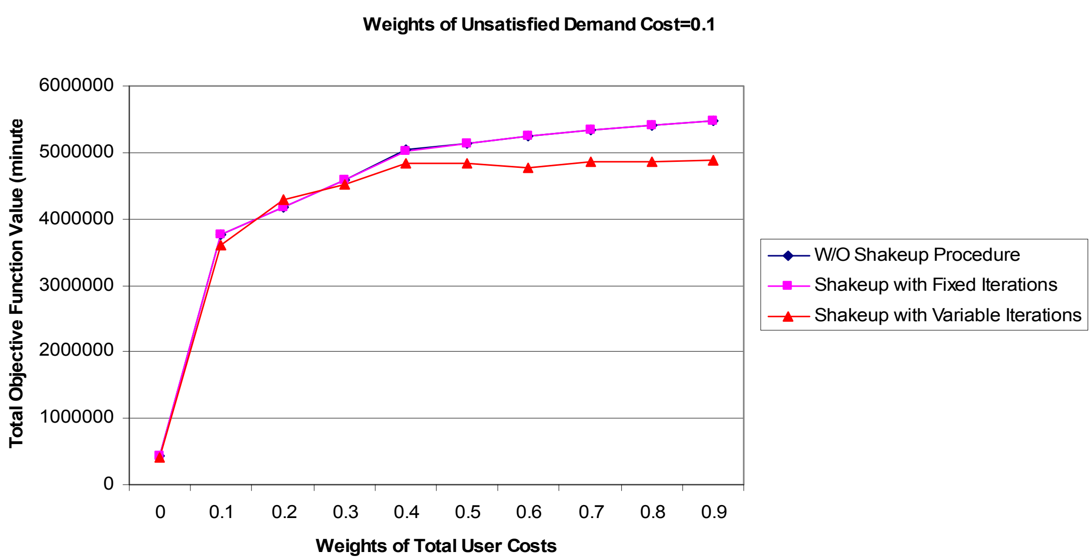


Figure 9.16(1)

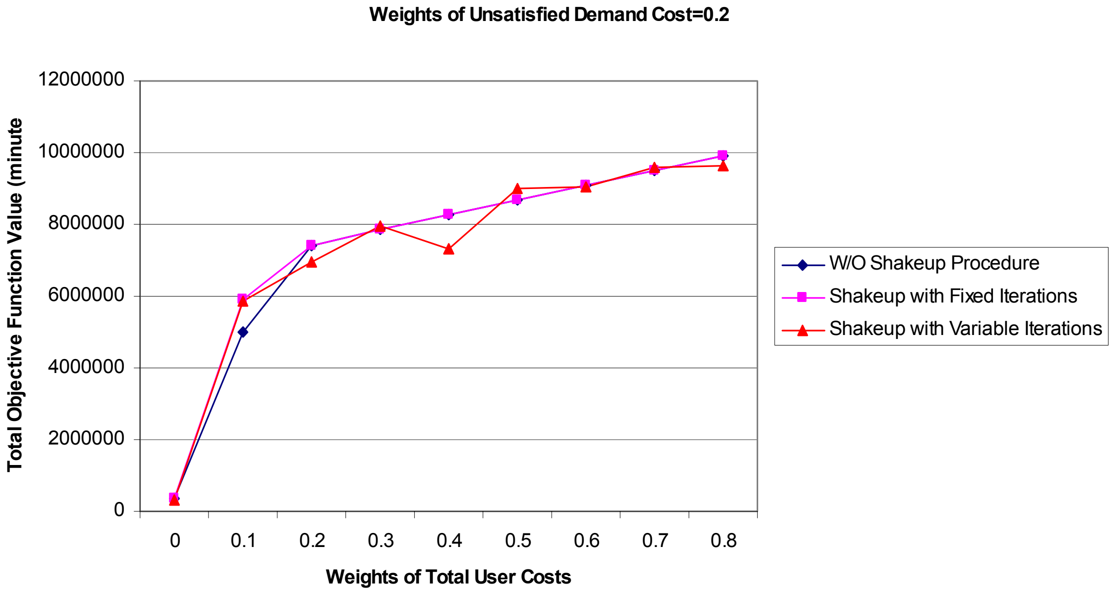


Figure 9.16(2)

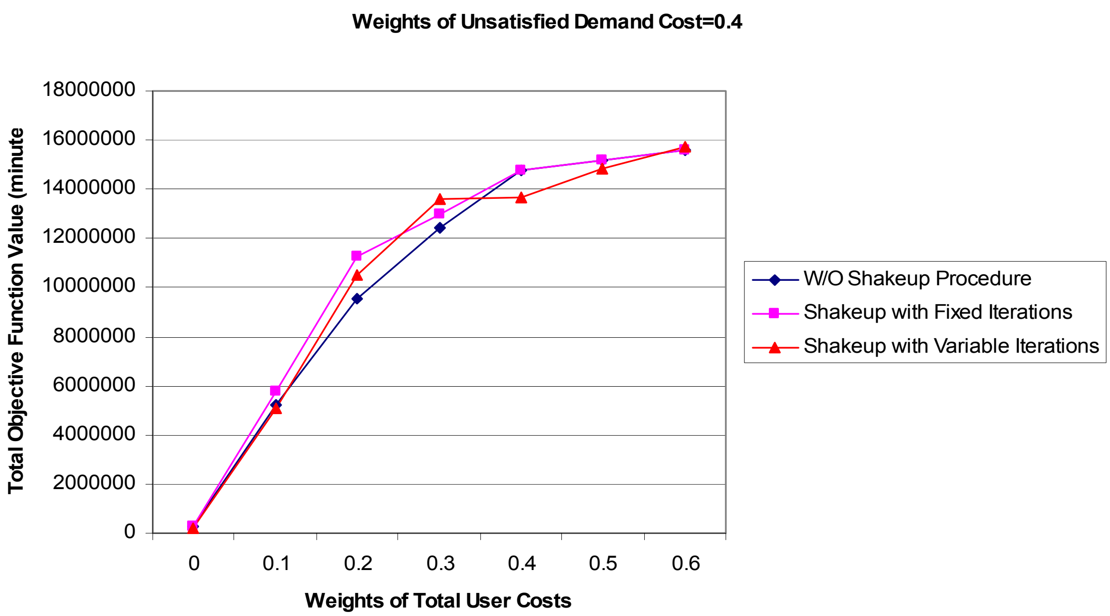


Figure 9.16(3)

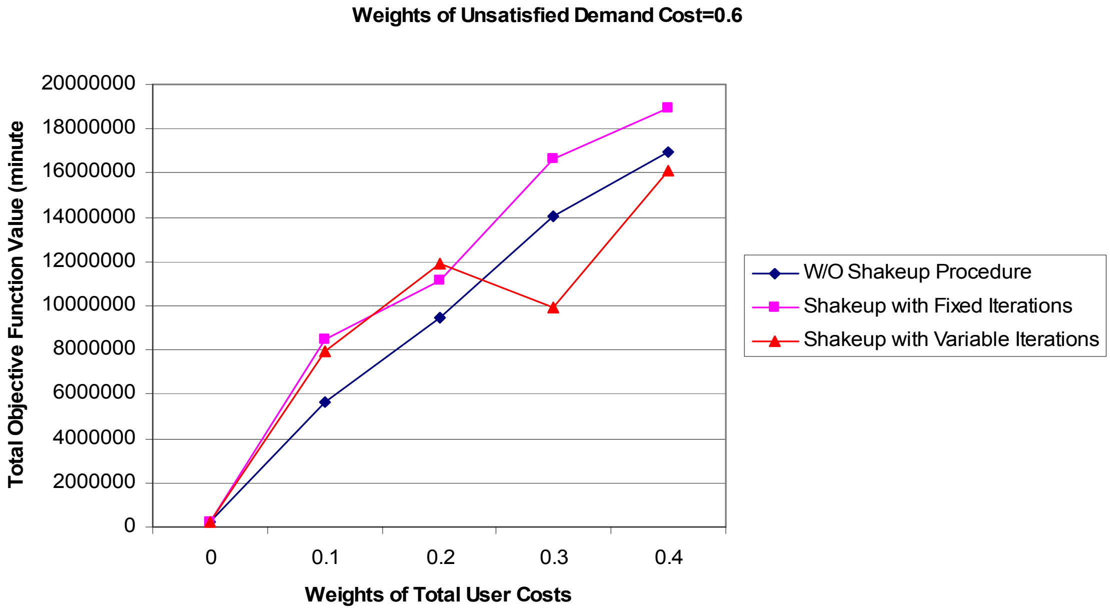


Figure 9.16(4)

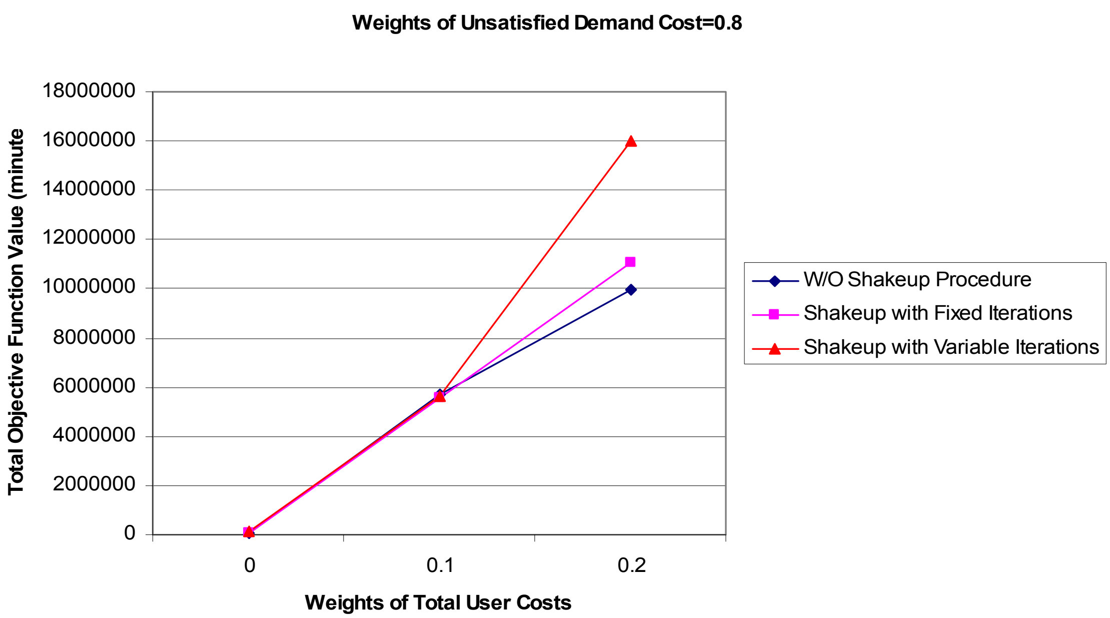


Figure 9.16(5)

Figure 9.16 Tabu Search Algorithm Comparisons using Medium Network for the BTRNDP with Fixed Demand

One can see from Figure 9.15 that Tabu Search with shakeup and fixed tenures (i.e., fixed iterations) clearly seems to outperform other tabu search algorithms using the small network with fixed demand at any weight set level. Therefore, this tabu algorithm is chosen as the best tabu search algorithm. However, one can see from Figure 9.16 that the choice becomes much less distinct using the medium network for the BTRNDP with fixed demand. In other words, this might suggest that as network size increases, big differences among these three tabu search algorithms are unlikely. Put another way, these three Tabu Search algorithms seem to converge in solution quality in some sense so that either one can be used to represent the Tabu Search algorithm to find the optimal transit route set from the solution space. For comparisons with other heuristic algorithms, the tabu search without shakeup is selected and comparison details are presented as follows.

9.4.8.2 Heuristic Search Algorithms Comparisons

Figure 9.17 presents numerical results of Heuristic Search Algorithm Comparisons using the Small Network with Fixed Demand. For each graph, the weight of total unsatisfied demand cost is set at a specific level between 0.1 and 0.8. The *x*-axis denotes the weight of total user cost and the *y*-axis is the objective function value measured in minutes. Note that each point shown for each algorithm in each graph is a decision making problem with a particular weight set for the three components contained in the objective function, where the weight of total operator cost can be obtained at each point by subtracting 1.0 from the weight sum of total unsatisfied demand cost and user cost.

As can be seen for each algorithm from any graph, as the weight of total user cost increases, the optimal objective function value obtained by using that algorithm tend to increase. This is expected because the user cost is usually greater than the operator cost and the increase in total user cost due to a 0.1 unit increase in the weight of total user cost outweighs the decrease in total operator cost due to a 0.1 unit decrease in the weight of total operator cost. As a result, the total objective function value increases.

One interesting phenomenon is that the optimal objective function value obtained from the random search method (and also for the local search method in most cases) seems to follow a linear pattern. This is expected because the optimal number of generations for the random search is found to be 5000, which probably results in a very large chance (if not always) of capturing the same optimal solution network due to the small-sized network and the small overall solution space. However, it is hypothesized that as the network size becomes larger, the random search graph might fluctuate rather than following a linear pattern because the optimal solution network might be different each time (as will be shown in Figure 9.18). In addition, it is noted that the genetic algorithm seems to be more variable than any other algorithm in terms of the optimal objective function value (Figure 9.17(1) to 9.17(5) and Figure 9.18.) This might suggest that, compared to other algorithms, genetic algorithm may largely depend on the chosen parameters at any particular level. If the chosen parameters inherent in GA are fixed, the solution quality for the BTRNDP might be unstable. Therefore, to achieve the optimal solution network at each weight set level, one might need to run the programming code and get the optimal parameter set at that level although the computational burden would become larger.

Furthermore, for each graph (i.e., for each weight level for the total unsatisfied demand cost), simulated annealing seems to consistently outperform any other algorithm in terms of the quality of solution (i.e., it always results in the lowest objective function value). This might reach the conclusion that compared to other heuristic algorithms, simulated annealing performs best. Random search seems to be the second best algorithm for solving the BTRNDP. The overall model performance of the local search method is the third best algorithm. Also note that the local search method is a little worse than the simulated annealing algorithm, which is expected due to the inherent characteristics and relationship between these two model structures. The genetic algorithm, however, seems to be the most undesirable model. This might be possible because although genetic algorithm might achieve some better solutions by learning from the previous solutions through a genetic approach, it might take much more time inside the algorithm itself to look for this achievement (compared to the random search method) while does not take much more effort looking for possibly better solutions from other "neighborhood" solutions in the candidate solution space (compared to the local search and simulated annealing algorithm). Conversely, the simulated annealing algorithm not only can look for a good solution with a specific origin-destination node pair through "random search" in its early stage, but also can fully explore possibly better neighborhood solutions. Note that the tradeoffs between route coverage and the route directness might be well balanced between chosen shortest paths or *k*-th shortest paths between specific origin-destination node pairs. It is expected that this inherent characteristics of the simulated annealing algorithm might make it particularly suited for the BTRNDP and therefore outperform any other algorithm for this small network.

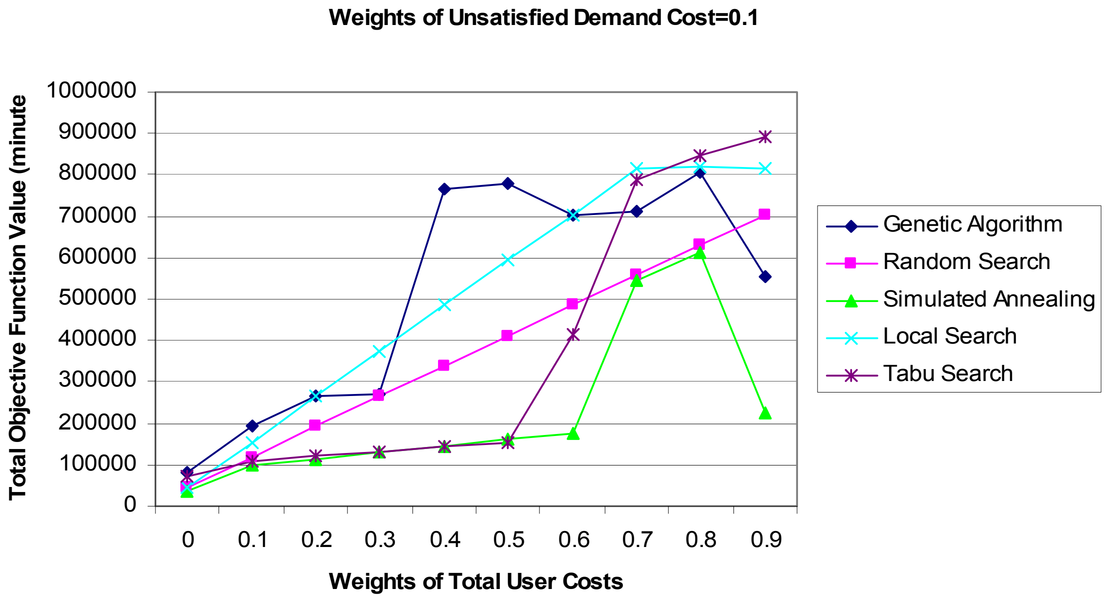


Figure 9.17(1)

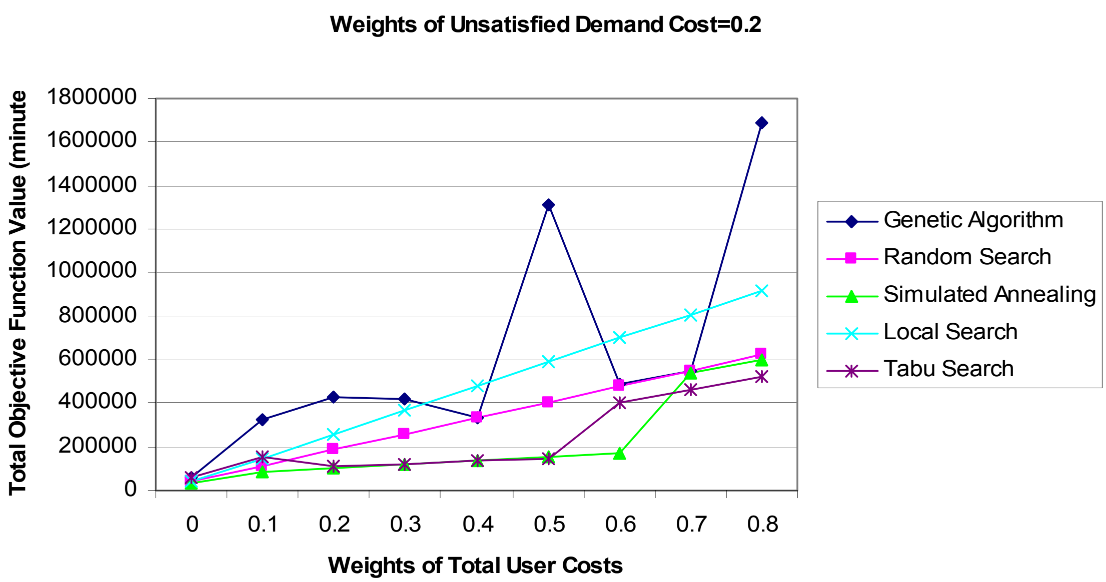


Figure 9.17(2)

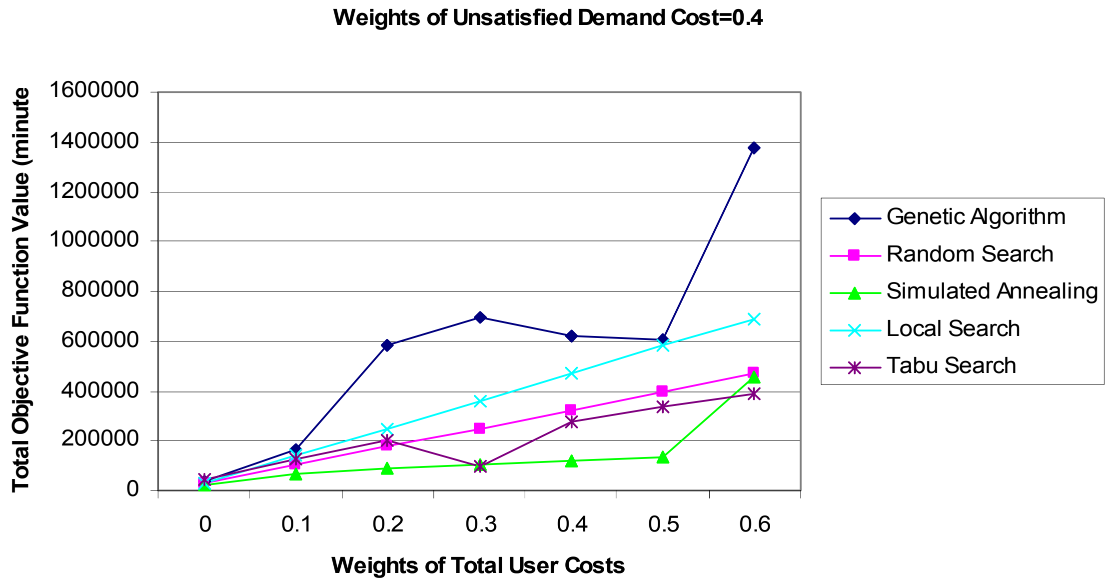


Figure 9.17(3)

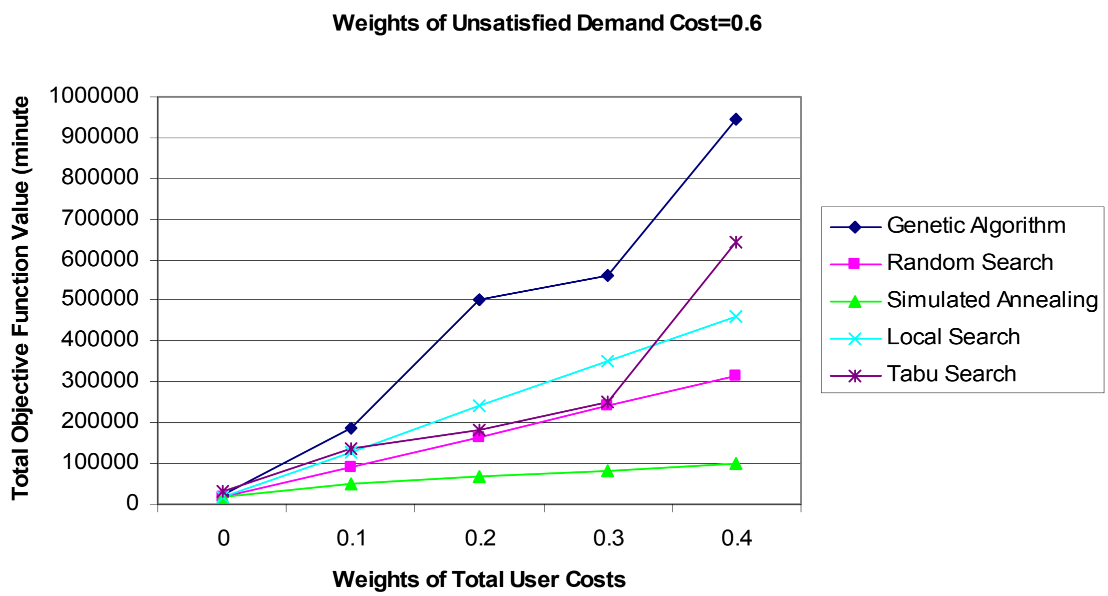


Figure 9.17(4)

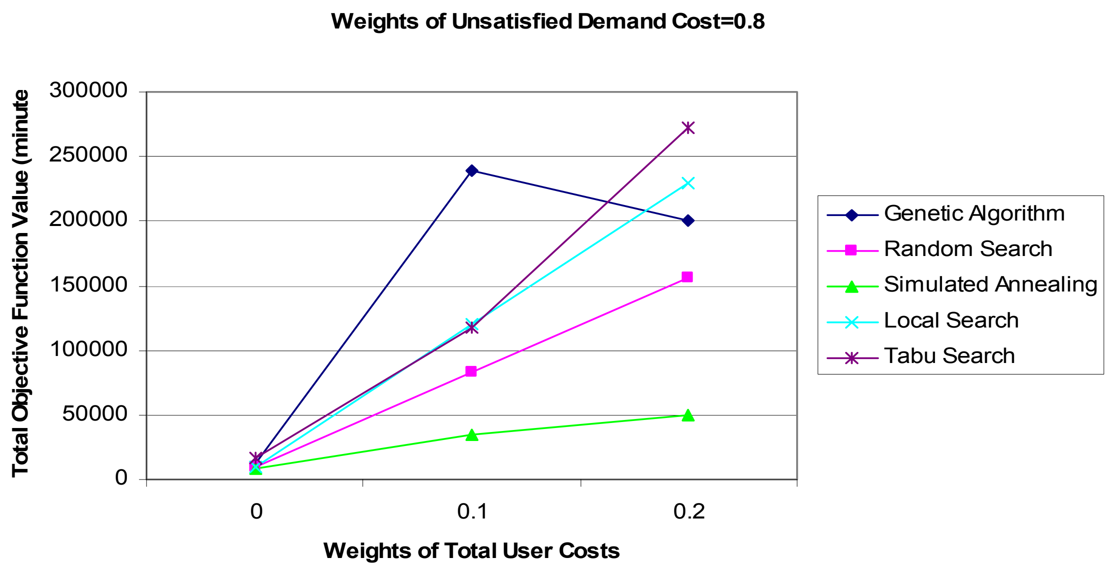


Figure 9.17(5)

Figure 9.17 Heuristic Search Algorithm Comparisons using Small Network for the BTRNDP with Fixed Demand

Figure 9.18 present the numerical results of Heuristic Search Algorithm Comparisons using Medium Network for the BTRNDP with Fixed Demand.

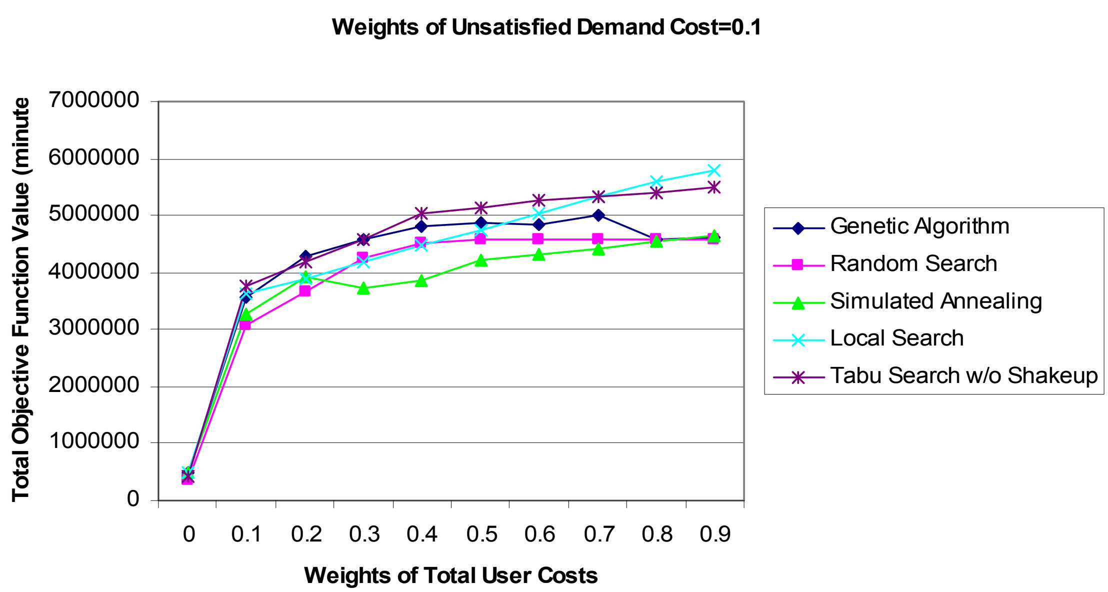


Figure 9.18(1)

Weights of Unsatisfied Demand Cost=0.2

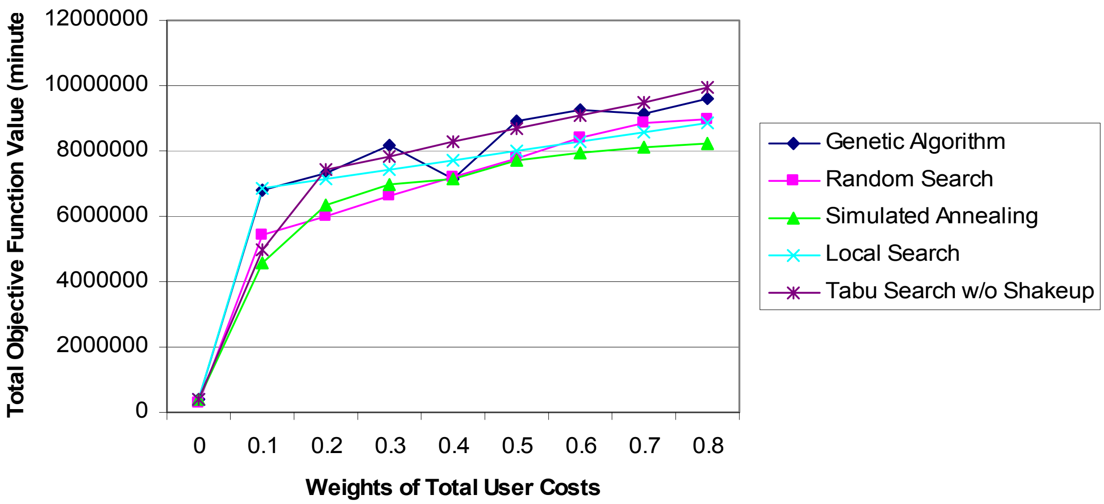


Figure 9.18(2)

Weights of Unsatisfied Demand Cost=0.4

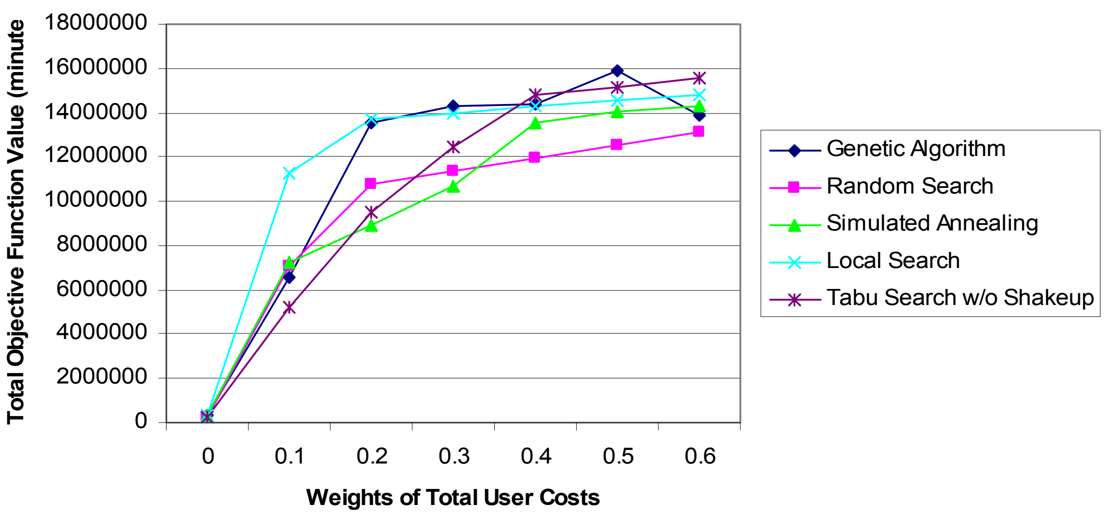


Figure 9.18(3)

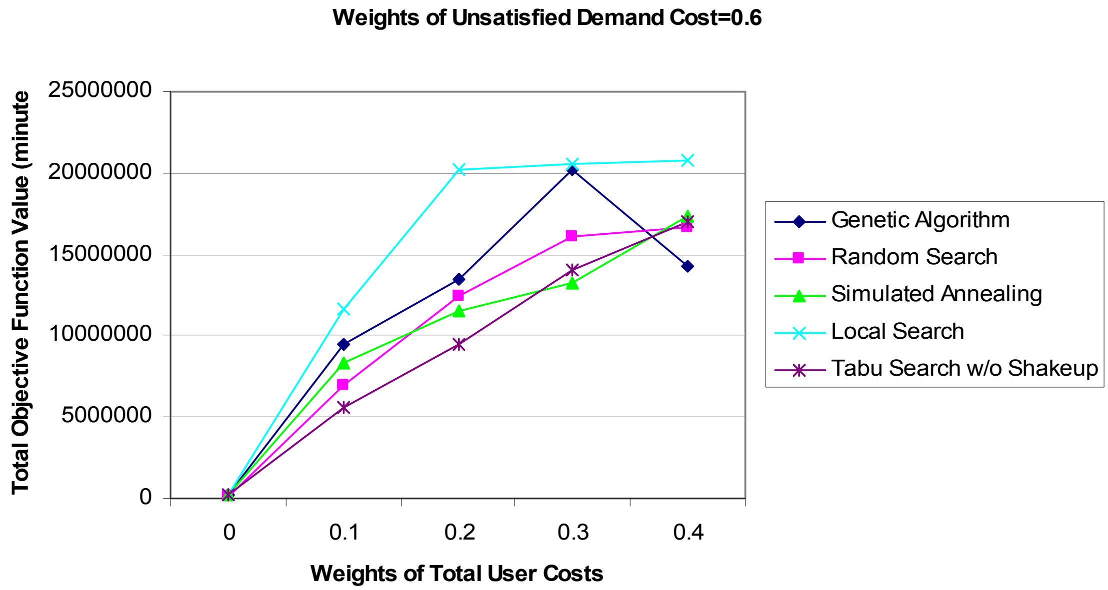


Figure 9.18(4)

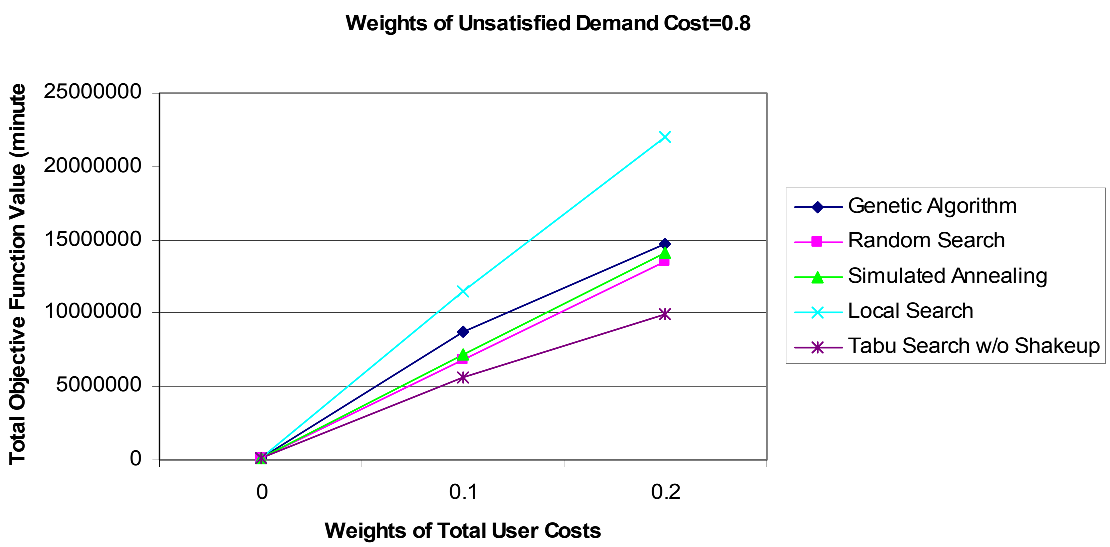


Figure 9.18(5)

Figure 9.18 Heuristic Search Algorithm Comparisons using Medium Network for the BTRNDP with Fixed Demand

As can be seen from Figure 9.18, the local search seems to be the most undesirable algorithm for solving the BTRNDP with fixed demand using this medium network because it produces the worst result in terms of objective function value in most cases at any weight set level. Furthermore, compared to that in Figure 9.17, it seems that the differences among all heuristic algorithms except the local search method become much less distinct and any of these algorithms seems to yield quite consistent solutions. This might suggest that as the network size increases, the heuristic algorithms except the local search method tend to yield transit route network solutions for the BTRNDP whose qualities are almost are similar. Put another way, when one uses this software or wants to develop a new solution framework, any of these heuristic algorithms can be used.

The above sections present the applications of these heuristic algorithms to the BTRNDP with fixed demand. As mentioned in previous chapters, however, there exists a variable relationship between the transit demand and the transit route network. Therefore, the BTRNDP with variable transit demand is studied and the numerical results are presented in the following sections.

9.5 The TRNDP with Variable Transit Demand

Using an approach similar to that for fixed demand, the following sections present the sensitivity analyses for each algorithm and comparisons for all heuristic algorithms for the BTRNDP with variable transit demand.

9.5.1 Algorithm Sensitivity Analyses

For sensitivity analyses regarding the BTRNDP with variable transit demand, the same procedure was followed and the following table provides a summary of these sensitivity analyses for each algorithm for the BTRNDP with variable transit demand (including the small and medium networks).

When the sensitivity analyses were conducted for the BTRNDP with variable demand, the optimal parameter set values for each algorithm also changed as the network size increased from a small network to medium one. Furthermore, comparing Table 9.1 and 9.2, one can see that the parameter changes in Table 9.2 seem to be larger than that in Table 9.1. This might suggest that for the BTRNDP with variable transit demand, the optimal parameter set might greatly depend on specific network characteristics including the network size and its configuration. Therefore, it is strongly recommended that when one uses this software or wants to develop a solution framework to design an optimal transit route network, one should perform the sensitivity analyses for the studied network and use these values as general guidelines.

As before, the following sections first compare the three employed tabu search algorithms. Based on the comparison results, the best tabu algorithm is chosen and compared with other heuristic algorithms including the GA, LS, SA and RS method. Related numerical results are included.

Table 9.2 Summary of Algorithm Sensitivity Analyses for the BTRNDP with Variable Demand

			Variable Demand	
			Small Network	**Medium Network**
Genetic Algorithm		Population Size	80	100
		Generations	20	70
		Crossover Probability	0.8	0.8
		Mutation Probability	0.1	0.01
Local Search		Generations	5000	2000
Simulated Annealing		Temperature	5000	2000
		Generations	100	20
		Alpha Value	0.6	0.6
		Repetition Counter	10	10
Random Search		Generations	5000	10000
Tabu Search	**Tabu w/o Shakeup and with Fixed Tenures**	Generations	10	80
		Tenures	10	10
		Search Neighbors	40	10
	Tabu w/t Shakeup and Fixed Tenures	Generations	50	50
		Tenures	10	10
		Search Neighbors	10	10
		Shakeup Number	50	50
	Tabu w/t Shakeup and Variable Tenures	Generations	10	50
		Search Neighbors	40	10
		Shakeup Number	50	50

9.5.2 Algorithm Comparisons

9.5.2.1 Tabu Search Algorithm Comparisons

Three versions of TS algorithms have been developed and their sensitivity was examined for the BTRNDP with variable demand. To determine which variation of TS is the most suitable for BTRNDP with variable demand, they are compared using both the small and medium networks. Numerical results show that the tabu search algorithms behave like that for the BTRNDP with fixed demand using the medium network. In other words, it seems that there are no big differences among these three tabu search algorithms in terms of solution quality using both the small and medium networks for the BTRNDP with variable demand. Therefore, for conciseness, these graphs are not included in this research. Moreover, this might suggest that any of these three algorithms can be used to represent the Tabu Search algorithm. For simplicity, the tabu search without shakeup is selected as a representative for the tabu search algorithms and details of comparisons among all heuristic algorithms are presented as follows.

9.5.2.2 Heuristic Search Algorithms Comparisons

Figure 9.19 present numerical results of Heuristic Search algorithm comparisons using Small Network for the BTRNDP with Variable Transit Demand.

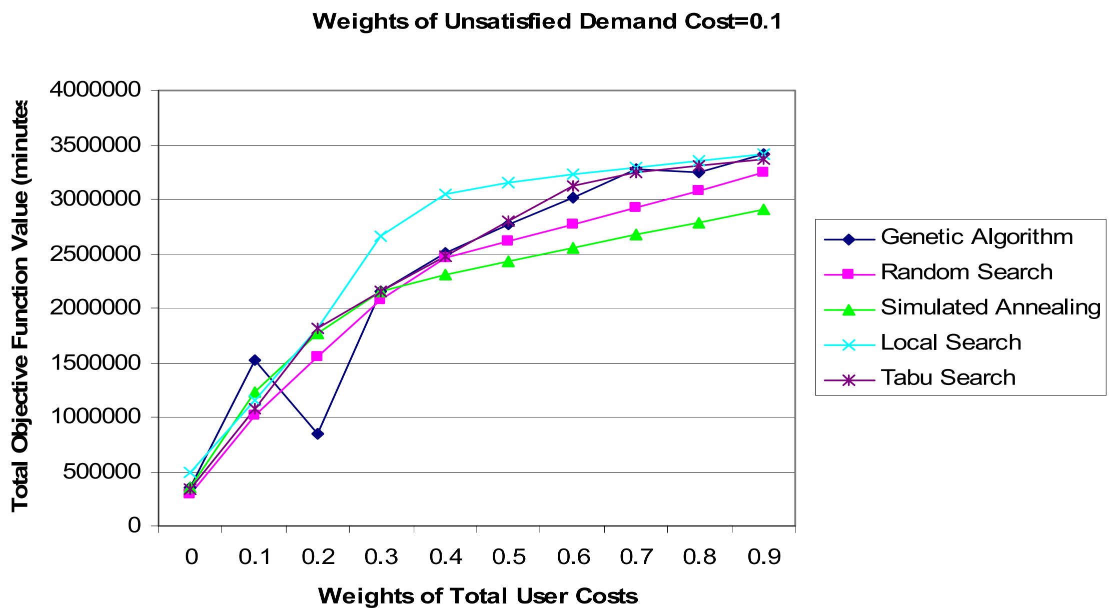


Figure 9.19(1)

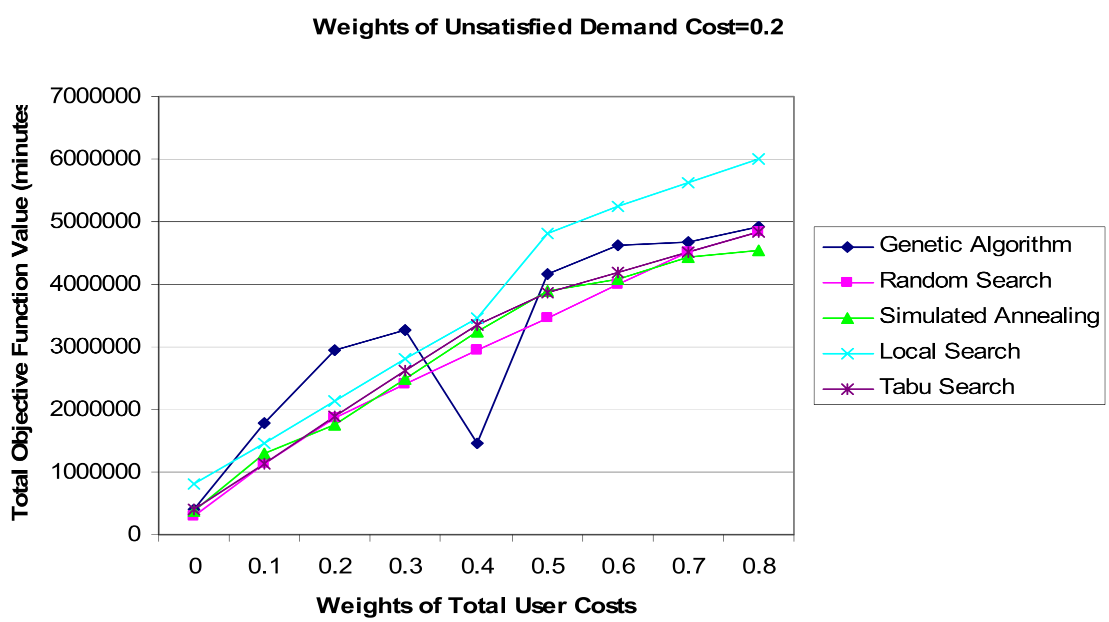


Figure 9.19(2)

Weights of Unsatisfied Demand Cost=0.4

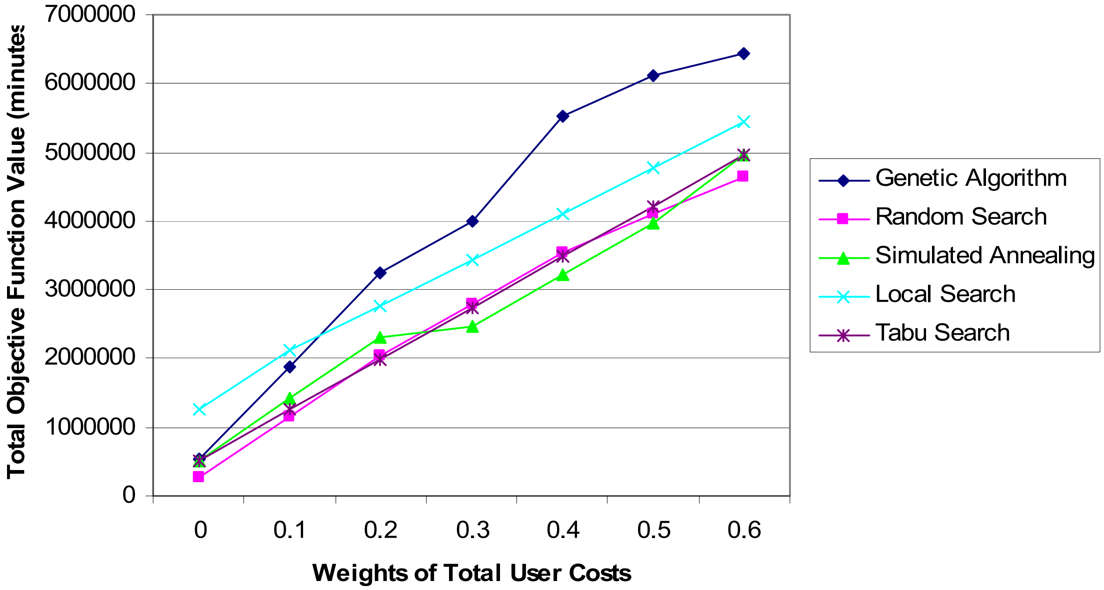


Figure 9.19(3)

Weights of Unsatisfied Demand Cost=0.6

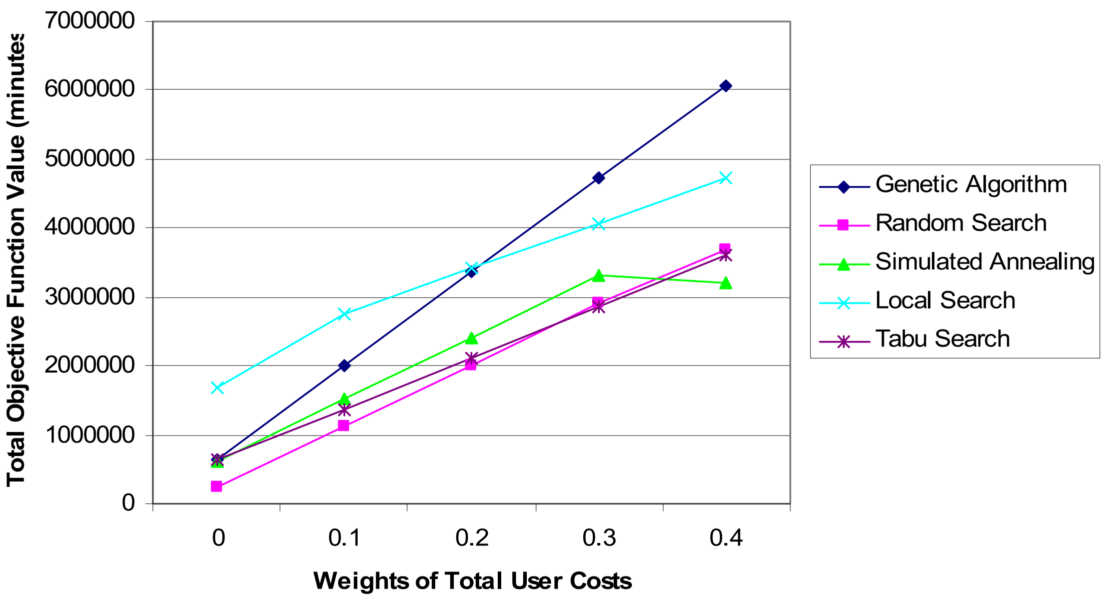


Figure 9.19(4)

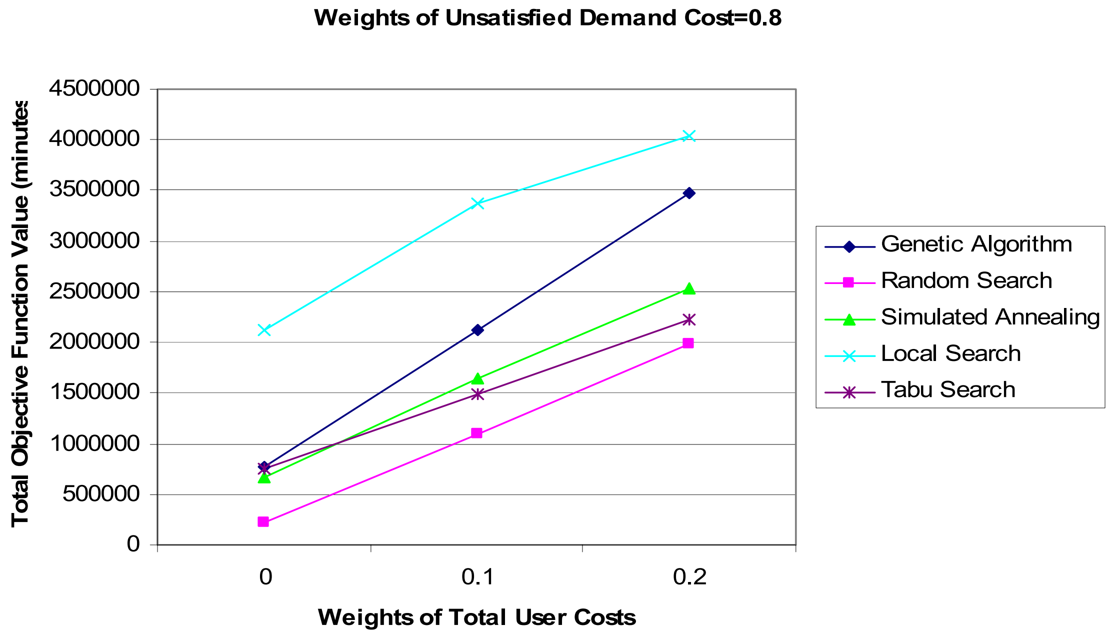


Figure 9.19(5)

Figure 9.19 Heuristic Search Algorithm Comparisons using Small Network for the BTRNDP with Variable Transit Demand

Figure 9.20 present numerical results of Heuristic Search algorithm comparisons using Medium Network for the BTRNDP with Variable Transit Demand.

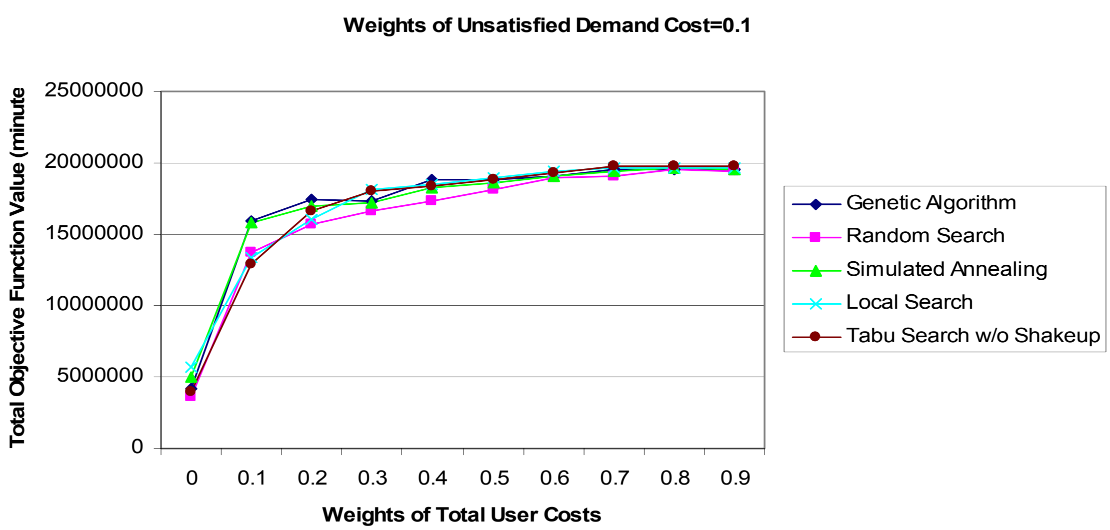


Figure 9.20(1)

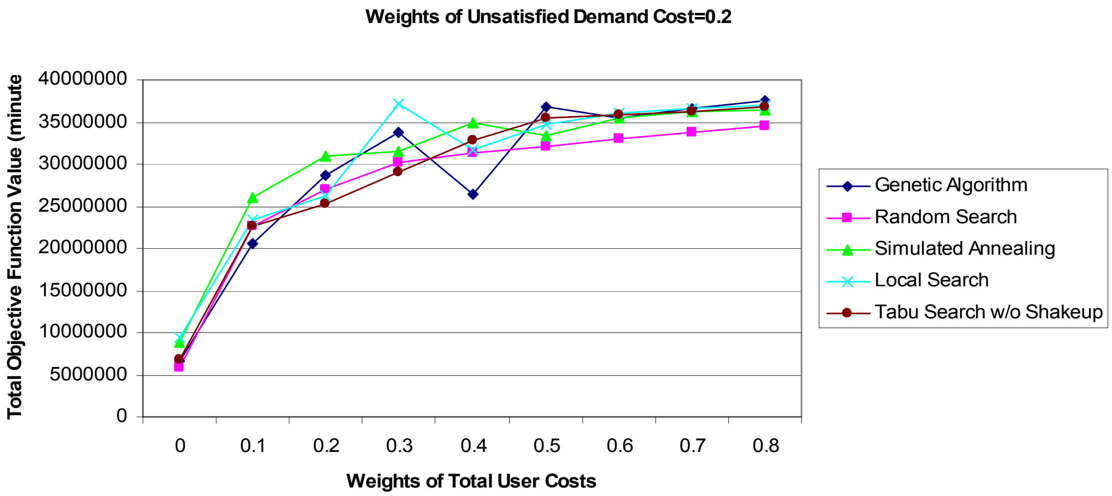


Figure 9.20(2)

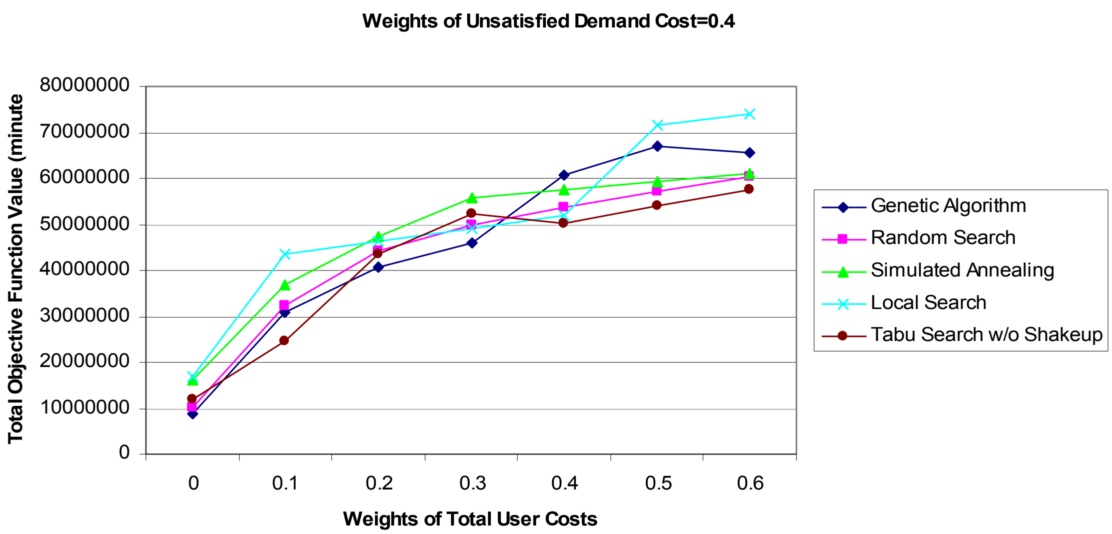


Figure 9.20(3)

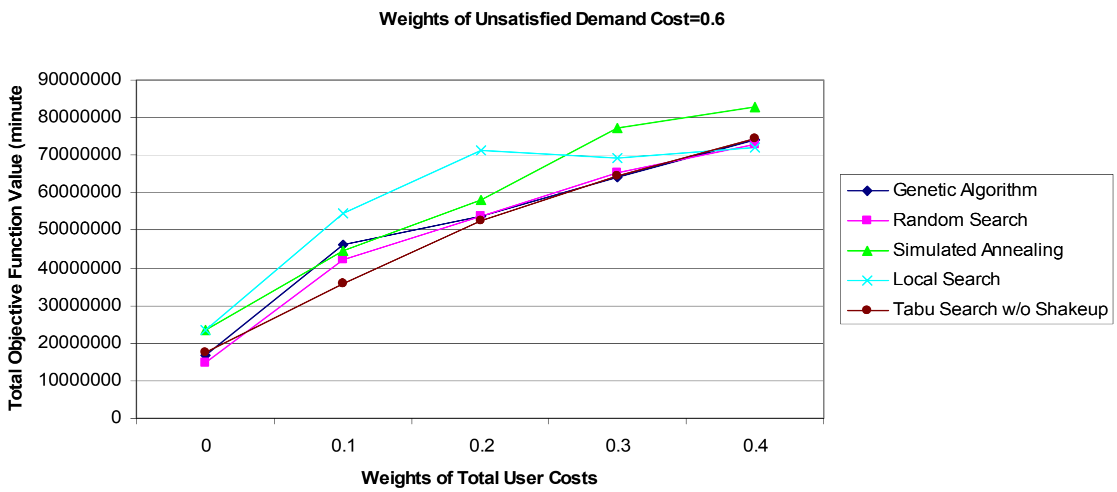


Figure 9.20(4)

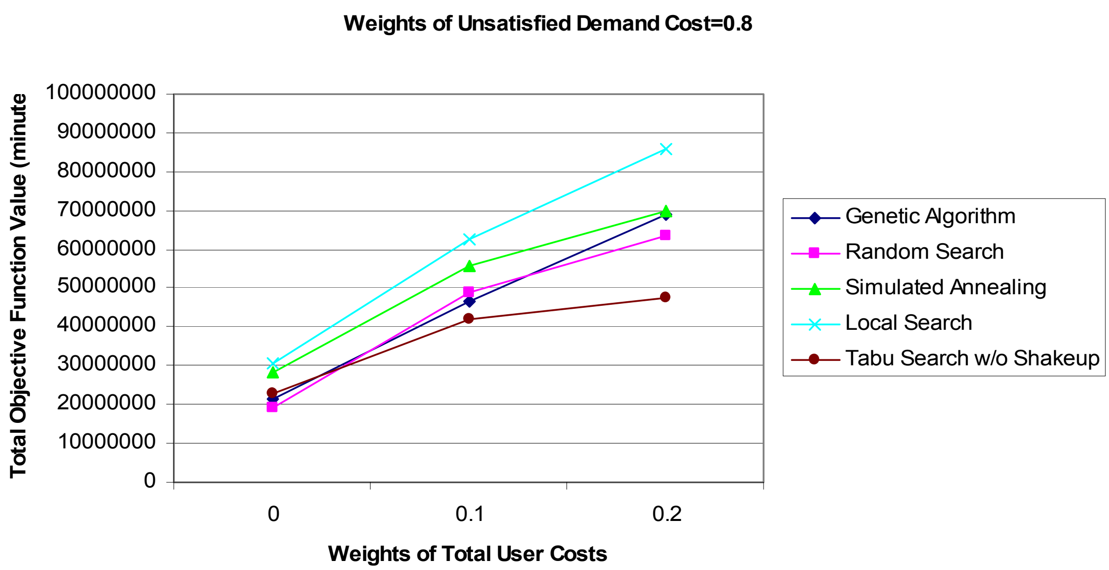


Figure 9.20(5)

Figure 9.20 Heuristic Search Algorithm Comparisons using Medium Network for the BTRNDP with Variable Transit Demand

As can be seen from Figures 9.19 and 9.20, the numerical results for the BTRNDP with variable demand both for the small and medium networks seem to be consistently like that for the BTRNDP with fixed demand using the medium network. That is to say, it seems that the differences among all heuristic algorithms except the local search method are not large and any of these algorithms seems to yield quite consistent solutions. This might suggest that as the variable transit demand characteristics of the BTRNDP are considered for any network size, the heuristic algorithms except the local search method tend to yield transit route network solutions for the BTRNDP whose qualities are almost at the same level. Put another way, when one uses this software or wants to develop a new solution framework, any of these heuristic algorithms (GA, SA, RS and TS) can be used.

9.6 Characteristics of the TRNDP

The characteristics of the BTRNDP are very extensive due to its multi-decision making nature and the variety of parameters and procedures involved. These characteristics might depend upon the network size, the assumption of fixed or variable transit demand, the chosen parameters in the solution process, the chosen algorithm and the chosen weight set level for each component of the objective function. In this sense, it is very hard to generalize all characteristics of the BTRNDP. However, it is expected that in most cases, the BTRNDP characteristics should be similar. Therefore, the following sections provide broad statements about BTRNDP both with fixed and variable transit demand based upon the medium network. Furthermore, since the numerical results based upon weights of 0.4, 0.4 and 0.2 for the user cost, operator cost and unsatisfied demand cost respectively seem to be very representative, these are chosen for presenting related BTRNDP characteristics.

The effect of the number of proposed routes in the transit network solution is investigated by varying it from 1 to 10 and the results for the BTRNDP with fixed and variable demand are provided in Figures 9.21 and 9.22 respectively. The values of each performance measure of the optimal network at each route set size level including the user cost, the operator cost, the fleet size required, the percentage of the satisfied transit demand and the total objective function value are shown in Figures 9.21(1) through 9.21(6).

9.6.1 Effects of Route Set Size for the TRNDP with Fixed Demand

For the BTRNDP with fixed transit demand, the characteristics can be presented as follows. Generally speaking, as the number of routes provided in the network increases, more passengers will be served by transit and therefore, the satisfied transit demand increases. Furthermore, if one assumes fixed transit demand, the percentage of satisfied transit demand also increases as shown in Figure 9.21(5). Also as a result, the unsatisfied demand cost decreases. However, the operator cost tends to increases because the fleet size required for the network generally increases. In addition, the user cost generally increases because more transit users get involved and the total objective function value also increases. The reason might be that although service might be better in some sense (such as more passengers get direct route service) as more routes are provided, the headway might be longer on some routes. Therefore, the transit user cost as a whole might actually increase.

In conclusions, the numerical results in Figure 9.21 indicate that as a whole, as the route set size increases, the solution improved initially because more demand was satisfied and unsatisfied demand costs decrease. However, the lowest objective function value is achieved with 4 routes for this scenario and increases in the fleet size (i.e., operator cost) produces underutilization of routes and does not result in an improved objective function value.

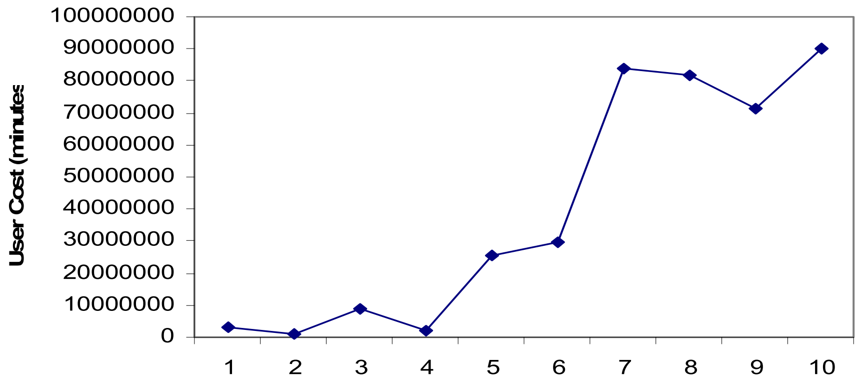


Figure 9.21(1)

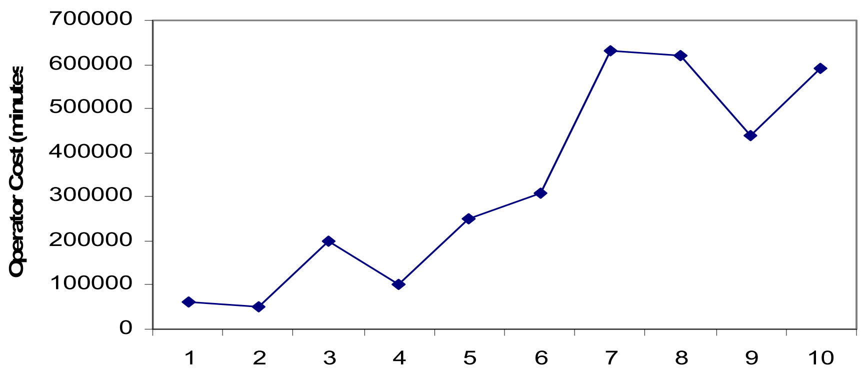


Figure 9.21(2)

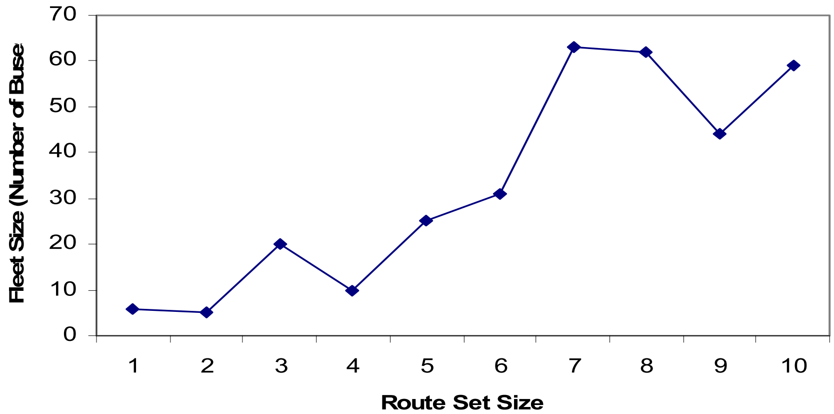


Figure 9.21(3)

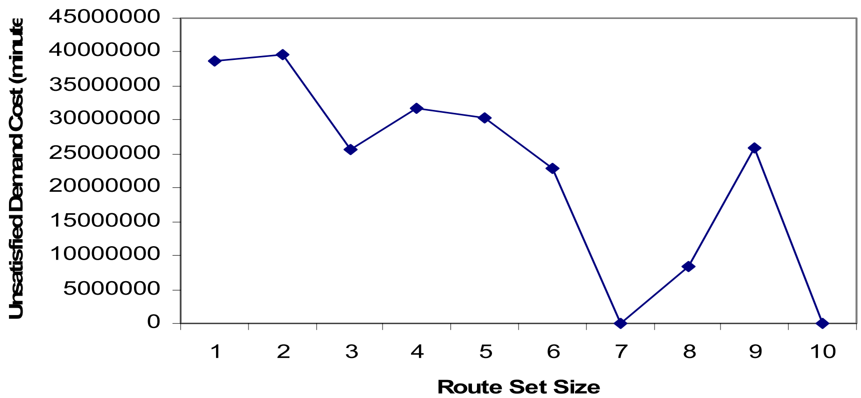


Figure 9.21(4)

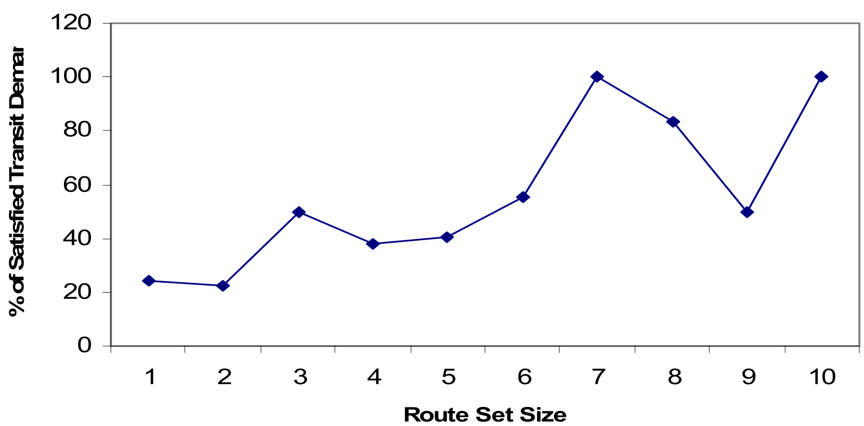


Figure 9.21(5)

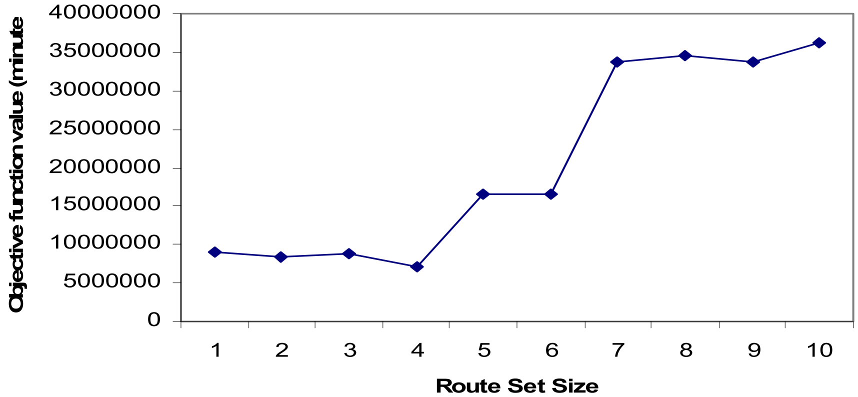


Figure 9.21(6)

Figure 9.21 Effect of Route Set Size on Objective Function and its Components for the BTRNDP with Fixed Demand

9.6.2 Effects of Route Set Size for the TRNDP with Variable Demand

For the BTRNDP with variable transit demand, the characteristics seem to follow a very similar if not the same pattern as that with fixed transit demand and they can be presented as follows. Generally speaking, as the number of routes provided in the network increases, more passengers seem to choose transit over auto mode and therefore, the percentage of satisfied transit demand out of the total travel demand increases as shown in Figure 9.22(5). If one assumes fixed total demand, this might suggest that the satisfied transit demand increases. Also as a result, the unsatisfied demand cost decreases. However, the operator cost tends to increase because the fleet size required for the network generally increases. In addition, the user cost generally increases because more transit users get involved and the total objective function value tends to increase. The reason might be that although more passengers might get direct transit route service as more routes are provided, the headway might be longer on some routes. As the operator cost generally also increases as the number of routes increases, the total objective function is more likely to increase.

In conclusion, the numerical results in Figure 9.22 indicate that as a whole, as the route set size increases, the solution improved initially because more transit demand is assigned to the network and unsatisfied demand costs decrease. However, the lowest objective function value is achieved with 5 routes for the studied network and increases in the fleet size (i.e., operator costs) produce underutilization of routes and does not result in an improved objective function value. This situation is the same for the BTRNDP with variable demand as that with fixed transit demand.

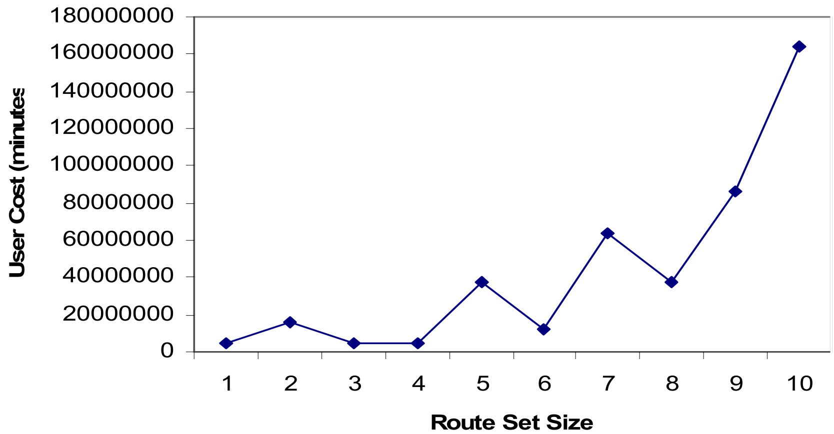


Figure 9.22(1)

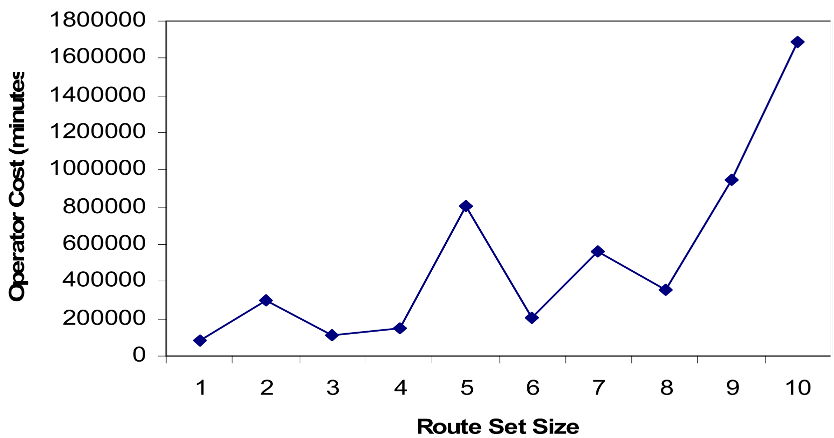


Figure 9.22(2)

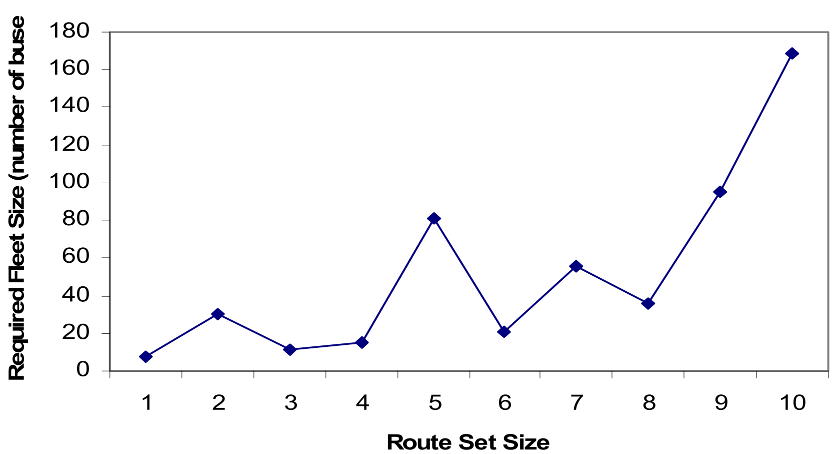


Figure 9.22(3)

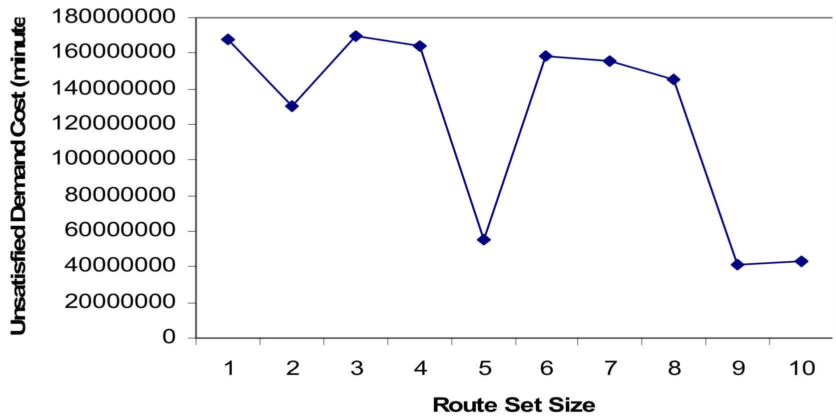


Figure 9.22(4)

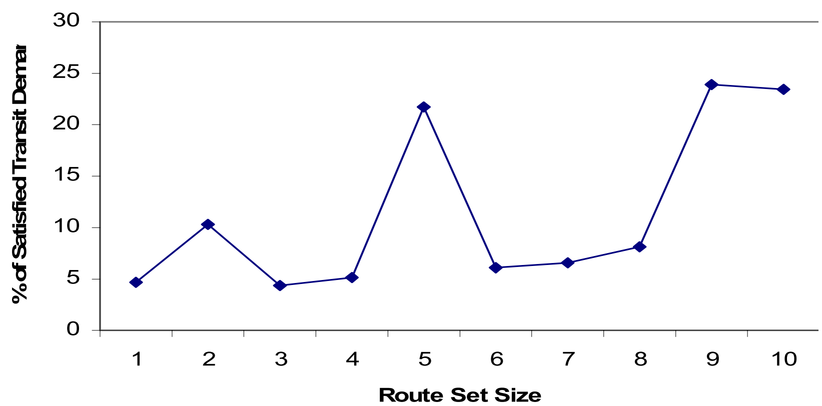


Figure 9.22(5)

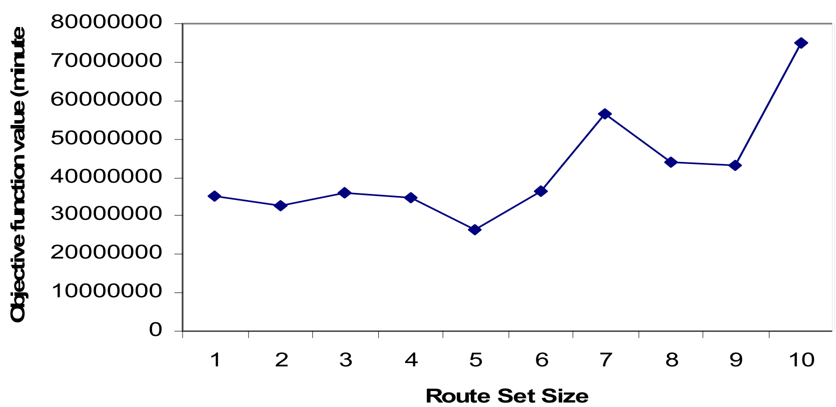


Figure 9.22(6)

Figure 9.22 Effect of Route Set Size on Objective Function and its Components for the BTRNDP with Variable Demand

9.6.3 Comparisons between The TRNDP with Fixed and Variable Demand

9.6.3.1 Characteristic Changes in User Cost, Operator Cost and Unsatisfied Demand Cost

Although it seems that the effect of the route set size on the objective function and its components follows the same pattern for the BTRNDP both with fixed and variable transit demand, there are still some differences between them. That is, the impact of route set size on the objective function including user cost, operator cost, fleet size required, the percentage of satisfied transit demand and total objective function value for the BTRNDP with fixed transit demand seems to be relatively larger than that for variable transit demand. This can be seen by comparing fixed and variable demand figures. In other words, the effect of the route set size for variable transit demand seems to be less distinct than that for the fixed transit demand. This is expected, since the discrete choice modeling structure employed in this research consists of a two-staged model, which has a nonlinear binary logit curve and an inversely proportional model. Both models act as a buffer to reduce the impact of the route set size on the objective function, making the BTRNDP with variable transit demand smoother than that for the fixed transit demand.

9.6.3.2 Comparisons between Optimal Solution Networks in Two Scenarios

One can see from Figures 9.21 and 9.22 that the optimal transit route network with fixed transit demand is different from that with variable transit demand at the 0.4, 0.4 and 0.2 weight set level (user cost, operator cost and unsatisfied demand cost respectively). The optimal solution networks at any weight set level for these two different scenarios are distinct from one scenario to another. As a result, the BTRNDP with variable transit demand should be considered.

9.7 Larger Network Extensions

In previous chapters, the small and medium networks are used for the BTRNDP with fixed and variable transit demand. Numerical results are presented and algorithm comparisons are performed. In this research, the large network as illustrated before is introduced to examine the effect of the network size on the computing speed and to investigate the effect of demand aggregations both on the computing speed and solution qualities compared to that without demand aggregations. In addition, since all heuristic algorithms seem to perform quite efficiently, the random search algorithm is chosen as the representative for all heuristic algorithms here due to its relative simplicity. The following sections present related numerical results.

9.7.1 Effects of Network Size on Computing Speed

As mentioned in Chapter 4, as network size increases, the computational time for the BTRNDP either with fixed or variable transit demand grows exponentially. Figure 9.23 and 9.24 present this effect for the BTRNDP with fixed and variable transit demand respectively (The 1, 2 and 4 in the x-axis indicate the small, medium and large networks respectively). One can see from these two Figures and conclude that the effect seems to be consistent that the computation time does change nearly proportionally (i.e., almost linearly rather than exponentially) as the network size increase from the small to medium to large network if the heuristic algorithms are used. For example, the computation time for the small network in this research is about 30 minutes, about 8 hours for the medium network and about 70 hours for the large network. However, if the exhaustive search method is used, as one can see from Figure 9.14(4), the computation time grows exponentially as the network size increases. From another perspective, this might suggest that the BTRNDP is an NP-hard problem and indicates the validity and necessity of employing the heuristic search algorithms to solve the BTRNDP.

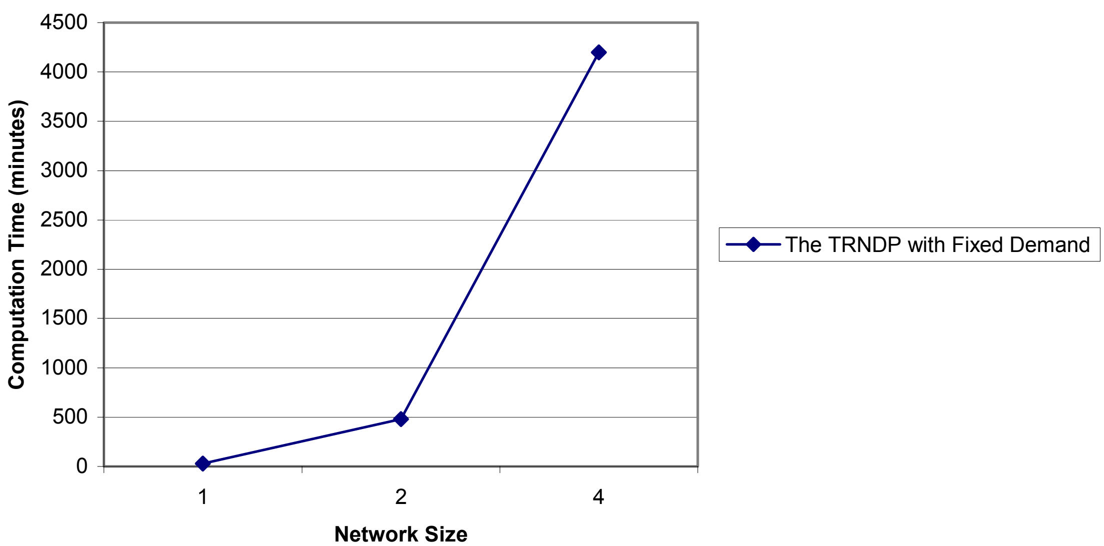


Figure 9.23 The Effect of Network Size on Computation Time for the BTRNDP with Fixed Transit Demand

In addition, it should be noted that for any network size, the computation speed for the BTRNDP with fixed transit demand is always slightly faster than that with variable transit demand if chosen parameters are the same under both scenarios. This is expected because one more procedure, (i.e., the transit demand equilibration procedure) and the two-staged BLM-IPM model are added in the BTRNDP with variable transit demand to accommodate this variable transit demand characteristic. However, as shown in Figure 9.23 and 9.24, the differences in the computing speed between these two scenarios are not very significant because they share the same solution framework except the added transit demand equilibration procedure.

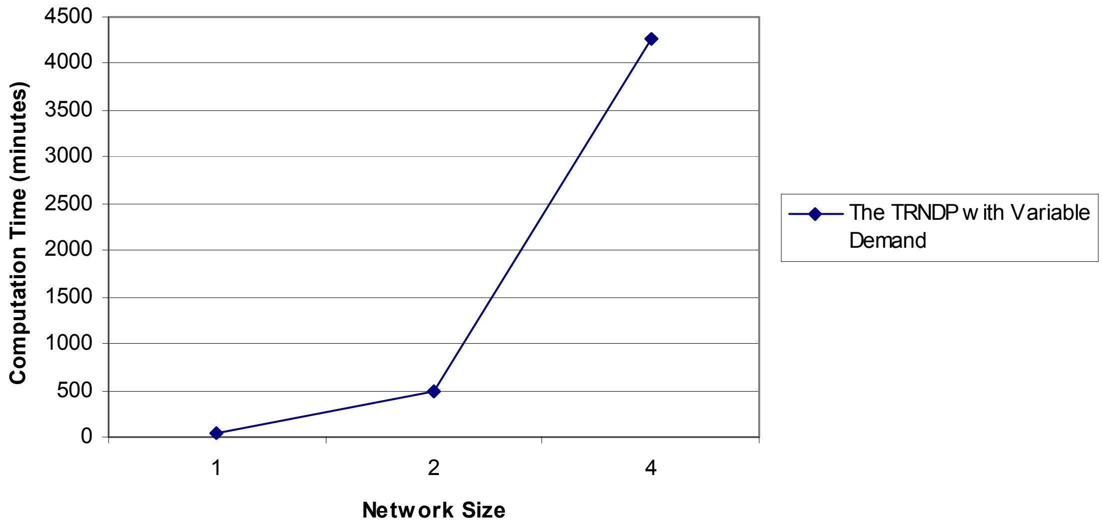


Figure 9.24 The Effect of Network Size on Computation Time for the BTRNDP with Variable Transit Demand

9.7.2 Effects of Demand Aggregation

The impact of demand aggregation on the computation time and solution quality was examined using the large network as an example. Basically speaking, the demands are aggregated at the distribution node level so that there is only one possible bus stop for each centroid. Additionally, note that the chosen location might be obvious (such as there is only one entrance/exit for a particular travel zone) or less obvious (in case two or more than two locations can be chosen). For example, one could set the demand aggregation point to the bus stop that is on the main street. The locations chosen for bus stops could have a great impact on the solution quality. An example of demand aggregations for the large network is shown in Figure 9.25 and this network is used to examine the effect of demand aggregation on the computation time and solution quality.

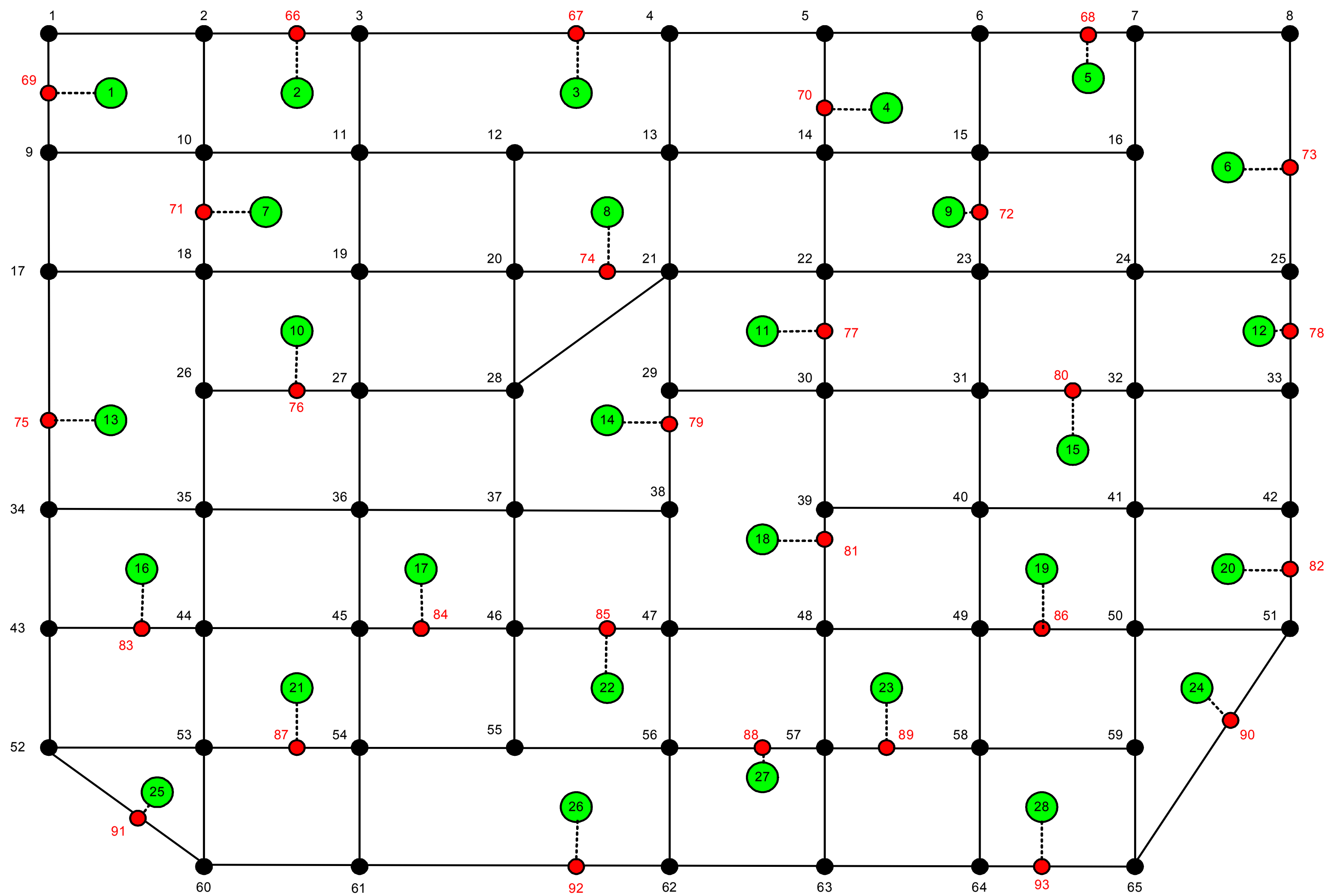


Figure 9.25 Large Network Graphical Representation with Demand Aggregations

9.7.2.1 Impact on Solution Quality

The following numerical results are obtained by setting the weight of the unsatisfied demand cost to be 0.4 and varying the weights of the user and operator cost respectively both with fixed and variable transit demand. Note that the results are very representative because similar results have been obtained at other weight set levels. Figure 9.26 and 9.27 show the effect of demand aggregations on solution quality both with fixed and variable transit demand respectively.

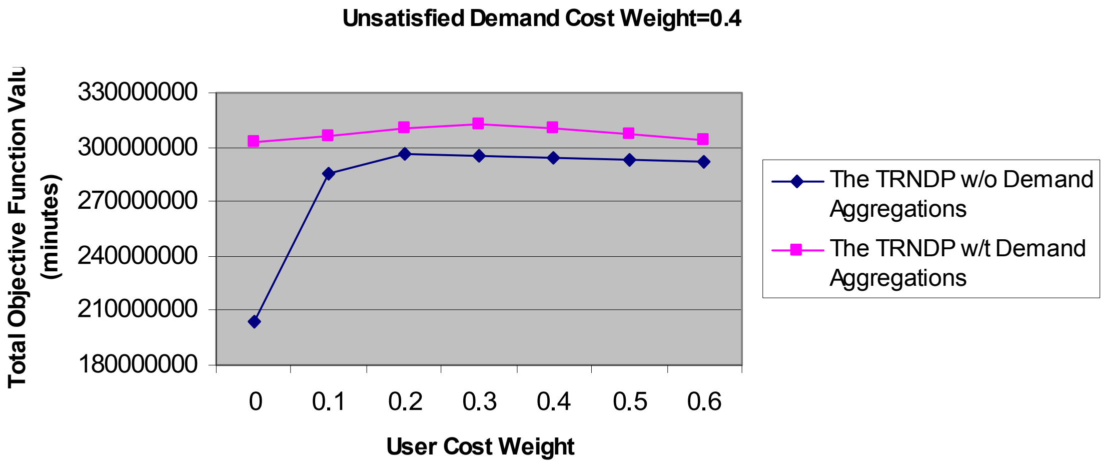


Figure 9.26 The Effect of Demand Aggregations on Solution Quality for the BTRNDP with Fixed Transit Demand

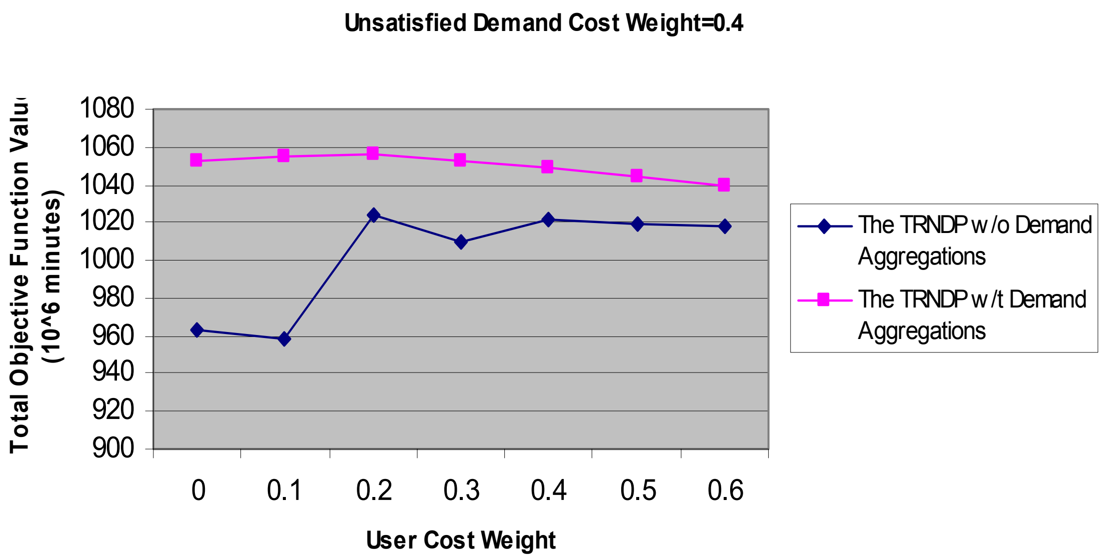


Figure 9.27 The Effect of Demand Aggregations on Solution Quality for the BTRNDP with Variable Transit Demand

As can be seen from both figures, the objective function values are smaller (more desirable) without demand aggregation than with demand aggregation under both scenarios. This is expected because the BTRNDP with demand aggregation is a constrained optimization version of the BTRNDP without demand aggregation. Therefore, the optimal objective function that can be achieved in the latter case is usually less than that in the former case (it will be always less if the BTRNDP is a convex problem). If one tends to emphasize the solution quality and wants an optimal solution network, the BTRNDP without demand aggregation is the better choice.

However, the better solution quality from the BTRNDP without demand aggregation comes at the cost of a slower computation speed as shown in the following sections.

9.7.2.2 Impacts on Solution Efficiency

Figure 9.28 presents the effect of demand aggregation on computation time for the BTRNDP with fixed and variable transit demand. The BTRNDP with demand aggregation requires much less computation time than that without demand aggregation under both scenarios. For example, the computing time for the BTRNDP with fixed demand and demand aggregation is about 3 hours while that for the BTRNDP with fixed demand and without demand aggregation is about 72 hours. This is expected because the BTRNDP with demand aggregation has a much smaller solution space than that without demand aggregation and the number of overlapping nodes among different routes is much less. Therefore, the computation time for the NAP for each proposed solution network is much less for the BTRNDP with demand aggregation. Therefore, the total computation time in the former case is much less the latter case.

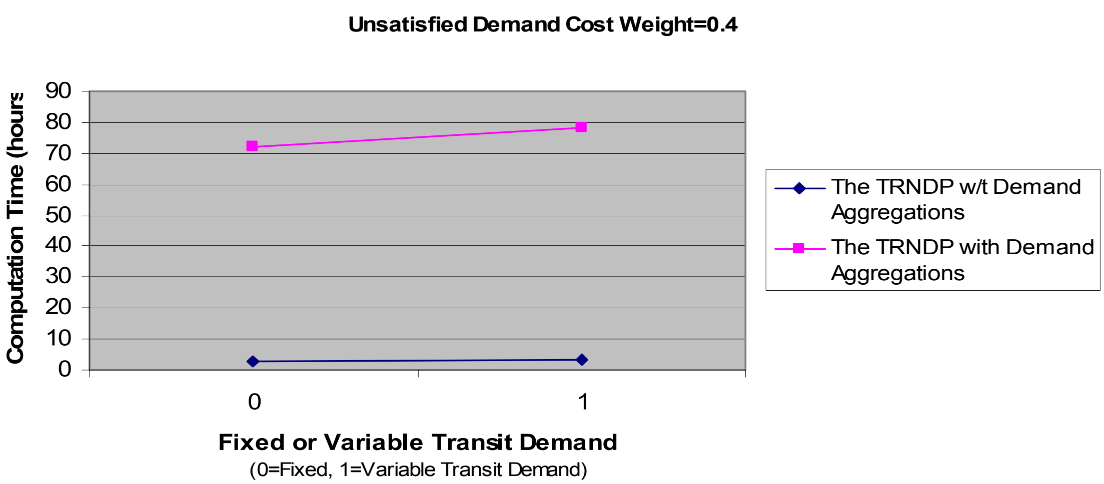


Figure 9.28 The Effect of Demand Aggregations on Computation Time for the BTRNDP with Fixed Transit Demand and that with Variable Transit Demand

9.8 The Redesign of the Existing Transit Network Issues

9.8.1 Design Strategy and Corresponding Implementation Changes

Previous chapters mentioned the redesign of existing transit networks. This situation is very common because in the early stages of the BTRNDP, transit planners had only their experience and judgment to design the current transit route network. However, these networks contain inefficient characteristics and redesign of the existing transit network is potentially beneficial.

Generally speaking, if one redesigns a transit route network, one would usually keep some routes that are very efficient as the network skeleton and find some optimal routes to serve the transit demand that cannot be satisfied by these kept skeleton routes. Moreover, this problem is investigated under the situation that the demand matrix has already achieved equilibrium (at least for demand between any node pairs on the kept existing routes). Therefore, essentially speaking, the redesign of the transit route network can be studied under the assumption of fixed transit demand. (Although transit demand on other demand pairs or routes might be variable, for simplicity, fixed transit demand is assumed here). Based on these considerations, some minor modifications should be made to the current C++ codes. That is to say, the whole skeleton remains unchanged. However, at each iteration when the heuristic search algorithm is used to propose the new transit route network, one should always keep the efficient routes and store them in a specific location in the arrays that are used to store all the proposed solution routes. This way, the kept routes will be always included in the optimal transit solution route set at any route set size level.

Based on the above descriptions, the computer codes were modified and tested. The numerical results are presented as follows.

9.8.2 Numerical Results

The following uses the random search method for the BTRNDP with fixed demand for the medium network as a typical example for the redesign of the existing transit network. Note that the genetic algorithm is not chosen because its selection procedure is expected to change the proposed transit route network according to the objective function values of previous solution networks. Keeping some routes, in some sense, is the same as adding constraints and this might destroy the characteristics of the genetic algorithm and therefore might result in misleading or at least unrepresentative numerical results. Therefore, the random search rather than the genetic algorithm is used. Also note that the weight set level 0.4, 0.4 and 0.2 as described before are used and the numerical results are given in the following sections.

Figure 9.29 provides the current optimal transit route solution network for the medium network with fixed demand using the random search method. Note that the headway is 20 minutes on the black route and 19 minutes on the pink route respectively.

In addition, Figure 9.30 and 9.31 graphically show the optimal transit route solution network configuration after keeping one good (efficient) route and that after keeping one bad (inefficient) route respectively. Note that the definition of "efficient" and/or "inefficient" might affect the numerical results. The two examples as shown here are based on the author's judgment and the numerical results are presented as follows.

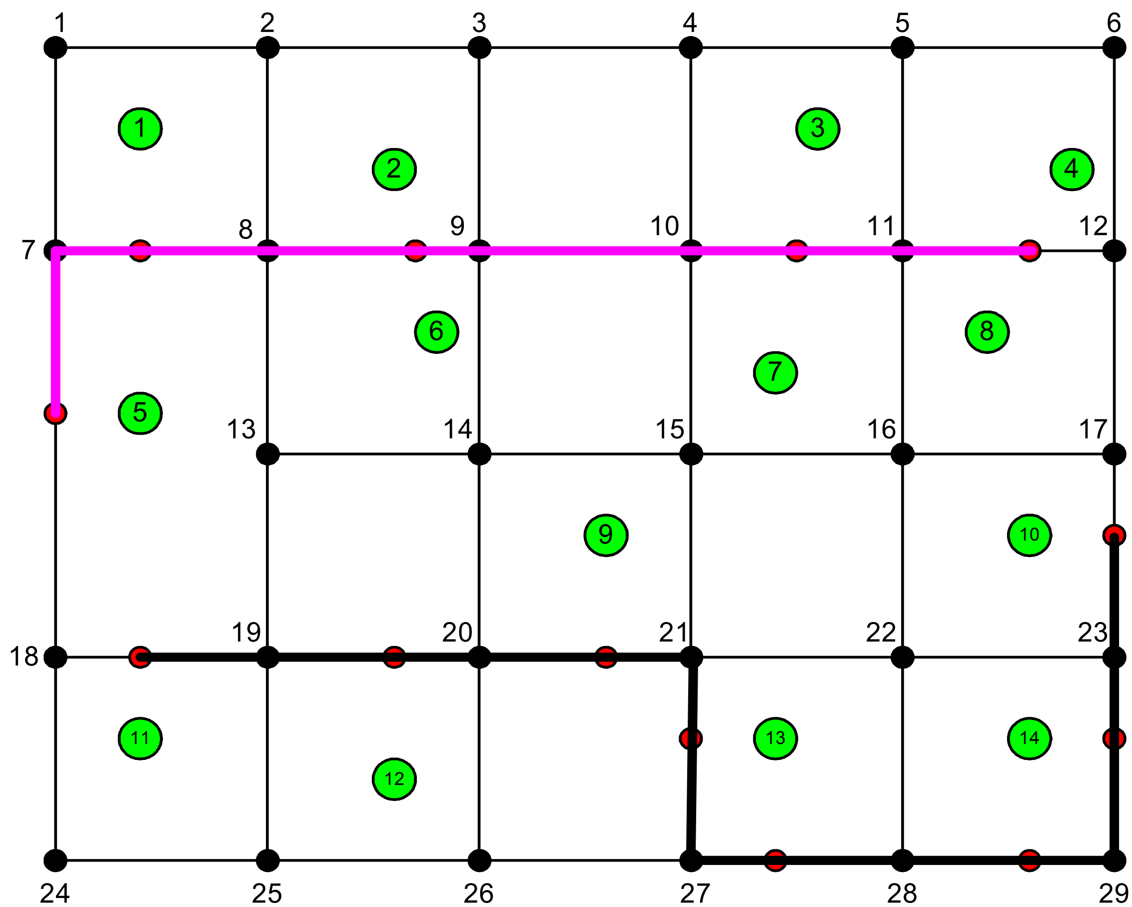


Figure 9.29 Current Optimal Transit Route Solution Network for the Medium Network with Fixed Demand

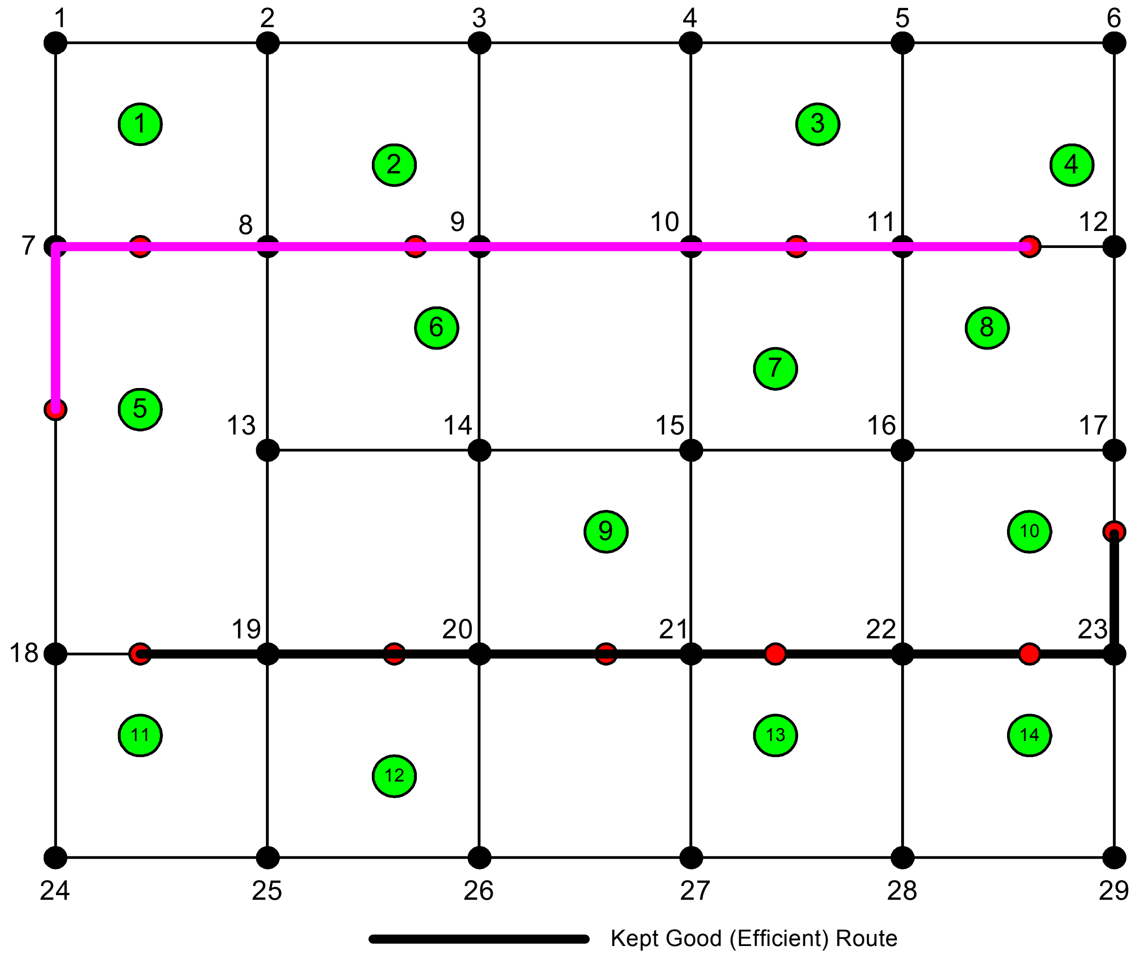


Figure 9.30 The Optimal Transit Route Solution Network Configuration after Keeping Good (Efficient) Routes

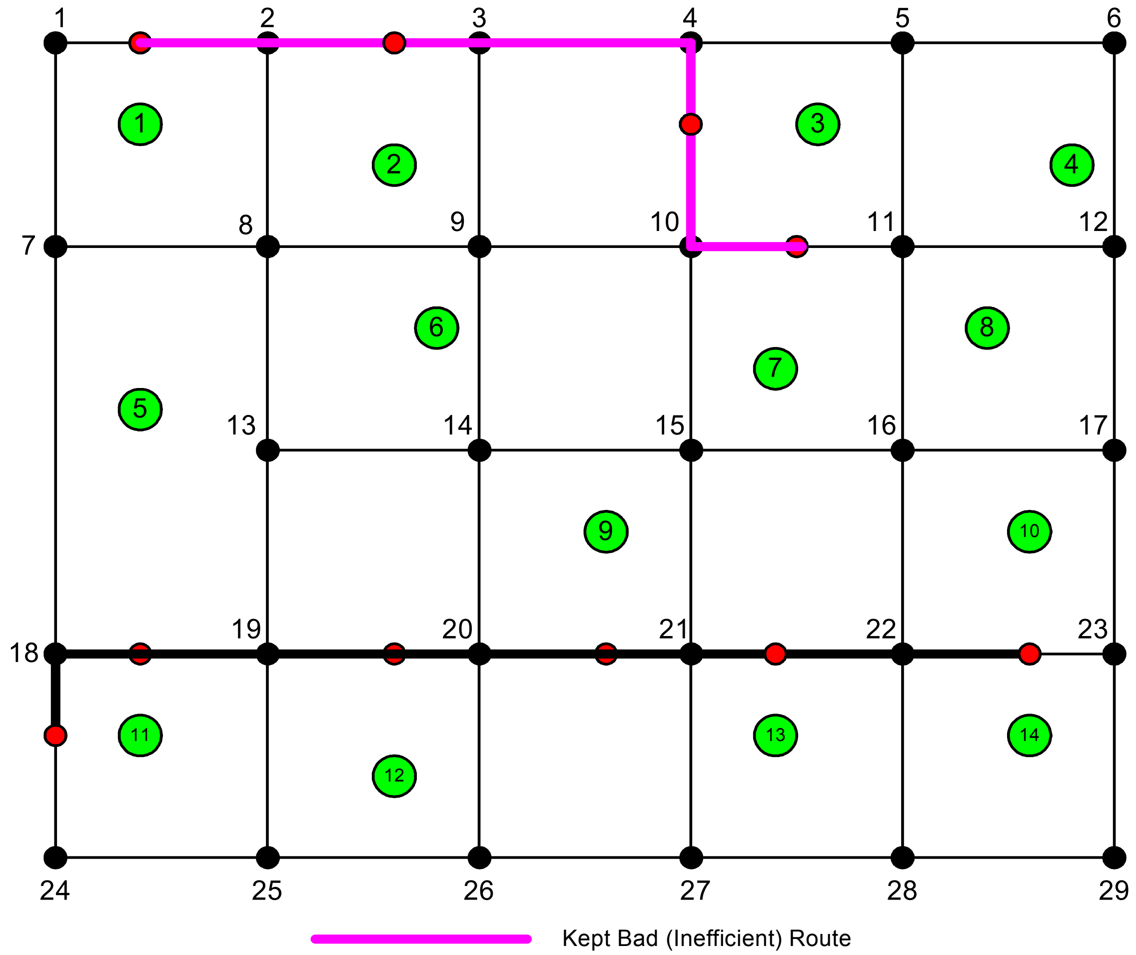


Figure 9.31 The Optimal Transit Route Solution Network Configuration after Keeping Bad (Inefficient) Routes

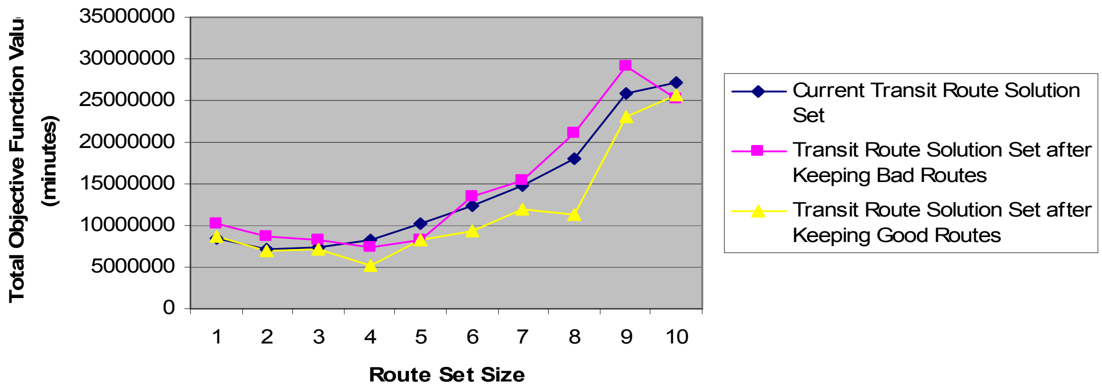


Figure 9.32(1)

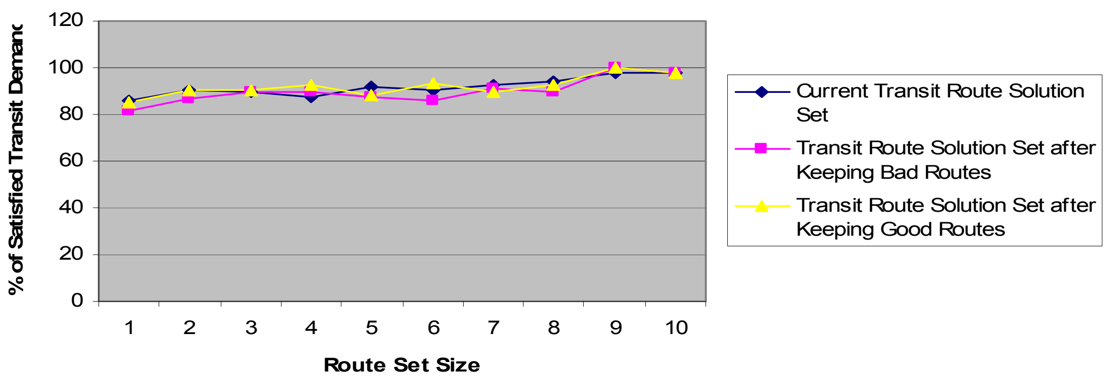


Figure 9.32(2)

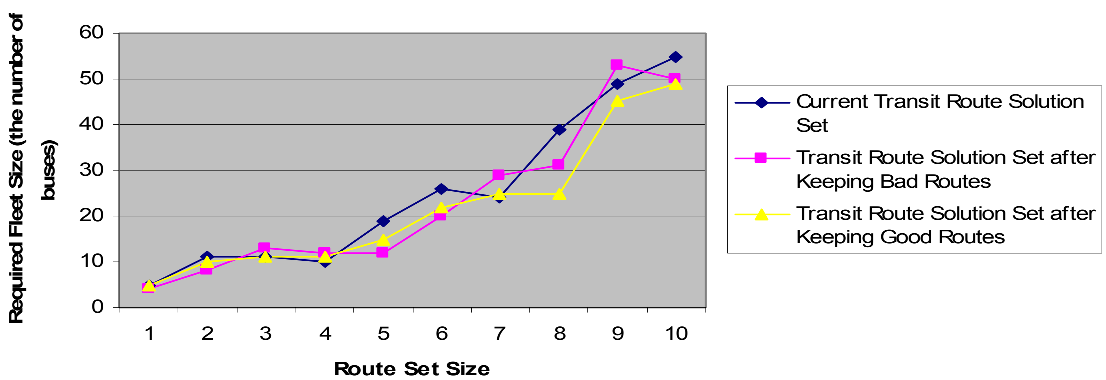


Figure 9.32(3)

Figure 9.32 The Effect of Keeping Good (Efficient) and Bad (Inefficient) Routes on Solution Quality for the BTRNDP with Fixed Transit Demand

As can be seen from Figure 9.32, the lowest objective function value can always be achieved for the BTRNDP with fixed transit demand after keeping good (efficient) routes for any route set size. However, the BTRNDP after keeping bad (inefficient) routes seem to result in the worst "optimal" transit route structure for almost every route set size. Based on these results, one can conclude that the redesign of the existing transit route network might depend on the characteristics of the kept routes. Put another way, it might depend on the specific route that one wants to keep. If it's very efficient, solution quality increases and using this algorithm will yield a very good solution network.

9.9 Summary and Conclusions

This chapter focuses on the algorithm implementation and numerical results for the proposed solution methodologies. Details of the C++ program codes that were developed to implement the six algorithms for the BTRNDP both with fixed and variable transit demand are presented.

Three experimental networks are designed and successfully tested for the BTRNDP. The network representation and algorithm implementation details of the C++ programming codes for the proposed methodology to the BTRNDP are presented. Comprehensive numerical results of the six proposed solution methods, including the genetic algorithm, local search, simulated annealing, random search, tabu search algorithm and exhaustive search method, are presented for the BTRNDP with fixed and variable transit demand respectively using three example networks. Sensitivity analyses for each algorithm are conducted and algorithms are compared based on the multi-objective decision making nature of the BTRNDP. The characteristics underlying the BTRNDP are identified and effects of route set size on the objective function and its components are examined and compared. In the large network extensions, effects of network size and demand aggregation on the computation speed and solution quality are studied. Redesign of the existing network issues is discussed and related numerical results are presented. The next chapter concludes this research with a summary and future directions for research are also given.

CHAPTER TEN
SUMMARY AND CONCLUSIONS

10.1 Introduction

Public transit has been widely recognized as a potential way of reducing air pollution, lowering energy consumption, improving mobility and lessening traffic congestion. Designing an operationally and economically efficient bus transit network is very important for the urban area's social, economic and physical structure.

The primary objective of this research is to develop a robust optimization tool using systematical heuristic approaches and demonstrate these tools using comprehensive experimental networks for the computer-aided design of the optimal transit route networks. The TRNDP involves the minimization of some generalized costs subject to a variety of constraints, which reflect system performance requirements and/or resource limitations. The decision one seeks to make is the determination of a transit route network configuration including a set of transit routes and their associated bus frequencies. Although several research efforts have examined this subject in the past decade, the present work is the first effort to apply systematic heuristic solution approaches to the TRNDP.

The following sections are organized as follows. In section 10.2, the principal features of the solution approaches designed for the TRNDP are reviewed and a summary of conclusions for the numerical results derived from computational tests is discussed. Section 10.3 presents a brief discussion of the limitations of the current approaches and possible directions for further research are also given.

10.2 Summary and Conclusions

As mentioned, the optimal transit route network design problem addressed in this research involves finding a bus transit route network configuration and associated service frequencies that achieve a desired objective with a variety of given constraints. The literature describing previous solution approaches to the TRNDP has been reviewed. As mentioned by several researchers including Baaj (1990), six main sources of complexity often preclude finding a unique optimal solution for the TRNDP and these are the following: (1) great difficulty in defining the decision variables and expressing the objective function; (2) non-convexities and non-linearity are involved in the cost associated with the transit network configuration; (3) combinatorial complexity arises from the discrete nature of the route design problem, making the TRNDP an NP-hard one; (4) many important tradeoffs among conflicting objectives need to be addressed, making the TRNDP an inherently multi-objective decision making problem; (5) spatial layout of routes makes it very hard to design an acceptable and operationally feasible set of routes with the need to address many important design criteria; and (6) the nature of variable transit demand even with a given total travel demand makes the already-difficult TRNDP more complex. These sources of complexity render the solution search space computationally intractable and the computational burden of the problem grows exponentially with the size of the studied transit network.

Previous approaches that were used to solve the TRNDP can be classified into three categories: practical guidelines and ad hoc procedures; analytical optimization models for idealized situations; and meta-heuristic approaches for more practical problems. Principal shortcomings of previous approaches include the failure to consider the TRNDP where the demand should be distributed among the bus stops serving an origin (destination) rather than to the centroid (single point); failure to address the inherent multiple objective nature of the transit route network design problem; failure to consider the transit route network design problem in the context of variable demand; failure to incorporate practical service guidelines; and failure to consider essential aspects of the problem in the solution process. Building on several previous approaches, the solution methodology proposed in this research includes the following major features: 1) Systematic heuristic methods for transit route generation and improvement; 2) a transit network evaluation model to compute a variety of system performance measures; 3) systematic use of context-specific knowledge to guide the search technique; and 4) accommodate different design requirements for the TRNDP under different scenarios.

The proposed solution approach consists of three main components: an Initial Candidate Route Set Generation Procedure (ICRSGP) that generates all feasible routes incorporating practical guidelines that are commonly used in the bus transit industry; a Network Analysis Procedure (NAP) that achieves the transit demand, assigns the transit trips, determines the service frequencies on each route and computes many performance measures; and a Heuristic Search Procedure (HSP) that scientifically guides the feasible solution generation and search process. In the third step, five heuristic algorithms, including the Genetic Algorithm, Local Search, Simulated Annealing, Random Search and Tabu Search Methods, along with the Exhaustive Search Method as a benchmark to examine the efficiency and measure the quality of the solutions obtained from these heuristic algorithms for the TRNDP with small network, are applied to select an optimum set of routes from the huge solution space.

Comprehensive experiments are performed, sensitivity analyses for each algorithm are conducted and algorithms are compared based on the multi-objective decision making nature of the TRNDP. The numerical results showed that the values in the optimal parameter set for each algorithm changes slightly as the network size changes, suggesting that the optimal parameter set might also depend on specific network including the network size and its configuration. Although one can use the optimal values in the parameter set for each algorithm as a general guideline, it is recommended to perform sensitivity analyses for the employed algorithm for the studied specific network. The exhaustive search method is examined for the TRNDP using a small network. The results showed that although this method can find the global optimum, it requires extremely long computation time to get this unique optimal solution. Since the computation time for this method grows exponentially, the heuristic search algorithms, rather than the exhaustive search method, are recommended for use for the TRNDP with the large networks.

To determine which tabu search version is most suitable for the TRNDP both with fixed and variable demand, tabu search algorithms are compared from a multi-objective decision making perspective using different sized networks. The results showed that for any network size, there are no big differences among the three tested algorithms so that either one can be used for the TRNDP. Similar conclusions result from the comparison among all five heuristic algorithms, including the genetic algorithm, local search, simulated annealing, random search and tabu search. In other words, the heuristic algorithms except the local search method tend to yield transit route network solutions for the TRNDP whose qualities are almost equal. Therefore, any of these heuristic algorithms (such as the GA, SA, RS and TS) can be used for the TRNDP.

Characteristics underlying the TRNDP are also identified and effects of the route set size on the objective function and its components are examined and compared. The numerical results show that for the TRNDP both with fixed and variable demand, as the route set size increases, the solution improves initially because more demand is assigned to the network routes and unsatisfied demand costs decreases. However, after a threshold point at the optimal route set size, increases in the fleet size (i.e., operator cost) produce underutilization of routes and does not result in an improved objective function value. Moreover, the effect of route set size with variable transit demand seems to be less distinct than for fixed transit demand due to its inherent discrete choice modeling structures. Also, the solution networks for the TRNDP with fixed demand (including the transit route network structure and service frequencies) are distinct from those with variable demand. As a result, the TRNDP with variable transit demand is recommended to be used for real-world applications.

In the large network extensions, the effects of network size and demand aggregation on computation speed and solution quality are studied. The numerical results indicated that as network size increases, the computation time for any heuristic algorithm increases dramatically, suggesting the necessity of employing an heuristic search algorithm to solve the TRNDP. In addition, the TRNDP without demand aggregation produces significantly more desirable objective function values than that with demand aggregation. The redesign of existing networks is discussed and related numerical results are also presented. It is concluded that the redesign of the existing transit route network might depend on the characteristics of the kept routes. That is to say, if the retained route skeleton is very efficient, solution quality increases and using the employed solution methodology will yield a very good solution network.

The principal unique contributions of this work result from the following aspects:

1) In the present version, the route generation procedure is mainly used for a long range planning tool which generates a new set of bus transit routes to provide associated service for transit users. However, the methodology developed in this work can be also used to support short or medium range planning because it allows for the transit planner to specify some routes based on his/her experience or keeping a subset of routes of the existing transit system and then design or modify the transit network taking into account the presence of such predetermined routes. This point can be seen from the numerical results as already shown in Chapter 9. This is extremely important because any transit planning tool intended to be used in practice should offer this capability and the current developed transit route network design optimization tool can help meet this requirement. Furthermore, the current optimization tool can be easily modified to be used for subway transit route system design, where there are fewer routes than bus networks, higher service frequencies, larger passenger transit capacities and more transferring activity among routes. Essentially, the subway network design problem is just a variant of the optimal transit route network design problem and therefore with some minor modifications, the overall framework would be applicable in this context. Also, the developed heuristic methodology can also be used for many other research topics related to transportation routing problems such as locating a single light rail line although minor modifications still need to be performed. Lastly, the current version can also be extended to integrated design or redesign of bus and rail transit route network systems.

2) It considers the TRNDP in a more real world general context where the transit demand is distributed among bus stops serving an origin (destination) rather than aggregating them to a zone centroid (single point). This enables more applicability of the transit assignment process when the transit transfers occurred.

3) It can explicitly address the TRNDP as a multi-objective decision problem and generates different sets of routes and determines their associated service frequencies corresponding to built-in tradeoffs between conflicting objectives.

4) A computer-based route generation procedure that can explicitly incorporate several practical guidelines and industry rules of thumb and form a complete solution space consisting of all feasible routes while doesn’t heavily rely on user experience or the demand matrix. The optimal transit route network is described using the network evaluation procedure and heuristic search methods.

5) A transit network analysis procedure with the following important features: i) It can accommodate different route design requirements for the TRNDP under two different scenarios, namely both with fixed and variable transit demand and it uses an iterative procedure to obtain total transit trip demand for the TRNDP with variable transit demand; ii) It uses a two-staged binary logit model and inversely proportional model to reflect the choice process for the trip makers and compute the number of transit users; iii) It incorporates a trip assignment procedure which explicitly considers transfer and long-walk related characteristics and assigns trips among related transit routes. This enables

the evaluation and design of the TRNDP under a much more real situation; iv) It computes system performance measures reflecting the quality of service, user cost, the resources required by the operator and unsatisfied demand cost and can be used as a sensitivity analysis tool for system performance measures and a variety of variables and parameters; v) It uses C++ programming language as the software development tool, which greatly facilitates the implementation of the path enumeration, route evaluation and solution search techniques.

6) It uses systematic heuristic algorithms to scientifically guide the route set solution generation and search process and identify the most applicable algorithms for the TRNDP of different network size and different service deign requirements in each scenario. The results demonstrate the robustness of these algorithms as an optimization tool for the TRNDP.

7) The developed transit route optimization tool explicitly considers the variable relationship between the generated set of bus routes and the transit trip demand matrix, and this makes this work much more applicable than any previous.

10.3 Directions for Further Research

In this section, some of the limitations of current solution approaches for the TRNDP are presented and directions for further research are also discussed.

The solution approaches tackle the locations of bus stops around a specific transit demand zone along any route as exogenous data (i.e., they are given through an input file). However, in the real-world bus planning process, the optimal number and locations of bus stops should be determined as part of the TRNDP rather than predetermined. The incorporation of determining bus stops would add realism for the solution approach. The literature shows that possible solution approaches for this consideration include an integrated application of the dynamic programming and heuristic algorithm. In addition, after setting the bus stops, the work of driver scheduling and bus scheduling on each route in the whole transit route network should be performed because the transit trip assignments are essentially schedule-based once the schedules are set. Existing literatures indicate that heuristic algorithms, such as tabu search methods, can be used for driver scheduling and bus scheduling, while branch and bound algorithm-based programming can be used for solving the schedule-based transit trip assignments. Furthermore, it is noted that proposing the bus routes, determining the number and locations of bus stops, setting the bus and driver schedules and assigning the transit trip based on the determined schedules should all be included as sequential decisions in the TRNDP. Also, current algorithms other than the employed heuristic ones for solving combinatorial algorithms, such as set-covering algorithm, can also be considered for solving the TRNDP. As extensions of these works, they might be potentially good and possibly produce quality solutions. To make the current solution approach more complete and usable, these work need to be added.

The capability to display the solutions of transit route sets represented by node connectivity lists is extremely important and useful for the TRNDP. Put another way, if the changes (at least the improvements) in the current optimal solution set of routes' layout can be showed graphically and dynamically along the searching process, the transit planners are therefore able to see an instant picture of route layouts and the corresponding user and operator cost as well as the measures of performance of the developing network dynamically. Also, the graphical display technology for several network descriptors would enable them to develop a 'feel' for the performance of the route design and quickly notice the sensitivity of the resulting solutions to different user input parameters. Especially when the solution approaches are applied to different time periods, the operations on the generated solution route set, such as identifying which routes to disconnect or detecting when the frequency of service on some routes to increase or decrease, a graphical display of the sets of routes' layouts is almost necessitated.

The solution approach requires further testing on different transit networks and different transit demand matrices. The solution approach provides alternative design features that are applicable to the TRNDP both with fixed and variable transit demand. However, the conclusion remarks about the solution approached are derived from the numerical results using limited designed networks. In addition, the sensitivity analyses parts are performed sequentially although one should try all combinations of all the parameters to determine the optimal parameter set for each algorithm. To make these results more conclusive, more comprehensive (especially as large as the real-world sized) experimental networks are still needed for testing these algorithms.

Lastly, to make the solution approach more versatile, other service choice dimensions need to be incorporated. Examples of these service choices that might improve the design performance of bus transit systems can be seen from the concepts of express bus service that serves two bus terminals non-stop or with limited stops in the middle and the transit transfer center that facilitates transit transfers. In this regard, some minor modifications to the solution approach are needed to help achieve these requirements.

The numerical results were tested using the computers available with the Microsoft Windows XP System, 1.03GHz clock speed, 512MB of RAM and a 35GB hard drive. As hardware improves, the authors believe that the heuristic algorithms and the optimization software derived from this research will become more viable and practicable for real sized large network. As a result, an operationally and economically efficient optimal transit route network with associated service frequencies can be obtained, making the development of our neighborhoods more sustainable. The reduced air pollution, lowered energy consumption, improved mobility and less traffic congestion can be beneficial for the social, economic and physical structure of our next generations' world.

APPENDIX

The input files for all designed three experimental networks are included as follows:

Small Network:

Network.dat:

7	35	82	(total number of zones			total number of nodes		total number of arcs)
1	2	3	4	5	6	7	(Zone_IDs)	
1	0	(Node_ID		"0" means this node is an intersection node)				
2	0							
3	0							
4	0							
5	0							
6	0							
7	0							
8	0							
9	0							
10	0							
11	0							
12	0							
13	0							
14	0							
15	0							
16	1	1	(Node_ID		this number of zones this node is in			Zone_IDs)
17	1	1						
18	2	1	2					
19	2	1	3					
20	1	2						
21	2	2	5					
22	2	2	3					
23	1	3						
24	2	3	6					
25	2	3	4					
26	1	4						
27	2	4	7					
28	1	4						
29	1	5						
30	1	5						
31	2	5	6					
32	1	6						
33	2	6	7					
34	1	7						
35	1	7						
1	1	16	(Link_ID		Link_origin		Link_destination)	
2	1	17						
3	2	16						
4	2	19						
5	2	23						
6	3	23						
7	3	25						
8	3	26						
9	4	26						

10	4	28
11	5	17
12	5	18
13	5	20
14	6	18
15	6	19
16	6	22
17	7	11
18	7	22
19	7	24
20	8	24
21	8	25
22	8	27
23	8	33
24	9	27
25	9	28
26	9	35
27	10	20
28	10	21
29	10	29
30	11	7
31	11	21
32	11	31
33	12	29
34	12	30
35	13	30
36	13	31
37	13	32
38	14	32
39	14	33
40	14	34
41	15	34
42	15	35
43	16	1
44	16	2
45	17	1
46	17	5
47	18	5
48	18	6
49	19	2
50	19	6
51	20	5
52	20	10
53	21	10
54	21	11
55	22	6
56	22	7
57	23	2
58	23	3
59	24	7
60	24	8
61	25	3
62	25	8
63	26	3
64	26	4
65	27	8
66	27	9
67	28	4
68	28	9

69	29	10
70	29	12
71	30	12
72	30	13
73	31	11
74	31	13
75	32	13
76	32	14
77	33	8
78	33	14
79	34	14
80	34	15
81	35	9
82	35	15

XY.dat

1	0	900	(Node_ID	X_Coordinate	Y_Coordinate)
2	400	900			
3	800	900			
4	1200	900			
5	0	700			
6	400	700			
7	400	500			
8	800	500			
9	1200	500			
10	0	300			
11	400	300			
12	0	100			
13	400	100			
14	800	100			
15	1200	100			
16	100	900			
17	0	800			
18	200	700			
19	400	800			
20	0	600			
21	200	300			
22	400	625			
23	600	900			
24	550	500			
25	800	725			
26	1050	900			
27	1025	500			
28	1200	800			
29	0	200			
30	100	100			
31	400	225			
32	500	100			
33	800	275			
34	1000	100			
35	1200	300			

Zone.dat

7	(total number of zones)			
1	100	800	4	(Zone_ID X_Coordinate Y_Coordinate total number of distribution nodes)
2	300	600	4	
3	600	650	5	
4	1050	800	4	
5	100	200	4	
6	500	250	4	
7	1000	300	4	

Small Network & Fixed Transit Demand:

Transit Demand Matrix:

Demand.dat:

7							
0	1	2	3	4	5	6	7
1	0	400	200	600	800	500	800
2	400	0	1200	200	400	600	200
3	200	1200	0	600	1000	400	1000
4	600	200	600	0	400	800	500
5	800	400	1000	400	0	400	600
6	500	600	400	800	400	0	200
7	800	200	1000	500	600	200	0

Small Network & Variable Transit Demand:

Total Demand Matrix:

Demand.dat:

7							
0	1	2	3	4	5	6	7
1	0	2136	1108	9488	10333	4991	13350
2	2136	0	5424	2919	3257	3539	3060
3	1108	5424	0	4600	9379	2190	8079
4	9488	2919	4600	0	8829	7732	4486
5	10333	3257	9379	8829	0	2751	10362
6	4991	3539	2190	7732	2751	0	1467
7	13350	3060	8079	4486	10362	1467	0

Medium Network:

Network.dat:

14	72	180									
1	2	3	4	5	6	7	8	9	10	11	12
	13	14									
1	0										
2	0										
3	0										
4	0										
5	0										
6	0										
7	0										
8	0										
9	0										
10	0										
11	0										
12	0										
13	0										
14	0										
15	0										
16	0										
17	0										
18	0										
19	0										
20	0										
21	0										
22	0										
23	0										
24	0										
25	0										
26	0										
27	0										
28	0										
29	0										
30	1	1									
31	1	2									
32	1	3									
33	1	4									
34	1	1									
35	2	1	2								
36	1	2									
37	1	3									
38	2	3	4								
39	1	4									
40	2	1	5								
41	2	2	6								
42	2	3	7								
43	2	4	8								
44	1	5									
45	2	5	6								
46	1	6									
47	1	7									
48	2	7	8								
49	1	8									
50	1	6									

51	1	9	
52	1	7	
53	2	8	10
54	1	9	
55	1	9	
56	1	10	
57	1	10	
58	2	5	11
59	1	12	
60	1	9	
61	1	13	
62	2	10	14
63	1	11	
64	2	11	12
65	1	12	
66	1	13	
67	2	13	14
68	1	14	
69	1	11	
70	1	12	
71	1	13	
72	1	14	
1	1	30	
2	1	34	
3	2	30	
4	2	31	
5	2	35	
6	3	31	
7	3	36	
8	3	4	
9	4	3	
10	4	37	
11	4	32	
12	5	32	
13	5	38	
14	5	33	
15	6	33	
16	6	39	
17	7	34	
18	7	40	
19	7	44	
20	8	35	
21	8	40	
22	8	45	
23	8	41	
24	9	41	
25	9	36	
26	9	46	
27	9	10	
28	10	9	
29	10	37	
30	10	42	
31	10	47	
32	11	38	
33	11	42	
34	11	43	
35	11	48	
36	12	39	
37	12	43	

38	12	49
39	13	45
40	13	50
41	13	19
42	14	46
43	14	50
44	14	51
45	14	54
46	15	47
47	15	51
48	15	55
49	15	52
50	16	48
51	16	52
52	16	53
53	16	56
54	17	49
55	17	53
56	17	57
57	18	44
58	18	58
59	18	63
60	19	13
61	19	58
62	19	59
63	19	64
64	20	54
65	20	59
66	20	60
67	20	65
68	21	55
69	21	60
70	21	61
71	21	66
72	22	56
73	22	61
74	22	62
75	22	67
76	23	57
77	23	62
78	23	68
79	24	63
80	24	69
81	25	64
82	25	69
83	25	70
84	26	65
85	26	70
86	26	27
87	27	66
88	27	26
89	27	71
90	28	67
91	28	71
92	28	72
93	29	68
94	29	72
95	30	1
96	30	2

97	31	2
98	31	3
99	32	4
100	32	5
101	33	5
102	33	6
103	34	1
104	34	7
105	35	2
106	35	8
107	36	3
108	36	9
109	37	4
110	37	10
111	38	5
112	38	11
113	39	6
114	39	12
115	40	7
116	40	8
117	41	8
118	41	9
119	42	10
120	42	11
121	43	11
122	43	12
123	44	7
124	44	18
125	45	8
126	45	13
127	46	9
128	46	14
129	47	10
130	47	15
131	48	11
132	48	16
133	49	12
134	49	17
135	50	13
136	50	14
137	51	14
138	51	15
139	52	15
140	52	16
141	53	16
142	53	17
143	54	14
144	54	20
145	55	15
146	55	21
147	56	16
148	56	22
149	57	17
150	57	23
151	58	18
152	58	19
153	59	19
154	59	20
155	60	20

156	60	21
157	61	21
158	61	22
159	62	22
160	62	23
161	63	18
162	63	24
163	64	19
164	64	25
165	65	20
166	65	26
167	66	21
168	66	27
169	67	22
170	67	28
171	68	23
172	68	29
173	69	24
174	69	25
175	70	25
176	70	26
177	71	27
178	71	28
179	72	28
180	72	29

XY.dat

1	0	1700
2	400	1700
3	800	1700
4	1200	1700
5	1600	1700
6	2000	1700
7	0	1300
8	400	1300
9	800	1300
10	1200	1300
11	1600	1300
12	2000	1300
13	400	900
14	800	900
15	1200	900
16	1600	900
17	2000	900
18	0	500
19	400	500
20	800	500
21	1200	500
22	1600	500
23	2000	500
24	0	100
25	400	100
26	800	100
27	1200	100
28	1600	100
29	2000	100
30	150	1700
31	650	1700

32	1450	1700
33	1925	1700
34	0	1550
35	400	1500
36	800	1450
37	1200	1550
38	1600	1500
39	2000	1450
40	150	1300
41	700	1300
42	1400	1300
43	1838	1300
44	0	950
45	400	1050
46	800	1150
47	1200	1050
48	1600	1100
49	2000	1150
50	750	900
51	1050	900
52	1350	900
53	1800	900
54	800	750
55	1200	750
56	1600	750
57	2000	750
58	150	500
59	650	500
60	1050	500
61	1350	500
62	1850	500
63	0	350
64	400	300
65	800	250
66	1200	350
67	1600	350
68	2000	350
69	150	100
70	650	100
71	1350	100
72	1850	100

Zone.dat

14

1	150	1550	4
2	650	1450	4
3	1450	1550	4
4	1925	1450	4
5	150	950	4
6	750	1150	4
7	1350	1050	4
8	1750	1150	4
9	1050	750	4
10	1850	750	4
11	150	350	4
12	650	250	4
13	1350	350	4
14	1850	350	4

Medium Network & Fixed Transit Demand:

Transit Demand Matrix:

Demand.dat:

14

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	0	400	200	600	800	500	800	300	600	500	200	400	800	500
2	400	0	1200	200	400	600	200	600	400	200	1000	500	400	800
3	200	1200	0	600	1000	400	1000	500	400	800	400	300	1200	200
4	600	200	600	0	400	800	500	400	800	300	500	700	500	900
5	800	400	1000	400	0	400	600	400	1000	500	300	600	300	1000
6	500	600	400	800	400	0	200	300	300	600	400	500	900	700
7	800	200	1000	500	600	200	0	600	400	400	300	800	500	800
8	300	600	500	400	400	300	600	0	500	300	400	1000	1400	200
9	600	400	400	800	1000	300	400	500	0	400	1200	600	500	800
10	500	200	800	300	500	600	400	300	400	0	400	200	500	700
11	200	1000	400	500	300	400	300	400	1200	400	0	600	200	1000
12	400	500	300	700	600	500	800	1000	600	200	600	0	800	500
13	800	400	1200	500	300	900	500	1400	500	500	200	800	0	400
14	500	800	200	900	1000	700	800	200	800	700	1000	500	400	0

Medium Network & Variable Transit Demand:

Total Demand Matrix:

Demand.dat:

14

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	0	2019	2609	12444	10619	3261	9097	5113	7522	10334	4328	8371	18585	17131
2	2019	0	8578	2577	4239	2718	1436	6948	3335	2843	18069	7336	7113	23115
3	2609	8578	0	3058	16596	2883	5029	2961	3623	8498	9922	5443	18079	4080
4	12444	2577	3058	0	9410	7962	3105	2266	8836	2941	16449	14705	7323	14220
5	10619	4239	16596	9410	0	2049	5634	5403	8431	7921	2725	6282	5745	26455
6	3261	2718	2883	7962	2049	0	1115	2879	1627	8518	6364	4956	14203	21509
7	9097	1436	5029	3105	5634	1115	0	2598	1841	2497	5032	8954	4198	9997
8	5113	6948	2961	2266	5403	2879	2598	0	3752	1844	9285	15639	13849	4731
9	7522	3335	3623	8836	8431	1627	1841	3752	0	2332	10308	3359	2705	6854
10	10334	2843	8498	2941	7921	8518	2497	1844	2332	0	8741	2957	3172	4348
11	4328	18069	9922	16449	2725	6364	5032	9285	10308	8741	0	3198	2274	16387
12	8371	7336	5443	14705	6282	4956	8954	15639	3359	2957	3198	0	4796	5514
13	18585	7113	18079	7323	5745	14203	4198	13849	2705	3172	2274	4796	0	2069
14	17131	23115	4080	14220	26455	21509	9997	4731	6854	4348	16387	5514	2069	0

Large Network:

Network.dat:

28	160	418									
1	2	3	4	5	6	7	8	9	10	11	12
13	14	15	16	17	18	19	20	21	22	23	24
25	26	27	28								
1	0										
2	0										
3	0										
4	0										
5	0										
6	0										
7	0										
8	0										
9	0										
10	0										
11	0										
12	0										
13	0										
14	0										
15	0										
16	0										
17	0										
18	0										
19	0										
20	0										
21	0										
22	0										
23	0										
24	0										
25	0										
26	0										
27	0										
28	0										
29	0										
30	0										
31	0										
32	0										
33	0										
34	0										
35	0										
36	0										
37	0										
38	0										
39	0										
40	0										
41	0										
42	0										
43	0										
44	0										
45	0										
46	0										
47	0										
48	0										
49	0										

50	0	
51	0	
52	0	
53	0	
54	0	
55	0	
56	0	
57	0	
58	0	
59	0	
60	0	
61	0	
62	0	
63	0	
64	0	
65	0	
66	1	1
67	1	2
68	1	3
69	1	4
70	1	5
71	1	6
72	1	1
73	2	1
74	2	2
75	1	3
76	1	4
77	2	4
78	2	5
79	1	1
80	2	2
81	2	3
82	2	4
83	1	5
84	1	6
85	1	7
86	1	7
87	1	8
88	1	8
89	1	9
90	1	9
91	1	6
92	1	13
93	2	7
94	1	8
95	1	11
96	1	9
97	2	6
98	1	10
99	1	10
100	1	14
101	1	11
102	1	11
103	1	12
104	1	12
105	1	10
106	2	11
107	1	15
108	1	12

109	1	13
110	1	13
111	1	14
112	2	14
113	1	15
114	1	15
115	2	13
116	1	17
117	1	14
118	2	15
119	1	20
120	1	16
121	1	16
122	1	17
123	1	17
124	1	18
125	1	18
126	1	19
127	2	19
128	1	20
129	1	16
130	1	21
131	1	17
132	1	22
133	1	18
134	1	23
135	1	19
136	2	20
137	1	21
138	1	21
139	1	22
140	1	22
141	1	23
142	1	23
143	1	24
144	1	25
145	1	21
146	2	22
147	1	27
148	1	23
149	1	28
150	1	24
151	1	25
152	1	25
153	1	26
154	2	26
155	1	27
156	1	28
157	2	24
158	1	26
159	1	27
160	1	28
1	1	66
2	1	72
3	2	66
4	2	67
5	2	73
6	3	67
7	3	68

8	3	74
9	4	5
10	4	68
11	4	75
12	5	4
13	5	69
14	5	76
15	6	69
16	6	70
17	6	77
18	7	70
19	7	71
20	7	78
21	8	71
22	8	84
23	9	17
24	9	72
25	9	79
26	10	73
27	10	79
28	10	80
29	10	85
30	11	12
31	11	74
32	11	80
33	11	86
34	12	11
35	12	81
36	12	87
37	13	14
38	13	75
39	13	81
40	13	88
41	14	13
42	14	76
43	14	82
44	14	89
45	15	77
46	15	82
47	15	83
48	15	90
49	16	78
50	16	83
51	16	91
52	17	9
53	17	92
54	17	109
55	18	85
56	18	92
57	18	93
58	18	98
59	19	20
60	19	86
61	19	93
62	19	99
63	20	19
64	20	28
65	20	87
66	20	94

67	21	88
68	21	94
69	21	95
70	21	100
71	21	101
72	22	89
73	22	95
74	22	96
75	22	102
76	23	24
77	23	31
78	23	90
79	23	96
80	24	23
81	24	91
82	24	97
83	24	103
84	25	84
85	25	97
86	25	104
87	26	98
88	26	105
89	26	110
90	27	28
91	27	36
92	27	99
93	27	105
94	28	20
95	28	27
96	28	100
97	28	111
98	29	101
99	29	106
100	29	112
101	30	31
102	30	39
103	30	102
104	30	106
105	31	23
106	31	30
107	31	32
108	31	113
109	32	103
110	32	107
111	32	108
112	32	114
113	33	42
114	33	104
115	33	108
116	34	109
117	34	115
118	34	120
119	35	36
120	35	110
121	35	115
122	35	121
123	36	27
124	36	35
125	36	116

126	36	122
127	37	111
128	37	116
129	37	117
130	37	123
131	38	112
132	38	117
133	38	124
134	39	30
135	39	40
136	39	125
137	40	39
138	40	113
139	40	118
140	40	126
141	41	114
142	41	118
143	41	119
144	41	127
145	42	33
146	42	119
147	42	128
148	43	52
149	43	120
150	43	129
151	44	121
152	44	129
153	44	130
154	44	137
155	45	122
156	45	130
157	45	131
158	45	138
159	46	123
160	46	131
161	46	132
162	46	139
163	47	124
164	47	132
165	47	133
166	47	140
167	48	125
168	48	133
169	48	134
170	48	141
171	49	126
172	49	134
173	49	135
174	49	142
175	50	127
176	50	135
177	50	136
178	50	143
179	51	128
180	51	136
181	51	150
182	52	43
183	52	144
184	52	151

185	53	137
186	53	144
187	53	145
188	53	152
189	54	55
190	54	138
191	54	145
192	54	153
193	55	54
194	55	139
195	55	146
196	56	140
197	56	146
198	56	147
199	56	154
200	57	141
201	57	147
202	57	148
203	57	155
204	58	142
205	58	148
206	58	149
207	58	156
208	59	143
209	59	149
210	59	157
211	60	61
212	60	151
213	60	152
214	61	60
215	61	153
216	61	158
217	62	154
218	62	158
219	62	159
220	63	64
221	63	155
222	63	159
223	64	63
224	64	156
225	64	160
226	65	157
227	65	160
228	65	150
229	66	1
230	66	2
231	67	2
232	67	3
233	68	3
234	68	4
235	69	5
236	69	6
237	70	6
238	70	7
239	71	7
240	71	8
241	72	1
242	72	9
243	73	2

244	73	10
245	74	3
246	74	11
247	75	4
248	75	13
249	76	5
250	76	14
251	77	6
252	77	15
253	78	7
254	78	16
255	79	9
256	79	10
257	80	10
258	80	11
259	81	12
260	81	13
261	82	14
262	82	15
263	83	15
264	83	16
265	84	8
266	84	25
267	85	10
268	85	18
269	86	11
270	86	19
271	87	12
272	87	20
273	88	13
274	88	21
275	89	14
276	89	22
277	90	15
278	90	23
279	91	16
280	91	24
281	92	17
282	92	18
283	93	18
284	93	19
285	94	20
286	94	21
287	95	21
288	95	22
289	96	22
290	96	23
291	97	24
292	97	25
293	98	18
294	98	26
295	99	19
296	99	27
297	100	21
298	100	28
299	101	21
300	101	29
301	102	22
302	102	30

303	103	24
304	103	32
305	104	25
306	104	33
307	105	26
308	105	27
309	106	29
310	106	30
311	107	31
312	107	32
313	108	32
314	108	33
315	109	17
316	109	34
317	110	26
318	110	35
319	111	28
320	111	37
321	112	29
322	112	38
323	113	31
324	113	40
325	114	32
326	114	41
327	115	34
328	115	35
329	116	36
330	116	37
331	117	37
332	117	38
333	118	40
334	118	41
335	119	41
336	119	42
337	120	34
338	120	43
339	121	35
340	121	44
341	122	36
342	122	45
343	123	37
344	123	46
345	124	38
346	124	47
347	125	39
348	125	48
349	126	40
350	126	49
351	127	41
352	127	50
353	128	42
354	128	51
355	129	43
356	129	44
357	130	44
358	130	45
359	131	45
360	131	46
361	132	46

362	132	47
363	133	47
364	133	48
365	134	48
366	134	49
367	135	49
368	135	50
369	136	50
370	136	51
371	137	44
372	137	53
373	138	45
374	138	54
375	139	46
376	139	55
377	140	47
378	140	56
379	141	48
380	141	57
381	142	49
382	142	58
383	143	50
384	143	59
385	144	52
386	144	53
387	145	53
388	145	54
389	146	55
390	146	56
391	147	56
392	147	57
393	148	57
394	148	58
395	149	58
396	149	59
397	150	51
398	150	65
399	151	52
400	151	60
401	152	53
402	152	60
403	153	54
404	153	61
405	154	56
406	154	62
407	155	57
408	155	63
409	156	58
410	156	64
411	157	59
412	157	65
413	158	61
414	158	62
415	159	62
416	159	63
417	160	64
418	160	65

XY.dat

1	0	2900
2	400	2900
3	800	2900
4	1600	2900
5	2000	2900
6	2400	2900
7	2800	2900
8	3200	2900
9	0	2500
10	400	2500
11	800	2500
12	1200	2500
13	1600	2500
14	2000	2500
15	2400	2500
16	2800	2500
17	0	2100
18	400	2100
19	800	2100
20	1200	2100
21	1600	2100
22	2000	2100
23	2400	2100
24	2800	2100
25	3200	2100
26	400	1700
27	800	1700
28	1200	1700
29	1600	1700
30	2000	1700
31	2400	1700
32	2800	1700
33	3200	1700
34	0	1300
35	400	1300
36	800	1300
37	1200	1300
38	1600	1300
39	2000	1300
40	2400	1300
41	2800	1300
42	3200	1300
43	0	900
44	400	900
45	800	900
46	1200	900
47	1600	900
48	2000	900
49	2400	900
50	2800	900
51	3200	900
52	0	500
53	400	500
54	800	500
55	1200	500
56	1600	500

57	2000	500
58	2400	500
59	2800	500
60	400	100
61	800	100
62	1600	100
63	2000	100
64	2400	100
65	2800	100
66	160	2900
67	640	2900
68	1360	2900
69	2160	2900
70	2680	2900
71	3120	2900
72	0	2700
73	400	2700
74	800	2700
75	1600	2700
76	2000	2650
77	2400	2700
78	2800	2750
79	160	2500
80	600	2500
81	1360	2500
82	2240	2500
83	2680	2500
84	3200	2460
85	400	2300
86	800	2300
87	1200	2300
88	1600	2300
89	2000	2300
90	2400	2300
91	2800	2460
92	160	2100
93	600	2100
94	1440	2100
95	1840	2100
96	2320	2100
97	3080	2100
98	400	1900
99	800	1900
100	1440	2000
101	1600	1900
102	2000	1900
103	2800	1900
104	3200	1900
105	640	1700
106	1840	1700
107	2640	1700
108	3120	1700
109	0	1620
110	400	1620
111	1200	1620
112	1600	1620
113	2400	1500
114	2800	1500
115	200	1300

116	960	1300
117	1440	1300
118	2600	1300
119	3040	1300
120	0	1100
121	400	1100
122	800	1100
123	1200	1100
124	1600	1220
125	2000	1220
126	2400	1100
127	2800	1100
128	3200	1100
129	240	900
130	640	900
131	960	900
132	1440	900
133	1840	900
134	2160	900
135	2560	900
136	3000	900
137	400	700
138	800	700
139	1200	700
140	1600	700
141	2000	700
142	2400	700
143	2800	700
144	280	500
145	640	500
146	1400	500
147	1840	500
148	2160	500
149	2560	500
150	3040	580
151	220	300
152	400	350
153	800	300
154	1600	350
155	2000	400
156	2400	300
157	2800	300
158	1360	100
159	1840	100
160	2560	100

Zone.dat

28

1	160	2700	4
2	640	2700	4
3	1360	2700	4
4	2160	2650	4
5	2680	2750	4
6	3040	2460	5
7	560	2300	4
8	1440	2300	4

9	2320	2300	4
10	640	1900	4
11	1840	1900	4
12	3120	1900	4
13	160	1620	4
14	1440	1620	4
15	2640	1500	4
16	240	1100	4
17	960	1100	4
18	1840	1220	5
19	2560	1100	4
20	3040	1100	4
21	640	700	4
22	1440	700	4
23	2160	700	4
24	2960	700	4
25	280	350	3
26	1360	300	4
27	1840	400	4
28	2560	300	4

Large Network & Fixed Transit Demand:

Transit Demand Matrix:

Demand.dat:

28

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	0	400	200	600	800	500	800	300	600	500	200	400	800	500
2	400	0	1200	200	400	600	200	600	400	200	1000	500	400	800
3	200	1200	0	600	1000	400	1000	500	400	800	400	300	1200	200
4	600	200	600	0	400	800	500	400	800	300	500	700	500	900
5	800	400	1000	400	0	400	600	400	1000	500	300	600	300	1000
6	500	600	400	800	400	0	200	300	300	600	400	500	900	700
7	800	200	1000	500	600	200	0	600	400	400	300	800	500	800
8	300	600	500	400	400	300	600	0	500	300	400	1000	1400	200
9	600	400	400	800	1000	300	400	500	0	400	1200	600	500	800
10	500	200	800	300	500	600	400	300	400	0	400	200	500	700
11	200	1000	400	500	300	400	300	400	1200	400	0	600	200	1000
12	400	500	300	700	600	500	800	1000	600	200	600	0	800	500
13	800	400	1200	500	300	900	500	1400	500	500	200	800	0	400
14	500	800	200	900	1000	700	800	200	800	700	1000	500	400	0
15	300	500	1400	500	500	200	400	500	600	400	800	600	700	500
16	500	800	400	1000	200	400	300	1000	400	800	600	500	700	400
17	1200	600	700	400	1500	800	700	1000	500	400	800	600	300	1400
18	400	300	500	700	600	300	200	400	400	300	800	500	1000	800
19	1000	400	800	600	500	700	800	200	200	1000	500	600	800	400
20	500	1000	800	500	400	500	400	500	1400	400	1000	1000	300	200
21	400	600	200	800	200	300	500	600	1200	800	400	800	500	1000
22	800	400	200	1000	600	700	300	500	300	400	300	400	1400	600
23	1000	500	1000	400	1200	600	700	400	600	700	1400	300	500	700
24	700	300	400	500	200	500	600	800	400	300	1000	700	500	300
25	800	700	800	600	1000	800	500	500	400	400	400	1100	200	500
26	200	600	600	400	600	1000	800	400	800	500	500	300	600	900
27	800	500	500	800	200	600	1000	800	1000	200	800	900	500	800
28	700	800	700	400	600	200	700	300	300	500	300	1000	800	200

0	15	16	17	18	19	20	21	22	23	24	25	26	27	28
1	300	500	1200	400	1000	500	400	800	1000	700	800	200	800	700
2	500	800	600	300	400	1000	600	400	500	300	700	600	500	800
3	1400	400	700	500	800	800	200	200	1000	400	800	600	500	700
4	500	1000	400	700	600	500	800	1000	400	500	600	400	800	400
5	500	200	1500	600	500	400	200	600	1200	200	1000	600	200	600
6	200	400	800	300	700	500	300	700	600	500	800	1000	600	200
7	400	300	700	200	800	400	500	300	700	600	500	800	1000	700
8	500	1000	1000	400	200	500	600	500	400	800	500	400	800	300
9	600	400	500	400	200	1400	1200	300	600	400	400	800	1000	300
10	400	800	400	300	1000	400	800	400	700	300	400	500	200	500
11	800	600	800	800	500	1000	400	300	1400	1000	400	500	800	300
12	600	500	600	500	600	1000	800	400	300	700	1100	300	900	1000
13	700	700	300	1000	800	300	500	1400	500	500	200	600	500	800
14	500	400	1400	800	400	200	1000	600	700	300	500	900	800	200
15	0	500	800	200	900	1000	700	800	200	800	700	1000	500	400
16	500	0	600	500	800	500	700	500	900	500	700	800	200	1000
17	800	600	0	600	1200	600	700	400	1500	800	700	500	500	400
18	200	500	600	0	1000	400	1200	600	700	400	600	400	800	700
19	900	800	1200	1000	0	900	1200	800	1000	600	200	800	1000	1000
20	1000	500	600	400	900	0	400	1200	600	700	800	400	600	400
21	700	700	700	1200	1200	400	0	500	800	1000	1000	300	800	800
22	800	500	400	600	800	1200	500	0	900	400	600	1200	400	600
23	200	900	1500	700	1000	600	800	900	0	300	200	1200	500	500
24	800	500	800	400	600	700	1000	400	300	0	600	300	300	700
25	700	700	700	600	200	800	1000	600	200	600	0	600	700	800
26	1000	800	500	400	800	400	300	1200	1200	300	600	0	1400	200
27	500	200	500	800	1000	600	800	400	500	300	700	1400	0	200
28	400	1000	400	700	1000	400	800	600	500	700	800	200	200	0

Large Network & Variable Transit Demand:

Total Demand Matrix:

Demand.dat:

28

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	0	1329	781	2436	3307	2174	2723	1183	2609	1924	877	1786	3319	2312
2	1329	0	4047	761	1581	2571	609	2092	1600	727	3993	2229	1680	3409
3	781	4047	0	2069	3716	1704	3593	1491	1468	3274	1493	1371	5744	856
4	2436	761	2069	0	1219	2933	2001	1334	2200	1256	1661	2812	2346	3531
5	3307	1581	3716	1219	0	1344	2664	1561	3193	2297	1135	2435	1479	4408
6	2174	2571	1704	2933	1344	0	947	1303	1098	2948	1662	1785	4531	3412
7	2723	609	3593	2001	2664	947	0	1850	1486	1106	1082	3613	1874	3136
8	1183	2092	1491	1334	1561	1303	1850	0	1465	952	1049	4079	5775	621
9	2609	1600	1468	2200	3193	1098	1486	1465	0	1563	3160	2162	2374	2863
10	1924	727	3274	1256	2297	2948	1106	952	1563	0	1334	877	1645	2413
11	877	3993	1493	1661	1135	1662	1082	1049	3160	1334	0	2123	840	2610
12	1786	2229	1371	2812	2435	1785	3613	4079	2162	877	2123	0	3604	2139
13	3319	1680	5744	2346	1479	4531	1874	5775	2374	1645	840	3604	0	1557
14	2312	3409	856	3531	4408	3412	3136	621	2863	2413	2610	2139	1557	0
15	1464	2377	6672	2031	2105	848	1930	2054	2168	1849	2637	2048	3148	2099
16	2176	3413	1908	4583	958	1978	1215	4238	1885	2930	2519	2240	2279	1554
17	5356	2538	3277	1800	7505	4383	2707	3722	2169	1322	2791	2812	1083	3930
18	2147	1493	2526	3108	2952	1638	962	1555	1640	1310	2313	2190	4516	2329
19	4814	1912	3929	2553	2205	3168	3898	859	802	4565	1774	2263	3605	1725
20	2299	4641	3890	2179	1769	2194	1946	2246	5832	1879	3927	3620	1343	970
21	1858	2677	1016	3892	1035	1648	2135	2647	5949	3040	1700	3860	1934	3875
22	3913	1876	1034	4687	3061	3866	1378	2122	1397	1650	1127	1885	5940	1986
23	4990	2458	5079	1796	5721	3071	3473	1755	2576	3150	5035	1295	2274	2650
24	3190	1397	1999	2254	912	2322	2996	3837	1807	1474	4341	2856	2339	1504
25	3445	2984	3652	2631	4407	3616	2153	2195	1829	1671	1755	4797	828	2120
26	1020	2978	3321	2016	3155	5494	4055	2011	4248	2392	2402	1472	2927	4056
27	3778	2330	2462	3602	945	2999	4734	3556	4530	892	3195	4019	2272	3141
28	3170	3683	3461	1804	2747	955	3439	1420	1375	2428	1300	4322	3730	942

0	15	16	17	18	19	20	21	22	23	24	25	26	27	28
1	1464	2176	5356	2147	4814	2299	1858	3913	4990	3190	3445	1020	3778	3170
2	2377	3413	2538	1493	1912	4641	2677	1876	2458	1397	2984	2978	2330	3683
3	6672	1908	3277	2526	3929	3890	1016	1034	5079	1999	3652	3321	2462	3461
4	2031	4583	1800	3108	2553	2179	3892	4687	1796	2254	2631	2016	3602	1804
5	2105	958	7505	2952	2205	1769	1035	3061	5721	912	4407	3155	945	2747
6	848	1978	4383	1638	3168	2194	1648	3866	3071	2322	3616	5494	2999	955
7	1930	1215	2707	962	3898	1946	2135	1378	3473	2996	2153	4055	4734	3439
8	2054	4238	3722	1555	859	2246	2647	2122	1755	3837	2195	2011	3556	1420
9	2168	1885	2169	1640	802	5832	5949	1397	2576	1807	1829	4248	4530	1375
10	1849	2930	1322	1310	4565	1879	3040	1650	3150	1474	1671	2392	892	2428
11	2637	2519	2791	2313	1774	3927	1700	1127	5035	4341	1755	2402	3195	1300
12	2048	2240	2812	2190	2263	3620	3860	1885	1295	2856	4797	1472	4019	4322
13	3148	2279	1083	4516	3605	1343	1934	5940	2274	2339	828	2927	2272	3730
14	2099	1554	3930	2329	1725	970	3875	1986	2650	1504	2120	4056	3141	942
15	0	2144	3240	685	2485	3214	3081	3136	690	3112	3044	4618	1979	1592
16	2144	0	1877	1874	3191	2050	2217	1797	3512	2154	2699	3406	823	4293
17	3240	1877	0	1747	4193	2313	1899	1059	4839	3345	2711	1885	1778	1657
18	685	1874	1747	0	3086	1471	4515	1662	1812	1594	2545	1676	2667	2754
19	2485	3191	4193	3086	0	2634	4631	2666	2751	2006	825	3264	3558	3639
20	3214	2050	2313	1471	2634	0	1646	4522	2028	2321	3334	1711	2339	1535
21	3081	2217	1899	4515	4631	1646	0	1454	2716	4075	3515	1060	2844	3203
22	3136	1797	1059	1662	2666	4522	1454	0	2473	1502	2259	3643	1136	2200
23	690	3512	4839	1812	2751	2028	2716	2473	0	1009	796	4184	1420	1611
24	3112	2154	3345	1594	2006	2321	4075	1502	1009	0	2458	1203	1115	2457
25	3044	2699	2711	2545	825	3334	3515	2259	796	2458	0	2271	2699	3298
26	4618	3406	1885	1676	3264	1711	1060	3643	4184	1203	2271	0	4148	752
27	1979	823	1778	2667	3558	2339	2844	1136	1420	1115	2699	4148	0	667
28	1592	4293	1657	2754	3639	1535	3203	2200	1611	2457	3298	752	667	0

REFERENCES

Aashtiani, H.Z. and Iravani, H. (2002), "Application of Dwell Time Functions in Transit Assignment Model," Presented at *the 81st Transportation Research Board*, Preprint CD-ROM, Washington D.C., 2002.

Abkowitz, M., Eiger, A. and Engelstein, I. (1986), "Optimal Control of Headway Variation on Transit Routes," *Journal of Advanced Transportation*, Vol. 20, No. 1, P 73-88.

Abkowitz, M. and Tozzi, J. (1986), "Transit Route Characteristics and Headway-Based Reliability Control," *Journal of Transportation Research Record*, No. 1078, P 11-16.

Abkowitz, M., Josef, R. and Driscoll, M. (1987), "Operational Feasibility of Timed Transfer in Transit Systems," *Journal of Transportation Engineering*, Vol. 113, No. 2, P 68-177, ASCE.

Adamski, A. (1992), "Probabilistic Models of Passengers Service Processes at Bus Stops," *Transportation Research Part B*, Vol. 26, P 253-259.

Ahuja, R.K., Magnanti, T.L., and Orlin, J.B. (1993), *Network Flows: Theory, Algorithms and Applications*, Prentice Hall.

American Public Transportation Association, Public Transportation Ridership Statistics http://www.apta.com/stats/ridershp/

American Public Transit Association (1998), *Transit Fact Book*, Washington, D.C., February, 2000.

Andersson, P.A. and Scalia-Tomba, G.P. (1981), "A Mathematical Model of an Urban Bus Route," *Transportation Research Part B: Methodological*, Vol. 15, No. 4, P 249-266.

Baaj, M.H. (1990), The Transit Network Design Problem: An AI-Based Approach, Ph.D. thesis, Department of Civil Engineering, University of Texas, Austin, Texas.

Baaj, M.H. and Mahmassani, H.S. (1990), "TRUST: A Lisp Program for the Analysis of Transit Route Configuration," *Transportation Research Record*, No. 1283, Transportation Research Board, Washington, D.C., P 125-135.

Baaj, M.H. and Mahmassani, H.S. (1991), "An AI-Based Approach for Transit Route System Planning and Design," *Journal of Advanced Transportation*, Vol. 25, No. 2, P 187-210.

Baaj, M.H. and Mahmassani, H.S. (1992), "Artificial Intelligence-Based System Representation and Search Procedures for Transit Route Network Design," *Transportation Research Record*, No. 1358, Transportation Research Board, Washington, D.C., P 67-70.

Baaj, M.H. and Mahmassani, H.S. (1995), "Hybrid Route Generation Heuristic Algorithm for the Design of Transit Networks," *Transportation Research Part C*, Vol. 3, Issue 1, P 31-50.

Bakker, J.J., Calkin, J., and Sylvester, S. (1988), "Multi-Centered Timed-Transfer System for Capital Metro, Austin, Texas," *Transportation Research Record*, No. 1202, Transportation Research Board, Washington, D.C., P 22-28.

Barnes, J.W. (2002), *Metaheuristics*, Class Notes, The University of Texas at Austin, Austin, Texas, 2002.
Bhat, C. (2003), "Discrete Choice Modeling," Class notes, UT-Austin, 2003.
Bielli, M., Caramia, M. and Carotenuto, P. (2002), "Genetic Algorithms in Bus Network Optimization," *Transportation Research Part C*, Vol. 10, Issue 1, P 19-34.
Bly, P.H. and Oldfield, R.H. (1986), "Competition between Minibuses and Regular Bus Service," *Journal of Transport Economics and Policy*, Vol. 20, No. 1, P 47-68.
Bookbinder, J.H. and Desilets, A. (1992), "Transfer Optimization in a Transit Network," *Transportation Science*, Vol. 26, No. 2, P 106-118.
Bowerman, R., Hall, B. and Calamai, P. (1995), "A Multi-Objective Optimization Approach to Urban School Bus Routing: Formulation and Solution Method," *Transportation Research Part A: Policy and Practice*, Vol. 29, No. 2, P 107-123.
Bowman, L.A. and Turnquist, M.A. (1981), "Service Frequency, Schedule Reliability and Passenger Wait Times at Transit Stops," *Transportation Research Part A*, Vol. 15, No. 6, P 465-471.
Boyan, J.A. and Mitzenmacher M. (2001), "Improved Results for Route Planning in Stochastic Transportation Networks," *12th Annual Symposium on Discrete Algorithms* (SODA), 2001.
Ceder, A. (1984), "Bus Frequency Determination using Passenger Count Data," *Transportation Research Part A*, Vol. 18, No. 5/6, P 439-453.
Ceder, A. (1985), "Computer Application for Determining Bus Headways and Timetables," *Journal of Transportation Research Record*, No. 1011, P 76-87.
Ceder, A. (1986), "Methods for Creating Bus Timetables," *Transportation Research Part A*, Vol. 21, No. 1, P 59-83.
Ceder, A. (1991), "Transit Scheduling," *Journal of Advanced Transportation*, Vol. 25, No. 2, P 137-160.
Ceder, R.B. and Wilson, N.H. (1986), "Bus Network Design," *Transportation Research Part B*, Vol. 20, No. 4, P 331-344.
Ceder, A. and Israeli, Y. (1998), "User and Operator Perspective in Transit Network Design", Paper No. 980267, 77th Annual Meeting of TRB, Washington, DC.
Ceder, A. (2001), "Operational Objective Functions in Designing Public Transport Routes," *Journal of Advance Transportation*, Vol. 35, No. 2, P 125-144.
Chakraborty, P., Deb, K. and Subrahmanyam, S. (1995), "Optimal Scheduling of Transit Systems using Genetic Algorithms," *Journal of Transportation Engineering*, ASCE, Vol. 121, No. 6, P 544-552.
Chambers, L., (1995), *Practical Handbook of Genetic Algorithms: Applications: Volume I*. Boca Raton, FL, CRC Press.
Chandrasekar, P., Cheu, R.L. and Chin, H.C. (2002), "Simulation Evaluation of Route-Based Control of Bus Operations," *Journal of Transportation Engineering*, Vol. 128, No. 6, P 519-527.
Chang, S.K. and Schonfeld, P.M. (1991), "Integration of Fixed and Flexible Route Bus System," *Transportation Research Record*, No. 1308, Transportation Research Board, Washington, D.C., P 51-57.

Chang, S.K. and Schonfeld, P.M. (1991), "Multiple Period Optimization of Bus Transit Systems," *Transportation Research Part B*, Vol. 25, No. 6, P 453-478.
Cherkassky, B.V., Goldberg, A.V. and Radzik, T. (1996), "Shortest Paths Algorithms: Theory and Experimental Evaluation," *Mathematical Programming*, Vol. 73, P 129-174.
Chien, S. and Schonfeld, P. (1997), "Optimization of Urban Grid Transit System in Heterogeneous Urban Environmental," *Journal of Transportation Engineering*, ASCE, Vol. 123, No. 1, P 28-35.
Chien, S. and Schonfeld, P. (1998), "Joint Optimization of a Rail Transit Line and its Feeder Bus System," *Journal of Advanced Transportation*, Vol. 32, No. 3, P 253-284.
Chien, S. and Yang, Z. (2000), "Optimal Feeder Bus Routes with Irregular Street Networks," *Journal of Advanced Transportation*, Vol. 34, No. 2, P 213-248.
Chien, S., Yang, Z. and Hou, E. (2001), "A Genetic Algorithm Approach for Transit Route Planning and Design," *Journal of Transportation Engineering*, ASCE, Vol. 127, No. 3, P 200-207.
Chien, S., Dimitrijevic, B.V. and Spasovic, L.N. (2001), "Bus Route Planning in Urban Grid Commuter Networks," Transportation Research Board 80th Annual Meeting, Washington, D.C., January, 2001.
Lee, C. (1998), Combined Traffic Signal Control and Traffic Assignment: Algorithms, Implementation and Numerical Results, Ph.D. thesis, Department of Civil Engineering, University of Texas, Austin, Texas.
Cominetti, R. and Correa, J. (2001), "Common-Lines and Passenger Assignment in Congested Transit Networks," *Transportation Science*, Vol. 35, No. 3, P 250-267.
Constantin, I. and Florian, M. (1995), "Optimizing Frequencies in a Transit Network: a Nonlinear Bi-level Programming Approach," *International Transactions in Operational Research*, Vol. 2, No. 2, P 149-164.
Corberan, A., Fernandez, E., Lagunay, M. and Marti, R. (2000), "Heuristic Solutions to the Problem of Routing School Buses with Multiple Objectives," *Journal of the Operational Research Society*, Vol. 53, P 427-435.
Cury, J.E., Gomide, F.A.C. and Mendes, M.J., (1980), "A Methodology for Generation of Optimal Schedules for an Underground Railway System," *IEEE Transactions on Automatic Control*, Vol. AC-25, No. 2, P 217-222.
Dial, R.B. (1967), "Transit Pathfinder Algorithm," *Highway Research Record*, No. 205, P 67-85.
Dial, R.B., Glover, F., Karney, D., and Klingman, D. (1979), "A Computational Analysis of Alternative Algorithms and Labeling Techniques for Finding Shortest Path Trees," *Networks*, Vol. 9, P 215-248.
Dijkstra, E.W. (1959), "A Note on Two Problems in Connection with Graphs," *Numeriche Mathematik*, Vol. 1, P 269-271.
Ding, Y. and Chien, S.I. (2001), "Improving Transit Service Quality and Headway Regularity with Real-time Control," *Transportation Research Board*, Paper No. 01-2150, Transportation Research Board, Washington, D.C., 2001.
Dubois, D., Bell, G., and Llibre, M. (1979), "A Set of Methods in Transportation Network Synthesis and Analysis," *Journal of Operations Research Society*, Vol. 30, P 797-808.

Dufourd, H., Gendreau, M. and Laporte, G. (1996), “Locating a transit line using tabu search,” *Location Science*, Vol. 4, Issues 1-2, P 1-19.
Eglese, R.W. (1990), “Simulated annealing: A Tool for Operational Research,” *European Journal of Operational Research*, Vol. 46, P 271-281.
Erkut, E. (1996), “The Road Not Taken,” *ORMS Today*, Vol. 23, No. 6, P 22-28.
Fan, W. and Machemehl, R.B. (2002), “Characterizing Bus Transit Passenger Waiting Times,” 30th Annual Canadian Society of Civil Engineering (CSCE) Conference, Montreal, Quebec, Canada, June, 2002.
Fan, W. and Machemehl, R.B. (2003), “Hybrid Transit Trip Assignment Models for Transit Network Systems,” *working paper*.
Fan, W. and Machemehl, R.B. (2003), “A Genetic Algorithm Approach for Optimal Bus Transit Route Network Design Problem,” Submitted to the 32nd Annual CSCE Conference, June, 2004.
Fan, W. and Machemehl, R.B. (2003), “Optimal Bus Transit Route Network Design Problem with Variable Transit Demand: A Genetic Algorithm Approach,” Submitted to *Transportation Research Record of the Transportation Research Board*, 2003.
Fan, W. and Machemehl, R.B., “Using Simulated Annealing Algorithm to Solve the Transit Route Network Design Problem,” Submitted to *Transportation Research Record of the Transportation Research Board*, 2003.
Fan, W. and Machemehl, R.B. (2003), “Tabu Search Methods for Optimal Transit Route Network Design Problem,” Submitted to 9th International Conference on Computer Aided Scheduling of Public Transport, San Diego, California, August, 2004.
Fan, W. and Machemehl, R.B. (2003), “Heuristic Search Strategies for Optimal Transit Route Network Design Problem: Solution Algorithms and Numerical Results,” *working paper*.
Florian, M. and Speiss, H. (1983), “On Two Mode Choice/Assignment Models,” *Transportation Science*, No. 17, P 32-47.
Friedrich, M., Hofsaess, I. and Wekeck, S. (2001), “Timetable-Based Transit Assignment Using Branch and Bound Techniques,” *Journal of Transportation Research Record*, No. 1752, P 100-107.
Furth, P.G. and Wilson, N.H.M. (1981), “Setting Frequencies on Bus Routes: Theory and Practice,” *Transportation Research Record*, No. 818, Transportation Research Board, Washington, D.C., P 1-7.
Furth, P.G. and Day, F.B. (1985), “Transit Routing and Scheduling Strategies for Heavy Demand Corridors,” *Journal of Transportation Research Record*, No. 1011, P 23-26.
Furth, P.G. and Rahbee, A.B. (2000), “Optimal Bus Stop Spacing Through Dynamic Programming and Geographic Modeling,” *Journal of Transportation Research Record*, No. 1731, P 15-22.
Gallo, G., and Pallottino, S. (1988), “Shortest Paths Algorithms,” *Annals of Operations Research*, Vol. 13, P 3-79.
Gen, M., Liu, B. and Ida, K. (1996), “Evolution Program for Deterministic and Stochastic Optimizations,” *European Journal of Operational Research*, Vol. 94, No. 3, P 618-625.
Glover, F. (1977), “Heuristics for Integer Programming Using Surrogate Constraints,” *Decision Sciences*, Vol. 8, No. 1, P 156-166.

Glover, F. (1989), “Tabu Search, Part I,” *ORSA Journal on Computing*, Vol. 1, P 190-206.
Glover, F. (1990), “Tabu Search, Part II,” *ORSA Journal on Computing*, Vol. 2, P 4-32.
Glover, F. and Laguna, M. (1997), *Tabu Search*, Kluwer Academic Publishers, 1997.
Goczyla, K. and Cielatkowski, J. (1995), “Optimal Routing in a Transportation Network,” *European Journal of Operational Research*, Vol. 87, P 214-222.
Golenberg, M. and Pernaw, S. (1979), “Demand-Estimating Model for Transit Route and System Planning in Small Urban Areas,” *Journal of Transportation Research Record*, No. 730, P 14-23.
Goldberg, D.E. (1989), Genetic Algorithms in Search, Optimisation, and Machine Learning, Addison-Wesley Reading, London.
Haghani, A.E. and Daskin, M.S. (1983), “Network Design Application of an Extraction Algorithm for Network Aggregation,” *Journal of Transportation Research Record*, No. 944, P 37-46.
Han, A.F. and Wilson, N.H.M. (1982), “The Allocation *of* Buses in Heavily Utilized Networks with Overlapping Routes,” *Transportation Research Part B*, Vol. 16, No. 3, P 221-232.
Han, A.F. and Hwang, C.H. (1993), “Efficient Search Algorithms for Route Information Services of Direct and Connecting Transit Trips,” *Transportation Research Record*, No. 1358, P 1-5.
Hasselstrom, D. (1981), *Public Transportation Planning -- A Mathematical Programming Approach,* Ph.D. thesis, Department of Business Administration, University of Gothenburg, Sweden.
Henk, R.H. and Hubbard, S.M. (1996), “Developing an Index of Transit Service Availability,” *Transportation Research Record*, Vol. 1521, P 12-19.
Hennan, B. and Ardekani, S. (1984), “Characterizing Traffic Conditions in Urban Areas,” *Transportation Science*, Vol. 18, No. 2, P 101-140.
Hickman M.D. and Wilson N.H.M. (1995), “Passenger Travel Time and Path Choice Implications of Real-time Transit Information,” *Transportation Research Part C: Emerging Technologies*, Vol. 3, Issue 4, P 211-226.
Holland, J.H. (1975), Adaptation in Natural and Artificial Systems. The University of Michigan Press, Ann Arbor, MI.
Horn, M.E. (2000), “Efficient Modeling of Travel in Networks with Time-Varying Link Speeds,” *Networks,* Vol. 36, No. 2, P 80-90
Horowitz, A.J. (1987), “Extensions of Stochastic Multipath Trip Assignment to Transit Networks,” *Transportation Research Record*, No. 1108, P 66-72.
Hsu, J. and Surti, V.H. (1975), “Framework of Route-Selection in Bus Network Design,” *Transportation Research Record*, No. 546, P 44-57.
Huang, R. and Peng, Z.R. (2002), “Schedule-Based Path Finding Algorithms for Transit Trip Planning Systems,” *Transportation Research Record*, No. 1783, P 142-148.
Hurdle, V.F. (1973), “Minimum Cost Schedules for a Public Transportation Route – I. Theory,” *Transportation Science*, Vol. 7, No. 2, P 109-137.
Hurdle, V.F. (1973), “Minimum Cost Schedules for a Public Transportation Route – I. Examples,” *Transportation Science*, Vol. 7, No. 2, P 138-158.

Imam, M.O. (1998), "Optimal Design of Public Bus Service with Demand Equilibrium," *Journal of Transportation Engineering*, Vol. 124, No. 5, P 431-436.
Israeli, Y. and Ceder, A. (1991), "Transit Network Design," Presented at *the 70th Annual Meeting of Transportation Research Board*, Washington, D.C.
Jayakrishnan, R., McNally, M.G. and Marar, A.G. (1995), "Recursive Structure for Exact Line Probabilities and Expected Waiting Times in Multipath Transit Assignment," *Journal of Transportation Research Record*, No. 1493, P 178-187.
John, R.K. (1993). Genetic Programming, MIT Press, London.
Jung, S. and Haghani A. (2000), "A Genetic Algorithm for Pick-up and Delivery Problem with Time Window," *79th Transportation Research Board Meeting*, Preprint CD-ROM, No. 00-0212, Washington, D.C., 2000.
Jung, S. and Haghani A. (2001), "A Genetic Algorithm for the Time-Dependent Vehicle Routing Problem," *80th Transportation Research Board Meeting*, Preprint CD-ROM, Washington, D.C., 2001.
Kaufman, D.E., and Smith, R. L. (1993), "Fastest Paths in Time-Dependent Networks for Intelligent Vehicle-Highway Systems Application," *IVHS Journal*, Vol. 1, No. 1, P 1-11.
Khisty, C.J. and Lall, B.K. (1998), Transportation Engineering: An Introduction, Second edition, Prentice Hall, NJ.
Kim, K.H., Moon, K.C. (2003), "Berth scheduling by simulated annealing," *Transportation Research Part B*, Vol. 37, P 541-560.
Koncz, N., Greenfeld, J. and Mouskos, K. (1996), "A Strategy for Solving Static Multiple Optimal Path Transit Network Problems," *Journal of Transportation Engineering*, Vol. 122, No. 3, P 218-225.
Koulamas, C., Antony, S.R. and Jaen, R. (1994), "A Survey of Simulated Annealing Applications to Operations Research Problems," *Omega International Journal of Management Science*, Vol. 22, No. 1, P 41-56.
Lee, K.K.T. and Schonfeld, P. (1991), "Optimal Slack Time for Timed Transfers at a Transit Terminal," *Journal of Advanced Transportation*, Vol. 25, No. 3, P 281-308.
Lam, W.H.K., Gao, Z.Y., Chan K.S. and Yang H. (1999), "A Stochastic User Equilibrium Assignment Model for Congested Transit Networks," *Transportation Research Part B: Methodological*, Vol. 33, No. 5, P 351-368.
Lam, W.H.K., Zhou, J. and Sheng, Z.H. (2002), "A Capacity Restraint Transit Assignment with Elastic Line Frequency," *Transportation Research Part* B, Vol. 36, P 919-938.
Lampkin, W. and Saalmans. P.D. (1967), "The Design of Routes, Service Frequencies and Schedules for a Municipal Bus Undertaking: A Case Study," *Operation Research Ouarterly,* No. 18, P 375-397.
Larson, R.C. and Odoni, A.R. (1981), *Urban Operations Research,* Prentice-Hall Inc., Englewood Cliffs, New Jersey.
Law, A.M. and Kelton, W.D. (2003), *Simulation Modeling & Analysis*, Third Edition, McGraw Hill, 2003.
Lawler, E.L. (1977), "Comment on computing the k shortest paths in a graph," *Communication, ACM*, Vol. 20, P 603–604.

LeBlanc, L.J. (1975), "An Algorithm for the Discrete Network Design Problem," *Transportation science*, No. 9, P 183-199.
LeBlanc. L.J. (1988), "Transit System Network Design," *Transportation Research Part B*, Vol. 22, No. 5, P 383-390.
Lee, Y.J. and Vuchic, V.R. (2000), "Transit Network Design with Variable Demand," 79th Annual Meeting of TRB, Washington, DC.
Li, L.Y.O. and Fu, Z. (2002), "The School Bus Routing Problem: A Case Study," *Journal of the Operational Research Society*, Vol. 53, No. 5, P 552-558.
List, G.F. (1990), "Toward Optimal Sketch-Level Transit Service Plans," *Transportation Research Part B*, Vol. 24, No. 5, P 325-344.
Liu, G. and Wirasinghe, S.C. (2001), "A Simulation Model of Reliable Schedule Design for a Fixed Transit Route," *Journal of Advanced Transportation*, Vol. 35, No. 2, P 145-174.
Liu, Y.H. (1994), An Approach for the Characterization and Classification of Bus Transit Network Structure, Master Thesis, Department of Civil Engineering, University of Texas at Austin, Austin, Texas.
Loureiro, C.F., and Ralston, B. (1996), "Investment Selection Model for Multicommodity Multimodal Transportation Networks," *Transportation Research Record*, No. 1522, P 38-46.
Luenberger, D.G. (1984), *Linear and Nonlinear Programming*, Second Edition, Addison-Wesley Publishers.
Mandl, C.E. (1980), "Evaluation and Optimisation of Urban Public Transport Networks," *European Journal of Operational Research*, Vol. 6, P 31-56.
Martins, C.L. and Pato, M.V. (1998), "Search Strategies for the Feeder Bus Network Design Problem," *European Journal of Operational Research*, Vol. 106, No. 2-3, P 425-440.
Marwah, B.R., Umrigar, F.S., and Patnaik, S.B. (1984), "Optimal Design of Bus Routes and Frequencies for Ahmedabad," *Transportation Research Record*, No. 99, Transportation Research Board, Washington D.C., P 12-25.
Michalewicz, Z. (1999), Genetic Algorithms + Data Structure = Evolution Programs, Third Edition, Springer-Verlag, New York.
NCHRP Synthesis of Highway Practice 69 (1980), *Bus Route and Schedule Planning Guidelines,* Transportation Research Board, National Research Council, Washington, D.C.
Newell, G.F., (1979), "Some Issues Relating to the Optimal Design of Bus Routes," *Transportation Science*, Vol. 13, No. 1, P 20-35.
Newton, R., and Thomas, W. (1974), "Bus Routing in a Multi-School System," *Computers and Operations Research*, Vol. 1, P 213-222.
Ngamchai, S. and Lovell, D.J. (2003), "Optimal Time Transfer in Bus Transit Route Network Design using a Genetic Algorithm," *ASCE Journal of Transportation Engineering*, Vol. 129, No. 5, P 510-521.
Nielsen, O.A. (2000), "A Stochastic Transit Assignment Model Considering Differences in Passengers Utility Functions," *Transportation Research Part B: Methodological*, Vol. 34, No. 5, P 377-402.

Nuzzolo, A., Russo, F. and Crisalli, U. (2001), "A Doubly Dynamic Schedule-based Assignment Model for Transit Networks," *Transportation Science*, Vol. 35, No. 3, P 268-285.
Oldfield, R.H. and Bly, P.H. (1988), "An Analytic Investigation of Optimal Bus Size," *Transportation Research Part B*, Vol. 22, No. 5, P 319-337.
Osuna, E.E. and Newell, G.F. (1972), "Control Strategies for an Idealized Public Transportation System," *Transportation Science*, Vol. 6, P 57-72.
Papacostas, C.S. and Prevedouros, P.D. (2001), Transportation Engineering and Planning, 3rd Edition, Prentice Hall, New Jersey, 2001.
Pattnaik, S.B., Mohan, S. and Tom, V.M. (1998), "Urban Bus Transit Network Design Using Genetic Algorithm," *Journal of Transportation Engineering*, Vol. 124, No. 4, P 368-375.
Peng, Z.R. (1997), "A Methodology for Design of GIS-Based Automatic Transit Traveler Information Systems," *Computers, Environment and Urban Systems*, Vol. 21, No. 5, P 359-372
Ralston, B., Tharakan, G., and Liu, C. (1994), "A Spatial Decision Support System for Transportation Policy Analysis," *Journal of Transport Geography*, Vol. 2, P 101-110.
Rapp, M.M. and Gehner, C.D. (1976), "Transfer Optimization in an Interactive Graphic System for Transit Planning," *Transportation Research Record*, No. 619, P 22-29.
Rapp, M.H., Mattenberger, P., Piguet, S., and Robert-Grandpierre, A. (1976), "Interactive Graphic System for Transit Route Optimization," *Transportation Research Record*, Vol. 559, P 73-88.
Rea, J.C. (1971), "Designing Urban Transit Systems: An Approach to the Rout-Technology Selection Problem," PB 204881, University of Washington, Seattle, W A.
Salzborn, F.J.M. (1972), "Optimum Bus Scheduling," *Transportation Science*, Vol. 6, No. 2, P 137-148.
Schneider, J. and Smith, S. (1981), "Redesigning Urban Transit Systems: A Transit Center Based Approach," *Transportation Research Record*, No. 798, Transportation Research Board, Washington, D.C., P 56-65.
Schwefel, H.P. (1981), *Numerical Optimization of Computer Models*, John Wiley & Sons Ltd, Chichester, U.K.
Sheffi, Y. and Sugiyama, M. (1982), "Optimal Bus Scheduling on a Single Route," *Journal of Transportation Research Record*, No. 895, P 46-52.
Shier, D.R. (1979), "On Algorithms for Finding the k Shortest Paths in a Network," *Networks*, Vol. 9, P 195 - 214.
Shih, M.C. and Mahmassani, H.S. (1994), "A Vehicle Sizing Model for Bus Transit Systems," *Presented at the 73rd Annual Meeting of the Transportation Research Board*, Washington, D.C.
Shih, M.C. (1994), A Design Methodology for Bus Transit Route Networks with Coordinated Operations, Ph.D. thesis, Department of Civil Engineering, University of Texas, Austin, Texas.
Shih, M.C., Mahmassani, H.S. and Baaj M.H. (1997), "Trip Assignment Model for Timed-Transfer Transit Systems," *Transportation Research Record*, No. 1571, P 24-30.

Shih, M., Mahmassani, H.S. and Baaj, M. (1998), "A Planning and Design Model for Transit Route Networks with Coordinated Operations," Paper No. 980418, *77th Annual Meeting of TRB*, Washington, DC.
Shrivastava, P. and Dhingra, S.L. (2002), "Development of Coordinated Schedules using Genetic Algorithms," *Journal of Transportation Engineering*, Vol. 128, No. 1, P 89-96.
Silman, L.A., Barzily, Z., and Passy, U. (1974), "Planning the Route System for Urban Buses," *Computers and Operations Research*, Vol. 1, P 201-211.
Simon, J. and Furth, P.G. (1985), "Generating a Bus Route O-D Matrix from On- Off Data," *Journal of Transportation Engineering*, Vol. 111, No. 6, P 583- 593.
Solanki, R.S., Gorti, J.K. and Southworth, F. (1998), "Using Decomposition in Large-scale Highway Network Design with a Quasi-optimization Heuristic," *Transportation Research Part B: Methodological*, Vol. 32, No. 2, P 127-140.
Spasovic, L.N. and Schonfeld, P.A. (1993), "Method for Optimizing Transit Service Coverage," *Transportation Research Record*, No. 1402, P 28-39.
Speiss, H. and Florian, M. (1989), "Optimal Strategies: A New Assignment Model for Transit Networks," *Transportation Research Part B*, Vol. 23, No. 2, P 83- 102.
Taylor-Harris, A. and Stone, T. J. (1980), "Transit Center: A Means of Improving Transit Services," *Transportation Research Record*, No. 760, P 39-42.
Tom, V.M. and Mohan, S. (2003), "Transit Route Network Design Using Frequency Coded Genetic Algorithm," *Journal of Transportation Engineering*, Vol. 129, No. 2, P 186-195.
Tong, C.O. and Wong, S.C. (1999), "A Stochastic Transit Assignment Model Using a Dynamic Schedule-based Network," *Transportation Research Part B*, Vol. 33, P 107-121.
Tong, C.O. and Wong, S.C. (1999), "A Schedule-based Time-dependent Trip Assignment Model for Transit Networks," *Journal of Advanced Transportation*, Vol. 33, P 371-388.
Tong, C.O. and Wong, S.C. (2000), "A Predictive Dynamic Traffic Assignment Model in Congested Capacity-Constrained Road Networks," *Transportation Research Part B*, Vol. 34, P 625-644.
Tong, C.O., Wong, S.C., Poon, M.H. and Tan, M.C. (2001), "A Schedule-based Dynamic Transit Network Model - Recent Advances and Prospective Future Research," *Journal of Advanced Transportation*, Vol. 35, No. 2, P 175-195.
Van Nes, R., Hamerslag, R., and Immer, B.H. (1988), "The Design of Public Transport Networks," *Transportation Research Record*, No. 1202, Transportation Research Board, Washington, D.C., P 74-83.
Van Nes, R. and Bovy, P.H.L. (2000), "Importance of Objectives in Urban Transit Network Design," *Transportation Research Record*, No. 1735, P 25-34.
Van Nes, R. (2002), "Multi-level Network Optimization for Public Transport Networks," Presented in 81st TRB Annual Meeting, Paper 02-2493, Washington, D.C.
Victor, D.J. and Santhakumar, S.M. (1986), "Simulation study of bus transit," *Journal of Transportation Engineering*, American Society of Civil Engineers, Vol. 112, No. 2, P 199-211.

Vuchic, V.R. (1981), Urban Public Transportation: Systems and Technology, Prentice-Hall, Inc., Englewood Cliffs, N.J. 07632.
Wellman, M.P., Ford, M. and Larson, K. (1995), "Path Planning under Time-Dependent Uncertainty," *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence* (*UAI-95*), Montreal, Quebec, Canada, August, 1995.
Willoughby, K.A. (2002), "A Mathematical Programming Analysis of Public Transit Systems," *Omega*, Vol. 30, No. 3, P 137-142.
Willoughby, K.A. and Uyeno, D.H. (2001), "Resolving Splits in Location/Allocation Modeling: a Heuristic Procedure for Transit Center Decisions," *Transportation Research Part E: Logistics and Transportation Review*, Vol. 37, No. 1, P 71-83.
Uyeno, D.H. and Willoughby, K.A. (1995), "Transit Center Location-Allocation Decisions," *Transportation Research Part A: Policy and Practice*, Vol. 29, No. 4, P 263-272.
Wilson, N.H.M. and Gonzalez, S.L. (1982), "Methods for Service Design," *Transportation Research Record*, No. 862, Transportation Research Board, Washington, D.C., P 74-83.
Wolsey, L.A. (1998), *Integer Programming*, John Wiley & Sons, New York, 1998.
Wong, S.C. and Tong, C.O. (1998), "Estimation of Time-dependent Origin-destination Matrices for Transit Networks," *Transportation Research Part B*, Vol. 32, P 35-48.
Woodhull, J., Simon, J. and Shoemaker, D.A. (1985), "Goals for Bus Transit Scheduling," *Journal of Transportation Research Record*, No. 1011, P 72-76.
Xiong, Y. and Schneider, J.B. (1992), "Transportation Network Design using a Cumulative Genetic Algorithm and Neural Network," *Transportation Research Record*, No. 1364, Transportation Research Board, Washington D.C., P 37-44.
Yen, J.Y. (1971), "Finding the K Shortest Loopless Paths in a Network," *Management Science*, Vol. 17, No. 11, P 712-716.
Yuratovac, D.G. (1982), "Bus Route-Level Demand Modeling," *Journal of Transportation Research Record*, No. 862, P 16-22.
Zhan, F.B. (1997), "Three Fastest Shortest Path Algorithms on Real Road Networks: Data Structures and Procedures," *Journal of Geographic Information and Decision Analysis*, Vol. 1, No. 1, P 69-82.
Zhan, F.B., and Noon, C.E. (1998), "Shortest Path Algorithms: An Evaluation Using Real Road Networks," *Transportation Science*, Vol. 32, P 65-73.
Zhan, F.B., and Noon, C.E.A (2000), "Comparison Between Label-Setting and Label-Correcting Algorithms for Computing One-to-One Shortest Paths," *Working paper.*
Zhao, F., Zeng, X. and Ubaka, I. (2003), "Transit Network Optimization-Minimizing Transfers and Optimizing Route Directness," *Working paper*, Department of Civil & Environmental Engineering, Florida International University, Miami, Florida.