LOCATION OF ALTERNATIVE-FUEL REFUELING STATIONS ON TRANSPORTATION NETWORKS CONSIDERING VEHICLE DEVIATIONS AND GREENHOUSE GAS EMISSIONS

A Dissertation in
Industrial and Manufacturing Engineering

by
Sang Jin Kweon

© 2017 Sang Jin Kweon

Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

August 2017
The dissertation of Sang Jin Kweon was reviewed and approved* by the following:

Jose A. Ventura  
Professor of Industrial and Manufacturing Engineering  
Dissertation Advisor  
Chair of Committee

Vittaladas V. Prabhu  
Professor of Industrial and Manufacturing Engineering

Uday V. Shanbhag  
Professor of Industrial and Manufacturing Engineering

Venky N. Shankar  
Professor of Civil and Environmental Engineering

Janis P. Terpenny  
Head of Department of Industrial and Manufacturing Engineering  
Professor of Industrial and Manufacturing Engineering

*Signatures are on file in the Graduate School
ABSTRACT

Burning conventional fossil fuels including gasoline and diesel mainly results in over 90% of greenhouse gas emissions from transportation. To reduce these emissions from the ground transportation sector, the use of alternative-fuel vehicles is being spotlighted. As a result, refueling station location problems for alternative-fuel vehicles have received attention as well. These refueling station location problems can be classified into two types depending on the set of candidate sites: when a preliminary (finite) set of candidate sites is given, this problem is called discrete; when the stations can be located anywhere along the network, the problem is called continuous. This dissertation considers one discrete and two continuous location problems for alternative-fuel refueling stations on transportation networks.

First, the discrete location problem is addressed with two conflicting objectives of maximizing the total vehicle-miles traveled per time unit covered by the stations and minimizing the capital cost for constructing the refueling infrastructure. Two versions of bi-criteria binary linear programming models are proposed and validated with an application to the Pennsylvania Turnpike System regarding the location of liquefied natural gas refueling stations on existing service plazas.

Next, assuming that a vehicle cannot deviate from the preplanned path, the continuous location problem for a single refueling station is addressed on a tree network with the objective of maximizing the traffic flow (in round trips per time unit) covered by the station. Two reduction properties regarding the problem size and some optimality conditions are derived. Then, an exact polynomial algorithm is developed to determine the set of optimal locations for the refueling station.

Lastly, the continuous location problem introduced above is extended to a version where a given portion of drivers are willing to deviate to be able to refuel if the station is not located
along their preplanned routes. Optimality properties and exact polynomial algorithms are suggested to determine the deviation options and find the set of optimal station locations that maximizes the traffic flow (in round trips per time unit) covered by the station.
# TABLE OF CONTENTS

List of Figures .................................................................................................................. vii
List of Tables ..................................................................................................................... ix
Acknowledgements .......................................................................................................... xi

Chapter 1 INTRODUCTION ......................................................................................... 1
  1.1 Location Problems ............................................................................................... 2
  1.2 Network Location Problems .............................................................................. 3
  1.3 Motivation ........................................................................................................... 7
  1.4 Goal, Contribution, Objective, and Research Direction ..................................... 8
  1.5 Body of the Dissertation .................................................................................... 10

Chapter 2 LITERATURE REVIEW ........................................................................... 12
  2.1 General Network Location Problems ............................................................... 12
    2.1.1 p-Center Problems ...................................................................................... 13
    2.1.2 p-Median Problems ................................................................................... 16
    2.1.3 Covering Problems .................................................................................... 19
    2.1.4 Flow Capturing Location Problems .......................................................... 23
  2.2 Refueling Station Location Problems on Transportation Networks .................. 27
    2.2.1 Vertex-Based Demand Problems .............................................................. 27
    2.2.2 Flow-Based Demand Problems ................................................................ 31

Chapter 3 ENERGY POLICY CONSIDERATIONS IN THE DESIGN OF AN ALTERNATIVE-FUEL REFUELING INFRASTRUCTURE TO REDUCE GREENHOUSE GAS EMISSIONS ON A TRANSPORTATION NETWORK .......... 41
  3.1 Introduction ......................................................................................................... 41
  3.2 Location Optimization Model ............................................................................. 46
    3.2.1 Notation ..................................................................................................... 48
    3.2.2 Bi-objective Functions ............................................................................... 49
    3.2.3 Constraints .................................................................................................. 50
    3.2.4 Solution Approach ..................................................................................... 53
  3.3 Case Study .......................................................................................................... 55
    3.3.1 Pennsylvania Turnpike ............................................................................... 55
    3.3.2 Regression Model to Estimate Construction Costs of Refueling Stations .... 58
    3.3.3 Results and Analysis ................................................................................. 62
    3.3.4 Estimation of Greenhouse Gas Emission Savings ...................................... 71
  3.4 Concluding Remarks ........................................................................................... 79
Chapter 4 A CONTINUOUS NETWORK LOCATION PROBLEM FOR A SINGLE REFUELING STATION ON A TREE

4.1 Problem Statement ........................................................................................................ 84
4.2 Refueling Segment ......................................................................................................... 90
4.3 Endpoint Optimality Theorem ...................................................................................... 92
4.4 The Single Refueling Point Algorithm .......................................................................... 93
4.5 Numerical Example ....................................................................................................... 97
4.6 Concluding Remarks ..................................................................................................... 106

Chapter 5 A CONTINUOUS DEVIATION-FLOW LOCATION PROBLEM FOR AN ALTERNATIVE-FUEL REFUELING STATION ON A TREE NETWORK ........................................... 107

5.1 Introduction ................................................................................................................... 107
5.2 Problem Statement ....................................................................................................... 114
5.3 Set of Candidate Points ............................................................................................... 121
  5.3.1 Refueling Segment ................................................................................................. 122
  5.3.2 Cycle Starting Vertex ............................................................................................ 123
  5.3.3 Cycle Returning Point ......................................................................................... 125
  5.3.4 Refueling Sub-tree ............................................................................................... 130
  5.3.5 Candidate Point Optimality Theorem ................................................................ 133
5.4 Deviation-Flow Optimal Candidate Point Algorithm .................................................. 134
5.5 Convex Combination Property .................................................................................... 138
  5.5.1 Optimality Conditions for a Convex Combination of Consecutive Optimal Candidate Points ................................................................................................. 138
  5.5.2 Complete Set of Optimal Solutions .................................................................... 142
5.6 Numerical Example and Analysis of Results ................................................................ 143
  5.6.1 Numerical Example for a Given Vehicle Driving Range ....................................... 145
  5.6.2 Sensitivity Analysis for Coupled Effects of Deviation Portion and Vehicle Driving Range ........................................................................................................... 151
  5.6.3 Performance Analysis .......................................................................................... 154
5.7 Concluding Remarks ..................................................................................................... 156

Chapter 6 CONCLUSIONS AND FUTURE RESEARCH ......................................................... 158

6.1 Conclusions ................................................................................................................... 158
6.2 Future Research ............................................................................................................ 160
  6.2.1 Future Research Regarding Chapter 3 .................................................................. 160
  6.2.2 Future Research Regarding Chapters 4 and 5 ...................................................... 162

REFERENCES ...................................................................................................................... 164
LIST OF FIGURES

Figure 1-1. Three types of network location problems (West 2001) ........................................5
Figure 1-2. Examples of interstate highways and toll roads in a form of tree networks ..........6
Figure 2-1. Classification of the network location problem .........................................................13
Figure 2-2. Two example tree networks for the vertex $k$-center problem .................................15
Figure 2-3. Example network to explain the concept about NIPS (Church and Meadows 1979) ..................................................................................................................23
Figure 2-4. The characteristics of the flow-refueling location model ........................................35
Figure 3-1. PA Turnpike with 48 interchanges and 19 service plazas (Hwang et al. 2017a) ..56
Figure 3-2. Spread-level plots for the two initial regression models ...........................................61
Figure 3-3. Results for the final regression model ........................................................................61
Figure 3-4. Trade-off between construction cost and effective VMT coverage for three different truck driving ranges (R=300, R=450, and R=600) .................................................65
Figure 3-5. Comparison of effective VMT coverage between M2 and Hwang et al. model for R=300 ..................................................................................................................................69
Figure 3-6. Estimate of truck weight distribution ........................................................................73
Figure 3-7. Cost benefit analysis of annual GHG emissions reduction and SCC savings .......76
Figure 3-8. Cost benefit analysis of annual SCC savings .............................................................79
Figure 4-1. A simple example tree to illustrate the difference between edge and path in definition and the fourth assumption ..................................................................................86
Figure 4-2. Three types of refueling segments (Ventura et al. 2015) .............................................91
Figure 4-3. Structure of tree $T$ (Ventura et al. 2015) .................................................................98
Figure 4-4. Two sub-trees from the original tree (Ventura et al. 2015) .......................................99
Figure 4-5. Refueling segments and endpoints in the first sub-tree (Ventura et al. 2015) .....101
Figure 4-6. One refueling segment and two endpoints in the second sub-tree (Ventura et al. 2015) .........................................................................................................................103
Figure 5-1. Simple deviation path ..............................................................................................118
Figure 5-2. Structure of undirected tree network $T$.................................................................144

Figure 5-3. Refueling segments, symmetric cycles, and candidate points in two sub-trees ....150

Figure 5-4. Change of the optimal solutions for the original tree network as the value of the deviation portion changes.............................................................................................................150

Figure 5-5. Trade-off between deviation portion and maximum traffic flow covered for the five different vehicle driving ranges (R=40, R=60, R=80, R=100, and R=120)........153

Figure 5-6. Locations of the optimal candidate points listed in Table 5-6..............................154

Figure 5-7. Coupled effects of vehicle driving range and deviation portion on the maximum traffic flow covered for the three models (FRLM, Ventura et al. model, and proposed model)..................................................................................................................156
LIST OF TABLES

Table 1-1. Reduction of greenhouse gas emissions benefits from the use of alternative-fuel vehicles (California Energy Commission 2007) ........................................................................... 7

Table 2-1. Assessment of Network (a) and Network (b) ................................................................. 16

Table 3-1. Climate impacts of GHGs for targeted gases .............................................................. 43

Table 3-2. List of the interchanges, toll plazas, and aggregated interchanges on the PA Turnpike .......................................................................................................................... 57

Table 3-3. Parameters and cost estimates for the 19 candidate station locations (service plazas) ............................................................................................................................... 60

Table 3-4. Effective VMT coverage for each budget and truck range combination for Models M1 and M2.............................................................................................................................. 64

Table 3-5. Optimal LNG station locations considering different budget levels and truck range combinations .................................................................................................................. 66

Table 3-6. Effective VMT coverage and the optimal locations by the Hwang et al. (2015) model........................................................................................................................................ 70

Table 3-7. Total Fuel Cycle Emissions for Target Fuels and GHGs (Meyer et al. 2011) ......72

Table 4-1. Round trip traffic flows in tree $T$ (Ventura et al. 2015) .............................................. 98

Table 4-2. Sets of the paths with positive flows covered by endpoints in the first sub-tree (Ventura et al. 2015) ................................................................................................................. 100

Table 4-3. Total traffic flow for endpoints in the first sub-tree (Ventura et al. 2015) ..........102

Table 4-4. Summary of results for the numerical example (Ventura et al. 2015)....................... 105

Table 5-1. Tree networks of highways in the U.S. ................................................................. 110

Table 5-2. Comparison between the proposed model and the existing path-based demand models considering deviation............................................................................................... 113

Table 5-3. Round trip traffic flows corresponding to all simple paths in tree $T$ ...............144

Table 5-4. Sets of paths and the traffic flows (in round trips per time unit) covered by candidate points in the first sub-tree ................................................................. 147

Table 5-5. Sets of paths and the traffic flows (in round trips per time unit) covered by candidate points in the second sub-tree ................................................................. 149
Table 5-6. Optimal locations and the corresponding traffic flows covered for each vehicle driving range (and deviation portion combination) for three models .......................... 152
ACKNOWLEDGEMENTS

It is my great honor to thank all of those who made this dissertation clear and specific. Above all, I would like to show my gratitude to my academic adviser, Dr. Ventura. I am wholeheartedly indebted to him for all the guidance and inspiration he has provided for me over the years. He encouraged me to find solutions with his valuable comments whenever I faced difficulties. In addition, he gave me an excellent example as a professor. He is my role model not only as an academic adviser but also as a great life mentor. I will remember him forever as my father in academia.

I would also like to show my gratitude to my committee members, Dr. Vittaldas Prabhu, Dr. Uday Shanbhag, and Dr. Venky Shankar for their warm encouragement and insightful advice not only on my dissertation but also about my future career.

In addition, I would like to thank Matthew Tormay and Chenxi Li for their help in Chapter 3 of the dissertation to estimate greenhouse gas emissions and construction costs of the liquefied natural gas refueling stations at the service plazas in the Pennsylvania Turnpike.

I am also grateful to my friends in Penn State. In particular, I want to note that I owe many thanks to Dr. Seong Wook Hwang and Kiseok Sung. When I first came to Penn State, I did not even know how to cook, but I had no problem because they cooked healthy and delicious foods to me for years. Moreover, Seong Wook has been my best colleague in research. I believe that, without Seong Wook and Dr. Ventura, I could not have had a good career in research.

Last but not the least, I am grateful for my parents’ eternal love and infinite support. Thanks to my parents supporting me, I have been able to study in a rich environment without financial difficulties. If my parents did not aid me both materially and spiritually, I would not be here writing my acknowledgements. I will never forget my parents’ love for me forever.
Chapter 1

INTRODUCTION

Alternative-fuel vehicles such as vehicles powered by bio-fuel, electric motors, hydrogen, and natural gas have received great attention because of their potential to reduce tailpipe emissions in the ground transportation sector. To introduce these alternative-fuel vehicles to the customer market successfully, a well-designed refueling infrastructure in transportation networks is a prerequisite since potential customers will hesitate to purchase alternative-fuel vehicles if they cannot refuel their vehicles around their travel paths. Construction of an alternative-fuel refueling infrastructure requires, however, a substantial investment. Thus, locating alternative-fuel refueling stations in the right places is a key issue so that more customers will be able to refuel their alternative-fuel vehicles. To find the potential locations for alternative-fuel refueling stations on transportation networks, new types of location problems have been addressed. The existing literature, however, does not consider the capital cost for constructing the refueling infrastructure or the sub-optimality of the solutions found by the models. In this dissertation, three models are proposed in Chapters 3-5 to address these issues in finding the potential sites for alternative-fuel refueling stations on transportation networks.

In the first chapter of the dissertation, Section 1.1 and 1.2 briefly review location problems and network location problems, respectively, to lay the foundation for this dissertation. Then, with benefits of alternative-fuel vehicles, Section 1.3 addresses the necessity of new types of network location models to determine optimal locations for alternative-fuel refueling stations. Section 1.4 then specifies the overall objective and research direction of this dissertation. Lastly, Section 1.5 introduces the body of this dissertation concisely after the first chapter.
1.1 Location Problems

Location problems have received considerable attention from many diverse application areas for centuries. In particular, location problems have directly affected all government agencies and public and private companies, including medium and small enterprises. The capability of agencies and enterprises to procure, supply, and distribute their products or to provide a high-quality service to customers relies on their location. Poor location decisions would weaken their competitiveness, and furthermore, they may increase the possibility of destruction to life and property. In other words, it is not too much to say that the success or failure of business is dependent on the location of their facilities.

Thus, it is necessary to approach location problems in a mathematical sense and define quantifiable objectives for the locations of the facilities in order to find optimal or desirable facility locations. For example, local government authorities consider the maximum time from each demand area to any of its closest facilities when they make a decision about fire station location planning, so as to minimize the worst time to reach the demand area and extinguish a fire. On the other hand, when company executives are planning to build a new distribution center for logistics, the preferred location should be where the total cost to travel between the distribution center and demand points (e.g., retailers) is minimized.

The type of quantifiable objectives usually depends on the purpose of the facilities. The facilities in the two examples above are different in purpose, thus they have different objectives. Recall that the objective of fire station location problem focuses on minimizing the maximum time because it is critical to save people’s lives and properties as soon as possible against any emergency, and even in the most undesirable situation. However, the objective of a distribution center location problem is to minimize the total or sum of weighted distances because the main purpose of a distribution center is to distribute materials or products to retailer in an effective way
(e.g., decreased cost, decreased delivery time). Like the fire station location problem, the location problem whose objective is to minimize the maximum or worst of weighted distance from each demand point to one of its closest facilities is called the $p$-center problem, where $p$ denotes the number of facilities to be placed. On the other hand, the location problem whose objective is to minimize the sum of weighted distances from each demand point to one of its closest facilities, such as the distribution center location problem, is regarded as the $p$-median problem (Francis et al. 1992).

### 1.2 Network Location Problems

Location problems can be classified in many ways. One of these classifications is based on the way in which facility locations and demand points are described, e.g., planar location problems versus network location problems (Chhajed et al. 1993). In planar location problems, facilities and demand points are located anywhere on a plane with any given coordinate. Since demands or servers can move freely in the plane without a predetermined route, planar location problems consider the Euclidean distance between a demand point and a facility in analytic geometry. On the other hand, in network location problems, facilities and demand points are located only at vertices or edges on a network, and demands or servers can only have their trip along the predetermined routes in the network. Note that the Euclidean distance by the planar location problem is always shorter than or equal to the predetermined route by the network location problem.

As seen above examples in Section 1.1 (i.e., fire station location problem and distribution center location problem), many location problems aim to optimize some function of transportation time, cost, or distance (Francis et al. 1983), and the transportation of interest has usually occurred on networks because demands or servers can only have their trips along the
predetermined routes rather than Euclidean distance. Thus, the location problems with predetermined routes can be solved using network location problems. Network location problems can be more useful than analogous planar location problems in that a solution algorithm is usually developed on the basis of an underlying network (Daskin 2011).

Network location problems are often categorized into problems that occur on tree networks and those that occur on more general networks. In a tree network, a tree or a forest is often considered. A tree is defined as a connected graph with no cycles. If a tree has \( n \) vertices, it will have \( n-1 \) edges. Thus, a tree will have a cycle and will not be a tree anymore if we add one more edge that connects two vertices in it. A forest consists of more than one tree, but still there are no cycles in it. Since a forest is simply composed of a group of trees, each tree is not linked to any other tree; each tree is isolated. It means that some vertices are never connected in a forest. In a tree network, however, any pair of vertices is connected by a unique path; thus, there exist no cycles. It implies that every edge of a tree network is a cut-edge, meaning that removing the edge from a network results in more components. Part (a) and Part (b) of Figure 1-1 show examples of a tree network. Part (a) is one tree that has six vertices and five edges. Every edge in Part (a) is a cut-edge, thus losing any of them will lead to more components, while adding any edge to Part (a) will make a cycle. Part (b) has one forest consisting of two trees. Each tree of Part (b) has 3 vertices and 2 edges. Note that vertices of the left tree are not connected to any of those in the right tree. On the contrary, a general network includes at least one cycle. It implies that there is more than one path between some vertices; thus, every edge cannot be a cut-edge. Part (c) of Figure 1-1 shows an example of a general network which contains one cycle.

As compared with general networks, many network location models have been developed based on tree networks for many reasons. One of the reasons a tree network is preferred is that it is easier to get insight into a tree network location problem prior to more general networks (Francis et al. 1992). Recall that every edge of a tree network is a cut-edge, meaning that there is
only one path between each pair of vertices. Thus, a solution algorithm for location problems considers only one scenario for each pair of vertices, and it will reduce the complexity enough to be practical. Furthermore, we can easily observe that many real-life network location problems form tree or tree-like networks. In particular, a tree network is a common structure for ground transportation networks such as interstate highways or toll roads. If we regard beltways within metropolitan areas as single vertices, interstate highways form tree networks in many states, including Arizona, Missouri, North Dakota, Washington, Montana, Idaho, Wyoming, New Mexico, Mississippi, Nevada, and Utah (Federal Highway Administration 2016a). Tree or tree-like networks are also easily observed from the toll roads in many states (Federal Highway Administration 2016b). For example, the Oklahoma and Pennsylvania Turnpikes are described as forests that consist of several tree networks (Oklahoma Turnpike 2014, Pennsylvania Turnpike 2014). Also, Indiana Toll Road and the Maine and Ohio Turnpikes form a single path tree structure (Indiana Toll Road 2014, Maine Turnpike 2014, Ohio Turnpike 2014). Figure 1-2 illustrates five examples of interstate highways and toll roads in a form of tree networks. In case of Part (a) of Figure 1-2 that shows the interstate highway in Mississippi, it forms a forest which consists of two sub-trees. The other cases, Parts (b) – (e) of Figure 1-2, forms one trees, respectively.

![Diagram]

(a) a tree  (b) a forest with two trees  (c) a general network having a cycle

Figure 1-1. Three types of network location problems (West 2001)
Figure 1-2. Examples of interstate highways and toll roads in a form of tree networks.
1.3 Motivation

Location problems on road networks are currently one of the interesting application areas in transportation areas. In particular, as alternative-fuel vehicles are being spotlighted to replace the internal-combustion engine vehicles in the near future (Shukla et al. 2011), a number of models to locate alternative-fuel refueling stations on road networks have been developed.

The necessity of network location models for alternative-fuel refueling stations originates from the public interest in alternative-fuel vehicles to solve environmental and economic issues (MirHassani and Ebrazi 2013). According to the Environmental Protection Agency annual report, burning conventional fossil fuels including gasoline and diesel mainly results in over 90% of greenhouse gas emissions from transportation (Environmental Protection Agency 2014). These greenhouse gas emissions cause air pollution and global warming problems often called “the greenhouse effect”. As global concerns over air quality and climate change intensify, there has been increasing pressure to reduce greenhouse gas emissions from ground transportation sector. Using alternative-fuels in the transportation sector can reduce tailpipe greenhouse gas emissions compared to gasoline and diesel. Table 1-1 provides greenhouse gas emissions benefits for alternative-fuels in midsize automobiles and urban buses across the evaluation timeframe.

<table>
<thead>
<tr>
<th>Vehicle type</th>
<th>Natural gas</th>
<th>Hydrogen</th>
<th>Electric</th>
</tr>
</thead>
<tbody>
<tr>
<td>Midsize automobiles</td>
<td>20-30% reduction compared to gasoline</td>
<td>45% reduction compared to gasoline</td>
<td>68% reduction compared to gasoline</td>
</tr>
<tr>
<td>Urban buses</td>
<td>23-24% reduction compared to diesel</td>
<td>23-24% reduction compared to diesel</td>
<td>55% reduction compared to diesel</td>
</tr>
</tbody>
</table>

Table 1-1. Reduction of greenhouse gas emissions benefits from the use of alternative-fuel vehicles (California Energy Commission 2007)
Despite of all the benefits from alternative-fuel vehicles, these vehicles are not yet as popular as the internal-combustion engine vehicles. One of the barriers to a wide use of alternative-fuel vehicles is a shortage of a refueling structure infrastructure (MirHassani and Ebrazi 2013, Rodrigue et al. 2006). Drivers take refueling availability into consideration when buying a vehicle (Kitamura and Sperling 1987). Thus, a sufficient number of refueling stations should be prepared to satisfy an initial market for alternative-fuel vehicles (Kurani et al. 1996), but construction and maintenance of an alternative-fuel refueling infrastructure requires a substantial investment (Melaina and Bremsen 2008). Thus, locating alternative-fuel refueling stations in the right places is a key issue so that more customers will be able to refuel their alternative-fuel vehicles.

1.4 Goal, Contribution, Objective, and Research Direction

The goal of this dissertation is to contribute to the development of the refueling infrastructure by proposing models that consider bi-objective functions and improve the sub-optimality of the solutions with vehicle deviations that the existing literature has not addressed.

First, bi-objective functions are required to consider various factors when finding the potential locations for alternative-fuel refueling stations. However, the existing literature in the area of the network location problems for alternative-fuel refueling stations (also called “refueling station location problem”) usually has a single objective; that is, the objective is to either maximize the total traffic flow covered by the stations or minimize the number of stations to cover all the traffic flow, assuming that the capital cost for constructing the refueling infrastructure is equal regardless of the region. In reality, this capital cost may differ regionally; thus, a bi-objective version of the refueling station location problem is required to consider the different capital cost. In Chapter 3 of this dissertation, a refueling station location model is
developed as a form of a bi-criteria binary linear programming model to locate alternative-fuel
refueling stations on a directed transportation network with two conflicting objectives of
maximizing the total vehicle-miles traveled per time unit covered by the stations and minimizing
the capital cost for constructing the refueling infrastructure. Two versions of bi-criteria binary
linear programming models are proposed and validated with an application to the Pennsylvania
Turnpike System regarding the location of liquefied natural gas refueling stations on existing
service plazas.

Next, a novel approach is required to improve the sub-optimality of the locations found
by the models in the existing literature. Network location problems can be classified into two
types depending on the set of candidate sites: when a preliminary (finite) set of candidate sites is
given, this problem is called discrete; when the stations can be located anywhere along the
network, the problem is called continuous. For the network location problems where demand has
no limited driving range, a finite dominating set is established considering only the vertices of a
network; thus, the results of the discrete version of the problem are the same to those of the
continuous version. However, in practice, demand in the refueling station location problem has a
limited driving range per refueling; thus, a finite dominating set cannot be established including
only the vertices of a network in this case. This implies that the optimal locations of the discrete
version of the problem may be suboptimal to those of the continuous version. Nevertheless, the
existing literature prefer using the discrete version, since the continuous version is more
challenging than its discrete counterpart due to the size of the search space. Thus, polynomial
algorithms need to be developed to address the continuous version of the refueling station
location problem. In Chapter 4, the continuous version of the single refueling station location
model is developed on a tree network, assuming that vehicle deviation is not allowed. In this
model, the objective is maximizing the traffic flow in round trips per time unit covered by the
station. Two reduction properties regarding the problem size and optimality conditions are
derived. Then, an exact polynomial algorithm is developed to determine the set of optimal locations for the refueling station. In Chapter 5, this model is extended to a version where a given portion of drivers are willing to deviate to be able to refuel if the station is not located along their preplanned routes. Optimality properties and exact polynomial algorithms are suggested to determine the deviation options and find the set of optimal station locations that cover the maximum traffic flow in round trips per time unit covered by the station.

For given networks in the proposed models, key inputs are the travel distances and the traffic flows between all origin and destination pairs, the fuel tank levels at all origin and destination points, and the alternative-fuel vehicle driving range.

1.5 Body of the Dissertation

The body of this dissertation is organized as follows. In Chapter 2, the literature about both general network location problems and the network location problems specialized in refueling stations is reviewed to provide the theoretical foundation for this dissertation. In Chapter 3, a discrete version of the refueling station location model is suggested with two conflicting objectives of maximizing the total vehicle-miles traveled per time unit covered by the stations and minimizing the capital cost for constructing the refueling infrastructure, and further analysis is performed on the Pennsylvania Turnpike to estimate the impact of alternative-fuel refueling station construction on the greenhouse gas emissions reductions and the social cost of carbon savings. In Chapter 4, a continuous version of the refueling station location model is developed to find the potential sites for the single refueling station on a tree network, given that vehicle deviation is not allowed. In Chapter 5, the continuous location problem proposed in Chapter 4 is extended to a version where a given portion of drivers are willing to deviate to be able to refuel if
the station is not located along their preplanned routes. Lastly in Chapter 6, conclusions and future research topics are discussed.
Chapter 2

LITERATURE REVIEW

In this chapter we review the relevant literature in the area of network location problems. In Section 2.1, we have a brief look at general network location problems in terms of motivation, property, modeling, application, complexity, and algorithm. In Section 2.2, among the network location problems, we then specially focus on the literature about refueling station location problems on transportation networks.

2.1 General Network Location Problems

The network location problem addresses finding the potential sites for facilities on a given network. This problem basically consists of demands and facilities (or equivalently, servers). Demands receive service at one of the facilities; in other words, facilities provide services to demands. Classical network location problems generally assume that the goal of a trip initiated by demands or servers is to receive or provide service. This trip has been called the one-stop tour between a demand point and a facility site (Berman et al. 1992). In this dissertation, it will be referred to as one-stop trip. In the location problems following this one-stop trip assumption, demands are located at the vertex of the network (i.e., vertex-based demands). Daskin (2011) categorizes these network location problems into the specific property of the problems, such as \( p \)-center problems, \( p \)-median problems, and covering problems. On the other hand, some network location problems assume that demands receive service in a discretionary way on the way to their preplanned trips (Berman et al. 1992, Hodgson 1990). In these problems, demands are described as a flow between their origin and destination pair (i.e., flow-based demands). In this section, we deal with literature about general network location problems for both vertex-based demands (i.e.,
\( p \)-center problems, \( p \)-median problems, and covering problems) and flow-based demands (i.e., flow capturing location model). Figure 2-1 summarizes the classification of literature in this section.

Figure 2-1. Classification of the network location problem

### 2.1.1 \( p \)-Center Problems

In this section, we provide a brief review of the network location problem called the \( p \)-center problem, where \( p \) represents the number of facilities, also called centers, to be located. In this problem, each demand location is given as a vertex of the network, and an edge between vertices represents the shortest distance between demands. The number of facilities, denoted as \( p \), is also generally predetermined based on the capability of public and private enterprises (such as financial and other reasons). The \( p \)-center problem is also called the minimax problem because it aims at finding facility locations that minimize the maximum edge between a demand node and its closest facility, on the assumption that every demand is served by its closest facility (Minieka 1970). Since its objective is to minimize the maximum edge, this problem basically assumes that all demands are served by at least one facility. If there existed at least one demand which is not served by any facility, the objective function would have an infinite value.
According to the permissible level of facility location in the network, $p$-center problems can be categorized as vertex $p$-center problems or absolute $p$-center problems. While vertex $p$-center problems allow facilities to be located only on the vertices of the network, absolute $p$-center problems allow facilities to be located anywhere on the network. This implies that the vertex $p$-center problem may be better than the absolute $p$-center problem in terms of the complexity, whereas the absolute $p$-center problem would be better than the vertex $p$-center problem at optimal solution. Fortunately, both the vertex $p$-center problem and the absolute $p$-center problem may derive optimal solutions in polynomial time if $p$ is fixed (Daskin 2011). On the other hand, if $p$ is not fixed and a given network includes at least one cycle, both the $p$-center problem and the absolute $p$-center problem are NP-hard (Kariv and Hakimi 1979a).

To find the optimal solution of the $p$-center problem in polynomial time, a number of algorithms have been developed. In particular, Hochbaum and Shmoys (1985) suggest a 2-approximation algorithm for the $p$-center problem which satisfies the triangle inequality. This algorithm has one $until$ loop, which takes $O(\log|E|)$ time for binary search until $p$-centers are located, and each binary search contains $while$ loop, which takes $O(|E|)$ time, where $E$ denotes the edge set in a graph, and $|E|$ is the cardinality of $E$. Thus, this algorithm solves the $p$-center problem which satisfies the triangle inequality in polynomial (i.e., $O(|E| \log|E|)$) time. On the other hand, Mladenović et al. (2003) build three metaheuristics (i.e., Tabu search, Variable neighborhood search, and Multistart local search) for solving the $p$-center problem without the triangle inequality. All of these three metaheuristics have been verified to be significantly better than the binary search heuristic both at running time and quality of solutions, through the 40 OR-Lib test Problems (Mladenović et al. 2003).

The $p$-center problem has been applied widely, in particular, to the emergency stations location (Mladenović et al. 2003, Shier 1977). This is because the objective of the $p$-center problem, i.e., minimizing the maximum distance, naturally corresponds to minimizing the
most undesirable situation to respond to emergency scenarios, in order to save people’s lives and properties.

Nevertheless, the $p$-center problem has some intrinsic issues in its wide usage. Most of all, its formulation does not consider any benefit associated with the distance between each demand point and one of its closest facilities, while there exist clear benefits associated with the distance in many cases. Additionally, the $p$-center problem does not consider the priority of demand, even if every demand has different weight in most real world problems. Furthermore, the $p$-center problem only focuses on minimax but not average distance, which often leads to inappropriate decision making in assessing networks. For example, Figure 2-2 shows two networks for the vertex $l$-center problem. The number on each edge denotes the (shortest) distance between demand nodes. In assessing the two networks in Figure 2-2, it is clear that Network (a) is better than Network (b) in maintaining $l$-center in terms of average driving time, cost, etc. But, $l$-center problem prefers Network (b) for the sole reason that the min-max distance of Network (b), i.e., 99, is shorter than that of Network (a), i.e., 100.

Figure 2-2. Two example tree networks for the vertex $l$-center problem
Table 2-1. Assessment of Network (a) and Network (b)

<table>
<thead>
<tr>
<th></th>
<th>Network (a)</th>
<th>Network (b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average distance</td>
<td>13.375</td>
<td>99.000</td>
</tr>
<tr>
<td>Min-Max distance</td>
<td>100.000</td>
<td>99.000</td>
</tr>
</tbody>
</table>

In the next section, we present a brief review of the $p$-median problem that remedies shortcomings of the $p$-center problem.

2.1.2 $p$-Median Problems

While the $p$-center problem focuses on minimizing the maximum distance between each demand point and one of its closest facilities, the $p$-median problem aims at finding the $p$-facility locations that minimize the (weighted) sum of distances between demand nodes and their closest facilities in the network. The $p$-median problem overcomes the weakness of the $p$-center problem pointed out in the above section. Recall from the above example in Figure 2-2 that the vertex $I$-center problem prefers Network (b) to Network (a). The conclusion of the $I$-center problem that Network (b) is better than Network (a), however, cannot be acceptable in many cases, in particular, in business. On the other hand, if we apply the $I$-median problem to the same example, it leads to the opposite result. In the $I$-median problem, the total distance minimized by the $I$-median problem is 107 in Network (a) and 792 in Network (b). Thus, the $I$-median problem prefers Network (a) to Network (b) in terms of driving time or cost, which is usually a more realistic preference in many cases.

As seen above, the economic property for which the $p$-median problem seeks makes it widely useful for many location problems in business. In particular, the $p$-median problem is very useful for locating hub facilities in the transportation sector or communication systems. Campbell (1996) discusses the properties of the hub location problem and applies the $p$-median problem to
derive an integer programming formulation of the $p$-hub location problem. Since it would take tremendous time theoretically to solve the $p$-median problem for large $p$ (we discuss the complexity of the $p$-median problem at the next paragraph), Campbell (1996) introduces two heuristics, i.e., MAXFLO and ALLFLO, for solving single $p$-hub location problems. First, MAXFLO heuristic allocates every demand to the hub such that the maximum flow from and to a demand occurs between a demand and the hub. Then, ALLFLO heuristic finds $p$-hub location that minimizes the total transportation cost. The $p$-median problem has been also used in dynamic facility location problems, introduced by Wesolowsky (1973) and Wesolowsky and Truscott (1975). In the dynamic facility location problem, each demand is assumed to change over the time horizon, but the facilities are located over the time horizon. On the assumption that relocation of a facility is prohibited in the model, Drezner (1995) applies the $p$-median problem to the dynamic location problem in order to find the optimal facility locations that produce the least total cost in the transportation sector over the time horizon.

One of the main properties of the $p$-median problem is that there exists at least one optimal solution which is composed of $p$ facilities located only on vertices of the network (Hakimi 1965). This property reduces the complexity of the $p$-median problem by finding an optimal solution among the combination of $p$ facilities out of all candidate sites i.e., $\binom{n}{p}$, where $n$ is the number of candidate sites. For $p \ll n$, the $p$-median problem is $O(n^p)$. It represents that it is solved in polynomial time theoretically for a given value of $p$.

The problem is, however, the number of combinations would be very large as the value of $p$ increases, and it would take astronomical time even if its complexity is polynomial (Daskin 2011). Besides the number of combinations, the $p$-median problem becomes NP-hard and it may take exponential time to find the optimal solution if $p$ is variable or the problem is defined on a general network (Garey and Johnson 1979, Kariv and Hakimi 1979b). Most of $p$-median
problems of realistic size have fixed but large values of $p$ and $n$ or variable $p$ and $n$ with a general network including cycles rather than a simple tree network. These factors may increase the complexity of solving optimal solutions. In order to reduce its complexity, there have been a number of studies to find the heuristic algorithms that approximate the optimal solution in polynomial time. Teitz and Bart (1968) suggest vertex substitution heuristic algorithm for the $p$-median problem, which is regarded as one of the exchange heuristic algorithms and classified as the improvement algorithm (Golden et al. 1980). The exchange heuristic algorithm improves an initial solution that is randomly derived by changing facility locations at each iteration. The solution depends on the initial solution and substitution procedure. Thus, its solution and running time would be good if the problem size is small, but it may take more than a reasonable amount of time and may not be sufficient for a global optimum solution if the problem size is large. To solve large size $p$-median problems, Densham and Rushton (1992) present the global/regional interchange algorithm by applying two-phase search heuristic to the exchange heuristic algorithm in order to guarantee less computation for large size problems. On the other hand, Narula et al. (1977) propose the heuristic algorithm by applying Lagrangian relaxation and subgradient method to the $p$-median problem. It is worth noticing that the core requirement of their method is a linear increasing function with the number of candidate sites, meaning that this algorithm is applicable to large size problems. Modern heuristic algorithms also have been developed for the $p$-median problem. Murray and Church (1996) present a simulated annealing method based on a heuristic algorithm to solve the $p$-median problem and compare it to the interchange heuristic algorithm. Chiyoshi and Galvão (2000) combine the vertex substitution method developed by Teitz and Bart (1968) with a simulated annealing method to solve the $p$-median problem and develop a regression equation about computing times for the $p$-median problems ranging from 500 to 700 vertices. Alp et al. (2003) use a genetic algorithm for the $p$-median problem. In particular, they use a greedy selection heuristic to facilitate new solutions at each iteration in
order to generate a number of better solutions quickly, instead of using a crossover operator. This leads to generating solutions within 0.1% of the optimal solutions for 85% of their 80 test problems, and the worst solution records only 0.41% away from the optimal solution.

2.1.3 Covering Problems

In this section, we briefly review covering problems to locate facilities on the network. The critical factor in covering problems is the shortest distance between each demand point and one of its closest facilities. In covering problems, on the assumption that every demand is assigned to one of its closest facilities, the closest facility can provide service or “cover” a demand if the shortest distance between them is within a given coverage. The expression of “covering” comes from this concept. To find the optimal location under distance restriction between the facilities and the demands, covering problems include coverage constraints with some binary variables, which are equal to 1 if an assigned facility covers demands (i.e., the shortest distance between them does not exceed the coverage); 0 otherwise.

2.1.3.1 Set Covering Problems

According to an objective type for the facility location, covering problems can be classified into the set covering problem or the maximum covering problem (Schilling et al. 1993). The set covering problem finds the optimal location for the facilities to cover all demand points with the minimum cost (Toregas 1971). They have been used not only in a broad range of location problems but also beyond the scope of location concerns (Daskin 2011). Kolesar and Walker (1974), Walker (1974), and Plane and Hendrick (1977) propose a set covering formulation to minimize the total time on locating fire stations or the fire trucks having a tower ladder. Daskin
and Stern (1981) modify the conventional set covering model to find the optimal location for emergency medical service vehicles that satisfies the two following objectives simultaneously: (i) minimize the number of vehicles required to cover all demand, and (ii) maximize the multiple coverage of demand. Set covering problems also have been well-applied to flight crew scheduling (Rubin 1973, Marsten and Shepardson 1981, Graves et al. 1993). Reggia et al. (1983) apply a set covering model to diagnostic expert systems by using the model to resolving multiple simultaneous disorders and checking past work on the systems.

Since the set covering problem is NP-hard on a general graph (Garey and Johnson 1979), there has been steady research on heuristic algorithms to solve the problem more efficiently. Beasley (1990) suggests a Lagrangian heuristic which is based on Lagrangian relaxation and improves the optimal solution found by the set covering problem at each iteration of a subgradient optimization procedure. As an alternative to a Lagrangian relaxation method, Lorena and Lopes (1994) use a surrogate relaxation method and formulate the problem as a knapsack problem to present a heuristic algorithm which reduces the number of iterations and leads to a shorter computing time on finding a final solution, compared to the heuristics based on a Lagrangian relaxation. Jacobs and Brusco (1995) introduce a local-search heuristic for the set covering problem on the basis of the simulated annealing algorithm. Beasley and Chu (1996) present a heuristic based on a genetic algorithm to efficiently find the high-quality solution for the non-unicost set covering problem. However, the sequential process of the standard genetic algorithm is lamented for a waste of processing time towards the high-quality solution. To improve a genetic algorithm in terms of computing time and quality of solutions simultaneously, Solar et al. (2002) propose a parallel genetic algorithm that reduces the processing time when solving the set covering problem.

In spite of its wide usage in a variety of fields, the set covering model has two main weak points in its formulation: (i) Every demand node is treated identically; thus, the model cannot
prioritize demands, and (ii) the model does not stipulate the maximum number of facilities allowed (for financial and other reasons); therefore, there is no way to keep the optimal number of facilities from exceeding the capability of public and private enterprises (Daskin 2011).

### 2.1.3.2 Maximum Covering Problems

To overcome these vulnerable points of the set covering problem discussed at the end of the previous section, Church and ReVelle (1974) present the maximum covering problem. While the set covering problem searches for the facility locations to minimize the number of facilities to cover all demand nodes, the maximum covering problem aims at finding the optimal facility locations to maximize the extent of covered demands within capability of companies (i.e., the maximum number of facilities allowed). The maximum covering problem also has been applied in many ways, as well as the set covering problem (Chung 1986, Daskin 1983, Galvão et al. 2000).

In theory, the time required to solve the maximum covering problem is $O\left(DP^2\right)$, which is NP-hard (Megiddo et al. 1983), where D is the number of demand nodes, K is the number of candidate sites for a facility, and P is the number of facilities to be located. To avoid its huge complexity and solve the maximum covering problem more efficiently, many heuristic algorithms have been studied. The greedy adding algorithm locates the first facility in the place where the facility covers the most demands. Then, it locates the next facility in the place where this facility covers the most demands among the demands not covered by the first facility. It repeats this process until all facilities are located or all demands are covered. By doing this, the complexity decreases to $O(DPK)$ (Church and ReVelle 1974). However, there is no way to verify (i) if the solution resulting from the greedy adding algorithm is optimal, and (ii) how far the optimal solution is from the solution of the greedy adding algorithm. The heuristic algorithm
based on Lagrangian relaxation is one of the alternatives to the greedy adding algorithm in that it improves lower and upper bounds at every iteration and derives a better approximate to the optimal solution, especially for large size problems (Galvão and ReVelle 1996). Note that neither of them (i.e., the greedy adding algorithm and the Lagrangian relaxation) is guaranteed to derive an optimal solution. An optimal solution would be acquired by using the branch and bound method, but this is an \(O(2^n)\) algorithm and may need an exponential time to obtain the solution.

2.1.3.3 Network Intersect Point Set

While the original set covering model and maximum covering model restrict facility location only at demand nodes, Church and Meadows (1979) consider the case in which facility location is allowed anywhere on the network for both problems. This is attractive since it may improve the optimal solutions. For example, there are two demand nodes, \(N_1\) and \(N_2\), with equal weight, and the distance between them is 5, shown in Figure 2-3. If coverage distance for each demand node is 3 and facility location is restricted only at nodes, two facilities need to be located on each node for the set covering problem. But, if facility location can be located anywhere on the network, only one facility on any intersection point of two coverage distances is enough to cover all demand nodes. By definition of Church and Meadows (1979), each of two points, A and B, is defined as a network intersect point (NIP) because the distance between A and \(N_2\) and the distance between B and \(N_1\) are the longest possible coverage for demand nodes \(N_2\) and \(N_1\), respectively. Also by definition of Church and Meadows (1979), the set of all NIPs (i.e., A and B) plus all demand nodes (i.e., \(N_1\) and \(N_2\)) are defined as network intersect point set (NIPS). NIP at least covers the same demand nodes as the interior points between two NIPS cover, shown in Figure 2-3. Church and Meadows (1979) claim that it is possible for some NIPs to cover more demand nodes than their interior points as the network in Figure 2-3 is extended to that with more
demand nodes. Based on those properties, Church and Meadows (1979) derive and prove the theorem that “if the set covering problem or maximum covering problem allows facility location anywhere on the network, then there exists at least one optimal solution to the problem that consists entirely of points belonging to the NIPS.”

Figure 2-3. Example network to explain the concept about NIPS (Church and Meadows 1979)

2.1.4 Flow Capturing Location Problems

The previous network location problems (i.e., $p$-center problems, $p$-median problems, and covering problems) we have reviewed in Sections 2.1.1 to 2.1.3 assume that demands (e.g., drivers, customers) or servers (e.g., ambulances, fire trucks, home delivery service vehicles) have a one-stop trip between a demand and a facility in order to consume or provide the service and return to their original places. This assumption is valid for the trip in which the service itself is the goal. However, many services are also provided on the way to their preplanned trips. For example, a driver has a preplanned trip from New York to Pittsburgh and maybe need to refuel his car on his way to the destination (i.e., Pittsburgh). In this example, the original goal of the trip is to arrive at the destination, and the service (i.e., refueling) is provided in a discretionary way
(Berman et al. 1992). Thus, this cannot be defined as one-stop trip, and one-stop trip assumption is not valid for discretionary service activities.

For discretionary service activities, traffic flows are a key factor that needs to be considered in locating facilities. But, facility location decision making simply on the basis of passing flows by traffic counts would locate facilities so close to each other by multiple counting of flows and may lead to flow cannibalization (Ghosh and McLafferty 1987). To avoid cannibalization in locating facilities, Hodgson (1990) distinguishes passing flow into origin and destination (OD) pair and introduces the flow capturing location model (FCLM) that finds $p$ facility locations to maximize the total amount of flows between OD pairs “captured” by the facilities. In his model, the term “capture” means covering all the flow that passes by the facility.

In the FCLM, the traffic flow between each origin and destination, denoted as $q \in Q$, represents a demand, and origins and destinations are described as vertices of the network. The integer programming formulation of the FCLM is structurally identical to that of the maximum covering location model (MCLM), except for the definition of $N_q$. While $N_i$ is defined as the set of vertices where a facility can cover a demand vertex $i$ in the MCLM (Church and ReVelle 1974), $N_q$ is used in the FCLM to define the set of vertices (i.e., candidate facility sites) which can capture OD pair $q$. The FCLM counts an OD pair $q$ as captured if a facility which belongs to $N_q$ is located anywhere along $q$.

Recall that the formulation of the FCLM is structurally identical to that of the MCLM. Thus, FCLM is NP-hard as well as the MCLM (Megiddo et al. 1983). Hodgson (1990) and Berman et al. (1992) suggest a greedy heuristic algorithm to solve large size FCLM problems in a reasonable amount of time. The main concept of the greedy heuristic algorithm for FCLM is very similar to the greedy heuristic algorithm for MCLM. It locates the first facility at the vertex which can capture the maximum OD pairs. Then, it repeats locating the next facilities at the vertices which can capture the next maximum flow of OD pairs until either all OD pairs are captured or
all the $p$ facilities are located in the network. Note that Berman et al. (1992) are another authors who independently study the network location problem for discretionary service activities and present similar approach, formulation, and conclusions to those of the FCLM. Instead of using the term “capture” in the FCLM, Berman et al. (1992) use the term “intercept” as a form of capturing the flow in their model, and therefore, their model is called the flow intercepting location model (FILM).

The FCLM has been applied in a variety of fields. Hodgson and Rosing (1992) consider two types of different commercial demands (i.e., demands under one-stop trip and demands under preplanned trip) at the same time in the network. Since the $p$-median problem is suitable for commercial demands under a one-stop trip and the FCLM for demands under a preplanned trip, they build a hybrid model by combining the $p$-median problem and the FCLM together. Berman et al. (1995c) extend the FCLM to the model that has probabilistic demand and considers facility setup cost. They use the Markov chain to deal with probabilistic data and formulate a nonlinear integer programming. Since the computation is unfavorable for the nonlinear integer programming formulation, they transform it into the equivalent linear integer programming formulation. Then they use the greedy heuristic algorithm presented by Berman et al. (1992) to approximate the optimal solution. Berman et al. (1995b) apply the FCLM to the refueling problem when flows of OD pairs can deviate from their shortest paths to refuel, on the assumption that their maximum deviation is fixed as $\Delta$. Hodgson et al. (1996) apply FCLM to a preventive inspection model that locates vehicle inspection stations to prevent hazardous vehicles such as trucks carrying explosive items in a dangerous way or vehicles driven under the influence from driving the public road networks. The main difference of a preventive inspection model from the classic FCLM (Hodgson 1990) is the concept of capturing. While the classic FCLM captures vehicles if an inspection station is located anywhere between a vehicle’s OD pair, a preventive inspection model captures vehicles if an inspection location is located only at the
origin vertex because this model aims to prevent hazardous vehicles from using the road networks from the first. Thus, the objective is not to maximize the number of inspections in order to punish hazardous vehicles but to maximize prevention of hazardous vehicles at their origins. Hodgson and Berman (1997) allow multiple counting in applying FCLM to the billboard location problem because exposure of a billboard to drivers several times is better than exposure only once. Berman and Krass (1998) combine the spatial interaction model with the FCLM in order to consider competition with the existing facilities and preplanned trip demand at the same time. The spatial interaction model itself assumes that a demand is located at the vertex which represents one-stop trip between a demand and a facility. Thus, the combined model presented has two types of demands (i.e., demands under one-stop trip and demands under preplanned trip), like Hodgson and Rosing (1992), under competitive condition. Boccia et al. (2009) discuss five issues in applying the FCLM according to the specific objective of the problem and present two flow oriented problems (i.e., maximization of the intercepted flow and minimization of the number of flow intercepting facilities) and two gain oriented problems (i.e., maximization of the achievable gain and minimization of the number of facilities for gain maximization). They also suggest the applications of these models to vehicle inspection station location for protecting a road network (Hodgson et al. 1996), variable message signs location for providing traffic guidance (Huynh et al. 2003), and traffic counting sensor location for estimating OD flow matrix (Bianco et al. 2001). The FCLM has been also applied to the pickup problem suggested by Zeng et al. (2009) to consider four types of customer’s location preference and its benefit according to a product type. The objective of this problem is to maximize the total benefit of flows captured that arises from various customer’s location preferences.
2.2 Refueling Station Location Problems on Transportation Networks

Compared to the number of published papers addressing the network location problems in the other areas, the number of published papers considering refueling station location problems on transportation networks is relatively fewer due to the omnipresence of gas stations throughout the world and the dominance of the existing vehicles refueled at these gas stations (Kuby and Lim 2005). Since environmental and economic issues related to these existing vehicles powered by gasoline or diesel have gotten more serious, alternative-fuel vehicles have received great attention because of their potential to increase the environmental and economic sustainability. As a result, finding potential locations for alternative-fuel refueling stations has been steadily studied as an effort to accelerate a wide usage of alternative-fuel vehicles. In this section, we classify literature that addresses refueling station location problems on transportation networks into the demand type, i.e., the vertex-based demand problems and the flow-based demand problems. Then we have a look at both of the two types of problems.

2.2.1 Vertex-Based Demand Problems

This section focuses on the literature that addresses vertex-based demand problems. The $p$-median model and its related limitations are presented in Section 2.2.1.1. Then, the maximum covering model, which has the potential to overcome the limitations of the $p$-median model, is discussed in Section 2.2.1.2. After that, we briefly point out the limitations of both of these models in Section 2.2.1.3.
2.2.1.1 Models Based on \( p \)-Median Model Approach

Some refueling station location models with vertex-based demand are based on the \( p \)-median model approach, which has been reviewed in Section 2.1.2. Nicholas et al. (2004) use the \( p \)-median model to determine the number and the locations of hydrogen fuel refueling stations in Sacramento County, California, so as to reduce the average driving time from driver’s home to the closest station within a given time. As a follow-up study, Nicholas and Odgen (2006) apply their model to find the potential locations for the four metropolitan areas in California and compare the results. Similarly, Chan et al. (2007) use the \( p \)-median model to find the optimal locations for the refueling station so as to minimize the total expected traveling costs to the refueling station of all drivers in Singapore.

Another application of the \( p \)-median model is a “fuel-travel-back” approach. The “fuel-travel-back” refers to the time from a starting point of the refueling trip to its closest refueling station, and one of the main assumptions of this approach is that “where you drive more is where you more likely need refueling” (Lin et al. 2008). Lin et al. (2008) use this approach to locate hydrogen fuel stations in Southern California. Since the fuel-travel-back model is structurally identical to the \( p \)-median model, a demand (i.e., a driver) of this model is vertex-based and is assumed to have one-stop trip from his/her start to a hydrogen fuel station. The objective of the model is to find the set of hydrogen fuel station locations on the vertex in the transportation network so as to minimize the average refueling travel time of a driver. However, this approach cannot exclude unpractical solutions for refueling station locations, such as location continuity or multiple stations for one node.

To implement the \( p \)-median model, spatial and traffic data are required, such as population, passing traffic flow, and distances between demands and their closest candidate sites. The fact that these data can be easily obtained is a benefit of this model. Nevertheless, the \( p \)-
median model has a limitation in solving the refueling station location problem in that the \( p \)-median model does not consider the covering distance of each demand (i.e., vehicle driving range). Without the vehicle driving range, this model cannot consider a multiple refueling scenario for a long trip, which often happens to many drivers. To address the vehicle driving range in the refueling station location problem, we have a look at the literature based on the covering problem in the next section.

2.2.1.2 Models Based on Maximum Covering Model Approach

The covering model is well-suited when a demand has the covering distance for the service. For a refueling station location model, the demand is a vehicle which has a limited driving range per refueling. Thus, the covering model can be said to be well-suited for refueling station location models. This section reviews the literature in the area of refueling station location problem based on the maximum covering model approach.

Bapna et al. (2002) apply a maximum covering/shortest path model, which is formulated by Current et al (1985), to find the potential locations for unleaded gas stations, and therefore, to make an inter-city trip available in India. This is a variant of the maximum covering problem; thus, a demand, i.e., a driver who plans to make an inter-city trip, is vertex-based, and each vertex represents a city where the drivers are gathered. Also, the covering distance in their model represents the vehicle driving range. Their approach has two conflicting objectives to include various factors in managing gas stations. The first objective is to minimize the total cost in locating and operating gas stations, and the second objective is to maximize the extent of covering (i.e., refueling opportunity), which is identical to the objective of the original maximum covering model.
Regarding the first objective of their model, the total cost is divided into the fixed cost and the traveling cost. First, the fixed cost occurs when locating a gas station. India is consistently converting leaded gas into unleaded gas in order to reduce air pollution; thus, there are two possible options to locate unleaded gas stations. The first option is simply converting the existing leaded gas station into the unleaded gas station. In this case, the cost is saved and its location does not change. The other option is to build a new gas station at the new location. In this case, it costs more money than the case of simply converting existing refueling station, but it may cover more drivers to refuel. Next, the traveling cost is experienced by the driver who makes an inter-city trip, and this cost is assumed to be linearly proportional to the distance the driver has traveled. Regarding the second objective of the Bapna et al. (2002) model, the objective function maximizes the drivers along the paths covered by the unleaded gas station, not the total traffic flow of OD pairs (Kuby and Lim 2005). Thus, their approach may not provide a global optimal solution from a whole system (i.e., the Indian national highway network) viewpoint.

2.2.1.3 Limitations of the Vertex-Based Demand Model

Even if the maximum covering model overcomes the weak point of the $p$-median model by considering the covering distance of demands, these two models are based on vertex-based demands models, and therefore, they have an inherent limitation in being applied to the refueling station location problem. Recall that a refueling service is usually provided in a discretionary way on the way to driver’s preplanned trip (Berman et al. 1992, Hodgson 1990). However, vertex-based demand models assume that driver’s refueling behavior is the one-stop trip; that is, a refueling itself is the goal of the trip, which does not reflect driver’s refueling behavior well. To consider the refueling behavior as a discretionary service activity on the way to driver’s preplanned trip, Hodgson (1990) and Berman et al. (1992) suggest a flow-based demand. In the
next section, we take a look at literature that uses a flow-based demand for refueling station
location models.

### 2.2.2 Flow-Based Demand Problems

A flow-based demand approach has been established to overcome the limitation of the vertex-
based demand approach in describing discretionary service activities on the way to preplanned
trips. According to the method to collect data about flow-based demands, flow-based demand
models are categorized as passing flow-based demand models and path flow-based demand
models. Data about passing flow-based demands are collected by counting traffic flows at each
ing the network. This data is easily obtained, but it has a multiple counting issue by counting
traffic flow for every edge. Thus, the optimal location by using passing flow-based demand may
incur cannibalization by locating the facilities excessively close to each other in the area with the
high traffic flows. To overcome this issue, data about path flow-based demands are collected by
investigating amounts of origin and destination for every path in the network. It is sometimes
difficult to obtain this data, but it constructs a global optimal location from a whole system
viewpoint. In this section, we briefly review the applications of both passing flow-based demand
models and path flow-based demand models to the refueling station location problems.

#### 2.2.2.1 Passing Flow-Based Demand Model

A passing flow is defined as the amount of traffic flows that pass an edge. This data is easily
obtained by counting traffic flows at each edge of the network. As an effort to overcome the
weakness of the vertex-based demand model in dealing with the network location problem for
discretionary service activities, passing flow-based demand models has been used. Goodchild and
Noronha (1987) combine passing flow-based demands with vertex-based demands which are weighted according to population size and suggest which refueling stations need to be kept open in order to maximize the market share after merging with others. It can be said that their model lies on the border between the vertex-based demand model and the flow-based demand model in that they mix both of two types of demand. Melendez and Milbrandt (2005) also develop a passing flow-based demand model to locate a minimum number of refueling stations that support wide usage of hydrogen vehicles in the interstate highways, on the assumption that a driving range and a maximum distance between two stations are known. Then, with real data about hydrogen vehicle demand for each sub-region (Melendez and Milbrandt 2006), Melendez and Milbrandt (2008) apply their methodology to the sub-regions in the United States – such as Atlanta, Northern California, Southern California, Chicago, Denver, Detroit, Minneapolis - St. Paul, Philadelphia - New York - Boston, Phoenix, St. Louis, East Texas, Salt Lake City, Seattle, Washington - Baltimore – in order to propose the optimal location of the hydrogen refueling stations for each sub-region.

Recall that flow-based demands are well-suited for describing discretionary service activities on the way to preplanned trips, and passing flow-based demands, which are one of the flow-based demand types, can be easily obtained by counting traffic flows at each edge. One big problem with the passing flow-based demand model is, however, that it incurs a multiple counting issue and may lead to cannibalization by locating the refueling stations excessively close to each other in the same area with high passing traffic flows (Ghosh and McLafferty 1987). This issue motivates the necessity of the path flow-based demand model in locating refueling stations. In the next section, we talk about the applications of path flow-based demand models to locate refueling stations in the road network.
2.2.2.2 Path Flow-Based Demand Model

A path flow is defined as the amount of traffic flows for each path (i.e., each OD pair) in the network. Thus, it basically prevents the multiple counting issue and allows the model to find the global optimal location from a whole system viewpoint. Path flow data can be easily obtained in the toll road network since every demand’s OD pair is collected at their origin and destination of the toll road network in order to charge a toll based on their OD. On the other hand, it is usually difficult to obtain the exact path flow data in the general road network because we cannot investigate every demand’s OD in the general road.

To overcome this issue in the general road network, one could estimate the flows of every OD pair or could use the distance of OD pairs instead of using the flow of OD pairs. Wang and Lin (2009) present the second case to use the distance matrix of OD pairs, not the flow matrix of OD pairs in their set covering model. Before taking a review of Wang and Lin (2009), recall the literature in Section 2.2.1.2 that Bapna et al. (2002) apply one of the covering problems, named the maximum covering/shortest path model, with vertex-based demand. The covering problem is useful for presenting the demand that has the covering distance to be served. But, the classical covering problem is not well-suited for the facility location problems in which the facility provides service activities in a discretionary way, e.g., refueling service, because a demand of the classical covering problem is vertex-based and is assumed to have a one-stop trip only for the service itself without considering a preplanned trip. On the other hand, for the discretionary service purchases, a flow-based demand is more preferred to reflect preplanned trips (Berman et al. 1992). To overcome this weak point of the classical covering problem, Wang and Lin (2009) propose the flow-based set covering model using the distance matrix of OD pairs, not the flow matrix of OD pairs, and find the optimal refueling station location to cover all alternative-fuel vehicles with a minimum cost. However, this model does not consider the maximum number of
refueling stations allowed. Thus, the optimal cost to cover all demands may exceed the budget (Daskin 2011).

While Wang and Lin (2009) use the distance matrix of OD pairs, many models use the flow of OD pairs, on the assumption that they can obtain or estimate these data, in order to develop the model that locates the optimal refueling stations location so as to maximize the total path flows from a whole system viewpoint. In particular, Kuby and Lim (2005) are the first who suggest the flow-refueling location model (FRLM) for the alternative-fuel vehicles. The FRLM is an extension of the FILM/FCLM (Berman et al. 1992, Hodgson 1990) to maximize the total path flows captured. The objective of the FRLM is to locate a given number of refueling stations at the vertex in the road network to maximize the total path flows (i.e., flows of OD pairs) refueled, and it is formulated as a mixed integer programming. The formulation of the FRLM is structurally identical to that of the FCLM, but the FRLM is different from FCLM in that FRLM considers a covering distance of a demand (i.e., driving range) while the FCLM does not. The FCLM counts a demand as captured if there exists only one facility anywhere along the path regardless of the distance of the path because it never considers a covering distance of a demand. On the other hand, the FRLM counts a demand as captured by a refueling station only if the station is located on a demand’s preplanned route and within a demand’s covering distance. This enables a demand to refuel multiple times at different locations if necessary. This implies that a driving range of the alternative-fuel vehicle is a critical factor to the optimal location of the refueling stations. Note that the covering distance concept of the FRLM is identical to that of the covering model reviewed in Section 2.1.3. In addition, recall from the literature in Section 2.1.4 that the formulation of the FCLM is also structurally identical to that of the MCLM, but the FCLM is distinct from the MCLM in that the FCLM is based on path-flow based demands while the MCLM is based on vertex-based demands. To sum up, the FRLM structurally belongs to the same family with the FCLM and the MCLM because the MCLM extends to the FCLM, and the
FCLM again extends to the FRLM. Furthermore, the FRLM has both the distinct characteristics of the FCLM and the MCLM at the same time. That is, the FRLM is path-flow based like the FCLM and considers a covering distance of demands like the MCLM. Figure 2-4 illustrates the characteristics of the FRLM graphically.

![Venn diagram showing the relationship between Maximum covering location model, Flow-refueling location model, and Flow capturing/intercepting location model.]

Figure 2-4. The characteristics of the flow-refueling location model

One more distinct property of the FRLM is that this model ensures a sustainable round trip after refueling. That is, every path flow refueled by the stations can be repeated the next day, or another round trip to a different destination can be completed in the FRLM. To ensure the sustainable round trip in the model, the FRLM assumes that the remaining range of an alternative-fuel vehicle both at the origin and destination is the half full tank if there is no station at the origin or destination. The half full tank assumption assures that, if a demand (i.e., a driver or a vehicle) can make it from the refueling station to the destination with at least the half full tank, it can return to a station on the same or different trip. Therefore, the main contribution of the FRLM is the use of refueling station combinations located at the vertices so as to maximize the total path flows refueled in the undirected road network.
Like the MCLM and the FCLM, the FRLM is also NP-hard because all of them are structurally identical to each other. In particular, the FRLM needs to generate all combinations of refueling stations to refuel demands. As a result, the FRLM is found to be computationally intractable to solve the refueling station location problem in large networks. To solve this issue, three heuristic algorithms such as greedy-adding, greedy-adding with substitution, and genetic algorithms have been developed to solve the FRLM problem (Lim and Kuby 2010). Also, these heuristics are applied to the metropolitan scale of Orlando and the statewide scale of Florida with real data to find the optimal location of initial refueling stations for each scale (Kuby et al. 2009). Upchurch and Kuby (2010) compare the $p$-median model with the FRLM for locating alternative-fuel refueling stations both at the metropolitan scale of Orlando and the statewide scale of Florida and find that the FRLM derives a better and more stable result than the $p$-median model does, especially at the state scale.

To avoid the high complexity of the FRLM in generating all the combinations of refueling stations to refuel demands, Capar and Kuby (2012) introduce an efficient mixed integer programming model that does not need pre-generation of combination as input data. For this, the authors add new variables that check availability of a refueling station for each path and feasibility of the set of solutions with limited driving range. To reduce the complexity, the authors also prioritize the candidate sites by checking previous solutions and they set a cutoff value. The main contribution of their model is that it provides an efficient way to solve the large-scale FRLM problems to optimality as well as or better than the three heuristic algorithms by Lim and Kuby (2010). However, their model needs building sets of candidate sites for every path, which would still result in the high complexity, prior to solving the FRLM problem.

As another effort to avoid combinations of refueling stations in the formulation and to solve the FRLM more efficiently, Capar et al. (2013) also suggest an edge cover-path-cover formulation of the generalized FRLM which is based on covering edges. Capar et al. (2013) is
similar to Capar and Kuby (2012) in that both of them do not need combinations of the refueling station. However, Capar et al. (2013) focuses on covering the edge of each path, while Capar and Kuby (2012) focus on covering the vertex of each path. For the generalized model, the authors assume that the starting fuel level of a vehicle is computed on the basis of the distance from the refueling station location. This assumption assures that, if there exists the refueling station at the origin vertex, then the starting fuel level is full, but if there exists no station at the origin, then the starting fuel level is the remaining fuel from the last refueling. To allow the FRLM to be solved without the use of combinations of candidate refueling station sites, the authors formulate that a flow is refueled only if each directional edge on a path is travelable after refueling at one of the available refueling stations. The cover set consists of candidate vertices that make a demand possible to complete the round trip. Jochem et al. (2015) use the Capar et al. (2013) model to solve the rapid electric charging station allocation problem on the German autobahn.

Since the FRLM has demonstrated superiority for the refueling station location problem, various extensions based on the FRLM have been published. Upchurch et al. (2009) suggest the capacitated FRLM in order to consider the finite refueling service at each refueling station. It releases the assumption of the original FRLM that a refueling station can provide infinite refueling service for all vehicles that pass through the station. For the capacitated model, the authors allow the decision variable for locating the refueling station at a vertex to have an integer value, while the original FRLM set it as a binary variable. Also, multiple combinations of refueling stations are required for every path in order to consider the possibility that one combination of the stations would not provide enough refueling service for all vehicles. Under these modified constraints, the capacitated FRLM locates the refueling stations with the objective of maximizing the total vehicle-miles traveled refueled by the set of stations. Then, it is validated with an application to an Arizona case study.
Kim and Kuby (2012) suggest the deviation-FRLM in order to consider the necessary deviations. Drivers usually make deviations from their preplanned trips in order to refuel their vehicles when there is no available refueling station on their preplanned trip within their remaining fuel level. Thus, the objective of the deviation-FRLM is to locate a given number of refueling stations so as to maximize the total path flows refueled at the stations on the preplanned trips or on the deviation paths. For describing the fraction of flows on deviation paths, the authors assume that the flow refueled at the station decreases as the distance of deviation increases. However, it is extremely hard to generate and solve the formulation of the deviation-FRLM with a realistic problem size. Thus, the heuristic algorithm, such as greedy-substitution heuristic, has been developed for the deviation-FRLM (Kim and Kuby 2013). Huang et al. (2015) suggest a multipath refueling location model, where multiple detours are available to drivers, with the objective of minimizing the total cost of allocating stations to cover all the traffic flow.

MirHassani and Ebrazi (2013) reformulate the refueling location model using an expanded network by adding two new vertices – such as a source (i.e., a virtual supply) and a sink (i.e., a virtual demand) – to the ends of the best path, based on the FRLM (Kuby and Lim 2005) and the flow-based set covering problem (Wang and Lin 2009). Their model has two main advantages, i.e., flexibility and efficiency. The first advantage is that their model is flexible in switching between the flow-based set covering problem with the objective of minimizing the cost of building refueling stations on all paths and the flow-based maximum covering problem (i.e., the FRLM) with the objective of maximizing the total path flow refueled by the station. The next advantage is that their model reduces the complexity in deriving an optimal solution both for the flow-based set covering problem and the FRLM because it does not require making the combinations of refueling stations for each path, which creates a huge complexity. Chung and Kwon (2015) apply the MirHassani and Ebrazi (2013) model to solve the multi-period electric charging station location problem on the Korean Expressway. Hosseini and MirHassani (2015)
also extend the MirHassani and Ebrazi (2013) model to consider uncertain refueling demand and two types of refueling stations (permanent and portable stations).

Hwang et al. (2015) suggest a refueling station location model on a symmetric transportation network with some candidate sites for single-access facilities covering traffic flow in one direction and other sites for dual-access facilities covering traffic flow in both directions, which is a distinctive situation in toll roads and some highways. Hwang et al. (2017a) expand the original Hwang et al. (2015) model to a version in which vehicles have different driving ranges and different fuel tank levels at their origins and destinations. Hwang et al. (2017b) further generalize the original model by allowing vehicle deviations and non-symmetric round trips between all OD pairs with positive flows.

Some refueling station location problems are bi-objective. Wang and Wang (2010) extend a set covering problem to a bi-criteria model for locating refueling stations. The two objectives of this model are minimizing the total cost of locating stations to cover all intra-city trips and maximizing the coverage for inter-city trips, using OD distance data. Brey et al. (2016) present a bi-objective model to search for the optimal set of the given number of refueling station locations. In their model, the two objectives are minimizing the average distance of the population to one of the nearest stations and maximizing the population refueled by the set of stations. Tu et al. (2016) use taxi global positioning system (GPS) data and propose a bi-objective model to solve the location problem for electric taxi charging stations. The two objectives of their model are minimizing the total charging wait time of the electric taxis and maximizing the total distance of these taxis traveled with customers. Note that none of these studies does not consider single-access refueling stations because these studies are based on undirected networks, where all stations can cover traffic flow in both directions (i.e., all refueling stations are dual-access). In addition, none of them does not address the construction cost of the refueling station for each candidate site well. Thus, none of these models is not well-suited for precisely estimating the
effect of investment in refueling infrastructure on greenhouse gas emissions reduction and relating the investment to the social cost of carbon savings on highway systems.

While the original FRLM (Kuby and Lim 2005) consider locating refueling stations only at the vertices in the road network, Kuby et al. (2005) mention that a network location problem can have better candidate sites along edges in capturing more path flows and suggest two added-node dispersion problems (ANDP) – such as the minimax ANDP model with the objective of minimizing the maximum sub-edge length and the maximin ANDP model with the objective of maximizing the minimum sub-arc length – for solving the continuous network location problem. With these two methods, Kuby and Lim (2007) develop a mid-path segment algorithm and compare the sample network test results of these three methods (i.e., the mid-path segment, the minimax ANDP, the maximum ANDP) for three different driving ranges. While the mid-path segment algorithm is based on the vehicle’s driving range, the other two ANDP methods do not count the driving range. Thus, the mid-path segment algorithm works better than either of the two ANDP methods for the small number of refueling stations and short driving ranges. However, the mid-path segment algorithm does not consider the combination of candidate segments for the trip that needs consecutive refueling but selects a single non-vertex candidate segment for each path independently. Thus, this method does not improve the optimal solution of the FRLM for larger numbers of refueling stations.
Chapter 3

ENERGY POLICY CONSIDERATIONS IN THE DESIGN OF AN ALTERNATIVE-FUEL REFUELING INFRASTRUCTURE TO REDUCE GREENHOUSE GAS EMISSIONS ON A TRANSPORTATION NETWORK

3.1 Introduction

In recent decades, science and technology (S&T) has progressed very rapidly, and dramatically contributed to the development and progress of people’s lives, the economy, and society. For instance, the development of transportation has significantly extended the range of human activities, and information and communications technology (I&CT) has minimized time and distance in communications. However, S&T has brought not only these benefits but also some detriments. To achieve such advances in S&T, developed countries have relied on fossil fuels, such as coal, oil, and natural gas, for generating energy. These energy resources have enabled large-scale economic activities, mass production, and global transportation. However, fossil fuels have a downside for humankind. Consumption of oil is responsible for greenhouse gas (GHG) emissions to the atmosphere, climate change, and air pollution. And because oil is a limited resource, it is subject to great price fluctuations. In this context, S&T must help make economic growth compatible with sustainability, and one recent challenge is to develop sources and supply infrastructures for alternative energy in transportation networks (Omi 2009).

Since the passage of the Air Pollution Control Act in 1955 (U.S. Congress 1955), the U.S. government has passed a number of laws and regulations to address the importance of environmental air quality and the ability of the institution to research the quantity of pollutants

---

emitted by stationary and mobile sources of air pollution. The subsequent passage of the Clean Air Act in 1963 (U.S. Congress 1963), as well as significant amendments to the Act in 1970, 1977, and 1990, expanded the government’s ability to research, monitor, control, and to establish standards to regulate the ongoing enforcement effort and the reduction of air pollutants. These national laws and standards are reinforced through international agreements as well, as the goal to reduce GHG emissions and prevent the consequences of climate change have been established by the 1992 United Nations (UN) Framework Convention on Climate Change (UN 1992) and the 1997 Kyoto Protocol (UN 1998). More recently, the 2015 Paris Agreement (UN 2015) extended the UN Framework to mitigate GHG emissions through adaptation to new sustainable development and the financing required to bring those strategies to reality. The legislative steps taken in these past decades are unsurprising given the wide-ranging research on, and well-documented impacts of, air pollution.

In this study, we investigate the reduction of GHG emissions for the case of establishing liquefied natural gas (LNG) refueling stations on the Pennsylvania (PA) Turnpike System, with the goal of converting Federal Highway Administration (FHWA) Classes 6-10 heavy-duty vehicles (classified by number of axles) (FHWA 2012), which correspond to the U.S. Environmental Protection Agency (EPA) Class 8a and 8b combination trucks (EPA and Department of Transportation (DoT) 2016, where class 6 vehicles correspond to EPA class 8a vehicles and the class 7-10 FHWA vehicles correspond to EPA class 8b) from their most popular current fuel, diesel, to LNG as an alternative-fuel. When considering the impact of GHG emissions from the transportation sector, we follow the lead of the EPA and National Highway Traffic Safety Administration (NHTSA) and focus on four main gases: carbon dioxide (CO₂), methane (CH₄), nitrous oxide (N₂O), and hydrofluorocarbons (HFCs). The first three gases are produced through the fuel cycle as byproducts of fossil fuel combustion and we will consider the impact of these emissions when evaluating the shift from diesel-fueled heavy-duty vehicles
(HDVs) to LNG-fueled HDVs. The fourth category of GHG, HFCs, is typically emitted from air conditioning units within the vehicles, and we will assume that any improvements that can be made to reduce the emissions from this source could be applied equally across trucks which consume any fuel type. The European Union has already begun to phase down the use of fluorinated gases through regulation (European Commission 2016). Furthermore, N$_2$O emissions can be mitigated using on-vehicle emission control technologies, and the EPA expects that both diesel and LNG vehicles can be fitted with the same emission control technologies, eliminating any significant difference between the N$_2$O emissions for the two fuel types (EPA 2003).

When discussing the impact of GHG emissions, it is common to use a scaling factor to compare the potential for different GHGs to have an impact on climate change. This factor is referred to as the global warming potential (GWP) for a gas, and the factor enables the comparison between the impact of a gas and the reference GHG, CO$_2$. This allows the measurement of GHG emissions to be standardized from the collection of emitted GHGs to a single measure of carbon dioxide equivalence (CO$_2$Eq). Table 3-1 shows the GWP values for the three main GHGs we will consider in this chapter for transportation emissions resulting from fossil fuel combustion.

<table>
<thead>
<tr>
<th>GHG</th>
<th>Lifetime in Atmosphere (years)</th>
<th>20 year$^1$ (scaled to CO$_2$Eq)</th>
<th>100 year$^1$ (scaled to CO$_2$Eq)</th>
<th>% Responsible for Man-Made Global Warming$^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carbon Dioxide (CO$_2$)</td>
<td>30-95</td>
<td>1</td>
<td>1</td>
<td>64%</td>
</tr>
<tr>
<td>Methane (CH$_4$)</td>
<td>12</td>
<td>72</td>
<td>25</td>
<td>17%</td>
</tr>
<tr>
<td>Nitrous Oxide (N$_2$O)</td>
<td>114</td>
<td>289</td>
<td>298</td>
<td>6%</td>
</tr>
</tbody>
</table>

$^1$EPA (2016a)  
$^2$European Commission (2016)

GHG emissions in the U.S. in 2014 totaled 6,870.5 million metric tons (MMT) CO$_2$Eq (EPA 2016a), with 76% of this total coming from the combustion of fossil fuels like coal, natural
gas, and petroleum products (diesel and gasoline), predominantly for energy, industrial, and transportation purposes. In 2014, the transportation sector emitted 1,810.3 MMT CO₂Eq, accounting for 26.3% of the total GHG inventory in the country for that year. Of this total, 95.9% (1,737.6 MMT CO₂Eq) was the result of burning fossil fuels like diesel (typically used in HDVs and long-haul trucks), gasoline (typically used in passenger cars and light trucks), compressed natural gas (CNG, typically used in vocational vehicles like local busing and delivery vehicles), and liquefied natural gas (LNG, which is an emerging fuel in this sector for long-haul trucks and HDVs and is the focus of this study). The transportation sector is one of the largest contributors of GHG emissions, trailing only the energy sector, which accounts for 30.3% of all GHG production in pursuit of electricity generation in stationary sources (EPA 2016a). Despite being only 4% of the registered vehicles on the road in the U.S., HDVs accounted for nearly 23% of the GHG emissions from transportation in the U.S. (White House 2014), so reducing the consumption of impactful GHG-emitting fuels from this sector would represent a significant overall reduction in GHG emissions.

The adoption of LNG as an alternative to diesel fuel offers the opportunity to reduce CO₂ emissions due to its lower carbon content, but natural gas, comprised mostly of methane, creates an additional problem that needs to be addressed. As shown in Table 3-1, 1 kg of methane emissions has the same GWP as 25 kg of CO₂ over a 100-year study period. Although methane spends less time in the atmosphere, the 25- and 100-year potentials suggest that methane is arguably a more hazardous gas to emit in the near term. Fugitive emissions from vehicle drivetrains, refueling operations, and boil-off of LNG expanding to CNG in a fuel tank may offset the CO₂ reduction potential of LNG by contributing more impactful GHG emissions from methane (EPA and DoT 2016). We will return to this consideration in Section 3.3, where we will estimate the total GHG emissions impacts and in Section 3.4, where we will suggest policy and mitigation strategies to address this concern.
Current literature offers an avenue to relate the location of alternative-fuel refueling stations to the problem of mitigating the effects of GHG emissions from transportation sources. In this chapter, we adopt the modeling framework proposed by Hwang et al. (2015) in the formulation of a bi-criteria binary linear programming model for locating alternative-fuel refueling stations on a directed transportation network, where the objectives are maximizing the total vehicle-miles traveled (VMT) covered by the stations and minimizing the capital cost of constructing the refueling infrastructure. The proposed model is applied to the PA Turnpike concerning the development of an LNG refueling infrastructure considering service plazas as potential station locations. In general, the construction cost of an LNG station depends on construction worker wages, land value, additional paved area to access the station, and length of a pipeline connecting the LNG station with the conventional fuel station in the service plaza. Two versions of the model are implemented and compared. In the first version, the normalized weighting method is used to combine the two objectives. In the second version, the second objective is relegated to the constraint set and handled as an infrastructure budget constraint. In the proposed model, instead of considering the number of stations as an input parameter as is often done in the literature, either the weight assigned to the capital cost in the objective function of the first version of the model or the infrastructure budget is the specified parameter. That is, in the first version, the weight assigned to the capital cost in the objective function is the specified parameter and in the second version of the model the infrastructure budget is the known parameter.

In the succeeding sections of this chapter, Section 3.2 describes the development of the two versions of the proposed model with the goals of maximizing VMT and minimizing infrastructure cost, where either the two objectives are consolidated into a single one through a weighting method or the second objective is relegated to the constraint set as a budget constraint. Section 3.3 presents the case study on the PA Turnpike, where we analyze the resulting LNG
refueling station locations for a variety of construction budgets while maximizing VMT and estimating the GHG emissions changes for CO₂ and CH₄. The construction costs for the candidate LNG refueling stations, which are optimally located at selected service plazas on the turnpike, are estimated using a regression model developed with estimates from a previous study (Myers et al. 2013). In Section 3.4, we draw conclusions and recommend policy actions that may be taken to advance the development of a natural gas refueling infrastructure and potential research and development targets to maximize the environmental benefits realized through the transition to LNG vehicles.

3.2 Location Optimization Model

This section presents a bi-criteria binary linear programming model to locate alternative-fuel refueling stations on a directed transportation network with the objectives of maximizing the total VMT that can be covered by the stations and minimizing the capital cost for setting up the refueling infrastructure. Demand in this study is treated as the traffic flow between origin and destination interchange pairs; thus, the traffic flow rates for all OD pairs are input data for the proposed model.

To precisely formulate and solve the problem, the following assumptions are made. First, the transportation network, denoted as G, in this study is connected and symmetric, so that edges are separated into two driving directions. In addition, vehicles can enter and exit the network only at interchanges and are not allowed to exit the network for the purpose of refueling during the trip. Second, there exist a finite number of candidate refueling station locations in the network; prospective stations in some candidate locations can only provide refueling service to vehicles on one side of the road (single-access stations) and potential stations in other sites can service vehicles on both sides (dual-access stations). Third, construction costs of candidate refueling
stations are predetermined and stations have infinite capacities, i.e., an incoming vehicle to a station will always be able to fully refuel. Four, all vehicles have the same limited driving range per refueling, denoted as $R$, which is not affected by road condition, topography, or any other endogenous or exogenous factors. Fifth, vehicles make complete (symmetric) round trips between their origin and destination interchanges using their shortest paths without considering deviation options. Note that, if there are multiple shortest paths for a given OD pair, then we arbitrarily choose one of the paths. A round trip between a pair of origin and destination interchanges consists of an original trip from the origin to the destination and a return trip from the destination to the origin. Sixth, all vehicles are assumed to have their fuel tank level at least half-full every time they enter the network at their origin interchange in the original trip or at their destination interchange in the return trip. The coverage conditions for traffic flow between an OD pair depend on the length of the path. If the length of the path does not exceed $R/4$, then we consider the trip to be short and traffic flow is covered if drivers can refuel their vehicles at least once throughout their round trip. Otherwise, if the length of the path exceeds $R/4$, we consider the trip to be medium-to-long and traffic flow is covered if drivers can refuel their vehicles along both the original trip and return trip, under the conditions that fuel tanks are at least half-full each time they enter and also when they exist the network.

The next sections present the notation, bi-objective functions and constraints used in the model, as well as the solution approaches.
3.2.1 Notation

Sets:

- \( G \) transportation network, i.e., \( G = (N, A) \), which is connected, symmetric, and divided into two driving directions.
- \( N \) set of nodes on \( G \), i.e., \( N = \{n_1, ..., n_l\} \), which consists of two different types of nodes, i.e., \( N = P \cup K \).
- \( P \) set of all interchanges on \( G \), i.e., \( P = \{p_1, ..., p_s\} \).
- \( K \) set of all candidate sites for refueling stations on \( G \), i.e., \( K = \{k_1, ..., k_h\} \).
- \( A \) set of arcs on \( G \), i.e., \( A = \{(n_u, n_v) \mid \text{for some } n_u, n_v \in N\} \).
- \( Q \) set of all OD pairs in \( G \), i.e., \( Q = \{(p_o, p_d) \mid p_o, p_d \in P, o < d\} \).

Parameters:

- \( R \) vehicle’s limited driving range (in miles) per refueling,
- \( P(n_u, n_v) \) original trip from \( n_u \) to \( n_v \) between \( n_u \in N \) and \( n_v \in N \),
- \( P(n_v, n_u) \) return trip from \( n_v \) to \( n_u \) between \( n_u \in N \) and \( n_v \in N \),
- \( d(n_u, n_v) \) shortest distance (in miles) between node \( n_u \in N \) and node \( n_v \in N \). Since \( G \) is symmetric, \( d(n_u, n_v) = d(n_v, n_u) \).
- \( f(p_o, p_d) \) average traffic flow (in round trips per year) on the shortest path between \( p_o \) and \( p_d \); that is, the average traffic flow in the original trip from \( p_o \) to \( p_d \) and the return trip from \( p_d \) to \( p_o \), and is defined only for \( o < d \),
- \( c_{k_u} \) capital cost (in million dollars ($M)) for building a refueling station in candidate site \( k_u \in K \),
- \( B \) refueling infrastructure budget (in $M).
**Decision Variables:**

- \( z_1 \) total VMT (in miles per year) covered by refueling stations in selected candidate sites,
- \( z_2 \) total capital cost (in $M) of setting up refueling stations in selected candidate sites,
- \( y_{p_o p_d} \) \( \begin{cases} 1, & \text{if traffic flow in round trip } (p_o, p_d) \in Q \text{ is covered by new refueling stations,} \\ 0, & \text{otherwise,} \end{cases} \)
- \( x_{k_u} \) \( \begin{cases} 1, & \text{if a refueling station is located at candidate site } k_u \in K, \\ 0, & \text{otherwise.} \end{cases} \)

**3.2.2 Bi-objective Functions**

The proposed model considers two conflicting objective functions. One objective is to maximize the total VMT covered by the refueling stations and the other objective is to minimize the total capital cost of setting up the selected refueling stations.

In the first objective function, for OD pair \((p_o, p_d) \in Q\), the corresponding VMT (in miles per year) is calculated by multiplying the average traffic flow \(f(p_o, p_d)\) (in round trips per year) by the shortest round trip distance (in miles), which is calculated as \(2 \times d(p_o, p_d)\) due to the symmetric nature of the distance, i.e., \(d(p_o, p_d) = d(p_d, p_o)\), for OD pair \((p_o, p_d) \in Q\). Then, based on this objective, denoted as Objective (3.1), the model determines the values of binary variables \(y_{p_o p_d}\), for all \((p_o, p_d) \in Q\), that maximize the sum of all covered VMT:

\[
\text{maximize } z_1 = \sum_{(p_o, p_d) \in Q} 2f(p_o, p_d) d(p_o, p_d) y_{p_o p_d}. \tag{3.1}
\]

In the second objective function, the construction cost of setting up a refueling station for candidate site \(k_u \in K\) is denoted as \(c_{k_u}\). Thus, based on this objective, denoted as Objective (3.2), the model assigns the values of binary variables \(x_{k_u}\), for all \(k_u \in K\), that minimize the total infrastructure cost for constructing the refueling infrastructure:
minimize $z_2 = \sum_{k_u \in K} c_{k_u} x_{k_u}$. \hfill (3.2)

### 3.2.3 Constraints

In the proposed model, a vehicle’s remaining driving distance to the next candidate sites for refueling stations is a key restriction to be considered. By the sixth assumption, every vehicle entering the network has at least a half-full tank and, if a vehicle makes a medium-to-long trip, when the vehicle leaves the network it must also have at least a half-full tank. These conditions require that some refueling stations must be located within a distance of $R/2$ from its origin and other stations within the same distance from its destination on both the original and return trips. These candidate sites are called end sites. In addition, some intermediate candidate sites may be required to provide refueling service every $R$ miles. For OD pair $(p_o, p_d) \in Q$, the sets of end candidate sites at $p_o$ and $p_d$ on the original and return trips are denoted as $E_{1,i}(p_o, p_d)$ and $E_{2,i}(p_o, p_d)$, for $i = 0, 1$, where $i = 0$ indicates an original trip and $i = 1$ an return trip. If $d(p_o, p_d) > R$, candidate sites that belong to neither $E_{1,i}(p_o, p_d)$ nor $E_{2,i}(p_o, p_d)$ are called intermediate sites, and the sets of intermediate sites on the original and return trips are denoted as $I_i(p_o, p_d)$, for $i = 0, 1$. These sets are defined as follows:

**Sets of end and intermediate candidate sites:**

- $E_{1,0}(p_o, p_d) = \{ k_u \in K \mid k_u \in P(p_o, p_d) \text{ and } d(p_o, k_u) \leq R/2 \}$,
- $E_{1,1}(p_o, p_d) = \{ k_u \in K \mid k_u \in P(p_d, p_o) \text{ and } d(p_d, k_u) \leq R/2 \}$,
- $E_{2,0}(p_o, p_d) = \{ k_u \in K \mid k_u \in P(p_o, p_d) \text{ and } d(k_u, p_d) \leq R/2 \}$,
- $E_{2,1}(p_o, p_d) = \{ k_u \in K \mid k_u \in P(p_d, p_o) \text{ and } d(k_u, p_o) \leq R/2 \}$,
- $I_0(p_o, p_d) = \{ k_u \in K \mid k_u \in P(p_o, p_d), d(p_o, k_u) > R/2, \text{ and } d(k_u, p_d) > R/2 \}$,
- $I_1(p_o, p_d) = \{ k_u \in K \mid k_u \in P(p_d, p_o), d(p_d, k_u) > R/2, \text{ and } d(k_u, p_o) > R/2 \}$.
Since vehicles have a limited driving range $R$ per refueling, traffic flow in OD pair $(p_o, p_d) \in Q$ has different coverage requirements depending on their travel distance $d(p_o, p_d)$. According to $R$ and $d(p_o, p_d)$, we partition the set of OD pairs $Q$ into three subsets of OD pairs to define the coverage conditions for short, medium, and long trips. The main difference between medium and long trips is that coverage for medium trips only requires end candidate sites while coverage for long trips require both end and intermediate candidate sites.

Three subsets of OD pairs:

- OD pairs for short trips: $Q^{(1)} = \{(p_o, p_d) \in Q: 0 < d(p_o, p_d) \leq R/4\}$,
- OD pairs for medium trips: $Q^{(2)} = \{(p_o, p_d) \in Q: R/4 < d(p_o, p_d) \leq R\}$,
- OD pairs for long trips: $Q^{(3)} = \{(p_o, p_d) \in Q: R < d(p_o, p_d)\}$.

When vehicles take a short trip in OD pair $(p_o, p_d) \in Q^{(1)}$, by the sixth assumption, a single refueling stop during the round trip is sufficient to cover the trip. For OD pair $(p_o, p_d) \in Q^{(1)}$, since vehicles are able to travel up to $R/2$ miles when they enter the network, and $d(p_o, p_d) \leq R/4$, $E_{1,i}(p_o, p_d) = E_{2,i}(p_o, p_d)$, for $i = 0, 1$, and they can be refueled at any selected candidate site either on their original or return trip. Thus, the sets of candidate sites on the original and return trips for OD pair $(p_o, p_d) \in Q^{(1)}$, denoted as $E_{i}^{(1)}(p_o, p_d)$, for $i = 0, 1$, are defined as follows:

$$E_{i}^{(1)}(p_o, p_d) = \{k_u \in E_{1,i}(p_o, p_d) \mid (p_o, p_d) \in Q^{(1)}\}, \text{ for } i = 0, 1.$$

Then, in Constraint set (3.3), if at least one candidate site is chosen for a refueling station on the round trip between OD pair $(p_o, p_d) \in Q^{(1)}$, the corresponding round trip is covered.

$$\sum_{k_u \in \bigcup_{i=0,1} E_{i}^{(1)}(p_o,p_d)} x_{k_u} \geq y_{p_o p_d} \quad \forall (p_o, p_d) \in Q^{(1)}.$$  (3.3)
When vehicles make a medium-to-long trip in OD pair \((p_o, p_d) \in Q^{(2)} \cup Q^{(3)}\), by the sixth assumption, at least one candidate site needs to be selected within a distance of \(R/2\) from \(p_o\) and another candidate site within the same distance from \(p_d\), on both their original and return trips. For OD pair \((p_o, p_d) \in Q^{(t)}\), for \(t = 2, 3\), the sets of end candidate sites of \(E_{c, l}(p_o, p_d)\), for \(c = 1, 2\), for \(i = 0, 1\), are denoted as \(E^{(t)}_{c, l}(p_o, p_d)\) and determined as follows:

\[
E^{(t)}_{c, l}(p_o, p_d) = \{k_u \in E_{c, l}(p_o, p_d) \mid (p_o, p_d) \in Q^{(t)}\}, \text{ for } t = 2, 3, \text{ for } c = 1, 2, \text{ for } i = 0, 1.
\]

Using \(E^{(t)}_{c, l}(p_o, p_d)\), Constraint set (3.4) imposes two conditions for the sets of end candidate sites for \(p_o\) and \(p_d\) on the original and return trips between OD pair \((p_o, p_d) \in Q^{(t)}\), for \(t = 2, 3\). This constraint set enforces that either two stations in \(E^{(t)}_{1, i}(p_o, p_d)\) and \(E^{(t)}_{2, i}(p_o, p_d)\), or one station in \(E^{(t)}_{1, i}(p_o, p_d) \cap E^{(t)}_{2, i}(p_o, p_d)\) for \(i = 0, 1\), is selected to cover the round trip between OD pair \((p_o, p_d) \in Q^{(t)}\), for \(t = 2, 3\).

\[
\sum_{k_u \in E^{(t)}_{c, l}(p_o, p_d)} x_{k_u} \geq y_{p_o p_d} \quad \forall (p_o, p_d) \in Q^{(t)}; \ t = 2, 3; \ c = 1, 2; \ i = 0, 1. \tag{3.4}
\]

Constraint set (3.4) is enough to identify the sets of refueling stations that can cover a medium trip in OD pair \((p_o, p_d) \in Q^{(2)}\). However, when we consider a long trip in OD pair \((p_o, p_d) \in Q^{(3)}\), additional conditions are required to include intermediate sites in sets \(l_i(p_o, p_d)\), for \(i = 0, 1\), in the original and return trips, respectively. To identify pairs of two candidate sites \(k_u, k_v \in K\), such that \(k_v\) follows \(k_u\) along the path and \(d(k_u, k_v) \leq R\), we first define two sets of candidate sites for refueling stations on the original and return trips, respectively:

\[
S^{(3)}_{1, i}(p_o, p_d) = \{k_u \in E^{(3)}_{1, i}(p_o, p_d) \cup l_i(p_o, p_d) \cup E^{(3)}_{2, i}(p_o, p_d)\}, \text{ for } i = 0, 1,
\]

\[
S^{(3)}_{2, i}(p_o, p_d) = \{k_v \in l_i(p_o, p_d) \cup E^{(3)}_{2, i}(p_o, p_d)\}, \text{ for } i = 0, 1.
\]
Consider two candidate sites \( k_u \) and \( k_v \), such that \( k_u \in S_{1,i}^{(3)}(p_o, p_d) \) and \( k_v \in S_{2,i}^{(3)}(p_o, p_d) \), in either the original trip \((i = 0)\) or the return trip \((i = 1)\). In order to represent mathematically whether a vehicle refueled at \( k_u \) is able to reach \( k_v \), we define two matrices, one for the original trip and the other for the return trip, for each OD pair \((p_o, p_d) \in Q^{(3)}\). The dimensions of these matrices are \( |S_{1,i}^{(3)}(p_o, p_d)| \times |S_{2,i}^{(3)}(p_o, p_d)| \), \( i = 0, 1 \), and their elements \( e_{k_u,k_v} \) for \( k_u \in S_{1,i}^{(3)}(p_o, p_d) \) and \( k_v \in S_{2,i}^{(3)}(p_o, p_d) \), for \( i = 0, 1 \), are binary; if \( k_v \) follows \( k_u \) along the path and \( d(k_u,k_v) \leq R \), then \( e_{k_u,k_v} = 1 \); otherwise, \( e_{k_u,k_v} = 0 \). Constraint set (3.5) searches for all pairs of two stations within the limited driving range \( R \) on the original and return trips in OD pair \((p_o, p_d) \in Q^{(3)}\). Thus, the set of refueling stations that satisfy Constraint sets (3.4) and (3.5) is able to cover a long trip between OD pair \((p_o, p_d) \in Q^{(3)}\).

\[
\sum_{k_u \in S_{1,i}^{(3)}(p_o, p_d)} e_{k_u,k_v} x_{k_u} \geq y_{p_o p_d} \quad \forall k_v \in S_{2,i}^{(3)}(p_o, p_d); \forall (p_o, p_d) \in Q^{(3)}; \ i = 0, 1. \tag{3.5}
\]

Finally, Constraint set (3.6) ensures that all the decision variables are binary.

\[
x_{k_u} \in [0, 1], \forall k_u \in K; \quad y_{p_o p_d} \in [0, 1], \forall (p_o, p_d) \in Q. \tag{3.6}
\]

### 3.2.4 Solution Approach

The proposed model is a bi-objective model, where the purpose is to identify trade-off solutions using multi-criteria optimization techniques. One popular technique is to use the normalized weighting method where each of the objectives is normalized by their own maximum magnitude and then the normalized terms are consolidated into one using non-negative weights that sum to one (Cohon 2013, Ehrigott 2006); in our model, we first find the maximum values, \( \alpha \) and \( \beta \), for each of the two objectives individually, then normalize them by dividing the objectives by \( \alpha \) and \( \beta \), respectively, and finally sum up all weighted normalized terms as one dimensionless objective.
using weights \( w \) and \( 1 - w \), for \( w \in [0, 1] \). Using this method, Objectives (3.1) and (3.2) are normalized and combined into Objective (3.7) in model M1 with Constraint sets (3.3) – (3.6).

Varying the weight \( w \) in Objective (3.7) from 0 to 1, M1 can identify different trade-off solutions.

\[
(M1) \quad \text{maximize} \quad \frac{w}{\alpha} \sum_{(p_o, p_d) \in Q} 2 f(p_o, p_d) d(p_o, p_d) y_{p_o, p_d} - \frac{(1 - w)}{\beta} \sum_{k_u \in K} c_{k_u} x_{k_u}, \\
\text{subject to} \quad \text{Constraint sets (3.3) - (3.6)}. 
\]  

Another popular technique to solve bi-objective problems is the constraint method. By converting one objective into a constraint to the model, the other objective can be optimized subject to this additional constraint (Cohon 2013, Ehrgott 2006). Assuming that \( B \) is the known parameter for the infrastructure budget capital cost, Objective (3.2) can be transformed into Constraint (3.8), which enforces the model to select candidate sites for refueling stations within the budget.

\[
\sum_{k_u \in K} c_{k_u} x_{k_u} \leq B. 
\]  

By iterating over all selected values of \( B \) in Constraint (3.8), model M2 having a single Objective (3.1) with Constraint sets (3.3) – (3.6) and (3.8) is able to find different trade-off optimal solutions.

\[
(M2) \quad \text{maximize} \quad \sum_{(p_o, p_d) \in Q} 2 f(p_o, p_d) d(p_o, p_d) y_{p_o, p_d}, \\
\text{subject to} \quad \text{Constraint sets (3.3) - (3.6), (3.8)}. 
\]
3.3 Case Study

In this section, we apply the two proposed models, M1 and M2, to the PA Turnpike with the annual traffic flow data from 2011 to determine the optimal locations for LNG refueling stations and evaluate the GHG emission savings for our optimal solutions.

The remainder of this section is organized as follows. In Section 3.3.1, we briefly summarize the PA Turnpike. In Section 3.3.2, we estimate construction costs for LNG refueling stations on the turnpike. In Section 3.3.3, we run the two proposed models (M1 and M2) to determine the optimal locations for LNG refueling stations on the turnpike, and also compare their results with Hwang et al. (2015) model. In Section 3.3.4, we finally analyze the estimated GHG emission savings from trucks using the PA Turnpike.

3.3.1 Pennsylvania Turnpike

The PA Turnpike road network comprises two main segments, the East-West mainline between Pittsburgh and Philadelphia (I-70, I-76, and I-276) and the Northeast extension between Philadelphia and Scranton (I-476). This chapter refers to these two segments as the PA Turnpike, as shown in Figure 3-1, which forms a tree network with 43 interchanges with ramps, where vehicles can enter or exit the network, of which 36 of the 43 interchanges also have toll plazas. There are 5 additional toll plazas without ramps in the travel lanes of the turnpike, where vehicles cannot enter or exit the network and do not affect the data used in this study. Throughout the remainder of this chapter, when we refer to the interchanges on the network, we are referring only to the 43 interchanges where vehicles can enter or exit the network. Table 3-2 provides information about the interchanges and the toll plazas on the turnpike. The distance between Pittsburgh and Scranton is approximately 400 miles, which is the longest distance between any
two points on the turnpike. There are 17 currently open service plazas, and two additional service plazas which are temporarily or permanently closed (Myers et al. 2013). All service plazas have conventional refueling stations for diesel fuel and unleaded gasoline. The PA Turnpike has two types of service plazas: single-access plazas, which provide refueling service only to vehicles traveling on the same side of the roadway where the plaza is located, and dual-access plazas, which provide refueling service to vehicles traveling on both directions of the roadway. In this case study, we consider 19 candidate sites for the LNG refueling stations, including the 17 open service plazas and 2 temporarily closed service plazas: 3 service plazas are dual-access and the other 16 service plazas are single-access. Figure 3-1 and Table 3-3 provide information about the locations of interchanges and active service plazas, as well as code information for each service plaza.

![Figure 3-1. PA Turnpike with 48 interchanges and 19 service plazas (Hwang et al. 2017a)](image)

On the PA Turnpike network, some vehicles enter and exit the network without passing through any refueling station. A trip that can be made without encountering any service plaza, it is
referred to as an uncoverable trip. We eliminate these uncoverable trips in the PA Turnpike by reducing the 43 original interchanges down to 19 aggregated interchanges by consolidating each longest consecutive sub-sequence of interchanges without any service plaza within the sub-sequence into a single interchange. The location of each aggregated interchange is calculated by using distance-weighted means, where the interchange weights for each sub-sequence are the annual in/out traffic counts (see Hwang et al (2015) for further details). In addition, we reduce the original OD matrix of annual truck trips provided in the 2011 PA Turnpike Commission Report down to a 19×19 aggregated matrix of truck trips using only aggregated interchanges. Table 3-2 summarizes information about the aggregated interchanges.

Table 3-2. List of the interchanges, toll plazas, and aggregated interchanges on the PA Turnpike

<table>
<thead>
<tr>
<th>Aggregated Interchange</th>
<th>Original Interchanges/ Toll Plazas</th>
<th>Aggregated Interchange</th>
<th>Original Interchanges/ Toll Plazas</th>
</tr>
</thead>
<tbody>
<tr>
<td>p₁</td>
<td>T2, T10, T13</td>
<td>p₁₁</td>
<td>T312, T320*</td>
</tr>
<tr>
<td>p₂</td>
<td>T28, T30*, T39, T48</td>
<td>p₁₂</td>
<td>T326</td>
</tr>
<tr>
<td>p₃</td>
<td>T57, T67, T75</td>
<td>p₁₃</td>
<td>T333</td>
</tr>
<tr>
<td>p₄</td>
<td>T91, T110</td>
<td>p₁₄</td>
<td>A20</td>
</tr>
<tr>
<td>p₅</td>
<td>T146</td>
<td>p₁₅</td>
<td>T339, T340, T343, T351</td>
</tr>
<tr>
<td>p₆</td>
<td>T161</td>
<td>p₁₆</td>
<td>T352, T358, T359</td>
</tr>
<tr>
<td>p₇</td>
<td>T180, T189, T201</td>
<td>p₁₇</td>
<td>A31, A44</td>
</tr>
<tr>
<td>p₈</td>
<td>T226, T236, T242, T247</td>
<td>p₁₈</td>
<td>A56, A74</td>
</tr>
<tr>
<td>p₉</td>
<td>T266, T286</td>
<td>p₁₉</td>
<td>A95, A105, A112*, A115, A121*, A122, A130*, A131</td>
</tr>
<tr>
<td>p₁₀</td>
<td>T298</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Toll plaza without entry and exit ramps

The number of round trips between each pair of interchanges is determined as the average of the number of trips in each direction. By converting the 2011 annual in/out truck traffic counts (in round trips per year) into the 2011 annual truck VMT, about 4.66% of the annual VMT correspond to vehicles that have no chance of passing through any service plazas on the turnpike.
because there are no service plazas within the sub-sequence of interchanges in their path. In contrast, 847.84 million of the 889.31 million of annual VMT (95.34%) pass by at least one service plaza on the PA Turnpike. In our analysis, we denote as *effective VMT coverage* the proportion of truck VMT per year that is covered with respect to the 847.84 million miles of annual VMT that can be covered by placing LNG refueling stations in all 19 candidate sites.

### 3.3.2 Regression Model to Estimate Construction Costs of Refueling Stations

In this section, we describe the methodology used to estimate the construction costs of the LNG refueling stations at the 19 pre-selected service plazas in the PA Turnpike. Myers et al. (2013) provided cost estimates for 8 of the service plazas in Table 3-3 without a “*” in the last column, based on a competitive bidding process, including the cost of the LNG refueling equipment, storage, dispensers, canopy, and site-related amenities. The service plaza real estate is owned by the PA Turnpike Commission (PTC), so land lease related to constructing an LNG refueling station was not considered when estimating construction costs.

Based on the estimated construction costs for the LNG refueling stations at the 8 service plazas in Table 3-3, we need to determine the construction costs for the 11 remaining service plazas to be able to run Models M1 and M2 to determine the optimal set of station locations considering the VMT coverage and available infrastructure budget. Hence, we decided to use the estimates from Myers et al. (2013) to create a linear regression model to estimate the construction cost for any new LNG refueling stations at the remaining service plazas along the turnpike. All new LNG refueling stations are assumed to have the same area, which is 3200 ft². Myers et al. (2013) considered one to three possible locations to setup an LNG refueling station at the 8 service plaza they studied. For the 11 remaining service plazas, we used service plaza maps to select appropriate locations for the LNG refueling stations. Then, we estimate the size of the new
pavement area that would be required to access the station and construct the parking area for LNG trucks as well as the length of the required natural gas supply line from the conventional station to the new LNG station for the 17 currently open service plazas. For the two temporarily closed single-access service plazas, Zelienople and North Neshaminy, we used the mean values of the corresponding two nearby single-access operational service plazas to estimate their parameters. We used estimates from the Oakmont-Plum and New Stanton service plazas for Zelienople, and the estimates from Valley Forge and King of Prussia service plazas for North Neshaminy. Furthermore, we augmented the estimates for the closed service plazas by doubling the cost predicted by the regression model as a conservative estimate on the cost of both installing LNG infrastructure and preparing the station for public servicing.

The linear regression model is developed with the available information from the 8 service plazas estimated in the Myers et al. (2013) feasibility study. Note that, when multiple construction options are provided by Myers et al. (2013), we pick the one that better fits the linear regression model. In the regression model, the dependent variable is the construction cost (in $M), and the independent variables are the length of the natural gas supply line (in ft), pavement size (in ft²), construction labor cost (in $/year), estimated for each locality using PA Center for Workforce Information & Analysis (Pennsylvania Center for Workforce Information & Analysis, 2016) data, and real estate cost (in $/ft²), estimated for each locality using Trulia databases (Trulia 2016). The estimates for these independent factors are shown in Table 3-3. The results from the first regression model in Figure 3-2(a) show that the real estate cost is negative (-0.0006262) with Adjusted R-squared value of 0.8980, so we remove this factor from the model and fit a new model where the real estate cost is considered a fixed value. Then, we fit a second regression model using the remaining three variables as independent variables, with Adjusted R-squared value of 0.9231. In the results for this model shown in Figure 3-2(b), we see that the construction labor cost is not a significant predictor of the LNG facility construction cost at the service plazas,
Table 3-3. Parameters and cost estimates for the 19 candidate station locations (service plazas)

<table>
<thead>
<tr>
<th>Code</th>
<th>Service Plaza</th>
<th>Milepost</th>
<th>Natural Gas Supply Line Length (ft)</th>
<th>Pavement Size (ft²)</th>
<th>Construction Labor Cost ($/year)</th>
<th>Real Estate Cost ($/ft²)</th>
<th>LNG Infrastructure Development Cost ($M)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZEN</td>
<td>Zelienople (closed)</td>
<td>T21.70 EB</td>
<td>669.65</td>
<td>0</td>
<td>43,910</td>
<td>74</td>
<td>4.10*</td>
</tr>
<tr>
<td>OMP</td>
<td>Oakmont Plum</td>
<td>T49.30 EB</td>
<td>208.40</td>
<td>0</td>
<td>50,030</td>
<td>65</td>
<td>1.80</td>
</tr>
<tr>
<td>NST</td>
<td>New Stanton</td>
<td>T77.60 WB</td>
<td>1130.90</td>
<td>0</td>
<td>44,900</td>
<td>87</td>
<td>2.30</td>
</tr>
<tr>
<td>SSS</td>
<td>South Somerset</td>
<td>T112.30 EB</td>
<td>292.70</td>
<td>20800</td>
<td>37,900</td>
<td>85</td>
<td>3.60*</td>
</tr>
<tr>
<td>NSS</td>
<td>North Somerset</td>
<td>112.37 WB</td>
<td>339.30</td>
<td>19200</td>
<td>37,900</td>
<td>85</td>
<td>3.50*</td>
</tr>
<tr>
<td>SMW</td>
<td>South Midway</td>
<td>T147.31 EB</td>
<td>218.20</td>
<td>0</td>
<td>40,050</td>
<td>85</td>
<td>1.80</td>
</tr>
<tr>
<td>NMW</td>
<td>North Midway</td>
<td>T147.32 WB</td>
<td>409.80</td>
<td>0</td>
<td>40,050</td>
<td>85</td>
<td>1.80</td>
</tr>
<tr>
<td>SLH</td>
<td>Sideling Hill</td>
<td>T172.27 EB/WB</td>
<td>199.60</td>
<td>22400</td>
<td>47,010</td>
<td>76</td>
<td>3.40</td>
</tr>
<tr>
<td>BMT</td>
<td>Blue Mountain</td>
<td>T202.63 WB</td>
<td>392.90</td>
<td>25600</td>
<td>45,740</td>
<td>87</td>
<td>4.10*</td>
</tr>
<tr>
<td>CLV</td>
<td>Cumberland Valley</td>
<td>T219.12 EB</td>
<td>359.20</td>
<td>9600</td>
<td>45,740</td>
<td>108</td>
<td>2.70*</td>
</tr>
<tr>
<td>HSP</td>
<td>Highspire</td>
<td>T249.70 EB</td>
<td>159.60</td>
<td>12800</td>
<td>46,930</td>
<td>89</td>
<td>2.80*</td>
</tr>
<tr>
<td>LAW</td>
<td>Lawn</td>
<td>T258.77 WB</td>
<td>335.40</td>
<td>6400</td>
<td>40,090</td>
<td>116</td>
<td>2.40*</td>
</tr>
<tr>
<td>BMV</td>
<td>Bowmansville</td>
<td>T289.87 EB</td>
<td>399.10</td>
<td>12800</td>
<td>43,000</td>
<td>119</td>
<td>3.00*</td>
</tr>
<tr>
<td>PJC</td>
<td>Peter J. Camiel</td>
<td>T304.84 WB</td>
<td>300.10</td>
<td>0</td>
<td>53,410</td>
<td>144</td>
<td>1.80</td>
</tr>
<tr>
<td>VFG</td>
<td>Valley Forge</td>
<td>T324.55 EB</td>
<td>162.20</td>
<td>12800</td>
<td>58,590</td>
<td>232</td>
<td>2.80*</td>
</tr>
<tr>
<td>KPR</td>
<td>King of Prussia</td>
<td>T328.40 WB</td>
<td>443.50</td>
<td>0</td>
<td>57,140</td>
<td>136</td>
<td>1.90</td>
</tr>
<tr>
<td>NNM</td>
<td>North Neshaminy</td>
<td>T351.90 WB</td>
<td>302.85</td>
<td>6400</td>
<td>52,850</td>
<td>158</td>
<td>4.70*</td>
</tr>
<tr>
<td>NNMC</td>
<td>North Neshaminy (closed)</td>
<td>T351.90 WB</td>
<td>302.85</td>
<td>6400</td>
<td>52,850</td>
<td>158</td>
<td>4.70*</td>
</tr>
<tr>
<td>ALT</td>
<td>Allentown</td>
<td>A55.90 NB/SB</td>
<td>279.30</td>
<td>19200</td>
<td>49,480</td>
<td>89</td>
<td>3.80</td>
</tr>
<tr>
<td>HKR</td>
<td>Hickory Run</td>
<td>A86.14 NB/SB</td>
<td>346.80</td>
<td>12800</td>
<td>35,150</td>
<td>88</td>
<td>3.00*</td>
</tr>
</tbody>
</table>

* Cost estimate from the final regression model
so we remove this variable as well and fit a final regression model using only natural gas supply line length and pavement size as independent variables. The details of the final regression model used to estimate the construction costs for LNG facilities are shown in Figure 3-3 with Adjusted R-squared value of 0.9368.

![Figure 3-2. Spread-level plots for the two initial regression models](image)

(a) First regression model  
(b) Second regression model

Model 3: “LNG Length” and “Pavement Size”.

Coefficients:

|                | Estimate | Std. Error | t value | Pr(>|t|) |
|----------------|----------|------------|---------|----------|
| (Intercept)    | 1.642e+00| 1.454e-01  | 11.295  | 9.51e-05 |
| CNG_Length     | 5.893e-04| 2.634e-04  | 2.238   | 0.075431 |
| Pavement_Size  | 8.620e-05| 8.433e-06  | 10.221  | 0.000154 |

Residual standard error: 0.2042 on 5 degrees of freedom  
Multiple R-squared: 0.9548, Adjusted R-squared: 0.9368  
F-statistic: 52.86 on 2 and 5 DF, p-value: 0.0004333

Regression Equation:

\[
\text{Construction Cost} = (0.0005893 \times \text{LNG Supply Line Length}) + (0.0000862 \times \text{New Pavement size}) + 1.642
\]

![Figure 3-3. Results for the final regression model](image)

Using the final regression model, we estimate the cost for constructing LNG stations at the remaining 11 service plazas, which are shown in Table 3-3. In the next section, these construction cost estimates are used in Models M1 and M2 to determine optimal station locations considering different budget levels and VMT coverage.
3.3.3 Results and Analysis

In this section, using the 2011 O/D traffic flow matrix (in trucks/year) for FHWA truck classes 6-10, the O/D distance matrix (Myers et al. 2013), and the milestone information for service plazas in Table 3-3, we run proposed Models M1 and M2, and compared their results in terms of effective VMT coverage and optimal locations for LNG refueling stations on the turnpike. We also compared the performance of the new models with the original Hwang et al. (2015) model. The mathematical models in this section are built using MATLAB R2015a and solved with CPLEX Version 12 on an Intel Core i7-4790S CPU at 3.2GHZ 16GB RAM. The total computational time to build and solve each mathematical model is less than one minute.

For truck driving ranges in the case study, we consider 300, 450, and 600 miles. The truck driving range of 300 miles is a conservative range for current and older models of LNG trucks with a single fuel tank (Burnham 2013), but with improving technologies, the range of an LNG truck should increase substantially, justifying our inclusion of trucks with longer driving ranges. Also, the truck driving range of 600 miles is a reasonable range for LNG trucks with a dual fuel tank system (Park 2014).

For Model M1 using the normalized weighting method, we first set $\alpha = 847.84$ million VMT and $\beta = 55.30$M, which are the total VMT that can be covered per year and the total construction cost for a full LNG infrastructure development at the service plazas in the turnpike, respectively. They normalize each of the objectives in Objective (3.7). Then, we run M1 with a variety of different weight values from 0 to 1. Table 3-4 summarizes the weight values for Model M1 that produce results for the objective of maximizing the normalized, combined values for each independent experiment. These normalized results are converted into effective VMT coverage and construction cost pairs, which are also shown in Table 3-4.
For Model M2 using the constraint method, the budget levels used in the independent experiments range from $2M to $56M in $3M increments. The smallest budget of $2M allows construction of an LNG station at the least expensive service plaza and the largest budget of $56M enables construction of LNG stations in all 19 service plazas. Model M2 produces results for the objective of maximizing VMT for each budget level, which are also shown in Table 3-4.

Both Models M1 and M2 for \( R = 300 \) lead to the same solutions, including effective VMT coverage and sets of optimal station locations for different construction costs. While Model M2 can be efficiently managed to generate the effective VMT coverage and optimal solution for a given budget level, Model M1 is not as intuitive to manage and a bisection search on weight \( w \), for \( w \in [0, 1] \), may be necessary to obtain the effective VMT coverage and optimal solution that meets a desired budget level, since Objective (3.7) in M1 combines and normalizes the two different objectives. In our experiments, we run the model for weight values from 0 to 1 in 0.01 increments. If M1 missed an optimal solution identical to a solution found in M2, then we rerun M1 with weight values in smaller increments until M1 found the same optimal solution. Based on our experiments, the second column of Table 3-4 provides ranges of weight values for M1 that produces identical solutions to M2. Since using M1 is time-consuming, we decided to use M2 of the remainder of this case study.

Figure 3-4 shows the trade-off between construction cost and effective VMT coverage for the three different truck driving ranges \( (R = 300, R = 450, \text{and} \ R = 600) \) in Model M2, and Table 3-5 summarizes the corresponding optimal locations for the LNG refueling stations. From these results, we first notice that the effective VMT coverage increases as the construction cost increases and, at the beginning, a large percentage of effective VMT can be covered with a relatively small construction budget. After that, the rate of increase of effective VMT coverage decreases as the construction budget increases. In addition, the effective VMT coverage for trucks with driving ranges of 450 and 600 miles is nearly the same at every budget level, implying that
<table>
<thead>
<tr>
<th>Budget Level for M2 ($M)</th>
<th>Weight Ranges for M1</th>
<th>Construction Cost ($M)</th>
<th>Number of Stations</th>
<th>Effective VMT coverage (%)</th>
<th>Construction Cost ($M)</th>
<th>Number of Stations</th>
<th>Effective VMT coverage (%)</th>
<th>Construction Cost ($M)</th>
<th>Number of Stations</th>
<th>Effective VMT coverage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.14988431361111</td>
<td>1.90</td>
<td>1</td>
<td>6.55</td>
<td>1.80</td>
<td>1</td>
<td>12.61</td>
<td>1.80</td>
<td>1</td>
<td>15.47</td>
</tr>
<tr>
<td>5</td>
<td>0.14988431361112 - 0.24122</td>
<td>3.60</td>
<td>2</td>
<td>36.92</td>
<td>3.60</td>
<td>2</td>
<td>45.38</td>
<td>4.10</td>
<td>2</td>
<td>47.88</td>
</tr>
<tr>
<td>8</td>
<td>0.24123 - 0.24124</td>
<td>7.40</td>
<td>3</td>
<td>53.84</td>
<td>7.70</td>
<td>4</td>
<td>65.76</td>
<td>7.70</td>
<td>4</td>
<td>69.72</td>
</tr>
<tr>
<td>11</td>
<td>0.286 - 0.289</td>
<td>10.00</td>
<td>5</td>
<td>68.79</td>
<td>11.00</td>
<td>5</td>
<td>81.35</td>
<td>11.00</td>
<td>5</td>
<td>84.62</td>
</tr>
<tr>
<td>14</td>
<td>0.40 - 0.59</td>
<td>13.80</td>
<td>6</td>
<td>85.71</td>
<td>13.80</td>
<td>6</td>
<td>86.72</td>
<td>13.80</td>
<td>6</td>
<td>89.44</td>
</tr>
<tr>
<td>17</td>
<td>0.64 - 0.65</td>
<td>16.80</td>
<td>7</td>
<td>88.66</td>
<td>16.20</td>
<td>7</td>
<td>90.33</td>
<td>16.80</td>
<td>7</td>
<td>92.40</td>
</tr>
<tr>
<td>20</td>
<td>0.66 - 0.67</td>
<td>19.60</td>
<td>8</td>
<td>91.61</td>
<td>19.15</td>
<td>8</td>
<td>93.29</td>
<td>19.70</td>
<td>8</td>
<td>95.19</td>
</tr>
<tr>
<td>23</td>
<td>0.67 - 0.76</td>
<td>22.60</td>
<td>9</td>
<td>94.57</td>
<td>22.60</td>
<td>9</td>
<td>95.77</td>
<td>21.50</td>
<td>9</td>
<td>96.05</td>
</tr>
<tr>
<td>26</td>
<td>0.77 - 0.78</td>
<td>25.50</td>
<td>10</td>
<td>95.83</td>
<td>26.00</td>
<td>10</td>
<td>97.49</td>
<td>25.60</td>
<td>10</td>
<td>98.04</td>
</tr>
<tr>
<td>29</td>
<td>0.79 - 0.8507</td>
<td>28.50</td>
<td>10</td>
<td>97.53</td>
<td>27.40</td>
<td>10</td>
<td>98.04</td>
<td>28.50</td>
<td>10</td>
<td>98.66</td>
</tr>
<tr>
<td>32</td>
<td>0.8508 - 0.8509</td>
<td>32.00</td>
<td>12</td>
<td>98.19</td>
<td>31.40</td>
<td>11</td>
<td>99.24</td>
<td>30.30</td>
<td>11</td>
<td>99.53</td>
</tr>
<tr>
<td>35</td>
<td>0.88 - 0.89</td>
<td>35.00</td>
<td>12</td>
<td>99.34</td>
<td>34.10</td>
<td>12</td>
<td>99.68</td>
<td>33.70</td>
<td>12</td>
<td>99.68</td>
</tr>
<tr>
<td>38</td>
<td>0.94 - 0.97</td>
<td>37.70</td>
<td>13</td>
<td>99.77</td>
<td>37.60</td>
<td>13</td>
<td>99.82</td>
<td>37.60</td>
<td>13</td>
<td>99.82</td>
</tr>
<tr>
<td>41</td>
<td>0.98</td>
<td>40.30</td>
<td>14</td>
<td>99.91</td>
<td>40.30</td>
<td>14</td>
<td>99.91</td>
<td>40.30</td>
<td>14</td>
<td>99.91</td>
</tr>
<tr>
<td>44</td>
<td>0.99 - 1.00</td>
<td>43.10</td>
<td>15</td>
<td>100.00</td>
<td>43.10</td>
<td>15</td>
<td>100.00</td>
<td>43.10</td>
<td>15</td>
<td>100.00</td>
</tr>
</tbody>
</table>
current and developing technologies that enhance the driving range of LNG vehicles do not significantly improve the coverage of LNG trucks using any optimal set of LNG station locations considering different budget levels. Furthermore, when the budget level becomes $41M, all three truck driving ranges converge to the same effective VMT coverage of 99.91% with an identical set of optimal station locations and, when $43.10M are invested, the effective VMT coverage for all three truck driving ranges reaches 100% with an identical set of optimal station locations. Note that the cost of constructing LNG refueling stations at all 19 service plazas is $55.30M, but the analysis shows that an investment beyond $43.10M leads to the same result of 100% coverage.

Figure 3-4. Trade-off between construction cost and effective VMT coverage for three different truck driving ranges (R=300, R=450, and R=600)

We now compare the performance of Model M2 with that of the Hwang et al. (2015) model in terms of effective VMT coverage for $R = 300$. Note that the mathematical model proposed by Hwang et al. (2015) maximizes the coverage of annual traffic flow in vehicle round
Table 3-5. Optimal LNG station locations considering different budget levels and truck range combinations

(a) Truck range: R=300

<table>
<thead>
<tr>
<th>Budget Level ($M)</th>
<th>Construction Cost ($M)</th>
<th>Effective VMT Coverage (%)</th>
<th>Optimal Locations</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1.90</td>
<td>6.55</td>
<td>KPR</td>
</tr>
<tr>
<td>5</td>
<td>3.60</td>
<td>36.92</td>
<td>SMW, NMW</td>
</tr>
<tr>
<td>8</td>
<td>7.40</td>
<td>53.84</td>
<td>SMW, NMW, ALT</td>
</tr>
<tr>
<td>11</td>
<td>10.00</td>
<td>68.79</td>
<td>OMP, SMW, NMW, HSP, PJC</td>
</tr>
<tr>
<td>14</td>
<td>13.80</td>
<td>85.71</td>
<td>OMP, SMW, NMW, HSP, PJC, ALT</td>
</tr>
<tr>
<td>17</td>
<td>16.80</td>
<td>88.66</td>
<td>OMP, SMW, NMW, HSP, PJC, ALT, HKR</td>
</tr>
<tr>
<td>20</td>
<td>19.60</td>
<td>91.61</td>
<td>OMP, NST, SMW, SLH, HSP, PJC, KPR, ALT</td>
</tr>
<tr>
<td>23</td>
<td>22.60</td>
<td>94.57</td>
<td>OMP, NST, SMW, SLH, HSP, PJC, KPR, ALT, HKR</td>
</tr>
<tr>
<td>26</td>
<td>25.50</td>
<td>95.83</td>
<td>OMP, NST, SSS, NMW, CLV, HSP, PJC, KPR, ALT, HKR</td>
</tr>
<tr>
<td>29</td>
<td>28.50</td>
<td>97.53</td>
<td>ZEN, OMP, NST, SSS, SLH, HSP, PJC, KPR, ALT, HKR</td>
</tr>
<tr>
<td>32</td>
<td>32.00</td>
<td>98.19</td>
<td>ZEN, OMP, NST, SSS, NMW, CLV, HSP, LAW, PJC, KPR, ALT, HKR</td>
</tr>
<tr>
<td>35</td>
<td>35.00</td>
<td>99.34</td>
<td>ZEN, OMP, NST, SSS, NMW, SLH, HSP, PJC, KPR, NNM, ALT, HKR</td>
</tr>
<tr>
<td>38</td>
<td>37.70</td>
<td>99.77</td>
<td>ZEN, OMP, NST, SSS, NMW, SLH, CLV, HSP, PJC, KPR, NNM, ALT, HKR</td>
</tr>
<tr>
<td>41</td>
<td>40.30</td>
<td>99.91</td>
<td>ZEN, OMP, NST, SSS, SMW, SLH, CLV, LAW, BMV, PJC, KPR, NNM, ALT, HKR</td>
</tr>
<tr>
<td>44</td>
<td>43.10</td>
<td>100.00</td>
<td>ZEN, OMP, NST, SSS, SMW, SLH, CLV, LAW, BMV, PJ C, VFG, KPR, NNM, ALT, HKR</td>
</tr>
</tbody>
</table>
(b) Truck range: R=450

<table>
<thead>
<tr>
<th>Budget Level ($M)</th>
<th>Construction Cost ($M)</th>
<th>Effective VMT Coverage (%)</th>
<th>Optimal Locations</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1.80</td>
<td>12.61</td>
<td>PJC</td>
</tr>
<tr>
<td>5</td>
<td>3.60</td>
<td>45.38</td>
<td>SMW, NMW</td>
</tr>
<tr>
<td>8</td>
<td>7.70</td>
<td>65.76</td>
<td>OMP, NST, SMW, PJC</td>
</tr>
<tr>
<td>11</td>
<td>11.00</td>
<td>81.35</td>
<td>OMP, SMW, NMW, PJC, ALT</td>
</tr>
<tr>
<td>14</td>
<td>13.80</td>
<td>86.72</td>
<td>OMP, SMW, NMW, HSP, PJC, ALT</td>
</tr>
<tr>
<td>17</td>
<td>16.20</td>
<td>90.33</td>
<td>OMP, NST, SMW, HSP, PJC, KPR, ALT</td>
</tr>
<tr>
<td>20</td>
<td>19.15</td>
<td>93.29</td>
<td>OMP, NST, SMW, HSP, PJC, KPR, ALT, HKR</td>
</tr>
<tr>
<td>23</td>
<td>22.60</td>
<td>95.77</td>
<td>OMP, NST, SMW, SLH, HSP, PJC, KPR, ALT, HKR</td>
</tr>
<tr>
<td>26</td>
<td>26.00</td>
<td>97.49</td>
<td>ZEN, OMP, NST, SMW, CLV, HSP, PJC, KPR, ALT, HKR</td>
</tr>
<tr>
<td>29</td>
<td>27.40</td>
<td>98.04</td>
<td>ZEN, OMP, NST, SMW, BMT, HSP, PJC, KPR, ALT, HKR</td>
</tr>
<tr>
<td>32</td>
<td>31.40</td>
<td>99.24</td>
<td>ZEN, OMP, NST, SMW, SLH, HSP, PJC, KPR, NNM, ALT, HKR</td>
</tr>
<tr>
<td>35</td>
<td>34.10</td>
<td>99.68</td>
<td>ZEN, OMP, NST, SMW, SLH, CLV, HSP, PJC, KPR, NNM, ALT, HKR</td>
</tr>
<tr>
<td>38</td>
<td>37.60</td>
<td>99.82</td>
<td>ZEN, OMP, NST, NSS, SMW, SLH, CLV, HSP, PJC, KPR, NNM, ALT, HKR</td>
</tr>
<tr>
<td>41</td>
<td>40.30</td>
<td>99.91</td>
<td>ZEN, OMP, NST, SSS, SMW, SLH, CLV, LAW, BMV, PJC, KPR, NNM, ALT, HKR</td>
</tr>
<tr>
<td>44</td>
<td>43.10</td>
<td>100.00</td>
<td>ZEN, OMP, NST, SSS, SMW, SLH, CLV, LAW, BMV, PJC, VFG, KPR, NNM, ALT, HKR</td>
</tr>
</tbody>
</table>
(c) Truck range: R=600

<table>
<thead>
<tr>
<th>Budget Level ($M)</th>
<th>Construction Cost ($M)</th>
<th>Effective VMT Coverage (%)</th>
<th>Optimal Locations</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1.80</td>
<td>15.47</td>
<td>PJC</td>
</tr>
<tr>
<td>5</td>
<td>4.10</td>
<td>47.88</td>
<td>NST, SMW</td>
</tr>
<tr>
<td>8</td>
<td>7.70</td>
<td>69.72</td>
<td>OMP, NST, SMW, PJC</td>
</tr>
<tr>
<td>11</td>
<td>11.00</td>
<td>84.62</td>
<td>OMP, SMW, NMW, PJC, ALT</td>
</tr>
<tr>
<td>14</td>
<td>13.80</td>
<td>89.44</td>
<td>OMP, NMW, CLV, PJC, KPR, ALT</td>
</tr>
<tr>
<td>17</td>
<td>16.80</td>
<td>92.40</td>
<td>OMP, NMW, CLV, PJC, KPR, ALT, HKR</td>
</tr>
<tr>
<td>20</td>
<td>19.70</td>
<td>95.19</td>
<td>OMP, NST, CLV, LAW, PJC, KPR, ALT, HKR</td>
</tr>
<tr>
<td>23</td>
<td>21.50</td>
<td>96.05</td>
<td>OMP, NST, SMW, CLV, LAW, PJC, KPR, ALT, HKR</td>
</tr>
<tr>
<td>26</td>
<td>25.60</td>
<td>98.04</td>
<td>ZEN, OMP, NST, NMW, CLV, LAW, PJC, KPR, ALT, HKR</td>
</tr>
<tr>
<td>29</td>
<td>28.50</td>
<td>98.66</td>
<td>ZEN, OMP, NST, CLV, LAW, PJC, KPR, NN, ALT, HKR</td>
</tr>
<tr>
<td>32</td>
<td>30.30</td>
<td>99.53</td>
<td>ZEN, OMP, NST, NMW, CLV, LAW, PJC, KPR, NN, ALT, HKR</td>
</tr>
<tr>
<td>35</td>
<td>33.70</td>
<td>99.68</td>
<td>ZEN, OMP, NST, NMW, SLH, CLV, LAW, PJC, KPR, NN, ALT, HKR</td>
</tr>
<tr>
<td>38</td>
<td>37.60</td>
<td>99.82</td>
<td>ZEN, OMP, NST, SSS, SMW, SLH, CLV, HSP, PJC, KPR, NN, ALT, HKR</td>
</tr>
<tr>
<td>41</td>
<td>40.30</td>
<td>99.91</td>
<td>ZEN, OMP, NST, SSS, SMW, SLH, CLV, LAW, BM, PJC, KPR, NN, ALT, HKR</td>
</tr>
<tr>
<td>44</td>
<td>43.10</td>
<td>100.00</td>
<td>ZEN, OMP, NST, SSS, SMW, SLH, CLV, LAW, BM, PJC, VFG, KPR, NN, ALT, HKR</td>
</tr>
</tbody>
</table>
trips for a given number of refueling stations. We convert their results (annual round trips covered) into the annual VMT covered, and compute the effective VMT coverage in Table 3-6. Figure 3-5 compares the effective VMT coverage between Model M2 and the Hwang et al. model with respect to the construction cost. Clearly, Model M2 performs better than Hwang et al. model, since the objective of Model M2 is maximizing the total VMT covered. Note, for example, the effective VMT coverage by Model M2 is 36.92% when $3.60M are invested in construction, while the effective VMT coverage by the Hwang et al. model is only 16.90% when the construction investment is $3.80M. The coverage gap between Model M2 and Hwang et al. model in Figure 3-5 is noticeable when the construction cost is within $17.50M. This gap, however, suddenly decreases and becomes small beyond $17.70M of construction cost, and then finally becomes zero beyond a construction cost of $41.90M.

Figure 3-5. Comparison of effective VMT coverage between M2 and Hwang et al. model for R=300
Table 3-6. Effective VMT coverage and the optimal locations by the Hwang et al. (2015) model

<table>
<thead>
<tr>
<th>No. of Stations</th>
<th>Annual Round Trips Covered (millions)</th>
<th>Annual VMT Covered (millions)</th>
<th>Effective VMT Coverage (%)</th>
<th>Construction Cost ($M)</th>
<th>Optimal Locations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.07</td>
<td>143.25</td>
<td>16.90</td>
<td>3.80</td>
<td>ALT</td>
</tr>
<tr>
<td>2</td>
<td>3.58</td>
<td>365.02</td>
<td>43.05</td>
<td>7.20</td>
<td>SLH, ALT</td>
</tr>
<tr>
<td>3</td>
<td>4.93</td>
<td>417.51</td>
<td>49.24</td>
<td>9.10</td>
<td>SLH, KPR, ALT</td>
</tr>
<tr>
<td>4</td>
<td>5.97</td>
<td>465.88</td>
<td>54.95</td>
<td>10.90</td>
<td>OMP, SLH, KPR, ALT</td>
</tr>
<tr>
<td>5</td>
<td>7.00</td>
<td>619.73</td>
<td>73.09</td>
<td>13.60</td>
<td>OMP, SLH, HSP, PJC, ALT</td>
</tr>
<tr>
<td>6</td>
<td>8.02</td>
<td>641.63</td>
<td>75.68</td>
<td>17.50</td>
<td>OMP, SLH, HSP, PJC, NNM, HKR</td>
</tr>
<tr>
<td>7</td>
<td>8.96</td>
<td>748.57</td>
<td>88.29</td>
<td>17.70</td>
<td>OMP, SMW, NMW, HSP, PJC, NNM, HKR</td>
</tr>
<tr>
<td>8</td>
<td>9.77</td>
<td>768.76</td>
<td>90.67</td>
<td>19.60</td>
<td>OMP, SMW, NMW, HSP, PJC, KPR, NNM, HKR</td>
</tr>
<tr>
<td>9</td>
<td>10.33</td>
<td>785.59</td>
<td>92.66</td>
<td>23.70</td>
<td>ZEN, OMP, SMW, NMW, HSP, PJC, KPR, NNM, HKR</td>
</tr>
<tr>
<td>10</td>
<td>10.86</td>
<td>818.56</td>
<td>96.55</td>
<td>28.30</td>
<td>ZEN, OMP, NST, SMW, BMT, HSP, PJC, KPR, NNM, HKR</td>
</tr>
<tr>
<td>11</td>
<td>11.25</td>
<td>833.58</td>
<td>98.32</td>
<td>32.10</td>
<td>ZEN, OMP, NST, SMW, BMT, HSP, PJC, KPR, NNM, ALT, HKR</td>
</tr>
<tr>
<td>12</td>
<td>11.41</td>
<td>844.58</td>
<td>99.62</td>
<td>35.70</td>
<td>ZEN, OMP, NST, SSS, SMW, BMT, HSP, PJC, KPR, NNM, ALT, HKR</td>
</tr>
<tr>
<td>13</td>
<td>11.46</td>
<td>845.35</td>
<td>99.71</td>
<td>38.50</td>
<td>ZEN, OMP, NST, SSS, SMW, BMT, HSP, PJC, VFG, KPR, NNM, ALT, HKR</td>
</tr>
<tr>
<td>14</td>
<td>11.51</td>
<td>847.11</td>
<td>99.91</td>
<td>41.90</td>
<td>ZEN, OMP, NST, SSS, SMW, SLH, BMT, HSP, PJC, VFG, KPR, NNM, ALT, HKR</td>
</tr>
<tr>
<td>15</td>
<td>11.55</td>
<td>847.84</td>
<td>100.00</td>
<td>43.10</td>
<td>ZEN, OMP, NST, SSS, NMW, SLH, CLV, LAW, BMV, PJC, VFG, KPR, NNM, ALT, HKR</td>
</tr>
</tbody>
</table>
3.3.4 Estimation of Greenhouse Gas Emission Savings

In this section, we discuss the impact of refueling station construction on GHG emissions reduction and relate the cost of construction to the social cost of carbon (SCC) savings. To evaluate the GHG emission differences between the current state, where we assume all HDVs are currently using diesel fuel, to a potential future state, where we assume all HDVs could be transitioned to using LNG, we must first determine the emission factors for both fuels for our target GHGs and the CO₂Eq for the total emitted GHGs.

Table 3-7 shows the emissions for both diesel fuel and LNG for CO₂, CH₄, and N₂O in three stages of the supply chain for both fuel types where emissions are created. In the feedstock stage, unprocessed fuel is recovered from a natural source. In the fuel processing stage, the raw materials are converted into usable fuel and transported through distribution channels, and in the tailpipe (vehicle use) stage, the fuel is consumed through combustion. In conducting this analysis, we use estimates from (Meyer et al. 2011) on the total fuel cycle (well-to-wheels), as opposed to strictly limiting our analysis to tailpipe emissions. These factors show that for transition from diesel to LNG, a 5% reduction in CO₂ can be realized, but the emissions of CH₄ will be increased 2.9 times. Diesel emits almost no CH₄ through tailpipe emissions (<1 mg/ton-mile), but as with both fuels the CH₄ emissions are predominantly experienced in the feedstock, fuel processing, and vehicle use stages (<1% of diesel CH₄ emissions are realized in tailpipe emissions, while about 17% of LNG CH₄ emissions are realized from the vehicle tailpipe). As N₂O emissions are considered a function of emissions control technologies, we conclude that the reduction in N₂O is negligible between fuel types. We calculate the CO₂Eq of CH₄ using a GWP of 25 (for a 100-year GWP) and the EPA relation for CO₂Eq (EPA 2016b):

$$\text{CH}_4 \text{ CO}_2 \text{Eq (mg/ton-mile)} = \text{CH}_4 \text{ (mg/ton-mile)} \times (\text{GWP for } \text{CH}_4)$$  (3.9)
Table 3-7. Total Fuel Cycle Emissions for Target Fuels and GHGs (Meyer et al. 2011)

<table>
<thead>
<tr>
<th>Feedstock Stage</th>
<th>Fuel Processing Stage</th>
<th>Tailpipe Emissions</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diesel</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CO₂ (g/ton-mile)</td>
<td>5</td>
<td>10</td>
<td>85</td>
</tr>
<tr>
<td>CH₄ (mg/ton-mile)</td>
<td>90</td>
<td>10</td>
<td>&lt; 1</td>
</tr>
<tr>
<td>N₂O (mg/ton-mile)</td>
<td>&lt; 1</td>
<td>&lt; 1</td>
<td>&lt; 1</td>
</tr>
<tr>
<td>CH₄CO₂Eq (g/ton-mile)</td>
<td>2.25</td>
<td>0.25</td>
<td>0</td>
</tr>
<tr>
<td>Total CO₂Eq (g/ton-mile)</td>
<td>7.25</td>
<td>10.25</td>
<td>85</td>
</tr>
<tr>
<td>LNG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CO₂ (g/ton-mile)</td>
<td>5</td>
<td>10</td>
<td>80</td>
</tr>
<tr>
<td>CH₄ (mg/ton-mile)</td>
<td>145</td>
<td>90</td>
<td>50</td>
</tr>
<tr>
<td>N₂O (mg/ton-mile)</td>
<td>&lt; 1</td>
<td>&lt; 1</td>
<td>&lt; 1</td>
</tr>
<tr>
<td>CH₄CO₂Eq (g/ton-mile)</td>
<td>3.75</td>
<td>2.25</td>
<td>1.25</td>
</tr>
<tr>
<td>Total CO₂Eq (g/ton-mile)</td>
<td>8.75</td>
<td>12.25</td>
<td>81.25</td>
</tr>
</tbody>
</table>

Note: Values in the table may not sum due to rounding to the nearest 5 in the original data.

In calculating the differences between diesel and LNG by the weight of the emitted gases, we consider the VMT from the optimal solution to the model presented in the previous section, as well as the weight of the trucks. Due to a lack of data correlating a truck trip with a truck weight, we estimate the weight of any given truck using truck weight distribution data. We calculate the expected value of a truck from data accrued from a 15-state study of weight frequency data (National Research Council 2010). The summary graph along with the estimates of the proportion of trucks observed at each weight point is shown in Figure 3-6.

Using estimates of the proportion of trucks at a specific weight point within EPA weight class 8 used by the EPA for emission regulations (EPA and Department of Transportation 2016), the expected weight of a truck is calculated to estimate the total amount of GHG emissions reduction for the miles traveled measurements from the results produced by the model in the previous section. With our approximation for a truck weight of 31.4 tons, we multiply this weight by the VMT saved through each combination of budget (up to $44M) and truck range, calculate the difference in GHG emissions for our target GHGs, the overall CO₂Eq, and the SCC savings in
<table>
<thead>
<tr>
<th>Weight (lbs)</th>
<th>Count (Estimated)</th>
<th>Probability (Weight)</th>
</tr>
</thead>
<tbody>
<tr>
<td>46300</td>
<td>900</td>
<td>0.116</td>
</tr>
<tr>
<td>50700</td>
<td>850</td>
<td>0.109</td>
</tr>
<tr>
<td>55100</td>
<td>900</td>
<td>0.116</td>
</tr>
<tr>
<td>59500</td>
<td>950</td>
<td>0.122</td>
</tr>
<tr>
<td>63900</td>
<td>975</td>
<td>0.125</td>
</tr>
<tr>
<td>68330</td>
<td>1075</td>
<td>0.138</td>
</tr>
<tr>
<td>72800</td>
<td>1250</td>
<td>0.161</td>
</tr>
<tr>
<td>77200</td>
<td>500</td>
<td>0.064</td>
</tr>
<tr>
<td>81600</td>
<td>225</td>
<td>0.029</td>
</tr>
<tr>
<td>86000</td>
<td>75</td>
<td>0.010</td>
</tr>
<tr>
<td>90400</td>
<td>50</td>
<td>0.006</td>
</tr>
<tr>
<td>94800</td>
<td>25</td>
<td>0.003</td>
</tr>
<tr>
<td>99200</td>
<td>10</td>
<td>0.001</td>
</tr>
<tr>
<td>103600</td>
<td>1</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Note: Frequency values for the target weights are estimated from National Research Council (2010).

\[
E[\text{Weight(lbs)}] = \sum [\text{Weight(lbs)} \times \text{Probability(Weight)}] = 62809.3 \text{ lbs} \approx 31.4 \text{ tons}
\]

Figure 3-6. Estimate of truck weight distribution
relation to the amount of CO$_2$Eq prevented from entering the atmosphere.

The EPA estimates the SCC to quantify the magnitude of the financial impact of CO$_2$ emissions, including, but not limited to, “changes in net agricultural productivity, human health, property damages from increased flood risk, and the value of ecosystem services due to climate change” (Interagency Working Group on Social Cost of Carbon 2016). The EPA estimates the SCC for every 5-year increment, and we use the 2020 estimate of $12 per ton CO$_2$ (in 2007 dollars, 5.0% discount rate average value) in our calculations for the social impact of the emissions reduction we would expect to see from constructing our optimal LNG refueling infrastructure.

As shown in Figures 3-7 (a)-(d), CO$_2$, CO$_2$Eq, and SCC savings all follow the same pattern as VMT. This is expected given the linear nature of the savings in relation to VMT. The exception is the total full cycle analysis of CH$_4$, which is also linearly related to VMT, but shows a negative emission savings. This is due to the increase in CH$_4$ emissions from both LNG vehicles and the total fuel cycle for natural gas fuel products. Although CH$_4$ has a larger carbon footprint, the overall CO$_2$Eq savings is still positive, given the much smaller mass of CH$_4$ emitted by LNG vehicles when compared to the mass of carbon emitted by diesel vehicles.
(a) CO$_2$ Savings based on construction cost

(b) CH$_4$ Savings based on construction cost
Figure 3-7. Cost benefit analysis of annual GHG emissions reduction and SCC savings

(c) CO$_2$Eq Savings based on construction cost

(d) SCC Savings based on construction cost
Now, we perform the cost benefit analysis of annual SCC savings. Figure 3-8 (a) shows the marginal SCC improvements per year due to the last $1M invested for each actual construction cost when considering the estimates of social benefits conferred through reduction in CO\textsubscript{2}Eq. Considering the intervals between budget levels in our analyses, we can see that investing in LNG construction confers a cost-benefit ratio greater than 1 for investments up to around $17M, but then the investment becomes less cost-effective. This makes sense, given that each additional station beyond a certain level covers fewer VMT than those covered by the prior station, and we are already aware that any investment over $43.10M produces no additional benefits to VMT, GHG emissions, or SCC. At the $17M investment level, the trucks in this study with even the assumed minimum driving ranges are covered in over 88% of their trips, and trucks with longer ranges are covered for more than 90% of their studied trips. If the investment were to proceed to the next budget level, $20M, the longer-range trucks achieve approximately 95% effective coverage. As originally observed in Hwang et. al. (2015), we also note the apparent inconsistency with the marginal SCC improvement for trucks with a 300-mile range. When the optimal solutions for each budget level are produced for the 300-mile truck range, the results show that the optimal locations for refueling facilities at low budget levels are not shared between adjacent budget levels. This means that the results for a budget of $8M and the results for budget level of $11M have a distinctly different set of facility locations (specifically, the Allentown (ALT) service plaza is part of the optimal set of refueling station locations at the $8M budget level, but is replaced by three other service plazas at a $11M budget level, shown in Table 3-5 (a)), resulting in an increase in the marginal SCC benefit for higher budget levels, contrary to intuition. Figure 3-8 (b) shows the average SCC savings per year for each $1M invested in construction realized through emissions reduction for each actual construction cost. Initial investments in infrastructure development show a high annual SCC savings, but eventual increases in construction cost lead to a decreasing growth rate in social benefits and lower
average SCC savings per year. Knowing that there is no additional SCC savings benefit after the $43.10M of construction cost, our analysis shows that the SCC savings at that construction cost the SCC benefits exceed 180% of the cost of construction, and at an effective coverage level of over 98% for all truck ranges, a budget constraint of $32M produces more than 245% return on investment in annual SCC savings.

(a) Marginal SCC improvement for the last $1M invested based on construction cost
3.4 Concluding Remarks

In this chapter, we have applied a modified version of Hwang et al. (2015) model to locate LNG refueling stations in a directed (symmetric) transportation network. The results demonstrate significant benefits including:

- Reduction in dependence on foreign suppliers of fossil fuels furthering the goals of energy independence;
- Domestic economic benefits in the form of transitioning fuel suppliers to local sources, including both the reduced supply chain costs when purchasing fuel from local sources and increased revenue generated by domestic suppliers;
- Reduction in operating costs for fueling vehicles running on natural gas products due to abundant, inexpensive fuel;
• Economic benefits including research and development of supply chain infrastructure for alternative fuels, enabling the creation of jobs in research, development, manufacturing, and supply chain operations and maintenance;

• Reductions in CO\textsubscript{2} emissions when considered independently from other GHGs;

• Potential for GHG emission reduction when all GHG emissions are considered in totality; and

• Reduction in exposure to price volatility for fuel supply from potentially unstable foreign sources.

This application of Hwang et al. (2015) model shows that the traffic flows observed in a transportation network can be used to design an alternative-fuel infrastructure to maximize GHG emission savings from HDVs. However, there are still some major obstacles to overcome:

• While the adoption of LNG as a fuel in a supply chain that includes heavy-duty, long-haul trucks shows the benefit of reducing CO\textsubscript{2} emissions, LNG suffers from the problematic issue of venting a potentially more damaging GHG, CH\textsubscript{4}, into the atmosphere, potentially defeating the gains made by reducing CO\textsubscript{2} emissions.

• The adoption and implementation of an LNG infrastructure suffers from the problem that suppliers of natural gas products are unlikely to invest in a publicly available LNG or CNG infrastructure unless major potential consumers of natural gas products for transportation are willing to transition their fleet vehicles to these alternative fuels to ensure profitability. Likewise, fleet operators are unlikely to invest in conversion to natural gas vehicles unless there is an infrastructure available to supply fuel for fleet vehicles to ensure fuel cost savings.

• The conversion to vehicles that operate on an alternative-fuel has a steep cost. Moving to a new fuel type would require either a conversion of current vehicles to LNG or the purchase of new vehicles that operate on LNG technologies, and both options have a high cost of adoption.
To leverage the benefits of LNG as a transportation fuel, there are clearly issues to overcome. Initial investments in LNG technologies and infrastructure are significant, and the coordination of adoption between natural gas suppliers, fleet operators, and drivers is critical to establishing a foundation for the ongoing expansion of an alternative-fuel infrastructure. We recommend the following policy strategies to aid in the adoption of LNG as an alternative to diesel fuel for HDVs:

- **Mitigate the burden of converting to LNG vehicles through collaboration with industry stakeholders and leveraging economic tools to create a cost-effective adoption strategy.** To promote the adoption of LNG fuels, the substantial financial burden of conversion should be shared by federal, state, and local regulatory agencies as well as natural gas suppliers and fleet operators. Reducing the financial commitment required for private organizations to adopt LNG technologies through rebate structures, tax considerations, and carbon trading credits can lower the barriers to adoption of beneficial new technologies.

- **Streamline the logistics of LNG infrastructure by collaborating with industry stakeholders to mitigate the complexity of transitioning to LNG through managed adoption practices.** By identifying a target timeframe for adoption and creating agreements on the adoption of LNG, the difference between the time that it takes to setup an LNG infrastructure and the realization of operational economic benefits for the suppliers and consumers of natural gas products can be minimized.

- **Offer reductions in transportation costs for adopters of LNG technologies.** In this chapter, we have studied the adoption of LNG in a toll road. The cost of transporting goods is linked to the toll rate for HDVs on the PA Turnpike. We suggest offering reductions in tolls for adopters of alternative fuels. On many congested public highways, lanes are established for certain types of vehicles like buses, motorcycles, and hybrid cars. While the cost of laying...
new lanes on the turnpike is surely prohibitive, adopting reduced fares on the highway can improve the economic attractiveness of a vehicle conversion.

Furthermore, we consider the adoption of improving technologies to reduce fugitive CH₄ emissions:

- **Regulate the adoption of technologies that reduce fugitive GHG emissions, particularly CH₄.** Delgado and Muncrief (2015) showed that CH₄ emissions are not merely a tailpipe phenomenon, but can be emitted through the drivetrain of a vehicle, through boil-off CNG vented directly from a fuel tank to the atmosphere, from leaks during refueling operations, and through the entire supply chain from harvesting to combustion. We echo the sentiments in Delgado and Muncrief (2015) to invest in research and development of solutions to these problems such that fugitive CH₄ emissions from on-vehicle sources and delivery transactions can be minimized.

- **Engage fleet operators and natural gas suppliers to train drivers and service station attendants in proper refueling procedures and establish standard tools and practices to mitigate fugitive emissions.** In fueling and maintenance operations, it is critical that drivers and maintainers of LNG vehicles are aware of the risks posed by CH₄ emissions and are trained to take steps to prevent any leakage from occurring when vehicles fuel tanks contain LNG.

The adoption of alternative fuels, particularly natural gas products, has many potential benefits, and with the adoption and implementation of these recommendations it is possible to increase the impacts LNG offers over diesel fuel.
Chapter 4

A CONTINUOUS NETWORK LOCATION PROBLEM FOR A SINGLE REFUELING STATION ON A TREE²

In this chapter we define and analyze the continuous location problem for a single refueling station on a tree network to provide its properties and an exact solution algorithm that finds the potential set of locations for the station. From previous literature in Section 2.2, we have found that the flow-based demand model is more suitable for a refueling station location problem (Berman et al. 1992, Hodgson 1990). Thus, this dissertation develops modeling and algorithm based on the flow-based demand. In particular, demand in this dissertation is described in a form of the origin and destination (OD) flow pair to avoid a multiple counting issue and to construct the global optimal solution from a whole system viewpoint. Each vertex represents an origin/destination of a demand, and each edge means a path between a vertex pair. Since this chapter focuses on a tree network, the path between any vertex pair is unique. Note that many network location models have been developed based on tree networks in that it is easier to get insight into a tree network location problem prior to more general networks (Francis et al. 1992). In particular, tree-like networks have been usually considered when having a cycle complicates the problem such as a location problem on the interstate highway networks (Tansel et al. 1983). Besides, as discussed in Section 1.2, we can easily observe that interstate highways and toll roads form tree or tree-like networks in many states. Thus, using a tree network for the continuous location model in this chapter may be not only a realistic model but also a stepping stone to greater continuous location models on a general network.

Given the maximum travel distance of vehicles without refueling, our objective in this chapter is to locate the single alternative-fuel refueling station, so as to maximize the total OD flow pairs in round trips per time unit that can be covered by the refueling station. For this, we first provide the theoretical background needed to determine an optimal location of a single alternative-fuel refueling station on a tree network in Section 4.1 to Section 4.3. Then, we present the proposed algorithm in Section 4.4. Section 4.5 describes a simple test network and illustrates the algorithm by solving the continuous location problem for a single refueling station on a tree network.

### 4.1 Problem Statement

In Chapter 4, we aim at locating a single refueling station for alternative-fuel vehicles on a tree network. Let $T(V, E)$ be an undirected (symmetric) tree network with a set of $n$ vertices $V = \{v_1, ..., v_n\}$ and a set $E$ of $n - 1$ edges. An edge $(v_i, v_j) \in E$ is defined as the edge connecting two vertices $v_i$ and $v_j$. Since the network $T$ is undirected and symmetric, the distance between $v_i$ and $v_j$ is denoted as $d(v_i, v_j) = d(v_j, v_i)$. Similarly, $d(v_i, x)$ or $d(x, v_i)$ denotes the distance between $v_i$ and any point $x \in T$. We also denote $P(v_i, v_j) \in T$ as the unique (round trip) path between vertices $v_i$ and $v_j$, $i < j$, for all $v_i, v_j \in V$. Note that an edge $(v_i, v_j)$ is strictly different from $P(v_i, v_j)$ in definition. For example in Figure 4-1, we can define $P(v_i, v_k)$ but not edge $(v_i, v_k)$ because there is no edge connecting vertices $v_i$ and $v_k$ directly.

Before providing more details of the problem statement, the following assumptions are held in this chapter, which are inspired by Boccia et al. (2009), Capar and Kuby (2012), Hwang et al. (2015), Kim and Kuby (2012, 2013), Kuby and Lim (2005, 2007), Kuby et al. (2009), and Lim and Kuby (2010):
**Assumptions**

1. Any vehicle having a full fuel tank is able to travel by a distance $R$ before the next refueling, where $R$ refers to the safe travel distance of any vehicle.

2. A vehicle has at least half of its fuel tank full, i.e., fuel level for traveling a distance $R/2$, when it enters and exits the road network.

3. For their OD pair, vehicles make a complete round trip.

4. No vehicle is able to make a detour from its shortest path.

5. Vehicles traveling in both directions of the segment where a refueling station is located are able to stop by the station for refueling service.

6. The location of refueling station does not influence OD flow paths.

7. We have a complete knowledge of all OD flow paths that carry non-zero flows.

8. Energy consumption is equal in both directions of an OD pair and proportional to mileage.

The first assumption stipulates the maximum covering distance of a demand (vehicle). The first and the second Assumptions imply that either the trip origin of a vehicle is located within a distance $R/2$ of its entrance to the road network and has a private refueling station or there exists a public service refueling station located within a distance $R/2$ before the entrance, where the driver can fill up the tank before entering the road network. In addition, either the trip destination of a vehicle is within a distance $R/2$ from its exit point in the road network and has a private refueling station or there exists a public service refueling station located within a distance $R/2$ from the exit point, where the driver can fill up the tank. Note that the second assumption guarantees that the trip could be repeated the next day, or another trip to a different destination could be completed. The second assumption assures that, if a driver can make the trip from the refueling station to the destination with at least a fuel tank for a distance $R/2$ in the tank, the driver can go back to the refueling station on the same or different trip. In the third assumption, a
“complete round trip” means that, each vehicle returns to its origin by restarting the trip from its destination. For instance, if a vehicle makes a complete round trip between \( v_i \) and \( v_j \), then the total path of a complete round trip is \( v_i \rightarrow v_j \rightarrow v_i \). Next, the fourth assumption implies that a vehicle can only be refueled by a refueling station located within its path. For example, if a vehicle has \( P(v_i, v_j) \) and a refueling station is located between \( v_j \) and \( v_k \), such as \( i < j < k \), as shown in Figure 4-1, then this refueling location cannot cover \( P(v_i, v_j) \) because any deviation from \( P(v_i, v_j) \) is not available by the fourth assumption.

![Figure 4-1. A simple example tree to illustrate the difference between edge and path in definition and the fourth assumption](image)

The fifth assumption stipulates that it is an undirected location problem. Thus, the station can serve vehicles traveling in either direction of a network. The sixth assumption represents that the OD flow pairs data is exogenous (Hodgson 1990). The seventh assumption excludes uncertainty of input data for simplicity of the model (Boccia et al. 2009). Lastly, the aim of the eighth assumption is to ignore hills, winds, or congestion between an OD pair that could affect the energy consumption of a vehicle but to make the energy consumption only depend on the distance between an OD pair. This assumption ensures that if round trip of \( v_i \rightarrow v_j \rightarrow v_i \) is able to be refueled, then the opposite round trip of \( v_j \rightarrow v_i \rightarrow v_j \) is also able to be refueled.

Based on the above definitions and assumptions, we define \( L \) as the set consisting of all possible paths in \( T \), that is,

\[
L = \{P(v_i, v_j) \mid i < j, \text{ for all } v_i, v_j \in V\}.
\]
Note that we do not distinguish between $d(v_i, v_j)$ and the length of $P(v_i, v_j)$ in this chapter since any path $P(v_i, v_j)$ in $T$ is a simple path.

We now define the continuous location problem for a single refueling station on a tree network $T$. Given a safe travel distance $R$, this problem is to find the optimal set of locations for a single refueling station at any point $x^*$ in $T$ so as to maximize the total OD flow pairs (in round trips per time unit) that can be covered by the station. In this chapter, a vehicle in its round trip between $v_i$ and $v_j$ is said to be covered if the refueling station is placed at a point $x$ on the path $P(v_i, v_j)$ within a distance $R/2$ from both $v_i$ (origin of the road network) and $v_j$ (destination of the road network), or equivalently, $x \in P(v_i, v_j)$, $d(v_i, x) \leq R/2$, and $d(x, v_j) \leq R/2$.

Let $f(v_i, v_j)$ denote the average traffic flow (in round trips per time unit) corresponding to path $P(v_i, v_j)$. Since all trips are assumed to be round trips, by the third and the eighth assumptions, we can infer that $f(v_i, v_j) = f(v_j, v_i)$. For convenience, in this chapter we define $f(v_i, v_j)$ only for $i < j$. Then, $S(x)$ can be defined as the set of the paths with positive flows covered by a station located at point $x$:

$$S(x) = \{ P(v_i, v_j) \mid x \in P(v_i, v_j), f(v_i, v_j) > 0, d(v_i, x) \leq R/2, \text{ and } d(x, v_j) \leq R/2, i < j, \text{ for all } v_i, v_j \in V \}.$$ 

Now, let $F(x)$ denote the total traffic amount of OD flow pairs in round trips per time unit covered by a station located at $x \in T$. $F(x)$ can be computed as follows:

$$F(x) = \sum_{P(v_i, v_j) \in S(x)} f(v_i, v_j).$$

By using $F(x)$, we can determine an optimal location point, denoted as $x^*$, to solve this problem as follows:

$$x^* = \max_{x \in T} F(x).$$
By the first assumption, round trips between \( v_i \) and \( v_j \), such that \( d(v_i, v_j) > R \), cannot be covered by a single refueling station. Thus, \( T \) becomes a forest by deleting those edges whose distance is greater than a distance \( R \) from \( T \), and the Forest Property is derived as follows.

**Property 4.1 (Forest Property).** Let \( \hat{E} = \{(v_i, v_j) \mid d(v_i, v_j) > R, \text{for } (v_i, v_j) \in E\} \) and set \( t = |\hat{E}| + 1 \). Then, by removing the set of edges \( \hat{E} \) from tree \( T(V, E) \), \( T \setminus \hat{E} \) becomes a forest \( F(V_F, E_F) \) with \( t \) trees, denoted as \( T_q(V_q, E_q), q = 1, ..., t \), where \( T_q \) is the \( q^{th} \) tree with a set of \( n_q \) vertices \( V_q \subseteq V \) and a set of \( n_q - 1 \) edges \( E_q \subseteq E \). Thus, \( F(V_F, E_F) = \bigcup_{q=1,...,t} T_q(V_q, E_q) \) with \( V_F = \bigcup_{q=1,...,t} V_q \) and \( E_F = \bigcup_{q=1,...,t} E_q \). Furthermore, let \( x^* \) be an optimal solution to the continuous location problem for a single refueling station in tree \( T \). The following two cases can be considered:

(a) If \( f(v_i, v_j) = 0 \), for all \( P(v_i, v_j) \subseteq L \) such that \( d(v_i, v_j) \leq R \), then \( 1 \leq t \leq n \), any point in \( T \) is optimal, and \( F(x^*) = 0 \).

(b) Otherwise, \( 1 \leq t < n \), \( x^* \) must be located on a tree \( T_q \) that includes a path \( P(v_i, v_j) \subseteq L \) such that \( d(v_i, v_j) \leq R \) and \( f(v_i, v_j) > 0 \), and \( F(x^*) > 0 \).

**Proof.** Every edge in a tree is a cut-edge. This implies that, if an edge is removed from a tree, the tree is split into two sub-trees. Thus, if all the edges in \( \hat{E} \) are removed from \( T(V, E) \), the resulting network becomes a forest with \( t \) trees: \( T_q(V_q, E_q), q = 1, ..., t \).

In case (a), if all the edges in \( E \) are longer than \( R \), then \( \hat{E} = E \), forest \( F \) becomes a set of \( n \) isolated vertices, and \( t = n \). Otherwise, there exists at least one edge \( (v_i, v_j) \in E \) such that \( d(v_i, v_j) \leq R \); thus, forest \( F \) contains a tree \( T_q \) such that \( n_q > 1 \), which implies that \( 1 \leq t < n \). In addition, in case (a), since \( f(v_i, v_j) = 0 \), for all \( P(v_i, v_j) \subseteq L \) such that \( d(v_i, v_j) \leq R \), \( S(x) = \emptyset \), for all \( x \in T \). Thus, any point in \( T \) is optimal and \( F(x^*) = 0 \).
Now, in case (b), there exists at least one path \( P(v_i, v_j) \) \( \in L \) such that \( d(v_i, v_j) \leq R \) and \( f(v_i, v_j) > 0 \). In this case, there is a tree \( T_q \) in forest \( F \) that contains path \( P(v_i, v_j) \) with all its edges. This implies that \( n_q > 1 \) and \( 1 \leq t < n \). Note that, if the refueling station is placed at a point \( x \) along an edge in \( E \) or at an isolated vertex in \( F \), then \( S(x) = \emptyset \) and \( F(x) = 0 \). However, if the station is placed in the middle point of path \( P(v_i, v_j) \), denoted as \( y \), then \( P(v_i, v_j) \in S(y) \), which implies that \( S(y) \neq \emptyset \) and \( F(x^*) \geq f(v_i, v_j) > 0 \). Thus, \( x^* \) must be located on a tree \( T_q \) that includes a path \( P(v_i, v_j) \) \( \in L \) such that \( d(v_i, v_j) \leq R \) and \( f(v_i, v_j) > 0 \). □

Since a path \( P(v_i, v_j) \subseteq T_q \) with \( d(v_i, v_j) > R \) cannot be covered by a single refueling station, we state the Reduced Path Property to eliminate those paths from consideration.

**Property 4.2 (Reduced Path Property).** For any tree \( T_q, q = 1, \ldots, t \), a station located at \( x \in T_q \) can only cover paths in \( L_q = \{ P(v_i, v_j) \mid d(v_i, v_j) \leq R, f(v_i, v_j) > 0, \text{and } i < j, \text{for all } v_i, v_j \in V_q \} \). Thus, \( S(x) \subseteq L_q \).

*Proof.* By definition of \( S(x) \), for \( x \in T_q \), \( x \) can only cover a path \( P(v_i, v_j) \), \( v_i, v_j \in V_q \), if

\[
f(v_i, v_j) > 0 \quad \text{and} \quad d(v_i, v_j) \leq R. \]

However, if one of the two vertices \( v_i \) or \( v_j \in V_q \), then

\[
d(v_i, v_j) > R. \]

Also, if none of the two vertices belong to \( V_q \), then \( d(v_i, x) > R \) and \( d(x, v_j) > R \). Thus, \( S(x) \subseteq L_q \). □

The Forest Property and the Reduced Path Property are helpful to reduce the size of the continuous location problem for a single refueling station on a tree.
4.2 Refueling Segment

Recall that in Section 2.1.3.3 Church and Meadows (1979) define a set of points that are potential solutions to the set-covering location problem and maximal covering location problem. By applying their concept to the continuous location problem for a single refueling station on a tree, we introduce a segment that contains all possible locations for a single refueling station that can cover the traffic flow in a path. By the second assumption, a vehicle in a path \( P(v_i, v_j) \) can refuel in a station located within a distance \( \frac{R}{2} \) from \( v_i \). If the distance between \( v_j \) and the station is less than or equal to \( \frac{R}{2} \), the vehicle can also arrive at \( v_j \) with its fuel tank at least half-full. Note that the vehicle is able to fill up at the same station in its return trip from \( v_j \) to \( v_i \). From this observation, the line segment containing the set of station locations that cover path flow \( f(v_i, v_j) \) is defined as follows:

**Definition.** The Refueling Segment of a path \( P(v_i, v_j) \in L_q \) in a tree \( T_q(V_q, E_q) \) of \( T \), denoted as \( RS(v_i, v_j) \), is the segment of \( P(v_i, v_j) \) that contains the potential locations of a refueling station that covers its traffic flow \( f(v_i, v_j) \). That is,

\[
RS(v_i, v_j) = \{ x \in P(v_i, v_j) | d(v_i, x) \leq \frac{R}{2} \text{ and } d(x, v_j) \leq \frac{R}{2} \}.
\]

We describe three different types of \( RS(v_i, v_j) \) regarding \( d(v_i, v_j) \leq R \). Figure 4-2 illustrates the three types of the refueling segment. First, as seen in Figure 4-2(a), if \( 0 < d(v_i, v_j) \leq \frac{R}{2} \), a refueling station located at any point in \( P(v_i, v_j) \) covers \( f(v_i, v_j) \). That is, \( P(v_i, v_j) \) is identical to \( RS(v_i, v_j) \). Next, as shown in Figure 4-2(b), if \( \frac{R}{2} < d(v_i, v_j) < R \), \( RS(v_i, v_j) \) is a line segment in the interior of \( P(v_i, v_j) \). Lastly, as described in Figure 4-2(c), if \( d(v_i, v_j) = R \), \( RS(v_i, v_j) \) is the middle point of \( P(v_i, v_j) \). The refueling segment has one or two endpoints, denoted as \( w^k_{i,j} \) of \( RS(v_i, v_j) \) for \( k = 1, 2 \). The endpoints are calculated as follows:
(a) If \(0 < d(v_i, v_j) \leq R/2\), then \(w^1_{i,j} = v_i\) and \(w^2_{i,j} = v_j\).

(b) If \(R/2 < d(v_i, v_j) < R\), then \(w^1_{i,j}\) is such that \(d(v_i, w^1_{i,j}) = d(v_i, v_j) - R/2\), and \(w^2_{i,j}\) is such that \(d(v_i, w^2_{i,j}) = R/2\). \hfill (4.1)

(c) If \(d(v_i, v_j) = R\), then \(w^1_{i,j} = w^2_{i,j}\) is the middle point of \(P(v_i, v_j)\).

Since each path in \(L_q\) has one or two endpoints, the set of endpoints of \(T_q, q = 1, ..., t\), denoted as \(EP_q\), can be defined as follows:

\[
EP_q = \{w^1_{i,j}, w^2_{i,j} \mid w^1_{i,j} \text{ and } w^2_{i,j} \text{ are endpoints of } RS(v_i, v_j), \text{for all } P(v_i, v_j) \in L_q\}. \hfill (4.2)
\]

\[\]

Figure 4-2. Three types of refueling segments (Ventura et al. 2015)
4.3 Endpoint Optimality Theorem

In this section, we establish a relationship between $S(x)$, for $x \in T_q$, and $S(w)$, for $w \in EP_q$, and derive the Endpoint Optimality Theorem.

**Lemma 4-1.** Given any point $x \in T_q$, for $q = 1, \ldots, t$, there exists an endpoint $w \in EP_q$ in $T_q$ such that $S(x) \subseteq S(w)$.

*Proof.* First, suppose that $x$ does not belong to any refueling segment in $T_q$. Then, by the definition of refueling segment, $x$ does not cover any path in $L_q$, i.e., $S(x) = \emptyset$. Thus, $S(x) \subseteq S(w)$, for any $w \in EP_q$.

Now, suppose that $x$ belongs to a refueling segment $RS(v_i, v_j)$ in $T_q$. $RS(v_i, v_j)$ must have one or two endpoints. If $x$ is one of the endpoints of $RS(v_i, v_j)$, then $S(x) = S(w_{i,j}^1)$ or $S(x) = S(w_{i,j}^2)$. Otherwise, $x$ is interior of all the refueling segments generated by paths in $S(x)$. Thus, $x$ has at least one endpoint on each side. Let $w_1$ and $w_2$ be the two closest endpoints on each side of $x$. Since there are not vertices between $w_1$ and $x$, and between $x$ and $w_2$, all paths covered by $x$ are also covered by both endpoints $w_1$ and $w_2$, and all paths covered both by $w_1$ and $w_2$ are also covered by $x$. This implies that $S(x) = S(w_1) \cap S(w_2)$. Thus, $S(x) \subseteq S(w_1)$ and $S(x) \subseteq S(w_2)$. □

**Theorem 4-1 (Endpoint Optimality Theorem).** There always exists an optimal location $x^*$ for the continuous location problem for a single refueling station on a tree $T$, such that $x^* \in EP_q$, for some $q = 1, \ldots, t$.

*Proof.* (By contradiction) Suppose that none of the endpoints in $EP_q$ is optimal. Let $x^*$ be an optimal point covering $F(x^*)$ path flows (in round trips per time unit). By Lemma 4-1, there
exists an endpoint \( w \in EP_q \) of a refueling segment such that \( S(x^*) \subseteq S(w) \). This implies that \( F(x^*) \leq F(w) \). This contradicts the hypothesis that none of the endpoints in \( EP_q \) is optimal. Therefore, there must exist an endpoint in \( EP_q \) that is optimal. □

Based on Theorem 4-1, the next section provides an algorithm to find the set of optimal endpoints at which the total traffic flow (in round trips per time unit) covered is maximized in tree \( T \).

### 4.4 The Single Refueling Point Algorithm

Theorem 4-1 shows that there exists an endpoint of a refueling segment in a tree \( T_q, q = 1, \ldots, t \), that solves the continuous location problem for a single refueling station on a tree. In this section, we suggest an algorithm to find the set of endpoints in \( T \) that maximizes the total number of round trips per time unit covered. This algorithm is composed of three steps. In the first step, the algorithm finds the set of trees in forest \( F(V_F, E_F) \) from an original tree \( T \). In the next step, for each tree \( T_q(V_q, E_q) \), it determines the set of local optimal endpoints. In the last step, the algorithm selects the set of optimal endpoints of \( T \) considering the optimal local solutions (endpoints) of trees in forest \( F \). The specific steps of the algorithm are provided below.

**Single Refueling Point Algorithm (Ventura et al. 2015)**

Step 1: Determine the set of trees in \( F(V_F, E_F) = \bigcup_{q=1,\ldots,t} T_q(V_q, E_q) \) such that \( V_F = V, E_F = E \setminus \hat{E} \), where \( \hat{E} = \{(v_i, v_j) \mid d(v_i, v_j) > R, \text{ for } (v_i, v_j) \in E\} \), and set \( t = |\hat{E}| + 1 \).

Step 1.1: Construct set \( \hat{E} \) and remove all the edges in \( \hat{E} \) from \( T(V, E) \).

Step 1.2: Identify the set of trees \( T_q(V_q, E_q) \), for \( q = 1, \ldots, t \), where \( n_q \) is the number of vertices in \( V_q \subseteq V_F \) and \( n_q - 1 \) is the number of edges in \( E_q \subseteq E_F \).
Step 2: For each tree $T_q$, $q = 1, ..., t$, with $n_q > 1$, find the set of local optimal points in $EP^*_q$.

(Note that, if $n_q = 1$, then set $EP^*_q = V_q$ is a singleton and the optimal set of endpoints is $EP^*_q = \{x^*_q\}$ with $F(x^*_q) = 0$, where $x^*_q$ is the only endpoint in $EP^*_q$.)

Step 2.1: Generate $L_q = \{P(v_i, v_j) \mid d(v_i, v_j) \leq R, \ f(v_i, v_j) > 0, \text{ and } i < j,\}$
for all $v_i, v_j \in V_q$. 

Step 2.2: Construct the set of endpoints of refueling segments, $EP_q$, using Expression

(4.1).

Step 2.3: Determine $S(w)$ for all $w \in EP_q$:

$$S(w) = \{P(v_i, v_j) \in L_q \mid w \in P(v_i, v_j), d(v_i, w) \leq R/2, \text{and } d(w, v_j) \leq R/2\}.$$ 

Step 2.4: Determine the optimal set of endpoints $EP^*_q$, which consists of a set of
alternative optimal endpoints or a single optimal endpoint, by computing $F(w)$, for $w \in EP_q$:

$$EP^*_q = \arg \max \{F(w) \mid w \in EP_q\}, q = 1, ..., t.$$ 
where $F(w) = \sum_{P(v_i, v_j) \in S(w)} f(v_i, v_j), w \in EP_q$.

Step 3: Determine the optimal set of endpoints $EP^*$ for the original tree $T$ and the total number of
round trips per time unit covered ($F^*$):

$$EP^* = \arg \max \{F(x^*_q) \mid x^*_q \in EP^*_q, q = 1, ..., t\},$$

and $F^* = F(x^*), \text{ for any } x^* \in EP^*$.

Now, the following theorem proves the computational complexity of the proposed
algorithm.

Theorem 4-2. The complexity of the Single Refueling Point Algorithm is of polynomial time of $O(n^4)$, where $n = |V|$. 
Proof. Step 1 of the algorithm first generates edge set $\hat{E}$ with all the edges in $T$ that are longer than $R$ and then generates the set of trees $T_q, q = 1, \ldots, t$, by the Forest Property. This step takes $O(n)$ time because the process searches all the edges in $T$.

Step 2 requires four sub-steps to determine the set of local optimal endpoints for each tree $T_q$. In Sub-Step 2.1, it takes $O(n^2)$ time to check all vertex pairs in each tree $T_q$ to generate the sets of coverable paths, $L_q, q = 1, \ldots, t$. In Sub-Step 2.2, it also takes $O(n^2)$ time to form the sets of endpoints of refueling segments, $EP_q$ in $T_q, q = 1, \ldots, t$. In Sub-Step 2.3, $O(n^4)$ time is required to construct the sets of covered paths for each endpoint in $EP_q, q = 1, \ldots, t$. The last sub-step also requires $O(n^4)$ time to determine the total number of path flows (in round trips per time unit) that can be covered by each endpoint, compare the covered traffic flows, and find the set of local optimal solutions for each tree. Thus, Step 2 runs in $O(n^4)$ time.

In Step 3, comparing the best local solutions to identify the optimal set of endpoints $EP^*$ in $T$ takes $O(n^2)$ time. Thus, since each step only has to be executed one time, the computational complexity of the Single Refueling Point Algorithm is $O(n^4)$. □

In addition to the optimal set of endpoints found by the Single Refueling Point Algorithm, some interior points of the refueling segments corresponding to the paths that are covered by some optimal endpoints can also be optimal solutions. From this observation, two optimality properties of certain line segments defined by consecutive endpoints in $EP^*$ (Lemmas 4-2 and 4-3) are provided as follows.

Lemma 4-2 (Optimal Convexity Property). All points in the line segment $C(w^*_1, w^*_2)$ with endpoints $w^*_1, w^*_2 \in EP^*$, such that $S(w^*_1) = S(w^*_2)$, are also optimal to the continuous location problem for a single refueling station on $T$. 
Proof. Let \( C(w_1^*, w_2^*) = \{ w \mid w = \alpha w_1^* + (1 - \alpha) w_2^*, \ 0 \leq \alpha \leq 1 \} \). Since \( S(w_1^*) = S(w_2^*) \), and by Lemma 4-1, \( S(w) = S(w_1^*) \cap S(w_2^*) \), for any \( w \in C(w_1^*, w_2^*) \). Thus, \( S(w) = S(w_1^*) = S(w_2^*) \).

Therefore, \( F(w) = F(w_1^*) = F(w_2^*) \), for any \( w \in C(w_1^*, w_2^*) \). \( \square \)

The following lemma shows that at most two optimal endpoints in \( EP^* \) cover the same set of paths. This result is useful to characterize the set of all optimal solutions to the continuous location problem for a single refueling station on a tree.

Lemma 4-3 (Path Coverage Property). The number of optimal endpoints in \( EP^* \) with an identical set of covered paths is at most two.

Proof. (By contradiction) Suppose that there exist three optimal endpoints \( w_1^*, w_2^*, w_3^* \in EP^* \), such that \( S(w_1^*) = S(w_2^*) = S(w_3^*) \). Since \( S(w_1^*) = S(w_2^*) \), the line segment \( C(w_1^*, w_2^*) \) can be represented as the intersection of all refueling segments \( RS(v_i, v_j) \), such that \( P(v_i, v_j) \in S(w_k^*), k = 1 \text{ or } 2 \). Thus, since \( w_3^* \) is different than \( w_1^* \) and \( w_2^* \), but covers the same paths, \( w_3^* \) must be in the interior of \( C(w_1^*, w_2^*) \), which is a contradiction to the hypothesis that \( w_3^* \) is an optimal endpoint in \( EP^* \). Thus, the maximum number of optimal endpoints in \( EP^* \) with an identical set of covered paths is at most two. \( \square \)

Now, the following theorem proves the optimality of the Single Refueling Point Algorithm. Based on Lemmas 4-2 and 4-3, this theorem also characterizes and proves the complete set of optimal solutions for the continuous location problem for a single refueling station on a tree.

Theorem 4-3. The Single Refueling Point Algorithm finds the optimal set of endpoints \( EP^* \). Then, \( EP^* \) can be used to construct the complete set of optimal solutions (\( OS^* \)):

\[
OS^* = EP^* \cup \{ Int(C(w_1^*, w_2^*)) \mid S(w_1^*) = S(w_2^*), w_1^*, w_2^* \in EP^* \}
\]
where \(\text{Int}(C(w_1^*, w_2^*))\) is the interior of \(C(w_1^*, w_2^*)\).

**Proof.** By Theorem 4-1, there exists an endpoint of a refueling segment corresponding to a path in \(T\) that solves the continuous location problem for a single refueling station on a tree. Step 1 of the algorithm generates all the trees in forest \(F(V_F, E_F)\). Step 2 finds the set of local optimal endpoints for each tree by complete enumeration. Lastly, Step 3 generates the optimal set of endpoints \(EP^*\) that solves the problem by checking all local optimal endpoints determined in Step 2. Therefore, \(EP^*\) includes all optimal endpoints of \(T\).

Lemmas 4-2 and 4-3 prove that optimal solutions which are not endpoints must lie in the interior of a line segment bounded by consecutive optimal endpoints \(w_1^*, w_2^* \in EP^*\), such that \(S(w_1^*) = S(w_2^*)\). Thus, since the above definition of set \(OS^*\) includes set \(EP^*\) and the interiors of all line segments whose endpoints are consecutive optimal endpoints in \(EP^*\) with identical sets of covered paths, \(OS^*\) must include all optimal solutions to the continuous location problem for a single refueling station on a tree \(T\). □

### 4.5 Numerical Example

In this section, we describe a numerical example of tree \(T\) with seven vertices. The set of optimal endpoints for the continuous location problem for a single refueling station on a tree \(T\) is able to be found using the Single Refueling Point Algorithm. Tree \(T\) is depicted in Figure 4-3. In this example, we assume that the safe travel distance of vehicles \(R = 8\), and the average path flows (in round trips per time unit), \(f(v_i, v_j)\), are given in Table 4-1, for all vertex pairs \(v_i, v_j\), \(i < j\).
Table 4-1. Round trip traffic flows in tree $T$ (Ventura et al. 2015)

<table>
<thead>
<tr>
<th>Origin/Destination</th>
<th>$v_1$</th>
<th>$v_2$</th>
<th>$v_3$</th>
<th>$v_4$</th>
<th>$v_5$</th>
<th>$v_6$</th>
<th>$v_7$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$v_1$</td>
<td>-</td>
<td>30</td>
<td>10</td>
<td>70</td>
<td>15</td>
<td>10</td>
<td>40</td>
</tr>
<tr>
<td>$v_2$</td>
<td>-</td>
<td>-</td>
<td>40</td>
<td>55</td>
<td>20</td>
<td>10</td>
<td>35</td>
</tr>
<tr>
<td>$v_3$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>20</td>
<td>90</td>
<td>25</td>
<td>60</td>
</tr>
<tr>
<td>$v_4$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>30</td>
<td>55</td>
<td>80</td>
</tr>
<tr>
<td>$v_5$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>90</td>
<td>95</td>
</tr>
<tr>
<td>$v_6$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>120</td>
</tr>
<tr>
<td>$v_7$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Unit: average number of round trips per time unit

Figure 4-3. Structure of tree $T$ (Ventura et al. 2015)

In Step 1 of the algorithm, we determine the set of trees in forest $F$. In Sub-Step 1.1, we generate the edge set $\hat{E}$ with all the edges longer than $R = 8$: $\hat{E} = \{(v_5, v_6)\}$ because $d(v_5, v_6) = 9 > R$. Then, in Sub-Step 1.2, forest $F$ is determined: $F(V_F, E_F) = \bigcup_{q=1,2} T_q(V_q, E_q)$ such that
$V_F = V = \{v_1, v_2, v_3, v_4, v_5, v_6, v_7\}$ and $E_F = E \setminus \hat{E} = \{(v_1, v_2), (v_2, v_3), (v_3, v_4), (v_3, v_5), (v_6, v_7)\}$. The two sub-trees are then: $T_1(V_1, E_1)$, where $V_1 = \{v_1, v_2, v_3, v_4, v_5\}$, $E_1 = \{(v_1, v_2), (v_2, v_3), (v_3, v_4), (v_3, v_5)\}$, and $n_1 = 5$, and $T_2(V_2, E_2)$, where $V_2 = \{v_6, v_7\}$, $E_2 = \{(v_6, v_7)\}$, and $n_2 = 2$. Figure 4-4 shows $T_1$ and $T_2$.

![Diagram of sub-trees](image)

(a) Sub-tree $T_1$ from $T$

(b) Sub-tree $T_2$ from $T$

Figure 4-4. Two sub-trees from the original tree (Ventura et al. 2015)

In Step 2, we find the set of local optimal endpoints for each tree $T_q$, for $q = 1, 2$. For tree $T_1$, in Sub-Step 2.1, we determine path set $L_1$ containing all the paths that can be covered by a single refueling station. These paths must have positive flow and cannot be longer than $R = 8$: 
\[ L_1 = \{P(v_1, v_2), P(v_2, v_3), P(v_3, v_4), P(v_3, v_5), P(v_4, v_1), P(v_2, v_4)\}. \]

Note that four paths in tree \( T_1 \), \( P(v_1, v_4), P(v_1, v_5), P(v_2, v_5), \) and \( P(v_4, v_5) \), are excluded from \( L_1 \) because \( d(v_1, v_4) = 10, d(v_1, v_5) = 13, d(v_2, v_5) = 9, \) and \( d(v_4, v_5) = 9 \) are greater than \( R=8 \). In Sub-Step 2.2, we construct the set of endpoints, \( EP_1 \), corresponding to the refueling segments by using Expression (4.1). Since we have six paths in \( L_1 \), we find six refueling segments. Set \( EP_1 \) is defined by the six pairs of endpoint of the refueling segments:

\[ EP_1 = \{w_{1,2}^1, w_{1,2}^2, w_{2,3}^1, w_{2,3}^2, w_{3,4}^1, w_{3,4}^2, w_{3,5}^1, w_{3,5}^2, w_{1,3}^1, w_{2,4}^1, w_{2,4}^2\}. \]

Figure 4-5 graphically shows the location of the refueling segments and corresponding endpoints in tree \( T_1 \). In Sub-Step 2.3, we determine the sets of paths covered by the endpoints in \( EP_1, S(w) \), for \( w \in EP_1 \). Table 4-2 shows the results. Figure 4-5 also graphically summarizes the paths covered by each endpoint in \( EP_1 \).

Table 4-2. Sets of the paths with positive flows covered by endpoints in the first sub-tree

(Ventura et al. 2015)

<table>
<thead>
<tr>
<th>( w )</th>
<th>the set of the paths with positive flows covered by ( w \in EP_1 )</th>
<th>( w )</th>
<th>the set of the paths with positive flows covered by ( w \in EP_1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( w_{1,2}^1 )</td>
<td>( S(w_{1,2}^1) = {P(v_1, v_2)} )</td>
<td>( w_{3,5}^1 )</td>
<td>( S(w_{3,5}^1) = {P(v_3, v_5)} )</td>
</tr>
<tr>
<td>( w_{1,2}^2 )</td>
<td>( S(w_{1,2}^2) = {P(v_1, v_2), P(v_2, v_3), P(v_1, v_3)} )</td>
<td>( w_{3,5}^2 )</td>
<td>( S(w_{3,5}^2) = {P(v_3, v_5)} )</td>
</tr>
<tr>
<td>( w_{2,3}^1 )</td>
<td>( S(w_{2,3}^1) = {P(v_2, v_3), P(v_1, v_2), P(v_1, v_3)} )</td>
<td>( w_{1,3}^1 )</td>
<td>( S(w_{1,3}^1) = {P(v_1, v_3), P(v_1, v_2)} )</td>
</tr>
<tr>
<td>( w_{2,3}^2 )</td>
<td>( S(w_{2,3}^2) = {P(v_2, v_3), P(v_3, v_4), P(v_2, v_4)} )</td>
<td>( w_{1,3}^2 )</td>
<td>( S(w_{1,3}^2) = {P(v_1, v_3), P(v_1, v_2), P(v_2, v_3)} )</td>
</tr>
<tr>
<td>( w_{3,4}^1 )</td>
<td>( S(w_{3,4}^1) = {P(v_3, v_4), P(v_2, v_3), P(v_2, v_4)} )</td>
<td>( w_{2,4}^1 )</td>
<td>( S(w_{2,4}^1) = {P(v_2, v_4), P(v_2, v_3)} )</td>
</tr>
<tr>
<td>( w_{3,4}^2 )</td>
<td>( S(w_{3,4}^2) = {P(v_3, v_4)} )</td>
<td>( w_{2,4}^2 )</td>
<td>( S(w_{2,4}^2) = {P(v_2, v_4), P(v_3, v_4)} )</td>
</tr>
</tbody>
</table>
Sub-Step 2.4 determines the local optimal set of endpoints $EP_1^*$ by computing $F(w)$, for $w \in EP_1$.

A summary of the calculations is provided in Table 4-3.

Figure 4-5. Refueling segments and endpoints in the first sub-tree (Ventura et al. 2015)
By comparing $F(w)$, for $w \in E_{P_1}$, in Table 4-3, we find the local optimal set of endpoints $E_{P_1}^*$ at which the total traffic flow (in round trips per time unit) covered is maximized in tree $T_1$. That is,

$$E_{P_1}^* = \arg \max \{F(w_{1,2}), F(w_{1,2}), F(w_{2,3}), F(w_{2,3}), F(w_{1,3}), F(w_{2,4}), F(w_{2,4})\},$$

$$F(w_{1,2}), F(w_{1,3}), F(w_{2,3}), F(w_{1,3}), F(w_{2,4}), F(w_{2,4})\}$$

$$= \arg \max \{30, 80, 80, 115, 115, 20, 90, 90, 40, 80, 95, 75\} = \{w_{2,3}, w_{3,4}\}.$$

Thus, for tree $T_1$, two endpoints, $w_{2,3}$ and $w_{3,4}$, which cover 115 round trips per time unit in tree $T_1$, define the local optimal set of endpoints $E_{P_1}^*$. 

Table 4-3. Total traffic flow for endpoints in the first sub-tree (Ventura et al. 2015)

<table>
<thead>
<tr>
<th>$w$</th>
<th>Total traffic flow (in round trips per time unit) covered by endpoint $w$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w_{1,2}^1$</td>
<td>$F(w_{1,2}) = f(v_1, v_2) = 30$</td>
</tr>
<tr>
<td>$w_{1,2}^2$</td>
<td>$F(w_{1,2}) = f(v_1, v_2) + f(v_2, v_3) + f(v_1, v_3) = 30 + 40 + 10 = 80$</td>
</tr>
<tr>
<td>$w_{2,3}^3$</td>
<td>$F(w_{2,3}) = f(v_2, v_3) + f(v_1, v_2) + f(v_1, v_3) = 40 + 30 + 10 = 80$</td>
</tr>
<tr>
<td>$w_{2,3}^3$</td>
<td>$F(w_{2,3}) = f(v_2, v_3) + f(v_3, v_4) + f(v_2, v_4) = 40 + 20 + 55 = 115$</td>
</tr>
<tr>
<td>$w_{3,4}^3$</td>
<td>$F(w_{3,4}) = f(v_3, v_4) + f(v_2, v_3) + f(v_2, v_4) = 20 + 40 + 55 = 115$</td>
</tr>
<tr>
<td>$w_{3,4}^3$</td>
<td>$F(w_{3,4}) = f(v_3, v_4) = 20$</td>
</tr>
<tr>
<td>$w_{3,5}^3$</td>
<td>$F(w_{3,5}) = f(v_3, v_5) = 90$</td>
</tr>
<tr>
<td>$w_{3,5}^3$</td>
<td>$F(w_{3,5}) = f(v_3, v_5) = 90$</td>
</tr>
<tr>
<td>$w_{1,3}^3$</td>
<td>$F(w_{1,3}) = f(v_1, v_3) + f(v_1, v_2) = 10 + 30 = 40$</td>
</tr>
<tr>
<td>$w_{1,3}^3$</td>
<td>$F(w_{1,3}) = f(v_1, v_3) + f(v_1, v_2) + f(v_2, v_3) = 10 + 30 + 40 = 80$</td>
</tr>
<tr>
<td>$w_{2,4}^3$</td>
<td>$F(w_{2,4}) = f(v_2, v_4) + f(v_2, v_3) = 55 + 40 = 95$</td>
</tr>
<tr>
<td>$w_{2,4}^3$</td>
<td>$F(w_{2,4}) = f(v_2, v_4) + f(v_3, v_4) = 55 + 20 = 75$</td>
</tr>
</tbody>
</table>
Similarly, we repeat Step 2 for tree $T_2$. There is no path $P(v_i, v_j)$ in tree $T_2$ such that $d(v_i, v_j) > R$. Thus, Sub-Step 2.1 generates $L_2 = \{P(v_6, v_7)\}$. Since $L_2$ has one path, there exists a single refueling segment in tree $T_2$. Sub-Step 2.2 constructs the set of two endpoints corresponding to $RS(v_6, v_7)$. Since $d(v_6, v_7) \leq R/2$, based on Part (a) of Expression (4.1), $RS(v_6, v_7) = P(v_6, v_7)$. Thus, the endpoints are the vertices of this path: $w_{6,7}^1 = v_6$ and $w_{6,7}^2 = v_7$. The location of the refueling segment and corresponding endpoints in tree $T_2$ is illustrated in Figure 4-6. Sub-Step 2.3 determines $S(w)$ for $w \in EP_2$, that is, $S(w_{6,7}^1) = S(w_{6,7}^2) = \{P(v_6, v_7)\}$.

![Figure 4-6. One refueling segment and two endpoints in the second sub-tree (Ventura et al. 2015)](image)

In Sub-Step 2.4, we first compute $F(w)$, for $w \in EP_2$. That is, $F(w_{6,7}^1) = F(w_{6,7}^2) = f(v_6, v_7) = 120$. Then, the local optimal set of endpoints $EP_2^*$ is determined as

$$EP_2^* = \operatorname{arg\,max}\{F(w_{6,7}^1), F(w_{6,7}^2)\} = \operatorname{arg\,max}\{120, 120\} = \{w_{6,7}^1, w_{6,7}^2\}.$$ 

Thus, the two endpoints, $w_{6,7}^1$ and $w_{6,7}^2$, which cover 120 round trips per time unit in tree $T_2$, define the local optimal set of endpoints $EP_2^*$. 
In Step 3, we determine the optimal set of endpoints $EP^*$ for the original tree $T$ that maximizes the total number of round trips per time unit covered ($F^*$):

$$EP^* = \arg \max \{F(x_1^*), F(x_2^*)\}$$

$$= \{ \arg \max \{F(w_{2,3}^2), F(w_{3,4}^2), F(w_{6,7}^1), F(w_{6,7}^2)\} \}$$

$$= \{ \arg \max \{115, 115, 120, 120\} \}$$

$$= \{w_{6,7}^1, w_{6,7}^2\}.$$

This represents that the two endpoints, $w_{6,7}^1$ and $w_{6,7}^2$, are the optimal points for the original tree $T$ with the maximum traffic flow covered of 120 round trips per time unit. Since $w_{6,7}^1$ and $w_{6,7}^2$ cover the same path, i.e., $S(w_{6,7}^1) = S(w_{6,7}^2) = P(v_6, v_7)$, all points in the line segment $C(w_{6,7}^1, w_{6,7}^2)$ are also optimal for tree $T$ by Lemma 4-2. Therefore, we conclude that the two endpoints, $w_{6,7}^1$ and $w_{6,7}^2$, and all the interior points within $C(w_{6,7}^1, w_{6,7}^2)$ are the complete set of optimal solutions; that is,

$$OS^* = C(w_{6,7}^1, w_{6,7}^2).$$

Table 4-4 summarizes the results of the Single Refueling Point Algorithm for the example in Section 4.5.
### Table 4-4. Summary of results for the numerical example (Ventura et al. 2015)

<table>
<thead>
<tr>
<th>Vertex Pair</th>
<th>$T_q$</th>
<th>$L_q$</th>
<th>Refueling Segment</th>
<th>Endpoints</th>
<th>Total Traffic Flow Covered (in round trips per time unit)</th>
<th>Reason for Elimination</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(v_1, v_2)$</td>
<td>$T_1$</td>
<td>$L_1$</td>
<td>$\text{RS}(v_1, v_2)$</td>
<td>$w_{1,2}^1, w_{1,2}^2$</td>
<td>$F(w_{1,2}^1) = 30$, $F(w_{1,2}^2) = 80$</td>
<td>-</td>
</tr>
<tr>
<td>$(v_3, v_4)$</td>
<td>$T_1$</td>
<td>$L_1$</td>
<td>$\text{RS}(v_3, v_4)$</td>
<td>$w_{3,4}^1, w_{3,4}^2$</td>
<td>$F(w_{3,4}^1) = 40$, $F(w_{3,4}^2) = 80$</td>
<td>-</td>
</tr>
<tr>
<td>$(v_5, v_6)$</td>
<td>$T_1$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Reduced Path Property</td>
</tr>
<tr>
<td>$(v_7, v_8)$</td>
<td>$T_1$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Forest Property</td>
</tr>
<tr>
<td>$(v_9, v_{10})$</td>
<td>$T_1$</td>
<td>$L_1$</td>
<td>$\text{RS}(v_9, v_{10})$</td>
<td>$w_{9,10}^1, w_{9,10}^2$</td>
<td>$F(w_{9,10}^1) = 80$, $F(w_{9,10}^2) = 115$</td>
<td>-</td>
</tr>
<tr>
<td>$(v_{11}, v_{12})$</td>
<td>$T_1$</td>
<td>$L_1$</td>
<td>$\text{RS}(v_{11}, v_{12})$</td>
<td>$w_{11,12}^1, w_{11,12}^2$</td>
<td>$F(w_{11,12}^1) = 95$, $F(w_{11,12}^2) = 75$</td>
<td>-</td>
</tr>
<tr>
<td>$(v_1, v_7)$</td>
<td>$T_1$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Reduced Path Property</td>
</tr>
<tr>
<td>$(v_5, v_8)$</td>
<td>$T_1$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Forest Property</td>
</tr>
<tr>
<td>$(v_9, v_{10})$</td>
<td>$T_1$</td>
<td>$L_1$</td>
<td>$\text{RS}(v_9, v_{10})$</td>
<td>$w_{9,10}^1, w_{9,10}^2$</td>
<td>$F(w_{9,10}^1) = 20$, $F(w_{9,10}^2) = 115$</td>
<td>-</td>
</tr>
<tr>
<td>$(v_{11}, v_{12})$</td>
<td>$T_1$</td>
<td>$L_1$</td>
<td>$\text{RS}(v_{11}, v_{12})$</td>
<td>$w_{11,12}^1, w_{11,12}^2$</td>
<td>$F(w_{11,12}^1) = 90$, $F(w_{11,12}^2) = 90$</td>
<td>-</td>
</tr>
<tr>
<td>$(v_1, v_7)$</td>
<td>$T_1$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Forest Property</td>
</tr>
<tr>
<td>$(v_5, v_8)$</td>
<td>$T_1$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Forest Property</td>
</tr>
<tr>
<td>$(v_9, v_{10})$</td>
<td>$T_1$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Reduced Path Property</td>
</tr>
<tr>
<td>$(v_{11}, v_{12})$</td>
<td>$T_1$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Forest Property</td>
</tr>
<tr>
<td>$(v_1, v_7)$</td>
<td>$T_1$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Forest Property</td>
</tr>
<tr>
<td>$(v_{11}, v_{12})$</td>
<td>$T_1$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Forest Property</td>
</tr>
<tr>
<td>$(v_{13}, v_{14})$</td>
<td>$T_2$</td>
<td>$L_2$</td>
<td>$\text{RS}(v_{13}, v_{14})$</td>
<td>$w_{13,14}^1, w_{13,14}^2$</td>
<td>$F(w_{13,14}^1) = 120$, $F(w_{13,14}^2) = 120$</td>
<td>-</td>
</tr>
</tbody>
</table>

**Set of optimal solutions:** $OS^* = C(w_{6,7}^1, w_{6,7}^2)$

**Total traffic flow:** $F(x^*) = 120$ round trips per time unit, $x^* \in OS^*$
4.6 Concluding Remarks

In this chapter, the continuous location problem for a single refueling station on a tree network has been addressed. To identify the optimal set of locations on the network with the objective of maximizing the total traffic flow (in round trips per time unit) covered by the station, we have first derived two properties, i.e., the Forest Property and the Reduced Path Property, which are able to contribute to reduce the problem size. Next, the concept of the refueling segment of a path has been defined, and the Endpoint Optimality Theorem has been derived and proved to theoretically verify that at least one optimal solution is found at one of the endpoints of the refueling segments for this problem. Based on this theorem, the Single Refueling Point Algorithm, which runs on polynomial time, has been developed to determine the set of optimal endpoints where the total traffic flow (in round trips per time unit) covered by the station is maximized. Lastly, the complete set of optimal solutions has been built by identifying additional optimal points located in the interior of a line segment surrounded by some optimal endpoints. A small test example has been solved to illustrate the proposed algorithm.
Chapter 5

A CONTINUOUS DEVIATION-FLOW LOCATION PROBLEM FOR AN ALTERNATIVE-FUEL REFUELING STATION ON A TREE NETWORK

In Chapter 5, we address the continuous location problem for a single refueling station on a tree-type transportation network when a given portion of drivers are willing to deviate to be able to refuel their vehicles. A number of optimality properties regarding deviation are derived. Based on these properties, a novel algorithm is suggested to determine the deviation options and the set of potential locations for the refueling station covering the traffic flow for each origin-destination pair. Then, an exact polynomial-time algorithm is proposed to find the set of optimal station locations that cover the maximum traffic flow in the network. We also construct the complete set of optimal solutions by identifying additional optimal points located in the interior of paths which endpoints belong to the preliminary set of optimal points. Lastly, we solve a numerical example to illustrate the proposed approach, examine the coupled effects of deviation portion and vehicle driving range on the set of optimal locations and the maximum traffic flow covered, and analyze its performance.

5.1 Introduction

The U.S. national highway system significantly contributes to the country’s economic development. In particular, it plays a major role in providing mobility for goods and services. Using highways, trucks carry the largest shares by value, tons, and ton-miles for shipments moving 750 or fewer miles in the U.S. (Chambers et al. 2015). As concerns over climate change

---

are increasingly becoming a global issue, logistics companies are interested in replacing their
diesel trucks by alternative-fuel trucks (Wisniewski 2016). However, to introduce alternative-fuel
trucks to logistics companies successfully, a well-designed refueling infrastructure in the highway
system is a prerequisite.

According to the cyclomatic number that indicates the number of circuits in a network,
highway systems can be classified into two types: circuit networks and tree networks (Haggett
and Chorley 1969, Xie and Levinson 2007). A circuit is defined as a closed walk that begins and
ends on the same vertex without passing over the same edge twice. A circuit network contains at
least one circuit in its structure and has a positive cyclomatic number. On the other hand, a tree
network does not contain any circuit in its structure and has a zero cyclomatic number. A set of
tree networks is called a forest. Every edge in a tree network is a cut-edge which removal forms a
forest with two components.

While the U.S. national highway system consists of a set of circuit networks, if it is
partitioned into local highway systems by operating authority, then many of them form trees or
tree-like networks. Table 5-1 summarizes local highway systems in the U.S. by operating
authority that have tree or tree-like network structures. As shown in this table, 10 turnpikes and 8
portions of interstate highways within certain states constitute tree structures (Federal Highway
Administration 2016b). Furthermore, if we consolidate beltways and cycles within metro areas
into single vertices, then 11 more portions of interstate highways within certain states can also be
considered as tree-like structures (Federal Highway Administration 2016a, Francis et al. 1992).
We can also observe that toll roads and highways in other countries, such as Chile, Croatia
(southern and northeast regions), Indonesia, Ireland (excluding the Dublin region), Malaysia,
Norway (northern region), Philippines, Slovakia, Spain (excluding the Catalan region), and
Serbia, to name a few, have tree or tree-like network structures. Note also that road networks in
sparsely settled areas are generally trees, since tree road networks are the cheapest to construct.
Trees are central to the structural understanding of networks and graphs, and often occur with additional attributes such as roots and vertex-ordering. They have a wide range of applications, including data storage, searching, information processing, and facility location (Francis et al. 1992, West 2001). Because it is easier to get insight into tree network problems, several papers in facility location problems on transportation networks without cycles have been published (Chan and Francis 1976, Chen et al. 1988, Erkut et al. 1988, Tansel et al. 1980, Tansel et al. 1982).

Due to the financial risk, it is unusual to build multiple alternative-fuel refueling stations on a transportation network at the same time. Instead, an operating authority may be more inclined in setting up a single station first in a high traffic area and planning the gradual future construction of additional stations with the goal of maximizing the coverage of traffic flow. For example, the Pennsylvania Turnpike Commission (PTC) opened the first refueling station at New Stanton Service Plaza in 2014 to serve compressed natural gas (CNG) vehicles and it currently remains as the only CNG station on the turnpike (Pennsylvania Turnpike Commission 2014). As a second example, the Oklahoma Turnpike Authority opened the first CNG station at Stroud Travel Plaza on the turnpike in 1991, which was renovated in 2014, and then 23 years later, the authority opened the second CNG station at McAlester Travel Plaza on the turnpike in 2014 (Beaty 2014, Marks 2014, United Press International 1991). Note that both of these turnpikes form tree networks.

Drivers (customers) using a toll road regard such transportation network as being independent from others due to the toll payment system; that is, since drivers need to pay tolls when they exit the road, they try to avoid intermediate exits for refueling service before arriving to their preplanned destinations. Due to such driving behavior, operating authorities of toll roads are interested in providing refueling service for the drivers without considering competition with other existing refueling stations located outside their own networks (Hwang et al. 2015, Myers et
Table 5-1. Tree networks of highways in the U.S.

<table>
<thead>
<tr>
<th>State</th>
<th>Road Type (*)</th>
<th>Type of Network Structure</th>
<th>Component Highways</th>
<th>Operating Authority</th>
</tr>
</thead>
<tbody>
<tr>
<td>AL</td>
<td>I</td>
<td>Tree except beltway in Birmingham</td>
<td>I-20, I-22, I-59, I-65, I-85</td>
<td>AL Department of Transportation (DOT)</td>
</tr>
<tr>
<td>AK</td>
<td>S</td>
<td>Tree in Anchorage</td>
<td>AK-1, AK-3</td>
<td>AK DOT and Public Facilities</td>
</tr>
<tr>
<td>AZ</td>
<td>I</td>
<td>Tree except beltway and cycles in Phoenix</td>
<td>I-8, I-10, I-15 I-17, I-19, I-40</td>
<td>AZ DOT</td>
</tr>
<tr>
<td>AR</td>
<td>I</td>
<td>Tree except cycles in Little Rock</td>
<td>I-30, I-40, I-49, I-55, I-530, I-555</td>
<td>AR Highway and Transportation Department</td>
</tr>
<tr>
<td>CO</td>
<td>T</td>
<td>Tree except beltway in Denver</td>
<td>I-25, I-70, I-76</td>
<td>CO DOT</td>
</tr>
<tr>
<td>GA</td>
<td>T</td>
<td>Tree except beltway and cycles in Atlanta, Augusta, Columbus, and Macon</td>
<td>I-16, I-20, I-59, I-75, I-85, I-185, I-285, I-675, I-985</td>
<td>GA DOT</td>
</tr>
<tr>
<td>ID</td>
<td>I</td>
<td>Tree</td>
<td>I-15, I-84, I-86, I-90</td>
<td>ID Transportation Department</td>
</tr>
<tr>
<td>IN</td>
<td>T</td>
<td>Tree</td>
<td>I-80, I-90</td>
<td>IN Finance Authority/ ITR Concession Company LLC</td>
</tr>
<tr>
<td>KS</td>
<td>T</td>
<td>Tree</td>
<td>I-35, I-70, I-335, I-470</td>
<td>KS Turnpike Authority</td>
</tr>
<tr>
<td>ME</td>
<td>T</td>
<td>Tree</td>
<td>I-95</td>
<td>ME Turnpike Authority</td>
</tr>
<tr>
<td>MA</td>
<td>T</td>
<td>Tree</td>
<td>I-90</td>
<td>MA DOT</td>
</tr>
<tr>
<td>MN</td>
<td>I</td>
<td>Tree except cycles in Minneapolis</td>
<td>I-35, I-90, I-94</td>
<td>MN DOT</td>
</tr>
<tr>
<td>MS</td>
<td>I</td>
<td>Tree</td>
<td>I-20, I-22, I-55, I-59</td>
<td>MS DOT</td>
</tr>
<tr>
<td>MT</td>
<td>I</td>
<td>Tree</td>
<td>I-15, I-90, I-94</td>
<td>MT DOT</td>
</tr>
<tr>
<td>NE</td>
<td>I</td>
<td>Tree except cycles in Omaha</td>
<td>I-76, I-80</td>
<td>NE Department of Roads</td>
</tr>
<tr>
<td>NV</td>
<td>I</td>
<td>Tree except cycles in Las Vegas</td>
<td>I-15, I-80, I-515</td>
<td>NV DOT</td>
</tr>
<tr>
<td>NH</td>
<td>T</td>
<td>Tree</td>
<td>I-93, I-95, NH-3, NH-16</td>
<td>NH DOT</td>
</tr>
<tr>
<td>NM</td>
<td>I</td>
<td>Tree</td>
<td>I-10, I-25, I-40</td>
<td>NM DOT</td>
</tr>
<tr>
<td>NY</td>
<td>T</td>
<td>Tree</td>
<td>I-87, I-90</td>
<td>NY State Thruway Authority</td>
</tr>
<tr>
<td>ND</td>
<td>I</td>
<td>Tree</td>
<td>I-29, I-94</td>
<td>ND DOT</td>
</tr>
<tr>
<td>OH</td>
<td>T</td>
<td>Tree</td>
<td>I-76, I-80, I-90</td>
<td>OH Turnpike Commission</td>
</tr>
<tr>
<td>OR</td>
<td>I</td>
<td>Tree except cycles in Portland</td>
<td>I-5, I-84</td>
<td>OR DOT</td>
</tr>
<tr>
<td>SD</td>
<td>I</td>
<td>Tree</td>
<td>I-29, I-90</td>
<td>SD DOT</td>
</tr>
<tr>
<td>UT</td>
<td>I</td>
<td>Tree except cycles in Salt Lake City</td>
<td>I-15, I-70, I-80, I-84</td>
<td>UT DOT</td>
</tr>
<tr>
<td>VT</td>
<td>I</td>
<td>Tree</td>
<td>I-89, I-91, I-93</td>
<td>VT Agency of Transportation</td>
</tr>
<tr>
<td>WV</td>
<td>T</td>
<td>Tree</td>
<td>I-64, I-77</td>
<td>WV Parkways Authority</td>
</tr>
<tr>
<td>WY</td>
<td>I</td>
<td>Tree</td>
<td>I-25, I-80, I-90</td>
<td>WY DOT</td>
</tr>
</tbody>
</table>

(*) I: Interstate highway, S: State highway, T: Turnpike.
For example, the PTC has been operating the first and only CNG station on the turnpike at New Stanton Service Plaza, even if another CNG station is available nearby New Stanton Service Plaza, but outside the turnpike (Alternative Fuels Data Center 2016). Considering the operating authorities’ approach, in this chapter we try to find the set of optimal locations for the refueling station within the transportation network of interest and without considering any possible competition from existing stations located nearby the given network.

Regarding the single facility location problem on tree networks, a number of papers have been published (Chan and Francis 1976, Dearing 1977, Dearing and Francis 1974, Francis 1976, Goldman 1971, Goldman 1972, Goldman and Witzgall 1970, Halfin 1974, Handler 1973, Kariv and Hakimi 1979, Lin 1975, Rosenthal and Pino 1989, Slater 1981, Zelinka 1968). Such single facility location problems have been extended by Averbakh (2003), Barahona and Jensen (1998), Batta (1988 and 1989), Berman et al. (1985), Chen (2001), Gao (2012), Pelegrín et al. (2006), Puerto and Rodríguez-Chía (2011), Üster and Love (2002), and Zaferanieh et al. (2008) to consider more various network environments. None of them is, however, suitable for solving the single refueling station location problem because the demand in their models is node-based without a limited driving range and/or their objectives are to minimize the (demand-weighted) average, total, or maximum distance between demand points and the new facility. On the contrary, considering refueling behavior, demand in the refueling station location problem needs to be path-based with a limited driving range. Furthermore, the objective of maximizing the total traffic flow covered by the station is more desirable by the operating authority interested in locating the station.

Refueling station location problems can be classified into two types depending on the set of candidate sites: when a preliminary (finite) set of candidate sites is given, this problem is called discrete; when the stations can be located anywhere along the network, the problem is called continuous. Church and Meadows (1979) suggested the continuous version of the set-
covering and maximal covering location problems when demand is node-based, and Ventura et al. (2015) presented the continuous version of the refueling station location problem when demand is path-based. Since an optimal solution to the discrete version of the problem is always a feasible solution to the continuous version, an optimal solution to the continuous refueling station location problem can achieve a considerable improvement in terms of demand coverage. The continuous version of the problem is, however, more challenging than its discrete counterpart due to the size of the search space.

In this study, we consider the continuous version of the single refueling station location problem on a tree network, where drivers have the option of selecting a deviation path if the refueling station is not located along their preplanned routes. We believe that this is the first research work on the continuous version of the problem where the deviation option is considered. The deviation option is expected to be very common for drivers with alternative-fuel vehicles during the introductory period of alternative-fuel vehicles to the market since alternative-fuel refueling infrastructures are usually underdeveloped. The deviation option is also available to drivers of conventional fossil-fuel vehicles (powered by gasoline and diesel engines) when they cannot find a refueling station on their routes. Thus, considering deviation will make the problem more practical in deciding the potential location for the refueling station.

By allowing vehicle deviation in the continuous refueling station location problem, the solution approach proposed in this chapter is capable of finding all possible alternative paths for any origin-destination pair and improving suboptimal solutions obtained by the existing path-based demand models. The existing path-based demand models that consider deviation (Berman et al. 1995a, Huang et al. 2015, Kim and Kuby 2012) choose locations from a preliminary (finite) set of candidate sites (vertices). Thus, they enable deviation only if such vertices can be reached within the vehicle driving range. Since they treat the problem as a discrete location problem, there is a chance that their solutions do not contain a global optimum (Kuby et al. 2005). On the other
hand, the proposed model considers all possible deviation paths in the network regardless of the location of vertices on detours and the station can be located anywhere along any detour edge. Thus, the proposed approach improves any possible sub-optimality produced by other existing models and generates the entire set of global optima. The main differences between the proposed model and the existing path-based demand models are summarized in Table 5-2.

Table 5-2. Comparison between the proposed model and the existing path-based demand models considering deviation

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Base Model</strong></td>
<td>Flow-capturing location model <em>p</em>-median model</td>
<td>Flow-refueling location model</td>
<td>Set-covering model</td>
<td>Continuous refueling station location model</td>
</tr>
<tr>
<td><strong>Objective</strong></td>
<td>Maximize covered demand Minimize total deviation distance traveled</td>
<td>Maximize covered demand</td>
<td>Minimize total cost of locating stations</td>
<td>Maximize covered demand</td>
</tr>
<tr>
<td><strong>Main Constraints</strong></td>
<td>Number of facilities</td>
<td>Number of refueling stations Vehicle driving range</td>
<td>Cover all demand. Vehicle driving range</td>
<td>Single refueling station. Vehicle driving range</td>
</tr>
<tr>
<td><strong>Set of Candidate Sites</strong></td>
<td>Predetermined Finite</td>
<td>Predetermined Finite</td>
<td>Predetermined Finite</td>
<td>Transportation network Infinite</td>
</tr>
<tr>
<td><strong>Deviation Paths</strong></td>
<td>Only deviation paths with vertices that can be reached within driving range</td>
<td>Only deviation paths with vertices that can be reached within driving range</td>
<td>Only deviation paths with vertices that can be reached within driving range</td>
<td>All deviation paths</td>
</tr>
<tr>
<td><strong>Global Optimum</strong></td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Based on the characteristics of the continuous station location problem with deviation, we first develop a methodology to determine the complete set of optimal solutions for a single refueling station on a tree network that covers the maximum traffic flow (in round trips per time unit), when the deviation option is available. Later, the effect of the portion of drivers that are willing to deviate on the set of optimal station locations is analyzed, and the results of the proposed model are compared with those of the existing models to verify its performance. This research aims to become the stepping stone to more general continuous station location problems including locating multiple refueling stations on circuit networks.
The rest of this chapter is organized as follows. Section 5.2 includes the problem statement and derives a property related to the deviation option. Section 5.3 identifies a set of candidate points for potential locations of the refueling station. Section 5.4 determines optimal points from the set of candidate points for the refueling station that covers the maximum traffic flow when a given portion of drivers select the deviation option. Section 5.5 builds the complete set of optimal solutions by identifying additional optimal points located in the interior of paths which endpoints belong to the preliminary set of optimal points. Section 5.6 provides a numerical example to illustrate the proposed approach, examine the coupled effects of deviation portion and vehicle driving range on the set of optimal locations and the maximum traffic flow covered, and analyze its performance. Lastly, Section 5.7 presents a summary of this study and a list of topics for future research.

5.2 Problem Statement

Now, we address the deviation-flow continuous location problem on a tree network. Considering that a portion of drivers are willing to deviate from their paths, the objective of this problem is to determine the set of locations for a single refueling station that maximizes the total traffic flow covered (in round trips per time unit) by the station. The following key assumptions are considered in this problem:

Assumptions

(i) A single uncapacitated refueling station can provide service to all vehicles driving through both directions of the road segment where it is located.

(ii) All vehicles have the same fuel tank size. Fuel consumption is also the same for all vehicles and is a linear function of the driving distance.
(iii) Vehicles perform a complete round trip between their origin and destination points on the network.

(iv) Each vehicle enters and exits the network with a fuel tank that is at least half-full.

(v) The deviation option from a simple path is available to all drivers for their refueling service. However, only a certain portion $\alpha$ of the drivers selects this option, where $0 \leq \alpha \leq 1$.

The first assumption assures that all vehicles passing through the refueling station in either direction of the road segment can be refueled at the station. The second assumption means that all vehicles have the same driving range per refueling, denoted as $R$, regardless of road conditions, climate, congestion or any other variables. A complete round trip in the third assumption ensures that every vehicle goes back to its origin after arriving at its intended destination. A trip from origin to destination is called an original trip, and a trip from destination to origin is called a return trip. A round trip consists of an original trip and a return trip. The fourth assumption infers that drivers should be able to refuel their vehicles within a distance $R/2$ from their origin and also destination points to complete a round trip. Lastly, the fifth assumption implies that $\alpha \times 100\%$ of traffic flow deviates from their simple paths for refueling service. We call this flow the deviation-flow.

Let $T(V, E)$ be an undirected tree network consisting of a set $V$ with $n$ vertices and a set $E$ with $n - 1$ edges, where $n \geq 2$; otherwise, the network is trivial. An edge $(v_i, v_j) \in E$ is defined if $v_i \in V$ and $v_j \in V$ are directly connected. We also denote $P(v_i, v_j)$ as the unique simple path between $v_i$ and $v_j$ for $i < j$, for all $v_i, v_j \in V$. Set $L$ is defined as the set of all possible paths in $T$, i.e., $L = \{P(v_i, v_j) \mid i < j, \text{for all } v_i, v_j \in V\}$. The average traffic flow along $P(v_i, v_j)$ is denoted as $f(v_i, v_j)$. We note that $f(v_i, v_j)$ refers to the average of traffic flow of the original trip from $v_i$ to $v_j$ and the return trip from $v_j$ to $v_i$, and is defined only for $i < j$. We use $f(v_i, v_j)$ to represent the round trips per time unit between $v_i$ and $v_j$. The length of $P(v_i, v_j)$ is
denoted as \(d(v_i, v_j)\), and \(d(v_i, v_i) = d(v_j, v_i)\). Similarly, \(P(v_i, x)\) denotes the unique simple path between \(v_i \in V\) and any point \(x \in T\), and the length of this path is denoted as \(d(v_i, x) = d(x, v_i)\).

To address the deviation-flow continuous location problem, we distinguish between the traffic flow and the deviation-flow on a path. The traffic flow between \(v_i\) and \(v_j\) is “covered” by a single refueling station if the station is located at some point \(x\) along \(P(v_i, v_j)\) within a distance \(R/2\) from \(v_i\) and \(v_j\) (Ventura et al. 2015). Then, the set of paths with positive traffic flow covered by a point \(x \in T\), denoted as \(S(x)\), is defined as follows:

\[
S(x) = \{ P(v_i, v_j) \mid x \in P(v_i, v_j), f(v_i, v_j) > 0, d(v_i, x) \leq R/2, \text{ and } d(x, v_j) \leq R/2, i < j, \text{ for all } v_i, v_j \in V \}.
\]

Alternatively, the deviation-flow originating in path \(P(v_i, v_j)\) is covered by the station if it is placed at a point \(x\) somewhere in the network within a distance \(R/2\) from \(v_i\) and \(v_j\), but not along path \(P(v_i, v_j)\). Then, the set of paths with positive deviation-flow covered by a point \(x \in T\), denoted as \(S_D(x)\), is defined as follows:

\[
S_D(x) = \{ P(v_i, v_j) \mid x \in T \setminus P(v_i, v_j), \alpha \times f(v_i, v_j) > 0, d(v_i, x) \leq R/2, \text{ and } d(x, v_j) \leq R/2, i < j, \text{ for all } v_i, v_j \in V \}.
\]

Now, the total traffic flow (in round trips per time unit), including \(f(v_i, v_j)\) and \(\alpha \times f(v_i, v_j)\), for all \(v_i, v_j \in V\), covered by a single refueling station at \(x \in T\), denoted as \(F(x)\), is calculated as follows:

\[
F(x) = \sum_{P(v_i, v_j) \in S(x)} f(v_i, v_j) + \sum_{P(v_i, v_j) \in S_D(x)} \alpha \times f(v_i, v_j).
\]

Thus, an optimal point \(x^*\) that covers the maximum traffic flow (in round trips per time unit) in tree \(T\) can be obtained by comparing the values of \(F(x)\) for all \(x \in T\); that is, \(x^* \in \arg \max \{ F(x) \mid x \in T \} \).
When a refueling station is not located in a simple path, vehicles traveling along this path can deviate to a separate sub-path for their refueling service. In a deviation path, this separate sub-path is called a symmetric cycle, because the sub-path begins at a vertex in the simple path, reaches the station location, and returns to the same vertex using the same sub-path in the opposite direction. This closed walk includes at least one repeating vertex. Since we consider a single refueling station to be located on the tree network, vehicles can take only one symmetric cycle originating at a vertex located on their simple path and the entire path is called a simple deviation path. That is, each simple deviation path consists of one simple path and one symmetric cycle.

Figure 5-1 shows an example of an undirected tree network to help understand the concept of simple deviation path, as well as sets $S(x)$ and $S_D(x)$. The numerical value next to each edge of the network indicates the distance between the two vertices or between a vertex and a point. In this example, let us assume $R = 80$ and $\alpha = 0.20$. If drivers make a round trip between $v_1$ and $v_3$ with $f(v_1, v_3) = 100$, the simple path is $P(v_1, v_3)$, the vertex visitation sequence between $v_1$ and $v_3$ is $\{v_1, v_2, v_3\}$, and 100 round trips (per time unit) use that vertex visitation sequence. However, if a refueling station is located at point $r$ in the middle of edge $(v_4, v_6)$, 20 round trips from $f(v_1, v_3)$ deviate to a symmetric cycle with vertex visitation sequence $\{v_2, v_4, r, v_4, v_2\}$. Thus, the entire vertex visitation sequence for the simple deviation path between $v_1$ and $v_3$ is $\{v_1, v_2, v_4, r, v_4, v_2, v_3\}$. If there exists another traffic flow $f(v_4, v_6) = 400$, then $S(r) = \{P(v_4, v_6)\}$ and $S_D(r) = \{P(v_1, v_2)\}$. Thus, the total traffic flow covered by a refueling station located at $r$ is calculated as $F(r) = \sum_{P(v_i, v_j) \in S(r)} f(v_i, v_j) + \sum_{P(v_i, v_j) \in S_D(r)} \alpha \times f(v_i, v_j) = f(v_4, v_6) + 0.20 \times f(v_1, v_3) = 420$ round trips (per time unit).
When there exists an edge in the network whose length is greater than $R$, no traffic flow going through the edge can be covered by a single refueling station. This is because vehicles must have at least a half-full tank at their origin and destination vertices in their round trips by Assumptions (iii) and (iv). Furthermore, no deviation-flow going through this edge can be covered by a station located in any symmetric cycle. Thus, we need to split the tree network into sub-trees by removing all edges whose distance is greater than $R$, and solve the station location problem in each sub-tree separately. Let $\hat{E} = \{(v_i, v_j) | d(v_i, v_j) > R, (v_i, v_j) \in E\}$. Then, we eliminate all edges in $\hat{E}$ from $T(V, E)$ to form a forest $F(V_F, E_F)$ with $V_F = V$ and $E_F = E\setminus \hat{E}$, such that $F(V_F, E_F) = \bigcup_{q=1}^{t} T_q(V_q, E_q)$, where $T_q$ is the $q^{th}$ sub-tree that consists of a set of vertices $V_q$ and a set of edges $E_q$, and $t = |\hat{E}| + 1$ is the total number of sub-trees $T_q$ in forest $F(V_F, E_F)$ (Ventura et al. 2015).

Now, for sub-tree $T_q$, $q = 1, ..., t$, if there exists a symmetric cycle originating at a vertex located on a simple path in $T_q$ that can reach some segments in the removed edges in $\hat{E}$, the corresponding deviation-flow can be covered by a point located on one of the segments in these edges. In this case, we need to construct an expanded sub-tree $\hat{T}_q$ from $T_q$ by adding these reachable edges into $T_q$. Thus, for each sub-tree $T_q$, we first need to identify all vertices in $V_q$ that
are endpoints of the edges in $\hat{E}$. Let $v_k$ be one of the vertices in $V_q$ that is adjacent to $v_l \in V \setminus V_q$ by $(v_k, v_l) \in \hat{E}$. Then, the set of vertices $v_k$ in $V_q$ is defined as $V_q' = V(\hat{E}) \cap V_q$, where $V(\hat{E})$ is the set of vertices in $\hat{E}$. After identifying $V_q'$, the set of edges in $\hat{E}$ that are connected to $v_k$ in $V_q'$ is defined as $E(v_k) = \{(v_k, v_l) \in \hat{E} \mid v_l \in V \setminus V_q\}$. In addition, the set of vertices $v_l \in V \setminus V_q$ that are adjacent to $v_k$ in $V_q'$ is defined as $V(v_k) = V(E(v_k)) \setminus \{v_k\}$, where $V(E(v_k))$ is the set of vertices in $E(v_k)$. For $v_k \in V_q'$, if any positive deviation-flow $\alpha \times f(v_i, v_j)$ from path $P(v_i, v_j)$ in $T_q$ can be covered by some point $x$ located on one of the edges in $E(v_k)$, then we add the edge in $E(v_k)$ into $E_q$ and the corresponding end point (vertex) in $V(v_k)$ into $V_q$. That is, for sub-tree $T_q(V_q, E_q), q = 1, \ldots, t$, we first initialize set $V_q$ as an empty set, where $V_q$ is used to collect all $v_k \in V_q'$ such that $\alpha \times f(v_i, v_j), v_i, v_j \in V_q$, can be covered by some point $x$ located on one of the edges in $E(v_k)$. Next, for $v_k \in V_q'$, if we find at least one $\alpha \times f(v_i, v_j) > 0$, for $P(v_i, v_j) \in L$ in $T_q$, such that $\max \{d(v_i, v_k), d(v_j, v_k)\} < R/2$, then we add $v_k$ into $V_q$. Finally, we construct expanded sub-tree $\hat{T}_q(V_q, \hat{E}_q)$, such that $V_q = V_q \cup \{\cup_{v_k \in V_q} V(v_k)\}$ and $E_q = E_q \cup \{\cup_{v_k \in V_q} E(v_k)\}$. Based on the expanded sub-tree $\hat{T}_q$ constructed above for each sub-tree $T_q$, we can derive the following property:

**Property 5-1 (Expanded Sub-tree Property).** Let $x^*$ denote an optimal point to the deviation-flow continuous location problem for a single refueling station on $T$. Regarding the number of expanded sub-trees from $T$ and the location of $x^*$ as well as the value of $F(x^*)$, we consider the following two cases:

(a) If $f(v_i, v_j) = 0$, for all $P(v_i, v_j) \in L$ such that $d(v_i, v_j) \leq R$, then the number of expanded sub-trees $t \leq n, x^*$ can be located anywhere in $T$, and $F(x^*) = 0$. 


(b) Otherwise, the number of expanded sub-trees $t < n$, and $x^*$ has to be placed on an expanded sub-tree $\hat{T}_q$ that contains some path $P(v_i, v_j) \in L$ such that $f(v_i, v_j) > 0$ and $d(v_i, v_j) \leq R$.

In this case, $F(x^*) > 0$.

**Proof.** In case (a), if $d(v_i, v_j) > R$, for all $(v_i, v_j) \in E$, then $\hat{E} = E$. This implies that $n$ expanded sub-trees $\hat{T}_q$ are created, where each $\hat{T}_q$ is composed of an isolated vertex. Otherwise, at least an edge $(v_i, v_j) \in E$ such that $d(v_i, v_j) = R$ exists. Thus, there exists an expanded sub-tree $\hat{T}_q$ that contains edge $(v_i, v_j) \in \hat{E}_q$. This implies that $\hat{T}_q$ includes at least two vertices, i.e., $v_i, v_j \in \hat{V}_q$; thus, the number of expanded sub-trees $t < n$. Besides, since $f(v_i, v_j) = 0$, for all $P(v_i, v_j) \in L$ such that $d(v_i, v_j) \leq R$, $S(x) = S_D(x) = \emptyset$, and $F(x) = 0$, for all $x \in T$. Thus, for any point $x \in T$, $F(x) = 0$, which means that $x^*$ can be located anywhere in $T$ and $F(x^*) = 0$.

In case (b), there exists at least one path $P(v_i, v_j) \in L$, such that $f(v_i, v_j) > 0$ and $d(v_i, v_j) \leq R$. Thus, there exists an expanded sub-tree $\hat{T}_q$ that contains $P(v_i, v_j)$ and at least two vertices $v_i, v_j \in \hat{V}_q$. This means that the number of expanded sub-trees $t < n$. If the refueling station is located at a point $x$ at an isolated vertex in $\hat{T}_q$, then $S(x) = S_D(x) = \emptyset$, and $F(x) = 0$. If the station is located at a point $x$ along a symmetric cycle of $P(v_i, v_j)$, then $P(v_i, v_j) \in S_D(x)$; that is, $S_D(x) \neq \emptyset$, and $F(x) \geq \alpha \times f(v_i, v_j) \geq 0$. If the station is in the middle point $x$ of $P(v_i, v_j)$, then $P(v_i, v_j) \in S(x)$, which indicates that $S(x) \neq \emptyset$, and therefore, $F(x) \geq f(v_i, v_j) > 0$. Since $x^*$ is an optimal point, $F(x^*) \geq F(x) \geq f(v_i, v_j) \geq \alpha \times f(v_i, v_j) \geq 0$.

Hence, $x^*$ has to be placed on an expanded sub-tree $\hat{T}_q$ that contains a path $P(v_i, v_j) \in L$ such that $f(v_i, v_j) > 0$ and $d(v_i, v_j) \leq R$. Then, $F(x^*) > 0$. □

Note that, for any expanded sub-tree $\hat{T}_q(\hat{V}_q, \hat{E}_q)$, $q = 1, \ldots, t$, a path $P(v_i, v_j) \in \hat{T}_q$ such that $d(v_i, v_j) > R$ cannot be covered by a single station; in addition, a path $P(v_i, v_j) \in \hat{T}_q$
without flow, i.e., \( f(v_i, v_j) = 0 \), does not need to be considered to maximize traffic flow coverage (Ventura et al. 2015). Thus, the set of paths that need to be considered for coverage in \( \mathcal{T}_q \) is defined as \( \mathcal{L}_q = \{ P(v_i, v_j) \in \mathcal{T}_q | d(v_i, v_j) \leq R, f(v_i, v_j) > 0, \text{and } i < j, \text{for all } v_i, v_j \in \mathcal{V}_q \} \). Then, by definition of \( S(x) \), for \( x \in \mathcal{T}_q \), \( x \) covers any path \( P(v_i, v_j) \), for \( v_i, v_j \in \mathcal{V}_q \), such that \( d(v_i, v_j) \leq R \) and \( f(v_i, v_j) > 0 \). Thus, \( S(x) \subseteq \mathcal{L}_q \). Similarly, by definition of \( S_D(x) \), for \( x \in \mathcal{T}_q \setminus P(v_i, v_j) \), \( x \) covers every path \( P(v_i, v_j) \), for \( v_i, v_j \in \mathcal{V}_q \), such that \( d(v_i, v_j) \leq R \) and \( \alpha \times f(v_i, v_j) > 0 \). Therefore, \( S_D(x) \subseteq \mathcal{L}_q \) as well.

5.3 Set of Candidate Point

In this section, we establish a set of candidate points for each expanded sub-tree \( \mathcal{H}_q \) to locate a single refueling station on a tree network when a given portion of drivers are willing to deviate to be able to refuel their vehicles. Then, we prove that this set includes at least one optimal location for the deviation-flow continuous location problem for a single refueling station. In addition, if multiple optimal locations exist, these candidate points can be used to generate the entire set of optimal locations.

The rest of this section is organized as follows. For path \( P(v_i, v_j) \in \mathcal{L}_q \), \( q = 1, \ldots, t \).

Section 5.3.1 defines a segment that includes all potential locations for a single refueling station to cover the traffic flow in the path. Based on each segment identified in Section 5.3.1, Section 5.3.2 searches for points where a symmetric cycle can start. Given any point found in Section 5.3.2, Section 5.3.3 introduces an algorithm that identifies the farthest point for each symmetric cycle. By integrating the theoretical background regarding the deviation option suggested in Sections 5.3.1 to 5.3.3, Section 5.3.4 determines the set of candidate points when \( \alpha \times 100\% \) of
drivers selects the deviation option. Lastly, Section 5.3.5 proves that this set contains at least one optimal location.

5.3.1 Refueling Segment

In order to obtain the set of candidate points, the first step we need to do is to identify the segment for each simple path \( P(v_i, v_j) \) that contains all station locations that cover the round trips in the path when the deviation option is not considered. This segment is called the refueling segment of path \( P(v_i, v_j) \), and is denoted as \( RS(v_i, v_j) \). Since drivers have to refuel the vehicles within a distance of \( R/2 \) from the origin and destination vertices during their round trips by Assumption (iv), \( RS(v_i, v_j) \) must contain all the points that are within \( R/2 \) from \( v_i \) and \( v_j \) (Ventura et al. 2015). That is,

\[
RS(v_i, v_j) = \{ x \in P(v_i, v_j) \mid d(v_i, x) \leq R/2 \text{ and } d(x, v_j) \leq R/2 \}.
\]

Depending on the value of \( d(v_i, v_j) \), \( RS(v_i, v_j) \) includes one or two endpoints, denoted as \( w^k_{i,j} \), for \( k = 1, 2 \). If \( d(v_i, v_j) < R \), \( RS(v_i, v_j) \) has two different endpoints, \( w^1_{i,j} \) and \( w^2_{i,j} \); otherwise, \( w^1_{i,j} = w^2_{i,j} \). Then, the set of endpoints of \( RS(v_i, v_j) \), denoted as \( EP(v_i, v_j) \), is defined as follows:

\[
EP(v_i, v_j) = \{ w^k_{i,j} \mid w^k_{i,j}, \text{ for } k = 1, 2, \text{ are endpoints of } RS(v_i, v_j) \}.
\]

In the example of the round trip between \( v_1 \) and \( v_3 \) in Figure 5-1, \( RS(v_1, v_3) \) is identical to \( P(v_1, v_3) \) since \( d(v_1, v_3) < R/2 \). Thus, the two endpoints of \( RS(v_1, v_3) \), \( w^1_{1,3} \) and \( w^2_{1,3} \), are located exactly at \( v_1 \) and \( v_3 \); that is, \( EP(v_1, v_3) = \{ v_1, v_3 \} \). Since \( RS(v_1, v_3) = P(v_1, v_3) \), a station located at any point \( x \in P(v_1, v_3) \) can cover round trips between \( v_1 \) and \( v_3 \).
5.3.2 Cycle Starting Vertex

As a next step to build the set of candidate points for $T_q$, $q = 1, \ldots, t$, considering the deviation option, in this section, we determine a vertex at which a symmetric cycle begins its deviation from a simple path. This vertex is called a cycle starting vertex and denoted as $v_s$. A cycle starting vertex is the only common point (vertex) between the simple path and the symmetric cycle.

In order to identify a cycle starting vertex of a path $P(v_i, v_j)$, we examine the minimum remaining travel distance of vehicles at a vertex within $P(v_i, v_j)$, as well as the degree of this vertex. First, the minimum remaining travel distance of vehicles at a cycle starting vertex $v_s$ is denoted as $\delta(v_i, v_j; v_s)$ and computed as follows:

$$\delta(v_i, v_j; v_s) = \frac{R}{2} - \max\{d(v_i, v_s), d(v_j, v_s)\}. \quad (5.1)$$

Intuitively, $\delta(v_i, v_j; v_s)$ measures the minimum remaining travel distance at $v_s$ that vehicles can drive up when they either enter the network at $v_i$ in the original trip or at $v_j$ in the return trip. $\delta(v_i, v_j; v_s)$ is calculated by subtracting the maximum distance between $d(v_i, v_s)$ and $d(v_j, v_s)$ from $R/2$ because any point within a symmetric cycle must be reachable from both $v_i$ and $v_j$ in order for vehicles to make a complete round trip between $v_i$ and $v_j$ in the network. In order for vehicles in path $P(v_i, v_j) \in \hat{L}_q$ to start a symmetric cycle originating at $v_s$, the value of $\delta(v_i, v_j; v_s)$ must be positive; otherwise, the deviation option is not available at $v_s$. The value of $\delta(v_i, v_j; v_s)$ depends on the location of $v_s$ and the value of $d(v_i, v_j)$. In case of $v_s \in P(v_i, v_j) \setminus RS(v_i, v_j)$, $\delta(v_i, v_j; v_s) < 0$, for any value of $d(v_i, v_j)$; thus, no deviation is available. In case of $v_s \in RS(v_i, v_j)$, if $R/2 \leq d(v_i, v_j) \leq R$, then $\delta(v_i, v_j; v_s) > 0$ for $v_s$ in the interior of $RS(v_i, v_j)$, and $\delta(v_i, v_j; v_s) = 0$ for $v_s$ at the endpoint of $RS(v_i, v_j)$. If $(v_i, v_j) < R/2$, then
\[ \delta(v_i, v_j; v_s) > 0 \text{ for } v_s \in V(RS(v_i, v_j)), \text{ where } V(RS(v_i, v_j)) \text{ is the set of vertices in } RS(v_i, v_j). \]

Next, the degree of a cycle starting vertex \( v_s \) is denoted as \( \text{deg}(v_s) \). If \( v_s \) of a path \( P(v_i, v_j) \) is placed at either origin \( v_i \) or destination \( v_j \), i.e., \( v_s \in \{v_i, v_j\} \), then \( v_s \) can be a cycle starting vertex if and only if \( \text{deg}(v_s) \geq 2 \). This implies that \( v_s \) has at least one adjacent edge that does not belong to path \( P(v_i, v_j) \) and a portion of this edge or the entire edge can form a sub-path for a symmetric cycle. If \( v_s \) is placed at neither origin \( v_i \) nor destination \( v_j \), i.e., \( v_s \notin \{v_i, v_j\} \), then \( v_s \) can only be a cycle starting vertex if \( \text{deg}(v_s) \geq 3 \). That is, besides the two sub-paths from \( v_s \) to \( v_j \) in the original trip and from \( v_s \) to \( v_i \) in the return trip, at least one more separate sub-path connected to \( v_s \) exists to initiate a symmetric cycle.

Since there can exist multiple cycle starting vertices for a path \( P(v_i, v_j) \in \hat{L}_\alpha \), we define \( CSV(v_i, v_j) \) as the set of cycle starting vertices of path \( P(v_i, v_j) \). Based on the observations of a cycle starting vertex discussed above, \( CSV(v_i, v_j) \) is determined as follows:

\[
CSV(v_i, v_j) = \left\{ v_s \in V(RS(v_i, v_j)) \mid \delta(v_i, v_j; v_s) > 0; \text{ and } \delta(v_i, v_j; v_s) \geq 2 \text{ if } v_s \in \{v_i, v_j\}, \text{ or } \delta(v_i, v_j; v_s) \geq 3 \text{ if } v_s \notin \{v_i, v_j\} \right\}
\]

We note that \( CSV(v_i, v_j) \) is defined for \( 0 \leq \alpha \leq 1 \). The reason we define \( CSV(v_i, v_j) \) for \( \alpha > 0 \) is clear for the deviation option. Alternatively, the reason we define \( CSV(v_i, v_j) \) for \( \alpha = 0 \) is discussed with more details in Section 5.3.4.

Now we recall that \( RS(v_1, v_3) = P(v_1, v_3) \) in the example of the round trip between \( v_1 \) and \( v_3 \) in Figure 5-1. Thus, \( V(RS(v_1, v_3)) = \{v_1, v_2, v_3\} \), and by using Expression (5.1), \( \delta(v_1, v_3; v_1) = 2, \delta(v_1, v_3; v_2) = 20, \) and \( \delta(v_1, v_3; v_3) = 2 \). However, \( v_1 \) and \( v_3 \) cannot be cycle starting vertices of \( P(v_1, v_3) \) because they are origin and destination vertices of \( P(v_1, v_3) \),
and $\deg(v_1) = \deg(v_3) = 1$. In contrast, $v_2$ is neither origin nor destination vertex of $P(v_1, v_3)$ and $\deg(v_2) = 3$. Thus, $CSV(v_1, v_3) = \{v_2\}$, meaning that $v_2$ is the only cycle starting vertex of $P(v_1, v_3)$.

### 5.3.3 Cycle Returning Point

In this section, for each symmetric cycle originating at $v_s \in CSV(v_i, v_j)$, we identify the farthest point vehicles can reach before their returning to $v_s$. This point is called a cycle returning point and denoted as $r$. Considering that the purpose for deviation in this chapter is to refuel vehicles, the cycle returning point is regarded as the farthest feasible site for the refueling station; that is, drivers that are willing to deviate can travel up to the cycle returning point, refuel their vehicles at the refueling station located at the cycle returning point, and then return to their simple path head to their destination.

The cycle returning point has distinct characteristics compared to the cycle starting vertex. First, the cycle returning point is defined for $0 < \alpha \leq 1$, while the cycle starting vertex is defined for $0 \leq \alpha \leq 1$. Next, the cycle returning point belongs to the symmetric cycle, but does not belong to the simple path, while the cycle starting vertex is the only common point between the simple path and the symmetric cycle. In addition, the cycle returning point can be a vertex or any point on an edge, whereas the cycle starting vertex is always a vertex literally.

Multiple symmetric cycles can begin at the same cycle starting vertex according to the network structure. This implies that one cycle starting vertex may lead to multiple cycle returning points. Thus, we let $CRP(v_i, v_j; v_s)$ be the set of all cycle returning points $r$ arising from a given $v_s \in CSV(v_i, v_j)$. Then, $|CRP(v_i, v_j; v_s)|$ is the number of cycle returning points corresponding to a given $v_s \in CSV(v_i, v_j)$, or equivalently, the number of symmetric cycles originating at a
given \( v_s \in CSV(v_i, v_j) \). In addition, the number of all cycle returning points of path \( P(v_i, v_j) \) can be computed as \( \sum_{v_s \in CSV(v_i, v_j)} |CRP(v_i, v_j; v_s)| \). We note that, since cycle returning points are defined for \( 0 < \alpha \leq 1 \), set \( CRP(v_i, v_j; v_s) \) is also defined for \( 0 < \alpha \leq 1 \); otherwise, this set is empty.

The location of a cycle returning point \( r \in CRP(v_i, v_j; v_s) \) is determined by comparing the value of \( \delta(v_i, v_j; v_s) \) to the length of the separate sub-path originating at \( v_s \). If the value of \( \delta(v_i, v_j; v_s) \) is less than the length of the separate sub-path, then the cycle returning point \( r \) is located at a distance \( \delta(v_i, v_j; v_s) \) from \( v_s \) since drivers at cycle starting vertex \( v_s \) can travel up to this distance from \( v_s \) before refueling their vehicles; otherwise, the cycle returning point \( r \) is located at the leaf (end vertex) of the separate sub-path because that vertex is the farthest point from \( v_s \) in the symmetric cycle originating at \( v_s \).

From the observation above regarding the location of a cycle returning point, we propose a novel algorithm, called the Cycle Returning Point Algorithm, to identify the locations of all cycle returning points \( r \in CRP(v_i, v_j; v_s) \) for a given cycle starting vertex \( v_s \in CSV(v_i, v_j) \). This algorithm systematically explores the edges along separate sub-paths originating at \( v_s \) and computes the minimum remaining travel distance at each reachable vertex. Note that it is a generalized version of an algorithm suggested by Kweon et al. (2017b).

When exploring an edge \((v_u, v_r)\) along a separate sub-path such that \( d(v_i, v_u) < d(v_i, v_r) \) or \( d(v_j, v_u) < d(v_j, v_r) \), we call \( v_u \) the parent of \( v_r \), and \( v_r \) the child of \( v_u \). To establish a parent and child relationship between vertices in the course of scanning the edges, we let \( PARENT(v_r) \) be the set of parents of \( v_r \), and \( CHILDREN(v_u) \) the set of children of \( v_u \). In a tree network, every child \( v_r \) has a single parent \( v_u \), i.e., \( PARENT(v_r) = \{v_u\} \). In contrast, every parent \( v_u \) can have several children \( v_r \) according to the tree network structure. Thus, set \( CHILDREN(v_u) \) is determined as follows:
The algorithm iterates to explore each edge \((v_u, v_r)\) along separate sub-paths originating at a given vertex \(v_s \in CSV(v_i, v_j)\) until all cycle returning points \(r \in CRP(v_i, v_j; v_s)\) are identified. At a particular iteration exploring an edge \((v_u, v_r)\), the minimum remaining travel distance of vehicles at \(v_u\), denoted as \(\delta(v_i, v_j; v_u)\), is known. That is, if \(v_u = v_s\), the value of \(\delta(v_i, v_j; v_u)\) is given by Expression (5.1); otherwise, this value is given from the previous iteration. Then, we can compute the minimum remaining travel distance of vehicles at \(v_r\) as \\
\[
\delta(v_i, v_j; v_r) = \delta(v_i, v_j; v_u) - d(v_u, v_r).
\]
If \(\delta(v_i, v_j; v_r) < 0\), a cycle returning point \(r\) is located along edge \((v_u, v_r)\) such that \(d(v_u, r) = \delta(v_i, v_j; v_u)\). If \(\delta(v_i, v_j; v_r) = 0\) or if \(\delta(v_i, v_j; v_r) > 0\) and \(deg(v_r) = 1\), then \(r\) is placed exactly at \(v_r\). If \(\delta(v_i, v_j; v_r) > 0\) and \(deg(v_r) \geq 2\), then the drivers that are willing to deviate can travel further along the separate sub-path; in this case, the algorithm assigns a child \(v_r\) of the current iteration to a new parent \(v_u\) at the next iteration and determines a new child \(v_r\) using Expression (5.2) to proceed the next iteration.

In order to keep track of children \(v_r\) from previous iterations that need to be parents in future iterations to search for cycle returning points, this algorithm uses a (infinite) first-in, first-out queue \(Q\) consisting of \(v_r\) such that \(\delta(v_i, v_j; v_r) > 0\) and \(deg(v_r) \geq 2\). Two functions are used to manage \(Q\). The first function named \(enqueue(Q; v_r)\) places \(v_r\) at the tail of queue \(Q\).
Conversely, the second function named `dequeue(Q)` selects the vertex at the head of queue `Q` and eliminates it from `Q`. The details of the Cycle Returning Point Algorithm are provided below.

**Cycle Returning Point Algorithm**

Step 1: For a given \( v_s \in CSV(v_i, v_j) \), initialize \( CRP(v_i, v_j; v_s) = \emptyset \). If \( \alpha = 0 \), terminate the algorithm; otherwise, compute \( \delta(v_i, v_j; v_s) \) using Expression (5.1).

Step 2: Initialize \( Q = \emptyset \).

Step 3: \( enqueue(Q; v_s) \).

Step 4: Repeat the following sub-steps as long as there remain vertices in `Q`:

- **Sub-step 4.1**: \( u = dequeue(Q) \).
- **Sub-step 4.2**: Determine \( CHILDREN(v_u) \) using Expression (5.2).
- **Sub-step 4.3**: For each \( v_r \in CHILDREN(v_u) \), perform the following sub-steps:
  - **Sub-step 4.3.1**: Compute \( \delta(v_i, v_j; v_r) = \delta(v_i, v_j; v_u) - d(v_u, v_r) \).
  - **Sub-step 4.3.2**: According to the sign of \( \delta(v_i, v_j; v_r) \), perform one of the following procedures:
    - (a) If \( \delta(v_i, v_j; v_r) < 0 \): \( r \) is located along edge \( (v_u, v_r) \), such that \( d(v_u, r) = \delta(v_i, v_j; v_u) \). Also, add \( r \) into \( CRP(v_i, v_j; v_s) \).
    - (b) If \( \delta(v_i, v_j; v_r) = 0 \) or if \( \delta(v_i, v_j; v_r) > 0 \) and \( deg(v_r) = 1 \): \( r \) is located exactly at \( v_r \). Also, add \( r \) into \( CRP(v_i, v_j; v_s) \).
    - (c) If \( \delta(v_i, v_j; v_r) > 0 \) and \( deg(v_r) \geq 2 \): \( enqueue(Q; v_r) \) and set \( PARENT(v_r) = \{v_u\} \).

**Theorem 5-1 (Complexity of the Cycle Returning Point Algorithm)**. The complexity of the Cycle Returning Point Algorithm is \( O(n) \), where \( n = |V| \).
Proof. Step 1 includes initialization of \( CRP(v_i, v_j; v_s) \), checking the value of \( \alpha \), and computation of \( \delta(v_i, v_j; v_s) \), which takes \( O(1) \). Initialization of \( Q \) in Step 2 and enqueuing \( v_s \) to \( Q \) in Step 3 also take \( O(1) \). Step 4 consists of three sub-steps to identify all \( r \in CRP(v_i, v_j; v_s) \) for a given \( v_s \in CSV(v_i, v_j) \). Sub-steps 4.1 and 4.2 dequeue \( v_u \) from \( Q \) and generate set \( CHILDREN(v_u) \).

Sub-step 4.3 computes \( \delta(v_i, v_j; v_r) \) for \( v_r \in CHILDREN(v_u) \), and then according to the sign of \( \delta(v_i, v_j; v_r) \), it locates a cycle returning point within edge \((v_u, v_r)\) or enqueues \( v_r \) to \( Q \) for the next iteration. Step 4 iterates until queue \( Q \) is empty. Since every vertex can be enqueued and dequeued at most once for a given \( v_s \in CSV(v_i, v_j) \), the operations in Step 4 can be applied to every vertex at most once; thus, the total time devoted to the operations in Step 4 is \( O(n) \).

Therefore, the computational complexity of the Cycle Returning Point Algorithm is \( O(n) \). \( \square \)

Now, we apply the Cycle Returning Point Algorithm to the example in Figure 5-1 to determine all cycle returning points of \( P(v_1, v_3) \); that is, we identify \( CRP(v_1, v_3; v_s) \), for all \( v_s \in CSV(v_1, v_3) \). Recall that \( R = 80, \alpha = 0.20 \), and \( v_s \in CSV(v_1, v_3) = \{v_2\} \). In Step 1 of the algorithm, for a given cycle starting vertex \( v_2 \), we set \( CRP(v_1, v_3; v_2) = \emptyset \) and, by Expression (5.1), compute \( \delta(v_1, v_3; v_2) = 40 - \max\{d(v_1, v_2), d(v_3, v_2)\} = 20 \). In Step 2, we set \( Q = \emptyset \).

In Step 3, we place \( v_2 \) at the tail of queue \( Q \), i.e., \( Q = \{v_2\} \). Now, we repeat Step 4 until finding all cycle returning points \( r \in CRP(v_1, v_3; v_2) \). In Sub-step 4.1, we set \( v_u = v_2 \) and eliminate \( v_2 \) from \( Q \). In Sub-step 4.2, since \( v_u = v_2 = v_s \), by Expression (5.2), \( CHILDREN(v_2) = N(v_2) \setminus V(P(v_1, v_3)) = \{v_1, v_3, v_4\} \setminus \{v_1, v_2, v_3\} = \{v_4\} \). For \( v_r = v_4 \), we perform Sub-steps 4.3.

In Sub-step 4.3.1, \( \delta(v_1, v_3; v_4) = \delta(v_1, v_3; v_2) - d(v_2, v_4) = 5 \). Since \( \delta(v_1, v_3; v_4) > 0 \) and \( deg(v_4) = 3 \), we select Procedure (c) in Sub-step 4.3.2, i.e., \( Q = \{v_4\} \) and \( PARENT(v_4) = \{v_2\} \). Since \( Q \neq \emptyset \), we repeat Step 4. Now, in Sub-step 4.1, \( v_u = v_4 \) and \( v_4 \) is removed from \( Q \). In
Sub-step 4.2, by Expression (5.2), \( CHILDREN(v_4) = N(v_4) \setminus PARENT(v_4) = \{v_2, v_5, v_6\} \setminus \{v_2\} = \{v_5, v_6\} \). For \( v_r \in \{v_5, v_6\} \), we implement Sub-steps 4.3. For \( v_r = v_5 \), in Sub-step 4.3.1, \( \delta(v_1, v_3; v_5) = \delta(v_1, v_3; v_4) - d(v_4, v_5) = -5 \). As \( \delta(v_1, v_3; v_5) < 0 \), we select Procedure (a) in Sub-step 4.3.2; that is, we locate \( r_1 \) along edge \((v_4, v_5)\), such that \( d(v_4, r_1) = \delta(v_1, v_3; v_4) = 5 \), and add \( r_1 \) to \( CRP(v_3, v_2) \), where \( r_1 \) indicates the first cycle returning point \( r \in CRP(v_1, v_3; v_2) \). By repeating the same process for \( v_r = v_6 \), we locate \( r_2 \) along edge \((v_4, v_6)\), such that \( d(v_4, r_2) = 5 \), and add \( r_2 \) to \( CRP(v_1, v_3; v_2) \), where \( r_2 \) refers to the second \( r \in CRP(v_1, v_3; v_2) \). Now, \( Q = \emptyset \), so we end the algorithm with \( CRP(v_1, v_3; v_2) = \{r_1, r_2\} \).

### 5.3.4 Refueling Sub-tree

In this section, we determine the refueling sub-tree for each path \( P(v_i, v_j) \in \tilde{L}_q \) that includes all possible locations for a single refueling station when a given portion of drivers select the deviation option. Then, we identify candidate points in this sub-tree.

First, let us define symmetric cycle \( SC(v_i, v_j; v_s, r) \) as the segment consisting of all station locations that cover the deviation-flow of path \( P(v_i, v_j) \) originating at cycle starting vertex \( v_s \in CSV(v_i, v_j) \) and ending at cycle returning point \( r \in CRP(v_i, v_j; v_s) \), i.e., \( SC(v_i, v_j; v_s, r) = P(v_s, r) \). Note that any point \( x \) in \( SC(v_i, v_j; v_s, r) \setminus \{v_s\} \) covers deviation-flow \( \alpha \times f(v_i, v_j) \), while cycle starting vertex \( v_s \) covers traffic flow \( f(v_i, v_j) \) since \( v_s \in SC(v_i, v_j; v_s, r) \cap RS(v_i, v_j) \).

Now, for any path \( P(v_i, v_j) \in \tilde{L}_q \), we can construct the refueling sub-tree, denoted as \( RST(v_i, v_j) \), which consists of refueling segment \( RS(v_i, v_j) \) and all symmetric cycles \( SC(v_i, v_j; v_s, r) \) of \( P(v_i, v_j) \). This refueling sub-tree contains all potential locations for a single
refueling station when a given portion of drivers are willing to deviate from the preplanned route $P(v_i, v_j)$ to be able to refuel their vehicles. That is,

$$\text{RST}(v_i, v_j) = \bigcup_{v_s \in CSV(v_i, v_j)} \left( \bigcup_{r \in CRP(v_i, v_j; v_s)} SC(v_i, v_j; v_s, r) \right) \cup RS(v_i, v_j).$$

In $\text{RST}(v_i, v_j)$, the number of $RS(v_i, v_j)$ is one, but the number of $SC(v_i, v_j; v_s, r)$ is $\sum_{v_s \in CSV(v_i, v_j)} |CRP(v_i, v_j; v_s)|$, because the number of symmetric cycles in $\text{RST}(v_i, v_j)$ is equal to the number of cycle returning points $r \in CRP(v_i, v_j; v_s)$, for $v_s \in CSV(v_i, v_j)$.

The amount of traffic flow in $P(v_i, v_j)$ covered by $x \in \text{RST}(v_i, v_j)$ depends on the value of $\alpha$. If $\alpha = 1$, any point $x$ in $\text{RST}(v_i, v_j)$ covers $f(v_i, v_j)$. If $0 < \alpha < 1$, any point $x$ in $\text{RST}(v_i, v_j) \setminus RS(v_i, v_j)$ covers $\alpha \times f(v_i, v_j)$, while any point $x$ in $RS(v_i, v_j)$ covers $f(v_i, v_j)$. If $\alpha = 0$, then $CRP(v_i, v_j; v_s) = \emptyset$ by the Cycle Returning Point Algorithm; thus, $\text{RST}(v_i, v_j) = RS(v_i, v_j)$, and any point $x$ in $\text{RST}(v_i, v_j)$ covers $f(v_i, v_j)$.

The endpoints of $RS(v_i, v_j)$ indicate the boundary points defining the segment containing all station locations that can cover the positive traffic flow in path $P(v_i, v_j)$. Similarly, the endpoints of $SC(v_i, v_j; v_s, r)$, which are $v_s$ and $r$, represent the boundary points for the segment encompassing all station locations that can cover the positive deviation-flow in path $P(v_i, v_j)$ originating at cycle starting vertex $v_s \in CSV(v_i, v_j)$ and ending at cycle returning point $r \in CRP(v_i, v_j; v_s)$. In $\text{RST}(v_i, v_j)$, the endpoints of $RS(v_i, v_j)$ and all the endpoints of $SC(v_i, v_j; v_s, r)$ are called the candidate points because these points indicate the boundaries for potential station locations to cover path $P(v_i, v_j)$ with positive traffic flow or deviation-flow. The set of candidate points of $RST(v_i, v_j)$, denoted as $CP\left(\text{RST}(v_i, v_j)\right)$, is defined as follows:
\[ CP \left( RST(v_i, v_j) \right) = \left\{ \bigcup_{v_s \in CSV(v_i, v_j)} CRP(v_i, v_j; v_s) \right\} \cup CSV(v_i, v_j) \cup EP(v_i, v_j), \]  
\[(5.3)\]

where \( \left\{ \bigcup_{v_s \in CSV(v_i, v_j)} CRP(v_i, v_j; v_s) \right\} \) and \( CSV(v_i, v_j) \) refer to the endpoints of all symmetric cycles of \( P(v_i, v_j) \) and \( EP(v_i, v_j) \) the endpoints of refueling segment of \( P(v_i, v_j) \). The \( k^{th} \) candidate point of \( RST(v_i, v_j) \) is denoted as \( c_{i,j}^k, k = 1, \ldots, \Sigma_{v_s \in CSV(v_i, v_j)} |CRP(v_i, v_j; v_s)| + |CSV(v_i, v_j)| + |EP(v_i, v_j)|. \)

In Expression (5.3), we recall that \( CRP(v_i, v_j; v_s) \) is defined for \( 0 < \alpha \leq 1 \), while \( CSV(v_i, v_j) \) and \( EP(v_i, v_j) \) are defined for \( 0 \leq \alpha \leq 1 \). Note that the vertices in \( CSV(v_i, v_j) \) still need to be considered although the deviation option is not available, i.e., for \( \alpha = 0 \), because some of these points are intersection vertices that may cover traffic flows from multiple paths (Kweon et al. 2015). Since an intersection vertex within a refueling segment has the same characteristics of a cycle starting vertex, we construct set \( CSV(v_i, v_j) \) for \( \alpha = 0 \) to identify this type of intersection vertices even though the deviation option is not available.

By using Expression (5.3), the set of candidate points in expanded sub-tree \( \hat{T}_q, q = 1, \ldots, t \), denoted as \( CP_q \), can be defined as follows:
\[ CP_q = \bigcup_{P(v_i, v_j) \in \hat{T}_q} CP \left( RST(v_i, v_j) \right). \]  
\[(5.4)\]

Now, we determine the refueling sub-tree for path \( P(v_1, v_3) \) in Figure 5-1. Recall that \( CSV(v_1, v_3) = \{v_2\} \) and \( CRP(v_1, v_3; v_2) = \{r_1, r_2\} \). Then, symmetric cycles of \( P(v_1, v_3) \) are \( SC(v_1, v_3; v_2; r_2) \) and \( SC(v_1, v_3; v_2, r_2) \). Thus, \( RST(v_1, v_3) = SC(v_1, v_3; v_2, r_1) \cup SC(v_1, v_3; v_2, r_2) \cup RS(v_1, v_3) \). If we assume \( \hat{L}_q = \{P(v_1, v_3)\} \), since \( \alpha = 0.20 \), then any point in \( RST(v_1, v_3) \) covers 20% of \( f(v_1, v_3) \), and any point in \( RS(v_1, v_3) \) covers...
\( f(v_1, v_3) \). In addition, by Expression (5.4), \( CP_q = CRP(v_1, v_3; v_2) \cup CSV(v_1, v_3) \cup \)
\( EP(v_1, v_3) = \{r_1, r_2\} \cup \{v_2\} \cup \{v_1, v_3\} \).

5.3.5 Candidate Point Optimality Theorem

In this section, we prove that the set of candidate points, \( CP = \bigcup_{q=1}^{t} CP_q \), contains at least one
optimal location to the problem.

**Lemma 5-1.** Consider a point \( x \in \bar{T}_q \), for some \( q = 1, \ldots, t \), such that \( x \notin CP_q \). Then,
(a) If \( x \notin \bigcup_{P(v_i,v_j)\in\bar{L}_q} RST(v_i, v_j) \), then \( S(x) = S_D(x) = \emptyset \) and \( F(x) = 0 \).
(b) If \( x \in \bigcup_{P(v_i,v_j)\in\bar{L}_q} RST(v_i, v_j) \), then there exists a candidate point \( c \in CP_q \) such that \( S(x) \subseteq \)
\[ S(c), S_D(x) \subseteq S_D(c) \cup S(c), \text{ and } F(x) \leq F(c). \]

**Proof.** In case (a), by definition of refueling sub-tree, \( x \) covers no path in \( \bar{L}_q \). Thus, the result
holds. In case (b), \( x \) is in the interior of all the refueling sub-trees determined by paths in \( S(x) \)
and \( S_D(x) \). Thus, \( x \) has at least one candidate point on each side. Let \( c_1 \) and \( c_2 \) be the two closest
candidate points on each side of \( x \). Because no candidate point is located between \( x \) and \( c_i \), \( i = 1, 2 \),
the paths with positive traffic flow covered by \( x \) are the same paths with positive traffic flow
covered by both \( c_1 \) and \( c_2 \), i.e., \( S(x) = S(c_1) \cap S(c_2) \), and thus, \( S(x) \subseteq S(c_i), i = 1, 2 \). On the
other hand, any path with positive deviation-flow covered by \( x \) must either be a path with positive
deviation-flow covered by both \( c_1 \) and \( c_2 \), or be a path with positive traffic flow covered by one
of the two candidate points (\( c_1 \) and \( c_2 \)) and with positive deviation-flow covered by the other
candidate point, i.e., \( S_D(x) = \{S_D(c_1) \cap S_D(c_2)\} \cup \{S(c_1) \cap S_D(c_2)\} \cup \{S_D(c_1) \cap S(c_2)\} \). This
implies \( S_D(x) \subseteq S_D(c_i) \cup S(c_i), i = 1, 2 \). Thus, for \( 0 \leq \alpha \leq 1, \sum_{P(v_i,v_j)\in S(x)} f(v_i, v_j) + \)
Theorem 5-2 (Candidate Point Optimality Theorem). For the deviation-flow continuous location problem for a single refueling station on tree $T$, there exists at least one optimal location $x^* \in CP$.

Proof. (By contradiction) Suppose that none of the candidate points in $CP$ is an optimal solution. Let $x^*$ be an optimal point that covers $F(x^*)$ round trips per time unit. Since $x^*$ is an optimal location, by case (a) of Lemma 5-1, $x^*$ must belong to $\bigcup_{P(v_i,v_j) \in L_q} RST(v_i,v_j)$, for some $q = 1, ..., t$. If $x^* \in \bigcup_{P(v_i,v_j) \in L_q} RST(v_i,v_j)$, then by case (b) of Lemma 5-1, there exists a candidate point $c \in CP_q$ such that $F(x^*) \leq F(c)$. This contradicts the initial hypothesis that none of the candidate points in $CP$ is an optimal solution. Thus, there exists at least one candidate point in $CP$ that is an optimal location to the problem. □

5.4 Deviation-Flow Optimal Candidate Point Algorithm

Based on the Candidate Point Optimality Theorem, in this section the Deviation-flow Optimal Candidate Point Algorithm is proposed to find the optimal set of candidate points for a single refueling station that covers the maximum traffic flow (in round trips per time unit) when a given portion of drivers select the deviation option. This algorithm is composed of three steps. Step 1 constructs expanded sub-trees $T_q$, for $q = 1, ..., t$, from the original tree network $T$. Step 2 determines the set of local maximum candidate points for each $T_q$. Step 3 finds the set of global maximum candidate points of $T$ by comparing the local maximum candidate points.
A general structure of this algorithm is based on the Single Refueling Point Algorithm that finds the optimal locations for a single refueling station when no deviation is available to all drivers (Ventura et al. 2015). The Deviation-flow Optimal Candidate Point Algorithm is, however, distinct from the Single Refueling Point Algorithm in the way that it considers the deviation option and identifies different optimal locations as the value of deviation portion $\alpha$ changes. In particular, this algorithm works for $0 \leq \alpha \leq 1$ with all types of paths, regardless of their length and the amount of traffic flow, while the Single Refueling Point Algorithm limits the length and the amount of traffic flow of the paths to solve the problem. By changing the value of $\alpha$ between 0 and 1 in the algorithm, we can analyze the effect of the portion of drivers choosing the deviation option on the optimal set of locations. The steps of the algorithm are specified below.

**Deviation-flow Optimal Candidate Point Algorithm**

Step 1: Construct expanded sub-trees $T_{\hat{q}}(\hat{V}_q, \hat{E}_q)$, for $q = 1, \ldots, t$, from original tree network $T(V, E)$.

Sub-step 1.1: Establish set $\hat{E} = \{ (v_k, v_l) \in E \mid d(v_k, v_l) > R \}$ and eliminate $(v_k, v_l) \in \hat{E}$ from $E$.

Sub-step 1.2: Determine forest $F(V_F, E_F)$ that consists of sub-trees $T_q(V_q, E_q)$, for $q = 1, \ldots, t$, such that $V_F = \bigcup_{q=1}^{t} V_q = V$ and $E_F = \bigcup_{q=1}^{t} E_q = E \setminus \hat{E}$, where $t = \left| \hat{E} \right| + 1$.

Sub-step 1.3: For sub-tree $T_q(V_q, E_q)$, $q = 1, \ldots, t$, determine expanded sub-tree $T_{\hat{q}}(\hat{V}_q, \hat{E}_q)$.

Sub-step 1.3.1: Initialize $\hat{V}_q = \emptyset$ and generate $V_q' = V(E) \cap V_q$. 
Sub-step 1.3.2: For \( v_k \in V_q' \), if there exists at least one \( \alpha \times f(v_i, v_j) > 0 \), for \( P(v_i, v_j) \in L \) in \( T_q \), such that

\[
\max \{ d(v_i, v_k), d(v_j, v_k) \} < R/2,
\]
then add \( v_k \) into \( \bar{V}_q \).

Sub-step 1.3.3: Construct expanded sub-tree \( \hat{T}_q(V_q, E_q) \) such that \( \hat{V}_q = V_q \cup \{ \bigcup_{v_k \in \bar{V}_q} V(v_k) \} \) and \( \hat{E}_q = E_q \cup \{ \bigcup_{v_k \in \bar{V}_q} E(v_k) \} \), where

\[
E(v_k) = \{ (v_k, v_l) \in \hat{E} \mid v_l \in V \backslash V_q \}, V(v_k) = V(E(v_k)) \backslash \{ v_k \}, \text{ and } V(E(v_k)) \text{ is the set of vertices in } E(v_k).
\]

Step 2: For each expanded sub-tree \( \hat{T}_q \), \( q = 1, \ldots, t \), with \( \vert \bar{V}_q \vert > 1 \), identify the set of local optimal candidate points, denoted as \( CP_q^* \), in \( CP_q \) (note that, for \( \hat{T}_q \) with \( \vert \bar{V}_q \vert = 1 \), set \( CP_q^* = \{ x_q^* \} \) and \( F(x_q^*) = 0 \), where \( x_q^* \) is located at the only vertex in \( \hat{T}_q \)):

Sub-step 2.1: Construct \( \hat{L}_q = \{ \hat{P}(v_i, v_j) \mid d(v_i, v_j) \leq R, f(v_i, v_j) > 0, \text{ and } i < j, \text{ for all } v_i, v_j \in \hat{V}_q \} \).

Sub-step 2.2: Determine the set of candidate points, \( CP_q \), using Expression (5.4).

Sub-step 2.3: For all \( c \in CP_q \), construct \( S(c) = \{ \hat{P}(v_i, v_j) \in \hat{L}_q \mid c \in \hat{P}(v_i, v_j), f(v_i, v_j) > 0, d(v_i, c) \leq R/2, \text{ and } d(c, v_j) \leq R/2 \} \) and \( S_D(c) = \{ \hat{P}(v_i, v_j) \in \hat{L}_q \mid c \in T_q \backslash \hat{P}(v_i, v_j), \alpha \times f(v_i, v_j) > 0, d(v_i, c) \leq R/2, \text{ and } d(c, v_j) \leq R/2 \} \).

Sub-step 2.4: Calculate \( F(c) = \sum_{\hat{P}(v_i, v_j) \in S(c)} f(v_i, v_j) + \sum_{\hat{P}(v_i, v_j) \in S_D(c)} \alpha \times f(v_i, v_j) \), for all \( c \in CP_q \), then determine the local optimal set of candidate points, \( CP_q^* \), in \( CP_q \):

\[
CP_q^* = \arg \max \{ F(c) \mid c \in CP_q \}.
\]
Step 3: Find the global optimal set of candidate points, denoted as $CP^*$, for the original tree $T$ and the corresponding maximum total traffic flow in round trips per time unit ($F^*$): 

$$CP^* = \arg \max \{ F(x_q^*) \mid x_q^* \in CP^*_q, \ q = 1, \ldots, t \} ,$$

and $F^* = F(x^*)$, for $x^* \in CP^*$.

**Theorem 5.3 (Complexity of the Deviation-flow Optimal Candidate Point Algorithm).** The Deviation-flow Optimal Candidate Point Algorithm runs in $O(n^2)$ time, where $n = |V|$. 

**Proof.** Step 1 of the algorithm contains three sub-steps to generate expanded sub-trees $\hat{T}_q(V_q, E_q)$, for $q = 1, \ldots, t$, from original tree $T(V, E)$. Sub-step 1.1 takes $O(n)$ to find and remove edges $(v_k, v_l) \in \hat{E}$ from $E$. To construct forest $F(V_F, E_F)$ containing sub-trees $T_q(V_q, E_q)$, for $q = 1, \ldots, t$, in Sub-step 1.2, we start with a forest where each vertex forms its own tree. Then, for each edge $(v_i, v_k) \in E \setminus \hat{E}$, two trees connected by an edge $(v_i, v_k)$ are merged into a single tree. We repeat this iteration for all edges in $E \setminus \hat{E}$, thus we obtain the final set of sub-trees in $O(n)$ time. To build expanded sub-tree $\hat{T}_q$ from $T_q, q = 1, \ldots, t$, Sub-step 1.3 takes $O(n^3)$ to identify $V_q'$ and compute $\max \{ d(v_i, v_k), d(v_j, v_k) \}$ for $v_k \in V_q'$ and $P(v_i, v_j) \in L$ in $T_q$. Step 2 consists of four sub-steps to find the set of local optimal candidate points $CP^*_q$ for each $\hat{T}_q$. Sub-step 2.1 takes $O(n^2)$ to measure $d(v_i, v_j)$, for all $v_i, v_j \in V_q$, and build set $\hat{L}_q, q = 1, \ldots, t$. To determine $CP_q, q = 1, \ldots, t$, Sub-step 2.2 first builds $RS(v_i, v_j)$ and $EP(v_i, v_j)$ for $P(v_i, v_j) \in \hat{L}_q$, which takes $O(n^2)$. Next, for each $RS(v_i, v_j)$, Sub-step 2.2 identifies $CSV(v_i, v_j)$, which takes $O(n)$. Lastly, for each $v_s \in CSV(v_i, v_j)$, Sub-step 2.2 finds $CRP(v_i, v_j; v_s)$, which takes $O(n)$ by Theorem 5.1. Thus, the total time devoted to Sub-step 2.2 is $O(n^4)$. Since we can find $O(n)$ candidate points for each path $P(v_i, v_j) \in \hat{L}_q$, and furthermore, there exists $O(n^2)$ paths in each $\hat{L}_q$, the number of candidate points in $CP_q$ is $O(n^3)$. Therefore, Sub-step 2.3 takes $O(n^5)$ to
generate the sets of paths with positive traffic flow and deviation-flow covered by all candidate points \( c \in CP_q, q = 1, ..., t \). Similarly, Sub-step 2.4 takes \( O(n^5) \) to compute the total traffic flow covered by all candidate points, compare the total traffic flows covered by candidate points in \( \hat{T}_q \), \( q = 1, ..., t \), and determine the local optimal set of candidate points for each \( \hat{T}_q \). Step 3 takes \( O(n^2) \) to compare the local optimal solutions and determine the global optimal set of candidate points in \( T \). The algorithm performs each step once. Thus, the total time devoted to operations in the algorithm is \( O(n^5) \). 

\[
\square
\]

5.5 Convex Combination Property

This section consists of two sections to identify additional optimal points that can be generated as convex combinations of certain pairs of optimal candidate points. First, a condition is identified in Section 5.5.1 under which any point on the line segment between two consecutive optimal candidate points is also optimal. Then, based on this condition, the complete set of optimal solutions for the problem is determined in Section 5.5.2.

5.5.1 Optimality Conditions for a Convex Combination of Consecutive Optimal Candidate Points

In this section, given any pair of consecutive optimal candidate points, \( c_1^*, c_2^* \in CP^* \), we identify conditions under which any point that can be written as a convex combination of \( c_1^* \) and \( c_2^* \) is also optimal.

To simplify the proof process to derive the conditions, we first partition \( S(c_1^*) \), \( S_D(c_1^*) \), \( S(c_2^*) \), and \( S_D(c_2^*) \) into newly defined subsets as shown below. Since there exists no candidate point between \( c_1^* \) and \( c_2^* \), \( S(c_1^*) \) is partitioned into \( S_1(c_1^*) \) and \( S_{1,2}(c_1^*) \), where \( S_1(c_1^*) \) denotes the
subset of paths with positive traffic flow covered by $c_1^*$ but not covered by $c_2^*$, and $S_{1,2}(c_1^*)$ denotes the subset of paths with positive deviation-flow covered by $c_1^*$ but not covered by $c_2^*$, and $S_{D,1,2}(c_1^*)$ denotes the subset of paths with positive deviation-flow covered by both $c_1^*$ and $c_2^*$. We also partition $S_D(c_1^*)$ into $S_{D_1}(c_1^*)$ and $S_{D_{1,2}}(c_1^*)$, where $S_{D_1}(c_1^*)$ denotes the subset of paths with positive deviation-flow covered by $c_1^*$ but not covered by $c_2^*$, and $S_{D_{1,2}}(c_1^*)$ denotes the subset of paths with positive deviation-flow covered by both $c_1^*$ and $c_2^*$. $S_{D_1}(c_1^*)$ can also be partitioned into $S_{D_{1,2}}(c_1^*)$ and $S_{D_1}(c_1^*) \setminus S_{D_{1,2}}(c_1^*)$, where $S_{D_{1,2}}(c_1^*)$ denotes the subset of paths with positive deviation-flow covered by $c_1^*$ and positive traffic flow not covered by $c_2^*$, and $S_{D_1}(c_1^*) \setminus S_{D_{1,2}}(c_1^*)$ denotes the subset of paths with positive traffic flow covered by $c_2^*$ and positive deviation-flow covered by $c_1^*$. Similarly, we can partition $S(c_2^*)$ and $S_D(c_2^*)$, such that $S(c_2^*) = S_2(c_2^* \cup S_{1,2}(c_2^*)$, $S_D(c_2^*) = S_{D_2}(c_2^*) \cup S_{D_{1,2}}(c_2^*)$, and $S_{D_2}(c_2^*) = S_{D_{2,1}}(c_2^*) \cup \{S_{D_2}(c_2^*) \setminus S_{D_{2,1}}(c_2^*)\}$. Then, by using these terms, $F(c_1^*)$ and $F(c_2^*)$ can be written as follows:

$$F(c_1^*) = \sum_{P(v_i,v_j) \in S_1(c_1^*)} f(v_i,v_j) + \sum_{P(v_i,v_j) \in S_{1,2}(c_1^*)} f(v_i,v_j) + \alpha \left( \sum_{P(v_i,v_j) \in S_{D_{1,2}}(c_1^*)} f(v_i,v_j) + \sum_{P(v_i,v_j) \in S_{D_1}(c_1^*) \setminus S_{D_{1,2}}(c_1^*)} f(v_i,v_j) + \sum_{P(v_i,v_j) \in S_{D_1}(c_1^*) \setminus S_{D_{1,2}}(c_1^*)} f(v_i,v_j) \right).$$

$$F(c_2^*) = \sum_{P(v_i,v_j) \in S_2(c_2^*)} f(v_i,v_j) + \sum_{P(v_i,v_j) \in S_{1,2}(c_2^*)} f(v_i,v_j) + \alpha \left( \sum_{P(v_i,v_j) \in S_{D_{2,1}}(c_2^*)} f(v_i,v_j) + \sum_{P(v_i,v_j) \in S_{D_2}(c_2^*) \setminus S_{D_{2,1}}(c_2^*)} f(v_i,v_j) + \sum_{P(v_i,v_j) \in S_{D_2}(c_2^*) \setminus S_{D_{2,1}}(c_2^*)} f(v_i,v_j) \right).$$

Note also that, by definition, $S_{1,2}(c_1^*) = S_{1,2}(c_2^*)$ and $S_{D_{1,2}}(c_1^*) = S_{D_{1,2}}(c_2^*)$. Similarly, $\{S_{D_2}(c_2^*) \setminus S_{D_{2,1}}(c_2^*)\} \subseteq S_1(c_1^*)$ and $\{S_{D_1}(c_1^*) \setminus S_{D_{1,2}}(c_1^*)\} \subseteq S_2(c_2^*)$.

The following theorem establishes optimality conditions for points that can be written as convex combinations of consecutive optimal candidate points.

**Theorem 5-4. (Optimality Conditions for the Line Segment Joining Consecutive Optimal Candidate Points)** Given a pair of consecutive optimal candidate points, $c_1^*, c_2^* \in CP^*$, let $\ell(c_1^*, c_2^*)$ denote the line segment joining $c_1^*$ and $c_2^*$; that is, $\ell(c_1^*, c_2^*) = \{x \mid x = \beta c_1^* + \gamma c_2^* \text{ for } \beta, \gamma \geq 0, \beta + \gamma = 1\}$.
\((1 - \beta) c_2^*, \ 0 \leq \beta \leq 1\). Then, any point \(x \in \ell(c_1^*, c_2^*)\) is also optimal if and only if one of the following conditions is satisfied:

(a) \(S(c_1^*) = S(c_2^*)\), for \(\alpha = 0\).

(b) \(S(c_1^*) = S(c_2^*)\) and \(S_D(c_1^*) = S_D(c_2^*)\), for \(0 < \alpha < 1\).

(c) \(S(c_1^*) \cup S_D(c_1^*) = S(c_2^*) \cup S_D(c_2^*)\), for \(\alpha = 1\).

**Proof.** (\(\Rightarrow\)) Assume that any point \(x \in \ell(c_1^*, c_2^*)\) is also optimal. This implies \(F(c_1^*) = F(x) = F(c_2^*)\), or equivalently, \(F(c_1^*) - F(x) = F(c_2^*) - F(x) = 0\). Since \(c_1^*\) and \(c_2^*\) are the two closest candidate points on each side of \(x\), \(S(x) = S_{1,2}(c_1^*)\) and \(S_D(x) = S_{D,1,2}(c_1^*) \cup \{S_D(c_1^*) \setminus S_{D,1,2}(c_2^*)\}\). Thus, \(F(x)\) can be written as follows:

\[
F(x) = \sum_{p(v_i,v_j) \in S_{1,2}(c_1^*)} f(v_i,v_j) + \alpha \left( \sum_{p(v_i,v_j) \in S_{D,1,2}(c_1^*)} f(v_i,v_j) + \sum_{p(v_i,v_j) \in S_D(c_2^*) \setminus S_{D,1,2}(c_2^*)} f(v_i,v_j) \right).
\]

We recall that \(S_{1,2}(c_1^*) = S_{1,2}(c_2^*)\) and \(S_{D,1,2}(c_1^*) = S_{D,1,2}(c_2^*)\). Thus, \(F(c_1^*) - F(x)\) and \(F(c_2^*) - F(x)\) can be expressed as follows:

\[
F(c_1^*) - F(x) = \sum_{p(v_i,v_j) \in S_1(c_1^*)} f(v_i,v_j) + \sum_{p(v_i,v_j) \in S_{D,1}(c_1^*)} f(v_i,v_j) - \sum_{p(v_i,v_j) \in S_{D,2}(c_2^*) \setminus S_{D,1,2}(c_2^*)} f(v_i,v_j),
\]

\[
F(c_2^*) - F(x) = \sum_{p(v_i,v_j) \in S_2(c_2^*)} f(v_i,v_j) + \sum_{p(v_i,v_j) \in S_{D,2,1}(c_2^*)} f(v_i,v_j) - \sum_{p(v_i,v_j) \in S_D(c_1^*) \setminus S_{D,1,2}(c_1^*)} f(v_i,v_j).
\]

In case (a), since \(\alpha = 0\), \(S_D(c_1^*) = S_D(c_2^*) = \emptyset\). Then, the equalities \(F(c_1^*) - F(x) = F(c_2^*) - F(x) = 0\) can only be satisfied if \(S_1(c_1^*) = S_2(c_2^*) = \emptyset\). This means that \(S(c_1^*) = S_{1,2}(c_1^*) = S(c_2^*)\).

In case (b), \(0 < \alpha < 1\). Recall that \(\{S_{D,2}(c_2^*) \setminus S_{D,2,1}(c_2^*)\} \subseteq S_1(c_1^*)\) and \(\{S_D(c_1^*) \setminus S_{D,1,2}(c_1^*)\} \subseteq S_2(c_2^*)\). This means that \(\Sigma_{p(v_i,v_j) \in S_1(c_1^*)} f(v_i,v_j) - \Sigma_{p(v_i,v_j) \in S_{D,2}(c_2^*) \setminus S_{D,2,1}(c_2^*)} f(v_i,v_j) = \ldots\)
\[ f(v_i, v_j) > 0 \text{ in } F(c_1) - F(x) \text{ and } \sum_{(v_i, v_j) \in S_2(c_2)} f(v_i, v_j) - \sum_{(v_i, v_j) \in [S_D(c_1) \setminus S_{D_2}(c_2)]} \alpha \times \]

\[ f(v_i, v_j) > 0 \text{ in } F(c_2) - F(x) \text{ if } S_1(c_1) \text{ and } S_2(c_2) \text{ are not empty. Thus, to satisfy equalities } \]

\[ F(c_1) - F(x) = F(c_2) - F(x) = 0, \] it is necessary to have \( S_1(c_1) = S_2(c_2) = S_{D_2}(c_1) = S_{D_2}(c_2) = \emptyset. \) Note also that, if \( S_1(c_1) = S_2(c_2) = \emptyset, \) then \( \{S_D(c_2) \setminus S_{D_2}(c_2)\} = \{S_D(c_2) \setminus S_{D_2}(c_2)\} \subseteq S_1(c_1) \text{ and } \{S_D(c_2) \setminus S_{D_2}(c_2)\} \subseteq S_2(c_2). \) Therefore, \( S(c_1) = S_{1,2}(c_1) = S(c_2) \) and \( S_D(c_1) = S_D(c_2). \)

In case (c), since \( \alpha = 1 \) and \( \{S_D(c_2) \setminus S_{D_2}(c_2)\} \subseteq S_1(c_1), \) to satisfy \( F(c_1) - F(x) = 0, \) it is necessary to have \( S_{D_2}(c_1) = \emptyset \) and \( \{S_D(c_2) \setminus S_{D_2}(c_2)\} = S_1(c_1). \) Similarly, \( F(c_2) - F(x) = 0 \) can only be satisfied if \( S_{D_2}(c_2) = \emptyset \) and \( \{S_D(c_2) \setminus S_{D_2}(c_2)\} = S_2(c_2). \) These results also imply that \( S(c_1) \cup S_D(c_1) = S(c_2) \cup S_D(c_2) \text{ and } S(c_2) \cup S_D(c_2) = S_2(c_2) \cup S_{1,2}(c_2) \cup S_{D_1}(c_1) \cup S_{D_2}(c_1). \) Since \( S_1(c_1) = S_{1,2}(c_1) \) and \( S_{D_1}(c_1) = S_{D_2}(c_1), \) we can finally conclude that \( S(c_1) \cup S_D(c_1) = S(c_2) \cup S_D(c_2). \)

(\( \Leftarrow \)) In case (a), we assume \( S(c_1) = S(c_2). \) This implies that \( S_1(c_1) = S_2(c_2) = \emptyset. \)

Also, since \( \alpha = 0, S_D(c_1) = S_D(c_2) = \emptyset. \) Thus, \( F(c_1) - F(x) = F(c_2) - F(x) = 0. \)

In case (b), we assume that \( S(c_1) = S(c_2) \text{ and } S_D(c_1) = S_D(c_2). \) This implies that \( S_1(c_1) = S_2(c_2) = \emptyset \) and \( S_{D_1}(c_1) = S_{D_2}(c_2) = \emptyset. \) Note also that \( \{S_D(c_2) \setminus S_{D_2}(c_2)\} = \{S_D(c_2) \setminus S_{D_2}(c_2)\} \subseteq S_1(c_1), \{S_D(c_2) \setminus S_{D_2}(c_2)\} \subseteq S_2(c_2), \text{ and } S_1(c_1) = S_2(c_2) = \emptyset. \) Thus, \( F(c_1) - F(x) = F(c_2) - F(x) = 0. \)

In case (c), we assume that \( S(c_1) \cup S_D(c_1) = S(c_2) \cup S_D(c_2). \) This implies that \( S(c_1) \cup S_{D_1}(c_1) \cup S_{D_2}(c_1) \cup \{S_D(c_1) \setminus S_{D_2}(c_1)\} = S_2(c_2) \cup S_{D_1}(c_2) \cup S_{D_2}(c_2) \cup \{S_D(c_2) \setminus S_{D_2}(c_2)\}. \) Since \( \{S_D(c_2) \setminus S_{D_2}(c_2)\} \subseteq S_1(c_1) \text{ and } \{S_D(c_1) \setminus S_{D_2}(c_1)\} \subseteq S_2(c_2), \) to satisfy \( S_1(c_1) \cup S_{D_1}(c_1) \cup S_{D_2}(c_1) \cup \{S_D(c_1) \setminus S_{D_2}(c_1)\} = S_2(c_2) \cup S_{D_1}(c_2) \cup S_{D_2}(c_2) \cup \{S_D(c_2) \setminus S_{D_2}(c_2)\}, \) it is necessary to
have $S_{D \backslash 2}(c_1^*) = S_{D \backslash 1}(c_2^*) = \emptyset$, $S_1(c_1^*) = \{S_{D_2}(c_2) \backslash S_{D_2}(c_2^*)\}$, and $S_2(c_2^*) = \{S_{D_1}(c_1^*) \backslash S_{D_1}(c_1^*)\}$. As a result, $F(c_1^*) - F(x) = F(c_2^*) - F(x) = 0$.

Since the conditions in cases (a), (b), and (c) lead to $F(c_1^*) = F(x) = F(c_2^*)$, and furthermore $c_1^*, c_2^* \in CP^*$, any point $x \in \ell(c_1^*, c_2^*)$ is optimal. □

5.5.2 Complete Set of Optimal Solutions

In this section, we determine the complete set of optimal solutions, denoted as $CS^*$, for the deviation-flow continuous location problem for a single refueling station.

As an intermediate step toward determining $CS^*$, we first construct the set pairs of consecutive candidate points in $CP^*$ that satisfy the conditions stated in Theorem 5-4, denoted as set $\overline{CP}^*$. Then, by Theorem 5-4, the line segments defined by these pairs of candidate points in $\overline{CP}^*$ are also optimal. Consequently, the complete set of optimal solutions, denoted as $CS^*$, can be defined as follows:

$$CS^* = \left\{ \bigcup_{(c_1^*, c_2^*) \in CP^*} \ell(c_1^*, c_2^*) \right\} \cup \{CP^* \backslash V(\overline{CP}^*)\},$$

(5.5)

where $V(\overline{CP}^*)$ is the set of optimal candidate points in $\overline{CP}^*$. The optimality of set $CS^*$ is proved in the following theorem.

**Theorem 5-5 (Optimality of the Complete Set of Optimal Solutions).** Set $CS^*$, defined in Expression (5.5), contains all optimal solutions for the problem.

*Proof.* (By contradiction) Suppose that there exists an optimal point $x^*$ that does not belong to $CS^*$. Because $CS^*$ includes all optimal candidate points in $CP^*$ and $x^* \notin CS^*$, $x^*$ should be a point in $T$ that is not a candidate point. First, suppose that $x^* \notin \bigcup_{(v_i, v_j) \in \mathcal{E}_q} RST(v_i, v_j)$. Then, by
case (a) of Lemma 5-1, $F(x^*) = 0$. This contradicts the hypothesis that $x^*$ is optimal. Thus, $x^*$ must belong to $\bigcup_{P(v_i,v_j)\in I_q} RST(v_i,v_j)$. Since $x^* \in \bigcup_{P(v_i,v_j)\in I_q} RST(v_i,v_j)$, by case (b) of Lemma 5-1, there exist two closest candidate points, $c_1$ and $c_2$, on each side of $x^*$, such that $S(x^*) \subseteq S(c_1), S_D(x^*) \subseteq S_D(c_i) \cup S(c_i), \text{ and } F(x^*) \leq F(c_i), i = 1, 2$. Because $x^*$ is optimal, $c_1$ and $c_2$ are also optimal, and therefore, Theorem 5-4 holds for the convex combination of $c_1$ and $c_2$. This means that $x^* \in \ell(c_1,c_2) \in CS^*$. This contradicts the initial hypothesis that there exists an optimal point $x^*$ that does not belong to $CS^*$. Thus, $CS^*$ contains all the optimal solutions. □

5.6 Numerical Example and Analysis of Results

In this section, a numerical example is provided to illustrate the proposed model and verify its performance. Figure 5-2 shows a numerical example of an undirected tree $T(V,E)$ with $n = 11$, and Table 5-3 provides the average traffic flows $f(v_i,v_j)$, for all $v_i, v_j \in V$ such that $i < j$. We note that the information about the actual portion of drivers who select the deviation option per time unit is difficult to obtain or likely to be inaccurately estimated. Thus, in this example we change the value of the deviation portion $\alpha$ from 0 to 1 and analyze the effect of the deviation portion on the set of optimal locations.

The remainder of this section is organized as follows. In Section 5.6.1, for a given value of the vehicle driving range $R$, we apply the Deviation-flow Optimal Candidate Point Algorithm to solve the numerical example. In Section 5.6.2, we further examine the coupled effects of deviation portion and vehicle driving range on the set of optimal locations and the maximum traffic flow covered. Finally, in Section 5.6.3, we compare the results of the proposed model with the existing models to verify its performance.
Table 5-3. Round trip traffic flows corresponding to all simple paths in tree $T$

<table>
<thead>
<tr>
<th></th>
<th>$v_1$</th>
<th>$v_2$</th>
<th>$v_3$</th>
<th>$v_4$</th>
<th>$v_5$</th>
<th>$v_6$</th>
<th>$v_7$</th>
<th>$v_8$</th>
<th>$v_9$</th>
<th>$v_{10}$</th>
<th>$v_{11}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$v_1$</td>
<td>-</td>
<td>80</td>
<td>40</td>
<td>30</td>
<td>20</td>
<td>0</td>
<td>10</td>
<td>8</td>
<td>5</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>$v_2$</td>
<td>-</td>
<td>-</td>
<td>70</td>
<td>60</td>
<td>30</td>
<td>15</td>
<td>20</td>
<td>17</td>
<td>14</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>$v_3$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>90</td>
<td>60</td>
<td>50</td>
<td>10</td>
<td>9</td>
<td>6</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>$v_4$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>100</td>
<td>30</td>
<td>155</td>
<td>100</td>
<td>80</td>
<td>50</td>
<td>15</td>
</tr>
<tr>
<td>$v_5$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>10</td>
<td>20</td>
<td>16</td>
<td>14</td>
<td>15</td>
<td>12</td>
</tr>
<tr>
<td>$v_6$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>6</td>
<td>5</td>
<td>1</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>$v_7$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>90</td>
<td>85</td>
<td>82</td>
<td>80</td>
</tr>
<tr>
<td>$v_8$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>140</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$v_9$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$v_{10}$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>80</td>
</tr>
<tr>
<td>$v_{11}$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Figure 5-2. Structure of undirected tree network $T$
5.6.1 Numerical Example for a Given Vehicle Driving Range

In this section, we fix the driving range \( R = 80 \) and solve the numerical example to illustrate the Deviation-flow Optimal Candidate Point Algorithm in detail.

To construct the expanded sub-trees from \( T \) in Step 1, Sub-step 1.1 finds \( d(v_7, v_8) > R \); thus, \( \hat{E} = \{(v_7, v_8)\} \). By removing \((v_7, v_8)\) from \( T \), Sub-step 1.2 forms forest \( F(V_F, E_F) \) containing two sub-trees \( T_1(V_1, E_1) \) and \( T_2(V_2, E_2) \), where

\[
V_1 = \{v_1, v_2, v_3, v_4, v_5, v_6, v_7\},
E_1 = \{(v_1, v_2), (v_2, v_3), (v_3, v_4), (v_4, v_5), (v_3, v_6), (v_4, v_7)\},
V_2 = \{v_8, v_9, v_{10}, v_{11}\}, \text{ and } E_2 = \{(v_8, v_9), (v_9, v_{10}), (v_{10}, v_{11})\}.
\]

For \( T_1 \), Sub-step 1.3 sets \( \tilde{V}_1 = \emptyset \) and generates \( V_1' = V(\hat{E}) \cap V_1 = \{v_7\} \). For \( v_k \in V_1' \), even if \( \alpha > 0 \), there is no \( \alpha \times f(v_i, v_j) > 0 \), for \( v_i, v_j \in V_1 \), such that \( \max\{d(v_i, v_k), d(v_j, v_k)\} < R/2 \). Thus, \( V_1 \) remains empty, and expanded sub-tree \( \hat{T}_1(V_1, E_1) = T_1(V_1, E_1) \).

For \( T_2 \), Sub-step 1.3 sets \( \tilde{V}_2 = \emptyset \) and forms \( V_2' = V(\hat{E}) \cap V_2 = \{v_8\} \). For \( v_k \in V_2' \), if \( \alpha > 0 \), there exists \( \alpha \times f(v_8, v_9) > 0 \), such that \( \max\{d(v_8, v_k), d(v_9, v_k)\} = 15 < R/2 \). Thus, \( v_8 \) is added to \( \tilde{V}_2 \), and expanded sub-tree \( \hat{T}_2(\tilde{V}_2, \hat{E}_2) \) is such that \( \tilde{V}_2 = V_2 \cup V(v_8) \) and \( \hat{E}_2 = E_2 \cup E(v_8) \), where \( E(v_8) = \{(v_7, v_8)\} \) and \( V(v_8) = V(E(v_8)) \setminus \{v_8\} = \{v_7\} \).

Now, for expanded sub-tree \( \hat{T}_q, q = 1, 2 \), Step 2 determines the set of local optimal candidate points. For \( \hat{T}_1 \), Sub-step 2.1 builds set \( \hat{L}_1 \) that consists of \( P(v_i, v_j) \) in \( \hat{T}_1 \) such that \( d(v_i, v_j) \leq 80 \) and \( f(v_i, v_j) > 0 \); that is, \( \hat{L}_1 = \{P(v_1, v_2), P(v_1, v_3), P(v_2, v_3), P(v_3, v_4), P(v_3, v_6), P(v_4, v_5), P(v_4, v_7)\} \). By using Expression (5.4), Sub-step 2.2 establishes \( CP_1 \), which refers to the set of candidate points of the refueling sub-trees corresponding to the paths in \( \hat{L}_1 \). For \( P(v_2, v_3) \), since \( CSV(v_2, v_3) = \{v_2, v_3\}, \alpha \times 100\% \) of traffic flow in \( P(v_2, v_3) \) can select the deviation option that starts from \( v_2 \) or \( v_3 \). By applying the Cycle Returning Point Algorithm to \( v_2 \) and \( v_3 \), three cycle returning points, \( r_1, r_2 \), and \( r_3 \), are obtained such that \( r_1 \in CRP(v_2, v_3; v_2) \) and \( r_2, r_3 \in CRP(v_2, v_3; v_3) \). Thus, \( RST(v_2, v_3) \), consisting of \( RS(v_2, v_3), SC(v_2, v_3; v_2, r_1), \)
SC(v₂, v₃; v₂, r₂), and SC(v₂, v₃; v₃, r₃), has seven candidate points cₖ²₃, for k = 1, ..., 7, such that c₁²₃, c₂²₃ ∈ EP(v₂, v₃), c³²₃, c⁴²₃ ∈ CSV(v₂, v₃), c⁵²₃ ∈ CRP(v₂, v₃; v₂), and c₆²₃, c₇²₃ ∈ CRP(v₂, v₃; v₃). Note that there are no cycle starting vertices in any path P(vᵢ, vⱼ) ∈ ̂L₁\{P(v₂, v₃)}; that is, the deviation option is not available. Thus, RST(vᵢ, vⱼ) = RS(vᵢ, vⱼ), and each RST(vᵢ, vⱼ) has one or two candidate points depending on the value of d(vᵢ, vⱼ). Then, CP₁ includes eighteen candidate points, i.e., CP₁ = \{c₁¹₂, c₁²₂, c₁³₂, c₂²₃, c₂³₂, c₂⁴₃, c₂⁵₂, c₃²₃, c₃³₄, c₃⁴₅, c₄₃₅, c₄⁵₇, c₅₆₇\}. The left side of Figure 5-3 illustrates the locations of RS(vᵢ, vⱼ), SC(vᵢ, vⱼ; vₛ, rₛ), and the corresponding c ∈ CP₁ in ̂T₁, for P(vᵢ, vⱼ) ∈ ̂L₁, vₛ ∈ CSV(vᵢ, vⱼ), r ∈ CRP(vᵢ, vⱼ; vₛ). For c ∈ CP₁, Sub-step 2.3 constructs S(c) and S₀(c) to identify all positive traffic flows covered by c. The left and middle sides of Table 5-4 summarize the results. Based on the paths covered by each candidate point c ∈ CP₁, Sub-step 2.4 first calculates F(c), for c ∈ CP₁. The right side of Table 5-4 shows the calculation processes. Sub-step 2.4 next compares F(c), for c ∈ CP₁, and determines the local optimal set of candidate points CP₁⁺ in ̂T₁, for 0 ≤ α ≤ 1, as follows:

\[ CP₁⁺ = \text{arg max} \left\{ F(c_{1,2}), F(c_{1,2}^2), F(c_{1,3}^2), F(c_{2,3}^1), F(c_{2,3}^2), F(c_{2,3}^3), F(c_{2,3}^4), F(c_{2,3}^5), F(c_{2,3}^6), F(c_{2,3}^7), F(c_{3,4}^1), F(c_{3,4}^2), F(c_{3,4}^3), F(c_{3,4}^4), F(c_{3,4}^5), F(c_{3,4}^6), F(c_{3,4}^7), F(c_{4,5}^1), F(c_{4,5}^2), F(c_{4,5}^3), F(c_{4,5}^4), F(c_{4,5}^5), F(c_{4,5}^6), F(c_{4,5}^7), F(c_{5,6}^1), F(c_{5,6}^2), F(c_{5,6}^3), F(c_{5,6}^4), F(c_{5,6}^5), F(c_{5,6}^6), F(c_{5,6}^7), F(c_{6,7}^1), F(c_{6,7}^2), F(c_{6,7}^3), F(c_{6,7}^4), F(c_{6,7}^5), F(c_{6,7}^6), F(c_{6,7}^7) \right\} \]

\[ = \text{arg max} \left\{ 80, 120 + 70α, 120 + 70α, 70, 70, 70, 70, 70, 70, 120 + 70α, 70α \right\} \]

\[ = \{c_{1,2}^1, c_{1,3}^1, c_{2,3}^2, c_{4,7}^5, c_{4,7}^7\}. \]

This result implies that, if α ≥ 0.5, then c_{1,2}^1, c_{1,3}^1, and c_{2,3}^2 are the local optimal locations to ̂T₁; otherwise, c_{4,7}^5 and c_{4,7}^7 are the local optimal locations that cover the maximum traffic flow (in round trips per time unit) in ̂T₁.
Table 5-4. Sets of paths and the traffic flows (in round trips per time unit) covered by candidate points in the first sub-tree

<table>
<thead>
<tr>
<th>c</th>
<th>S(c)</th>
<th>S_D(c)</th>
<th>F(c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(c_{1,2}^1)</td>
<td>(S(c_{1,2}^1) = {P(v_1, v_2)})</td>
<td>(S_D(c_{1,2}^1) = \emptyset)</td>
<td>(F(c_{1,2}^1) = f(v_1, v_2) = 80)</td>
</tr>
<tr>
<td>(c_{1,2}^2)</td>
<td>(S(c_{1,2}^2) = {P(v_1, v_2), P(v_1, v_3)})</td>
<td>(S_D(c_{1,2}^2) = {P(v_2, v_3)})</td>
<td>(F(c_{1,2}^2) = f(v_1, v_2) + f(v_1, v_3) + \alpha f(v_2, v_3) = 120 + 70\alpha)</td>
</tr>
<tr>
<td>(c_{1,3}^1)</td>
<td>(S(c_{1,3}^1) = {P(v_1, v_2), P(v_1, v_3)})</td>
<td>(S_D(c_{1,3}^1) = {P(v_2, v_3)})</td>
<td>(F(c_{1,3}^1) = f(v_1, v_2) + f(v_1, v_3) + \alpha f(v_2, v_3) = 120 + 70\alpha)</td>
</tr>
<tr>
<td>(c_{2,3}^1)</td>
<td>(S(c_{2,3}^1) = {P(v_2, v_3)})</td>
<td>(S_D(c_{2,3}^1) = \emptyset)</td>
<td>(F(c_{2,3}^1) = f(v_2, v_3) = 70)</td>
</tr>
<tr>
<td>(c_{2,3}^2)</td>
<td>(S(c_{2,3}^2) = {P(v_2, v_3)})</td>
<td>(S_D(c_{2,3}^2) = \emptyset)</td>
<td>(F(c_{2,3}^2) = f(v_2, v_3) = 70)</td>
</tr>
<tr>
<td>(c_{2,3}^3)</td>
<td>(S(c_{2,3}^3) = {P(v_2, v_3)})</td>
<td>(S_D(c_{2,3}^3) = \emptyset)</td>
<td>(F(c_{2,3}^3) = f(v_2, v_3) = 70)</td>
</tr>
<tr>
<td>(c_{2,3}^4)</td>
<td>(S(c_{2,3}^4) = {P(v_2, v_3)})</td>
<td>(S_D(c_{2,3}^4) = \emptyset)</td>
<td>(F(c_{2,3}^4) = f(v_2, v_3) = 70)</td>
</tr>
<tr>
<td>(c_{2,3}^5)</td>
<td>(S(c_{2,3}^5) = {P(v_1, v_2), P(v_1, v_3)})</td>
<td>(S_D(c_{2,3}^5) = {P(v_2, v_3)})</td>
<td>(F(c_{2,3}^5) = f(v_1, v_2) + f(v_1, v_3) + \alpha f(v_2, v_3) = 120 + 70\alpha)</td>
</tr>
<tr>
<td>(c_{2,3}^6)</td>
<td>(S(c_{2,3}^6) = \emptyset)</td>
<td>(S_D(c_{2,3}^6) = {P(v_2, v_3)})</td>
<td>(F(c_{2,3}^6) = \alpha \times f(v_2, v_3) = 70\alpha)</td>
</tr>
<tr>
<td>(c_{2,3}^7)</td>
<td>(S(c_{2,3}^7) = \emptyset)</td>
<td>(S_D(c_{2,3}^7) = {P(v_2, v_3)})</td>
<td>(F(c_{2,3}^7) = \alpha \times f(v_2, v_3) = 70\alpha)</td>
</tr>
<tr>
<td>(c_{3,3}^1)</td>
<td>(S(c_{3,3}^1) = {P(v_3, v_4)})</td>
<td>(S_D(c_{3,3}^1) = \emptyset)</td>
<td>(F(c_{3,3}^1) = f(v_3, v_4) = 90)</td>
</tr>
<tr>
<td>(c_{3,3}^2)</td>
<td>(S(c_{3,3}^2) = {P(v_3, v_4)})</td>
<td>(S_D(c_{3,3}^2) = \emptyset)</td>
<td>(F(c_{3,3}^2) = f(v_3, v_4) = 90)</td>
</tr>
<tr>
<td>(c_{3,3}^3)</td>
<td>(S(c_{3,3}^3) = {P(v_3, v_4)})</td>
<td>(S_D(c_{3,3}^3) = \emptyset)</td>
<td>(F(c_{3,3}^3) = f(v_3, v_4) = 90)</td>
</tr>
<tr>
<td>(c_{3,3}^4)</td>
<td>(S(c_{3,3}^4) = {P(v_3, v_4)})</td>
<td>(S_D(c_{3,3}^4) = \emptyset)</td>
<td>(F(c_{3,3}^4) = f(v_3, v_4) = 90)</td>
</tr>
<tr>
<td>(c_{3,3}^5)</td>
<td>(S(c_{3,3}^5) = {P(v_3, v_6)})</td>
<td>(S_D(c_{3,3}^5) = \emptyset)</td>
<td>(F(c_{3,3}^5) = f(v_3, v_6) = 50)</td>
</tr>
<tr>
<td>(c_{3,3}^6)</td>
<td>(S(c_{3,3}^6) = {P(v_3, v_6)})</td>
<td>(S_D(c_{3,3}^6) = \emptyset)</td>
<td>(F(c_{3,3}^6) = f(v_3, v_6) = 50)</td>
</tr>
<tr>
<td>(c_{3,3}^7)</td>
<td>(S(c_{3,3}^7) = {P(v_3, v_6)})</td>
<td>(S_D(c_{3,3}^7) = \emptyset)</td>
<td>(F(c_{3,3}^7) = f(v_3, v_6) = 50)</td>
</tr>
<tr>
<td>(c_{4,5}^1)</td>
<td>(S(c_{4,5}^1) = {P(v_4, v_5)})</td>
<td>(S_D(c_{4,5}^1) = \emptyset)</td>
<td>(F(c_{4,5}^1) = f(v_4, v_5) = 100)</td>
</tr>
<tr>
<td>(c_{4,5}^2)</td>
<td>(S(c_{4,5}^2) = {P(v_4, v_5)})</td>
<td>(S_D(c_{4,5}^2) = \emptyset)</td>
<td>(F(c_{4,5}^2) = f(v_4, v_5) = 100)</td>
</tr>
<tr>
<td>(c_{4,7}^1)</td>
<td>(S(c_{4,7}^1) = {P(v_4, v_7)})</td>
<td>(S_D(c_{4,7}^1) = \emptyset)</td>
<td>(F(c_{4,7}^1) = f(v_4, v_7) = 155)</td>
</tr>
<tr>
<td>(c_{4,7}^2)</td>
<td>(S(c_{4,7}^2) = {P(v_4, v_7)})</td>
<td>(S_D(c_{4,7}^2) = \emptyset)</td>
<td>(F(c_{4,7}^2) = f(v_4, v_7) = 155)</td>
</tr>
</tbody>
</table>
This result implies that, if \( \alpha \geq 0.5 \), then \( c^2_{1,2}, c^1_{1,3}, \) and \( c^5_{2,3} \) are the local optimal locations to \( \hat{T}_1 \); otherwise, \( c^4_{1,7} \) and \( c^2_{4,7} \) are the local optimal locations that cover the maximum traffic flow (in round trips per time unit) in \( \hat{T}_1 \).

Similarly, Step 2 is repeated for \( \hat{T}_2 \). Sub-step 2.1 constructs \( \mathcal{L}_2 = \{P(v_8, v_9)\} \). Sub-step 2.2 generates \( EP(v_9, v_9) = \{v_8, v_9\}, EP(v_{10}, v_{11}) = \{v_{10}, v_{11}\} \), \( CSV(v_9, v_9) = \{v_8\}, CSV(v_{10}, v_{11}) = \{v_{10}\}, CRP(v_8, v_9; v_8) = \{r_1, r_2\}, \) and \( CRP(v_{10}, v_{11}; v_{10}) = \{r_3\} \), where \( r_1, r_2, \) and \( r_3 \) are the three cycle returning points obtained by the Cycle Returning Point Algorithm. Then, by Expression (5.4), \( CP_2 = \{c^1_{8,9}, c^2_{8,9}, c^3_{8,9}, c^4_{8,9}, c^5_{9,11}, c^1_{10,11}, c^2_{10,11}, c^3_{10,11}, c^4_{10,11}\} \). The right side of Figure 5-3 graphically shows the locations of \( RS(v_i, v_j) \), \( SC(v_i, v_j; v_5, r) \), and the corresponding \( c \in CP_2 \) in \( \hat{T}_2 \), for \( P(v_i, v_j) \in \mathcal{L}_2, v_5 \in CSV(v_i, v_j), r \in CRP(v_i, v_j; v_5) \). For \( c \in CP_2 \), Sub-step 2.3 determines \( S(c) \) and \( S_D(c) \). They are summarized in the left and middle sides of Table 5-5. Next, Sub-step 2.4 first computes \( F(c) \), for \( c \in CP_2 \). The calculations are summarized in the right side of Table 5-5. Sub-step 2.4 then compares \( F(c) \), for \( c \in CP_2 \), and identifies the local optimal set of candidate points \( CP^*_2 \) in \( \hat{T}_2 \), for \( 0 \leq \alpha \leq 1 \), as follows:

\[
CP^*_2 = \arg \max \{F(c^1_{8,9}), F(c^2_{8,9}), F(c^3_{8,9}), F(c^4_{8,9}), F(c^5_{9,11}), F(c^1_{10,11}), F(c^2_{10,11}), F(c^3_{10,11}), F(c^4_{10,11})\} \\
= \arg \max \{140, 140, 140, 220\alpha, 140\alpha, 80, 80, 80, 220\alpha\} \\
= \{c^1_{8,9}, c^2_{8,9}, c^3_{8,9}, c^4_{8,9}, c^1_{10,11}\}.
\]

This means that, if \( \alpha \geq 0.64 \), then \( c^4_{8,9} \) and \( c^1_{10,11} \) are local optimal candidate points in \( \hat{T}_2 \); otherwise, \( c^1_{8,9} \), \( c^2_{8,9} \), and \( c^3_{8,9} \) become local optimal to \( \hat{T}_2 \).

Finally, by comparing \( F(x^*_1) \), for \( x^*_1 \in CP^*_1 \), and \( F(x^*_2) \), for \( x^*_2 \in CP^*_2 \), Step 3 derives the global optimal set of candidate points \( CP^* \) for the original tree \( T \) that covers the maximum traffic flow in round trips per time unit (\( F^* \)).
\[ CP^* = \arg \max \{ F(x^*_1), F(x^*_2) \mid x^*_1 \in CP^*_1 \text{ and } x^*_2 \in CP^*_2 \} \]

\[ = \arg \max \left\{ F(c^1_{1,2}), F(c^1_{1,3}), F(c^5_{2,3}), F(c^4_{4,7}), F(c^4_{4,7}), \right\} \]

\[ = \arg \max \left\{ \{120 + 70\alpha, 120 + 70\alpha, 120 + 70\alpha, 155, 155\} \right\} \]

\[ = \{ c^2_{1,2}, c^1_{1,3}, c^5_{2,3}, c^4_{4,7}, c^4_{4,7}, c^4_{8,9}, c^4_{10,11} \} \]

Table 5.5. Sets of paths and the traffic flows (in round trips per time unit) covered by candidate points in the second sub-tree

<table>
<thead>
<tr>
<th>c</th>
<th>S(c)</th>
<th>S_D(c)</th>
<th>F(c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>c^9_{8,9}</td>
<td>S(c^9_{8,9}) = (P(v_0, v_9))</td>
<td>S_D(c^9_{8,9}) = \emptyset</td>
<td>F(c^9_{8,9}) = f(v_0, v_9) = 140</td>
</tr>
<tr>
<td>c^9_{8,9}</td>
<td>S(c^9_{8,9}) = (P(v_0, v_9))</td>
<td>S_D(c^9_{8,9}) = \emptyset</td>
<td>F(c^9_{8,9}) = f(v_0, v_9) = 140</td>
</tr>
<tr>
<td>c^9_{8,9}</td>
<td>S(c^9_{8,9}) = (P(v_0, v_9))</td>
<td>S_D(c^9_{8,9}) = \emptyset</td>
<td>F(c^9_{8,9}) = f(v_0, v_9) = 140</td>
</tr>
<tr>
<td>c^9_{8,9}</td>
<td>S(c^9_{8,9}) = \emptyset</td>
<td>S_D(c^9_{8,9}) = (P(v_0, v_9), P(v_{10}, v_{11}))</td>
<td>F(c^9_{8,9}) = \alpha(f(v_0, v_9) + f(v_{10}, v_{11})) = 220\alpha</td>
</tr>
<tr>
<td>c^9_{8,9}</td>
<td>S(c^9_{8,9}) = \emptyset</td>
<td>S_D(c^9_{8,9}) = (P(v_0, v_9), P(v_{10}, v_{11}))</td>
<td>F(c^9_{8,9}) = \alpha(f(v_0, v_9) + f(v_{10}, v_{11})) = 220\alpha</td>
</tr>
</tbody>
</table>

The set of optimal candidate points \( CP^* \) changes as the deviation portion \( \alpha \) varies between 0 and 1. When \( 0 \leq \alpha < 0.5 \), the two candidate points in \( CP^* = \{ c^1_{4,7}, c^2_{4,7} \} \) are optimal locations in tree \( T \) covering a traffic flow of 155 round trips per time unit. Since \( c^1_{4,7} \) and \( c^2_{4,7} \) are consecutive optimal candidate points satisfying the conditions of Theorem 5-4, all interior points between \( c^1_{4,7} \) and \( c^2_{4,7} \) are also optimal, and therefore, the complete set of optimal solutions \( CS^* = \{ c^1_{4,7}, c^2_{4,7} \} \). If \( 0.5 \leq \alpha < 0.8 \), the three candidate points in \( CP^* = \{ c^2_{4,7}, c^1_{1,3}, c^5_{2,3} \} \) are optimal locations with a traffic flow coverage of 155 to 176 round trips per time unit depending on \( \alpha \). Since these three optimal candidate points are located in the same spot, \( CS^* = \{ c^2_{4,7}, c^1_{1,3}, c^5_{2,3} \} \).

Lastly, if \( 0.8 \leq \alpha \leq 1 \), the two candidate points in \( CP^* = \{ c^4_{8,9}, c^4_{10,11} \} \) are optimal solutions with a traffic flow coverage of 176 to 220 round trips per time unit. Since \( c^4_{8,9} \) and \( c^4_{10,11} \) are
consecutive optimal candidate points that satisfy the conditions of Theorem 5-4, $C_S^* = \ell(c_{8,9}^4, c_{10,11}^4)$. Figure 5-4 illustrates the change of the optimal solutions in tree $T$ as the value of $\alpha$ changes from 0 to 1.

Figure 5-3. Refueling segments, symmetric cycles, and candidate points in two sub-trees

Figure 5-4. Change of the optimal solutions for the original tree network as the value of the deviation portion changes
5.6.2 Sensitivity Analysis for Coupled Effects of Deviation Portion and Vehicle Driving Range

In this section, we analyze the coupled effects of deviation portion and vehicle driving range on the set of optimal locations and the maximum traffic flow covered in tree $T$. The actual deviation portion $\alpha$, which refers to the portion of drivers willing to deviate from their preplanned paths for refueling service per time unit, is difficult to obtain or likely to be inaccurately estimated. Thus, the values of $\alpha$ used in this analysis range from 0 to 1. The vehicle driving range $R$, which depends on the fuel consumption, is expected to increase substantially with improving technologies (Park 2014). To reflect about this point, the values of $R$ used in this section range from $R = 40$ to $R = 120$, which are $\pm 50\%$ of the value of $R = 80$ used in the previous section, in increments of 20.

The first and last three columns of Table 5-6 show the complete set of optimal solutions $CS^*$ and corresponding traffic flow covered $F^*$ (in round trips per time unit) for five different vehicle driving ranges ($R = 40$, $R = 60$, $R = 80$, $R = 100$, and $R = 120$) and deviation portions between 0 and 1 for the proposed model. In addition, Figure 5-5 displays the trade-off between deviation portion $\alpha$ and maximum traffic flow covered $F^*$ for the five different vehicle driving ranges. From these results, we first notice that the maximum traffic flow covered $F^*$ increases or at least stays the same as the deviation portion $\alpha$ increases for a given vehicle driving range. For example, when the vehicle driving range is fixed to $R = 80$, $F^* = 155$ as $\alpha$ increases from 0 to 0.5; then $F^*$ increases from 155 to 220 round trips per time unit as $\alpha$ increases from 0.5 to 1. Note also that the value of $F^*$ increases or stays the same as the vehicle driving range $R$ increases for a given value of $\alpha$ between 0 and 1. For example, when the deviation portion is fixed to $\alpha = 0.5$, $F^* = 140$ for $R = 40$; $F^* = 155$ for $R = 60$ and $R = 80$; $F^* = 190$ for $R = 100$; $F^* = 435$ for $R = 120$. 
Table 5-6. Optimal locations and the corresponding traffic flows covered for each vehicle driving range (and deviation portion combination) for three models

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>C. Set of Optimal Solutions (vertices)</td>
<td>Complete Set of Optimal Solutions (CS*)</td>
<td>Traffic Flow Covered (F*)</td>
</tr>
<tr>
<td></td>
<td>Traffic Flow Covered (F*)</td>
<td>Traffic Flow Covered (F*)</td>
<td>Deviation Portion (α)</td>
</tr>
<tr>
<td>40</td>
<td>{v_8, v_9}</td>
<td>140</td>
<td>ℓ(c_1, c_2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>α = 1</td>
</tr>
<tr>
<td>60</td>
<td>{v_8, v_9}</td>
<td>140</td>
<td>{c_5}</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.00 ≤ α &lt; 0.50</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.50 ≤ α &lt; 0.80</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.80 ≤ α ≤ 1.00</td>
</tr>
<tr>
<td>80</td>
<td>{v_8, v_9}</td>
<td>140</td>
<td>ℓ(c_6, c_7)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.57 ≤ α &lt; 1.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>α = 1</td>
</tr>
<tr>
<td>100</td>
<td>{v_2}</td>
<td>190</td>
<td>{c_11}</td>
</tr>
<tr>
<td>120</td>
<td>{v_4}</td>
<td>435</td>
<td>{c_14}</td>
</tr>
</tbody>
</table>

Note: The location of candidate points in the complete set of optimal solutions (CS*) are shown in Figure 5-6.
Figure 5-5. Trade-off between deviation portion and maximum traffic flow covered for the five different vehicle driving ranges (R=40, R=60, R=80, R=100, and R=120)

Figure 5-6 graphically presents the locations of the optimal candidate points listed in Table 5-6. It can be observed that, when the maximum traffic flow covered $F^*$ stays the same, the complete set of optimal solutions $CS^*$ may expand as the deviation portion $\alpha$ increases and the vehicle driving range $R$ remains constant. For example, when $R = 40$ and $\alpha$ increases from 0 to 1, the maximum traffic flow covered $F^*$ stays at 140 round trips per time unit, but $CS^*$ is expanded from $\ell(c_1, c_2)$, for $0 \leq \alpha < 1$, to $\ell(c_1, c_2) \cup \ell(c_1, c_3) \cup \ell(c_1, c_4)$, for $\alpha = 1$. Similarly, while $F^*$ remains contact, $CS^*$ may enlarge as $R$ increases and $\alpha$ stays the same. For example, given that $0 \leq \alpha < 0.5$, as $R$ increases from 60 to 80, $F^*$ stays at 155 round trips per time unit, but $CS^*$ is expanded from one candidate point $\{c_5\}$ to line segment $\ell(c_6, c_7)$ including $\{c_5\}$. 
5.6.3 Performance Analysis

In this section, the optimal solution of the proposed model for the single station location problem is compared with those of the FRLM (Kuby and Lim 2005) and the original continuous single refueling station location model (Ventura et al. 2015) using the same numerical example introduced at the beginning of Section 5.6. The main distinctions among these three models can be summarized briefly. First, the FRLM assumes that the set of candidate sites for refueling stations is finite and equal to the set of vertices of the network. On the other hand, Ventura et al. (2015) model considers that the station can be located anywhere along the network, but vehicle deviations are not allowed. Lastly, the proposed model also considers the continuous version of the refueling station location problem, but a portion of drivers are willing to deviate if the
refueling station is not located along their preplanned paths. Ventura et al. (2015) model can be seen as a special case of the proposed model with $\alpha = 0$.

The first column of Table 5-6 shows the five values of $R$ that are considered. The optimal locations and corresponding traffic flow covered $F^*$ (in round trips per time unit) for the FRLM are shown in columns 2 and 3, and for Ventura et al. (2015) model are provided in columns 4 and 5. The last three columns of Table 5-6 include the results of the proposed model as explained in Section 5.6.2. Figure 5-7 displays the coupled effects of vehicle driving range and deviation portion on the maximum traffic flow $F^*$ for the three models: FRLM, Ventura et al. model, and proposed model.

Comparison among the three models in Table 5-6 and Figure 5-7 verifies that the proposed model has better performance than the other two models in terms of maximum traffic flow covered and set of optimal locations. First, Figure 5-7 shows that the maximum traffic flow covered $F^*$ of the proposed model is always higher than or equal to those of the other two models for all values of $R$. For example, when $R = 80$, the proposed model with $\alpha = 1.00$ finds the optimal locations covering 220 round trips (per time unit), which is 57% and 42% higher than the round trips covered at the suboptimal locations found using the FRLM and Ventura et al. model, respectively. Next, Table 5-6 also demonstrates that, when all three models find the same value of $F^*$, the proposed model may be able to detect more optimal locations than the other two models.

For example, when $R = 40$, $F^* = 140$ for all three models, but the set of optimal locations found by the proposed model, which is $\mathcal{C}^* = \ell(c_1, c_2) \cup \ell(c_1, c_3) \cup \ell(c_1, c_4)$, includes the two optimal locations, $\{v_8, v_9\}$, found by the FRLM and the set of optimal locations, $\ell(c_1, c_2)$, found by the Ventura et al. model.
Figure 5-7. Coupled effects of vehicle driving range and deviation portion on the maximum traffic flow covered for the three models (FRLM, Ventura et al. model, and proposed model)

5.7 Concluding Remarks

This chapter has studied the continuous version of the single refueling station location problem on a tree network considering that a given portion of drivers have the option to deviate from their preplanned paths to be able to refuel their vehicles. The objective of the problem is to identify the set of optimal station locations that maximizes the total traffic flow covered (in round trips per time unit). To achieve this goal, we have first generated a set of sub-trees from the original tree network to reduce the problem size and consider the deviation option. Next, we have identified three sets of candidate points, including endpoints of refueling segments, cycle starting vertices, and cycle returning points, to locate the refueling station. We have proved that there exists at least one optimal solution in these sets of candidate points. Then, we have developed an algorithm to determine the set of optimal candidate points that maximizes the total traffic flow covered. Lastly, we have derived conditions under which the line segments defined by certain pairs of consecutive optimal candidate points are also optimal. Our performance analysis demonstrates that the
proposed model can significantly improve the sub-optimality of the solutions found by the existing refueling station location models.
Chapter 6

CONCLUSIONS AND FUTURE RESEARCH

6.1 Conclusions

As alternative-fuel vehicles have come into the spotlight with their potential to reduce tailpipe emissions in the ground transportation sector, refueling station location problems for alternative-fuel vehicles have received attention as well. Refueling station location problems can be classified into two types depending on the set of candidate sites: when a preliminary (finite) set of candidate sites is given, the problem is called discrete; when the refueling stations are able to be located anywhere along the network, the problem is called continuous. In this dissertation, one discrete and two continuous refueling station location problems have been addressed, so as to contribute to the development of alternative-fuel refueling infrastructure.

First in Chapter 3, a bi-criteria binary linear programming model has been developed to solve the discrete location problem for alternative-fuel refueling stations on a directed transportation network. Two conflicting objectives have been considered to maximize the total VMT per time unit covered by the stations and minimize the capital cost for building the refueling infrastructure. Two methods, i.e., the normalized weighting method and the constraint method, have been proposed to solve the bi-criteria binary linear programming models. These two methods then have been validated with an application to the Pennsylvania Turnpike System regarding the location of LNG refueling stations on the 19 existing service plazas. Based on the estimated construction costs for the LNG refueling stations at the 8 service plazas by Myers et al. (2013), a linear regression model has been developed to estimate the construction costs for the 11 remaining service plazas in the PA Turnpike. The effective VMT coverage has been determined as a function of the LNG infrastructure cost on the PA Turnpike. Finally, the effect of LNG
refueling station construction on GHG emissions reduction has been discussed, and the relationship between cost of construction and the SCC savings has also been derived.

Next in Chapter 4, the continuous location problem has been addressed on a tree network to determine the optimal set of locations for a single refueling station with the objective of maximizing the traffic flow covered, on the assumption that drivers do not deviate from their preplanned (shortest) paths. Some properties, lemmas, and theorems have been derived to reduce and solve the problem efficiently. Then, a polynomial-time algorithm has been developed to find the optimal set of locations for the station that covers the maximum traffic flow, and an additional theorem has been built to expand the optimal locations to a certain line segment.

Lastly in Chapter 5, the continuous location problem proposed in Chapter 4 has been expanded to the version, where drivers have the option of selecting a deviation path if a refueling station is not located along their shortest routes. A property and a polynomial-time algorithm related to the deviation option and a set of candidate points have been determined. Then, some optimality conditions have been derived to find the points that cover the maximum traffic flow when a given portion of drivers have the deviation options. Based on these conditions, a polynomial-time algorithm has been developed to find the optimal points of this problem. Next, the complete set of optimal solutions has been determined by identifying additional optimal points located in the interior of paths to which the endpoints belong to the preliminary set of optimal points. The proposed model has been validated with an application to a numerical example regarding the coupled effects of deviation portion and vehicle driving range on the set of optimal locations and the maximum traffic flow covered as well as its performance. The performance analysis in the numerical example has shown that the proposed model can significantly improve the sub-optimality of the solutions found by the existing refueling station location models.
6.2 Future Research

In this section, a few directions are suggested for future research based on the studies presented in this dissertation. Section 6.2.1 discusses future research to improve the accuracy of the estimates for GHG emission changes realized when converting to an LNG refueling infrastructure. This section also discusses a possible extension of the bi-criteria binary linear programming model introduced in Chapter 3 for future research. Next, Section 6.2.2 presents future research for the continuous location problems for a single alternative-fuel refueling station on the tree-type transportation network addressed in Chapters 4 and 5.

6.2.1 Future Research Regarding Chapter 3

In Chapter 3, given a preliminary (finite) set of candidate sites, a bi-criteria binary linear programming model has been developed, so as to find the optimal locations for LNG refueling stations and analyze the estimates for GHG emissions reduction realized when converting to LNG refueling infrastructure. There may be better ways to improve the estimates of the reductions in GHG emissions realized when switching from a conventional fossil fuel (i.e., gasoline or diesel) to an LNG refueling infrastructure. First, since we do not have the traffic flow data in the form of the OD pairs outside of the PA Turnpike, it is currently difficult to estimate the distance that a truck travels outside of the PA Turnpike system. Given the study assumptions that a truck will have at least a half-tank of fuel when it exits the network, along with the estimates that a truck could be carrying a fuel tank with up to a 600-mile driving range, there exists a significant amount of uncertainty regarding the true travel distance of an LNG truck, and the estimates provided for the VMT converted to LNG should be viewed as a lower bound for the true VMT if all diesel trucks whose VMT were part of this study were converted to LNG.
Next, the proposed studies in Chapter 3 contain uncertainty regarding the truck weights used. High-level data which were collected from a study of 15 states were used in Chapter 3 to estimate the overall proportion of trucks at a variety of weights. However, data relating the weight of a specific truck to the round-trip distance it travels in the network are not considered. If this data were available, we would have been able to more accurately estimate the GHG emission differences between the current and prospective future refueling strategies for long-haul HDVs (high duty vehicles). This data may be collectible in the future through collaboration with industry partners, particularly if the GHG emissions reduction for truck fleets were part of a rebate structure or carbon trading policy.

In addition, we can improve the regression model proposed in Section 3.3.2 to estimate the construction costs better for new LNG refueling stations at the service plazas along the PA Turnpike. Note that the proposed linear regression model uses the original values of two independent variables (natural gas supply line length and pavement size) as predictors to estimate the construction costs. Applying a logarithm transformation to independent and/or dependent variables may enable the linear regression model to fit better.

Lastly, the bi-objective model proposed in Chapter 3 can be extended to a number of topics for future research to consider various situations. As one of the future research recommendations, a capacitated version of this model can be proposed to consider the maximum amount of alternative-fuel that could be dispensed per time unit before the station would need to be replenished, as suggested by Upchurch et al. (2009) in their capacitated version of the FRLM. The proposed model can also be extended to a multi-period version to incorporate demand patterns considering a given transportation network over multiple periods into strategic refueling station location planning. A bi-objective stochastic model can be another follow-up study of the dissertation to address the refueling station location problem under time-varying demand on a
given transportation network. Incorporating these concepts into the design of an LNG refueling infrastructure could improve the practicability and usefulness of the results.

### 6.2.2 Future Research Regarding Chapters 4 and 5

In Chapters 4 and 5, on the assumption that a refueling station can be located anywhere along a transportation network, a continuous location problem for a single refueling station and its extension where drivers are allowed to deviate from their shortest paths for refueling service have been addressed on a tree-type transportation network. These problems can be expanded to a number of topics for future research. First, some of the assumptions made in Chapters 4 and 5 can be relaxed. For example, some vehicles may have different fuel tank levels at their origin and destination points, and may have different fuel consumption rates as well, which imply that vehicles can travel different distances without refueling. To incorporate these situations into the continuous location problem, simulation techniques can be applied.

Second, while the proposed algorithms in Chapters 4 and 5 are based on a tree-type transportation network, most real transportation networks are cyclic having multiple paths between most pairs of vertices. Thus, the algorithms based on the tree network that have been developed in Chapters 4 and 5 can be expanded to the cyclic networks.

Next, a continuous location problem for a single refueling station can be expanded to multiple refueling stations on a large-scale transportation network. Many vehicles on a large-scale network can make a long-distance trip and may require several refueling stops at different locations to complete their trips. The multiple refueling station location problem can address the issue of multiple refueling stops per trip to find the optimal locations for refueling stations that cover the maximum traffic flow.
In addition, as similar to Chapter 3, a bi-objective version of the continuous location problem for refueling stations can be considered to identify optimal locations that both maximize the total traffic flow covered by the stations and minimize construction and maintenance costs of the stations, assuming that these costs differ regionally and stations can be located anywhere along a transportation network.

Lastly, the continuous location problem suggested in Chapters 4 and 5 can also be applied to the other areas. For example, in communication and pipeline networks, demands such as radio signals, liquids, and gases have a limited travel range, and therefore, planners of these networks may require to place facilities such as repeaters and compressors in the networks to efficiently deliver demands from origin to destination points. If these facilities can be located anywhere along the networks, then we can consider these problems as the continuous facility location problem and apply the proposed algorithms to find the set of optimal locations for these facilities.
REFERENCES


Burnham, A. 2013. Liquefied Natural Gas. Case Study. Argonne National Laboratory, Argonne, IL.


Federal Highway Administration. 2016a. National Highway System. U.S. Department of Transportation. Available at:


Department of Industrial and Manufacturing Engineering, Pennsylvania State University, University Park, PA.


hydrogen energy 34(15) : 6045-6064.


VITA

Sang Jin Kweon

Education
Ph.D. Pennsylvania State University, University Park, USA Aug. 2017
Dual-title Degree Program in Industrial Engineering and Operations Research, GPA 3.88/4.00

M.S. Pennsylvania State University, University Park, USA Dec. 2013
Industrial and Manufacturing Engineering, GPA 3.88/4.00

B.S. Yonsei University, Seoul, South Korea Feb. 2007
Computer Science and Industrial System Engineering, GPA: 3.91/4.30 (High Honors Graduation)

Journal Publications


