1-10-2014

Multi-objective Approaches in Transit Stop Location Problem

Sha Mamun
msm08014@engr.uconn.edu

Multi-objective Approaches in Transit Stop Location Problem

Sha A. Mamun, PhD

University of Connecticut, 2014

An efficient bus network design is a difficult problem, and hence is usually considered as a series of sub-problems solved sequentially. The bus stop location problem (BSLP) is foremost in this series of problems. The BSLP in this research is formulated as a mixed integer problem (MIP) that consideres three important design sub-problems: stop location, demand allocation and service frequency. Given a set of candidate bus stops, the level of trip connectivity and origin-destination pair-wise trip demands, the problem is to answer the following related questions. How many bus stops should be required and where should they be located? Which users/demands are to be assigned to each bus stop? What should be the service frequency for each bus line while being feasible in fleet requirements? The BSLP is formulated as multi-objective approaches which address the tradeoffs between improving accessibility to transit service with more stops and increasing the efficiency of the transit system with fewer stops. Transit access is measured as the physical proximity to transit service assuming that additional stops can provide greater access to the service and reduce walking distance to stops. Trip connectivity is measured using a transit performance measure namely Transit Opportunity Index (TOI) for quantifying the ease of reaching a destination from a given location. TOI is formulated by integrating transit accessibility (spatial and temporal coverage) with topological network connections and travel time (trip coverage) to quantify public transit opportunity. The fewer the number of stops along a line can result passengers' faster travel to destinations with less dwell time at stops. General Algebraic Modeling System (GAMS) software is used for modeling this

problem and MIP CBC (COIN-OR Branch and Cut) solver is used to solve the problem for optimal solutions. Limitations on computational efficiency for resulting optimal solutions are identified. A heuristic (Randomized Feature Selection Algorithm) algorithm is developed to speed computation and provide near optimal solutions. Computation results for TOI, the BSLP model solutions for both the algorithms (i.e., Branch and Cut algorithm, and Randomized Feature Selection algorithm) are given for a bus network in City of New Haven, CT. The computational results of the heuristic algorithm and the CBC solver are provided to show the efficiency of the proposed heuristic algorithm. The randomized feature selection algorithm is shown to be very efficient in terms of time compared with the known optimal solutions generated by CBC solver.

Multi-objective Approaches in Transit Stop Location Problem

Sha A. Mamun

B.S., Bangladesh University of Engineering and Technology, 2006

M.S., University of Connecticut, 2011

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of Doctor of Philosophy

at the

University of Connecticut

2014

2014

APPROVAL PAGE

Doctor of Philosophy Dissertation

Multi-objective Approaches in Transit Stop Location Problem

Presented by

Sha A. Mamun, B.S., M.S.

Major Advisor ______________________________
Dr. Nicholas E. Lownes

Associate Advisor ______________________________
Dr. John N. Ivan

Associate Advisor ______________________________
Dr. Norman W. Garrick

Associate Advisor ______________________________
Dr. Jeffrey P. Osleeb

University of Connecticut
2014

ACKNOWLEDGMENTS

No volume of words is enough to express my gratitude towards my major advisor, Dr. Nicholas E. Lownes for his encouragement, time, patience and trust. I would like to thank him for his unconditional support and mentorship, and would like to express my sincere gratitude for his kind assistance throughout my doctoral education. I am very grateful to my other advisors, Dr. John N. Ivan, Dr. Norman W. Garrick and Dr. Jeffrey Osleeb for their valuable contributions to my doctoral research. Without their support it would have been impossible for me to achieve this goal successfully. I believe every one of them have taught me something that marked my path to achieve my goal. For that and many other reasons I am very grateful to them.

I very much appreciate the support and encouragement of my friends and colleagues in the transportation engineering program at the University of Connecticut who have been part of this journey. I wish the best to them and their families. To my friends everywhere they are I say thank you. They were always happy for me and had no doubt I will be successful in my endeavors. I am glad they share my happiness and wish them the best.

I would like to thank my family who has unconditionally believed in me and supported me all the way through this journey, they are the reason I always want to become a better person. I would like to thank my parents who raising me to be happy, healthy, motivated and loved. I wouldn't be here without their love and sacrifice. I am always thankful and extremely happy to have the most beautiful nephews and nieces: Rafi, Aishy, Raiyaan, Esha and Araf.

Special thanks to Payel my best friend and wife. She is my everyday inspiration to become a better human being. I thank Payel for being with me every time I needed support in the rough times and for helping me to realize that my dreams will come true if I truly believe.

This dissertation is dedicated to the memory of my beloved nephew Prince. He was the greatest inspiration for me to become the best in everything I do. May God have mercy on him and make him happy in heaven. I cannot wait until he runs into my arms again.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

1.1. Background

Developing a transit system to serve passengers in an effective and sustainable way is important for transport experts and urban planners. Public transit is considered a key component of a sustainable transportation system that creates livable communities and provides travelers with greater opportunity, choice, and access to a variety of economic and social activities. Provision of public transit infrastructure will not, in itself, fulfill public transit potential. The system needs to be accessible and available to trip makers and connected to activity centers. Accessibility, one of the most important components for evaluating public transit supply, measures peoples' ease and convenience in reaching transit services. In addition to accessibility, origin-destination (O-D) trip connectivity provides a critical transit supply measure as it considers trip coverage (i.e., whether a transit service is connected travelers' trip origins and destinations).

In terms of increasing transit network efficiency, assessment of service performance will help for better targeting of service investment/improvement policies. The efficiency of a transit system can be significantly improved if the location of bus stops is optimized considering the service performance evaluation measures (i.e., transit accessibility and trip connectivity) as the optimization criteria. For an efficient transit network design, the optimal number of transit stops need to maintain and adjust transit line frequencies to serve the transit demand have been considered as important objectives in transportation planning. Although accessible transit service with shorter access distance (more stops) is an important means for increasing transit ridership, poor service quality (higher travel time, less trip connectivity) resulting from more stops contributes to low transit patronage. More stops and greater access decreases opportunities/destinations reachable within a given travel time budget. Thus both the transit

access and service efficiency (less stops) are critical considerations in transit stop location problem. A mathematical modeling approach that maximizes the network service efficiency without sacrificing current access and adjusts the transit line frequencies simultaneously to meet transit demands will be an effectual design approach to support strategic and operational analysis of transit service efficiency.

A number of well-known options exist for implementing mathematical stop location models. In general, linear optimization problems for networks of small size or with one or two decision variables are modeled in different commercial optimization software, such as General Algebraic Modeling System (GAMS) and different built-in solvers in the package, such as CBC, CPLEX, Lingo, etc., are used to generate optimal solutions based on model formulation and optimization characteristics. However, modern real world problems tend to be very sophisticated in determining multiple decision variables and relate to analysis of large data sets for a network considering multi-objective optimization. Even if the commercial solvers can be used to solve the problem, sometimes its unacceptable run time and computational complexity may turn out inefficient modeling approach of the problem. Therefore, heuristic approach is necessary for the optimization of a network of realistic size in which many parameters need to be determined. In that case, the model can be programmed in a language such as C++, Visual Basic, Java or Python to implement heuristics or search algorithm. In general, heuristic algorithms can be used to generate a rapid solution close to the best possible solution; however, it is necessary to evaluate the quality of solution generated by heuristics with comparison to the solution generated by exact algorithms.

1.2. Objectives

Three distinct objectives will be considered in this research.

i. This research will develop a transit performance measure that considers both the transit accessibility and trip connectivity. A transit opportunity index (TOI) will be used to evaluate the transit service performance available for each O-D pair.

ii. The next objective of this proposed research is to correlate the transit performance measures to the transit network design problem. To this end, a multi-objective transit stop location problem will be formulated to minimize the walking distance that potential users travel while maximizing service efficiency. This problem will determine the optimal number and location of transit stops in a transit network, allocation of trip demand to the bus stops and the optimal transit line frequencies.

iii. Another objective of this research is to develop a heuristic algorithm through which the multi-objective transit stop location model can be solved effectively and efficiently for a large and real world network. This algorithm will seek to solve the problem with computational efficiency and to capable of generating good solutions.

1.3. Dissertation Outline

This dissertation has five chapters including this first introductory chapter. The remainder of this dissertation is organized as follows. The second chapter is devoted to the development of a transit performance measure. A paper titled "A Method to Define Public Transit Opportunity Space" published in the *Journal of Transport Geography* represents this chapter. This paper introduces a new transit performance measure, the Transit Opportunity Index (TOI), which considers both transit accessibility and trip connectivity.

The detail on mathematical formulations, solution method and results for the bus stop location problem (BSLP) are presented in Chapter 3. This chapter consists of a paper titled “Access and Connectivity Tradeoffs in Transit stop Location Problem”, which is under review for publication in *Transportation Research Record: Journal of the Transportation Research Board*. This paper details the multi-objective optimization formulation for selecting an optimal set of transit stops based on tradeoffs between transit user access and system connectivity. The problem is formulated to select stops from the existing set of stops serving any number of transit lines. It is assumed that the lines remain fixed – stops either continue to be maintained or not. The assignment of demand to stops is also provided through this formulation, assuming that passengers know and utilize the nearest stop with which to make their desired trip. CBC (COIN-OR Branch and Cut) solver is used to solve the problem for optimal solutions.

A randomized feature selection algorithm to solve the BSLP is introduced in chapter 4. The aim of this chapter is to demonstrate that the use of feature selection algorithm is much more efficient for solving the bus stop location problem than the Branch and Cut algorithm used for optimal solutions in chapter 3. It is shown that accuracy is close to already known optimal solutions using CBC solver but the saving in run time is immense.

Conclusions and future research are presented in the final chapter. A brief review of results and the potential applications of developed BSLP model are described. A wide variety of future research questions and suggestions for further refinement of the BSLP model are identified.

This paper is published in the* ***Journal of Transport Geography*

CHAPTER 2: A METHOD TO DEFINE PUBLIC TRANSIT OPPORTUNITY SPACE

Abstract

A public transit performance measure quantifying the ease of reaching a destination from a given location is important for describing the efficiency and convenience of public transit. In this paper, a new method for quantifying public transit performance, the Transit Opportunity Index (TOI), is presented. This measure accounts for both transit accessibility (the level of access to the transit system) and transit connectivity (the system's provision of services between origins and destinations) by combining measures of spatial coverage, temporal coverage, and trip coverage. Spatial and temporal coverage measures are calculated using an origin-destination (O-D) representation of the transit network and then combined to create a transit accessibility score for each O-D pair. Transit accessibility is weighted by a binary connectivity parameter and a connectivity decay factor. The connectivity decay factor is derived from a travel time-based logistic function to reflect the decreasing connectivity with increasing travel time. The binary connectivity parameter and the connectivity decay factor are used to account for trip coverage, or transit connectivity. The Transit Opportunity Index (TOI) is then applied to the bus network of the city of New Haven, Connecticut. The results of this case study suggest that the TOI is a more complete and practical measure of public transit service performance than previously established measures. This method also has the potential to identify transfer zones for public transit trips between O-D pairs without direct connections. However, the TOI is most powerful when used in conjunction with a public transit demand measure to identify underserved areas.

2.1. Introduction

A major concern in the public transit sector has been the adequate assessment of access to transit service. Measures of transit accessibility are important in evaluating existing transit services, allocating transportation investments, and making decisions on land development (*1*). Many transit accessibility metrics exist reflecting the various perspectives of their developers and advocates. Generally these metrics are constructed from a combination of service characteristics and physical proximity to public transit service. To varying degrees they characterize the ease of access to the service with regard to spatial coverage and therefore have been recognized as an important tool for improving the quality of transit service. With respect to spatial coverage, a service is accessible when a large proportion of the population lives within a reasonable walking distance of a public transit station while with temporal coverage a service is accessible when its available at times customers want to travel. Most of the existing tools for measuring transit accessibility account for service frequency and distance to a station within a buffer area, albeit at different resolutions. Very few consider trip coverage, whether a public transit service is available for specific trip origin/destinations. Therefore, a public transit performance measure which integrates trip coverage with spatial and temporal coverage provides a more powerful and practical description of transit quality of service.

The objective of this paper is to introduce a new public transit quality of service measure, the Transit Opportunity Index (TOI), which considers both transit accessibility and transit connectivity. The TOI provides a standard framework for evaluating public transit service on the basis of spatial and temporal coverage measures (transit accessibility) and trip coverage measures (origin-destination connectivity). This paper is organized into five sections, including this introduction section. Section 2.2 presents a detailed review of the relevant literature on

measuring public transit service performance and transit connectivity. Section 2.3 describes the methodology for quantifying transit opportunity using the TOI. A numerical example is then used to illustrate the computations required for each step. In Section 2.4, the public transit performance measure is applied to the transit network of the city of New Haven, Connecticut. Finally, Section 2.5 summarizes the findings of this study and provides suggestions for future improvements to the TOI.

2.2. Literature Review

2.2.1 Transit Accessibility

The performance of a public transit network can be measured as its ability to meet mobility and economic needs efficiently, equitably, and in an environmentally sound manner. Considerable research has been conducted on the use of transit accessibility measures to evaluate public transit service performance. In general, transit accessibility refers to the ability of travelers to reach transit facilities (*2*). Numerous measures have been suggested for quantifying transit accessibility, each using different definitions of service coverage as appropriate for their varying purposes.

The transit service catchment area, a measure of spatial coverage, is an important determinant for transit accessibility. The study by Beimborn et al. (*3*) suggests that transit users will consider other transit accessibility factors such as cost, comfort, security, etc., provided the presence of spatial accessibility to transit. However, most studies of spatial accessibility focus on physical access, usually walking distance to a transit stop (*4 - 9*). The Transit Capacity and Quality of Service Manual (TCQSM) (*10*) provides a standard method for assessing transit accessibility/availability. This method uses a static ¼ mile buffer around each bus stop to define

the spatial coverage of a bus service. This radius is modified for various features, including street connectivity, grade, % elderly in the population, and the number of pedestrian crossings encountered (*11*). According to Grava (*12*), one can expect a local bus service to attract travelers within a ¼ mile of a bus stop, whereas light rail can attract travelers within ½ mile. These standard walking thresholds are supported in most transit research (for example, references *2, 5, 7, 9, 13, 14, 15*).

There are, however, alternative methods for defining the spatial coverage of a transit system. Similar to the previously mentioned methods, the transit level-of-service (TLOS) indicator developed by Ryus et al. (*16*) assumes that people within a 5 minute or ¼ mile walk of a stop/station have access to that stop. However, instead of using a strict Euclidean buffer as the case of the previously mentioned performance measures, the TLOS considers the existence and eminence of pedestrian routes connected to transit stops. Kuby et al. (*14*) used a half mile walking distance to define light-rail station service areas. This study used a raster-based approach to calculate the shortest network paths instead of Euclidean distances. Polzin et al. (*17*) used a ½ mile buffer around a transit route, rather than the location of the transit stop, to calculate the service area. O'Neil et al. (*15*) developed the network-ratio method to measure accessibility to transit services which measures the transit accessibility of a zone as the proportion of the street network within walking distance of transit services.

Several existing methods for quantifying transit accessibility incorporate temporal coverage measures alongside the previously discussed spatial coverage measures. The Time-of-Day-Based Transit Accessibility Analysis Tool developed by Polzin et al. (17) is an example of a transit accessibility measure that incorporates temporal and spatial coverage into a single measure. This study suggested that both the supply and demand dimensions of temporal

coverage are important for evaluating transit accessibility. The supply side of temporal coverage measures includes service frequency, time span of service, service headway, etc., while the demand side measure of temporal coverage considers the importance of service provided in each time period of the day. Service frequency (vehicle runs per hour) is used to measure the supply side and the time-of-day travel demand distribution is incorporated to measure the demand side. The Local Index of Transit Availability (LITA), developed by Rood (*18*), measures the transit service intensity or transit accessibility in an area by integrating a weekly average of service frequency (temporal coverage), seat-miles per residential and employment population , and transit stops per square mile (spatial coverage). This tool incorporates transit service, comfort and convenience into its transit accessibility calculations. Schoon et al. (*19*) formulated another set of transit accessibility indices, the Travel Time Accessibility Index (AI) and the Travel Cost Accessibility Index (AI) for different modes between O-D pairs. The Travel Time AI of a mode between a specific O-D pair is the ratio of the travel time between the O-D pair using the specified mode to the average travel time across all modes. Travel Cost AIs are calculated using a similar method.

Recently, Mamun and Lownes (*1, 20*) developed a composite index for calculating transit accessibility. This approach combines the various components of existing methods to reflect various perspectives of transit accessibility (i.e., transit planner, transit operator, traveler and property developer). This composite index also simultaneously characterizes three important aspects of transit accessibility, spatial coverage, temporal coverage, and comfort.

A review of the above studies reveals that much of the research accounts for temporal coverage using service frequency for the evaluation of transit accessibility. Some literature considers trip coverage for measuring transit accessibility from an area to trip destinations, where

destinations were considered as disaggregate attractions or opportunities (*21, 22, 23*). However, few studies consider O-D pair-wise transit accessibility (e.g., census tract to census tract or parcel to parcel) where destinations are treated as an aggregated concept as opposed to disaggregate activity/opportunity locations. Incorporating O-D trip coverage into a transit accessibility measure is important as trip coverage is a primary consideration of travelers using public transportation and is an important factor for describing the efficiency and convenience of a transit network (*24*).

2.2.2. Transit Connectivity

Transit connectivity measures have been proposed representing different points of view. Some transit connectivity measures consider the comfort of transfer stations, some use topological connections and some consider travel time as the estimator of transit connectivity. Very few studies tie the level of transit accessibility to topological network connections and travel time to estimate transit connectivity for a multi-route transit network. Moreover, most of the transit connectivity measures are developed without considering origin-destination (O-D) pair-wise transit connectivity.

Lam and Schuler (*25*) formulated a transit connectivity index for the purpose of evaluating transit system design and transit performance. This measure integrated travel time and transit accessibility with transit route and schedule structure. Lee and Lee (*26*) developed a graph-theoretic nodal accessibility measure that identifies the total connectivity of a node. The gamma index, a measure of transit connectivity described by Rodrigue (*27*), considers the relationship between the number of observed links and the number of possible links. However, the gamma index only accounts for network topology in evaluating network performance (*28*).

Fu et al. (*29*) proposed an O-D based approach called the Transit Service Indicator (TSI) to evaluate transit network accessibility by combining the various temporal attributes into one composite measure. Derrible and Kennedy (*30*) measured transit connectivity by considering the comfort aspects of transfer stations and the topological connections. Hadas and Ceder (*31*) developed a network connectivity measure based on the physical and spatial characteristics of transit routes. They used service reliability and comfort of transfer as estimators for the transit connectivity performance indicator. Hadas et al. (*32*) developed a tool for assessing public transit connectivity based on time and transfer by mode.

The contributions of this paper are:

- A measure to quantify public transit opportunity or the ease of reaching a destination from a given location using public transit.
- Integrating transit accessibility (spatial and temporal coverage) with topological network connections and travel time (trip coverage) in a new transit service performance measure.
- Accounting for O-D pair-wise transit connectivity with binary connectivity and decay factors.

This paper synthesizes public transit performance measures and perspectives into an overall performance measure of transit opportunity. The goal of this tool is to measure the ability of a service to provide an origin with access to destinations.

2.3. Methodology

The development of the proposed transit opportunity index (*TOI*) requires several steps (Figure 2.1). The first step is to estimate transit accessibility (A_{ijl}) for each O-D pair in a transit network. One begins by calculating spatial coverage (R_{il}) for each origin or destination (census tracts in

this analysis) and service frequency (S_{ijl}) for each O-D pair. Transit accessibility is then measured using a combination of spatial coverage and per capita service frequency (temporal coverage). A binary connectivity parameter (δ_{ijl}) is then aggregated over all transit lines based on the existence of a direct route from an origin to a destination on a particular transit line. The next task is to develop a decay function (f_{ijl}) to reflect decreasing connectivity with increased travel time. The final step is to calculate the Transit Opportunity Index for each O-D pair (TOI_{ij}) using transit accessibility and the connectivity parameter with decay. The following subsections describe each of these steps in greater detail. The following notation is used throughout this paper.

<u>Notation</u>

i	Set of origins *(e.g., Census Tracts)*
j	Set of destinations *(e.g., Census Tracts)*
l	Set of transit lines
P_i	Population of origin *i*
U	Capacity of a vehicle
R_{il}	Spatial coverage score of origin *i* for transit line *l*
$B_{i,l,buffer}$	A quarter-mile buffer area of origin *i* for transit line *l (Square-miles)*
$B_{i,total}$	Total area of origin *i (Square-miles)*
S_{ijl}	Service frequency score from origin *i* to destination *j* for transit line *l*
V_{ijl}	Daily vehicle runs from origin *i* to destination *j* for transit line *l*
A_{ijl}	Transit accessibility score from origin *i* to destination *j* for transit line *l*
δ_{ijl}	Binary connectivity parameter, takes the value 1 if a transit line *l* directly connects origin *i* to destination *j*, and 0 otherwise.

TOI_{ij}	Transit opportunity index from origin *i* to destination *j*
TOI_i	Transit opportunity index of origin *i*
T_{ijl}	Travel time from origin *i* to destination *j* for transit line *l*
f_{ijl}	Connectivity decay factor from origin *i* to destination *j* for transit line *l*
T_{ij}^k	Travel time from origin *i* to destination *j* with transfer at *k*
f_{ij}^k	Connectivity decay factor from origin *i* to destination *j* with transfer at *k*
TOI_{ij}^k	Transit opportunity index from origin *i* to destination *j* with transfer at *k*

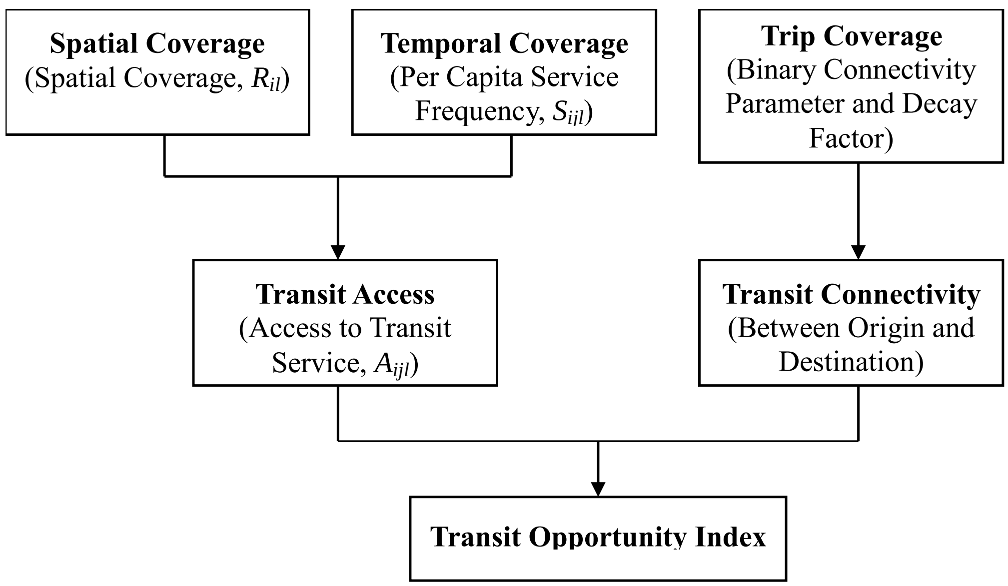

Figure 2.1. Transit opportunity index framework.

2.3.1. Transit Accessibility Score

2.3.1.1. *Spatial Coverage*

Spatial coverage *(R_{il})* is the proportion of the area of origin *i* served by the transit line *l*. It is computed using the ratio of the spatial coverage area of a transit line *($B_{i,l,buffer}$)* to the total area

($B_{i,total}$) (Eq. 1 below). A spatial coverage area is usually measured as the area covered by a particular route or by the area within a walking distance threshold of a transit stop or transit route (*10*). It is commonly accepted among transit planners and researchers that bus transit users are willing to walk up to ¼ mile (400 m) to reach their nearest stop (*2*, *4*, *9*, *33*). This paper also uses this definition of spatial coverage area and a threshold value of a ¼ mile buffer around bus lines for calculating the spatial coverage score. Line buffers were used as opposed to stop buffers to avoid overlapping coverage areas between closely spaced stops. A limitation of this approach is the overestimation of spatial coverage along lines where stops are spaced further apart.

$$R_{il} = \frac{B_{i,l,buffer}}{B_{i,total}} \quad (1)$$

2.3.1.2. *Per Capita Service Frequency (Temporal Coverage)*

This research only considers the supply side of temporal coverage in measuring relative transit service supply. Per Capita Service frequency (S_{ijl}) is measured by the *daily available seats per capita* from an origin *i* to a destination *j* (Eq. 2). First, the weekly average of vehicles runs (V_{ijl}) is determined for each transit line *l* from origin *i* to destination *j*. The daily available seats per capita is then calculated as the ratio of the product of daily bus runs from *i* to *j* and bus capacity (*U*) to the total population (P_i) of origin *i*.

$$S_{ijl} = \frac{V_{ijl}U}{P_i} \quad (2)$$

Spatial coverage and per capita service frequency are combined to obtain transit accessibility (A_{ijl}) for each *i-j* pair in the network connected by transit line *l* (Eq. 3).

$$A_{ijl} = R_{il}\, S_{ijl} \quad (3)$$

Transit accessibility accounts for two important access measures: the share of an area which has access to transit service and the level of service provided by public transit to those

people who have access to the service. Per capita service frequency estimates the level of service provided by the transit service for the total population of an area (e.g., census tract) not just for the people who have access to the service. For example, if two areas have same scores of per capita service frequency, one would expect the same level of access for the two areas. However, if the two areas have different spatial coverage of transit service, the result may be misleading. The Sample illustration below provides clarification. A weakness of this approach is the assumption that the population is uniformly distributed throughout the area. The method also assumes that all people within a buffer area have equal access to the transit service without considering the actual pedestrian network within the area.

2.3.2. Connectivity Parameter

Most existing transit accessibility indices measure the level of ease of using transit service from a particular stop or area. However, this study formulates transit accessibility for each O-D pair. A binary connectivity parameter (δ_{ijl}) is included with spatial coverage and per capita service frequency to measure access to possible trip destinations from a particular trip origin. This connectivity parameter represents the presence or absence of a direct transit route[1] (i.e., trip coverage) between O-D pairs. Thus, the connectivity parameter, δ_{ijl}, takes the value 1 if a transit line l directly connects origin i to destination j, and 0 otherwise.

2.3.3. Travel Time-based Logistic Function

The concept of connectivity decay introduced in this study is motivated by the literature on walking distance decay functions for transit demand estimation (*8*, *9*). Gutiérrez et al., (*33*)

[1] The term 'direct transit route' refers here a direct connection by a single transit line not considering any transfer between lines. The TOI calculation is performed for each individual line in the first step and then aggregate across all direct lines to obtain the TOI for each O-D pair. Using these TOI results we later present an approach to calculate the TOI for O-D pairs requiring transfer between lines.

developed a transit ridership forecast model using distance decay functions for delimiting the service area in terms of demand within station catchment areas. Unsurprisingly, these studies assume that the number of transit users decrease with an increase in walking distance to transit stops. Similarly, Wibowo and Olszewski (*34*) utilize a binary logit model to estimate the probability of an individual choosing walking to access transit service based on the difference in the utility between walking and non-walking. However, the connectivity decay function used in that study considered only the transit trip distance, not the access/egress distance.

Intuitively, passengers would have better transit connectivity between an O-D pair if it takes less travel time to make a trip. The binary connectivity parameter (δ_{ijl}) introduced earlier assigns the same level of connectivity to all destinations regardless of the trip distance, which is counterintuitive. To address this concern, a logistic decay function estimates the decreasing connectivity based on door-to-door travel time. The decay function used in this research implies that the level of transit connectivity decreases gradually up to certain travel time, then more rapidly in the middle range of travel time, and gradually again at the end (an inverted s-curve). Although actual populations are unlikely to follow the theoretical logistic decay curve exactly, the curve agrees with empirical evidence on access distance decay (*12*).

The travel time (T_{ijl}) calculation of a trip from origin i to destination j for transit line l considers access time, wait time, in-vehicle time, and egress time (Eq. 4). Both the access and egress times are assumed to be 5 minutes of walking time, which transit users are generally willing to undertake (*10*, *12*, *16*). Waiting time at transit stops is assumed to be one-half of the scheduled headway when the average headway of transit service is around 10 minutes (*35, 36*). For a scheduled headway greater than 10 minutes, the waiting time for transit service is taken as 10 minutes. The 10 minute default value represents the average wait for transit time from the

National Household Travel Survey (NHTS) data (*17*). The in-vehicle travel time is calculated using the scheduled arrival and departure time that is obtained from transit service schedules.

$$T_{ijl} = T_{access} + T_{wait} + T_{in-vehicle} + T_{egress} \quad (4)$$

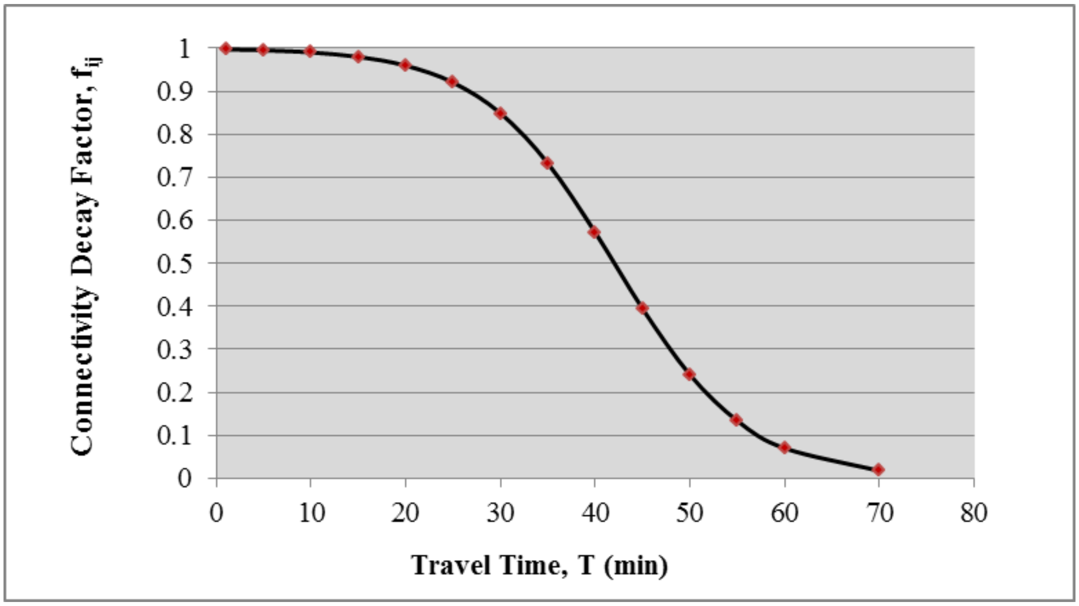

Figure 2.2. Estimation of decay function parameters.

The logistic decay curve is depicted graphically in Figure 2.2. The functional form of this curve is expressed in Eq. 5. Here, *T* represents the travel time in minutes; *L* is the upper limit of the connectivity factor which is assumed to be 1.0 in this study (and represents no decay in the connectivity of an O-D pair). In order to estimate coefficient values of *α* and *β*, we used average travel time data obtained from the 2009 American Community Survey (*37*)[2]. The survey data shows that in Connecticut 93% of the people make their work trips in less than 60 minutes. Therefore, we assumed that the connectivity from an origin to a destination that takes 60 minutes travel time would be only 0.07, where the upper limit of connectivity factor is 1.0. It is assumed

[2] Anyone having travel time data available for their city or region can estimate parameters *α* and *β* to best fit the decay curve for their application. This study used commute travel time data from Connecticut for the application to New Haven, Connecticut.

that the value of the connectivity factor for a trip of less than or equal to 10 minutes of travel time would be 1.0 as the first 10 minutes of total travel time comprises the 5 minutes walking time and 5 minutes waiting time only. The resulting estimates of the parameters using Connecticut commute data are α = 0.0024019 and β = - 0.1436361.

$$f_{ijl} = \frac{L}{1+\propto e^{-\beta T}} \tag{5}$$

2.3.4. Transit Opportunity Index

The binary connectivity parameter (δ_{ijl}) is multiplied by the connectivity decay factor (f_{ijl}) for estimating decreasing connectivity. This results in unconnecting (δ_{ijl} = 0) O-D pairs contributing nothing to the opportunity score and very distant trips (f_{ijl} approaching 0) which contribute very little. This product is multiplied by the transit accessibility score, A_{ijl} to weight the opportunity index by the transit accessibility of the O-D pair and shown in (Eq. 6). Products are summed for all transit lines for each O-D pair and divided by the sum across all O-D pairs to obtain a relative picture of transit opportunity for each O-D pair in the network. The TOI can also be used to quantify the level of opportunity of an origin *i* or a destination *j* as shown in Equations 7 and 8.

$$TOI_{ij} = \frac{\sum_l A_{ijl}\, \delta_{ijl} f_{ijl}}{\sum_i \sum_j \sum_l A_{ijl}\, \delta_{ijl} f_{ijl}} \tag{6}$$

$$TOI_i = \sum_j TOI_{ij} \tag{7}$$

$$TOI_j = \sum_i TOI_{ij} \tag{8}$$

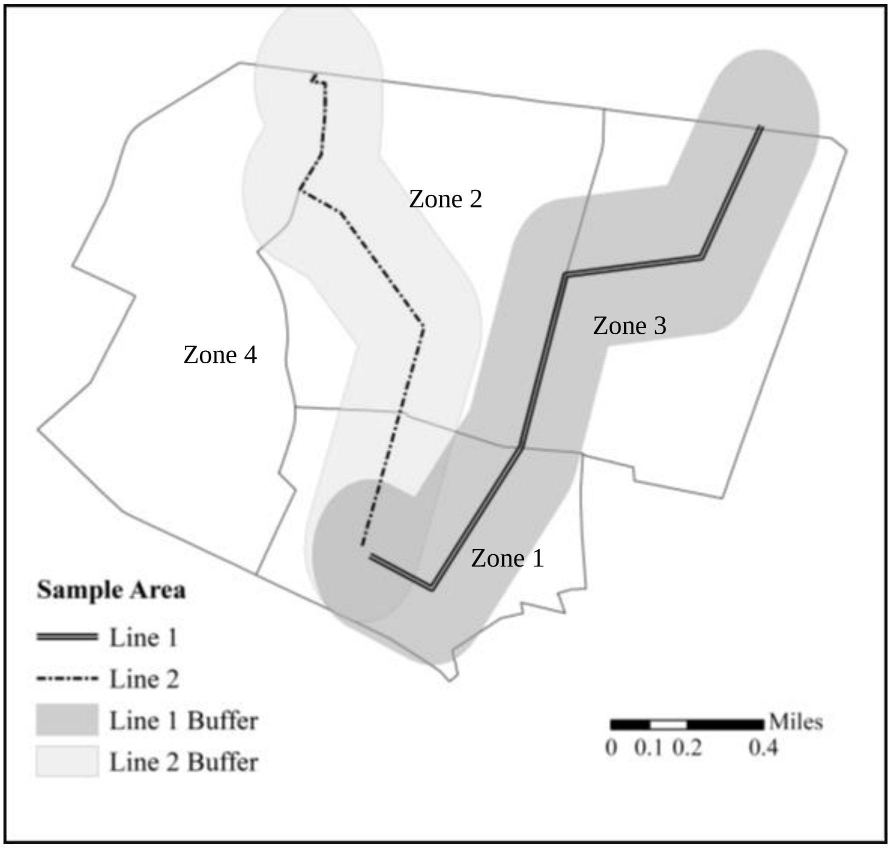

Figure 2.3. Sample zone structure with route system and buffers.

2.3.5. Sample Illustration

This section takes the reader through the steps of the process using a numerical example to illustrate the computation of the Transit Opportunity Index. Consider a transit network serving four zones with two transit lines (Line 1 & Line 2) as shown in Figure 2.3. The shaded areas represent ¼ mile buffers around each transit line.

Table 2.1 shows zonal data and results of the spatial coverage (R_i) calculations for this example. The zonal buffer area for each transit line is calculated using the ArcGIS analysis toolbox.

Table 2.1. Spatial coverage score calculation.

Zone, i	Population (P_i)	$B_{i,total}$ (mi^2)	Line 1		Line 2	
			$B_{i,l,buffer}$ (mi^2)	Spatial Coverage, R_i	$B_{i,l,buffer}$ (mi^2)	Spatial Coverage, R_i
1	**5150**	0.42	0.25	**0.595**	0.16	0.381
2	4675	0.64	0.13	0.203	0.27	0.421
3	4050	0.58	0.31	0.534	0	0
4	4965	0.63	0	0	0.1	0.158

Table 2.2 summarizes the TOI calculation steps for this sample transit network. Column 3 of Table 2.2 provides the number of vehicle operations per day (considering both weekday and weekend service) for each O-D pair. Column 4 shows per capita service frequency, S_{ijl}, using a bus capacity of 35 and column 5 shows the transit accessibility, A_{ijl} (using Eq. 3).

An example from the sample illustration may help to clarify this process. There are 42 vehicles running for each 1-3 and 2-3 origin-destination pairs for line 1 (Table 2.2). One would expect the same level of access for these two pairs using line 1. However, the spatial coverage for Zone 1 is higher than that of Zone 2, which translates into higher coverage of transit users for Zone 1 compared to Zone 2. Now, with the multiplication of the spatial coverage score by the per capita service frequency score, pair 1-3 results in higher transit accessibility score (Column 5, Table 2.2) than pair 2-3 although the same number of vehicles runs between each pair.

The travel times for each O-D pair and the connectivity decay factors obtained using Eq. 5 are shown in column 7 and 8, respectively.

For O-D pair 1-2 (see row bold, Table 2), the transit accessibility score and TOI for transit Line 1 are calculated as follows:

$$A_{12,\ Line\ 1} = R_1 S_{12} = 0.595 * \left(\frac{42 * 35}{5150}\right) = 0.170$$

$$TOI_{12,\ Line\ 1} = A_{12}\delta_{12}f_{12} = 0.170 * 1 * 0.930 = 0.158$$

Table 2.2. Transit opportunity index calculation.

1	2	3	4	5	6	7	8	9	10
				Line 1					
Origin	Destination	V_{ij}	S_{ij}	A_{ij}	δ_{ij}	T_{ij}	f_{ij}	$A_{ij}\,\delta_{ij} f_{ij}$	$\sum_j A_{ij}\,\delta_{ij} f_{ij}$
1	**2**	**42**	**0.285**	**0.170**	**1**	**24**	**0.930**	**0.158**	**0.302**
	3	42	0.285	0.170	1	30	0.848	0.144	
	4	0	0.000	0.000	0	∞	0.000	0.000	
2	1	20	0.150	0.030	1	25	0.920	0.028	0.086
	3	42	0.314	0.064	1	26	0.909	0.058	
	4	0	0.000	0.000	0	∞	0.000	0.000	
3	1	38	0.328	0.176	1	33	0.784	0.138	0.267
	2	30	0.259	0.139	1	24	0.930	0.129	
	4	0	0.000	0.000	0	∞	0.000	0.000	
4	1	0	0.000	0.000	0	∞	0.000	0.000	0.000
	2	0	0.000	0.000	0	∞	0.000	0.000	
	3	0	0.000	0.000	0	∞	0.000	0.000	
								$\sum_i \sum_j A_{ij}\,\delta_{ij} f_{ij}$	**0.655**
				Line 2					
Origin	Destination	V_{ij}	S_{ij}	A_{ij}	δ_{ij}	T_{ij}	f_{ij}	$A_{ij}\,\delta_{ij} f_{ij}$	$\sum_j A_{ij}\,\delta_{ij} f_{ij}$
1	2	50	0.340	0.129	1	22	0.946	**0.123**	**0.228**
	3	0	0.000	0.000	0	∞	0.000	0.000	
	4	50	0.340	0.129	1	32	0.808	0.105	
2	1	25	0.187	0.079	1	22	0.946	0.075	0.162
	3	0	0.000	0.000	0	∞	0.000	0.000	
	4	30	0.225	0.095	1	25	0.920	0.087	
3	1	0	0.000	0.000	0	∞	0.000	0.000	0.000
	2	0	0.000	0.000	0	∞	0.000	0.000	
	4	0	0.000	0.000	0	∞	0.000	0.000	
4	1	45	0.317	0.050	1	33	0.784	0.039	0.085
	2	45	0.317	0.050	1	26	0.909	0.046	
	3	0	0.000	0.000	0	∞	0.000	0.000	
								$\sum_i \sum_j A_{ij}\,\delta_{ij} f_{ij}$	**0.474**

Table 2.2 shows that transit accessibility from Zone 1 to Zones 2 & 3 have the same value; however, it takes 6 minutes longer to make a trip between 1 and 3 than 1 and 2. It is therefore reasonable to expect a higher level of connectivity for 1-2 than that of 1-3. Transit accessibility is multiplied by the decreasing connectivity parameter *($\delta_{ijl}f_{ijl}$)* resulting in a connectivity measure weighted by transit accessibility (Column 10, Table 2.2), the TOI.

The TOI for transit Line 2 for O-D pair 1-2 is calculated in the same way (i.e., $TOI_{12,\ Line\ 2} = 0.123$) and finally the TOI for both transit lines for this 1-2 pair is calculated below:

$$TOI_{12} = \frac{\sum_l A_{12l}\delta_{12l}f_{12l}}{\sum_i \sum_j \sum_l A_{ijl}\delta_{ijl}f_{ijl}} = \frac{0.158 + 0.123}{0.655 + 0.474} = 0.248$$

The denominator normalizes the index and provides a relative measure of transit service performance, or transit opportunity, for this pair compared to other O-D pairs. The TOI for Zone 1 as the origin is calculated as follows:

$$TOI_1 = \sum_j TOI_{1j} = \frac{\sum_j \sum_l A_{1jl}\delta_{1jl}f_{1jl}}{\sum_i \sum_j \sum_l A_{ijl}\delta_{ijl}f_{ijl}} = \frac{0.302 + 0.228}{0.655 + 0.474} = 0.469$$

Similarly, $TOI_2 = 0.220, TOI_3 = 0.236\ and\ TOI_4 = 0.076$

The results obtained for this example reflect the impact of spatial, temporal, and trip coverage for overall public transit service access and connectivity. Recall Figure 2.3, in which Zone 1 has access to both transit lines and it is connected to all other zones. Zone 2 has access to both transit lines and is connected to all zones, but this zone has poor route coverage for transit line 1 and therefore it receives a lower score than Zone 1. Zone 3 is exposed to transit line 1 only, has transit connectivity with Zones 1 and 2 and receives (as is expected) a lower score than Zones 2 and 1. Zone 4 has access to only one transit line (Line 2) and is connected only with

Zone 1 and 2. Zone 4 has very poor route coverage (only 15.8% of the population is exposed to transit service) resulting in a very low transit accessibility weight.

Table 2.3. Transit opportunity index *(Sample Illustration)*.

Origin, i	Destination, j				TOI_i
	1	2	3	4	
1	0	0.248	0.128	0.093	0.469
2	0.091	0	0.051	0.077	0.220
3	0.122	0.114	0	0	0.236
4	0.035	0.041	0	0	0.076
TOI_j	0.248	0.403	0.179	0.170	1.000

In this small sample network, the results (as shown in Table 2.3) are consistent with what would be expected. Pair 1-2 has the highest TOI score as pair 1-2 is connected by both the transit lines and it takes relatively lower travel time to make a trip which results in a higher level of connectivity (Table 2). In addition, Zone 1 has the greatest spatial coverage and has more frequent public transit service compared to other zones which also contribute the highest TOI for pair 1-2. The O-D pair 4-1 has the lowest TOI score because of a poor transit connectivity factor resulting from high travel time, the lower spatial coverage of Zone 4 and the limited service provision-from transit line 1 only. The diagonal TOI scores for pairs' 1-1, 2-2, 3-3 and 4-4 show zero transit opportunity scores, which could be misleading. For some instances, it is possible that transit service might be a viable option for a commute within a zone of larger size and the transit opportunity score for this zone should not in this case be zero. However, this research considers small spatial units (block group or smaller) and it is assumed that no transit service is going to be a serious option for trips within these small spatial units, as walking will almost always be more efficient. Furthermore, this study specifically assesses the TOI scores for O-D pairs, where trip origins and destinations are different, intrazonal demand was not part of the analysis.

O-D pairs 3-4 and 4-3 have TOI scores of zero because no direct public transit service exists between these two pairs within the network. A transfer at Zone 1 is required to make this trip possible. Identifying a transfer zone is easy for this simple public transit network consisting of only two lines with a single intersection in Zone 1. Section 2.4 will provide an approach for identifying transfer zones in more complex public transit networks served by multiple transit lines.

The method developed in this paper provides a way to quantify the TOI of O-D pairs that require a transfer between transit lines. Calculating the TOI *(*TOI_{ij}^{k}*)* for O-D pair *i-j* that is not directly connected and requires a transfer at zone *k* uses the TOI *(*TOI_{ik}*)* from origin *(i)* to transfer zone *(k)* and the TOI *(*TOI_{kj}*)* from transfer zone *(k)* to destination *(j)*. First, we calculate the average transit accessibility for the O-D pair *i-j* with transfer at *k* by dividing TOI by the corresponding decay factors (Eqn. 10). Then we multiply this by the connectivity decay factor *(*f_{ij}^{k}*)* based on total travel time *(*T_{ij}^{k}*)* from origin to destination including transfer penalty at *k*. In this case the total travel time has three parts (Eqn. 9): travel time from origin to transfer zone *(*T_{ik}*)*, a penalty time for transfer, and travel time from transfer zone to destination *(*T_{kj}*)*. A transfer penalty of 20 minutes is used in this calculation (*38*) but can vary by application and actual transfer distances.

$$T_{ij}^{k} = T_{ik} + 20 + T_{kj} \quad (9)$$

The formula for calculating the TOI (TOI_{ij}^{k}) of an O-D pair *i-j* having transfer at *k* is as follows:

$$TOI_{ij}^{k} = \frac{1}{2}\left(\frac{TOI_{ik}}{\sum_{l} f_{ikl}} + \frac{TOI_{kj}}{\sum_{l} f_{kjl}}\right) f_{ij}^{k} \quad (10)$$

where f_{ik} is the average connectivity decay factor of the *i-k* pair, f_{kj} is the average connectivity decay factor of the *k-j* pair, and f_{ij}^{k} is calculated using in Eqn. 5. A weakness of this approach is that it assumes all O-D pairs not directly connected by a transit line can be served by only one line transfer which may not be true for a comprehensive regional well-connected transit system. For O-D pair 3-4 in the sample, the total travel time and the connectivity decay factor are calculated as:

$$T_{34}^{1} = T_{31} + 20 + T_{14} = 33 + 20 + 32 = 75\ (from\ Table\ 2)$$

$$f_{34}^{1} = 0.006\ (using\ Eqn. 5)$$

Then, the TOI for O-D pair 3-4 is given by

$$TOI_{34}^{1} = \frac{1}{2}\left(\frac{TOI_{31}}{\sum_{l=1}^{2} f_{31}} + \frac{TOI_{14}}{\sum_{l=1}^{2} f_{14}}\right) f_{34}^{1} = \frac{1}{2}\left(\frac{0.122}{0.784 + 0} + \frac{0.093}{0 + 0.808}\right)(0.006) = 0.00035$$

The TOI results for O-D pairs requiring a transfer are noteworthy. The poor TOI value (0.00035) for pair 3-4 reflects the inconvenience of transfers and also the decay of connectivity associated with higher travel time. The consistency of transit performance measure (TOI) is further tested in the next section using a large-scale, real-world application from New Haven, Connecticut.

2.4. Case Study Application

In this section the Transit Opportunity Index is applied to the city of New Haven, Connecticut. The city of New Haven (Figure 2.4) had a 2000 population[3] of 123,626 and a land area of 19.22 square miles. The 2000 Census recorded 49,358 workers in the study area, giving an overall employment density of 4.01 workers per acre. Public transit bus service in the city of New Haven

[3] This research used US Census 2000 data for analysis. Census Tract level socioeconomic and demographic data for this study area was not yet available for 2010.

is provided by Connecticut Transit (CTTransit). CTTransit operates a fleet of 35 passenger vehicles on 19 fixed routes serving 981 transit stops (Figure 4) in this study area. The regular routes provide service every 15 to 20 minutes from 5 am to 12 am. Service information (i.e., service frequency, hours of operation, etc.) was obtained from the CTTransit website.

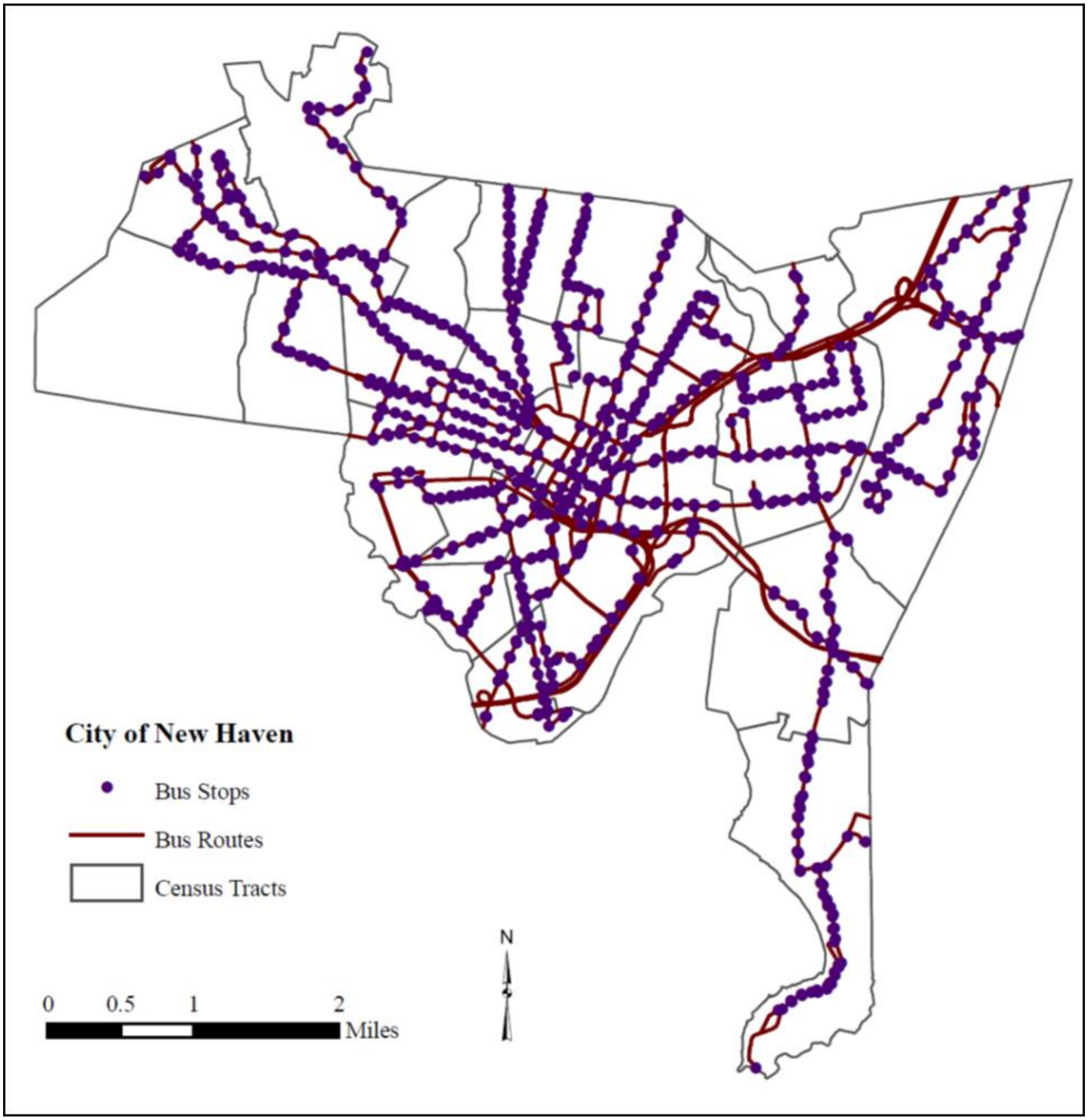

Figure 2.4. New Haven bus network.

This analysis was conducted for the 29 census tracts of New Haven. Table 2.4 shows TOI results for each O-D pair in the public transit network connected by direct routes. For ease of interpretation and comparison, TOI scores are rescaled such that all O-D TOI scores sum to 1000. This 29x29 TOI matrix provides an estimate of overall public transit supply provided for this area considering only direct public transit service. This TOI represents the degree to which travelers can make trips between origins and destinations using existing direct public transit service. The O-D pairs with TOIs of zero in Table 2.4 are identified as pairs that require a transfer to complete a trip. This leads to the important question: *which tract represents the best transfer zone for each O-D pair requiring transfer?* The TOI matrix allows us to identify the transfer tract that ensures the maximum aggregated TOIs from origin to transfer tract to destination tracts. At this point, this method does not account for transfer distances (i.e., stops separated by large walking distance) and therefore provides a best transfer point in terms of connectivity and accessibility. This shortcoming will be addressed in later evolutions of this methodology.

As an example, a trip from census tract 1412 to census tract 1418 is not feasible by direct public transit service, receiving a TOI of zero, and therefore requires a transfer between transit lines. To identify the best transfer tract for making a trip from 1412 to 1418, the TOIs for row 1412 are added to corresponding TOIs for column 1418 (as shown in Table 2.4 below). A symmetric TOI matrix is required for this method. Both the corresponding row and column values need to be nonzero for a tract to be a potential transfer zone. For example, the row value for the 1412-1403 pair and column value for the 1403-1418 are not feasible transfer zones as there is no direct public transit service from 1412 to 1403. This approach finds that tract 1401 is the best transfer tract for completing a trip from tract 1412 to tract 1418 with the maximum sum

of TOIs (i.e., 1.72 + 3.63 = 5.35). In this way public transit planners can identify trip pairs which are not directly connected by the existing public transit service and determine the optimal transfer tract for these O-D pairs.

Restated, the best transfer zone must satisfy the condition:

$$k = \{l: TOI_{il} + TOI_{lj} \geq TOI_{im} + TOI_{mj} \ \ \forall l, m\}$$

Where *k* is defined as the transfer zone.

This research also proposed an approach (as discussed in the sample illustration section) to calculate the TOI for the O-D pairs requiring transfer. The TOI for O-D pair 1412-1418 with transfer at tract 1401 is given by

$$TOI_{1412-1418}^{1401} = \frac{1}{2} * \left(\frac{1.72}{0.696} + \frac{3.63}{0.848}\right) * 0.001 = 0.0034$$

The pairs 1412-1401 and 1401-1418 are served directly by more than one transit line. In this case this approach estimated the sum of decay factors for pairs 1412-1401 and 1401-1418 and used those values (0.696 and 0.848) to calculate the TOI. This example illustrates how the TOI matrix can be used to identify O-D pairs that require line transfers and to determine the best transfer tracts for those O-D pairs based on TOI scores. Calculating TOI scores for all O-D pairs allows public transit providers/planners to examine the effects of service changes (e.g., changes in vehicle capacity and service frequency, new route alignment, etc.) and their estimated impact on public transit service performance.

Table 2.4. Transit opportunity index *(New Haven, CT)*.

	Destination																													
Origin	**1401**	**1402**	**1403**	**1404**	**1405**	**1406**	**1407**	**1408**	**1409**	**1410**	**1411**	**1412**	**1413**	**1414**	**1415**	**1416**	**1417**	**1418**	**1419**	**1420**	**1421**	**1422**	**1423**	**1424**	**1425**	**14261**	**14262**	**1427**	**1428**	***TOI_i***
1401	0	12.29	16.59	5.83	11.29	9.76	25.95	9.88	8.29	5.79	1.81	7.01	7.82	6.32	8.21	15.14	47.35	3.63	6.94	10.15	10.18	14.94	8.91	11.54	10.65	7.36	7.79	3.22	0.81	295.59
1402	9.92	0	3.97	4.35	1.50	0.24	3.83	0.15	0	0	0	2.12	2.78	3.36	0	3.55	6.56	1.11	1.34	1.55	0	4.22	0	0.22	0.19	0	0	0.52	0	51.57
1403	12.12	6.35	0	2.03	8.52	7.03	1.26	5.62	0	0	0	0	0	0	0	0	9.30	1.91	3.48	3.78	0	0	0	1.37	1.09	0	0	0	0	63.93
1404	2.68	2.84	1.71	0	1.72	0	1.10	0	0	0	0	0.45	0.68	0.93	0	1.03	1.12	0	0	0	0	1.10	0	0	0	0	0	0	0	15.40
1405	4.28	1.58	4.59	0.29	0	4.48	0.87	3.80	0	0	0	0	0	0	0	0	4.05	0	0.83	1.09	0	0	0	0.61	0.41	0	0	0	0	26.92
1406	2.96	0.31	3.07	0	2.92	0	0.97	2.77	0	0	0	0	0	0	0	0	2.98	0	0.25	0.28	0	0	0	0.20	0.15	0	0	0	0	16.93
1407	5.08	0.37	0.20	0.35	0.11	0.13	0	0.94	3.63	2.83	0.76	2.96	3.10	2.32	0.75	3.19	5.11	0	0.79	0.86	0.81	0.59	0.18	0.68	0.55	0.32	0.17	0.12	0.07	37.11
1408	2.36	0.11	1.76	0	1.94	1.86	2.11	0	0.74	0	0	0	0	0	0	0	2.41	0	0.05	0.08	0	0	0	0.03	0.02	0	0	0	0	13.53
1409	3.85	0	0	0	0	0	3.72	0.60	0	3.56	0.90	3.47	3.52	2.71	0	2.68	3.97	0	0.64	0.70	0.67	0	0	0.50	0.36	0.17	0	0	0	32.08
1410	1.21	0	0	0	0	0	1.36	0	1.39	0	0.91	1.42	1.45	0.51	0	0.50	1.30	0	0.52	0.59	0.56	0	0	0.38	0.26	0.11	0	0	0	12.56
1411	0.14	0	0	0	0	0	0.19	0	0.21	0.22	0	0.23	0.23	0	0	0	0.17	0	0.09	0.11	0.10	0	0	0.06	0.04	0.01	0	0	0	1.85
1412	1.72	0.32	0	0.21	0	0	1.97	0	1.69	1.73	0.83	0	2.31	1.42	0	1.36	1.85	0	0.31	0.37	0.34	0.32	0	0.21	0.13	0.05	0	0	0	17.22
1413	1.92	0.21	0	0.16	0	0	2.02	0	1.81	1.83	0.03	0.86	0	2.04	0	2.00	1.97	0	0.01	0.02	0.02	0.21	0	0.01	0	0	0	0	0	15.20
1414	1.58	0.49	0	0.42	0	0	1.62	0	1.11	1.07	0	1.45	1.47	0	0	1.64	1.61	0	0	0	0	0.49	0	0	0	0	0	0	0	12.98
1415	1.55	0	0	0	0	0	1.43	0	0	0	0	0	0	0	0	1.63	1.59	0.18	0	0.17	0	0.19	0.13	0	0	0	0.12	0.07	0.03	7.15
1416	3.86	0.65	0	0.58	0	0	3.53	0	0.88	0.87	0	1.40	1.49	1.56	2.37	0	3.93	0.41	0	0.41	0	1.11	0.36	0	0	0	0.35	0.23	0.13	24.21
1417	8.76	2.18	2.12	0.64	1.15	1.14	4.79	1.37	2.13	1.61	0.39	1.86	2.01	1.68	2.12	3.86	0	1.37	1.59	2.59	1.47	2.12	1.23	1.71	1.65	1.06	1.05	0.43	0.13	54.35
1418	1.04	0.46	0.47	0	0	0	0	0	0	0	0	0	0	0	0.62	0.61	1.08	0	0.52	1.10	0	0	0	0	0	0	0	0	0	5.94
1419	1.33	1.05	1.02	0	0.41	0.41	0.19	0.15	0.15	0.12	0.09	0.08	0.11	0	0	0	1.37	0.47	0	1.40	0.23	0	0	0.90	0.86	0.18	0	0	0	10.59
1420	4.36	2.58	2.81	0	1.14	1.14	1.02	0.24	0.84	0.73	0.56	0.51	0.64	0	0.27	0.29	4.71	1.29	3.68	0	1.21	0	0	2.53	2.32	0.85	0	0	0	33.82
1421	9.29	0	0	0	0	0	2.72	0	2.20	1.87	1.40	1.29	1.63	0	0	0	9.41	0	3.12	3.12	0	6.51	6.34	9.40	9.00	7.65	5.63	0	0	80.67
1422	12.25	4.57	0	0.98	0	0	3.50	0	0	0	0	0.60	0.79	0.96	1.89	3.37	11.71	0	0	0	5.27	0	7.86	5.19	5.03	4.44	7.31	5.38	1.80	82.98
1423	2.86	0	0	0	0	0	0.56	0	0	0	0	0	0	0	0.36	0.52	2.76	0	0	0	2.29	3.01	0	2.34	2.29	1.42	2.24	0.73	0.64	22.08
1424	2.54	0.33	0.30	0	0.17	0.17	0.55	0.05	0.39	0.31	0.21	0.19	0.25	0	0	0	2.57	0	1.15	1.11	2.27	1.55	1.56	0	2.71	2.10	1.44	0	0	22.02
1425	1.19	0.18	0.16	0	0.08	0.08	0.44	0.04	0.27	0.20	0.13	0.11	0.16	0	0	0	1.24	0	0.94	0.88	1.08	0.42	0.43	1.42	0	0.65	0.29	0	0	10.47
14261	1.43	0	0	0	0	0	0.10	0	0.05	0.03	0.02	0.02	0.03	0	0	0	1.38	0	0.24	0.22	1.69	1.49	1.61	1.87	1.98	0	1.77	0	0	14.00
14262	1.31	0	0	0	0	0	0.02	0	0	0	0	0	0	0	0.01	0.02	1.30	0	0	0	1.45	1.51	1.60	1.55	1.62	1.66	0	0.03	0.03	12.14
1427	0.57	0.40	0	0	0	0	0.17	0	0	0	0	0	0	0	0.08	0.14	0.54	0	0	0	0	0.65	0.30	0	0	0	0.31	0	0.36	3.55
1428	0.26	0	0	0	0	0	0.19	0	0	0	0	0	0	0	0.07	0.16	0.22	0	0	0	0	0.40	0.52	0	0	0	0.54	0.61	0	3.01
TOI_j	102.5	37.35	38.84	15.88	31.00	26.49	66.31	25.67	25.86	22.84	8.10	26.11	30.55	23.86	16.81	41.74	133.6	10.40	26.59	30.68	29.72	40.93	31.11	42.83	41.47	28.09	29.06	11.37	4.03	1000

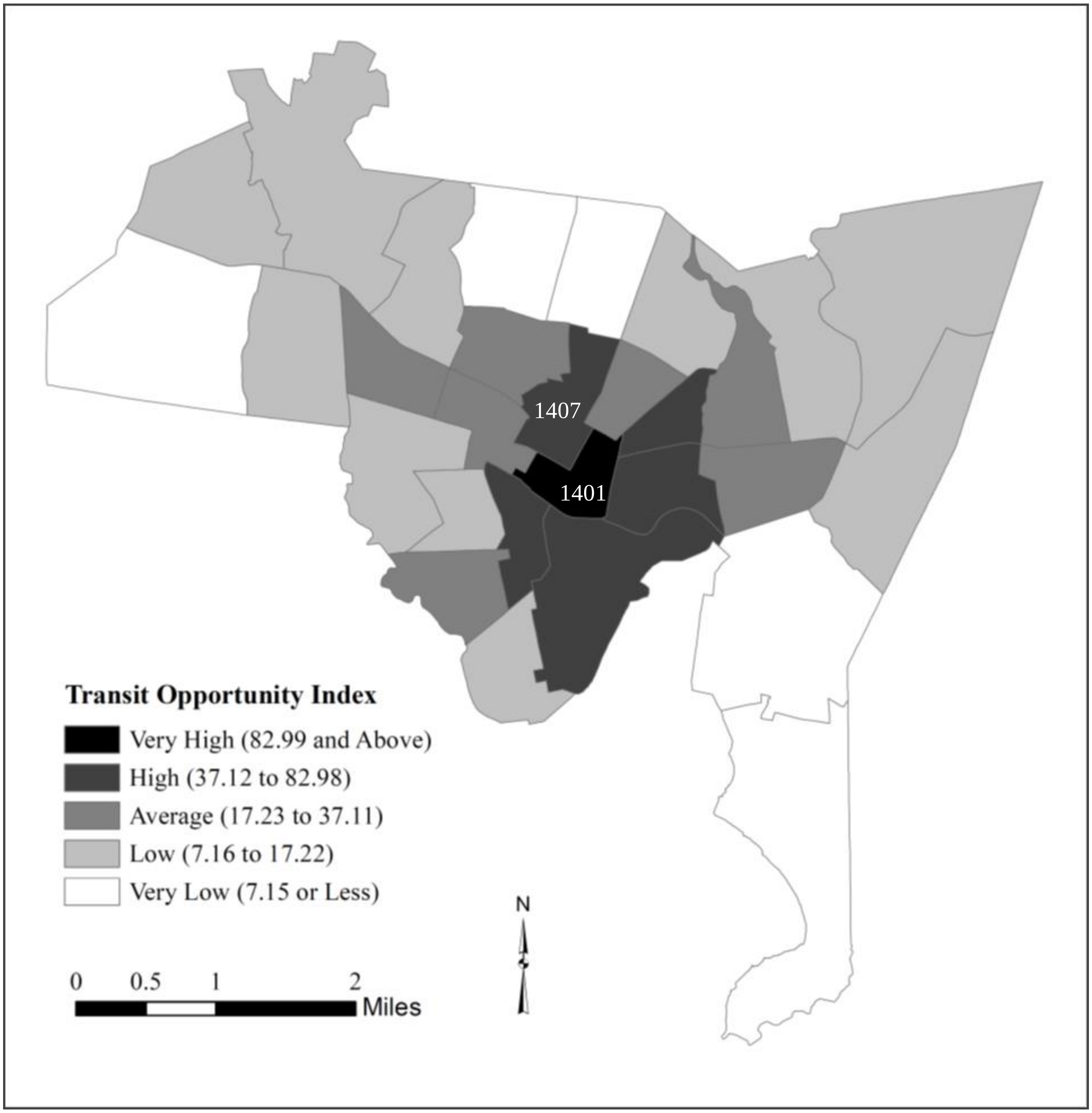

Figure 2.5. Tract TOI scores of New Haven, Connecticut.

Figure 2.5 shows the TOI obtained for each origin TOI_i (the last column in Table 2.4). The TOIs in this map are grouped according to the natural break (Jenk's optimization) classifications generated using GIS mapping functions (*39*). The areas with very high access to

public transit service and high transit connectivity to other areas are shown in the darkest shading, with lighter shades representing lower transit accessibility and transit connectivity. As shown in Figure 2.5, TOIs are highest in the central portion of the transit network. The central tracts (e.g., Tracts 1401, 1417, etc.) have high TOIs due to frequent bus service to and from the tracts (high temporal coverage). Tracts 1401 and 1417 are served by most New Haven bus routes and therefore have high route coverage. Thus, the TOI provides a strong indication of true transit service accessibility and connectivity consistent with the expectations for this transit network.

2.5. Conclusions and Future Research

The Transit Opportunity Index provides a robust tool for measuring public transit service performance. It builds on and expands the scope of existing public transit connectivity and accessibility measures. This tool provides the ability to explore the public transit opportunity between a single O-D pair, as well as the overall public transit opportunity from or to a specific origin or destination. Transit accessibility, a commonly used concept for measuring public transit service performance, measures ease of access to public transit service for an area (spatial coverage) and service frequency (supply aspect of temporal coverage). There is less literature considering trip coverage, or the transit connectivity of origins and destinations, where destinations are the aggregation of the multiple opportunities within a defined geographic zone. An aggregated destination concept might be of interest to travelers who want to satisfy most or all of their needs at one location. The TOI developed in this paper allows us to quantify the level of access provided by the service and the ability of this service to make trips from origin to destination. Another application of TOI is transit service monitoring for transit providers as required by the Title VI of the Civil Rights Act of 1964. Title VI seeks to identify and eliminate

discrimination in existing transit services or benefits based on race, color, income or national origin. According to Title VI, transit providers must evaluate future service changes or improvement options at the planning stages to determine whether the service changes have a discriminatory impact. TOI can be used to assess the social implications of transportation, specifically levels of access and opportunities experienced by low-income, minority, and transportation disadvantaged O-D pairs. Assessment of service performance for a single transit line can be performed for routes/services in minority/low income communities and this performance result can be compared to the established service policies and standards for identifying possible discrepancies. TOI is based on O-D connectivity in addition to spatial and temporal access, which may help to better assess the equity of access to activities via transit service.

There are several assumptions implicit in the TOI. The first assumption is the ¼ mile Euclidean buffer used along the public transit lines to estimate spatial coverage. We chose to use line buffers over stop buffers to avoid overlapping coverage areas between closely spaced stops. However, this method will overestimate spatial coverage along lines where the stops are spaced further apart. Another potential cause for overestimation is that this method neglects the actual pedestrian road network surrounding the public transit service. Future work will refine the method to eliminate these sources of overestimation.

The second assumption made is that the level of transit connectivity decreases in accordance with a logistic decay function (inverted s-curve). While in line with empirical studies and consistent with expectations, other functional forms may be appropriate for different modes, stop types and geographic locations. New functional forms and the flexibility to incorporate them in this methodology are subjects for future work.

A limitation of the TOI is that a maximum of one transfer is available to complete a trip between O-D pairs not directly connected, which may not be true for a comprehensive regional well-connected transit system. Expanding TOI to multiple transfers will improve generalizability and applications to dense, multi-modal (i.e. bus, train, and circulator) systems.

Another potential improvement that could be made to the model is to add a measure of pedestrian route connectivity to address the walking accessibility to transit service and its contribution to accessibility decay, at both the trip ends and at transfer points. The demand side of the temporal coverage could be incorporated in measuring transit accessibility, which would provide the relative value of service availability in each time period of the day. Lastly, Since the TOI provides a relative value of transit service supply available for each O-D pair it will be an important extension to incorporate a measure of demand into the index to identify service gaps and target investments towards underserved areas.

Acknowledgments

The authors wish to thank the Connecticut Cooperative Highway Research Program (Project 08-5) for providing funding for this research effort. This document, prepared in cooperation with the Connecticut Department of Transportation and the Federal Highway Administration, does not constitute a standard, specification, or regulation. The contents of this document reflect the views of the authors who are responsible for the ideas, facts and the accuracy of the data presented herein. The contents do not necessarily reflect the views of the Connecticut Department of Transportation or the Federal Highway Administration. We thank the reviewers for their helpful comments in improving this paper.

References

1. Mamun, S., and Lownes, N. A Composite Index of Public Transit Accessibility. *Journal of Public Transportation*, Vol. 14, No. 2, 2011, pp. 69-87.
2. Murray, A.T. Strategic Analysis of Transport Coverage. *Socio-Economic Planning Sciences*, Vol. 35, No. 3, 2001, pp. 175–188.
3. Beimborn, E., Greenwald, M., and Jin, X. Impacts of Transit Accessibility and Connectivity on Transit Choice and Captivity. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1835, Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 1–9.
4. Biba, S., Curtin, K., and Manca, G. A new method for determining the population with walking access to transit. *International Journal of Geographical Information Science*, Vol. 24, No. 3, 2010, pp. 347-364.
5. Currie, G. Quantifying spatial gaps in public transport supply based on social needs. *Journal of Transport Geography*, Vol. 18, 2010, pp. 31–41.
6. Delbosc, A., and Currie, G. Using Lorenz curves to assess public transport equity. *Journal of Transport Geography*, Vol. 19, 2011, pp. 1252–1259.
7. Hsiao, S., Lu, J., Sterling, J., and Weatherford, M. Use of geographic information system for analysis of transit pedestrian access. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1604, Transportation Research Board of the National Academies, Washington, D.C., 1997, pp. 50-59.
8. Kimpel, T., Dueker, K., and El-Geneidy, A. Using GIS to measure the effect of overlapping service areas on passenger boarding at bus stops. *Urban and Regional Information Systems Association Journal*, Vol. 19, No. 1, 2007, pp. 5-11.
9. Zhao, F., Chow, L., Li, M., Ubaka, I., and Gan, A. Forecasting Transit Walk Accessibility: Regression Model Alternative to the Buffer Method. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1835, Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 34– 41.
10. Transportation Research Board, Transit Capacity and Quality of Service Manual, 2nd Edition, Washington, DC, 2003, *Transportation Research Board, Transit Cooperative Research Program.*

11. Jiang, Y., Zegras, P., and Mehndiratta, S. Walk the line: station context, corridor type and bus rapid transit walk access in Jinan, China. *Journal of Transport Geography*, Vol. 20, 2012, pp. 1-14.
12. Grava, S. *Urban Transportation Systems: Choices for Communities*. McGraw-Hill, United States, 2003.
13. Peng, A., Dueker, K., Strathman, J., and Hopper, J. A simultaneous route level transit patronage model: demand, supply and inter-route relationship. In *Transportation*, Vol. 24, 1997, pp. 159–181.
14. Kuby, M., Barranda, A., and Upchurch, C. Factors influencing light-rail station boardings in the United States. *Transportation Research A*, Vol. 38, No. 3, 2004, pp. 223–247.
15. O'Neill, W., Ramsey, R., and Chou, J. Analysis of Transit Service Areas Using Geographic Information Systems. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1364, Transportation Research Board of the National Academies, Washington, D.C., 1992, pp. 131–138.
16. Ryus, P., Ausman, J., Teaf, D., Cooper, M., and Knoblauch, M. Development of Florida's transit level-of-service indicator. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1713, Transportation Research Board of the National Academies, Washington, D.C., 2000, pp. 123–129.
17. Polzin, S., Pendyala, R., and Navari, S. Development of time-of-day-based transit accessibility analysis tool. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1799, Transportation Research Board of the National Academies, Washington, D.C., 2002, pp. 35–41.
18. Rood, T. *The Local Index of Transit Availability: An Implementation Manual*, Local Government Commission, Sacramento, California, 1998.
19. Schoon, J., McDonald, M., and Lee, A. Accessibility Indices: Pilot Study and Potential Use in Strategic Planning. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1685, Transportation Research Board of the National Academies, Washington, D.C., 1999, pp. 29–38.
20. Mamun, S., and Lownes, N. Measuring Service Gaps: An Accessibility-based Transit Need Index. In *Transportation Research Record: Journal of the Transportation Research*

Board, No. 2217, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 153-161.

21. Chen, Ravulaparthy, Deutsch, Dalal, Yoon, Lei, Goulias, Pendyala, Bhat, and Hu. Development of Indicators of Opportunity-Based Accessibility. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2255, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 58-68.
22. Huang, R., and Wei, Y. Analyzing neighborhood accessibility via transit in a GIS environment. *Geographic Information Sciences*, Vol. 8, No. 1, 2002, pp. 39–47.
23. Mavoa, S., Witten, K., McCreanor, T., and O'Sullivan, D. GIS based destination accessibility via public transit and walking in Auckland, New Zealand. *Journal of Transport Geography*, Vol. 20, 2012, pp. 15-22.
24. Ceder, A., Y. Le Net, and C. Coriat, Measuring Public Transport Connectivity Performance Applied in Auckland, New Zealand. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2111, Transportation Research Board of the National Academies, Washington, D.C., 2009, pp. 139–147.
25. Lam, T., and Schuler, H. Connectivity Index for Systemwide Transit Route and Schedule Performance. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 854, Transportation Research Board of the National Academies, Washington, D.C., 1982, pp. 17–23.
26. Lee, K., and Lee, H. A New Algorithm for Graph-Theoretic Nodal Accessibility. *Measurement. Geographical Analysis*, Vol. 30, No. 1, 1998, pp.1–14.
27. Rodrigue, J. 2003. *Graph theory: measures and indices*. <http://people.hofstra.edu/geotrans/eng/ch2en/meth2en/ch2m2en.html>. Accessed July 5, 2011.
28. Scott, D., Novak, D., Aultman-Hall, L., and Guo, F. Network Robustness Index: A new method for identifying critical links and evaluating the performance of transportation networks. *Journal of Transport Geography*, Vol. 14, 2006, pp. 215–227.
29. Fu, L., and Xin, Y. A new performance index for evaluating transit quality of service. *Journal of Public Transportation*, Vol. 10, No. 3, 2007, pp. 47-69.

30. Derrible, S. and Kennedy, C. A Network Analysis *of* Subway Systems *in the* World using Updated Graph Theory. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2112, Transportation Research Board of the National Academies, Washington, D.C., 2009, pp. 17–25.

31. Hadas, Y., and Ceder, A. Public Transit Network Connectivity. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2143, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 1–8.

32. Hadas, Y., Ceder, A. and Ranjitkar, P. Modeling Public-Transit Connectivity with Quality-of-Transfer Measurements, Proceedings of *90th Annual Meeting of Transportation Research Board*, Washington DC, January 23-27, 2011 (Pre-print CD).

33. Gutiérrez, J., Cardozo, O., and García-Palomares, J. Transit ridership forecasting at station level: an approach based on distance-decay weighted regression. *Journal of Transport Geography*, Vol. 19, 2011, pp. 1081-1092.

34. Wibowo, S., and Olszewski, P. Modeling walking accessibility to public transport terminals: Case study of Singapore mass rapid transit, *Journal of the Eastern Asia Society for Transportation Studies*, Vol. 6, 2005, pp. 147-156.

35. Chang, S., and Hsu, C. Modeling Passenger Waiting Time for Intermodal Transit Stations. In *Transportation Research Record: Journal of the Transportation Research Board*, No 1753, Transportation Research Board of the National Academies, Washington, D.C., 2001, pp. 69-75.

36. Lam, W., and Morrall, J. Bus Passenger Walking Distances and Waiting Times: A Summer–Winter Comparison. *Transportation Quarterly,* Vol. 36, No. 3, 1982, pp. 407–421.

37. 2009 American Community Survey (Census). *http://www.census.gov/acs/www/data_documentation/2009_release/.* Accessed June 13, 2011.

38. Currie, G. Gap analysis of public transport needs: measuring spatial distribution of public transport needs and identifying gaps in the quality of public transport provision. In *Transportation Research Record: Journal of the Transportation Research Board*, No

1895, Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 137-146.

39. Murray, A., and Shyy, T. Integrating attribute and space characteristics in choropleth display and spatial data mining. *International Journal of Geographical Information Science,* Vol. 14, No. 7, 2000, pp. 649-667.

This paper is under review for publication in **Transportation Research Record: Journal of the Transportation Research Board.*

CHAPTER 3: ACCESS AND CONNECTIVITY TRADEOFFS IN TRANSIT STOP LOCATION

Abstract

This paper and the proposed formulation address the tradeoff between improving transit access and trip connectivity. Transit access is typically regarded as the physical proximity to transit service. Additional stops can provide greater access to the service and reduce walking distance to stops. However, closer stops do not guarantee that routes serving the stops are well connected to desired trip destinations. Transit riders may walk longer distances to access to a stop which has frequent transit services, requires less wait time and is well-connected to their desired destinations. Frequent stops on the way to these desired destinations, while providing shorter walking distance for some passengers, increase dwell time which results in a smaller portion of the network being connected within a certain travel time, which is an important element of the connectivity of the system. The proposed methodology considers both the impact of access distance to transit stops and trip connections to destinations for determining optimal stop locations and setting optimal transit line frequencies. The bus stop location model is formulated as a mixed integer program (MIP) and CBC (COIN-OR Branch and Cut) solver is used to solve the problem. Sensitivity analyses are performed and computational results are presented for an illustrative example. The formulation is applied to the bus transit network in New Haven, CT as a case study to illustrate the usefulness of the model.

3.1. Introduction

The purpose of this paper is to present a multi-objective optimization formulation for selecting an optimal set of transit stops based on tradeoffs between transit user access and system connectivity. The problem is formulated to select stops from the existing set of stops serving any number of transit lines. It is assumed that the lines remain fixed – stops either continue to be maintained or not. The assignment of demand to stops is also provided through this formulation, assuming that passengers know and utilize the nearest stop with which to make their desired trip. Access cost is defined as the combination of access/walking cost to transit stops and the unserved demand cost, whereas the benefit of system connectivity is measured as the savings in in-vehicle stopping time and wait time at stops. Three aspects of the problem in transit service planning are considered: optimal transit stop selection, allocation of demand to each stop, and assigning vehicle frequencies to transit lines given a fixed number of available vehicles.

3.1.1. Problem Motivation

The location and density of public transit stops are important elements of the service planning process for evaluating the system's level of accessibility. Accessibility to transit service significantly influences the service performance and efficiency, as the more people that reside and/or are employed in close proximity of transit, the greater the likelihood the service will be utilized (*1, 2*). In a transit system, adding a new transit stop onto a line has the potential to increase ridership by providing better access through shorter walking distances. However, adding stops can have the opposite effect if their addition increases travel times too greatly through the increased deceleration/dwell/acceleration cycles at the stops (*3*). More stops also result smaller geographic area that is reachable from a given trip origin under a given travel time budget (*4*).

A fast and efficient transit system is a prerequisite for augmenting transit ridership. The fewer the number of stops along a line, the faster one can travel to destinations with less dwell time at stops. From the operator's perspective, bus running times will decline with fewer stops and thus reduce operating costs that can, in turn, be translated into larger fleet sizes of service for a given operating budget. There is a fundamental tradeoff associated with the stop location problem between improving accessibility to transit service with more stops and increasing the efficiency of the transit system with fewer stops.

Furthermore, people will only use transit stops to access service if the stop is connected to their desired destinations, making connectivity another important aspect in identifying optimal stop locations (*5*). A transit stop is more likely to be used if it is served by multiple lines and the lines are connected to multiple destinations (*6*). In this case, transit riders are willing to walk further to access fast, frequent and directly served transit lines to their destinations as opposed to shorter walks to service requiring transfers. Transit line layouts, optimal location of transit stops, and the available number of vehicles are necessary to set service frequencies, which is another important transit network design activity for improving service efficiency (*7*).

This paper formulates a mathematical model for addressing both the accessibility and trip connectivity simultaneously using three variables. First, stop locations along a line are selected to balance access cost with additional in-vehicle stopping time associated with more frequent stops. Second, service frequencies of existing lines are set to maximize connectivity. Lastly, allocation of demand to stops is provided as a function of the new stop configuration and service frequency.

Section 3.2 reviews the transit stop location models and their optimization criteria. Section 3.3 presents the problem formulation and solution method for the transit stop location

model. A numerical example is then used to illustrate the problem formulations in section 3.4. Section 3.5 describes the application of the formulation to a study area. Conclusions noting potential future research and applications to current planning practices are presented in section 3.6.

3.2. Literature Review

In transit service planning, operational improvement issues have been receiving a high level of attention for competitive transportation funding and budgetary constraints. Transit planners and decision makers consider different operational issues for increasing service utilization (*8*). In general, transit planners consider consolidation of transit stops and changing route layouts in the physical improvement phase while in the operational improvement phase planners adjust service frequencies or vehicle/driver rescheduling. As most transit systems have relatively stable route structures and a politically-determined level of financial support to build new routes, the consolidation of transit stops has become a more desirable objective in the physical improvement phase rather than changing route layouts. Moreover, due to budgetary constraints/cuts in service operation, service providers are open to minimizing operational costs through fewer stop locations so that they can utilize savings to increasing fleet size. This paper aims to increase the service efficiency for a transit network where the transit routes have been defined a priori. Assuming that the route layout is defined, two important design problems will be considered in the proposed formulation: (a) locating the optimal number of transit stops (physical design problem) and (b) improving transit line frequencies (operational design problem).

The review of past research on these topics can be classified into three categories based on the optimization criteria: (a) studies that focused on locating bus stops based on user costs; (b)

studies focused on user costs and provider's operation costs; (c) studies focused on tradeoffs associated with the stop location problem, the focus of this paper.

Commonly, passenger's access and egress time required to travel to and from stops is emphasized in literature related to bus stop planning (*9*). Furth and Rahbee (*10*) focused on passenger's preference for minimizing access/egress times to/from stops, where it was assumed that the more frequent the stops, the shorter the walking time to access the service. This assumption is supported by the literature, which suggests a 5 min or 400 m (1/4 mile) walking distance is a reasonable access standard for bus stop transit in urban areas (*11-16*).

Some stop location studies focus on operating cost in addition to the walking/access distance to stops for determining optimal stop locations on bus routes. Ghoneim and Wirasinghe (*17*) proposed an optimization model for the total system costs in order to optimize the distance between the stops as well as the bus frequency and the length of route. Furth and Rahbee (*10*) employed a dynamic programming approach to determine the optimal number and location of bus stops for a heavily used route in the Massachusetts Bay Transportation Authority system. Saka (*18*) built a model for determining optimum bus stop spacing considering operation cost explicitly. El-Geneidy et al. (*9*) proposed a model aiming to reduce operating costs while maintaining service frequency for optimizing bus stop locations. They considered both the access and riding cost for minimizing the total costs. Li and Bertini (*19*) proposed a bus stop spacing model for minimizing transit operation cost without impact on transit accessibility. El-Geneidy and Surprenant-Legault (*20*) found that a limited-stop bus service can yield up to 4.6 minutes savings in running time compared to existing regular service.

Several models consider tradeoffs between different aspects of accessibility. Murray and Wu (*21*) look at tradeoffs between two important components of spatial accessibility: proximity

to stops (access through more stops) and geographic coverage (service efficiency through fewer stops). They developed a distance constrained p-median problem by minimizing the demand weighted access distance and used the constraint on number of stops to maintain. They considered the geographic coverage as the measure of service efficiency. Wu and Murray (*22*) proposed another model which allowed for a tradeoff between access (more stops) and service quality (lower travel time with fewer stops) in the selection of stops in an established multiple–route transit system. Total system travel time was utilized to reflect the public transit service quality/efficiency in this multi-objective model. The model results suggested that improved transit service quality is possible by eliminating redundant or underutilized service stops. Recently, Delmelle et al. (*4*) proposed a tradeoff model for identifying bus stop redundancy. They used a non-linear spatial interaction coverage (SIC) model for maximizing covered demand weighted by the physical attraction of the transit facilities demand. They defined connectivity as the number of destinations can be reached from a stop and destinations are considered as the different facilities that transit users are attracted.

All existing tradeoff approaches for the stop location problem examine service efficiency with different metrics. Some efficiency measures considered travel time, some used service coverage, or some considered facility connectivity as the estimator of service efficiency. In this paper, service efficiency is measured using a composite public transit performance measure named transit opportunity index (TOI) which integrates transit accessibility (spatial and temporal coverage), topological network connectivity, and travel time. This TOI measure accounts for O-D pair-wise transit connectivity with binary connectivity and travel-time based connectivity decay. Mamun et al. (*5*) provides more detail on the TOI its components. Earlier studies have constrained the number of stops to maintain to a predefined number – a restriction which we

relax. Instead, stop capacity is applied to determine the number of stops needed to serve travel demand. An additional contribution of this work is directional stop location and directional demand allocation to the selected stops.

3.3. BSLP Formulation

A multi-objective bus stop location problem (BSLP) formulation is considered in this paper that will minimize access cost, unserved demand cost, in-vehicle stopping cost, and maximize connectivity benefit. In maximizing transit connectivity, it would select the most connected and frequently serviced stops so that the in-vehicle stopping time will decrease for each transit user and they can travel to more destinations within their limited travel time budget. This proposed formulation will consider stop capacity, line capacity, demand coverage, proportional fleet distribution and fleet size budget constraints to obtain the optimal stops and optimal line frequencies, which is relatively scant in the literatures.

A formalization of the BSLP with associated descriptions of the formulation components is provided. This formulation includes the necessary sets and indices, parameters and data, assumptions, objective function, and the constraint set.

3.3.1. Sets and Indices

$i \in I$ existing stop locations,

$g \in G$ origin demand centroid locations,

$h \in H$ destination demand centroid locations, and

$l \in L$ set of transit lines.

The set I contains all existing stop locations for the bus network. The set G and H include the demand zone centroids for origin and destination respectively. The demand centroids are operationalized using population centroids at the block group level. The set L defines the set of bus lines for which the stop locations will be optimized and the available fleet size assigned.

3.3.2. Assumptions

The following assumptions are made:

- Transit lines have been defined a priori;
- People will tend to use stops that serve more lines and have connection to more destinations;
- Demand is located at population centroids;
- Demand represents the bus trips for home based work (HBW) and home based other (HBO) trips only for O-D pairs and is fixed for the morning peak period;

3.3.3. Given Data

C_a cost of access ($20/hour)

C_t cost of in-vehicle travel time ($30/hour)

b_c benefit of connectivity ($1/hour)

C_u cost of unserved demand ($10 per unserved passenger/hr)

d_{ig} rectilinear distance from stop node i to demand centroid g (feet)

v_{walk} pedestrian walking speed (14400 ft/hour)

U capacity of a vehicle (35 seats)

p_g population at demand centroid g (2010 US Census Bureau Data)

D Delay in in-vehicle travel time per stops (30 seconds)

The default values of access and in-vehicle time are pulled from TCRP Report 78 (*23*), shown in parentheses alongside their definitions. The benefit of connectivity is a difficult value to assign; different values can be calculated based on different connectivity aspects. The default value of connectivity benefit is based on the idea of savings in wait time for each additional bus available (increased frequency) per hour. The savings in wait time is calculated as the product of three elements: average reduction of wait time per passenger for each additional bus (≈30 seconds), average passengers carried by an additional bus (≈18 passengers), and hourly cost of wait time (≈20$/hour).

Unserved demand cost is calculated using the average household savings based on reduced vehicle miles traveled with the provision of public transportation service compared to no public transit service available. Bailey (24) reported a savings of $6,251 per year from public transportation use for two worker households. This savings is calculated based on the total annual cost difference between maintaining a single-vehicle and a two-vehicle household. It was assumed that providing transit service to a household worker allows that household to eliminate one vehicle and gain the annual savings. Not providing transit service to this household worker therefore forces the household workers to maintain the two-car and denies these savings (25). Total annual savings is divided by the annual mean travel time (≈240 working day * 2-hour round trip per day) (33) by public transit to calculate the unserved demand cost per passenger/hour (≈10 $/hour).

As mentioned earlier, demand is assumed to originate at the population centroids for each spatial unit (block group in this paper). ArcGIS proximity analysis tool is used to calculate the rectilinear distance (d_{ig}) from stop node i to demand centroid g. An average pedestrian walking

speed of 14400 feet per hour (*19*), and the capacity of 35 seats per bus are assumed as the default values in this paper. Dwell time is as assumed to be 30 seconds per stop. This can be estimated using more complex models using acceleration-deceleration velocities and the number of passengers alighting/boarding, however, dwell time estimation is not the focus of this research. Based on work by Levinson (*13*), a simple expression for dwell time can be applied, D = 4.0 + 1.7m_i, where m_i is the number of passengers alighting/boarding from bus, which one is greater. Applying this formula to our assumed dwell of 30 seconds, we are implicitly assuming 15-16 passengers boarding or alighting per stop (*19*).

3.3.4. Calculated Parameters

R_{gl} spatial coverage of origin *g* for transit line *l*

γ_{il} binary stop parameter indicating whether transit line *l* serves stop *i*

δ_{ghl} binary connectivity parameter, takes the value 1 if a transit line *l* directly connects origin *g* to destination *h*, and 0 otherwise

f_{ghl} connectivity decay factor from origin *g* to destination *h* for transit line *l*

μ_{igh} binary direction parameter, takes the value 1 if stop *i* serves trip demands from origin *g* to destination *h*, and 0 otherwise

a_{gh} peak hour trip demand from centroid *g* to centroid *h*

b capacity of the Transit stop (no. of peoples/stop)

θ proportion of passengers alighting service per trip for a transit line

N maximum number of buses available (buses)

U capacity of a vehicle

$$\rho_{igh} = \begin{cases} \mu_{igh}, & \mu_{igh} = 1 \\ M, & \mu_{igh} = 0 \end{cases}$$

M Arbitrary large number

The first parameter, R_{gl}, is the proportion of the area of origin g served by the transit line l. It is computed using the ratio of the intersection of the transit line buffer area and the origin block group to the total area of the origin block group. A 1/4 mile buffer around bus transit lines is used for calculating spatial coverage. The second parameter, γ_{il}, is a parameter that signifies whether bus stop, i, is served by a bus line l. The parameter takes the value "1" if a bus stop node i is served by bus line l, and 0 otherwise.

A binary connectivity parameter, δ_{ghl}, is calculated to measure access to possible trip destinations from a trip origin. This connectivity parameter represents the presence or absence of a direct transit line between O–D pairs. Thus, the connectivity parameter, δ_{ghl}, takes the value "1" if a passenger can make a trip from origin g to destination h using transit line l without making any transfer to other lines, and 0 otherwise. The parameter δ_{ghl} by itself assigns the same level of connectivity to all destinations regardless of the trip distance. To address this shortcoming, a connectivity decay factor, f_{ghl}, is calculated using a logistic decay function. This logistic decay function (an inverted s-curve) estimates the decreasing connectivity based on door-to-door travel time. Mamun et al. (*5*) provides additional detail on estimating the connectivity decay factor.

A binary direction parameter, μ_{igh}, is used to distinguish in-bound from out-bound stops and for assigning demand to the associated directional stops. It takes the value 1 if stop i serves demand from origin g to destination h, and 0 otherwise.

The peak hour demand (a_{gh}) is demand for travel from centroid g to centroid h. In this study, demand was gathered from the local planning model and distributed to the block groups from the original traffic analysis zone (TAZ) level proportional to the block group populations.

The parameter, θ, represents the proportions of passengers alighting transit service per trip. This parameter is introduced to model the total number of passengers that remains in bus service while travelling along a transit route. For the default or base value of θ (≈ 0.5), it is assumed that half of the passengers are alighting at each stop along a line. While restrictive, this parameter allows treatment of passenger turnover without conducting a full assignment of demand to the network. Further sensitivity analysis is presented in section 3.4 to examine the effect of different values of θ on decision variables.

3.3.5. Decision Variables

x_{igh} the number of passengers that travel from g to h using stop i;

y_i binary decision variable, takes the value 1 if a transit stop i is selected, and 0 otherwise;

B_l hourly vehicle runs for transit line l.

The variable x_{igh} assigns the number of passengers that travel from g to h using bus stop i. The assignment of passengers to bus stops considers walking distance and the connectivity of stops to the desired destinations. The second variable y_i provides the optimal stop locations needed to serve the trip demand for an area and the third variable B_l sets the number of vehicles assigned to each bus line.

3.3.6. Formulation

$$minimize\ Z = C_a \sum_i \sum_g \sum_h \frac{d_{ig}}{v_{walk}} x_{igh}\rho_{igh} + C_u \sum_g \sum_h \left(a_{gh} - \sum_i x_{igh} \right)$$

$$+ C_t \sum_i \sum_g \sum_h \sum_l a_{gh}\, \mu_{igh} y_i \gamma_{il} \frac{D}{3600}$$

$$- b_c \sum_i \sum_g \sum_h \sum_l \frac{B_l U}{p_g} R_{gl}\, \gamma_{il}\, \mu_{igh} \delta_{ghl}\, f_{ghl} \tag{1}$$

The objective function is defined in Eq. 1, which seeks to minimize the cost of access, the unserved demand, the dwell time cost cost for each additional stop, and maximize connectivity benefits. The first component concentrates on walking time to stops. The second cost component penalizes the demand that is left unserved by the bus system. The third cost component applies an hourly in-vehicle stopping time cost per passenger for each additional bus stop and identifies the savings/benefits in travel time with the optimal number of bus stops. The last component, connectivity benefit, considers the savings in wait time for better connectivity with an additional bus runs on a particular line. The four objectives can also be viewed as components of two larger competing objectives: (a) minimizing the cost of access, and (b) providing high quality service throughout the network (improve connectivity). In this paper, the cost of walk access and unserved demand are considered elements of access while the dwell and wait time pieces are part of connectivity.

The objective function is constrained by the following:

$$\sum_i x_{igh} \le a_{gh} \quad \forall g \in G, \forall h \in H \tag{2}$$

$$\sum_g \sum_h x_{igh} \le b\ y_i \quad \forall i \in I \tag{3}$$

$$\sum_g \sum_h x_{igh} \le \sum_l B_l\, \gamma_{il} U \quad \forall i \in I \tag{4}$$

Constraint (2) limits the total allocation of passengers for each O-D pair to the trip demand for that pair. Constraint (3) restricts the demand served at a stop based on stop capacity. A default value of 40 passengers boarding per stop *(b)* is used in this paper (*26*). Constraint (3) is dominant when a transit stop is served by low frequency transit lines, whereas constraint (4) dominates when high frequency transit serving a stop. Although constraints (3) and (4) capture elements of stop capacity, they are needed for accommodating differing line frequencies.

$$\sum_{i}\sum_{g}\sum_{h} x_{igh}\,\gamma_{il}\,\mu_{igh}\delta_{ghl}(1-\theta) \leq B_l\,U \quad \forall l \in L \tag{5}$$

$$\frac{B_l}{N} \geq \frac{\sum_i \sum_g \sum_h x_{igh}\mu_{igh}\delta_{ghl}}{\sum_g \sum_h a_{gh}} \quad \forall l \in L \tag{6}$$

$$\sum_{l} B_l \leq N \tag{7}$$

Constraint (5) places the capacity constraint for each line. It constrains the total number of passengers using a bus line to the total number of available seats. In this constraint, it is assumed that certain proportions of passengers (*θ*) are alighting service per trip using a transit line. Constraint (6) requires the proportional distribution of available buses to transit lines according to the demand along that line. Constraint (7) limits the numbers of buses assigned to the available buses for the bus system.

$$y_i \in \{0,1\} \quad \forall i \in I \tag{8}$$

$$x_{igh} \in Z^+ \quad \forall i \in I, \forall g \in G, \forall h \in H \tag{9}$$

$$B_l \in Z^+ \quad \forall l \in L \tag{10}$$

The final three constraints are definitional; restricting y_i to a binary variable, B_l and x_{igh} to positive integer values.

3.3.7. Solution Method

The proposed model is formulated as a mixed integer program (MIP). General Algebraic Modeling System (GAMS) software is used for modeling this problem and CBC (COIN-OR Branch and Cut) solver is used to solve the problem. The default values for the absolute optimality criteria is 0.0 (i.e., optca = 0.0) and the relative optimality criteria (i.e., optcr = 0.1) for CBC solver. However to obtain the optimal solution this research set the optcr value as 0.0 (i.e., CBC stops if the relative gap between the best known solution and the best possible solution is less than 0.0). Therefore, all the results reported in this paper are optimal solutions. CBC is part of the larger COIN-OR initiative (Computational Infrastructure for Operations Research) and is an open-source mixed integer programming solver written in C++. The solution processes applied branch-and-cut search techniques (*27*). The reader is referred to Martin (*28*), Mitchell (*29*) and Wolsey (*30*) for more details on the branch and cut algorithm.

3.4. Sample Illustration

We demonstrate the proposed model on the two bus routes: GS Av. and DD Av., in New Haven, CT shown in Figure 3.1. This example has 85 candidate bus stops serving 19 block group demand centroids. Five buses for GS Av. and four buses for DD Av. with capacity of 35 passengers are currently serving the peak hour demand. The BSLP model is used to optimize stop locations for this bus network, to assign available vehicles to each bus line, and to allocate the hourly trip demand to each stop for all O-D pairs. The GAMS code for this sample problem is provided in Appendix A.

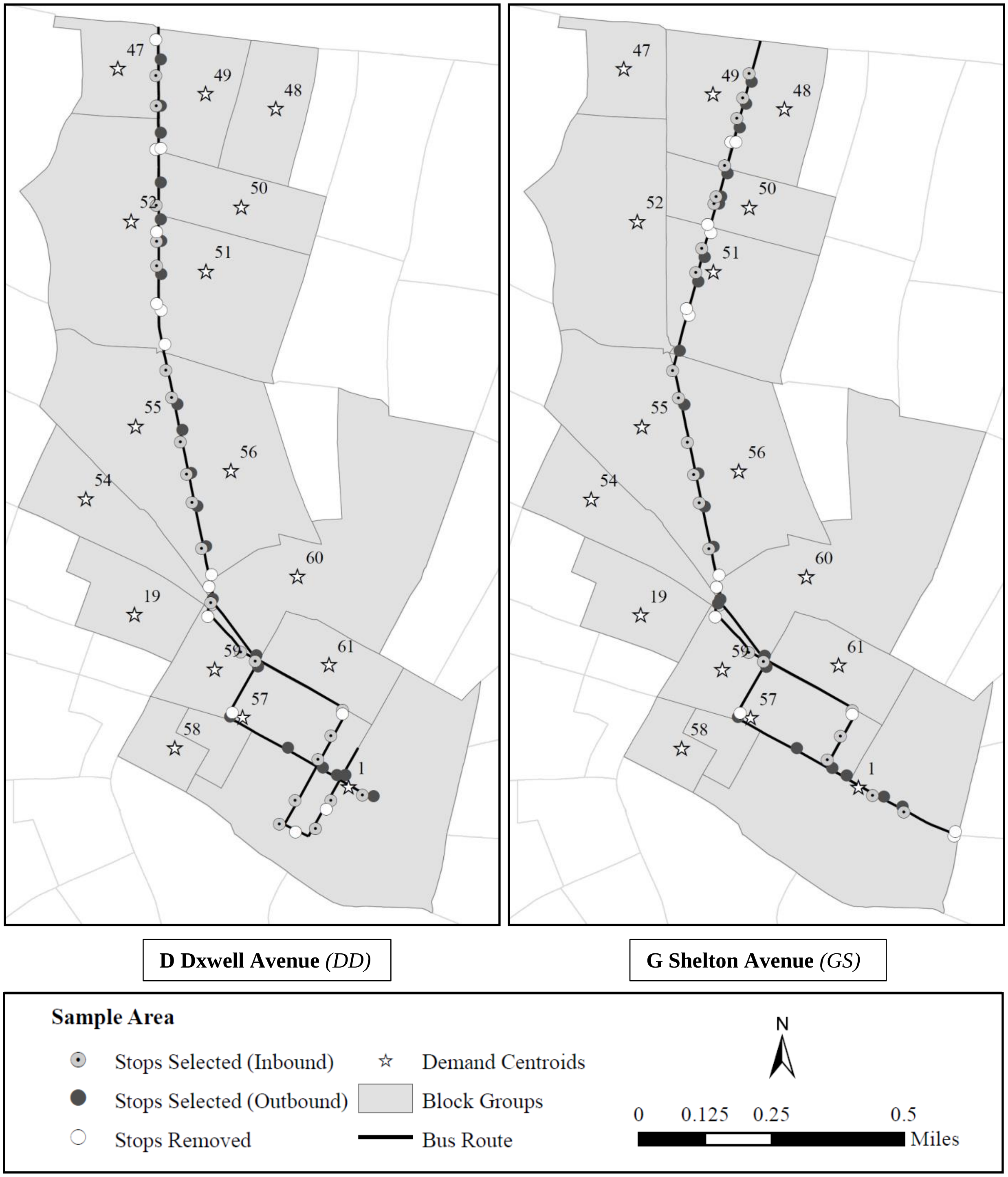

Figure 3.1. Sample bus service area with stop locations, demand centriods, and bus lines.

Table 3.1 summarizes the resulting stop changes, directional frequency changes and travel times for both bus lines. This result used the default values of cost parameters. Route DD

Avenue saw a removal of 11 inbound and 6 outbound stops, whereas route GS Avenue exhibited a reduction of 9 inbound and 6 outbound stops. The results obtained for this example reflect the impact of trip connectivity. Stop reduction for the shared route segment experienced fewer stops removed compared to single-service route segments. The reduction in stops in both directions led to a decrease in travel time between O-D pairs. Table 3.1 only shows the values of travel times between the O-D pairs with greatest separation.

Table 3.1. Summary of stop, frequency and travel time changes on bus lines.

Route	Direction	Stops Before/After	Stops Removed	Travel Time Before/After (to/from Downtown)	Vehicle Runs Before/After	Headway Before/After (min)
DD Avenue	Inbound	15/9 (17/12)	6 (5)	38/32.5[a]	4/3	15/20
DD Avenue	Outbound	10/6 (16/14)	4 (2)	39/36[a]	4/3	15/20
GS Avenue	Inbound	13/9 (17/12)	4 (5)	44/39.5[b]	5/6	12/10
GS Avenue	Outbound	14/10 (16/14)	4 (2)	48/45[b]	5/6	12/10

(): Values in the parentheses show the number of stops for the bus route segment that is shared by both routes.
[a] Travel times are calculated between centroid 47 and centroid 1.
[b] Travel times are calculated between centroid 49 and centroid 1.

It is also observed that the available buses are assigned according to the trip demands and level of trip connection for each route. It is expected that more buses would be assigned to lines that connects more origins and destinations (greater demand). Figure 3.1 shows that bus line GS Avenue connects more O-D pairs compared to DD Avenue, and so shifting a vehicle over to GS Avenue is a logical result.

3.4.1. Sensitivity Analysis

A sensitivity analysis is conducted to show the relationship among various cost parameters and design variables. The Problem was solved for a wide range of parameter values and combinations. The results are shown in Figures 3.2, 3.3 and 3.4.

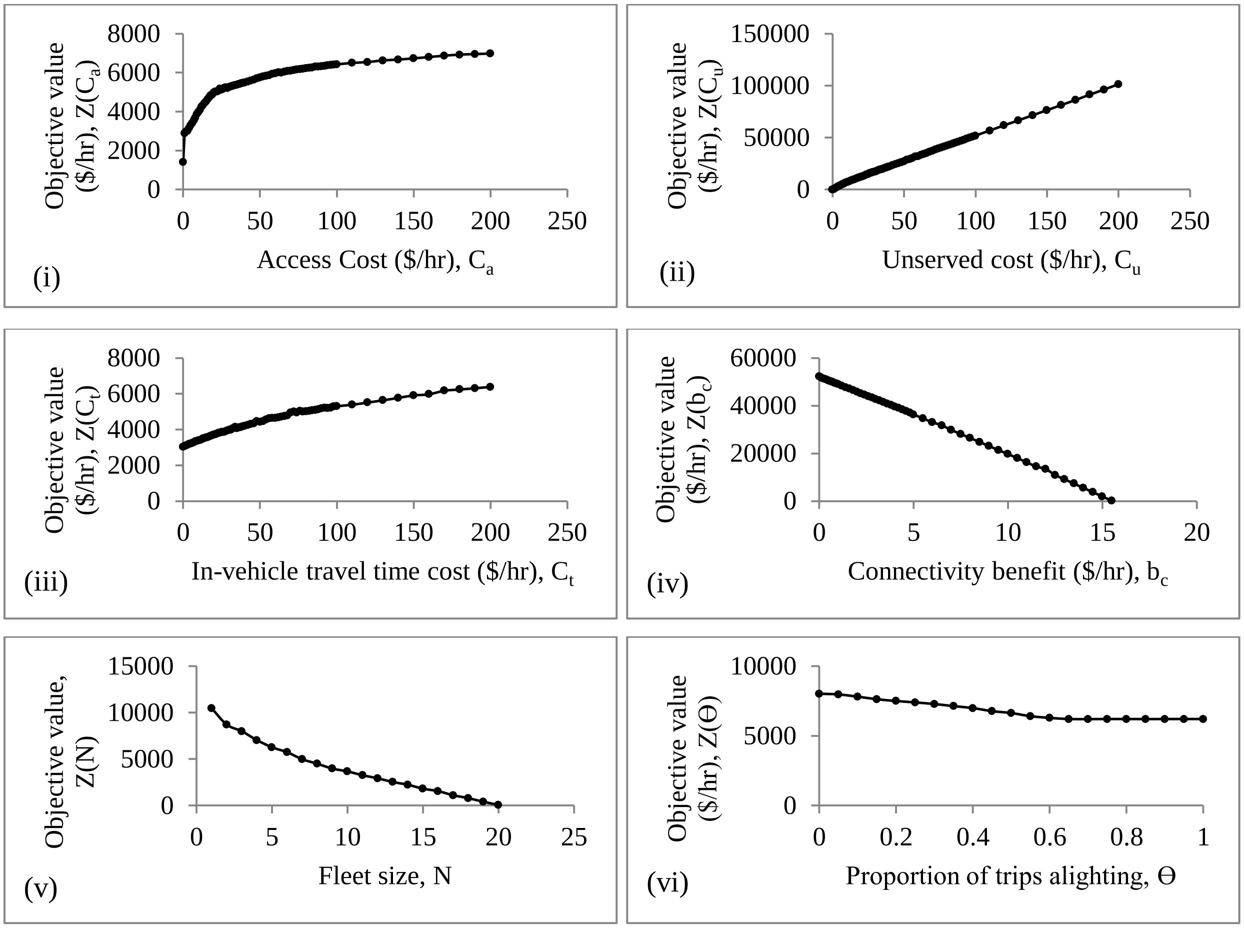

Figure 3.2. Total cost plotted against various values of (i) access cost; (ii) unserved cost; (iii) in-vehicle travel time cost; (iv) connectivity benefit and; (v) fleet size; (vi) proportion of trips alighting.

Figures 3.2(i), 3.2(ii) and 3.2(iii) show the objective values while altering the access/walking costs (C_a), unserved cost (C_u), and in-vehicle travel time cost (C_t). When the cost parameters increase, the objective function values increases and which is expected. Figure 3.2(iv) shows the objective values for different connectivity benefit values (b_c). It is found that the objective value decreases as the connectivity benefit value increases, as expected. Figure 3.2(v) depicts that the objective value decreases with the increased values of fleet size (N). With the increase in fleet size, the service frequency is getting higher and which increase the connectivity benefits and resulting lower objective values. Figure 3.2(vi) presents the variation

of objective values with the assumed proportions of trip alighting and it shows that objective values decreases with the increase in proportion values. A higher proportion of trips alighting increase the available seats for serving other demand. In other words, a higher proportion of trips alighting decreases unserved demand and favors a lower total cost.

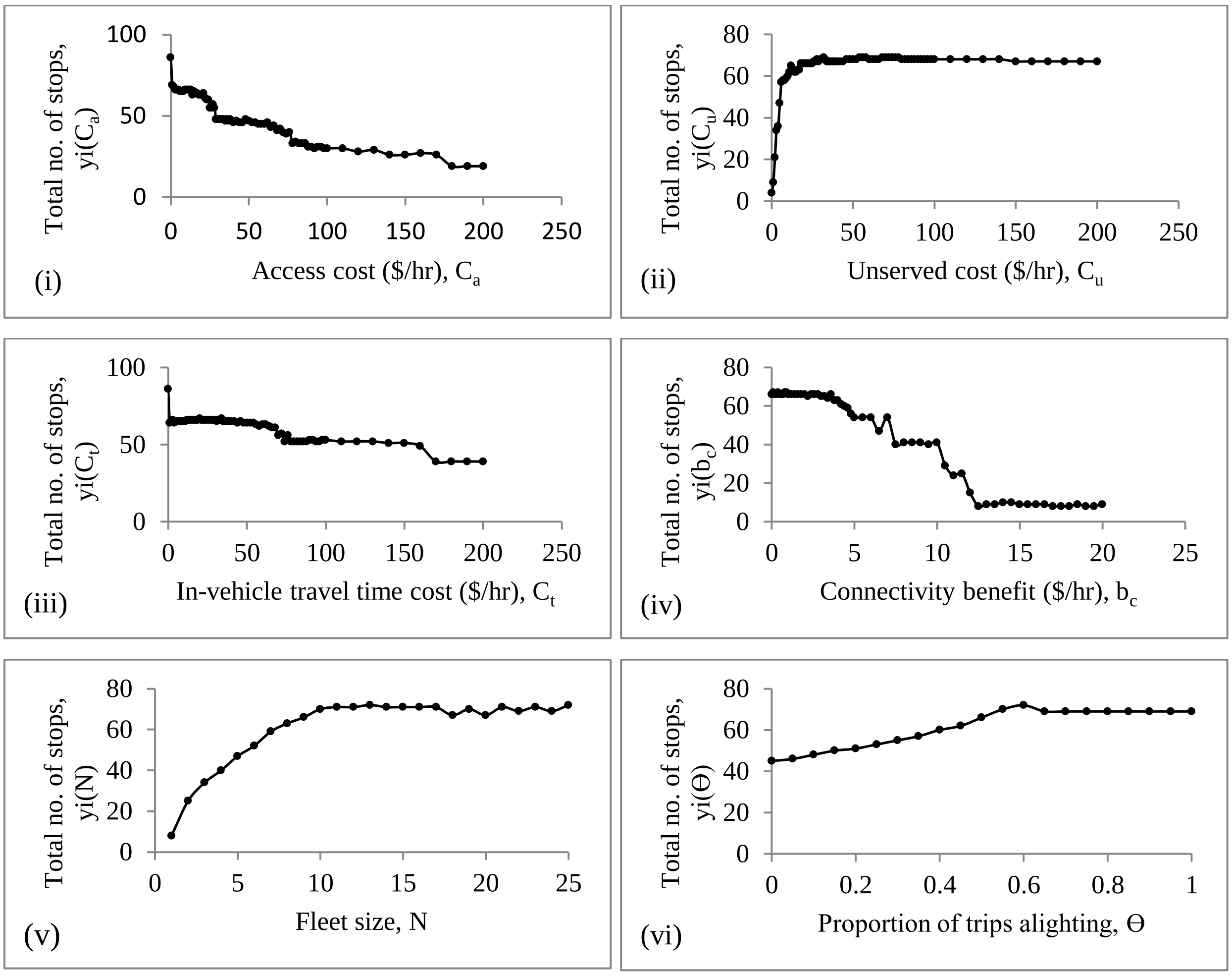

Figure 3.3. Total number of stops plotted against various values of (i) access cost; (ii) unserved cost; (iii) in-vehicle travel time cost; (iv) connectivity benefit and; (v) fleet size; (vi) proportion of trips alighting.

Figure 3.3(i) shows the number of stops selected while altering the access/walking costs (C_a). When C_a increases, the total number of stops required decreases, which is surprising until one looks at Figure 3.4(i) and realizes that demand is dropping as access cost increases, and

therefore fewer stops are needed to serve the demand. Figure 3.3(ii) shows the number of stops required for the different values of unserved cost (C_u) per stop – and suggests that the number of stops is quite stable once unserved demand cost exceeds about $10/hour. Figure 3.3(iii) shows that in-vehicle travel time cost increases results in solutions with fewer stops, which is what one would expect. Figure 3.3(iv) shows the variation in total number of stops selected as a function of connectivity benefit values (b_c). It is found that the optimal number of stops decreases as the connectivity benefit value increases, a result of the increased value on in-vehicle time implied by b_c. Figure 3.3(v) depicts that number of stops required increases with the values of fleet size – which is a result of the increased demand being served by a larger fleet. This model selects more stops when higher proportion of trips alighting (Figure 3.3(vi)), that is, more stops are required to meet the demand for boarding and alighting along a line.

Figure 3.4 shows the variations in total number of served demand with the variations of cost parameters, fleet size and proportion of trips alighting. It shows that the variations in served demand closely followed the same variations of total number of stops required. The results are consistent with the author's expectations as more stops are required to serve higher demand.

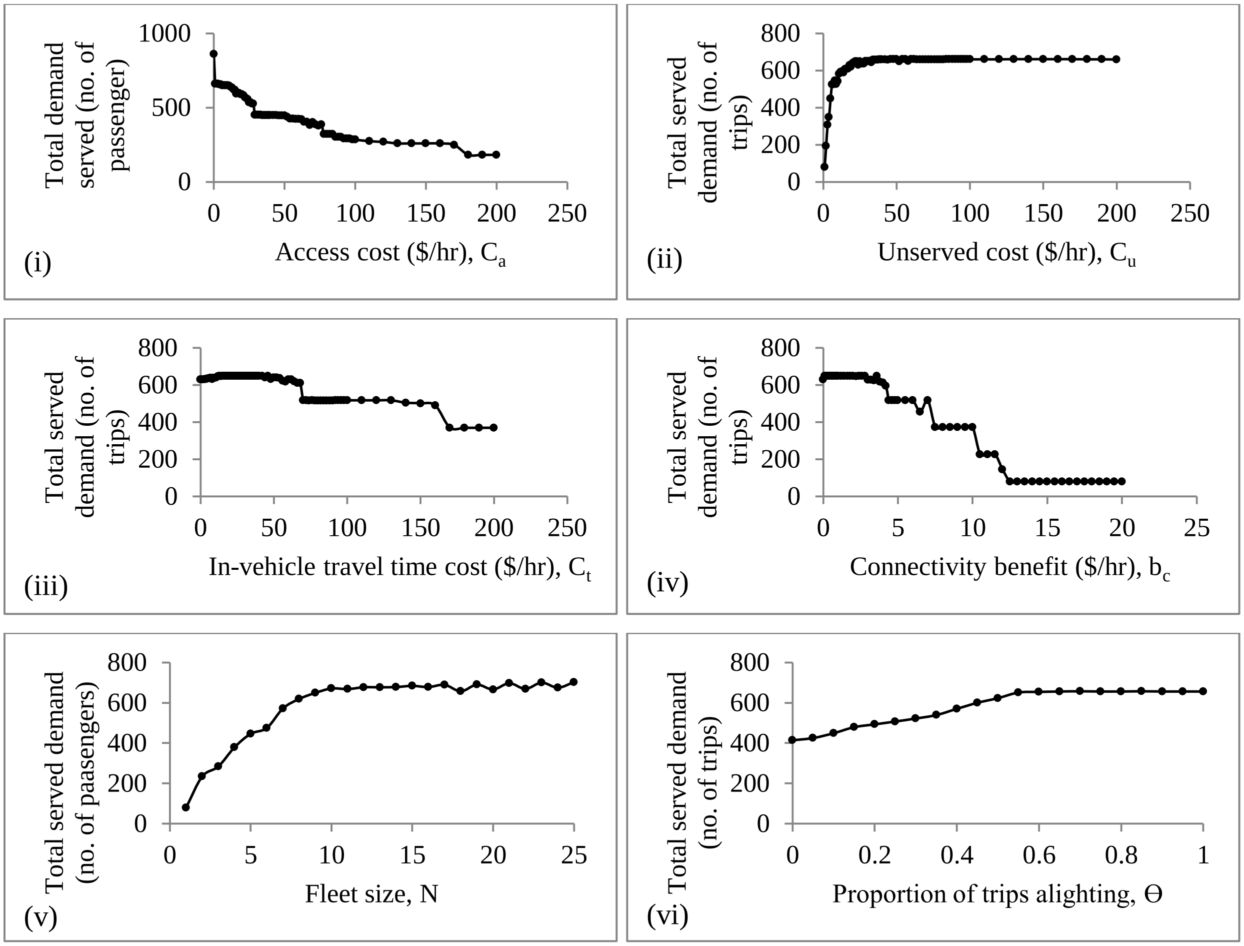

Figure 3.4. Total number of served trip demands plotted against various values of (i) access cost; (ii) unserved cost; (iii) in-vehicle travel time cost; (iv) connectivity benefit and; (v) fleet size; (vi) proportion of trips alighting.

3.4.2. Tradeoff Analysis

There is obviously a great deal of interaction between the parameter values and their impact on the solution. In multi-objective programming it is often desirable to find not only a single "optimal solution" but to establish a set of "non-dominated/pareto-optimal" solutions that create a boundary of best solutions for a range of weights on the multiple objectives. The most straightforward situation is the one in which there are two objectives. The tradeoff analysis presented in this section has organized the objective function into two primary objectives:

minimizing access cost (through walking cost and unserved demand cost) and maximizing connectivity to the network (through dwell time and wait time reduction). The tradeoff solution space between the two primary objectives for the bus service system of the sample illustration is shown in Figure 3.5. This point cloud depicts the solution of the BSLP for the example network for over 45000 combinations of the six parameters modified in the sensitivity analysis. The Java code to generate the solution points is provided in Appendix B. While not exhaustive over all possible parameter values, the point cloud captures solutions for all possible combinations of parameters within the following ranges:

- Cost of Access *(C_a)*: 5 – 40 $/hour in increment of 1
- Cost of unserved demand *(C_u)*: 5 – 30 $/hour in increment of 1
- Cost of in-vehicle travel time *(C_t)*: 20 – 50 $/hour in increment of 2
- Benefit of connectivity *(b_c)*: 1 – 10 $/hour in increment of 0.5
- Fleet size *(N)*: 5 – 15 buses in increment of 1
- Proportion of trips alighting *(θ)*: 0 – 1.0 in increment of 0.05

The ranges for parameter values are based on the sensitivity analysis and their practical implications. The y-axis (Access) in Figure 3.5 presents the total of access cost and unserved cost, and the x-axis (Connectivity) accounts for total in-vehicle stopping time and connectivity benefits form the objective function. The non-dominated solutions are those along the lower boundary of the cloud – often referred to as the "non-dominated frontier" or "Pareto frontier". The solutions along this frontier provide the trade-off between access cost minimization and connectivity benefit maximization for any combination of parameters within the ranges specified. A simplified search algorithm is proposed in this research based on the frontier generation algorithm by Baykasoglu et al. (*31*). The algorithm is coded in Java for generating the

"non-dominated/pareto optimal" solutions. First, all the solutions are sorted by x-values (i.e., access cost) starting with minimum. For solution points having similar x-values, solution with minimum y-value (i.e., maximum connectivity benefit) is selected as the pivot point for the first iteration. Next a set of neighbor solution (p,q) is generated within a circular neighbor area of pivot point having a step-size value as the radius. Now if there is any point (p,q), where x<p and y<q, (p,q) is removed from the result and the pivot point is updated for the next iteration. This step is performed repetitively until there is no point left to remove any more from this solution space. The Java code for the pareto-frontier generation algorithm is provided in Appendix C.

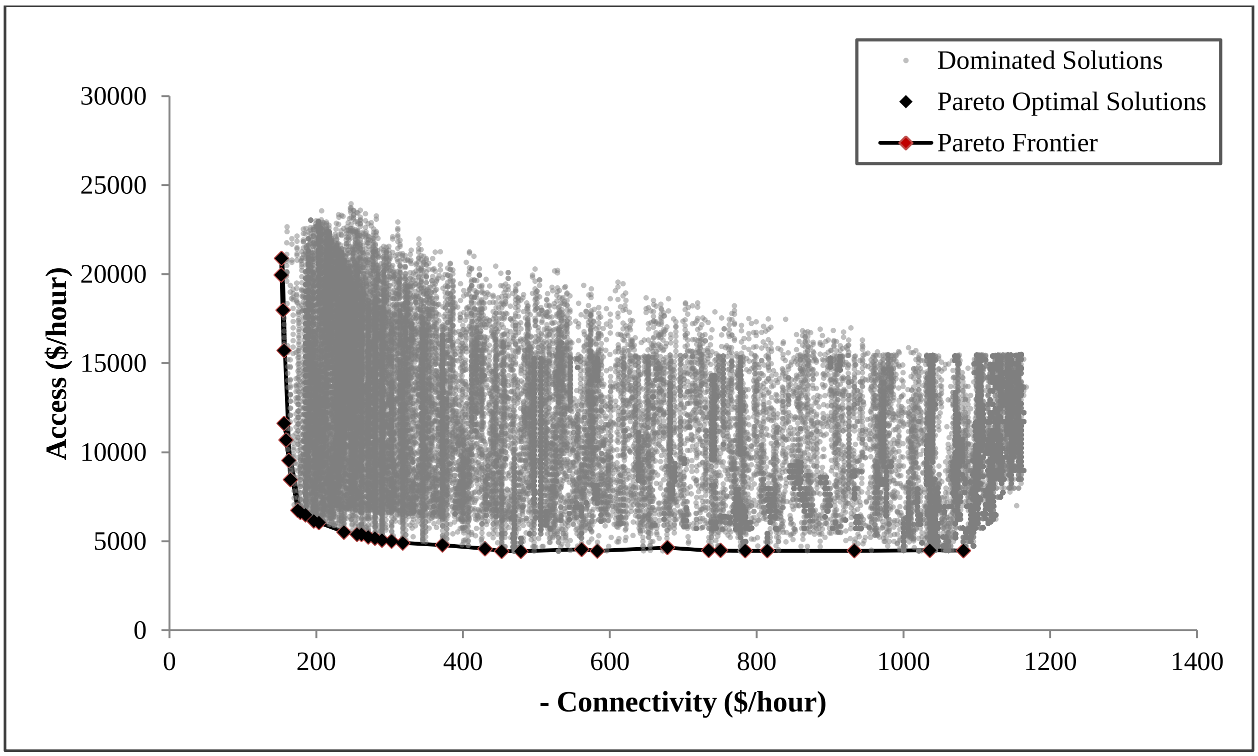

Figure 3.5. Access and connectivity tradeoff for the BSLP *(Sample Illustration).*

3.5. Case Study Application and Results

A case study application was conducted to demonstrate the proposed methodology using the full bus network in New Haven, Connecticut (Figure 6). Trip demand data come from the Connecticut Department of Transportation's Statewide Travel Model. This application focused on only bus HBW and HBO demand during the weekday morning peak hour. We consider a block group-level analysis, requiring the TAZ-level demand data be converted to census tract level trip demands and then again distributed to block group. The conversion was conducted based on proportional distribution of population for each block groups within a census tract.

Currently, 4.5% of all trips in Connecticut are made using public transit (*32*). The population of New Haven proper is 130,000, as reported in the 2010 census; though the Greater New Haven metropolitan area has a population near 860,000. This city is comprised of 106 block groups (Figure 3.6) and the average block group size is 0.2 square miles. Although public transit in this city consists of bus and rail service, the bus system is the dominant local service with 981 stops in service and 19 bus lines for local trips within the city. Some of these lines serve inter-city travel as well and the rail service is for inter-city travel only along the coastline.

This analysis used a partial bus network operated by CTtransit within the city of New Haven, CT. There are portions of both the local and regional transit network connected to New Haven that are not included in the application presented in this document. . This is a limitation of the application, not of the methodology. The method is general enough to allow application on a fuller representation of a network that would then capture more of the regional aspects of the problem. This model does provide the transit planner/provider with the flexibility to control the analysis boundaries: it can be carried out on a network of a single line, multiple lines serving a single city or a network serving multiple cities.

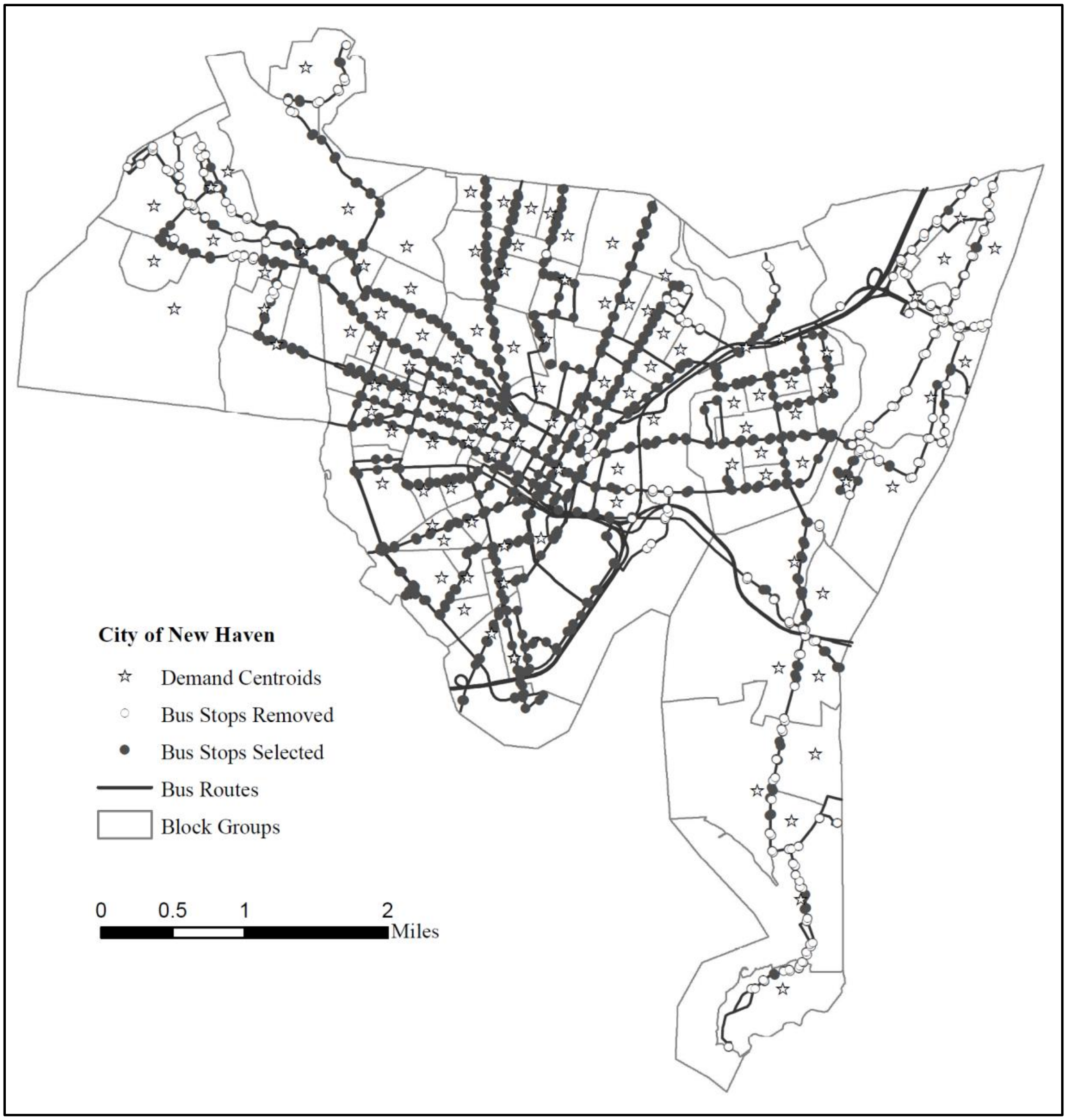

Figure 3.6. Demand centroids, bus stop locations and routes, city of New Haven, CT.

The BSLP formulation was applied to the New Haven study area on four different cases, keeping constant fleet size (70 buses), and trips alighting at each stop ($\theta = 0.5$). The cost parameters are used to distinguish the first three cases: In case 1, the model focused on

accessibility, using C_a= \$40 /hour and C_u= \$30 /hour. That is, access cost and unserved cost are increased from their base value so that spatial access can be improved through minimizing walk distance and more demand can be served by providing increased penalty for unserved demand. In the second case, default values from the literature are applied. In the third case, connectivity is given priority by weighting the cost of in-vehicle travel time saving and the benefit of system connectivity higher than the default. Case 4 was considered to evaluate existing network by considering no stops redundancy exists.

Table 3.2. Application to the City of New Haven, CT.

	Case 1 (Accessibility-focused)	**Case 2** (Default)	**Case 3** (Connectivity-focused)	**Case 4** (Existing)
Cost Parameters *(\$/hour)*	$\mathbf{C_a=40}$, $\mathbf{C_u=30}$, C_t=30, b_c=1	C_a=20, C_u=10, C_t=30, b_c=1	C_a=20, C_u=10, $\mathbf{C_t=50}$, $\mathbf{b_c=10}$	C_a=20, C_u=10, C_t=30, b_c=1
Objective value, Z *(in thousand \$/hour)*	369	374	371	395
Access cost *(in thousand \$/hour)*	379	390	401	398
Connectivity benefit *(in thousand \$/hour)*	10	16	30	3
Total number of stops ***(% stop reduction)***	812 ***(17.2)***	776 ***(20.9)***	747 ***(23.8)***	981 ***(0)***
Total served demand ***(% served demand)***	3214 ***(76.2)***	3,070 ***(72.8)***	2,932 ***(69.5)***	3290 ***(78.1)***

Table 3.2 summarizes the results of application for these three cases. For case 1, the number of stops required increases by 36 stops than default case, indicating that more demand is served if walk time is viewed as less burdensome. It shows a reduction in access cost

(approximately $11 thousand /hr) from the default case, but also a significant decrease ($6 thousand /hr) in connectivity benefits with the greater number of stops. In comparison with default case, case 3 results show that the connectivity benefit increased significantly, although the access cost increased with the fewer number of stops. In all the cases it seems there is substantial stop redundancy exists and anywhere from 17 – 25% of stops can be removed and still maintain good system connectivity. For connectivity-focused case, the model selects 747 stops (as shown in Figure 3.6) to serve approximately 70% of total trip demands much more efficiently compared to 981 stops to serve approximately 78% of total trip demands (Existing case). Access cost and connectivity benefits for all the cases show that case 1 (accessibility focused) dominates the existing system in both accessibility and connectivity whereas case 3 (connectivity-focused) dominates the existing system in connectivity only. Existing case dominates case 3 in access due to shorter access distance resulted with more stops compared to case 3. Although intuitive, the result demonstrates the tradeoffs between access and connectivity being made in this larger-scale application.

3.6. Conclusions

The bus stop location model (BSLP) focuses on locating stops in an existing public transit network – allowing for a reduction or reallocation of stops in the network to optimally balance the accessibility of the system with its connectivity. This model seeks to select the best number and location of stops in order to increase service efficiency and provide convenient access to the system. The model considers both walking distance and system connectivity. A mixed-integer multi-objective optimization approach is presented to identify optimal stop locations, assign demand to each stop to all O-D trip pairs, and set vehicle frequencies for all bus lines.

The model results suggest that wait time and dwell time savings at stops could be realized with adjustments to vehicle frequencies while still maintaining good system access. From a bus rider's perspective, the results indicate that the increase in access distance from stop consolidation is offset by the wait time and dwell time reductions, which thus improve the connectivity of the system.

The sensitivity analysis finds that the total cost, the optimal number of stops, and total served demand are affected by the users' access cost, unserved cost, connectivity benefit, in-vehicle stopping time cost, available fleet size, and the proportions of trip demands alighting. The tradeoff analysis between access and connectivity shows that average access cost is increased to some extent due to a reduction in the total number of bus stops but this is balanced by the improvements in service efficiency and network connectivity.

The BSLP model provides an alternate approach to model public transit access and connectivity. The BSLP model has the benefit of a detailed assignment of demand to stops. And while this study assumed an assignment of demand based on perfect information of the shortest paths and a fixed demand, the model is capable of accepting more sophisticated estimates of traveler assignment.

Assumptions in this paper in this development of this model to improve tractability and computational performance. Listed below are some ways of extending this method by relaxing assumptions, together with their implications and effects on solutions:

- This approach assumes transit lines have been defined a priori. This assumption is necessary to distinguish this facility location problem from the vehicle routing problem. However, this problem can be extended to the location-routing problem considering

vehicle routing aspects. These two problems can then be solved as sequential subproblems as studied by Barreto et al. (34) and Baldacci et al. (35).

- In this formulation, it is assumed that people are likely to use stops that serve more lines and have connection to more destinations. Existing research (6, 36, 37) on walking access to transit service have suggested that transit users are likely to use the nearest stop when all other available stops are equally attractive in service characteristics. However, they are willing to walk longer distances for a more attractive stop, which provides faster and more reliable service or requires fewer transfers to complete a trip.
- The demand is assumed to be located at the population centroid of each zone in order to minimize demand aggregation error. A common difficulty with modeling location problems that occur in urban or regional areas is that the number of demand points may be quite large, since each private residence might be a demand point. In this case it may be impossible to include every demand point in the corresponding model. Centroids were used to improve tractability, but other aggregation and interpolation methodologies can be considered to improve the spatial representation of demand. A study by Plastaria (42) suggested that aggregating the data at the centroid of the data set leads to much higher precision bounds on the aggregation errors. The exact locations of disaggregate demands and the actual distances between these demand locations to the bus stops can provide a good estimation of access cost and can be useful for a small scale network analysis.
- A limitation of this application is the O-D demand data, which are manipulated twice to get the trip matrix at the block group level. The O-D demand data might not represent the actual O-D demand for block groups. Employing a more robust demand modeling tool to

estimate the O-D pairwise demand at the block group level would improve the model results.

- Solving this model is computationally expensive. It would benefit from refined heuristics designed to speed computation while still providing good solutions. One potential improvement is to consider a more sophisticated representation of dwell time, one based on passengers' boarding/alighting at each stop.

References

1. Mamun, S., and Lownes, N. Measuring service gaps: an accessibility-based transit need index. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2217, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 153-161.
2. Murray, A., Davis, R., Stimson, R., and Ferreira, L. Public Transport Access. *Transportation Research Part D*, Vol. 3, 1998, pp. 319–328.
3. Chien, S., and Qin, Z. Optimization of bus stop locations for improving transit accessibility. *Transportation Planning and Technology*, Vol. 27, No. 3, 2004, pp. 211-227.
4. Delmelle, E., Li, S., and Murray, A. Identifying bus stop redundancy: A gis-based spatial optimization approach. *Computers, Environment and Urban Systems*, Vol. 36, 2012, pp. 445-455.
5. Mamun, S., Lownes, N., Osleeb, J., and Bertolaccini, K. A method to define public transit opportunity space. *Journal of Transport Geography*, Vol. 28, 2013, pp. 144-154.
6. O'Sullivan, S., and Morrall, J. Walking distances to and from light-rail transit stations. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1538, Transportation Research Board of the National Academies, Washington, D.C., 1996, pp. 19-26.
7. Ceder, A, Hassold, S., and Dano, B. Approaching even-load and even-headway transit timetables using different bus sizes. *Public Transportation*, Vol. 1, 2013, pp. 1-25.

8. Mauttone, A., and Urqhart, M. A multi-objective metaheuristic approach for the transit network design problem. *Public Transportation*, Vol. 1, 2009, pp. 253-273.

9. El-Geneidy, A., Strathman, J., Kimpel, T., and Crout, D. The effects of bus stop consolidation on passenger activity and transit operations. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1971, Transportation Research Board of the National Academies, Washington, D.C., 2006, pp. 32-41.

10. Furth, P., and Rahbee, A. Optimal bus stop spacing trough dynamic programming and geographic modeling. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1713, Transportation Research Board of the National Academies, Washington, D.C., 2000, pp. 15-22.

11. Demetsky, M., and Lin, B. Bus stop location and design. *Transportation Engineering Journal of ASCE*, Vol. 108, 1982, pp. 313–327.

12. Federal Transit Administration. Guidelines for the location and design of bus stops. Transit cooperative research program, report 19. Washington, DC: *National Academy Press*, 1996.

13. Levinson, S. Analyzing transit travel time performance. In *Transportation Research Record: Journal of the Transportation Research Board,* No. 915, Transportation Research Board of the National Academies, Washington, D.C., 1983, pp. 1-6.

14. Schobel, A. Locating stops along bus or railway lines – a bicriteria problem. *Annals of Operations Research*, Vol. 136, 2005, pp. 211–227.

15. Ceder, A., and Wilson, M. Bus Network Design. *Transportation Research Part B*, Vol. 20B, No. 4, 1986, pp. 331–344.

16. Murray, T. Strategic analysis of transport coverage. *Socio-Economic Planning Sciences*, Vol. 35, No. 3, 2001, pp. 175–188.

17. Ghoneim, N., and Wirasinghe, S. Optimum zone configuration for planned urban commuter rail lines. *Transportation Science*, Vol. 21, No. 2, 1987, pp. 106-l14.

18. Saka, A. Model for determining optimum bus-stop spacing in urban areas. *Journal of Transportation Engineering*, Vol. 127, 2001, pp. 195–196.

19. Li, H., and Bertini, R. Assessing a Model for Optimal Bus Stop Spacing with High-Resolution Archived Stop-Level Data. In *Transportation Research Record: Journal of*

the Transportation Research Board, No. 2111, Transportation Research Board of the National Academies, Washington, D.C., 2009, pp. 24-32.

20. El-Geneidy, A., and Surprenant-Legault, J. Limited-stop bus service: an evaluation of an implementation strategy. *Public Transportation*, Vol. 2, 2010, pp. 291-306.
21. Murray, A., and Wu, X. Accessibility tradeoffs in public transit planning. *Journal of Geography System*, Vol. 5, No.1, 2003, pp. 93–107.
22. Wu, C., and Murray, A. Optimizing public transit quality and system access: the multiple-route, maximal covering/shortest-path problem. *Environment and Planning B: Planning and Design*, Vol. 32, 2005, pp. 163–178.
23. ECONorthwest and Parsons Brinckerhoff Quade & Douglas Inc. TCRP Report 78: Estimating the Benefits and Costs of Public Transit Projects: A Guidebook for Practitioners. *Transportation Research Board*, National Research Council, Washington D.C., 2002.
24. Bailey, L. Public Transportation and Petroleum Savings in the U.S.: Reducing Dependence on Oil. ICF International, *American Public Transportation Association*, Washington DC: 35, 2004.
25. Lownes, N., and Machemehl, R. Commuter rail circulator network design and its implications for transit accessibility. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2042, Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 90-97.
26. Osleeb, J., Moellering, H., and Cromley, R. A Geographical Analysis for Bus Stop Location for Route Design: Method and Case Study, Discussion Paper No. 48, *Department of Geography*, Ohio State University, 1975.
27. CBC User Guide 2005. *http://www.coin-or.org/Cbc/index.html*
28. Martin, A. General Mixed Integer Programming: Computational Issues for Branch-and-Cut Algorithms. Computational Combinatorial Optimization: Lecture Notes in Computer Science, Vol. 2241, 2001, pp. 1-25.
29. Mitchell, J. Branch-and-Cut Algorithms for Combinatorial Optimization Problems. *Handbook of Applied Optimization*, Oxford University Press, 2000, pp. 1-17.
30. Wolsey, L. *Integer Programming*. John Wiley, New York, 1998.

31. Baykasoglu, A, Owen, S and Gindy, N. A taboo search based approach to find the Pareto optimal set in multiple objective optimisation. *Engineering Optimization*, Vol. 31, 1999, pp. 731-748.

32. U.S. Census, 2010. [*http://www.census.gov*](http://www.census.gov)

33. 2009 American Community Survey (Census). *http://www.census.gov/acs/www/data_documentation/2009_release/*. Accessed November 10, 2013.

34. S. Barreto, S., Ferreira, C. Paixao, J., Santos, B. Using Clustering Analysis in a Capacitated Location-routing Problem. *European Journal of Operational Research*, Vol. 127, No. 3, 2007, pp. 968–977.

35. Baldacci, R., Mingozzi, A., Calvo, R. An exact method for the capacitated location-routing problem. *Operations research*. Vol. 59, No. 5, 2011, pp. 1284-1296.

36. Farhan, B., Murray, A. T. Distance decay and coverage in facility location planning. *Annals of Regional Science*, Vol. 40, 2006, pp. 279–295.

37. Daniels, R., Mulley, C. Explaining walking distance to public transport: the dominance of public transport supply. *World Symposium on Transport and Land Use Research*. Whistler Canada, 28–30 July, 2011.

38. Francis, R., Lowe, T., Rushton, G., Rayco, M. (1999). A synthesis of aggregation methods for multi-facility location problems: strategies for containing error. *Geographical Analysis*, Vol. 31, 1999, pp. 67–87.

39. Francis, R., Lowe, T., Tamir, A. Demand point aggregation of location models. Z. Drezner & H. Hamacher (Eds.), *Facility location: applications and theory*. Berlin: Springer, 2002.

40. Zhao, P. Analysis of aggregation effects in location problems. *Ph. D. Dissertation, Dept. of Industrial Engineering, University at Buffalo (SUNY)*, Buffalo, NY, 1996.

41. Hale, T., Moberg, C. Location science research: a review. *Annals of Operations Research*, Vol. 123, 2003, pp. 21–35.

42. Plastria, F. New error bounds in continuous minisum location for aggregation at the gravity centre. *Studies in Locational Analysis*, Vol. 14, 2000, pp. 101–119.

*This paper is in preparation to submit for publication in **Public Transportation Journal**.

CHAPTER 4: A RANDOMIZED FEATURE SELECTION ALGORITHM FOR BUS STOP LOCATION PROBLEM

Abstract

A bus stop location model (BSLP) is formulated considering three important sub-problems: optimal stop location, demand allocation and service frequency. Given a set of candidate bus stops, the level of trip connectivity and origin-destination pair-wise trip demands, the problem is to answer the questions: How many bus stops should be required and where should they be located? Which users/demands are to be assigned to each bus stop? What should be the service frequency for each bus line while being feasible in fleet requirements? The BSLP is formulated as multi-objective capacitated facility location problem, which minimizes the total access cost while maximizing the connectivity benefits for satisfying all demands. A randomized feature selection algorithm is used to solve this problem. Computations of results are given for the bus network in City of New Haven, CT. The results have been compared with the optimal solution to the same problem, obtained by branch and cut algorithm. The feature selection algorithm is shown to be very efficient in terms of computation compared with the known optimal solutions. The average solution time for the feature selection algorithm is 140 CPU seconds compared to 480000 CPU seconds for the branch and cut algorithm. The maximum gap is 8.7% against the optimal solution.

4.1. Introduction

Bus stop locations are important to individuals for gaining access to transit service and completing trips to the activities the system connects. Access distance or time to a bus stop is an important factor that significantly influences the selection of stop locations as more residents within close proximity of bus stops greater bus service utilization is expected. Additionally, people will likely use bus stops served by multiple lines to have better trip/activity connectivity to multiple destinations. Another important consideration of stop location is travel time. Clearly, frequent stops increases dwell time, which results in a smaller portion of the network being connected within a travel time threshold. Therefore, it is necessary to select optimal bus stops in a system that will exclude the redundant and inefficient bus stops in order to increase service efficiency. However, stop placement should not be reduced to the extent that it decreases access.

A strategic approach for identifying the redundancy and inefficiency associated with the placement of bus stops may be structured using the facility location problem (FLP). FLP is a classical problem with vast literature and numerous applications in transit stop location. There are a variety of models representing a variety of facility location problems (*1, 2*). The Capacitated Facility Location Problem (CFLP) is a variant of the FLP, which includes capacities for the facilities. The CFLP involves the selection of locations where facilities with limited capacities are established and the assignment of demands to facilities so as to satisfy demands while minimizing total operating and transportation costs. In this paper, a multi-objective bus stop location problem (BSLP) is formulated as a capacitated facility location problem. This problem considers locating bus stops, allocating directional trip demands to the optimal bus stops and assigning service frequencies to the bus lines. The objective of BSLP is to optimize two cost components - access cost and connectivity benefit. Access cost is defined as the combination of

access/walking cost to transit stops and unserved demand cost, whereas the benefit of system connectivity is measured as the savings in in-vehicle stopping time and wait time at stops from total travel time.

Many exact algorithms and heuristic methods have been developed to solve the CFLP in the last 40 years (*3, 7, 8, 9, 16*). Exact algorithms usually exist only for small problems whereas heuristic procedures have been used for large practical problems. This paper focuses on a heuristic solution procedure for the BSLP. The most common heuristic approach in previous literatures to solving the CFLP is the use of Lagrangian relaxation. The traditional approach for solving this problem focuses on obtaining good Lagrangian duals, whose solutions improve the lower bounds provided by a linear programming (LP) relaxation (*4, 5, 6*).

Several greedy heuristics also have been developed to solve the CFLP. These greedy heuristics involve generating moves to explore the solution neighborhood effectively. Jacobsen (*7*) proposed two heuristic procedures for the CFLP: namely ADD and DROP. ADD starts with all facilities closed and then the facility that causes the maximum total cost reduction is opened. This search ends when no more facilities can be opened to further reduce the total cost (*8*). In DROP, all facilities are initially open and a facility is closed if closing it results in the maximum reduction in the total cost. This search ends when closing a facility does not result in any further reduction in the total cost. However, the drawbacks for these search procedures (as with all greedy methods) are being trapped in local optimality. Another widely used heuristic procedure is Tabu search (TS), which guides local heuristic search procedures to explore the solution space beyond local optimality. The TS procedures have been developed for more complicated facility location problems, with some TS heuristics application for the FLP found in (*9*) - (*15*). More

recently Sun (*16*) developed a TS heuristic procedure for the CFLP and found the procedure could result much better solutions using much less CPU time compared to Lagrangian relaxation.

In this study, a feature selection (FS) algorithm is proposed to solve the bus stop location problem. The optimal number of bus stops is selected with the objective of minimizing the total access cost and maximizing the connectivity benefits while satisfying all trip demand. The rest of the paper is organized as follows: the problem formulation is given in section 4.2 and a brief introduction to the feature selection algorithm is provided in section 4.3. The proposed algorithm is stated in section 4.4. Computational results and discussion are presented in section 4.5. Finally, conclusions are given in section 4.6.

4.2. Problem Formulation

A formalization of the BSLP with associated descriptions of the formulation components is provided in this section. This formulation includes the necessary sets and indices, parameters and data, assumptions, objective function, and the constraint set. The network consists of candidate bus stop nodes, $i \in I$; origin demand centroid nodes, $g \in G$; the destination demand centroid locations, $h \in H$ and set of transit lines, $l \in L$.

It is assumed that the transit lines have been defined a priori and people will tend to use stops that serve more lines and have connection to more destinations. Demand is assumed to be located at population centroids and demand is calculated as the bus trips for home based work (HBW) and home based other (HBO) trips only for O-D pairs and is fixed for the morning peak period.

The problem is constructed as a mixed integer program with decision variable x_{igh}, assigning the number of passengers that travel from g to h using bus stop i. The assignment of

passengers to bus stops considers walking distance and the connectivity of stops to desired destinations. The second variable y_i provides the optimal stop locations needed to serve the trip demand for an area and the third variable B_l sets the number of vehicles assigned to each bus line.

Four cost weights are used in constructing the objective function: the cost of access, C_a, the cost in in-vehicle travel time, C_t, the cost of unserved demand, C_u, and the benefit of connectivity, b_c. Additionally d_{ig} represents the rectilinear distance from stop node i to demand centroid g, v_{walk} is the average pedestrian walking speed in feet per hour, U is the capacity/seats per bus, p_g is the population at demand centroid g, and D is dwell time per stops. Mamun and Lownes (2013) provides additional detail on estimating the cost weights and parameter values.

The formulation also relies on several parameters: the first parameter, R_{gl}, is the proportion of the area of origin g served by the transit line l. It is computed using the ratio of the intersection of the transit line buffer area and the origin block group to the total area of the origin block group. A 1/4 mile buffer around bus transit lines is used for calculating spatial coverage. The second parameter, γ_{il}, is a parameter that signifies whether bus stop, i, is served by a bus line l. The parameter takes the value "1" if a bus stop node i is served by bus line l, and 0 otherwise.

A binary connectivity parameter, δ_{ghl}, is calculated to measure access to possible trip destinations from a trip origin. This connectivity parameter represents the presence or absence of a direct transit line between O–D pairs. Thus, the connectivity parameter, δ_{ghl}, takes the value "1" if a transit line l directly connects origin g to destination h, and 0 otherwise. The parameter δ_{ghl} by itself assigns the same level of connectivity to all destinations regardless of the trip distance. To address this shortcoming, a connectivity decay factor, f_{ghl}, is calculated using a logistic decay function. This logistic decay function (an inverted s-curve) estimates the decreasing connectivity

based on door-to-door travel time. Mamun et al. (*17*) provides additional detail on estimating the connectivity decay factor.

A binary direction parameter, μ_{igh}, is used to distinguish in-bound from out-bound stops and for assigning demand to the associated directional stops. It takes the value 1 if stop *i* serves demand from origin *g* to destination *h*, and 0 otherwise.

The peak hour demand (a_{gh}) is demand for travel from centroid *g* to centroid h. In this study, demand was gathered from the local planning model and distributed to the block groups from the original traffic analysis zone (TAZ) level proportional to the block group populations.

4.2.1. Formulation

$$minimize\ Z = C_a \sum_i \sum_g \sum_h \frac{d_{ig}}{v_{walk}} x_{igh} \rho_{igh} + C_u \sum_g \sum_h \left(a_{gh} - \sum_i x_{igh} \right)$$

$$+ C_t \sum_i \sum_g \sum_h \sum_l a_{gh}\, \mu_{igh} y_i \gamma_{il} \frac{D}{3600}$$

$$- b_c \sum_i \sum_g \sum_h \sum_l \frac{B_l U}{p_g} R_{gl}\, \gamma_{il}\, \mu_{igh} \delta_{ghl}\, f_{ghl} \qquad (1)$$

The objective function is defined in Eq. 1, which seeks to minimize the cost of access, the unserved demand, the dwell time cost for each additional stop, and maximize connectivity benefits. The first component concentrates on walking time to stops. The second cost component penalizes the demand that is left unserved by the bus system. The third cost component applies an hourly in-vehicle stopping time cost per passenger for each additional bus stop and identifies the savings/benefits in travel time with the optimal number of bus stops. The last component, connectivity benefit, considers the savings in wait time for better connectivity with an additional bus run on a particular line. The four objectives can also be viewed as components of two larger

competing objectives: (a) minimizing the cost of access, and (b) providing high connectivity throughout the network. In this paper, the cost of walk access and unserved demand are elements of access while the dwell and wait time pieces are part of connectivity.

The objective function is constrained by the following:

$$\sum_{i} x_{igh} \leq a_{gh} \quad \forall g \in G, \forall h \in H \tag{2}$$

$$\sum_{g} \sum_{h} x_{igh} \leq b \; y_i \quad \forall i \in I \tag{3}$$

$$\sum_{g} \sum_{h} x_{igh} \leq \sum_{l} B_l \, \gamma_{il} U \quad \forall i \in I \tag{4}$$

Constraint (2) limits the total allocation of passengers for each O-D pair to the trip demand for that pair. Constraint (3) restricts the demand served at a stop based on stop capacity. A default value of 40 passengers boarding per stop, *b,* is used in this paper (*26*). Constraint (3) is dominant when a transit stop is served by low frequency transit lines, whereas constraint (4) dominates when high frequency transit serves a stop. Although constraints (3) and (4) capture elements of stop capacity, they are needed for accommodating differing line frequencies.

$$\sum_{i} \sum_{g} \sum_{h} x_{igh} \, \gamma_{il} \, \mu_{igh} \delta_{ghl} (1-\theta) \leq B_l \, U \quad \forall l \in L \tag{5}$$

$$\frac{B_l}{N} \geq \frac{\sum_i \sum_g \sum_h x_{igh} \mu_{igh} \delta_{ghl}}{\sum_g \sum_h a_{gh}} \quad \forall l \in L \tag{6}$$

$$\sum_{l} B_l \leq N \tag{7}$$

Constraint (5) places the capacity constraint for each line. It constrains the total number of passengers using a bus line to the total number of available seats. In this constraint, it is

assumed that certain proportions of passengers (θ) alight at each stop. Constraint (6) requires the proportional distribution of available buses to transit lines according to the demand along that line. Constraint (7) limits the numbers of buses assigned to the available buses for the bus system.

$$y_i \in \{0,1\} \quad \forall i \in I \tag{8}$$

$$x_{igh} \in Z^+ \quad \forall i \in I, \forall g \in G, \forall h \in H \tag{9}$$

$$B_l \in Z^+ \quad \forall l \in L \tag{10}$$

The final three constraints are definitional; restricting y_i to a binary variable, B_l and x_{igh} to positive integer values.

4.3. Feature Selection Algorithm

Feature selection is defined as the process of selecting a subset of the most relevant features from a set of features, which involves discarding the irrelevant, redundant and noisy features. Feature selection algorithms are useful search algorithms as they provide shorter running time by eliminating irrelevant and redundant features. A generic feature selection algorithm is considered as the combination of search techniques for: (1) Selection of candidate subsets and; (2) Evaluation of selected/generated subset. Feature selection algorithms differ in the way the candidate subset are generated and in the evaluation criteria used to evaluate the candidate subset.

4.3.1. Selection of Candidate Subset

Subset selection begins with an initial subset that could be empty, the entire set of features, or some randomly chosen features. This initial subset can be changed in a number of ways. In

forward selection strategy, features are added one at a time. In backward selection the least important feature is removed based on some evaluation criterion (*19*). Random search strategy randomly adds, removes or swaps features to avoid being trapped in local optima.

Narendra and Fukunaga (*22*) developed a feature subset selection algorithm based on branch and bound algorithm, which selects the best subset by maintaining and traversing a tree, but stops the search along a particular branch if a predefined boundary value is exceeded. Greedy hill climbing is one of many popular search approaches, which iteratively evaluates a candidate subset of features, then modifies the subset and evaluates if the new subset is an improvement over the old (*23*). Sequential forward search (SFS), sequential backward search (SBS), and bidirectional search are some variations to the greedy hill climbing method (*18, 23, 24*). Some feature selection algorithms randomly pick subsets of features from the feature space by following some probabilistic steps and sampling procedures. Examples include evolutionary algorithms (EA), and simulated annealing (SA) that considers randomness in subset selection in order to avoid of getting trapped in local minima or maxima.

4.3.2. Evaluation of the Candidate Subset

After selecting the subsets from the original set of features, they are evaluated using the objective function. Existing subset selection algorithms are traditionally categorized as wrapper or filter methods (*20*). Wrappers use a search algorithm to search through the space of possible features and evaluate each subset by running a model on the subset. Although it is a computationally expensive procedure, wrapper method can find the subsets from the feature space with a high accuracy. Filter methods are computationally more efficient than wrapper as they evaluate the accuracy of a subset of features using only the objective criteria that can be tested quickly (*3*).

4.4. The Proposed Algorithm

In this paper, a randomized feature selection algorithm is proposed for solving the bus stop location problem. The algorithm selects a predetermined candidate subset of stops randomly and the candidate subset of stops is then evaluated by solving the restricted BSLP model, where the total number of stops is predetermined, using CBC (COIN-OR Branch and Cut) solver from General Algebraic Modeling System (GAMS) software. This proposed scheme falls into the class of wrapper method where the objective value in each step is determined by solving BSLP model for each subset. Next, a neighborhood subset of stops is generated and evaluated again by running the model. In this randomized feature selection algorithm, if the current subset of stops yields a better value for the objective function, the previous best solution is replaced with the current one. If not, the next candidate subset is generated. This process iterates over the search space until a stopping condition is satisfied. Finally, the best subset is validated by incorporating prior knowledge.

Figure 4.1 illustrates detail of the proposed algorithm. Consider the space of all possible subsets of bus stops. This algorithm starts with a random point I' (i.e., a random subset of the stops) in this space and calculates the objective value Z' corresponding to this subset. An unbiased three sided coin is flipped with possible outcome sides 1, 2, and 3. If the outcome of the coin flip is 1, the algorithm choose a random neighbor I'' of this point by removing one stop from I' and adding a new stop to I'. After choosing I'', the new objective value Z'' is computed. If $Z'' < Z'$ for this minimization problem then move to the point I'' and proceed with the search from I''. On the other hand, if $Z'' > Z'$, then stay with point I' (with some probability p) or move to point I'' with probability $(1-p)$. This step is done to ensure that search does not get stuck in a local minimum. If the outcome of the coin flip is 2, choose a random neighbor I'' by removing

one feature from *I'* and compute its objective value *Z''*. Consider the last case where the outcome of the coin flip is 3. A random neighbor *I''* is selected by adding one stop to *I'* and compute its objective value *Z''*. The search proceeds until no significant improvement in objective value is obtained. Stopping criteria can be set as a non-improving objective value, ε,between two consecutive iterations. . The algorithm shown in Figure 1 is coded in Java and this code is provided in Appendix D.

Input: The set ***I*** of all possible stops and BSLP model

Output: A near optimal subset ***I'*** of stops

Step 1: *(Initialization)*

Randomly select a subset ***I'*** of stops from ***I***

Run the BSLP model using subset ***I'***

Compute objective function value ***Z'***

Step 2: *(Neighborhood Generation using 3 Moves)*

Flip an unbiased three sided coin with sides 1, 2, and 3.

Step 2.1: If the outcome of the coin flip is 1 *(Swap Move)*

Choose a random stop ***i*** from ***I – I'*** and add it to ***I'***

Remove a random stop ***i'*** from ***I'*** to get ***I"***

Step 2.2: If the outcome of the coin flip is 2 *(Add Move)*

Choose a random stop ***i*** from ***I – I'*** and add it to ***I'*** to get ***I"***

Step 2.3: If the outcome of the coin flip is 3 *(Delete Move)*

Remove a random stop ***i'*** from ***I'*** to get ***I"***

Run the BSLP model using subset ***I"***

Compute objective function value ***Z"***

Step 3: *(Evaluation)*

Calculate Δ**Z** = |***Z' – Z"***|

If Δ**Z** $\neq$ **0;** Then go to Step 4, otherwise, go to Step 5.

Step 4: *(Updating current solution)*

If **Z"< Z'**, set ***I' := I"*** and ***Z' := Z"***; perform the search from ***I'***

If **Z"> Z'**; perform the search from ***I'*** with probability ***p*** and perform the search from ***I''*** with ***Z' := Z"*** with probability ***1 – p***

Step 5: *(Check stopping criteria)*

If Δ**Z** $\leq$ **ε**; Stop and go to Step 6; otherwise, go to Step 2.

Step 6: *(Termination)*

Output ***I'*** is the best solution found, and the corresponding objective value is ***Z'***.

Figure 4.1. Randomized feature selection algorithm for BSLP

4.5. Results and Discussions

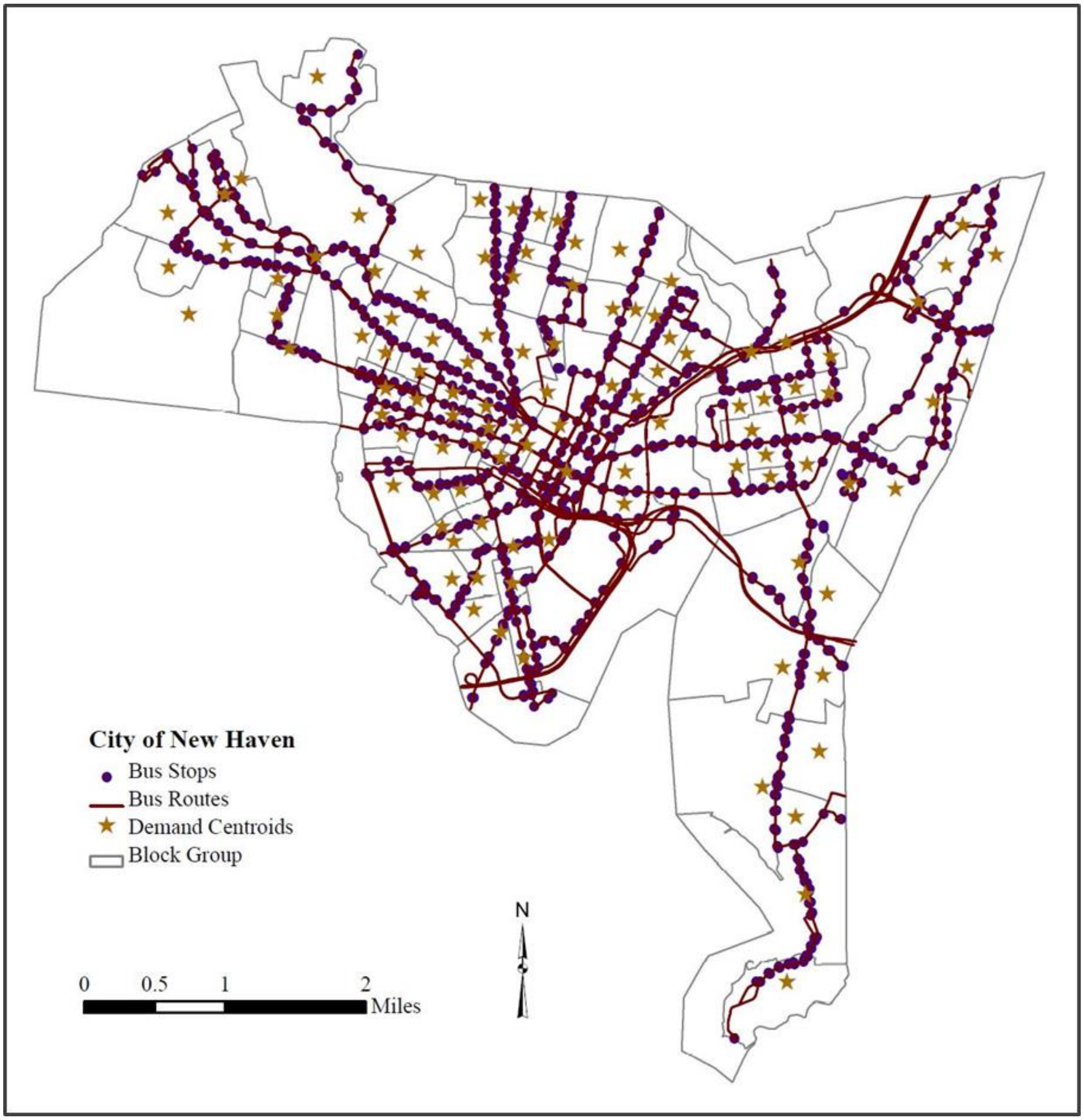

Figure 4.2. Demand centroids, bus stop locations and routes, city of New Haven, CT.

4.5.1. Application Network

The proposed heuristic procedure is applied to the BSLP model for the bus network of New Haven, Connecticut. The population of this city is 130,000, as reported in the 2010 census. The city is comprised 106 block groups (Figure 4.2) and the average block group size is 0.2 square

miles. Public transit in this city includes bus and rail service, however, the bus system is the dominant local service with 981 stops in service and 19 bus lines (Figure 4.2) for local trips within the city. The BSLP model seeks to select the best number and location of bus stops in order to increase service efficiency and provide convenient access to the service. The objective of this paper is to apply the FS algorithm to the BSLP model for obtaining near optimal solutions efficienciently.

4.5.2. Computational Performance and Efficiency

To evaluate the quality of the obtained solutions, the results have been compared to the best known values reported in chapter 3. The analysis in this paper was carried out on a Pentium 4 personal computer with 32GB memory. General Algebraic Modeling System (GAMS) software is used for modeling the bus stop location problem (BSLP) and CBC (COIN-OR Branch and Cut) solver was used to obtain the optimal solutions reported in Mamun and Lownes (*21*) (i.e., chapter 3).

The characteristics of the BSLP are extensive and complex due to its multi-objective decision making nature and the variety of parameters and data involved in the model formulations. The solution of the BSLP depends upon the chosen cost weights of the objective function and different parameters used in constraint equations. The BSLP optimal solutions using CBC solver for three different cases were presented in chapter 3 and are considered in this paper for the performance evaluation of the proposed FS algorithm. Table 4.1 presents the BSLP solutions obtained by FS algorithm and the CBC solver. The stopping criteria is $\varepsilon = 0.01$. The probability, $p = 0.9$, provides a means to escape local minima by allowing moves that worsen the objective function value with probability, $1 - p = 0.1$.

Table 4.1. Randomized FS algorithm solution comparison for the BSLP.

Cases		Solution by CBC Solver	Solution by Randomized FS Algorithm	Gap (%)
Case 1 (Accessibility-focused) **Ca=40, Cu=30,** Ct=30, bc=1	Objective value, Z (in thousand $/hour)	369	401	8.7
	Total no. of stops	812	856	-
	Execution time (CPU seconds)	49,736	106	-
	No. of iterations required	451461	1534	-
Case 2 (Default) Ca=20, Cu=10, Ct=30, bc=1	Objective value, Z (in thousand $/hour)	374	392	4.9
	Total no. of stops	776	812	-
	Execution time (CPU seconds)	52,723	117	-
	No. of iterations required	476732	1702	-
Case 3 (Connectivity-focused) Ca=20, Cu=10, **Ct=50**, **bc=10**	Objective value, Z (in thousand $/hour)	371	397	7.0
	Total no. of stops	747	807	-
	Execution time (CPU seconds)	53,674	168	-
	No. of iterations required	479239	2105	-

In comparison of the solutions generated by FS algorithm to the optimal solutions generated by CBC solver, it is clear that the difference is not great. The maximum percent gap of the best solution generated with the proposed approach from the optimal solution is reported 8.7% (case 1 in Table 1). Moreover, the execution time and the number of iterations required are significantly lower to obtain good solutions.

4.5.3. Tradeoff Analysis

This section presents the application of the BSLP and the tradeoff analysis for the large bus network of New Haven, CT. This analysis was not possible using exact methods because of the computational time required. The tradeoff analysis has two steps: generating dominated solution cloud with the application of a FS algorithm and generating a "Pareto-frontier" using the same pareto frontier generation algorithm used in Mamun and Lownes (*21*). In order to generate the solution cloud, this analysis used 1500 combinations of six parameters, which were generated based on prior application results and sensitivity analyses. The sensitivity analyses for this BSLP formulation are provided in chapter 3. The point cloud (Figure 3) captures solutions for all possible combinations of parameters within the following ranges:

- Cost of Access *(C_a)*: 15 – 40 $/hour in increment of 5
- Cost of unserved demand *(C_u)*: 10 – 30 $/hour in increment of 2
- Cost of in-vehicle travel time *(C_t)*: 20 – 50 $/hour in increment of 5
- Benefit of connectivity *(b_c)*: 1 – 10 $/hour in increment of 0.5
- Fleet size *(N)*: 50 – 75 buses in increment of 5
- Proportion of trips alighting *(θ)*: 0 – 1.0 in increment of 0.1

The objective of this tradeoff analysis is to establish a set of "non-dominated/pareto-optimal" solutions that create a boundary of best solutions for a range of weights on the multiple objectives. The tradeoff solution space[4] between the two primary objectives for the bus service system of New Haven is shown in Figure 4.3. The y-axis (Access Cost) in Figure 4.3 presents

[4] This tradeoff graph shows the solution space for this problem only.

the total of access cost and unserved cost, and the x-axis shows the reduction in connectivity benefits Connectivity benefit reduction is calculated by computing the maximum instance of total in-vehicle stopping time and wait time savings for the network. Then, the connectivity benefit value for a solution point is subtracted from the maximum value. The non-dominated solutions (in Figure 3) are those along the lower boundary of the cloud and the frontier constructed with these pareto optimal solutions provides the trade-off between access cost minimization and connectivity benefit maximization for any combination of parameters within the ranges specified.

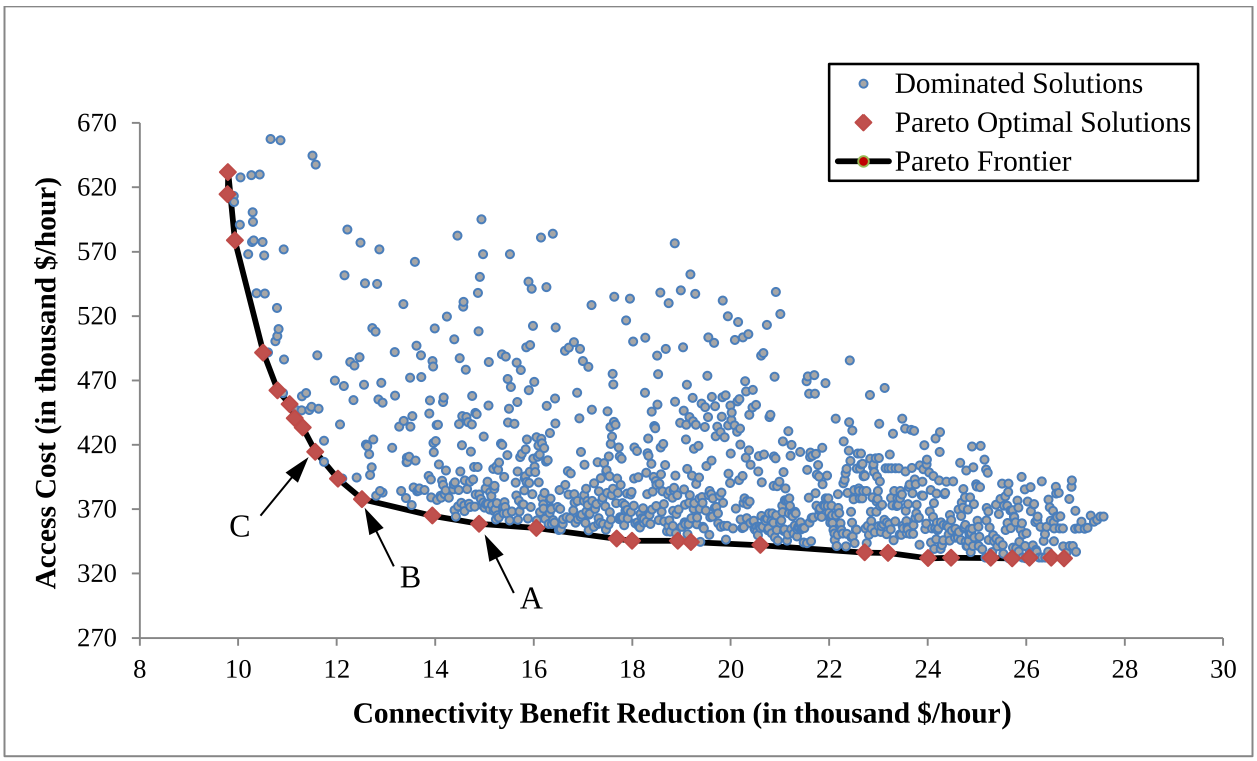

Figure 4.3. Access and connectivity tradeoff for the BSLP *(New Haven, CT)*.

Table 4.2. Solution points on Pareto frontier for New Haven, CT.

	Solution Point "A" (Accessibility-focused)	Solution Point "B"	Solution Point "C" (Connectivity-focused)
Cost Parameters ($/hour)	**Ca=37, Cu=26,** Ct=29, bc=2	C_a=19, C_u=10, C_t=28, b_c=1	C_a=18, C_u=12, **C_t=46, b_c=9**
Objective value, Z *(in thousand $/hour)*	338	364.5	390.5
Access cost *(in thousand $/hour)*	358	378	414
Connectivity benefit *(in thousand $/hour)*	20	22.5	23.5
Connectivity benefit reduction *(in thousand $/hour)*	15[a]	12.5	11.5
Total number of stops *(% stop reduction)*	841*(14.3)*	787*(19.8)*	796*(18.8)*

[a] The maximum connectivity benefit for this network was found as $35 thousand/hour, therefore the reduction for this solution point from the maximum is, 35-20= $15 thousand/hr.

Three solution points (Table 4.2) are identified on the "Pareto frontier" to distinguish the tradeoffs associated with access cost and connectivity benefit reductions. Point "A" presents the pareto optimal solution focused on accessibility. The higher cost weights C_a and C_u result in ahigher number of stops (Table 4.2) so that spatial access can be improved through minimizing walk distance and serving more demand. This solution point results in lower access cost and lower connectivity benefit (i.e., higher value of connectivity reduction as shown in Table 4.2) due to frequent stops. Point "C" describes the connectivity-focused (i.e., higher weights on b_c and C_t) solution which selects fewer stops which benefit from savings in in-vehicle stopping time and wait time. This fewer number of stops corresponds to higher access cost due to longer distance walks. Solution point "B" on the pareto frontier can be described as a neutral solution

point which balances between accessibility and connectivity. This point can also be described as the pivotal point from where the solutions to the right or left sacrifice either access or connectivity to improve the other. To summarize, for small values of access cost, higher number of stops are favored while for larger values of connectivity benefit, lower number of stops are favored.

4.6. Conclusions

This paper presented a randomized search technique for the bus stop location problem. The proposed feature selection algorithm for selecting a subset of the most relevant stops from the set of all possible stops is a wrapper method where the prediction accuracy in each step is determined by objective value obtained from a branch and cut algorithm. To demonstrate the validity of the proposed approach, it is applied to the bus system of New Haven, CT, which demonstrated that FS heuristic procedure can find good solutions using much less computational time. The maximum percent gap of the best solution generated with the proposed approach from the optimal solution is reported 8.7%. Moreover, the execution time and the number of iterations required are lower by several orders of magnitude. The tradeoff analysis between access and connectivity shows that the increase in access cost resulted due to the reduction in the total number of bus stops. However, the increased access cost can be balanced by the increased connectivity benefits due to the savings in in-vehicle stopping time and wait time with fewer stops.

In order to increase the quality of solutions, focusing on generating a good quality initial solution is an important future direction of this research. Other options to improve the performance of this algorithm include incorporating a probability function to override local

optima and the use of a more sophisticated probability function for neighborhood generation moves rather than using fixed probability values. Future research also includes application of this algorithm to other types of facility location problems, such as uncapaciated facility location problems and p-median problem. Finally, a comparison of the FS generated results to other existing heuristics (e.g., Tabu search, Simulated Annealing, etc.) can explore the evidence for FS as an important heuristic for BSLP and other facility location problems.

Acknowledgments

The authors wish to thank Subrata Saha, graduate student from Computer Science and Engineering department, University of Connecticut, for helping in Java code for the randomized feature selection algorithm.

References

1. Daskin, M., 1995. Network and Discrete Location, Models, Algorithms, and Applications, Wiley, New York.
2. Drezner, Z., and Hamacher, H., 2001. Facility Location: Theory and Algorithms, Springer, Berlin.
3. Sun, Y., Todorovic, S., and Goodison, S. A Feature Selection Algorithm Capable of Handling Extremely Large Data Dimensionality in 'SDM', SIAM, 2008, pp. 530-540.
4. Cornuejols, G., Sridharan, R., and Thizy, J. A comparison of heuristics and relaxations for the capacitated plant location problem. *European Journal of Operational Research,* Vol. 50, No. 3, 1991, pp. 280–297.
5. Beasley, J. Lagrangian heuristics for location problems. *European Journal of Operational Research,* Vol. 65, 1993, pp. 383–399.
6. Barahona, F., and Chudak, F. Near-optimal solutions to large scale facility location problems. *Discrete Optimization,* Vol. 2, No. 1, 2005, pp. 35–50.

7. Jacobsen, S. Heuristics for the capacitated plant location model. *European Journal of Operational Research,* Vol. 12, 1983, pp. 253–261.

8. Domschke, W., and Drexl, A., 1985. Add-heuristics starting procedures for capacitated plant location models. *European Journal of Operational Research,* Vol. 21, 1985, pp. 47–53.

9. Delmaire, H., Diaz, J., Fernandez, E., and Ortega, M. Reactive GRASP and tabu search based heuristics for the single source capacitated plant location problem. *Information Systems and Operations Research,* Vol. 37, No. 3, 1998, pp. 194–225.

10. Filho, V., and Galvão, R. A tabu search heuristic for the concentrator location problem. *Location Science,* Vol. 6, 1998, pp. 189–209.

11. França, P., Sosa, N., and Pureza, V. An adaptive tabu search approach for solving the capacitated clustering problem. *International Transactions in Operational Research,* Vol. 6, 1999, pp. 665–678.

12. Tuzun, D., and Burke, L. A two-phase tabu search approach to the location routing problem. *European Journal of Operational Research,* Vol. 116, No. 1, 1999, pp. 87–99.

13. Al-Sultan, K., and Al-Fawzan, M. A Tabu search approach to the uncapacitated facility location problem. *Annals of Operations Research*, Vol. 86, 1999, pp. 91–103.

14. Ghosh, D. Neighborhood search heuristics for the uncapacitated facility location problem. *European Journal of Operational Research*, Vol. 150, 2003, pp. 150–162.

15. Michel, L., and Van Hentenryck, P. A simple tabu search for warehouse location. *European Journal of Operational Research,* Vol. 157, 2004, pp. 576–591.

16. Sun, M. A tabu search heuristic procedure for the capacitated facility location problem. *Journal of Heuristics,* Vol. 18, No. 1, 2012, pp. 91-118.

17. Mamun, S., Lownes, N., Osleeb, J., and Bertolaccini, K. A method to define public transit opportunity space. *Journal of Transport Geography,* Vol. 28, 2013, pp. 144-154.

18. Ladha, L., and Deepa, T. Feature selection methods and algorithms. *International Journal on Computer Science and Engineering,* Vol. 3, 2011, pp. 1787-1797.

19. Pudil, P., and Novovicova, J. Novel methods for subset selection with respect to problem knowledge. *IEEE Intelligent System,* Vol. 13, 1998, pp. 66–74.

20. Kohavi, R., and John, G. Wrappers for feature subset selection. *Artificial Intelligence*, Vol. 97, 1997, pp. 273–324.

21. Mamun, S., and Lownes, N. Access and Connectivity Tradeoffs in Transit Stop Location. Accepted for presentation at *the 93rd Annual Meeting of the Transportation Research Board and publication in Transportation Research Record, 2013*.

22. Narendra, P., and Fukunaga, K. Branch and Bound Algorithm for Feature Subset Selection, *IEEE Trans. Computer*, Vol. 26, No. 9, 1977, pp. 917-922.

23. Caruana, R., and Freitag, D. Greedy attribute selection. *Proceedings 11th International Conference on Machine Learning,* New Brunswick, NJ, Morgan Kaufmann, San Mateo, CA, 1994, pp. 28–36.

24. Marcano-Cedeño, A., Quintanilla-dominguez, J., Cortina-Januchs, M. G., Andina, D. Feature selection using Sequential Forward Selection and classification applying Artificial Metaplasticity Neural Network. *IECON 2010 - 36th Annual Conference on IEEE Industrial Electronics Society,* 2010, pp. 2845-2850.

CHAPTER 5: CONCLUSIONS AND FUTURE RESEARCH

5.1. Conclusions

The bus stop location model (BSLP) has been developed for locating stops in an existing public transit network – allowing for a reduction or reallocation of stops in the network to optimally balance the accessibility of the system with its connectivity. This model seeks to select the best number and location of stops in order to increase service efficiency and provide convenient access to the system. The model considers both access and system connectivity. A mixed-integer multi-objective optimization approach is presented to identify optimal stop locations, assign demand to each stop to all O-D trip pairs, and set vehicle frequencies for all bus lines. For access, walking distance was modeled as one of the determinant for the location of stops as frequent stops have been considered as to provide greater access to stops from trip origins. For system connectivity, savings in travel time (i.e., dwell time and wait time at stops) due to fewer stops along a route was modeled. In order to measure the system connectivity, a transit opportunity index was formulated. This index provides the ability to explore the public transit opportunity between a single O-D pair, as well as the overall public transit opportunity from or to a specific origin or destination.

CBC, a commercial solver, was used to generate optimal solutions for the BSLP. It is found that the optimal number and locations of the stops will be affected by users' access cost and system connectivity. The model results suggest that wait time and dwell time savings at stops could be realized with adjustments to vehicle frequencies while still maintaining good system access. From a bus rider's perspective, the results indicate that the increase in access distance from stop consolidation is offset by the wait time and dwell time reductions, which thus improve the connectivity of the system. The tradeoff analysis between access and connectivity

shows that average access cost is increased to some extent due to a reduction in the total number of bus stops but this is balanced by the improvements in service efficiency and network connectivity.

Although CBC solver generated optimal solutions, its limitations of higher run time were noted. Therefore, to speed computation a randomized feature selection algorithm was proposed to solve the BSLP. Computational experiments have been performed on the bus network for City of New Haven. The behavior of these two algorithms has been compared in terms of solution quality and computational efficiency. Result showed that the proposed randomized feature selection algorithm returned near optimal solutions in a much faster time frame compared to Branch and Cut algorithm used in CBC solver. Due to low complexity and simple implementation of the proposed FS algorithm it is believed that the proposed algorithm may have potential for solving other facility location problem (i.e., the uncapaciated facility location problems, p-median problem).

The Transit Opportunity Index provides a robust tool for measuring public transit service performance. There is less literature considering trip coverage, or the transit connectivity of origins and destinations, where destinations are the aggregation of the multiple opportunities within a defined geographic zone. The TOI developed in this paper allows us to quantify the level of access provided by the service and the ability of this service to make trips from origin to destination. Incorporation of TOI in the bus stop location problem for modeling service efficiency is one of the distinguishing contributions of this research. This model gives the transit agencies decision making tools to decide on which redundant stops to eliminate and which strategic stops to maintain based on access distance and system connectivity. The BSLP model provides an alternative approach to model transit service efficiency and redundancy.

5.2. Future Research

Future research is needed to develop a more robust stop location problem to identify stop redundancy. For measuring access, only demand weighted walk/access distance was considered. One potential improvement to model access is addition of pedestrian route connectivity to stop for addressing the walking accessibility to transit service. The dwell time savings per stop in the BSLP objective function can be improved by considering more sophisticated formulation for the dwell time calculation based on passengers' boarding/alighting per stop. The connectivity component used in objective function measures the connectivity benefit considering savings in wait time for each additional bus available (increased frequency) per hour. An important extension to this objective can be considered to model the wait time savings at stop for the served demands. However, this extension can result a non-linear formulation of the BSLP. Another direction of improvement in BSLP formulation might be modeling other attributes at stop that could impact the selection of stops such as handicap access, bike racks, shelters, service maps, public phones at stops.

In regard to solution approach, several future research directions can be considered. Although randomized feature selection algorithm provided solutions in a faster time frame, more research is needed to understand the behavior of the algorithm with problem instances of different sizes. One could improve the performance of the proposed heuristic for getting a good initial solution. A non-linear mixed integer programming formulation for the BSLP is another possible direction for evaluating the proposed heuristics.

APPENDICES

Appendix A. GAMS code for the BSLP

This appendix provides GAMS code used for the mathematical formulation of the bus stop location problem used in chapter 2.

```
GAMS Rev 149 Copyright (C) 1987-2007 GAMS Developments.  All rights reserved
Licensee: Civil and Environmental Engineering           G120817:1356AP-WIN
          University of Connecticut                                 DC6887
          License for teaching and research at degree granting institutions

OPTIONS lp=coincbc, mip=coincbc;

set i intersections or candidate bus stops /
268,272,273,279,280,286,301,304,309,313,314,319,320,334,336,337,
341,348,355,365,383,387,391,394,441,444,449,453,481,493,499,507,
514,545,547,591,592,611,615,636,650,674,681,708,722,728,748,754,
755,759,772,777,780,786,794,800,801,813,814,823,826,843,844,845,
853,854,870,873,882,883,885,887,888,896,898,902,909,911,913,919,
923,924,926,936,945/
set g demand centroids / 1,47,48,49,50,51,52,54,55,56,57,58,59,60,61 /
set h demand centroids / 1,47,48,49,50,51,52,54,55,56,57,58,59,60,61 /
set l transit lines / 6, 9 /

Parameters population(g) number of population at demand centroid g in
cases
/         1         99
         47         81
         48         72
         49         86
         50         75
         51         59
         52         81
         54         108
         55         125
         56         104
         57         79
         58         49
         59         60
         60         46
         61         33 /;

$include gamma.txt
$include coverage.txt
$include mu.txt
$include delta.txt
$include demand.txt
$include distance.txt
```

```
Parameter rho(i,g,h) indicates walk coverage i to g in route;
          rho(i,g,h)$(mu(i,g,h) ge 1)=mu(i,g,h);
          rho(i,g,h)$(mu(i,g,h) le 0)=100000000000000;

scalars
        accesscost      usd in hourly weight                    /20/
        unservedcost    usd in hourly weight                    /10/
        ivttcost        usd in hourly weight                    /30/
        connectbenefit  usd in hourly weight                    /1/
        beta            available buses                         /9/
        theta           proportion of people alighting          /0.5/
        delay           in seconds                              /30/
        walkspeed       walking speed in feet per hour          /14400/
        capacity        in number                               /35/

;

variables
          z           objective variable
          y(i)        decision variable
          x(i,g,h)    decision variable
          B(l)        decision variable
          w           total access cost
          u           total unserved cost
          t           total invtt delay cost
          c           total wait cost savings
;

binary variable y;
integer variable x;
integer variable B;
positive variable w;
positive variable u;
positive variable t;
positive variable c;

equations
          cost                     define objective function
          demandeq(g,h)            demand need to satisfied
          supply(i)                stops needed to serve demand
          stopbudget               total number of stops need to select
          timeeq                   total invtt delay cost
          access                   total access cost
          connectivity             total connectivity benefit
          unserved                 total unserved cost
          stop                     stop constraint
          line                     line constraint
          budget                   budget constraint
          vehicle(l)               vehicle constraint

;
```

```
*objective function*

cost.. z =e= w + u + t + c;

access..w =e= accesscost* sum((i,g,h),((distance(i,g)/walkspeed)
*x(i,g,h)*rho(i,g,h)));

unserved..u =e= unservedcost* sum((g,h),(demand(g,h)-
sum(i,x(i,g,h))));

timeeq..t =e=ivttcost*
sum((i,g,h,l),mu(i,g,h)*y(i)*gamma(i,l)*delay/3600);

connectivity..c =e= connectbenefit*
sum((i,g,h,l),(coverage(g,l)*gamma(i,l)*mu(i,g,h)*delta(g,h,l)*y(i)*(c
apacity/population(g))));

*Constraint Equations*

demandeq(g,h).. sum(i,(x(i,g,h))) =l= demand(g,h);

supply(i).. sum((g,h),(x(i,g,h))) =l= 40*y(i);

stop(i).. sum((g,h),x(i,g,h)*mu(i,g,h)) =l=
sum(l,B(l)*gamma(i,l)*capacity);

line(l).. sum((i,g,h),x(i,g,h)*mu(i,g,h)*gamma(i,l)*delta(g,h,l))* (1-
theta) =l= B(l)*capacity;

vehicle(l).. B(l)/beta =g=
(sum((i,g,h),x(i,g,h)*mu(i,g,h)*delta(g,h,l)))/(sum((g,h),demand(g,h))
);

budget.. sum(l,B(l)) =l= beta;


model stopmodel / cost, demandeq, supply, stop, line, vehicle, budget,
access, unserved, timeeq, connectivity/;

Option reslim = 10000000
Option optcr = 0.0

solve stopmodel using mip minimizing z;

*output*

display z.l, y.l, x.l, B.l, w.l, u.l, t.l, c.l;
```

Appendix B. Java code for solution space generation

In this appendix section Java code for the generation of solution space through multiple running of BSLP for different combinations of model parameters is provided.

```
/* Copyright 2013, University of Connecticut, Sha Mamun */
/* please use and modify this program as you like */

package utilities;

import java.io.BufferedReader;
import java.io.BufferedWriter;
import java.io.DataInputStream;
import java.io.FileInputStream;
import java.io.FileNotFoundException;
import java.io.FileWriter;
import java.io.IOException;
import java.io.InputStreamReader;
import java.text.DecimalFormat;

public class Utils {

    public static final int primaryStartIndexForAccessCost = 5;
    public static final int primaryStartIndexForUnservedCost = 5;
    public static final int primaryStartIndexForIvttCost = 20;
    public static final int primaryStartIndexForConnectBenefit = 1;
    public static final int primaryStartIndexForTheta = 0;
    public static final int primaryStartIndexForBeta = 5;
    public static final int primaryStartIndexForDelay = 30;

    public static final int finalIndexForAccessCost = 40;
    public static final int finalIndexForUnservedCost = 30;
    public static final int finalIndexForIvttCost = 50;
    public static final int finalIndexForConnectBenefit = 10;
    public static final int finalIndexForTheta = 1;
    public static final int finalIndexForBeta = 15;
    public static final int finalIndexForDelay = 30;

    public static final double stepSizeForAccessCost = 1;
    public static final double stepSizeForUnservedCost = 1;
    public static final double stepSizeForIvttCost = 2;
    public static final double stepSizeForConnectBenefit = 0.5;
    public static final double stepSizeForTheta = 0.05;
    public static final double stepSizeForBeta = 1;
    public static final double stepSizeForDelay = 1;
```

```
    public static String accessCost = "accesscost        usd in hourly
weight ";
    public static String unservedCost = "unservedcost    usd in hourly
weight ";
    public static String ivttCost = "ivttcost            usd in hourly
weight ";
    public static String connectBenefit= "connectbenefit usd in hourly
weight ";
    public static String theta = "theta                  proportion of
people alighting ";
    public static String beta = "beta                    available
buses ";
    public static String delay = "delay                  in seconds ";
    public static String walkSpeed = "walkspeed          walking speed
in feet per hour              /14400/\n";
    public static String capacity = "capacity            in number
/35/\n";
    public static final String filePathWithFileName =
"D:\\StopModel1\\connectmodel7.gms";
    public static final String appResultFilePathWithFilaName =
"D:\\StopModel1\\connectmodel7.lst";
    public static final String resultFilePathWithFileName =
"D:\\StopModel1\\connect_result.txt";
    public static String inputData = "";

    public static String getRestOfTheContent() throws
FileNotFoundException, IOException {

        FileInputStream fstream = new
FileInputStream(filePathWithFileName);
        DataInputStream in = new DataInputStream(fstream);
        BufferedReader br = new BufferedReader(new
InputStreamReader(in));
        String restOfTheContent = "";
        boolean canBePut = false;
        String strLine = "";
        while ((strLine = br.readLine()) != null) {
            if (canBePut == false && strLine.startsWith("variables"))
{
                restOfTheContent = "variables" + "\n\n";
                canBePut = true;
            } else if (canBePut) {
                restOfTheContent += strLine + "\n";
            }
        }
        in.close();

        return restOfTheContent;
    }

    public static String getInitialContent() throws
FileNotFoundException, IOException {
```

```
        FileInputStream fstream = new
FileInputStream(filePathWithFileName);
        DataInputStream in = new DataInputStream(fstream);
        BufferedReader br = new BufferedReader(new
InputStreamReader(in));
        String initialContent = "";
        String strLine = "";
        while ((strLine = br.readLine()) != null) {
            if (!strLine.startsWith("scalars")) {
                initialContent += strLine + "\n";
            } else {
                break;
            }
        }
        in.close();

        return initialContent;
    }

    public static void createNewFileWithModifiedData(double
accessCost, double unservedCost, double ivttCost, double
connectBenefit, double beta, double theta, double delay, String
initialConetnt, String restOfTheContent) throws FileNotFoundException,
IOException {

        String primaryPart = "scalars\n\n";

        String tempAccessCost = Utils.accessCost + "   /" +
String.valueOf(accessCost) + "/" + "\n";
        String tempUnservedCost = Utils.unservedCost + "   /" +
String.valueOf(unservedCost) + "/" + "\n";
        String tempIvttCost = Utils.ivttCost + "   /" +
String.valueOf(ivttCost) + "/" + "\n";
        String tempConnectBenefit = Utils.connectBenefit + "   /" +
String.valueOf(connectBenefit) + "/" + "\n";
        DecimalFormat df = new DecimalFormat("#.###");
        String tempTheta = Utils.theta + "   /" + df.format(theta) +
"/" + "\n";
        String tempBeta = Utils.beta + "   /" + String.valueOf(beta) +
"/" + "\n";
        String tempDelay = Utils.delay + "   /" +
String.valueOf(delay) + "/" + "\n";

        inputData = tempAccessCost + tempUnservedCost + tempIvttCost +
tempConnectBenefit + tempTheta + tempBeta + tempDelay +
Utils.walkSpeed + Utils.capacity;
        String modifiedData = initialConetnt + primaryPart + inputData
+ restOfTheContent;

        BufferedWriter out = new BufferedWriter(new
FileWriter(filePathWithFileName));
```

```
        out.write(modifiedData);
        out.close();
    }

    public static void saveResultToFile(String result) throws
IOException {

        FileWriter fstream = new
FileWriter(resultFilePathWithFileName, true);
        BufferedWriter out = new BufferedWriter(fstream);
        out.write(result);
        out.close();
    }

    public static String getResult() throws FileNotFoundException,
IOException {

        FileInputStream fstream = new
FileInputStream(appResultFilePathWithFilaName);
        DataInputStream in = new DataInputStream(fstream);
        BufferedReader br = new BufferedReader(new
InputStreamReader(in));
        String strLine = "";
        String text = "";
        while ((strLine = br.readLine()) != null) {
            if (strLine.contains("VARIABLE w.L")) {
                String[] strings = strLine.split("\\s{1,}");
                text += strings[5] + ", ";
            } else if (strLine.contains("VARIABLE u.L")) {
                String[] strings = strLine.split("\\s{1,}");
                text += strings[4] + ", ";
            } else if (strLine.contains("VARIABLE t.L")) {
                String[] strings = strLine.split("\\s{1,}");
                text += strings[4] + ", ";
            } else if (strLine.contains("VARIABLE c.L")) {
                String[] strings = strLine.split("\\s{1,}");
                text += strings[4] + "\n";
            }
        }
        in.close();
        System.out.println(text);
        return text;
    }
}
```

Appendix C. Pareto-frontier generation algorithm

The Pareto-frontier generation algorithm is provided in this appendix section.

```
/* Copyright 2013, University of Connecticut, Sha Mamun */
/* please use and modify this program as you like */

namespace FrontierGeneration
{
    partial class Form1
    {
        /// <summary>
        /// Required designer variable.
        /// </summary>
        private System.ComponentModel.IContainer components = null;

        /// <summary>
        /// Clean up any resources being used.
        /// </summary>
        /// <param name="disposing">true if managed resources should
be disposed; otherwise, false.</param>
        protected override void Dispose(bool disposing)
        {
            if (disposing && (components != null))
            {
                components.Dispose();
            }
            base.Dispose(disposing);
        }

        #region Windows Form Designer generated code

        /// <summary>
        /// Required method for Designer support - do not modify
        /// the contents of this method with the code editor.
        /// </summary>
        private void InitializeComponent()
        {
            this.buttonCreatePoints = new
System.Windows.Forms.Button();
            this.SuspendLayout();
            //
            // buttonCreatePoints
            //
            this.buttonCreatePoints.Location = new
System.Drawing.Point(62, 31);
            this.buttonCreatePoints.Name = "buttonCreatePoints";
            this.buttonCreatePoints.Size = new
System.Drawing.Size(131, 23);
            this.buttonCreatePoints.TabIndex = 0;
```

```
            this.buttonCreatePoints.Text = "Create mapping points";
            this.buttonCreatePoints.UseVisualStyleBackColor = true;
            this.buttonCreatePoints.Click += new
System.EventHandler(this.buttonCreatePoints_Click);
            //
            // Form1
            //
            this.AutoScaleDimensions = new System.Drawing.SizeF(6F,
13F);
            this.AutoScaleMode =
System.Windows.Forms.AutoScaleMode.Font;
            this.ClientSize = new System.Drawing.Size(284, 261);
            this.Controls.Add(this.buttonCreatePoints);
            this.Name = "Form1";
            this.Text = "Form1";
            this.ResumeLayout(false);

        }

        #endregion

        private System.Windows.Forms.Button buttonCreatePoints;
    }
}
```

Appendix D. Java code for feature selection algorithm

This appendix section provides Java code for the randomized feature selection algorithm used in chapter 4 to generate the near optimal solutions.

```
/* Copyright 2013, University of Connecticut, Sha Mamun */
/* please use and modify this program as you like */


package civil;

import java.io.*;
import java.util.ArrayList;
import java.util.HashMap;
import java.util.List;
import java.util.Random;

public class RandomizedSearch {

    public String modelFileName = "modelEx.gms";
    public String resultFileName = "modModelEx.lst";
    public String terminalFileName = "bus_stopEx.txt";
    public String modifiedModelFileName = "modModelEx.gms";
    List<Integer> terminalList;
    String modelText = "";
    public HashMap hashMap;

    public RandomizedSearch() throws IOException {
        terminalList = getTerminalData(terminalFileName);
        hashMap = new HashMap();

        for (int i = 0; i < terminalList.size(); i++) {
            hashMap.put(new Integer(i), terminalList.get(i));
        }

        modelText = readModelFile(modelFileName);
    }

    public static void main(String[] args) throws IOException,
InterruptedException {
        double constant = 3.0;
        RandomizedSearch randomizedSearch = new RandomizedSearch();
        randomizedSearch.simulatedAnnealing(constant);
    }

    public double getResultFromGAMS(String modifiedModelFile, String
resultFileName, List<Integer> list) throws IOException {

        modifyModelFile(modifiedModelFile, list);
```

```
        String command = "C:\\GAMS\\win64\\23.9\\gams.exe
modModelEx";

        Runtime runtime = Runtime.getRuntime();
        Process process = runtime.exec(command);

        System.out.println("Leaving gams ...");

        try {
            process.waitFor();
        } catch (InterruptedException ex) {
            ex.printStackTrace();
        }

        return getScore(resultFileName);
    }

    public String readModelFile(String modelFileName) throws
IOException {
        FileInputStream fis = new FileInputStream(modelFileName);
        DataInputStream dataInputStream = new DataInputStream(fis);
        BufferedReader bufferedReader = new BufferedReader(new
InputStreamReader(dataInputStream));
        String textLine;
        String text = "";
        while ((textLine = bufferedReader.readLine()) != null) {
            text += textLine + "\n";
        }

        dataInputStream.close();
        return text;
    }

    public void modifyModelFile(String modifiedModelFileName,
List<Integer> list) throws IOException {

        String text = "OPTIONS lp=coincbc, mip=coincbc;" + "\n";
        text += "set i intersections or candidate bus stops /";
        String terminalList = "";

        for (int i = 0; i < list.size() - 1; i++) {
            if (hashMap.get(list.get(i)) != null) {
                terminalList += hashMap.get(list.get(i)) + ",";
            }
        }
        if (hashMap.get(list.size() - 1) != null) {
            terminalList += hashMap.get(list.size() - 1) + "/" + "\n";
        } else {
            terminalList += "/" + "\n";
        }
        String modifyText = text + terminalList + modelText;
```

```
        writeToFile(modifiedModelFileName, modifyText, false);
    }

    public void simulatedAnnealing(double constant) throws
IOException, InterruptedException {

        int numberOfFeatures = terminalList.size();
        List<Integer> finalList = new ArrayList<Integer>();

        int pickFeatures = generateRandomNumber(numberOfFeatures, 1) +
1;
        pickFeatures = 30;
        List<Integer> list = generateRandomNumbers(numberOfFeatures,
pickFeatures);
        double result = getResultFromGAMS(modifiedModelFileName,
resultFileName, list);
        System.out.println("FIRST ACCURACY: " + result + " SIZE: " +
list.size());

        double score = result;
        double tempScore = 0.0;
        double minScore = 999999;
        int size = 0;

        for (int i = 0; i < 500; i++) {

            int threeSidedCoin = generateRandomNumber(3, 0) + 1;
            if (threeSidedCoin == 1) {
                if (list.size() <= 5) {
                    System.out.println("GOING BACK...q = q");
                    continue;
                }
                // q = q
                int randomNumber = generateRandomNumber(list.size(),
0);
                int restRandomNumber =
generateRandomNumber(numberOfFeatures, list);
                int remove = list.get(randomNumber);
                list.remove(randomNumber);
                list.add(restRandomNumber);

                tempScore = getResultFromGAMS(modifiedModelFileName,
resultFileName, list);
                System.out.println("INTERIM ACCURACY: " + tempScore +
" SIZE: " + list.size());

                if (tempScore > score) {
                    double displacement = -constant *
Math.abs(tempScore - score);
                    double displacementProbability =
Math.exp(displacement);
```

```
                    displacementProbability = 0.1;

                    double rand = new Random().nextDouble();
                    if (rand > displacementProbability) {
                        list.remove((Integer) restRandomNumber);
                        list.add(remove);
                        tempScore = score;
                    } else {
                        // do nothing
                    }
                } else {
                    // do nothing
                }
            } else if (threeSidedCoin == 2) {
                // q = q + 1
                if (list.size() == numberOfFeatures) {
                    System.out.println("GOING BACK...q=q+1");
                    continue;
                }
                int randomFeature =
generateRandomNumber(numberOfFeatures, list);
                list.add(randomFeature);
                tempScore = getResultFromGAMS(modifiedModelFileName,
resultFileName, list);
                System.out.println("INTERIM ACCURACY: " + tempScore +
" SIZE: " + list.size());

                if (tempScore > score) {

                    double displacement = -constant *
Math.abs(tempScore - score);
                    double displacementProbability =
Math.exp(displacement);
                    displacementProbability = 0.1;

                    double rand = new Random().nextDouble();
                    if (rand > displacementProbability) {
                        boolean remove = list.remove((Integer)
randomFeature);

                        tempScore = score;
                    } else {
                        // do nothing
                    }
                } else {
                    // do nothing
                }

            } else if (threeSidedCoin == 3) {
                // q = q - 1
                int randomFeature = generateRandomNumber(list.size(),
0);
                int remove = list.get(randomFeature);
```

```
                list.remove(randomFeature);
                if (list.size() <= 5) {
                    System.out.println("GOING BACK...q=q-1");
                    continue;
                }
                tempScore = getResultFromGAMS(modifiedModelFileName,
resultFileName, list);
                System.out.println("INTERIM ACCURACY: " + tempScore +
" SIZE: " + list.size());

                if (tempScore > score) {

                    double displacement = -constant *
Math.abs(tempScore - score);
                    double displacementProbability =
Math.exp(displacement);
                    displacementProbability = 0.1;

                    double rand = new Random().nextDouble();
                    if (rand > displacementProbability) {
                        list.add(remove);
                        tempScore = score;
                    } else {
                        // do nothing
                    }
                } else {
                    // do nothing
                }
            }

            if (minScore > tempScore) {
                minScore = tempScore;
                size = list.size();
                finalList = list;
            }
            score = tempScore;

        }
        System.out.println("MIN_SCORE: " + minScore + " SIZE: " +
size);
        String terminals = getDecodedList(finalList);
        System.out.println(terminals);
    }

    public int generateRandomNumber(int argument, int specify) {
        Random randomGenerator = new Random();
        if (specify == 0) {
            return randomGenerator.nextInt(argument);
        }
        return (argument / 2 + (int) (Math.random() * argument / 2));
    }
```

```
    public int generateRandomNumber(int argument, List<Integer> list)
{
        Random randomGenerator = new Random();

        while (true) {
            int randomNumber = randomGenerator.nextInt(argument) + 1;
            if (!isExist(list, randomNumber)) {
                return randomNumber;
            }
        }
    }

    public String getDecodedList(List<Integer> list) {
        String terminals = "";
        for (int i = 0; i < list.size() - 1; i++) {
            if (hashMap.get(list.get(i)) != null) {
                terminals += hashMap.get(list.get(i)) + ",";
            }
        }

        return terminals;
    }

    public List<Integer> generateRandomNumbers(int argument, int
how_many) {
        List<Integer> list = new ArrayList<Integer>();
        Random randomGenerator = new Random();

        for (int i = 0; i < how_many; i++) {
            while (true) {
                int randomNumber = randomGenerator.nextInt(argument) +
1;
                if (!isExist(list, randomNumber)) {
                    list.add(randomNumber);
                    break;
                }
            }
        }
        return list;
    }

    public boolean isExist(List<Integer> list, int number) {
        for (int i = 0; i < list.size(); i++) {
            if (list.get(i).intValue() == number) {
                return true;
            }
        }
        return false;
    }

    public void writeToFile(String outFile, String text, boolean
append) throws IOException {
```

```
        FileWriter fileWriter = new FileWriter(outFile, append);
        BufferedWriter out = new BufferedWriter(fileWriter);
        out.write(text);
        out.close();
    }

    public List<Integer> getTerminalData(String fileName) throws
IOException {
        List<Integer> list = new ArrayList<Integer>();
        FileInputStream fis = new FileInputStream(fileName);
        DataInputStream dataInputStream = new DataInputStream(fis);
        BufferedReader bufferedReader = new BufferedReader(new
InputStreamReader(dataInputStream));
        String textLine;
        while ((textLine = bufferedReader.readLine()) != null) {
            String[] strings = textLine.split(",");
            for (int i = 0; i < strings.length; i++) {
                if (strings[i] != null) {
                    list.add(Integer.parseInt(strings[i]));
                }
            }
        }

        dataInputStream.close();
        return list;
    }

    public double getScore(String resultFileName) throws IOException {

        FileInputStream fileInputStream = new
FileInputStream(resultFileName);
        DataInputStream in = new DataInputStream(fileInputStream);
        BufferedReader br = new BufferedReader(new
InputStreamReader(in));
        String strLine = "";
        double score = 0.0;
        while ((strLine = br.readLine()) != null) {
            if (strLine.contains("VARIABLE w.L")) {
                String[] strings = strLine.split("\\s{1,}");
                score += Double.parseDouble(strings[5]);
                System.out.println(strings[5]);
            } else if (strLine.contains("VARIABLE u.L")) {
                String[] strings = strLine.split("\\s{1,}");
                score += Double.parseDouble(strings[4]);
                System.out.println(strings[4]);
            } else if (strLine.contains("VARIABLE t.L")) {
                String[] strings = strLine.split("\\s{1,}");
                score += Double.parseDouble(strings[4]);
                System.out.println(strings[4]);
            } else if (strLine.contains("VARIABLE c.L")) {
                String[] strings = strLine.split("\\s{1,}");
                score += Double.parseDouble(strings[4]);
```

```
                    System.out.println(strings[4]);
                }
            }
            in.close();
            return score;
        }
    }
```