A Framework for Designing of Electric Vehicle Charging Infrastructure Network

Shengyin Li
Clemson University, shengyl@clemson.edu

Follow this and additional works at: https://tigerprints.clemson.edu/all_dissertations
Part of the Transportation Commons

Recommended Citation
https://tigerprints.clemson.edu/all_dissertations/1546

This Dissertation is brought to you for free and open access by the Dissertations at TigerPrints. It has been accepted for inclusion in All Dissertations by an authorized administrator of TigerPrints. For more information, please contact kokeefe@clemson.edu.
A FRAMEWORK FOR DESIGNING OF ELECTRIC VEHICLE CHARGING INFRASTRUCTURE NETWORK

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Civil Engineering

by
Shengyin Li
December 2015

Accepted by:
Dr. Yongxi Huang, Committee Chair
Dr. Mashur (Ronnie) Chowdhury
Dr. Akshay Gupte
Dr. Zhen (Sean) Qian
Abstract

The lack of sufficient public charging stations for electric vehicles has long been recognized as a major hinder for massive adoption of electric vehicles (EV). This dissertation aims to develop a framework for designing charging station infrastructure networks that electric vehicle with limited travel range can be recharge en-route to complete trips to destinations and then would facilitate the adoption of electric vehicles.

The first part of this dissertation is concerned with modeling travel range of electric vehicle and users behavior of deviating from their most preferred routes when siting charging stations. The proposed multi-path refueling location model provides the most cost-effective deployment strategy of placing charging stations that are needed on the network to satisfy electric vehicle travel demand between all origin-destination (O-D) pairs. In the second part of the dissertation, heuristic based on greedy adding algorithms are developed to address the computational challenges of the multi-path refueling location model. The heuristics are tested on the Sioux Falls network and a real-life case study of South Carolina and compared with the exact solutions.

In reality, however, EV market matures gradually, in other words, not all the cities would become electric vehicle adopters at the current state. In third part of this dissertation, a multi-period multi-path refueling location model is developed to
expand EV charging network to dynamically satisfy O-D trips with the growth of EV market. The model captures the dynamics in the topological structure of network and determines the cost-effective station rollout scheme on both spatial and temporal dimensions. The multi-period location problem is formulated as a mixed integer linear program and solved by a heuristic based on genetic algorithm. The model and heuristic are justified using the benchmark Sioux Falls road network and implemented in a case study of South Carolina. The results indicate that the charging station rollout scheme is subject to a number of major factors, including geographic distributions of cities, vehicle range, and deviation choice, and is sensitive to the types of charging station sites.

The last part of this dissertation presents an extension of the multi-path refueling location model to integrate probabilities of cities becoming EV market into optimization of location decisions. This probability-based model differs from the multi-path model in two major aspects. First it maximizes the total weighted coverage of all cities with a given budget while the multi-path model minimize the cost of covering all the O-D pairs. Second, instead of only consider one way trips as in the multi-path model, this model extends to also satisfy the round trips from destinations back to origins. A genetic algorithm based heuristic is adopted to solve this probability-based model. Numerical experiments are conducted to justify the incorporation of probability information in optimally siting charging station.
I would like to thank Dr. Yongxi Huang, my advisor, mentor, and friend, without his guidance this dissertation would have been impossible. I am grateful to his generous support and persistent encouragement. He is a true role model to me, academically and personally. Also, I am very grateful to other committee members, Dr. Mashrur Chowdhury, Dr. Akshay Gupte and Dr. Zhen Qian for their valuable time, insightful discussions and suggestions during my graduate study at Clemson University.

I would like to show my great appreciation to Dr. Scott J. Mason from the Department of Industrial Engineering, for his support and valuable suggestions on research and professional development. It was a joy to work with him.

Because of many fellow students and friends, my graduate study at Clemson University has been very stimulating and delightful. I especially appreciate Dr. Fei Xie, Jie Lu, Kwaku Brown and Dr. Zhiyun Li for their insights and assistants on research projects.

Finally, I want to give my deepest appreciation to my family for their tolerance, patience, support and love. I hope I can repay some of the support, encouragement, and love they have given to me.
# Table of Contents

Title Page ................................................................. i
Abstract ................................................................. ii
Acknowledgments ....................................................... iv
List of Tables .......................................................... vii
List of Figures ......................................................... ix

1 Introduction ............................................................ 1
   1.1 Background ...................................................... 1
   1.2 Research contributions ...................................... 2
   1.3 Structure of the dissertation .................................. 3

2 Literature review ..................................................... 5
   2.1 Flow-based location models .................................. 5
   2.2 Solution methods for flow-based location models .......... 7
   2.3 Multi-period location models .................................. 9
   2.4 Solution methods for multi-period location models ........ 10
   2.5 Probability-based location models ......................... 12

3 Multi-path refueling location model for electric vehicle charging stations ........................................... 14
   3.1 Problem statement ............................................. 14
   3.2 Methods .......................................................... 16
   3.3 Results and discussions ...................................... 26
   3.4 Summary ........................................................ 40

4 Heuristic approaches for the multi-path refueling location model ...................................................... 41
   4.1 Problem statement ............................................. 41
   4.2 Methods .......................................................... 42
   4.3 Results and discussions ...................................... 49
   4.4 Summary ........................................................ 56
List of Tables

Table 2.1 Summary of differences between the proposed MPRLM and existing flow-based models ........................................ 7
Table 3.1 Number of paths served by different locations ..................... 30
Table 3.2 Comparisons of computational performances of models (P) and (P1) ................................................................. 33
Table 3.3 The average deviation and charging station locations under different deviation scenarios ........................................ 38
Table 3.4 Numbers of paths used in the Sioux Falls network under different deviation scenarios ........................................ 39
Table 3.5 Numbers of paths used in 25-node network under different deviation scenarios ........................................ 40
Table 4.1 Node weights in iteration #1 with shortest path scenario ...... 45
Table 4.2 Node weights in iteration #2 with shortest path scenario ...... 46
Table 4.3 Node weights in iteration #1 with deviation path scenario ...... 46
Table 4.4 Node weights in iteration #2 with deviation path scenario ...... 47
Table 4.5 Number of stations by heuristics and exact solutions for the Sioux Falls network .................................................. 50
Table 4.6 Solving times (in CPU seconds) of heuristic and exact solutions for Sioux Falls network ........................................ 51
Table 4.7 Statistics on results comparing heuristic solutions with the exact solution ......................................................... 52
Table 4.8 Results of SC network (15 cities) ....................................... 55
Table 4.9 Effects of deviations on SC network (vehicle range of 85 miles) 56
Table 5.1 M2PRLM solution ($K = 1$) ........................................... 72
Table 5.2 M2PRLM solution ($K = 3$) ........................................... 72
Table 5.3 Heuristic parameters and performance ................................ 73
Table 5.4 Comparisons between heuristic and CPLEX ....................... 74
Table 5.5 Myopic solution with relocation ($K = 1$) .......................... 75
Table 5.6 Myopic solution without relocation ($K = 1$) ....................... 75
Table 5.7 Comparison between heuristic and CPLEX ....................... 77
Table 5.8 Comparisons between M2PRLM and myopic solutions ........ 79
Table 5.9 Effects of weighting factors on rollout schemes .................. 81
Table 6.1  Probability of adopting EVs ........................................ 95
Table 6.2  Comparison of coverage for cities with high (≥ 0.80) or low
(≤ 0.20) probabilities ......................................................... 101
List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>A 7-node sample network</td>
<td>17</td>
</tr>
<tr>
<td>3.2</td>
<td>Two test networks</td>
<td>27</td>
</tr>
<tr>
<td>3.3</td>
<td>Location of charging stations on the Sioux Falls network with vehicle range of 100 under three deviation scenarios</td>
<td>29</td>
</tr>
<tr>
<td>3.4</td>
<td>Multiple location solutions for $K=1$ and vehicle range of 100</td>
<td>31</td>
</tr>
<tr>
<td>3.5</td>
<td>Effects of deviation caps and vehicle ranges on the minimum numbers of charging stations needed on the Sioux Falls network</td>
<td>32</td>
</tr>
<tr>
<td>3.6</td>
<td>Effects of $K$ and vehicle ranges on the minimum numbers of charging stations needed on the Sioux Falls network</td>
<td>33</td>
</tr>
<tr>
<td>3.7</td>
<td>Effects of the vehicle range deduction on the minimum numbers of charging stations needed on the Sioux Falls network</td>
<td>34</td>
</tr>
<tr>
<td>3.8</td>
<td>Locations of charging stations on the 25-node network with a vehicle range of 10 under three deviation scenarios</td>
<td>35</td>
</tr>
<tr>
<td>3.9</td>
<td>Effects of the deviation caps and vehicle range on the minimum numbers of charging stations needed on 25-node network</td>
<td>36</td>
</tr>
<tr>
<td>3.10</td>
<td>Effects of $K$ and the vehicle ranges on the minimum numbers of charging stations needed on 25-node network (note: $K=4$ and $K=5$ lines are completely overlapped)</td>
<td>37</td>
</tr>
<tr>
<td>3.11</td>
<td>Effects of the vehicle range deduction on the minimum numbers of charging stations needed on the 25-node network</td>
<td>38</td>
</tr>
<tr>
<td>4.1</td>
<td>A seven-node sample network (the circled numbers represent node indices, numbers on the links denote the lengths, and all links are bidirectional)</td>
<td>45</td>
</tr>
<tr>
<td>4.2</td>
<td>A diagram of the integrated GA-E algorithm</td>
<td>47</td>
</tr>
<tr>
<td>4.3</td>
<td>Effects of network size on the solution times</td>
<td>58</td>
</tr>
<tr>
<td>4.4</td>
<td>South Carolina state network (15 cities)</td>
<td>59</td>
</tr>
<tr>
<td>4.5</td>
<td>South Carolina state network (40 cities)</td>
<td>59</td>
</tr>
<tr>
<td>5.1</td>
<td>The South Carolina network</td>
<td>76</td>
</tr>
<tr>
<td>5.2</td>
<td>Charging station deployment over time ($K=3$ and VR=150 miles)</td>
<td>84</td>
</tr>
<tr>
<td>5.3</td>
<td>Cumulative distribution of percentage of O-D pairs covered under random tests</td>
<td>85</td>
</tr>
</tbody>
</table>
Figure 5.4 Charging station deployment with differentiation of the types of sites ($K = 3$ and VR=150) ............................................. 86

Figure 6.1 Definition of coverage for a city ............................................. 89
Figure 6.2 The Sioux Falls network ......................................................... 94
Figure 6.3 Optimality gap ................................................................. 95
Figure 6.4 Comparison of heuristic and CPLEX on solving time ............... 96
Figure 6.5 Detailed location and route choice decisions for VR=100 and $K=3$ ................................................................. 97
Figure 6.6 Detailed charging schedule for O-D pair (1,21) ......................... 97
Figure 6.7 Impact of budget and vehicle range ........................................ 98
Figure 6.8 Impact of initial battery status ............................................. 99
Figure 6.9 Coverage ($Z$) for node #4 and #24 when VR=100 and $K=3$ . 100
Figure 6.10 Detailed location and route choice decisions for VR=100 and $K=3$ with equal probability ......................................... 100
Chapter 1

Introduction

1.1 Background

Plug-in electric vehicles (PEVs, and for simplicity I will use EV instead for the remainder of this dissertation), including plug-in hybrid electric vehicles (PHEVs) and battery electric vehicles (BEVs), have long been recognized as one of the promising alternatives to supplant internal combustion engines (ICE) powered vehicles and as an effective way to alleviate the dependency on petroleum and to reduce greenhouse gas emissions. EVs cost as little as 2 to 3 cents per mile (compared to 13 cents per mile for ICE powered vehicles) and help reduce 50% CO$_2$ emissions per mile than ICE powered vehicles [63]. However, limited travel range and high life-cycle cost of ownership [65] have been identified as major barriers for massive market adoptions [31, 45]. Given that the travel range of most of EVs on market are less than 100 miles, drivers especially those traveling between cities are concerned about running out of power before reaching destinations and hesitate to by EVs unless either travel range of EVs is increased substantially (comparable with ICE vehicles) or sufficient public charging stations are located and available on transportation networks. However, due
to economic and technical concerns, improving vehicle range might be not affordable at least in the early stages. In fact, it is not even necessary to increase vehicle range to a very high level, as will be shown in this dissertation. Instead, to strategically place charging stations on transportation networks to ensure EV drivers would be to able to complete all the intercity trips within the network is much more promising and is attracting more interests from decision making perspective such as policy makers, EV manufacturers, energy corporations, and infrastructure planners.

1.2 Research contributions

This dissertation is focused on developing a framework that helps central planners to design an EV charging station infrastructure network with the lowest cost to satisfy all the intercity EV travel demand within the network. Both mathematical models and efficient solution methods are created to make the framework ready for application in real world large scale networks. Major research contributions of this study are listed as follows:

- **Integration of vehicle range, route choice and charging schedule into the design of charging station infrastructure networks.** A multi-path refueling location model (MPRLM) is formulated by taking into account the impact of vehicle range and deviated route choice. With the objective of minimizing cost of establishing the charging station network, the model determines location of charging station, route choice of drivers and detailed charging schedule on electrified paths for each O-D pair simultaneously.

- **Creation of an efficient heuristic to applying the proposed MPRLM to locate**
charging stations on a South Carolina network. Based on greedy adding algorithm, different node weighting strategies are considered to incorporate the impact of vehicle range and deviation choice when developing the heuristics.

- **Extension of the MPRLM to accommodate spatial expansion of EV travel demand over time.** A multi-period multi-path refueling location model (M2PRLM) with the consideration of relocating existing charging stations is created to optimize the rollout scheme of charging stations to accommodate spatial expansion of EV market. An efficient heuristic based on genetic algorithm is developed to solve the M2PRLM for real world cases.

- **Extension of the MPRLM to integrate probability of adopting EVs.** A probability-based multi-path refueling location model (P-MPRLM) is created to optimally siting charging stations by considering the different attitude of cities toward adopting EVs. An genetic algorithm based heuristic is developed to overcome the challenge of solving the problem.

These contributions together as a framework for designing of EV charging infrastructure network, answer questions about where to location charging stations, which routes in a network should be electrified, what are the detailed charging schedules for intercity EV drivers, and what is the best strategy to roll out charging stations distribution of EV travel demand are subject to substantial temporal or spatial variations.

### 1.3 Structure of the dissertation

The remainder of this dissertation is organized as follows. Existing studies in literature are first reviewed in Chapter 2, followed by the individual studies. In Chapter 7,
I will summarize the dissertation and will outline possible future research directions.
Chapter 2

Literature review

2.1 Flow-based location models

The fundamental question on how to deploy the discretionary facilities relates to the well-studied facility location problems [27], including the covering, center, and median problems, which all assume that there is a central planner who allocates supplies or services to satisfy demand on a spatial network. Many applications, including EV location problems [35, 36, 47, 64, 74], have treated demands as if they are located at specified nodes with cost measured by distance or travel time from these demand locations to the facilities.

However, it may be more realistic to model the demands as flows on the network if goods or services are obtained on the way, which leads to the evolutions of flow based location models. First, the FILM or FCLM [41, 12] is a maximal coverage model that entails facility locations to serve passing flows which are considered as captured if a facility is located on the flow paths. The model has been evolved to capture more realistic concerns, such as the different sized facilities [76] and flow uncertainty [78]. A critical issue—limited vehicle range, that was neglected in the FILM
or FCLM was incorporated in the FRLMs \[54, 51, 52, 79\], a new set of maximal flow coverage models that consider the effects of limited vehicle ranges for undertaking long-distance trips via multi-stop refueling. Distinct from the maximal flow coverage models, a series of flow-based set-covering models were developed to locate a minimum number of EVs while satisfying travel demands \[82, 83, 84\]. More recently, \[90\] further extended the model to consider heterogeneous types of EVs and fueling capacities. A comprehensive review on the flow-based facility location problems is referred to \[33, 91\]. Both the maximal coverage and set-covering models were recently generalized in \[62, 85\].

All the models were formulated based on a general assumption that drivers would only consider a shortest distance/time path between origins and destinations. Due to the sparse distribution of EV charging stations on networks, EV users, however, may be willing to take a slightly longer path (i.e., a deviation path) to ensure that they can refuel their vehicles en route, particularly for long-distance trips. This alternate routing consideration is realistic as drivers can now use available mobile map applications (e.g., Google Map) to familiarize themselves with the transportation network. Building upon the deviation assumption in the FILM-D \[12\], \[51\] formulated the DFRLM which extends the FRLM to maximize the total flows covered via at most one path (including deviation paths) for each origin destination pair that contributes most to the objective. As seen in the literature, both maximal coverage and minimum cost are possible objectives for different purposes. The maximal-coverage models provide budget-constrained location solutions which however do not intend to satisfy all demand. In contrast, minimum-cost models satisfy all demands, which can be used to provide cost assessment of long-term strategic plans of EV charging station placements. The major differences between these two types of models are summarized in Table 2.1.
Table 2.1: Summary of differences between the proposed MPRLM and existing flow-based models

<table>
<thead>
<tr>
<th>Models</th>
<th>Objectives</th>
<th>Major constraints</th>
<th>Paths considered between an O-D pair</th>
<th>Routes as decision variables</th>
<th>Applications (examples)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MPRLM (proposed)</td>
<td>Minimize the cost of locating refueling stations (set-covering)</td>
<td>Satisfy all travel demand Vehicle range</td>
<td>Multiple paths including shortest paths and deviation paths</td>
<td>Yes</td>
<td>Potential application for EV charging stations, hydrogen refueling stations, etc.</td>
</tr>
<tr>
<td>FCLM [12, 41]</td>
<td>Maximize covered flows (max coverage)</td>
<td>Number of facilities</td>
<td>One shortest path</td>
<td>No</td>
<td>Billboard inspection stations</td>
</tr>
<tr>
<td>FRLM [54]</td>
<td>Maximize covered flows (max coverage)</td>
<td>Number of charging stations Vehicle range</td>
<td>One shortest path</td>
<td>No</td>
<td>EV charging stations Hydrogen stations</td>
</tr>
<tr>
<td>DFRLM [51]</td>
<td>Maximize covered flows (max coverage)</td>
<td>Number of charging stations Vehicle range</td>
<td>At most one path including deviation paths</td>
<td>No</td>
<td>EV charging stations Hydrogen stations</td>
</tr>
<tr>
<td>Wang’s models</td>
<td>Minimize the cost of locating charging stations (set covering)</td>
<td>Satisfy all travel demand Vehicle range</td>
<td>One shortest path</td>
<td>No</td>
<td>EV charging stations Hydrogen stations</td>
</tr>
</tbody>
</table>

2.2 Solution methods for flow-based location models

Facility location problems are NP-hard [27]. Heuristics, especially greedy-adding heuristics, have been recognized as one of the most efficient heuristic approaches [17] for solving location problems. The heuristics iterate by selecting best nodes out of a set of candidate nodes one at a time and adding it to the solution set until a pre-defined stopping criterion (e.g., maximum number of locations or all demand satisfied) is reached. The method is called greedy because it does what it is believed to be the best at each iteration without foreseeing how the current decisions will impact later decisions [27]. There is a vast literature on greedy-adding algorithms for different types of facility location problems, including set-covering, p-median, p-center, and maximal coverage problems. In this study, I only focus on the heuristics for
set-covering problems, which are most relevant to the MPRLM.

The greedy-adding algorithm as an approximation algorithm for solving set-covering problems was firstly introduced in [48], in which each candidate location was weighted based on its coverage to demands. This weighting strategy (also called priority rules) was then revised by [24] to take into account the cost for launching facilities at nodes. Since then, progress has been made on designing different weighting strategies to better reflect the objectives [5, 16, 55], and more recently randomization schemes was introduced to further improve the solution quality [39, 55]. The greedy-adding algorithm has also been coupled with other solution methods, such as, subgradient optimization methods [5, 16] and Lagrangian relaxation methods [9, 13, 38, 88] to offer better initial solutions or to obtain solutions for relaxed sub-problems in each iteration.

All aforementioned algorithms are designated for node-based set-covering problems. For flow-based set-covering problems, demands are no longer based on nodes but associated with paths that connect O-D pairs instead. As a result, the old node weighting schemes cannot be directly applied to solve the flow-based models, as highlighted in the recent study [91].

The new heuristic approaches must also take into account vehicle range and deviation paths in an integrated manner. [58] used a similar node priority rule as in [91] with the combined effects of the vehicle range when solving the FRLM. Kim and Kuby developed a network transformation heuristic to solve the DFRLM [52], in which for the first time the deviation paths were considered. However, both heuristics were designed for maximal flow coverage models which are distinct from the MPRLM and other flow-based set-covering problems in that the models minimize the total cost (e.g., the total number of stations) while satisfying travel demand. Thus, a new heuristic solution is in need.
2.3 Multi-Period Location Models

Multi-period or dynamic facility location problems are not particularly new, which have been extensively studied in the past few decades. However, similar to the static flow-based location models, the majority of the multi-period/dynamic models are node-based and there are two general categories models: location and location-relocation models. The first category assumes that once a facility is in service, it will not be relocated [80, 86]. This type of models is well suited for capital intense infrastructure planning, such as refineries [43, 43]. The second category allows for facility’s relocation after location [87], which is particular suitable for mobile facilities, such as ambulance [19] and public service facilities [37]. Interested readers can refer to recent survey papers [2, 60] for detailed reviews.

These dynamic node-based facility location models are not well suited for dynamic location of charging stations for the same reason as for the static counterparts. Compared to the static flow-based location models, the literature on dynamic flow-based location problems is scarce. To the best my knowledge, the multi-period location model in [23], perhaps, is the only dynamic flow-based location model. The model was extended from the FRLM, in which a pre-specified number of charging stations were sequentially placed on a freeway network in Korea for a finitely many time stages. The goal is to maximize the total traffic flow covered over time. In their problem, the topological structure of freeway network is given with a fixed set of O-D pairs while traffic flows are time-dependent and increase with the growth of EV market. Distinct from this study, I tackle a different problem. First, the proposed model will take into account the topological dynamics of network, in which the origins and destinations of the network will be undertaken a sequential expansion with more cities becoming EV markets. Second, rather than seeking a maximum-flow-coverage
solution with a given number of stations, the goal is to find a least-cost solution that can dynamically satisfy all O-D trips over time. Third, an integrated view is taken to incorporate deviation paths and limited vehicle range into the model and further allow location and relocation of charging stations for the sake of lower cost. From the modeling perspective, the multi-period model proposed in this dissertation belongs to the second category of location-relocation models.

2.4 Solution methods for multi-period location models

There exist various solution methods, both exact and heuristic, for solving multi-period location problems, which in many cases are modeled as mixed-integer programs. The initial attempt to overcome the computational difficulties, perhaps, involves a myopic approach, which consists of solving the first-period problem without taking into account future-period demands, and then solving the second-period problem given the optimum facility locations identified in the first-period problem, and so forth. [23] extended it to consider both forward-myopic and backward-myopic methods, in which they demonstrated the resulting solutions were suboptimal compared with the optimum solutions from the multistage optimization model solved by CPLEX. In this dissertation, I compare the proposed multi-period optimization model solutions to the myopic solutions to elaborate the effects of taking into account the future demands in a multistage model and temporal trade-offs in terms of investment and deferral. Other solution methods that solve the problem as a whole could be dynamic programming method, which is naturally suitable for the multi-period problems by taking the advantage of its adaptive solution process [6, 15, 32]. However,
implementation of this method can be problematic if each single-period subproblem is already difficult to solve (e.g., the set covering location problem). In addition, the branch-and-bound approach, though solving (mixed) integer programs, is often limited to handling small sized problems [22], especially when it comes to solving multi-period location problems.

It is not of a surprise to note that there is a vast pool of studies developing heuristic methods to solve multistage location problems while balancing the solution efficiency and quality. One of them is the Lagrangian relaxation based heuristic solutions, which have been widely used in the location and inventory planning problems, such as the integration with branch-and-bound method [25], coupling with heuristics and subgradient optimization method for obtaining lower bounds [40, 53], and solving for the dynamic facility location problems [20]. However, the success of implementing the Lagrangian relaxation method for a particular problem depends on several factors, such as the constraint(s) identified to be relaxed, the goodness of bounds, and the solution efficiency of the relaxed problem. These are something that I readily understood after I tried and failed to solve the model. There are other heuristics that have been proven effective, especially for real-world large-scale problems involving hundreds of nodes and arcs, such as the Tabu-search [50], the Simulated Annealing (SA) algorithm [1, 8] and Genetic Algorithm (GA) [4].

To a broader extent, new methods can be elucidated by the solution methods applied to solving inventory routing problems that are inherently multi-period [14]. In particular, a two-phase approach decomposed the set of decisions into a delivery schedule first, followed by the construction of a set of delivery routes. As relating to the multi-period charging location problems, the first phase can utilize integer programming to identify the locations to be placed over time, whereas the second phase solves a linear program to identify optimal routes between O-D pairs. This method
2.5 Probability-based location models

Most of the literature on facility location problems has been focused on situations where the model structure is deterministic. Contrasted to deterministic assumptions, travel time on road links may not be fixed, travel cost may include more than one attribute, demand rates at the nodes may vary stochastically, and the volume of demand may be so high that the facilities may be busy and unavailable to service [61]. Facility location decisions without considering those uncertainties of network characteristics may result in a waste of resources. Among the factors that influence the public facility location decisions, the most apparent one is the stochastic nature of demand. There are many studies focused on situations where the demand quantity is a random variable. [18] studied public facility location problem when the number of users at each demand node is a random variable with a known distribution. Given a predefined probability level, the model minimizes the upper bound of the distance traveled by users. [69] developed a model for locating cleanup capability to response to oil spills, in which the volume of spill at risk points has a known probability mass function defined over a finite set. [11] incorporated the relative likelihood of spill occurrence in a partial covering approach to site response resources such that the probability to cover a spill event is maximized. Interested readers may refer to [29, 73] for more detailed reviews on facility location under uncertainty.

Though all the aforementioned models incorporate uncertainty in facility location problems, all of models assume that demand is node based, which might not be appropriate for location problems where services are provided along the paths of traffic flows, i.e., hydrogen refueling stations and electric vehicle charging stations. In the
past decade, flow based refueling station location problem has received increasing attention in literature. First appeared as a flow intercepting or capturing location model (FILM/FCLM) [12, 41], flow based location models have been evolved to take into account more realistic considerations, such as vehicle range, in the flow-refueling location problem (FRLP)[54] and flow-based set-covering models [44, 82, 83, 84].

However, none of the aforementioned flow based models capture uncertainty when layout facilities. Taking the example of EVs, the market penetrate rate is still very low and whether a city would become an EV adopter is affected by many factors. It be a waste of resources if charging stations are placed to satisfy EV trips between two cities that are with very low chance to become EV market. Aiming to bridge this gap, in this study I present a probability-based flow refueling location model for EV charging stations. Based on the previous multi-path refueling location model (MPRLM) [44] probability of each city to become an EV adopter is integrated in optimizing station deployment. Predicted by considering demographic and economic data, the probability of each city is used to represent the priority of that city in the objective to maximize the total coverage of all the cities in a network with a given budget. This probability-based multi-path refueling location model (P-MPRLM) also differs from the MPRLM in ensuring round trips for covered O-D pairs. Same as in MPRLM, vehicle range and deviation behaviors of EV drivers are also taken into account in optimizing station deployment.
Chapter 3

Multi-path refueling location
model for electric vehicle charging stations

3.1 Problem statement

In this chapter, I developed a novel, EV charging station location model, called the
Multipath Refueling Location Model (MPRLM), in which EV users could utilize
multiple deviation paths between all O-D pairs on the network. An O-D pair is
considered as covered if there is at least one path, either a shortest path or a path
with reasonable deviation, available between the O-D pair through which drivers
can complete a trip with single/multiple refueling stops. The model minimizes the
total cost of locating charging stations while satisfying travel demands between all
O-D pairs, subject to limited vehicle travel ranges. The MPRLM provides integrated
decisions on strategic refueling station locations and the feasible routes between O-D
pairs under a single framework. Note that all stations in this study are assumed
as uncapacitated and the effects of traffic flows and congestions are not explicitly addressed in this model. These assumptions are more defensible in the early EV adoption phase. When EVs are massively adopted, explicit station capacity design will be necessary.

By incorporating deviation paths into a flow-based set-covering model, the MPRLM considers multiple deviation paths available between an O-D pair. It is believed that this relaxed assumption offers a greater flexibility in siting stations on network than the existing studies.

I implemented the MPRLM on two test networks the Sioux Falls road network [56] and a 25-node network [72]. These have been widely used as representations of real networks for numerical experiments in the transportation network design (the Sioux Falls network) and facility location problems (the 25-node network). I use these test networks to draw insights into the interplay between locations, deviations, and vehicle ranges. The model can be applied towards different problems by customizing it to meet the specific technological requirements, e.g., electric vehicle charging stations, battery swapping stations, compressed natural gas stations, hydrogen refueling stations. The most cost-effective vehicle travel range is also identified as an elbow point, at which the marginal reduction in the total cost of building charging stations drops. This work will enhance the efforts of private industry to increase its service coverage in a strategic manner and help government agencies plan subsidies to generate public interest in buying EVs before the market matures.

The remainder of the paper is organized as follows. I will present the concept, assumptions, and formulation of the multi-path station location model in section 3.2, with in-depth discussions on modeling and algorithms in generating multiple paths. In section 3.3, I will present the numerical results of the model on the two test networks and the results of sensitivity analyses. I will conclude the study in section
3.4.

3.2 Methods

In this MPRLM, charging stations may not be located exactly on a pre-planned path (e.g., a shortest distance or shortest travel time path), but they can still serve drivers if they are fairly close within a reasonable deviation limit. It is believed that drivers would accept a moderate deviation for refueling en route. The model integrates the considerations of multiple paths and the limited vehicle ranges into the decision making process for charging station locations, aiming to provide a least-cost solution while satisfying trips between all O-D pairs with reasonable deviations.

The critical model inputs are deviation paths between O-D pairs and vehicle travel ranges. Without available empirical data as to what extent travelers would deviate for refueling from their pre-planned paths (normally the paths with which the travelers are most familiar), I assume a set of deviation tolerances in terms of distance, mainly because the vehicle range is directly related to the distance traveled. Note that travelers deviation choices can also be dominated by other factors, such as travel time, travel cost, and network topology. However, there has been no well-defined relationship between deviation acceptance and different factors and most studies still rely on some explorative methods as pointed out by [51]. In this study, I assume that one of the shortest paths is treated as the pre-planned path between an O-D pair, and that deviation paths are either other shortest paths (if any) or paths that are slightly longer than the pre-planned path. Deviation paths are exogenously generated by algorithms described in section 2.3 and dependent on the selected deviation tolerances. These deviation paths will then be the model inputs. Note that the model is not designed to satisfy all deviation paths. Instead, as long as there is at least one path
between an O-D pair that can be completed via single or multi-stop refueling, this pair is considered as covered. The EV travel range is another important model input that restricts the maximum distance traveled by a vehicle before refueling.

### 3.2.1 A sample network

A 7-node network (Figure 3.1) is used to demonstrate the concept of the multipath refueling location problem. Assume that nodes A and E are origins, nodes C and G are destinations, and nodes B, D, and F are intermediate nodes. There are four O-D pairs, i.e., A-C, A-G, E-C, and E-G. The numbers on the links are link lengths and vehicle range is assumed to be 15. If only shortest paths are allowed between O-D pairs, three stations in total are needed at nodes B, D, and F, in which nodes B and F will respectively serve the O-D pairs A-C and E-G and node D will serve both the pairs A-G and E-C. If a 20% deviation from a shortest path is acceptable, drivers going from node A to C can accept the path $A \rightarrow D \rightarrow C$. Similarly, the path $E-D-G$ is now acceptable for trips from node E to G. Because of this relaxation, only node D is needed, which covers all O-D pairs. Paths between the O-D pairs of A-G and E-C remain unchanged. However, a drop in the deviation to 10% eliminates the deviation path available for both pairs A-C and E-G, resulting in a solution that is identical to that of considering the shortest paths. The above example illustrates how deviations can help reduce the total number of stations needed. The charging
station locations are results of vehicle range and choices of deviation tolerances. A higher deviation from the increased flexibility in path choice would reduce the number of stations needed on the network while increasing the total distance traveled. This trade-off will be explicitly discussed in section 3.3.

3.2.2 Model formulation

Before introducing the proposed MPRLM, definition of decision variables and parameters is presented.

Indices:

\( i \) :: index of candidate sites, \( i \in \hat{N} \subset N \)

\( r \) :: an origin node in the network, \( r \in R \subset N \)

\( s \) :: a destination node in the network, \( s \in S \subset N \)

\( k \) :: index of the paths for an O-D pair, \( k = 1, 2, \ldots, K \)

\( a \) :: index of arc set \( A \), \( a = (i,j) \in A \)

Parameters:

\( \omega_i \) : the installing cost of a charging station at node \( i \), \( i \in \hat{N} \)

\( \beta \) : onboard fuel capacity (unified in travel distance), i.e., vehicle range

\( M \) : a sufficiently large number

\( P_{rs,k} \) : a sequence of nodes on the \( k^{th} \) path from \( r \) to \( s \) and then back to \( r \) by the same path, where \( k = 1, 2, \ldots, K \)

\( d_{ij} \) : distance between node \( i \) and \( j \)

\( \delta_{i}^{rs,k} \) : =1 if node \( i \) is in the set of node \( P_{rs,k} \), 0 otherwise; this is an outcome of the deviation paths that are exogenously generated
Variables:

$X_i$: =1 if a charging station is located at node $i$; 0 otherwise

$Y_{rs,k}$: =1 if the $k^{th}$ path between $r$ and $s$ can be selected to be electrified; 0 otherwise

$B_{i}^{rs,k}$: remaining onboard power at node $i$ on the $k^{th}$ path of O-D pair $r-s$

$l_{i}^{rs,k}$: amount of power recharged at node $i$ on the $k^{th}$ path of O-D pair $r-s$

Following assumptions are made: (1) candidate locations for charging stations are predetermined; (2) vehicles are homogenous and fully fueled at origins; (3) charging stations are uncapacitated; (4) energy (e.g., fuel or electricity) consumed and refueled is unified in terms of travel distance; (5) all drivers are fully informed about charging stations locations on the network; and (6) vehicle range (e.g., miles) is predetermined and homogenous for all charging stations. A mixed integer linear programming (MILP) model is formulated. The complete model (P) is provided in (3.1) - (3.10):

$$\min \sum_{i \in N} \omega_i X_i$$ (3.1)
Subject to:

\[
B_{rs,k}^{rs,k} + l_{rs,k}^{rs,k} \leq M(1 - Y_{rs,k}) + \beta, \forall r, s; i \in P_{rs,k}; k = 1, 2, ..., K \tag{3.2}
\]

\[
B_{rs,k}^{rs,k} + l_{rs,k}^{rs,k} - d_{ij} - B_{j}^{rs,k} \leq M(1 - Y_{rs,k}), \forall r, s; i, j \in P_{rs,k}; (i, j) \in A; k = 1, 2, ..., K \tag{3.3}
\]

\[-(B_{i}^{rs,k} + l_{i}^{rs,k} - d_{ij} - B_{j}^{rs,k}) \leq M(1 - Y_{rs,k}), \forall r, s; i, j \in P_{rs,k}; (i, j) \in A; k = 1, 2, ..., K \tag{3.4}
\]

\[
\sum_{r, s} \sum_{k} l_{rs,k} \delta_{i}^{rs,k} \leq MX_i, \forall i \in N \tag{3.5}
\]

\[
\sum_{k} Y_{rs,k} \geq 1, \forall r, s \tag{3.6}
\]

\[
B_{i}^{rs,k} = \beta, \forall r, s; i \in R; k = 1, 2, ..., K \tag{3.7}
\]

\[
X_i = \{0, 1\}, \forall i \in \hat{N} \tag{3.8}
\]

\[
Y_{rs,k} = \{0, 1\}, \forall r, s; k = 1, 2, ..., K \tag{3.9}
\]

\[
B_{i}^{rs,k} \geq 0, l_{i}^{rs,k} \geq 0, \forall r, s; i \in P_{rs,k} \tag{3.10}
\]

The objective is to minimize the total cost of locating charging stations on the network. When \(\omega_i = 1, \forall i \in \hat{N}\), it minimizes the total number of stations, which is essentially a flow-based set-covering problem with path deviations. Constraint set (3.2) assures that the total onboard fuel will not exceed the EV fuel capacity \((B_{i}^{rs,k} + l_{i}^{rs,k} \leq \beta)\) on those paths \(k\) that are taken by the vehicle (i.e., \(Y_{rs,k} = 1\)); otherwise no restriction is applied (i.e., \(Y_{rs,k} = 0\)), simply because no traveler will use that route. Constraints (3.3) and (3.4) work simultaneously to ensure that the energy consumption conservation \(B_{i}^{rs,k} + l_{i}^{rs,k} - d_{ij} - B_{j}^{rs,k} = 0\) holds for all links on the \(k^{th}\) path if the path is taken \((Y_{rs,k} = 1)\) by any EV. Otherwise, when \(Y_{rs,k} = 0\), the inequality becomes \(B_{i}^{rs,k} + l_{i}^{rs,k} - d_{ij} - B_{j}^{rs,k} \leq M\), i.e., no restraining effects. Constraint set (3.5) is a logic constraint, stating that refueling is only available at
node i if charging stations are available. Constraint set (3.6) states that there is at least one path, either a shortest or deviation path, available between an O-D pair. Constraint set (3.7) follows the assumption that all EVs are fully refueled at origins (i.e., \( i \in R \)). Constraints (3.8) - (3.10) are binary and nonnegativity constraints. Note that there are several constraints includes big \( M \), and in practice I tried to use the smallest value for big \( M \). For constraints 3.2, 3.3 and 3.4, we can variable \( B_{rs,k}^i \) and \( l_{rs,k}^i \) are the remaining and recharged vehicle, therefore both should be less than the range associated with battery capacity. Thus \( 2 \beta \) can be used as an estimation of big \( M \). While for constraints 3.5, we can simply imagine that all paths passing through node \( i \) would be fully charged and then get an estimation for the value of big \( M \). In other words, the value for big \( M \) can be different for different instances when vehicle range is different.

Note that the proposed MPRLM generalizes the assumption of the shortest paths between O-D pairs in the refueling location models, which is equivalent to the case of \( K=1 \) in this model. It is also distinct from the DFRLM [51] in that multiple deviation paths are considered in constraint (3.6). Moreover, the model provides the routing choices indicated by \( Y_{rs,k} \).

**Remark 3.1:** The constraints (3.2)-(3.5) in the model (P) involving sufficiently large number may raise computational concerns. Mixed-integer programming solvers (such as CPLEX) proceed by optimizing continuous relaxations of the constraints, but for very big \( M \) values a relaxation will yield very tiny values of the binary variable (i.e., in this model) that do not provide useful information about the effects of the variables on the left-hand side of an inequality. Generally, the presence of one group of coefficients that are larger by many orders of magnitude than any of the others is known to have potentially bad numerical effects of the all aspects of the solution process.

Now I convert problem (P) to an equivalent optimization problem (P1) that
will yield the same optimal value of the objective function by replacing the inequality (3.2) with a more restrictive set of constraints (3.11) and (3.12),

\[
B^r_{rs,k} + l^r_{rs,k} \leq \beta, \forall r, s; i \in P^{rs,k}; k = 1, 2, ..., K
\]  
\[
l^r_{rs,k} \leq \beta Y^{rs,k}, \forall r, s; i \in P^{rs,k}; k = 1, 2, ..., K
\]

Constraint set (3.11) assures that the total onboard energy will not exceed the capacity on all possible paths between an O-D pair. Constraint set (3.12) is a logic constraint, stating that the energy fueled \(l^r_{rs,k}\) has to be within the capacity for those paths taken (i.e., \(Y^{rs,k} = 1\)); otherwise it is zero (\(Y^{rs,k} = 0\)). With substitution of constraint sets (3.11) and (3.12) with constraint set (3.2), I define a new model (P1) as:

\[
\min \sum_{i \in N} \omega_i X_i
\]
Subject to:

\[ B_{rs,k}^{s,k} + l_{rs,k}^{s,k} \leq \beta, \forall r, s; i \in P^{rs,k}; k = 1, 2, ..., K \]  
(3.14)

\[ l_{rs,k}^{s,k} \leq \beta Y_{rs,k}, \forall r, s; i \in P^{rs,k}; k = 1, 2, ..., K \]  
(3.15)

\[ B_{rs,k}^{s,k} + l_{rs,k}^{s,k} - d_{ij} - B_{j}^{rs,k} \leq M (1 - Y_{rs,k}), \forall r, s; i, j \in P^{rs,k}; (i, j) \in A; k = 1, 2, ..., K \]  
(3.16)

\[-(B_{i}^{rs,k} + l_{i}^{rs,k} - d_{ij} - B_{j}^{rs,k}) \leq M (1 - Y_{rs,k}), \forall r, s; i, j \in P^{rs,k}; (i, j) \in A; k = 1, 2, ..., K \]  
(3.17)

\[ \sum \sum \sum l_{rs,k}^{s,k} \delta_{i}^{rs,k} \leq MX_{i}, \forall i \in N \]  
(3.18)

\[ \sum \sum Y_{rs,k} \geq 1, \forall r, s \]  
(3.19)

\[ B_{rs,k}^{s,k} = \beta, \forall r, s; i \in R; k = 1, 2, ..., K \]  
(3.20)

\[ X_{i} = \{0, 1\}, \forall i \in \hat{N} \]  
(3.21)

\[ Y_{rs,k} = \{0, 1\}, \forall r, s; k = 1, 2, ..., K \]  
(3.22)

\[ B_{rs,k}^{s,k} \geq 0, l_{rs,k}^{s,k} \geq 0, \forall r, s; i \in P^{rs,k} \]  
(3.23)

The model (P1) can significantly improve the solution efficiency, compared to the model (P) for two major reasons. First, it eliminates the big number \( M \) in constraint set (3.2). As discussed in Remark 1, the big \( M \) results in computational and numerical problems. Secondly, it improves the solution efficiency. In constraint set (3.14) if \( Y_{rs,k} = 0 \), then \( l_{rs,k}^{s,k} = 0 \) and \( B_{rs,k}^{s,k} \leq \beta \), which reduces the number of variables and constraints from the model (P) and essentially eliminates the big \( M \) in constraints (3.16) and (3.17) as well. I show the numerical comparisons between these two models in section 3.1. Constraint set (3.18) is a standard logic constraint.
and a similar transformation is not easy to derive. For numerical implementations, a least sufficiently large value will be selected. For example, it is possible to set the value equal to the total number of paths $\times \beta$ for the uncapacitated location model or to the charging station capacity for a capacitated model.

**Proposition 3.1:** Problem (P1) yields the same optimal value of the objective function as in Problem (P).

**Proof:** Let $Z^*$ and $Z_1^*$ respectively denote the optimal objective values of problems (P) and (P1). Since the constraint set in problem (P1) are more restrictive, then $Z^* \leq Z_1^*$. On the other hand, suppose $X^*$ and $Y^*$ are solutions of problem (P). It is then clear that for the $k^{th}$ path between O-D pair $r$ and $s$ such that $Y^* = 1$, constraint (3.2) is equivalent to constraints (3.11) and (3.12). For the $k^{th}$ path between O-D pair $r$ and $s$ such that $Y^* = 0$, because $B_{i}^{rs,k}$ and $l_{i}^{rs,k}$ are practically not bounded by constraints (3.2)-(3.4) due to a fairly large number $M$, I can always find $B_{i}^{rs,k}$ and $l_{i}^{rs,k}$ for any node $i \in P_{rs,k}$ such that both constraints (3.11) and (3.12) are satisfied, and the adoption of the new constraints does not change the feasibility of $X^*$ and $Y^*$. Therefore, $X^*$ and $Y^*$ are still feasible so that $Z_1^* \leq Z^*$. Overall, $Z_1^* = Z^*$, and $X^*$ and $Y^*$ is also an optimal solution for the problem (P1). ⊙

**Remark 3.2:** Problem P1 can be solved much more efficiently than the original problem. In the case where the weights $w$ equal one for all locations, the model leads to the minimum number of charging stations locations needed for the network. In another case when the weights $w$ represent installing and operation costs of charging stations and the values may be differentiated with different locations, the model results in the minimum total cost and the resulting total number of stations will be at least as many as the minimum number of stations of the problem when $w=1$. However, in either case, the station locations may not be unique, nor are the paths traversed between O-D pairs. More discussions on numerical experiments are presented in section 3.3.
3.2.3 Notes on the algorithms generating multiple paths

The multiple paths between an O-D pair reflect drivers tolerance in deviation from the shortest path and the deviation paths could be the $2^{nd}$, $3^{rd}$, $k^{th}$ shortest paths. There are two types of K-shortest path problems: one allows loops between node pairs in a network and the other does not allow any loops, which is also called the K shortest loopless paths. The first type is easier to implement with existing algorithms such as the N-path method (Hoffman and Pavley 1959). The second type is more challenging due to the additional loopless constraint that no repeated nodes are allowed on a path. First addressed by Yens algorithm [89], it has been further developed in other studies [59, 70]. In this study, I adopt the loopless Yens algorithm for two reasons. First, because a transportation network is a network without negative travel time or distance links, looped trips will not reduce travel cost. Second, for flow-based facility location problems including the refueling station location problem in this paper, specific looped trips to stations are avoided because the goods or services are modeled to be obtained on the way.

However, because this algorithm only specifies the number of alternative paths between an O-D pair, in some cases, the deviations from a shortest path can be too large to be realistic in some cases. I develop the K-shortest path with deviation cap algorithm by imposing a cap that restrains the alternative path within a predefined deviation limit, e.g., 10%, 20%, and 50%. The algorithm can be described as follows:

For each O-D pair $(r, s)$:

Step 1. Find the shortest path using any efficient shortest path algorithm, such as Dijkstra's algorithm [30] for $(r, s)$ (the length denoted as $L^{rs,1}$); set
Step 2. Compare the length of the $k^{th}$ path (denoted as $L^{rs,k}$) with $(1+p)L^{rs,1}$, where $p$ is a predefined deviation cap. If $L^{rs,k} \geq (1+p)L^{rs,1}$, then stop; otherwise, set $k = k + 1$, and go to step 3;

Step 3. Find the $k^{th}$ shortest path using Yens algorithm; Go to step 2.

Note that if the parameter $K$ in the algorithm is sufficiently large, the algorithm will find all deviation paths within the predefined deviation cap. Depending on the network structure, the number of deviation paths between an O-D pair may vary. In some cases (especially large-scale networks), I may need to use a finite $K$ and deviation cap jointly to restrict the number of deviation paths between O-D pairs.

### 3.3 Results and discussions

To validate the model and demonstrate its applicability, I implement the model on the two well-known test networks: the Sioux-Falls roadway network and the 25-node network shown in Figure 3.2. The numbers in the circles represent the node indices. The numbers on the links denote the test distances in miles or kilometers.

These two networks have different topological structures. In particular, the Sioux-Falls network is a closed-loop network, in which every node is interconnected with at least two other nodes, while in the 25-node network node #25 is only connected to #24, which is a tail that makes the network less flexible in station deployment. In the numerical study, I assume that all nodes are candidate sites for charging stations, i.e., $\hat{N} = N$ and I treat every candidate site equally costly, i.e., $\omega_i = 1, \forall i \in \hat{N}$. Every node on the network is an origin and a destination, i.e.,
Solving the model (P1) involves two major steps: (1) preparing the multipath sets for all O-D pairs as model inputs by using the K-shortest path algorithm and the K-shortest path with deviation cap algorithm, which are both implemented in MATLAB; and (2) programming the model in AMPL [34] and solving it by using a commercial solver CPLEX 12.4. All numerical experiments described run on a DELL desktop with 8 GB RAM and Intel Core i5-2500@3.30GHz processor under Windows 7 environment. Depending on the deviation tolerance, the number of decision variable and number of constraints may be expanding exponentially. Thus, solving time may vary dramatically.
3.3.1 The Sioux Falls network

The vehicle range is set to be 100 the length of the longest link on the network. Three different deviation scenarios—the shortest path ($K = 1$), 3-shortest paths ($K = 3$), and 20% deviation cap are considered to illustrate how different deviations affect the station siting strategies. Optimal charging station locations are represented by solid nodes in Figure 3.3.

Results indicate that deviations help reduce the number of stations required. In particular, the minimum numbers of stations needed following the solutions of shortest path ($K = 1$), $K = 3$ and 20% deviation cap are respectively 12, 7, and 11. The paths traversed between O-D pairs can be identified by revealing the decision variables $Y_{rs,k}$, which are determined simultaneously with the station location decisions. For example, in the $K = 1$ solution, path #1 $\rightarrow$ 2 $\rightarrow$ 6, highlighted in Figure 3.3a, is traversed between nodes #1 and #6 and EVs on that path have to stop at node #2 for refueling. When $K = 3$, two additional paths #1 $\rightarrow$ 3 $\rightarrow$ 4 $\rightarrow$ 5 $\rightarrow$ 6 and #1 $\rightarrow$ 3 $\rightarrow$ 12 $\rightarrow$ 11 $\rightarrow$ 4 $\rightarrow$ 5 $\rightarrow$ 6 become acceptable, which are the second and third shortest paths between that O-D pair. These two deviation paths shown in Figure 3.3b help eliminate charging stations at node #2. In particular, the path #1 $\rightarrow$ 3 $\rightarrow$ 4 $\rightarrow$ 5 $\rightarrow$ 6 is completed via refueling at node #5 and the path #1 $\rightarrow$ 3 $\rightarrow$ 12 $\rightarrow$ 11 $\rightarrow$ 4 $\rightarrow$ 5 $\rightarrow$ 6 is completed via multi-stop refueling at nodes #12, 11, and 5. The solution of the 20% deviation cap shows a similar location pattern as the solution of $K = 1$. This is because only very few paths other than the shortest are within a 20% deviation cap. For example, there is no deviation path within 20% cap between the O-D pair #1 to #6 and thus only the shortest path is taken. The results in Figure 3.3 also imply that new station siting plans with deviations are not simple processes of removing stations from the $K = 1$ solution, which
instead require re-optimizations of the entire network.

Figure 3.3: Location of charging stations on the Sioux Falls network with vehicle range of 100 under three deviation scenarios

The number of paths served by different stations (i.e., the number of paths passing through a particular station) under three deviations are provided in Table 3.1 to demonstrate the variability of the workload of each station (measured by the number of paths served). It is noted that even for the same location (e.g., the station at node #5), the number of paths served varies with deviation scenarios. Note also that the total number of paths served can be higher or lower than the 552 O-D pairs, i.e., the total number of O-D pairs on the Sioux Falls network. This could be due to duplicating counts of a path that goes through multiple stations, (more than 552) or multiple O-D pairs served by the same paths (less than 552). From this table, it is clear that the nodes #5, #8, #11, and #15 are critical locations as they are used in all three deviation scenarios.

As noted in Remark 3.2, the model yields the minimum number of stations but non-unique locations for a given vehicle range and a deviation. Figure 3.4b
Table 3.1: Number of paths served by different locations

<table>
<thead>
<tr>
<th>Stations (node ID)</th>
<th>Number of paths served by a refueling station</th>
<th>K=1</th>
<th>K=3</th>
<th>20% deviation cap</th>
</tr>
</thead>
<tbody>
<tr>
<td>#2</td>
<td></td>
<td>16</td>
<td>-</td>
<td>12</td>
</tr>
<tr>
<td>#3</td>
<td></td>
<td>58</td>
<td>-</td>
<td>63</td>
</tr>
<tr>
<td>#5</td>
<td></td>
<td>69</td>
<td>100</td>
<td>59</td>
</tr>
<tr>
<td>#8</td>
<td></td>
<td>90</td>
<td>93</td>
<td>84</td>
</tr>
<tr>
<td>#10</td>
<td></td>
<td>-</td>
<td>-</td>
<td>78</td>
</tr>
<tr>
<td>#11</td>
<td></td>
<td>54</td>
<td>45</td>
<td>68</td>
</tr>
<tr>
<td>#12</td>
<td></td>
<td>50</td>
<td>55</td>
<td>-</td>
</tr>
<tr>
<td>#13</td>
<td></td>
<td>-</td>
<td>-</td>
<td>42</td>
</tr>
<tr>
<td>#15</td>
<td></td>
<td>67</td>
<td>88</td>
<td>70</td>
</tr>
<tr>
<td>#16</td>
<td></td>
<td>91</td>
<td>-</td>
<td>91</td>
</tr>
<tr>
<td>#18</td>
<td></td>
<td>-</td>
<td>83</td>
<td>-</td>
</tr>
<tr>
<td>#20</td>
<td></td>
<td>38</td>
<td>-</td>
<td>47</td>
</tr>
<tr>
<td>#21</td>
<td></td>
<td>-</td>
<td>63</td>
<td>-</td>
</tr>
<tr>
<td>#23</td>
<td></td>
<td>13</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>#24</td>
<td></td>
<td>52</td>
<td>-</td>
<td>54</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>612</td>
<td>527</td>
<td>668</td>
</tr>
</tbody>
</table>

demonstrates another location solution for the same minimum 12 stations when \( K = 1 \) with a vehicle range of 100. This non-uniqueness in location solutions also helps explain why charging stations locations can alter when node weight \( \omega \) is differentiated. For example, if the weight of node #3 is set high at 10 (e.g., a high initial installation cost of a charging station in reality) while the weights of other nodes remain, a new location solution is found as shown in Figure 3.4c. The total number of stations increases to 13 and the node #3 is not selected due to the high cost. It implies that the node #3 can be replaced with other combinations of nodes on the network.

I further analyze the coupled effects of deviations and vehicle ranges on the number of charging stations sited. In particular, I test five different deviation caps at 0%, 10%, 15%, 20%, and 50% coupled with seven different vehicle ranges between 100 and 250 with a 25 interval, totaling 35 tests. The results, shown in Figure 5, indicate that the minimum number of stations needed declines with longer vehicle ranges,
which in turn makes deviations less appealing. For example, when the vehicle range is up to 225, only one station is needed and no user needs to deviate their routing. It drops to zero when the vehicle range reaches 250. I also analyze the effects of the \(K\) values in the same way and the results plotted in Figure 3.6 show similar conclusions.

Empirical vehicle ranges can be lower than the theoretical (or anticipated) values, due to a number of prevailing factors, such as traffic conditions, weather, and the variability in users anxiety to refuel. I conduct an analysis to understand the impacts of variations in vehicle ranges caused by these factors on the minimum number of stations needed. In particular, I assume a vehicle range of 200 and two levels of deductions, 25% and 50%, with resulting vehicle ranges of 150 and 100. The results plotted in Figure 3.7 show that the numbers of stations increases significantly from one station to as many as 12 stations with the deductions, varied by deviations. It is noted that a larger deviation (e.g., \(K = 3\)) helps reduce the sensitivity in siting strategies than a lower deviation (e.g., shortest path). In particular, following \(K = 3\)
deviation, there are five more stations needed when vehicle range is reduced by 50% while it requires 10 more stations if the shortest paths are considered.

Numerical experiments are conducted for comparing the computational performances by models (P) and (P1) under different deviation scenarios, the results of which are reported in Table 3.2. All these numerical experiments are solved by CPLEX to optimality (i.e., 0% gap). Here, all the solutions are in terms of number of stations and all computation times are in CPU seconds. From the results, higher deviations result in longer computing times due to the increased number of variables and constraints. Under all three deviations, both (P) and (P1) models result in identical optimal objective values (i.e., the same minimum number of stations) while model (P1) significantly reduces computing times than model (P) under all deviation scenarios.
Figure 3.6: Effects of K and vehicle ranges on the minimum numbers of charging stations needed on the Sioux Falls network

Table 3.2: Comparisons of computational performances of models (P) and (P1)

<table>
<thead>
<tr>
<th>Models</th>
<th>K=1 Solution</th>
<th>Time (s)</th>
<th>K=3 Solution</th>
<th>Time (s)</th>
<th>20% deviation Solution</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>12</td>
<td>0.2</td>
<td>7</td>
<td>476</td>
<td>11</td>
<td>7</td>
</tr>
<tr>
<td>P1</td>
<td>12</td>
<td>0.1</td>
<td>7</td>
<td>132</td>
<td>11</td>
<td>3</td>
</tr>
</tbody>
</table>

3.3.2 A 25-node network

I implement the model on a 25-node network by using the same three deviation scenarios. The optimal locations are represented by the solid nodes in Figure 3.8. With vehicle range of 10, the minimum of numbers of stations are 10, 8, and 9 for deviation scenarios K=1, K=3 and 20% deviation cap, respectively. Compared to the Sioux Falls network, the 25-node network shows lower variations in the station siting strategies, mainly because the link lengths are less varied and the tail of the network (i.e., the nodes #22, 23, 24, and 25) requires stations to be located at nodes #14 and #24 regardless.

Here I take the O-D pair from nodes #1 to #10 as an example to highlight the differences in the used paths under different deviation scenarios. The $K = 1$ solution
Figure 3.7: Effects of the vehicle range deduction on the minimum numbers of charging stations needed on the Sioux Falls network is illustrated in Figure 3.8a. In the $K = 3$ solution, this O-D pair is covered via the third shortest path with multi-stop refueling at nodes #4 and #8 as shown in Figure 3.8b. With the 20% deviation cap, both the shortest path $#1 \to 2 \to 3 \to 9 \to 10$ and the fourth shortest path $#1 \to 5 \to 7 \to 8 \to 10$ is used to cover the O-D pair as shown in Figure 3.8c.

The effects of deviations coupled with vehicle ranges on the station deployment on the 25-node network are also analyzed with the results plotted in Figures 3.9 and 10. Similar observations are made as for the Sioux Falls network. Particularly, an identical number of stations is used when vehicle range is 10 when deviation caps are relatively low (i.e., 10%, 15%, and 20%). A further investigation reveals that only the shortest paths are used in these deviation scenarios. It is also identified that the elbow point is at 20, implying that a vehicle range of 20 may be the most cost effective in terms of the total cost of building stations. The marginal savings in the
Figure 3.8: Locations of charging stations on the 25-node network with a vehicle range of 10 under three deviation scenarios

total cost (in this case, equivalent to the total number of stations) reduces when the vehicle range is higher than 20.

I conduct a similar analysis on the effects of deducted vehicle ranges on the 25-node network, with the results shown in Fig. 11. These results are based on the vehicle range of 40 and three levels of deductions, i.e., 25%, 50%, and 75%, with respectively resulting vehicle ranges of 30, 20, and 10. Similar to what I have observed for the Sioux Falls network, the deductions in vehicle range require more stations. There is, however, less variability among different deviation scenarios ($K = 1$, $K = 3$, and 20% deviations), due to the lower variability in both the link lengths and the topological structure.

3.3.3 Sensitivity analysis on user inconvenience due to deviations

In general, a reduced number of charging stations deployed on the network increase the average travel distance. I further explore the tradeoffs between the cost of locating
stations (in terms of number of stations) and user convenience (in terms of travel distance) with deviations. The results would provide managerial insights to planners.

As noted in Remark 3.2, the model (P) sets the objective as to minimize the number of stations, which may not yield unique station locations. For each deviation scenario, the resulting average deviation, defined as the average additional travel distance for the entire network due to deviations, may not be unique. In this analysis, I reformulate the problem to minimize the total deviation in the objective function (13), given the number of stations (i.e., the resulting minimum number of stations from the model (P)) in constraint (14). The new model is denoted as model (P-dev). The average deviation equals the total deviation (i.e., objective value) divided by the total number of paths (including all deviation paths) $\sum_k Y_{rs,k}$, which is unique with respect to the given number of stations.

$$\min \sum_{rs,k} \frac{c_{rs,k} - c_{rs,1}}{c_{rs,1}} Y_{rs,k}$$

(3.25)
Figure 3.10: Effects of $K$ and the vehicle ranges on the minimum numbers of charging stations needed on 25-node network (note: $K = 4$ and $K = 5$ lines are completely overlapped)

Subject to:

$$\sum_i X_i = T \quad (3.26)$$

Including (3.3)-(3.12)

Where

$c_{rs,k}^r$: a parameter, the length of the $k^{th}$ shortest path connecting $r$ and $s$,

$T$: a parameter, the minimum number of stations resulted from model (P)

All other notations are the same as in the model (P).

The results are summarized in Table 3.3, in which zeroes indicate that either no deviation (i.e., $K = 1$) or deviation paths with an identical minimum distance (in deviation cap scenarios of 10%, 15%, and 20%). The resulting station locations are displayed in the column with heading Station location #. In general, a relaxed deviation scenario (e.g., higher $K$ or deviation cap) helps reduce the number of stations
Figure 3.11: Effects of the vehicle range deduction on the minimum numbers of charging stations needed on the 25-node network

Table 3.3: The average deviation and charging station locations under different deviation scenarios

<table>
<thead>
<tr>
<th>Deviation scenarios</th>
<th>Sioux Falls network (vehicle range of 100)</th>
<th>25-node network (vehicle range of 10)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Min avg. Deviation</td>
<td>Station locations #</td>
</tr>
<tr>
<td>K=1</td>
<td>0</td>
<td>2,3,5,8,9,11,12,15,16</td>
</tr>
<tr>
<td>K=2</td>
<td>0.50%</td>
<td>2,5,8,11,12,15,16,20,24</td>
</tr>
<tr>
<td>K=3</td>
<td>1.95%</td>
<td>5,8,11,12,15,18,21</td>
</tr>
<tr>
<td>K=4</td>
<td>1.94%</td>
<td>5,8,11,12,15,18,21</td>
</tr>
<tr>
<td>K=5</td>
<td>3.33%</td>
<td>4,6,12,15,16,21</td>
</tr>
<tr>
<td>10%</td>
<td>0</td>
<td>2,3,5,8,10,11,13,15,16,20,24</td>
</tr>
<tr>
<td>15%</td>
<td>0</td>
<td>2,3,5,8,9,11,13,15,16,20,24</td>
</tr>
<tr>
<td>20%</td>
<td>0</td>
<td>1,2,5,8,9,11,12,15,16,20,24</td>
</tr>
<tr>
<td>50%</td>
<td>1.94%</td>
<td>5,8,11,12,15,16,21</td>
</tr>
</tbody>
</table>

with only a slight increase in the average deviation. This finding is most interesting to planners, as it implies that a high level of service to EV users can be maintained with a low cost through smart planning. For example, for deviation scenarios from $K = 1$ to $K = 5$, on the Sioux Falls network, half number of stations has been saved (from 12 to 6) but this 50% reduction only increases the average travel distance by about 3.3%. Similarly, on the 25-node network, 30% ($= (10 \cdot 7) / 10$) reduction in the number of stations only causes the average travel distance increased by about 2%.

I also notice that for some cases when the same numbers of stations are sited
on the network, a higher deviation helps reduce the average deviation on the network. For example, on the 25-node network, the average deviation with \( K=3 \) drops slightly by about 0.1\% (=1.42\% - 1.3\%) from \( K = 2 \). This can be explained as following. First, the model (P-dev) minimizes the total deviation on the network, given a fixed total number of charging stations. When the total number of charging stations is identical, the average deviation in a higher deviation scenario should be as least as good as the one in a lower deviation scenario. Second, a higher deviation permits the selection of more deviated paths for some O-D pairs while let other O-D pairs be traversed via shortest paths. The average deviation for the entire network is a result of the trade-offs. As seen from Tables 5 and 6, which report the numbers of shortest and deviation paths used on both networks under different deviation scenarios, the number of shortest paths increases by five (= 483 - 478) from \( K = 3 \) to \( K = 2 \) and the number of deviation paths drops by five (= 122-117). As a result, the overall average deviation is reduced. Similar explanations apply to the comparisons between \( K = 5 \) and \( K = 4 \).

Tables 3.4 and 3.5 also indicate that charging stations locations are primarily determined based on the use of shortest paths, which implies that a small number of deviation paths can substantially reduce the number of charging stations needed.

**Table 3.4:** Numbers of paths used in the Sioux Falls network under different deviation scenarios

<table>
<thead>
<tr>
<th>Deviation scenarios</th>
<th>Total number of used paths ( \sum_{r,x,k} Y_{r,x,k} )</th>
<th>Total number of used shortest paths ( \sum_{r,x,k} Y_{r,x,1} )</th>
<th>Total number of used deviation paths ( \sum_{r,x,k} Y_{r,x,k} - \sum_{r,x,k} Y_{r,x,1} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>K=1</td>
<td>552</td>
<td>552</td>
<td>0</td>
</tr>
<tr>
<td>K=2</td>
<td>552</td>
<td>524</td>
<td>28</td>
</tr>
<tr>
<td>K=3</td>
<td>552</td>
<td>478</td>
<td>74</td>
</tr>
<tr>
<td>K=4</td>
<td>554</td>
<td>479</td>
<td>75</td>
</tr>
<tr>
<td>K=5</td>
<td>553</td>
<td>432</td>
<td>121</td>
</tr>
<tr>
<td>10%</td>
<td>553</td>
<td>535</td>
<td>18</td>
</tr>
<tr>
<td>15%</td>
<td>552</td>
<td>540</td>
<td>12</td>
</tr>
<tr>
<td>20%</td>
<td>552</td>
<td>530</td>
<td>22</td>
</tr>
<tr>
<td>50%</td>
<td>552</td>
<td>470</td>
<td>82</td>
</tr>
</tbody>
</table>
Table 3.5: Numbers of paths used in 25-node network under different deviation scenarios

<table>
<thead>
<tr>
<th>Deviation scenarios</th>
<th>Total number of used paths</th>
<th>Total number of used shortest paths</th>
<th>Total number of used deviation paths</th>
</tr>
</thead>
<tbody>
<tr>
<td>K=1</td>
<td>600</td>
<td>600</td>
<td>0</td>
</tr>
<tr>
<td>K=2</td>
<td>600</td>
<td>498</td>
<td>122</td>
</tr>
<tr>
<td>K=3</td>
<td>600</td>
<td>141</td>
<td>117</td>
</tr>
<tr>
<td>K=4</td>
<td>600</td>
<td>140</td>
<td>135</td>
</tr>
<tr>
<td>K=5</td>
<td>600</td>
<td>141</td>
<td>121</td>
</tr>
<tr>
<td>10%</td>
<td>601</td>
<td>535</td>
<td>66</td>
</tr>
<tr>
<td>15%</td>
<td>600</td>
<td>536</td>
<td>64</td>
</tr>
<tr>
<td>20%</td>
<td>602</td>
<td>537</td>
<td>65</td>
</tr>
<tr>
<td>50%</td>
<td>600</td>
<td>414</td>
<td>186</td>
</tr>
</tbody>
</table>

3.4 Summary

In Chapter 3, I developed the multi-path refueling station location model (MPRLM) to seek the most cost-effective charging stations location strategies on the networks, considering limited vehicle ranges and allowing for multiple deviation paths between O-D pairs. The MPRLM minimizes the total cost of establishing new refueling stations on transportation networks while satisfying demand among all O-D pairs. This model integrates deviation paths into a flow-based set-covering problem. I implemented the model on two well-known test networks - the Sioux Falls and 25-node networks. The results indicate that the decisions on charging stations locations and paths traversed between O-D pairs are interdependent and thus should be determined simultaneously. The use of deviation paths can substantially reduce the total cost of establishing charging stations or the minimum number of stations with a reasonable compromise of users convenience. Charging stations are primarily located based on the use of shortest paths while a small number of deviation paths have substantially reduced the system cost. An elbow point rule is also used to identify the most cost effective vehicle range in terms of total cost of building charging stations.
Chapter 4

Heuristic approaches for the multi-path refueling location model

4.1 Problem statement

In this study, I am primarily concerned with developing greedy heuristic approximation solutions, particularly the greedy-adding and greedy-adding with extension (combined processes of pre-selection, substitution, and solution refining) algorithms. The research efforts were concerted on designing effective node weighting schemes, considering the effects of vehicle range and multiple deviation paths on transport networks. Distinct from the heuristics for maximal coverage problems, the proposed heuristic solutions will iterate until all the O-D pairs on the network are covered. I implement the heuristics on two networks: the Sioux Falls network [56] which is a well-regarded test network in transportation network modeling society, and a real-life highway network based on the state of South Carolina. Compared with the exact solutions, the proposed heuristic solutions have been demonstrated for quality solutions with substantially reduced solving times.
The remainder of this chapter is organized as follows. In section 4.2 I propose the greedy-adding and greedy-adding with extension algorithms. The results of numerical implementations of the heuristics on the two networks are presented in section 4.3. Then a summary of this chapter is provided in in section 4.4.

4.2 Methods

I present two heuristic methods for the MPRLM, which are greedy-adding (GA) and GA with extensions (GA-E) algorithms. The extensions combine the pre-selection, substitution, and refining processes. In this section, I first review prior studies on heuristics for general set covering problems, followed by the discussions on why new heuristic approaches are needed to solve the MPRLM in section 3.2. The GA and GA-E algorithms will be respectively presented in sections 4.2.1 and 4.2.2.

4.2.1 Greedy-adding algorithm for MPRLM

It is important to first stress that the greedy-adding (GA) algorithm, as aforementioned, is a myopic method, in the sense that it does the best for improving the solution for now but may result in suboptimal solution in a long run.

The GA heuristic developed in this paper is distinct from previous heuristics [52, 58, 91] for their different node weighting schemes and stopping criteria. In general, the proposed GA picks a node if that node can cover more uncovered O-D pairs than any other nodes in the set, given deviation paths and a vehicle range. It iterates until all O-D pairs are covered, which results in the minimum number of stations needed. With varied drivers refueling behaviors, I further assume that i) drivers are fully knowledgeable about locations of the charging stations (if existing) on the network and familiar with transportation network; ii) drivers behave consistently and
aggressively, that is, they will not refuel until they cannot reach the next station and when refueling, they will fuel a full tank; and iii) deviation paths are indifferent to drivers. These assumptions result in a minimum set of nodes to cover an O-D pair and eliminate possible supersets. However, for all O-D pairs, a mere union or merger of all the node sets will likely result in inferior solutions with redundant nodes, and thus a refining procedure is needed to possibly further improve the solution quality. The refining procedure is part of the GA-E which is discussed in details in section 3.4. Note that these assumptions may also eliminate other possible node combinations which might result in better solution quality in a long run. However, enumeration of all possible node combinations is impossible even for a median sized problem, as noted in [54]. The procedures of the GA algorithm are described as follows:

Step 0. Set the selected set $V$ empty;

Step 1. Check if all O-D pairs can be covered by the set $V$ given the vehicle range.

If yes, stop and it is the final solution; otherwise, proceed to step 2;

Step 2. Calculate weight for all the nodes

2.1. set $Temp = \emptyset$

2.2. for each uncovered O-D pair $m$, $m = 1, 2, ..., M$, where $M \leq |R \times S|$

On each path $k$, $k = 1, 2, ..., K$ that connects the O-D pair $m$, add a node $(i)$ to the set $Temp$ if $B_{i}^{rs,k} - d_{i,i+1} < 0$ and then set $B_{i}^{rs,k} = \beta$. Repeat it until the destination is reached. The weight of each node $i$ on the path $k$ for O-D pair $m$ is $w_{i}^{k,m} = 1/ |Temp|$, and the total weight of node $i$ is $w_{i} = \sum_{m=1}^{M} \sum_{k=1}^{K} w_{i}^{k,m}$.
Step 3. add the one with the highest weight (in the case of multiple nodes having equal weight, arbitrarily pick one) to the set $V$ and then go back to step 1.

**Remark 3.1**: The total solving time of the GA equals the solving time in each iteration multiplied by the number of iterations. The solving time in each iteration is polynomial to the number of O-D pairs and the number deviation paths considered for each O-D pair.

I use a seven-node sample network as shown in Figure 4.1 to elaborate the node weighting scheme and GA procedures. Starting with the set $V$ being empty in step 1, except O-D pairs (1, 3), (1, 6) and (1, 7), all other pairs can be covered without refueling, given the vehicle range of 15. Then, I proceed to step 2 to add best nodes to the set $V$ through iterations to cover the three O-D pairs. I discuss the uses of shortest paths and deviation paths separately.

**Shortest path scenario**: in iteration #1, the paths $1 \rightarrow 2 \rightarrow 3$, $1 \rightarrow 5 \rightarrow 6$, and $1 \rightarrow 5 \rightarrow 6 \rightarrow 7$ are the shortest paths for the O-D pairs (1, 3), (1, 6), and (1, 7), respectively. On the path $1 \rightarrow 2 \rightarrow 3$, the node #2 will cover the O-D pair (1, 3); thus node #2 weighs 1. Similarly, on the path $1 \rightarrow 5 \rightarrow 6$, the node #5 covers the O-D pair (1, 6); thus weighs 1. On the path $1 \rightarrow 5 \rightarrow 6 \rightarrow 7$, both nodes #5 and 6 need to be used. This O-D coverage will be split between the two nodes; thus each weighs 0.5. Only nodes #2, 5, and 6 have positive coverages with 1, 1.5, and 0.5, respectively as summarized in Table 2. The node #5 with the highest coverage is thus selected and added to the set $V$. By the end of the iteration, $V = \{5\}$ and the pair (1, 6) is covered.

At beginning of iteration #2, only the pairs (1, 3) and (1, 7) are uncovered. Following the same procedures, both nodes #2 and #6 weigh 1 in Table 4.2. In this
Figure 4.1: A seven-node sample network (the circled numbers represent node indices, numbers on the links denote the lengths, and all links are bidirectional.)

Table 4.1: Node weights in iteration #1 with shortest path scenario

<table>
<thead>
<tr>
<th>Uncovered O-D pairs</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1, 3)</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(1, 6)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(1, 7)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.5</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td>Node weights</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1.5</td>
<td>0.5</td>
<td>0</td>
</tr>
</tbody>
</table>

tie situation, I arbitrarily select a node (e.g., node #2) and add it to the set $V$. By the end of the iteration, $V = \{2, 5\}$ covers the pairs (1, 3) and (1, 6). In iteration #3, to cover the only uncovered pair (1, 7), node #6 is selected. In the final solution, $V = \{2, 5, 6\}$.

Deviation path scenarios: trips between an O-D pair can be completed via shortest and/or deviation paths. I use $K = 3$ as an example to illustrate how the node weighting scheme works with deviation paths. Taking the O-D pair (1, 3) as an example, the first three shortest paths are 1 → 2 → 3, 1 → 4 → 3, and 1 → 2 → 4 → 3 in an ascending order. I will then examine each of the paths, if charging stations are needed on that path. In particular, on the path 1 → 2 → 3, the node #2 is needed to cover this O-D pair. Similarly, on the path 1 → 4 → 3, the node #4 is needed and on the path 1 → 2 → 4 → 3, both nodes #2 and #4 are needed. As a result, the weights for nodes #2 and #4 are identical as 1.5 and 1.5 in
Table 4.2: Node weights in iteration #2 with shortest path scenario

<table>
<thead>
<tr>
<th>Uncovered O-D pairs</th>
<th>Node #</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 2 3 4 5 6 7</td>
</tr>
<tr>
<td>(1, 3)</td>
<td>0 1 0 0 0 0 0</td>
</tr>
<tr>
<td>(1, 7)</td>
<td>0 0 0 0 0 1 0</td>
</tr>
<tr>
<td>Node weights</td>
<td>0 1 0 0 0 1 0</td>
</tr>
</tbody>
</table>

Table 4.3: Node weights in iteration #1 with deviation path scenario

<table>
<thead>
<tr>
<th>Uncovered O-D pairs</th>
<th>Node #</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 2 3 4 5 6 7</td>
</tr>
<tr>
<td>(1, 3)</td>
<td>0 1.5 0 1.5 0 0 0</td>
</tr>
<tr>
<td>(1, 6)</td>
<td>0 1 0 1 1 0 0</td>
</tr>
<tr>
<td>(1, 7)</td>
<td>0 0.5 0 0.5 0.5 1.5 0</td>
</tr>
<tr>
<td>Node weights</td>
<td>0 3 0 3 1.5 1.5 0</td>
</tr>
</tbody>
</table>

Table 4.3. The same procedure is executed for the other two pairs (1, 6) and (1, 7), and the resulting node weights are shown in the same table. As seen, both nodes #2 and #4 have the highest equal weight; I arbitrarily pick node #2. By the end of the iteration, $V = \{2\}$ and the pair (1, 3) is covered via the path 1 → 2 → 3 and the pair (1, 6) is covered via the path 1 → 2 → 4 → 6, leaving only the pair (1, 7) uncovered. In iteration #2, the node #6, according to Table 4.4, is selected and the pair (1, 7) is covered via the path 1 → 2 → 4 → 6 → 7. The final solution under the scenario of deviation paths is $V = \{2, 6\}$.

4.2.2 Greedy-adding algorithm with extensions

To improve heuristic solution, I design an integrated GA-Extension (GA-E) algorithm, which includes three additional processes: pre-selection, substitution, and solution refining as highlighted in bold boxes in Figure 4.2. Referred to the numerical results in section 4.3, the GA-E improves the solution from the GA, however adds dramatically to the solving time.
Table 4.4: Node weights in iteration #2 with deviation path scenario

<table>
<thead>
<tr>
<th>Uncovered O-D pairs</th>
<th>Node # 1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1,7)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.5</td>
<td>0.5</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Node weights</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.5</td>
<td>0.5</td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>

Figure 4.2: A diagram of the integrated GA-E algorithm

**Pre-selection:** this process is to identify the nodes that have to be included to complete trips between some O-D pairs and executed once at the beginning of the GA procedures. The results of this process constitute an initial solution set $V$. The process is described as follows: On path $k$, $k = 1, 2, \ldots, K$, connecting an O-D pair $m = 1, 2, \ldots, |R \times S|$, if the sum distance of any two adjacent links on that path exceeds...
the vehicle range, the node where the two links meet will be selected and added to a temporary set $V_k$ (inner loop). After examining all $K$ paths for this O-D pair $m$, I take the intersection of the sets $V_k$, $k = 1, 2, \ldots, K$ to form the set $V_{pre}^m$ as the pre-selection node set for O-D pair $m$, i.e., $V_{pre}^m = \cap_{k=1}^{K} V_k$, where $m = 1, 2, \ldots, |R \times S|$. Repeat this process for all O-D pairs (outer loop) and take the union of $V_{pre}^m$, which constitutes the initial $V$ in Step 0, i.e., $V = \cap_{m=1}^{\mid R \times S \mid} V_{pre}^m$. Although this extension consumes additional computing time, the enhanced initial solution $V$ may help reduce the number of iterations and the total solution time.

Still take the sample network in Figure 4.1 as an example. With the shortest paths, all nodes #2, #5, and #6 have to be selected. Node #2 is included to serve charging stations traversing the path 123 between the O-D pair (1, 3). Similarly, nodes #5 and #6 have to be selected to cover the pairs (1, 6) and (1, 7). The initial set $V$ in Step 0 is $V = \{2, 5, 6\}$, by which all O-D pairs are already covered. With deviation paths (i.e., $K = 3$), node #6 is selected, because all charging stations have to refuel there in order to reach node #7. Thus, the initial set is $V = \{6\}$. As seen, the pre-selection can reduce the number of iterations.

*Substitution*: this process exchanges every node in set $V$ with every unselected nodes from set $\hat{N} \setminus V$. As soon as an exchanges that improve the solution is found, the exchange is made [27]. The process continues until either one successful substitution (or exchange) is completed or no substitution is made after all nodes have been examined. As noted in [27] and seen in vast numerical experiments [42, 52, 58], substitution improve solutions, but adds significantly to the execution time simultaneously.

*Solution Refining*: this process is executed once at the end of the entire GA algorithm to eliminate redundant nodes if any, due to the merger of the sets of selected nodes for all O-D pairs. The procedure is described as follows: randomly eliminate a node from a solution set $V$. If the elimination does not change the O-D pair coverage,
the subset \( V \setminus \{i\} \) preserves; otherwise continue on to the next node. It terminates when all the nodes in the set have been examined. If the solution set is ever reduced, the same refining process will be executed, which repeats until there is no more new solution set produced.

### 4.3 Results and discussions

I implemented the proposed GA and GA-E heuristics on the two networks, the Sioux-Falls road network [56] and a real-life aggregated highway network of the state of South Carolina. All deviation paths are exogenously generated using MATLAB and the run times are not reported in this paper. In this chapter, the exact solutions were obtained by using the CPLEX solver 12.6. All numerical implementations run on a desktop with 8 GB RAM and Intel Core i5-2500@3.30GHz processor under Windows 7 environment.

#### 4.3.1 Sioux Falls network

##### 4.3.1.1 Baseline numerical implementations

The network consists of 24 nodes and 76 directed links. In the baseline, I assume that all nodes are candidate sites for charging stations, i.e., \( \hat{N} = N \) and treat every candidate site equally important, i.e., \( w_i = 1, \forall i \in \hat{N} \). All 24 nodes are assumed as origins and destinations, i.e., \( R = S = N \), a total of 552 O-D pairs. I consider seven different vehicle ranges from 100 to 250 with an interval of 25, coupled with nine different deviation scenarios, i.e., \( K = 1, 2, 3, 4, \) and 5 and deviation cap (DC) = 10\%, 15\%, 20\%, and 50\%. Note that the scenario \( K = 1 \) is essentially the shortest paths, and the resulting model is equivalent to the prior flow-based set-
covering problems. The problem complexity increases with deviations. When $K = 5$, problem is most complex with 28,050 variables (including 2,784 binary variables) and 51,082 constraints. The results of heuristics and exact solutions (CPLEX) for the baseline are reported in Table 4.5 and their corresponding solving times are reported in Table 4.6.

**Table 4.5:** Number of stations by heuristics and exact solutions for the Sioux Falls network

<table>
<thead>
<tr>
<th>Deviation scenarios</th>
<th>Solution methods</th>
<th>Vehicle range</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>100</td>
</tr>
<tr>
<td>K=1</td>
<td>GA</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>GA-E</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>CPLEX</td>
<td>12</td>
</tr>
<tr>
<td>K=2</td>
<td>GA</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>GA-E</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>CPLEX</td>
<td>9</td>
</tr>
<tr>
<td>K=3</td>
<td>GA</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>GA-E</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>CPLEX</td>
<td>7</td>
</tr>
<tr>
<td>K=4</td>
<td>GA</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>GA-E</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>CPLEX</td>
<td>7</td>
</tr>
<tr>
<td>K=5</td>
<td>GA</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>GA-E</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>CPLEX</td>
<td>7</td>
</tr>
<tr>
<td>DC=10%</td>
<td>GA</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>GA-E</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>CPLEX</td>
<td>11</td>
</tr>
<tr>
<td>DC=15%</td>
<td>GA</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>GA-E</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>CPLEX</td>
<td>11</td>
</tr>
<tr>
<td>DC=20%</td>
<td>GA</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>GA-E</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>CPLEX</td>
<td>11</td>
</tr>
<tr>
<td>DC=50%</td>
<td>GA</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>GA-E</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>CPLEX</td>
<td>7</td>
</tr>
</tbody>
</table>

From Tables 4.5 and 4.6, both the number of stations and solving times decrease with the extended vehicle ranges. The reason is straightforward. A longer vehicle range helps reduce the refueling demand and thus requires fewer charging sta-

50
Table 4.6: Solving times (in CPU seconds) of heuristic and exact solutions for Sioux Falls network

<table>
<thead>
<tr>
<th>Deviation choice</th>
<th>Solution methods</th>
<th>Vehicle range</th>
<th>100</th>
<th>125</th>
<th>150</th>
<th>175</th>
<th>200</th>
<th>225</th>
<th>250</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K=1</td>
<td>GA</td>
<td></td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>&lt;1</td>
<td>&lt;1</td>
<td>&lt;1</td>
</tr>
<tr>
<td></td>
<td>GA-E</td>
<td></td>
<td>25</td>
<td>8</td>
<td>14</td>
<td>8</td>
<td>1</td>
<td>2</td>
<td>&lt;1</td>
</tr>
<tr>
<td></td>
<td>CPLEX</td>
<td></td>
<td>&lt;1</td>
<td>1</td>
<td>&lt;1</td>
<td>&lt;1</td>
<td>&lt;1</td>
<td>&lt;1</td>
<td>&lt;1</td>
</tr>
<tr>
<td>K=2</td>
<td>GA</td>
<td></td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>&lt;1</td>
</tr>
<tr>
<td></td>
<td>GA-E</td>
<td></td>
<td>50</td>
<td>38</td>
<td>26</td>
<td>9</td>
<td>1</td>
<td>1</td>
<td>&lt;1</td>
</tr>
<tr>
<td></td>
<td>CPLEX</td>
<td></td>
<td>18</td>
<td>17</td>
<td>21</td>
<td>9</td>
<td>1</td>
<td>1</td>
<td>&lt;1</td>
</tr>
<tr>
<td>K=3</td>
<td>GA</td>
<td></td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>GA-E</td>
<td></td>
<td>66</td>
<td>32</td>
<td>12</td>
<td>9</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>CPLEX</td>
<td></td>
<td>132</td>
<td>66</td>
<td>28</td>
<td>11</td>
<td>1</td>
<td>&lt;1</td>
<td>&lt;1</td>
</tr>
<tr>
<td>K=4</td>
<td>GA</td>
<td></td>
<td>7</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>GA-E</td>
<td></td>
<td>79</td>
<td>62</td>
<td>25</td>
<td>6</td>
<td>4</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>CPLEX</td>
<td></td>
<td>599</td>
<td>282</td>
<td>108</td>
<td>70</td>
<td>27</td>
<td>4</td>
<td>&lt;1</td>
</tr>
<tr>
<td>K=5</td>
<td>GA</td>
<td></td>
<td>7</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>GA-E</td>
<td></td>
<td>67</td>
<td>46</td>
<td>20</td>
<td>7</td>
<td>5</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>CPLEX</td>
<td></td>
<td>829</td>
<td>260</td>
<td>198</td>
<td>98</td>
<td>35</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>DC=10%</td>
<td>GA</td>
<td></td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>&lt;1</td>
<td>&lt;1</td>
<td>&lt;1</td>
</tr>
<tr>
<td></td>
<td>GA-E</td>
<td></td>
<td>52</td>
<td>22</td>
<td>28</td>
<td>5</td>
<td>1</td>
<td>&lt;1</td>
<td>&lt;1</td>
</tr>
<tr>
<td></td>
<td>CPLEX</td>
<td></td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>&lt;1</td>
<td>&lt;1</td>
<td>&lt;1</td>
</tr>
<tr>
<td>DC=15%</td>
<td>GA</td>
<td></td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>&lt;1</td>
<td>&lt;1</td>
</tr>
<tr>
<td></td>
<td>GA-E</td>
<td></td>
<td>71</td>
<td>17</td>
<td>19</td>
<td>18</td>
<td>1</td>
<td>1</td>
<td>&lt;1</td>
</tr>
<tr>
<td></td>
<td>CPLEX</td>
<td></td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>&lt;1</td>
</tr>
<tr>
<td>DC=20%</td>
<td>GA</td>
<td></td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>GA-E</td>
<td></td>
<td>108</td>
<td>26</td>
<td>9</td>
<td>9</td>
<td>1</td>
<td>1</td>
<td>&lt;1</td>
</tr>
<tr>
<td></td>
<td>CPLEX</td>
<td></td>
<td>3</td>
<td>5</td>
<td>20</td>
<td>17</td>
<td>8</td>
<td>2</td>
<td>&lt;1</td>
</tr>
<tr>
<td>DC=50%</td>
<td>GA</td>
<td></td>
<td>12</td>
<td>9</td>
<td>7</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>GA-E</td>
<td></td>
<td>239</td>
<td>126</td>
<td>34</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>CPLEX</td>
<td></td>
<td>202</td>
<td>237</td>
<td>198</td>
<td>365</td>
<td>136</td>
<td>77</td>
<td>17</td>
</tr>
</tbody>
</table>

...tions, which also helps reduce the total solving time. A higher deviation will make more paths available and thus save charging stations (with few exceptions in the GA, e.g., $K = 4$ with vehicle range of 100). On the other hand, the numbers of variables and constraints rise and the MILP is exponentially difficult to solve by using exact solutions (see the solution times with CPLEX). Nevertheless, the solution times with the GA seem remaining in a close range. This is because the total solving time of the GA is a combined result of the complexity of solving an iteration and the number of iterations (see Remark 2). A more relaxed deviation, though taking longer to solve...
in each iteration, may finish with fewer iterations and as a result would not change the solving time much. The GA-E is more time consuming, mainly due to the substitution process, however, it generally yields better solutions than the GA.

Table 4.7: Statistics on results comparing heuristic solutions with the exact solution

<table>
<thead>
<tr>
<th>Gap (number of stations)</th>
<th>GA-E</th>
<th>GA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of instances</td>
<td>Relative percentage</td>
</tr>
<tr>
<td>0</td>
<td>47</td>
<td>75%</td>
</tr>
<tr>
<td>1</td>
<td>15</td>
<td>24%</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2%</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0%</td>
</tr>
</tbody>
</table>

I compared the heuristic solutions with the exact solutions for all the 63 (= 9 deviations 7 vehicle ranges) instances and the gaps, defined as the differences in the number of stations, are reported in Table 4.7. The gaps are small in a range between zero and three, with average of 0.27 and 0.62 for the GA-E and the GA, respectively, which indicates that the heuristics yield quality solutions and the GA-E can further improve solutions. I also demonstrate how paths vary with deviations using the results of the Sioux Falls network. Fig. 4 plots the best charging station locations (green nodes) with the vehicle range of 100 and deviations \( K = 1 \) and \( K = 3 \). I use the paths between the O-D pair #1 and 6 to illustrate the variations between deviation paths as highlighted in figures. When \( K = 1 \), the path \#1 → 2 → 6 is used and EVs have to stop at node #2 for refueling. When \( K = 3 \), two additional paths become acceptable, which are the paths \#1 → 3 → 4 → 5 → 6 and \#1 → 3 → 12 → 11 → 4 → 5 → 6, respectively the second and third shortest paths, which help eliminate the node #2. The path \#1 → 3 → 4 → 5 → 6 is completed via refueling at node #5 and the path \#1 → 3 → 12 → 11 → 4 → 5 → 6 is completed via multi-stop refueling at nodes #12, 11, and 5. The noticeable variations between deviation paths are attributed to the dispersions of nodes (both O-D pairs and charging station locations) on the
network.

4.3.1.2 Effects of network size

Numerical experiments were conducted to analyze the effects of network size on the solution time. I construct another five networks of different sizes using the Sioux Falls network as follows: randomly select four nodes out of the 24 as origins and destinations (12 O-D pairs), then add another four random nodes (56 O-D pairs), and continue adding four random nodes each time until all the 24 nodes are used (552 O-D pairs). I implement the heuristics and exact solution on the five new networks with a vehicle range of 100 under three deviation scenarios ($K = 1$, 3, and 5). The solving times plotted in Figure 4.3 indicate that the exact solutions are much more sensitive to the network sizes and deviations than the GA solutions. When $K = 1$, the problem size is relatively small and CPLEX outperforms the heuristics (both the GA and GA-E). With relaxed deviations, the solving times of the exact solutions increase exponentially with the increased problem size while the solving times of the GA only increase in a non-discernible range. This is consistent with the observations from Table 4.6 and can be similarly explained. The GA-E in all cases take longer to solve than the GA, mainly due to the substation which also makes the solving time also dependable on the network size and deviation paths.

4.3.2 South Carolina network

I implement the model and solution methods on an illustrative case study, in which an EV fast-charging corridor is developed in the state of South Carolina (SC). The MPRLM is used to determine the minimum number of stations and locations on high-
ways so that travelers can drive EVs between major cities across the state. Figure 4.4 shows an aggregate highway network of SC that consists of 519 nodes (including 15 major cities, highway junctions, and rest areas) and 876 bidirectional links including interstate highways and US and state routes. In South Carolina, the 15 cities (represented by red dots in the figure) have public EV charging stations already placed within city boundaries. I treat these cities as both origins and destinations, constituting 210 O-D pairs, and all other 504 nodes are candidate locations for EV charging stations. Three realistic vehicle ranges - 85, 100, and 150 miles (projected), based on Nissan Leafs performances [46], are considered. A deviation scenario \( K = 3 \) is adopted since most mobile map services now provide users with three routes between any O-D pairs, which results in a total of 630 deviation paths. However, as some routes may deviate substantially (much longer routes) which are unlikely to be used, I impose a cap (e.g., 20%) to eliminate those routes. It forms the other deviation scenario, i.e., \( K = 3 \) with \( DC=20\% \) with fewer deviation paths.

Note that it was not possible to compare the performances of the heuristic solutions to the exact solutions because of the size and complexity of the network. For example, the computer took 4 hours to attain a solution with 72.23\% optimality gap on the instance of \( K = 3 \) with vehicle range of 85 miles, which has 56,066 constraints and 28,869 variables including 1,149 binary variables, and it was already running low on virtual memory.

I only compare the GA to the GA-E solutions on the SC network. The results displayed in Table 4.8 show that the GA-E heuristics improves the solution over the GA in all cases. The extensions, however, do incur much higher run times. For example, in the case with vehicle range of 85 miles and \( K = 3 \), the GA takes only 20 seconds to yield a solution of 28. Though reduced to 26, the GA-E consumed 14,402 seconds. In this case, extension yields a solution with 7\% fewer stations, but takes
about 720 times longer to solve. When comparing the two deviations, the resulting number of stations is close for the same solution method used, which implies that the deviation paths under both deviation scenarios are similar in shapes; however, as the number of deviation paths is reduced with the scenario of $K = 3$ with DC=20%, the total solving time decreases in all instances.

Different from the Sioux Falls network, in which deviation paths are in noticeable different shapes (see Figure 4), the majority of 15 cities on SC network are located on the ends of the interstate highways (see Figure 4.4). The resulting tree structure is less likely to have deviation paths with substantial different sequences of nodes visited, since the majority of the trips may be completed via the interstate highways, although it may also be in part contributed by the relatively low deviation ($K = 3$). To better understand the effects of deviation and network size on the SC case study, I consider a larger deviation ($K = 5$) and an extended network by adding another 25 cities that are projected as future PEV markets based on their socioeconomic characteristics (see Figure 4.5), resulting in a total of 1,560 O-D pairs. The total number of deviation paths varies with the deviation scenario and network size in a range between $630 (=210 \times 3)$ and $7,800 (=1,560 \times 5)$.

Table 4.9 reports the results of deviation $K=5$ compared with $K=3$ for both network sizes of 15 and 40 cities. The vehicle range of 85 miles is adopted. First, the total number of stations is reduced as a result of increased deviation for all cases. Second, when comparing the results over network expansions (horizontally), the number

<table>
<thead>
<tr>
<th>Deviations</th>
<th>Solution methods</th>
<th>85 miles</th>
<th>Number of stations</th>
<th>Times (s)</th>
<th>100 miles</th>
<th>Number of stations</th>
<th>Times (s)</th>
<th>150 miles</th>
<th>Number of stations</th>
<th>Times (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K=3$</td>
<td>GA</td>
<td>28</td>
<td>20</td>
<td>19</td>
<td>13</td>
<td>13</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$K=3$ with DC=20%</td>
<td>GA</td>
<td>26</td>
<td>14,402</td>
<td>16</td>
<td>4,033</td>
<td>11</td>
<td>3,261</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GA-E</td>
<td>26</td>
<td>12</td>
<td>19</td>
<td>9</td>
<td>13</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GA-E</td>
<td>25</td>
<td>7,489</td>
<td>17</td>
<td>4,246</td>
<td>11</td>
<td>2,045</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4.9: Effects of deviations on SC network (vehicle range of 85 miles)

<table>
<thead>
<tr>
<th>Deviations</th>
<th>Solution methods</th>
<th>15 cities</th>
<th>40 cities</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of stations</td>
<td>Times (s)</td>
<td>Number of stations</td>
</tr>
<tr>
<td>K=3</td>
<td>GA</td>
<td>28</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>GA-E</td>
<td>26</td>
<td>14,402</td>
</tr>
<tr>
<td>K=5</td>
<td>GA</td>
<td>26</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>GA-E</td>
<td>25</td>
<td>17,881</td>
</tr>
</tbody>
</table>

of stations with K=5 increases by 11 (=37-26) and 9 (=34-25) stations respectively for the GA and GA-E, which are lower than the 13 and 12 in the correspondents under K=3. This is attributed to the increased deviation paths with higher deviation. When comparing the results over the deviations (vertically), the number of stations on the 15-city network decreases by 2 (=28-26) and 1 (=26-25) stations respectively with the GA and GA-E, which are smaller than the 4 and 4 stations in the correspondents on the 40-city network. This implies that a wider distribution of O-D pairs takes a better use of deviation paths. Third, comparing the solving times, the GA over the GA-E is more applicable to solve real-world problems with high quality solutions.

4.4 Summary

I developed greedy heuristics, especially the GA and GA-E algorithms, to address the computational challenges for solving the MPRLM, a flow based set-covering problem with deviation paths. The heuristics, which are the first in the literature for solving the flow-based set-covering problems with deviation paths, successfully take into account the effects of vehicle range and deviation paths on a transportation network. The solutions are applicable for MPRLM and other flow-based set-covering models, which are equivalent to shortest-path scenario (i.e., K=1) in the model.
The heuristic approaches were implemented on the two networks - the Sioux Falls test network and the real-life South Carolina network. The results indicate that the GA can greatly reduce the solution times while retaining high quality solutions, especially with higher deviation scenarios and larger networks. Comparing the GA with the GA-E, the extensions can improve solutions (i.e., fewer charging stations in this study) but add significantly to the solution times mainly due to the substitution process.
Figure 4.3: Effects of network size on the solution times

(a) shortest path ($K = 1$)

(b) 3-shortest paths ($K = 3$)

(c) 20% deviation cap
Figure 4.4: South Carolina state network (15 cities)

Figure 4.5: South Carolina state network (40 cities)
Chapter 5

Multi-period multi-path refueling location model for electric vehicle charging stations

5.1 Problem statement

I propose a multi-period multi-path refueling location model (M2PRLM), which is built upon the MPRLM [44]. The model expands a PEV charging network to serve growing intercity trips. The objective is to minimize the total cost of installations of new stations and relocations of existing ones while satisfying every O-D trip via at least one path between the O-D pair. The path can be either a shortest path or a path that is deviated away from a pre-defined path within a reasonable tolerance (called deviation path). The M2PRLM is formulated as a mixed integer linear program. I adopt a heuristic based on the genetic algorithm (Vose, 1999) and justify the model and heuristic using a benchmark network - the Sioux Falls network [56]. With the success of numerical experiments, I demonstrate the model with a real-world case.
study based on the geographic settings of South Carolina and explore the interplay between major factors, including geographic distributions of cities, vehicle range, and deviation choice.

The remainder of this paper is organized as follows. The formulation of the M2PRLM and the heuristic are presented in section 5.2. In section 5.3, I first justify the model and heuristic using the Sioux Falls network and then demonstrate the model with the case study of South Carolina followed by result discussions. I’ll draw conclusions in section 5.4.

5.2 Methods

5.2.1 Formulation of M2PRLM

The M2PRLM is extended from the MPRLM to make sequential decisions considering both spatial and temporal distributions of intercity trips. The model simultaneously considers multiple paths, limited vehicle range, and dynamic topologies of network in the decision process of installing new charging stations and possibly relocating existing ones whenever advantageous. The multiple paths in this study are comprised of both shortest and deviation paths. I deem a traveler willing to take either a shortest path (a pre-planned path) or a path that is detoured from a pre-planned path within their deviation tolerances. The concepts and generations of deviation paths have been explicitly discussed in details in the previous chapters.

In this study, there are two trade-offs: (i) installing new charging stations and/or relocating existing stations, and (ii) doing it now or later. The future costs of installations and relocations will be discounted to the present worth by using the engineering economics equation: \( P = F \left( \frac{P}{F}, \ i, \ n \right) \), in which all future values (F) are
converted to the equivalent present worth \((P)\), \(i\) is the annual discount rate, and \(n\) is the number of years from now. The maintenance cost, between $1,000 and $2,000 per year [77], is less than 1% of the installation cost and too trivial to be included in the model.

PEV sales have been rising with more cities becoming PEV adopters [3, 75], from which new intercity trips are generated. From a network modeling perspective, new O-D pairs will be sequentially added to a transportation network, which requires the charging infrastructure network to be expanded and adaptive to this topological dynamics. Cities will be ranked and selected to be the next EV adopters according to the result of multivariate statistical analysis. I report the details of such selection process on the case study of South Carolina in section 5.3 as an example. In this section, I focus on the developing a multi-period optimization model.

Same as for MPRLM, let \((N,A)\) be a transportation network, where \(N\) and \(A\) are the sets of nodes and links, respectively. Let \(\hat{N}\) be the set of candidate charging station locations, \(\hat{N} \subset N\) and this set is assumed to fixed and unchanged over time. For example, they can be the rest areas on highway network and junctions of highways. Cities are both origins and destinations on the network and they increase over time. Let \(R_t\), index \(r\), be a set of origin nodes, and \(S_t\), index \(s\), be a set of destination nodes, where \(t\) is the index of time stages (or periods) \(t \in T\). Let \(K^{rs}\) be a predefined maximum number of deviation paths for O-D pair \(r - s\), which are exogenously generated [44]. I denote by \(P^{rs,k}\) a sequence of nodes on the \(k^{th}\) path for O-D pair \(r - s\), where \(k = 1, 2, , K^{rs}\). Denote a link by \(a\) or a pair of ending nodes, i.e., \(a = (i, j) \in A\).

I made the following assumptions to simplify the modeling without the loss of generality: (1) the number of time stages is predetermined and each time stage has an equal length; (2) vehicles are homogeneous and fully charged at origins; (3)
energy consumed is unified in terms of travel distance; (4) vehicle range is known and homogenous for all EVs and for all time; and (5) charging stations are uncapacitated. The notation used in the model is first presented, and followed is the complete mathematical formulation of the M2PRLM in (5.1)-(5.13).

Indices:

$i$: index of candidate sites, $i \in \hat{N} \subset N$
$t$: index of time stages, $t \in T$; $r$: an origin node in the network, $r \in R_t \subset N$
$s$: a destination node in the network, $s \in S_t \subset N$
$k$: index of the paths for an O-D pair, $k = 1, 2, ..., K_{rs}$
$a$: index of arc set $A$, $a = (i, j) \in A$

Parameters:

$c_{it}^b$: cost of building a new charging station at node $i$ in time stage $t$,
$c_{ijt}^r$: fixed cost of relocating an existing charging station from node $i$ to $j$ in time stage $t$,
$\beta$: onboard fuel capacity (unified in travel distance), i.e., vehicle range
$M$: a sufficiently large number
$P_{rs,k}$: a sequence of nodes on the $k^{th}$ path from $r$ to $s$ and then back to $r$ by the same path, where $k = 1, 2, ..., K$
$d_{ij}$: distance between node $i$ and $j$
$p_{ij}^r$: variable cost of relocating charging stations from node $i$ to $j$ (including both distance- and time- transportation cost). Assume that the cost is invariant with time stages, $\delta_{ij}^{rs,k} = 1$ if node $i$ is in the set of node $P_{rs,k}$, 0 otherwise; this is an outcome of the deviation paths that are exogenously generated
$\omega_i$: weighting factor that differentiates candidate sites, $i \in \hat{N}$,
Variables:

$X_{it}$: =1 if a charging station is available at node $i$ in time stage $t$; 0 otherwise,

$Z_{it}$: =1 if a charging station is newly built at node $i$ in time stage $t$; 0 otherwise,

$\bar{Z}_{ijt}$: =1 if a charging station is relocated from node $i$ to $j$ in time stage $t$; 0 otherwise,

$Y^{rs,k}$: =1 if the $k^{th}$ path between $r$ and $s$ is selected to be electrified; 0 otherwise

$B_{i}^{rs,k}$: remaining onboard power at node $i$ on the $k^{th}$ path of O-D pair $r - s$

$l_{i}^{r,s,k}$: amount of power recharged at node $i$ on the $k^{th}$ path of O-D pair $r - s$

$$\min \sum_{i \in \hat{N}} \sum_{t \in T} \omega_{i} c_{it}^{b} Z_{it} + \sum_{i \in \hat{N}} \sum_{j \in \hat{N}} \sum_{t \in T} (c_{ijt}^{r} + p_{ij}^{r}) \bar{Z}_{ijt} \quad (5.1)$$
Subject to:

\[ B_{rs,k}^i + l_{rs,k}^i \leq M(1 - Y_{rs,k}^i) + \beta, \quad \forall r \in R_t, s \in S_t; i \in P_{rs,k}; t \in T; k = 1, 2, ..., K^{rs} \]  
(5.2)

\[ B_{rs,k}^i + l_{rs,k}^i - d_{ij} - B_{rs,k}^j \leq M(1 - Y_{rs,k}^i), \quad \forall r \in R_t, s \in S_t; i \in P_{rs,k}; t \in T; k = 1, 2, ..., K^{rs} \]  
(5.3)

\[ -(B_{rs,k}^i + l_{rs,k}^i - d_{ij} - B_{rs,k}^j) \leq M(1 - Y_{rs,k}^i), \quad \forall r \in R_t, s \in S_t; i \in P_{rs,k}; t \in T; k = 1, 2, ..., K^{rs} \]  
(5.4)

\[ \sum_{r \in R_t} \sum_{s \in S_t} \sum_{k} l_{rs,k}^i \delta_{rs,k}^i \leq M X_{it}, \quad \forall i \in \hat{N}; t \in T \]  
(5.5)

\[ \sum_{k=1}^{K^{rs}} Y_{rs,k}^i \geq 1, \quad \forall r \in R_t, s \in S_t; t \in T \]  
(5.6)

\[ B_{rs,k} = \beta, \quad \forall r \in R_t, s \in S_t; k = 1, 2, ..., K^{rs} \]  
(5.7)

\[ X_{it} = X_{i,t-1} + Z_{it} - \sum_{j} \bar{Z}_{ij,t} + \sum_{j} \bar{Z}_{ij,t}, \forall t \in T \setminus 1; i \in \hat{N} \]  
(5.8)

\[ X_{i1} = Z_{i1}, \forall i \in \hat{N} \]  
(5.9)

\[ X_{it}, Z_{it} = \{0, 1\}, \forall t \in T; i \in \hat{N} \]  
(5.10)

\[ \bar{Z}_{ij,t} = \{0, 1\}, \forall t \in T; i, j \in \hat{N} \]  
(5.11)

\[ Y_{rs,k} = \{0, 1\}, \forall r \in R_t, s \in S_t; t \in T; k = 1, 2, ..., K^{rs} \]  
(5.12)

\[ B_{rs,k}^i \geq 0, l_{rs,k}^i \geq 0, \forall r \in R_t, s \in S_t; t \in T; i \in P_{rs,k}; k = 1, 2, ..., K^{rs} \]  
(5.13)

The objective is to minimize the total cost of new charging stations and relocations for a finite planning horizon. The cost of installation of new charging stations may be location specific and varies with site conditions (e.g., pre-wired). The relocation cost depends on the relocation distance and site conditions. The inequalities (5.2)-(5.7) are constraints for each time stage and describe the spatiality of the network and constraints (5.8)-(5.9) capture the temporality of the problem for a sequential expansion of charging network.

Constraint set (5.2) assures that the total onboard energy does not exceed battery capacity \((B_{rs,k}^i + l_{rs,k}^i \leq \beta)\) on paths that are taken (i.e., \(Y_{rs,k}^i = 1\); oth-
Otherwise no restriction is applied (i.e., \( Y_{rs,k} = 0 \)), simply because no traveler will use that route. Constraints (5.3) and (5.4) concur to ensure that the energy consumption conservation (i.e., \( B_i^{rs,k} + l_i^{rs,k} - d_{ij} - B_j^{rs,k} = 0 \)) holds for all links on the kth path if the path is taken (i.e., \( Y_{rs,k} = 1 \)). Otherwise, when \( Y_{rs,k} = 0 \), the inequality becomes \( B_i^{rs,k} + l_i^{rs,k} - d_{ij} - B_j^{rs,k} \leq M \), i.e., no restraining effects. Constraint set (5.5) is a logic constraint, stating that recharging is only available at node \( i \) if a charging station is open. Constraint set (5.6) states that there is at least one path, either shortest or deviation paths, available between an O-D pair. Constraint set (5.7) realizes the assumption that all PEVs are fully charged at origins. Constraint set (5.8) describes an adaptive relationship, at node \( i \in \hat{N} \) and in time stage \( t \), between status variable (i.e., \( X_{it} \)), which indicates the availability of charging station, and activity variables (i.e., \( Z_{it} \) and \( \bar{Z}_{ij,t} \)), which indicate if a new station is built or an existing station is relocated. In particular, availability of a charging station involves both spatial and temporal interactions: station available from time stage \( t - 1 \), (i.e., \( X_{i,t-1} \)), new station installed in time stage \( t \), (i.e., \( Z_{it} \)), and the stations relocated in time stage \( t \) (i.e., \( \sum_j \bar{Z}_{ij,t} - \sum_j \bar{Z}_{ij,t} \)). The boundary condition of the charging network is given in constraint set (5.9), which states that all charging stations available in the first period are newly built at the beginning of the planning horizon. Constraints (5.10)-(5.13) are binary and nonnegativity constraints.

**Remark 1.** Both charging and travel costs of paths are not included in the objective of the model. The M2PRLM is a spatial economics model, which determines the locations of charging stations only based on the spatial relationships between O-D pairs and roadway networks. Computing both costs is in need of the exact number of trips between O-D pairs. Crowdsourced data can be used, such as call detailed records (CRDs), to extract traces and estimate the distributions of trips, other than simulation as used in a recent study [49]. The inclusion of traffic flow or intercity
trips will lead to a new study.

The M2PRLM is a MILP. The number of decision variables and constraints increase exponentially with the number of deviation paths and time stages. The problem is NP-hard, because it reduces to the well-known set-covering location problem [27], when the planning horizon shrinks to one single period. Without an effective solution, this model is intractable even for a moderate sized problem.

Remark 2. The binary relocation variable \( \bar{Z}_{ijt} \) can be relaxed and it will not affect the solution of the model. This is because the \( \bar{Z}_{ijt} \) will naturally converge to binary due to the binary variables \( X_{it} \) and \( Z_{it} \) in constraint set (5.8). In this study, all numerical results obtained by CPLEX are result of relaxing variable \( \bar{Z}_{ijt} \).

5.2.2 A heuristic based on genetic algorithm for M2PRLM

In this subsection, I adopt a heuristic based on GA, which was developed by [10] for solving single-stage set-covering problems. I modify the operations to tune up the solution performance and add a procedure of feasibility check and solution refining. For completeness, the major procedures of the algorithm are reported and explained in the context of the problem formulation, which are representation and fitness function, parent selection, crossover operator and mutation, and feasibility check and solution refining. The notation used in solution is adopted from previous section.

Representation and fitness function: A 0-1 matrix \( X_{[N \times T]} \) is used to show the charging stations deployed on a spatio-temporal network, in which each row represents availability of charging stations for a time stage and each column represents how a charging station is used over time. For example, a cell \( X_{it} = 1 \) indicates that a charging station is available for service at node \( i \) in time stage \( t \). A matrix is called \( X_{[N \times T]} \) a feasible solution only if charging stations deployed over space and time
can satisfy all O-D trips. With the solution matrix of status variables, the activity decisions, i.e., $Z_{it}$ and $Z_{ijt}$, and routing decision $Y_{rs,k}$ can be readily retrieved. In this study, the fitness of each candidate solution (also called, individual) is defined as the value of the objective function (5.1) of the M2PRLM. As a cost minimization problem, a lower objective value indicates a better fitness, and vice versa.

**Parent selection**: The binary tournament selection [10] is used for parent selections. In particular, I initialize a large population of feasible solutions (e.g., 100), and randomly pick four individuals, to form two pools, each of which contains two individuals.

**Crossover operator and mutation**: The individual with better fitness in each pool will be selected as a parent to breed children based on fitness-based crossover operator [10]. Let $f_{P1}$ and $f_{P2}$ be the values of objective function for parents $P1$ and $P2$ respectively and let $C$ be child matrix. Given $i = 1, ..., |N|, t = 1, ..., |T|$:  

1) if $P_{1it} = P_{2it}$, then set $C_{it} := P_{1it}$ or $P_{2it}$;

2) if $P_{1it} \neq P_{2it}$, then $C_{it} := P_{1it}$ with probability $p = \frac{f_{P2}}{f_{P1} + f_{P2}}$, and $C_{it} := P_{2it}$ with probability $1 - p$.

Once a child matrix is formed, each cell in the matrix will be inverted based on a mechanism, called mutation [81]. In this study, I invert a cell (i.e., invert the value of the cell from zero to one, or vice versa.) if a randomly generated probability is less than a pre-defined threshold (e.g., 10%); otherwise the cell remains unchanged.

**Feasibility check and solution refining**: The crossover and mutation processes may inevitably cause infeasibility. A feasibility check is thus developed to examine if the resulting charging stations in every time stage can cover trips between every O-D pair, given a fixed vehicle range. In any time stage $t$, if the deployed charging stations cannot cover all O-D trips, this deployment solution will be replaced by an-
other solution, which is randomly selected from any later period \((t = t + 1, \ldots, |T|)\) of the initial set of feasible solutions. Such replacement warrants feasibility. This is because the set of O-D pairs of time stage \(t\) is a subset of a later period \(t\), as the network expands over time, so that a solution that is feasible in period \(t\) must also be feasible in period \(t\).

On the other hand, this feasibility remedy may introduce redundant charging stations. I develop a refining procedure to eliminate the redundancy as follows: randomly eliminate a node from a solution set \(V\) (for a single period). If the elimination does not cause infeasibility, the subset preserves; otherwise continue on to the next node. It terminates when all the nodes in the set have been examined. If the solution set is ever reduced, the same refining process will be executed, which repeats until there is no more new solution set produced. This process is applied for all the periods.

Population replacement: If a child solution is identical to any of the solutions in the initial population, this child solution will be neglected; otherwise it replaces the solution in the initial population with the worst fitness.

The GA procedure terminates when a maximum predetermined number of iterations \(M\) (e.g., 100) is reached. The final solution is the one with the best fitness in the population.

The procedure of the heuristic is also summarized as follows:

- Initialization: Randomly generate \(n\) solutions as the initial population;

- For iteration \(i = 1, 2, \ldots, M\)

Step 1. Pick four solutions from the population to form two pools, each of which contains two solutions;

Step 2. Select the solution with better fitness in each pool as one of the parents then do crossover to form a temp solution (may be infeasible);
Step 3. Mutation is applied to the temp solution with a mutation rate of $\alpha$;

Step 4. Check whether the temp solution is feasible. If yes, go to step 5; otherwise, use the feasibility remedy procedure to generate a feasible solution;

Step 5. Remove redundant stations in the temp solution to generate a child solution;

Step 6. If the child solution is not identical to any solutions in the population, replace the solution in the population with the worst fitness. If this is the last iteration, then terminate, otherwise, update the index of iteration and go to Step 1.

5.3 Results and discussions

5.3.1 The Sioux Falls network

Same as in previous two chapters, I first justified the M2PRLM and the genetic algorithm based heuristic on the Sioux Falls network. All nodes are candidate sites for charging stations, i.e., $\hat{N} = N$. There are six time stages with equal time intervals (e.g., one year). The O-D pairs are gradually added following the procedure: starts with four nodes that are randomly selected from the 24 nodes, resulting in $4 \times 3 = 12$ O-D pairs, and add another four new nodes in each time stage until all the 24 nodes are used up. In the end, there are a total of 552 O-D pairs. The vehicle range (VR) is assumed to be 100 miles in this case study. I assume that weighting factor on each candidate site is identical to unity, i.e., $w_i = 1, \forall i \in \hat{N}$.

The average costs of new fast charging stations and fixed relocation cost are $122,000 and $38,000, respectively, according to [26]. The new fast charging station cost includes the costs of equipment, installation, utility interconnection, as well
as host-site identification, analysis, screening and leases while the fixed relocation cost includes every component except for the equipment cost. The variable cost of relocation mainly occurred in transportation is $1.38 per mile [71]. A 5% annual discount rate is used for calculating the present worth of costs. To simplify the numerical tests, I assume that installation cost is identical to all sites.

5.3.1.1 Baseline case

I implement the M2PRLM for both the shortest path \((K = 1)\) and multi-path \((K = 3)\) scenarios for all O-D pairs and results are presented in Tables 5.1 and 5.2, respectively. The results were obtained by using CPLEX. The location IDs of new charging stations are highlighted in the tables.

Both tables show that there is no station relocation as the anticipation of future O-D pairs is incorporated in the M2PRLM. By comparing the results between the two deviation scenarios (i.e., \(K = 1\) to \(K = 3\)), the higher deviation reduces the total number of stations needed from 12 to 7 or a \((1.37-0.84)/1.37 = 38\%\) reduction in the total cost. This is because the deviation paths allows trips between an O-D pair to be completed via more than a shortest path (i.e., \(K = 1\)). Note also that in both cases there are more stations deployed in earlier stages, even though there are more O-D pairs introduced to the system in the later stages, which implies that paths used by the O-D trips in later stages largely overlap with the ones in early stages.

I test the performance of the heuristic and compare the solution quality and solving times to the counterparts of exact solutions obtained by CPLEX. As the heuristic may be sensitive to the choices of parameters in the heuristic, I conduct a series of numerical experiments of varying a few major parameters. In particular, I consider 18 different combinations from three different population size (i.e., 30, 50
Table 5.1: M2PRLM solution ($k = 1$)

<table>
<thead>
<tr>
<th>Time stage</th>
<th>Number of origins and destinations (O-D pairs)</th>
<th>Number of total available stations (station location IDs)</th>
<th>Number of new stations relocated</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>4 (#5,8,13,18)</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>-12</td>
<td>8</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>-56</td>
<td>9</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>-132</td>
<td>12</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>16</td>
<td>16</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>-240</td>
<td>12</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>20</td>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>24</td>
<td>24</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>-552</td>
<td>12</td>
<td>-</td>
</tr>
</tbody>
</table>

Total present worth of cost $1.37M

Table 5.2: M2PRLM solution ($k = 3$)

<table>
<thead>
<tr>
<th>Time stage</th>
<th>Number of origins and destinations (O-D pairs)</th>
<th>Number of total available stations (station location IDs)</th>
<th>Number of new stations relocated</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>4 (#11,12,18,21)</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>-12</td>
<td>7</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>-56</td>
<td>7</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>-132</td>
<td>7</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>16</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>-240</td>
<td>7</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>20</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>24</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>-552</td>
<td>7</td>
<td>-</td>
</tr>
</tbody>
</table>

Total present worth of cost $0.84M

and 100), three different mutation rates (i.e., 5%, 10% and 15%), and two different numbers of iterations (i.e., 50 and 100) for the deviation scenario $K = 1$. I run the heuristic for 200 times and report the results of the 95 percentiles of objective values and solving times in Table 5.3. From the table, it can be observe that as population size increases, the objective value decreases or solution quality improves. Mutation rate is crucial in controlling the solution quality and efficiency. A low rate may not suffice while a high rate may increase the solution time and make the problem too random. When the population size and mutation rate are fixed, more iterations
generally improve quality but in the meantime result in longer solving time. Within the 18 combinations, I pick the combination: 100 population, 10% mutation rate, and 100 iterations, for implementing the heuristic.

The results and computational performances of heuristic are reported in Table 5.3:

Table 5.3: Heuristic parameters and performance

<table>
<thead>
<tr>
<th>Population size</th>
<th>Mutation rate</th>
<th>Number of iterations</th>
<th>Objective value ($M$)</th>
<th>Solving time (CPU seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>5%</td>
<td>50</td>
<td>1.45</td>
<td>21.5</td>
</tr>
<tr>
<td>30</td>
<td>5%</td>
<td>100</td>
<td>1.43</td>
<td>43.3</td>
</tr>
<tr>
<td>30</td>
<td>10%</td>
<td>50</td>
<td>1.44</td>
<td>23.3</td>
</tr>
<tr>
<td>30</td>
<td>10%</td>
<td>100</td>
<td>1.43</td>
<td>45</td>
</tr>
<tr>
<td>30</td>
<td>15%</td>
<td>50</td>
<td>1.45</td>
<td>22.9</td>
</tr>
<tr>
<td>30</td>
<td>15%</td>
<td>100</td>
<td>1.44</td>
<td>44.1</td>
</tr>
<tr>
<td>50</td>
<td>5%</td>
<td>50</td>
<td>1.42</td>
<td>20.2</td>
</tr>
<tr>
<td>50</td>
<td>5%</td>
<td>100</td>
<td>1.4</td>
<td>38.4</td>
</tr>
<tr>
<td>50</td>
<td>10%</td>
<td>50</td>
<td>1.42</td>
<td>22.6</td>
</tr>
<tr>
<td>50</td>
<td>10%</td>
<td>100</td>
<td>1.42</td>
<td>43.8</td>
</tr>
<tr>
<td>50</td>
<td>15%</td>
<td>50</td>
<td>1.42</td>
<td>22.2</td>
</tr>
<tr>
<td>50</td>
<td>15%</td>
<td>100</td>
<td>1.42</td>
<td>44.3</td>
</tr>
<tr>
<td>100</td>
<td>5%</td>
<td>50</td>
<td>1.41</td>
<td>21.6</td>
</tr>
<tr>
<td>100</td>
<td>5%</td>
<td>100</td>
<td>1.4</td>
<td>42.2</td>
</tr>
<tr>
<td>100</td>
<td>10%</td>
<td>50</td>
<td>1.41</td>
<td>21.5</td>
</tr>
<tr>
<td>100</td>
<td>10%</td>
<td>100</td>
<td>1.4</td>
<td>41.7</td>
</tr>
<tr>
<td>100</td>
<td>15%</td>
<td>50</td>
<td>1.42</td>
<td>21.7</td>
</tr>
<tr>
<td>100</td>
<td>15%</td>
<td>100</td>
<td>1.41</td>
<td>42.5</td>
</tr>
</tbody>
</table>

4, compared with the counterparts of the exact solution (CPLEX). The size of problem dramatically increases with deviation paths. In particular, there are 10,792 variables (including 3,168 binary variables) and 6,890 constraints when $K = 1$. When $K = 3$, the numbers of variables and constraints increase to 40,746 (including 7,284 binary variables) and 52,178, respectively. From Table 5.4, the heuristic can yield high quality solutions, both within an average 3% gap of optimality. The solution is efficient especially when the problem is getting complex with deviations ($K = 3$). The results lend us confidence in implementing the heuristic for solving real-life case study of South Carolina.
Table 5.4: Comparisons between heuristic and CPLEX

<table>
<thead>
<tr>
<th>Deviation scenario</th>
<th>Heuristic</th>
<th>CPLEX</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total present worth of cost ($M)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K=1</td>
<td>$1.40</td>
<td>$1.37</td>
</tr>
<tr>
<td>K=3</td>
<td>$0.85</td>
<td>$0.84</td>
</tr>
<tr>
<td>Solving time (CPU seconds)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K=1</td>
<td>33</td>
<td>2</td>
</tr>
<tr>
<td>K=3</td>
<td>42</td>
<td>2318</td>
</tr>
</tbody>
</table>

5.3.1.2 Comparisons with myopic solutions

The model solutions are compared to myopic solutions. The myopic method is a so-called shortsighted approach in the sense that the method only does the best for now but neglects the future. This method is popular in engineering practice for its easy implementation. In this paper, although the complete information about future demand is assumed to be available, I compare the optimization solution to the myopic solution to highlight the differences in staged decision-making processes. In particular, the multistage optimization model takes into account the whole trajectory of future demand while the myopic method solves single-period MPRLM for each period successively. For illustration purpose, I only report the results of myopic solution based on $K = 1$ in Table 5.5.

The table presents a different sequence of locating and relocating stations compared to the optimization model solutions in Table 5.3. Although both myopic and optimization solutions result in the same total of 12 charging stations, station relocations occur in the last four time stages if myopic solution is adopted while there is no relocation by the optimization solution. Because of the relocations, the myopic solution yields a higher total cost by $(1.59-1.37)/1.37 = 16\%$. I run another test, in which no relocation is allowed. The results in Table 6 show that the total cost is even higher by $(1.69-1.37)/1.37 = 23\%$ from the optimization solution, because of the increased total number of stations from 12 to 15.
Table 5.5: Myopic solution with relocation ($K = 1$)

<table>
<thead>
<tr>
<th>Time stage</th>
<th>Number of origins and destinations (O-D pairs)</th>
<th>Number of total available stations (station location IDs)</th>
<th>Number of new stations</th>
<th>Stations relocated</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>3 (#6,13,18)</td>
<td>3</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>8 (#4,6,12,13,14,16,18,19)</td>
<td>5</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>56</td>
<td>9 (#4,6,10,12,13,14,15,16,18)</td>
<td>1</td>
<td>#19#15</td>
</tr>
<tr>
<td>4</td>
<td>12</td>
<td>11 (#2,3,4,6,10,12,13,14,17,18,22)</td>
<td>2</td>
<td>#16#17</td>
</tr>
<tr>
<td>5</td>
<td>132</td>
<td>12 (#2,3,4,6,10,11,13,14,17,18,21,22)</td>
<td>1</td>
<td>#12#21</td>
</tr>
<tr>
<td>6</td>
<td>20</td>
<td>24 (#2,3,5,8,10,11,13,14,18,19,21,22)</td>
<td>0</td>
<td>#6#8</td>
</tr>
</tbody>
</table>

Total present worth of cost $1.59M

Table 5.6: Myopic solution without relocation ($K = 1$)

<table>
<thead>
<tr>
<th>Time stage</th>
<th>Number of origins and destinations (O-D pairs)</th>
<th>Number of total available stations (station location IDs)</th>
<th>Number of new stations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>3 (#6,13,18)</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>8 (#4,6,10,12,13,14,18,19)</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>56</td>
<td>10 (#4,6,10,12,13,14,15,16,18,19)</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>12</td>
<td>13 (#2,3,4,6,10,12,13,14,15,16,18,20)</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>20</td>
<td>15 (#2,3,4,6,10,11,13,14,15,16,18,19,20,21)</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>132</td>
<td>15 (#2,3,4,6,10,11,13,14,15,16,18,19,20,21)</td>
<td>0</td>
</tr>
</tbody>
</table>

Total present worth of cost $1.69M

5.3.2 South Carolina Case Study

I use the case study of developing a public fast charging network in South Carolina to demonstrate the real-world application of M2PRLM. The following data are processed: roadway network, locations of cities, and candidate sites for fast charging...
stations. In this study, I use Geographic Information System (GIS) software packages (e.g., ArcGIS) to integrate location data with transportation network data [66]. An aggregate roadway network is shown in Figure 5.1, which consists of 519 nodes and 876 bidirectional links. The 519 nodes include cities, highway junctions, and rest areas, which are candidate sites for charging stations, and the links are interstate highway, US and state routes. I adopt the cost data of new charging stations and relocation from the report [26] and assume that the cost of installing new charging stations is identical to all candidate sites.

![Figure 5.1: The South Carolina network](image)

**Legend**
- Stage_1 cities
- Stage_2 cities
- Stage_3 cities
- Rest Area
- Junction
- Highway Network_Simplified
- Interstate Highway

0 15 30 60 Miles

76
5.3.2.1 Baseline case

I consider both shortest path (K=1) and multi-path (K=3) scenarios for all O-D pairs and two different vehicle ranges (VRs) of 100 and 150 miles. No optimal solutions can be attained within a reasonable amount of computing time by using CPLEX. I set the upper bound of run time to be four CPU hours and report the heuristic results and the best solution obtained by CPLEX in Table 7. In the case study, the heuristic uses population size of 100 and mutation rate of 10% and terminates after 50 iterations in each run. The heuristic results in Table 5.7 are of 95 percentiles of results of 50 runs. From Table 5.7, the heuristic attains lower total system cost with shorter solving times than the counterparts of the exact solutions (within four CPU hours) for all cases.

The results of this case study presented in this section are result of the heuristic. From the table, the extension of vehicle range by 50 miles reduces the total system cost by about 2/3 for both for $K = 1$ and $K = 3$. A deviation results in a lower total cost, but by a trivial extent. This is mainly because many cities are interconnected through freeways and alternative paths through secondary highways are not in favor. In other words, though available, the deviation paths do not render as much flexibility as I have seen in the case of the Sioux Falls network.

I illustrate the layouts of charging stations over the three phases based on the heuristic solutions for $K = 3$ with vehicle range of 150 miles in Figure 5.2. There

<table>
<thead>
<tr>
<th></th>
<th>VR = 100</th>
<th>VR = 150</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GA</td>
<td>CPLEX*</td>
</tr>
<tr>
<td>Total present worth of cost ($\text{M}$)</td>
<td>K=1</td>
<td>$3.87$</td>
</tr>
<tr>
<td>Solving time (CPU seconds)</td>
<td>K=1</td>
<td>321</td>
</tr>
</tbody>
</table>
are nine charging stations installed in Phase I, as shown in Figure 5.2a), which are
geographically dispersed and located along the freeways. The same nine stations re-
main to serve the O-D trips expanded for 13 new cities in Phase II (see Figure 5.2b).
This is because the additional O-D pairs generated by the 13 cities largely overlap
with the existing O-D pairs between the 15 cities in Phase I. In Phase III, all 94 cities
are considered, which are clustered around a few major cities, including Greenville,
Rock Hill, Columbia, Charleston and Myrtle Beach. Only two new charging stations
are placed as highlighted in Figure 5.2c, together with the existing nine stations, to
serve the 8,742 O-D trips. Throughout the planning horizon, no charging station is
relocated.

As a multi-stage decision making process, it is of interest to understand how
adding or relocating a station on a network may result in a different performance in
the current period and the expected performance in future periods. I measure the
effects of invest or defer options [21] by comparing the model solution to the myopic
solution. The decomposed costs in each phase and corresponding deployment strat-
egy are displayed in Table 5.8 in comparison with the results of the M2PRLM. Note
that all the costs in the table stated in net present values. In Phase I, in relative to
myopic solution, one more charging station is invested, which costs $0.122M (=1.098-
0.976) to the current period but saves $0.147M (=0.124-0+0.169-0.146) for the future
periods or about $25,000 less in the overall cost. In terms of deployment strategy, the
one more new station installed in the first phase helps eliminate the one new station
and one station relocation in phase II and one station relocation in phase III.

In Table 5.8, Deployment strategy consists of the number of new stations in-
stalled (the number before +) and number of stations relocated (the number after +).
For example, 8+0 means that there are eight new stations installed and no station
relocated. The results also indicate that there are more stations to be deployed earlier
than later periods, largely because of the demand city distributions. This insight is consistent with the findings in [23] that their optimization model emphasized more on earlier locations and associated flow coverages and consequently selected charging station locations to cover most frequently used paths in earlier time periods while considering the increase of traffic volume in the later time periods.

Table 5.8: Comparisons between M2PRLM and myopic solutions

<table>
<thead>
<tr>
<th>Phase</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>Overall cost</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cost</td>
<td>Deployment</td>
<td>Cost</td>
<td>Deployment</td>
</tr>
<tr>
<td>Net Present Value</td>
<td>Myopic</td>
<td>$0.98</td>
<td>8+0</td>
<td>$0.12</td>
</tr>
<tr>
<td>($M)</td>
<td>M2PRLM</td>
<td>$1.10</td>
<td>9+0</td>
<td>$0</td>
</tr>
</tbody>
</table>

5.3.2.2 Evaluation of charging station deployment strategy under demand uncertainty

I evaluate the performance by the deployed charging stations from the baseline case under demand uncertainty in terms of the number of O-D pairs completed or covered. A higher coverage implies a more robust station deployment strategy under uncertainty. The uncertainty mainly refers to the randomness that cities become demand cities (or EV adopters), which may be due to the factors, such as economic and population growth. In this analysis, the uncertainty only emerges in the second phase, whereas the 15 cities in first phase and 94 cities in the last (third) phase of planning process are fixed and given. I randomly select cities (except those already having EV charging stations) until the total population of the selected cities is at least 50% of the statewide population. The study is based on the same, fixed five-year phase. Note that the phase length can be uncertain as well. However, the variations of the phase length will neither affect the location strategy nor the O-D pair coverage, but only change the present worth of the cost. In this sensitivity analysis, I conduct an analysis
of 20 random sets of different cities in the Phase II for each of the four combinations of deviation choices (i.e., \( K = 1 \) and \( K = 3 \)) and vehicle ranges (i.e., VR=100 and VR=150), a total of 80 scenarios. Within the 20 random sets, the number of cities selected is ranged from 27 to 39 and the population coverage is between 50.04% and 53.72%.

The results are reported in Figure 5.3, in which the horizontal axis denotes O-D pair coverage achieved in percentages while the vertical axis is the cumulative probabilities following each of the four combinations of deviation choices and vehicle ranges. For example, following \( K = 1 \) and VR=100, the minimum and maximum coverages are about 90.5% and 98.5%, respectively, and there is about 48% of the chance that the coverage is between 90.5% and 94%. It is also observed that other combinations result in overall higher coverage, due to the increased flexibility that helps achieve a higher coverage, given a station deployment. When the vehicle range is extended to 150 miles, regardless the deviation choice, the coverage is at least 98% and there is a high chance to have all O-D pairs covered. Even for the same vehicle range of 100 miles, the deviation can substantially increase the coverage as shown in Figure 5.3. Further investigation reveals that the average O-D pair coverages are 94.58% and 99.34% for combinations of \( K = 1 \) with VR=100 and \( K = 3 \) with VR=150, respectively. These results indicate that the charging station deployment from baseline is quite robust in providing high coverage of O-D pairs under demand uncertainty, mainly due to the clustered cities along major highways in the case study of South Carolina. This conclusion may not hold for another different geographic setting.
5.3.2.3 Differentiations of types of candidate sites

In the baseline study, all candidate locations are assumed to be identical, which may not always reflect the reality. For example, rest areas and cities, because of being prewired, would be cheaper than highway junctions to install new charging stations. There could be other societal concerns. For example, rest areas and cities may be more preferable than highway junctions for siting charging stations as travelers may be more familiar with them. To differentiate the types of candidate sites, I set five different weighting factors (in the model formulation) for rest areas and cities between 0.2 and 1 with an increment of 0.2 while fixing the weighting factor for junctions at 1. The correspondent rollout schemes are presented in Table 9. Compared to the baseline (in the last column), the total number of stations located at rest areas and cities decreases with the increase of weighting factors, which implies that the junctions are more geographically favorable. Except for the baseline, relocations occur in Phase III, which is indicated by the reduced number of stations placed at rest areas and cities from Phases II to III. This is because for a low weighting factor, a relocation is as equivalently costly as a new installation.

I plot the layouts of stations of weighting factor \( \omega_i = 0.2 \) as an example in Figure 6 to demonstrate the effects of weighting factor on system planning. To be comparable with baseline case, the same combination of \( K = 3 \) and \( VR=150 \) is used.

<table>
<thead>
<tr>
<th>Weighting factors ( \omega_i )</th>
<th>0.2</th>
<th>0.4</th>
<th>0.6</th>
<th>0.8</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of stations located</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>at rest areas and cities</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase I</td>
<td>8/11</td>
<td>9/10</td>
<td>10/11</td>
<td>7/9</td>
<td>4/9</td>
</tr>
<tr>
<td>Phase II</td>
<td>10/12</td>
<td>10/11</td>
<td>10/12</td>
<td>8/11</td>
<td>4/9</td>
</tr>
<tr>
<td>/Total number of stations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase II</td>
<td>9/12</td>
<td>8/11</td>
<td>8/12</td>
<td>7/12</td>
<td>5/12</td>
</tr>
</tbody>
</table>

The new station rollout scheme is noticeably different from baseline. In Phase I, there
are ten stations installed, eight of which are placed at rest areas and cities (Figure 5.4a). In contrast, there are only four out of nine stations placed at rest areas and cities in the baseline case (Figure 5.2b). In Phase II, two more new stations (see Figure 5.4b) are added to serve increased O-D pairs. In the last phase, instead of adding new stations, three stations are relocated (Figure 5.4c).

5.4 Summary

I developed the M2PRLM for strategically expanding public charging network to facilitate intercity trips by PEVs. The model successfully captured the topological dynamics of network with the emerging PEV markets and integrally considered the effects of limited vehicle range and flexibility in selecting deviation paths. To make the model tractable for large-scale problems, I solve the problem by adopting a heuristic based on genetic algorithm.

I justified the model and heuristic using the benchmark Sioux Falls network by comparing the numerical results and solutions performance with the exact solutions obtained by CPLEX. With the success of numerical experiments on Sioux Falls network, the model was demonstrated on a real-world South Carolina case study, which provided us with the following major insights. First, the charging station rollout scheme was subject to the geographic distributions of cities, vehicle range, and deviation choice. Second, the anticipation of the increase of future demands in the multistage optimization model could help reduce the overall cost in a long run, although it could result in poorer performance for the current period. Third, the charging location strategy was robust in providing high coverage of O-D pairs under demand uncertainty. Last, the highway junctions were more geographically
favored than the rest areas or cities, and relocation could be a cost-effective alterna-
tive to new installation when installing new stations at rest areas or cities are cheaper.
Figure 5.2: Charging station deployment over time ($K = 3$ and VR=150 miles)
Figure 5.3: Cumulative distribution of percentage of O-D pairs covered under random tests
Phase I (10 stations)

Phase II (12 stations)

Phase III (12 stations)

Figure 5.4: Charging station deployment with differentiation of the types of sites ($K = 3$ and $VR=150$)
Chapter 6

Probability-based multi-path
refueling location model for EV
charging stations

6.1 Problem statement

In the earlier stage of adopting EVs, market penetrate rate is still very low and whether a city would become an EV adopter is affected by many factors. It be a waste of resources if charging stations are placed to satisfy EV trips between two cities that are with very low chance to become EV market. In this chapter I present a probability-based flow refueling location model for EV charging stations. Based on the previous multi-path refueling location model (MPRLM), probability of each city to become an EV adopter is integrated in optimizing station deployment. Predicted by considering demographic and economic data, the probability of each city is used to represent the priority of that city in the objective to maximize the total coverage of all the cities in a network with a given budget. This probability-based multi-
path refueling location model (P-MPRLM) also differs from the MPRLM in ensuring round trips for covered O-D pairs. Same as in MPRLM, vehicle range and deviation behaviors of EV drivers are also taken into account in optimizing station deployment.

The remainder of chapter is organized as follows. In section 6.2, the detailed formulation of the proposed model is presented. Section 6.2.3 is dedicated to the designing of a genetic algorithm based heuristic to overcome the challenge to solve the proposed problem. In section 6.3, numerical experiment is implemented on the Sioux Falls network and results are analyzed. Section 6.4 concludes the study and outlines possible future research opportunities.

6.2 Methods

In this section, I extend the previous MPRLM to incorporate probability associated with demand nodes adopting EVs and to consider round trips between origins and destinations. In P-MPRLM, multiple paths are considered as candidate paths to complete round trips between origin and destination with one or more recharging stops. To simplify the problem, I make the assumption that drivers would use the same path as traveling from r to s to complete the trip from s to r.

6.2.1 Definition of coverage for each city

Figure 6.1 is used to illustrate how coverage is defined for each city. There are five nodes in the network, where node r is the origin node while all other four nodes are destination nodes. The bold lines in the figure represents that there exists at least one path that could be used to complete trips from node r to the destination node and then back to node r via charging en-route. Note that in reality, there are immediate nodes and multiple paths connecting origin and destination nodes. As in
this example, if there is no path electrified between node \( r \) and \( c \), in other words, no sufficient charging stations located along the paths to ensure the trip \( r \rightarrow s \rightarrow r \), then the coverage of node \( r \) is defined as \( 0.75(= 3/4) \), namely the percent of O-D pairs originating from node \( r \) be covered.

![Figure 6.1: Definition of coverage for a city](image)

### 6.2.2 Model formulation

Before introducing the proposed P-MPRLM, definition of decision variables and parameters is presented.

**Indices:**

- \( i \): index of candidate sites, \( i \in \hat{N} \subset N \)
- \( r \): an origin node in the network, \( r \in R \subset N \)
- \( s \): a destination node in the network, \( s \in S \subset N \)
- \( k \): index of the paths for an O-D pair, \( k = 1, 2, ..., K \)
- \( a \): index of arc set \( A \), \( a = (i, j) \in A \)

**Parameters:**

- \( C_i \): the installing cost of a charging station, \( i \in N \)
- \( \beta \): onboard fuel capacity (unified in travel distance), \( i.e., \) vehicle range
- \( \beta_0 \): initial battery status
- \( M \): a sufficiently large number
Variables:

$X_i$: =1 if an AFS is located at node $i$; 0 otherwise

$Y^{rs,k}$: =1 if the $k^{th}$ path between $s$ and $s$ can be completed (taken); 0 otherwise

$y^{rs}$: =1 if O-D pair $r-s$ is covered; 0 otherwise

$Z_i$: a continuous variable, representing the portion of a demand city being covered

$B_i^{rs,k}$: remaining onboard power at node $i$ on the $k^{th}$ path of O-D pair $r-s$

$l_i^{rs,k}$: amount of power recharged at node $i$ on the $k^{th}$ path of O-D pair $r-s$

$$\max \sum_r p_r Z_r \quad (6.1)$$
Subject to:

\[ B^{rs,k}_i + l^{rs,k}_i \leq M(1 - Y^{rs,k}) + \beta, \forall r, s; i \in P^{rs,k}; k = 1, 2, ..., K \] (6.2)

\[ B^{rs,k}_i + l^{rs,k}_i - d_{ij} - B^{rs,k}_j \leq M(1 - Y^{rs,k}), \forall r, s; i, j \in P^{rs,k}; (i, j) \in A; k = 1, 2, ..., K \] (6.3)

\[-(B^{rs,k}_i + l^{rs,k}_i - d_{ij} - B^{rs,k}_j) \leq M(1 - Y^{rs,k}), \forall r, s; i, j \in P^{rs,k}; (i, j) \in A; k = 1, 2, ..., K \] (6.4)

\[ \sum_{r,s \in N} \sum_{k} l^{rs,k}_i \delta^{rs,k}_i \leq MX_i, \forall i \in N \] (6.5)

\[ \sum_k Y^{rs,k} \leq My^{rs}, \forall r, s \] (6.6)

\[ y^{rs} \leq \sum_k Y^{rs,k}, \forall r, s \] (6.7)

\[ Z_r = \frac{s}{n_r}, \forall r \] (6.8)

\[ \sum_i C_i X_i \leq m \] (6.9)

\[ B^{rs,k}_i = \beta_0, \forall r, s; i \in R; k = 1, 2, ..., K \] (6.10)

\[ X_i = \{0, 1\}, \forall i \in N \] (6.11)

\[ Y^{rs,k} = \{0, 1\}, \forall r, s; k = 1, 2, ..., K \] (6.12)

\[ y^{rs}, \forall r, s \] (6.13)

\[ B^{rs,k}_i \geq 0, l^{rs,k}_i \geq 0, \forall r, s; i \in P^{rs,k} \] (6.14)

The objective of the model is to maximize the expected coverage of demand cities. Constraint set (2) assures that the total onboard electricity each vehicle carries will not exceed the EV battery capacity \( (B^{rs,k}_i + l^{rs,k}_i \leq \beta) \) on those paths \( k \) that are taken to electrify; otherwise no restriction exists when \( Y^{rs,k} = 0 \). Constraints (6.3) and (6.4) work simultaneously to ensure that the energy consumption conservation \( B^{rs,k}_i + l^{rs,k}_i - d_{ij} - B^{rs,k}_j = 0 \) holds for all links traversed on the \( k^{th} \) path which is taken to deploy adequate stations \( (Y^{rs,k} = 1) \). Otherwise, if \( Y^{rs,k} = 0 \), then \( B^{rs,k}_i + l^{rs,k}_i - d_{ij} - B^{rs,k}_j \leq M \), namely no restraining effects. Constraints (6.5) implies a
logic that refueling is only available at node $i$ if there is a charging station. Constraints (6.6) and (6.7) establish the relationship between $Y_{rs,k}$ and $y_{rs}$. Constraints (6.8) are the definition of O-D based coverage for each node $r$. Budget is indicated by constraint (6.9). Constraints (6.10) represent the initial on battery status of all EVs. The remaining are binary and nonnegativity constraints on variables.

### 6.2.3 A heuristic based on genetic algorithm

Because of the introduction of round trip, the number of variables and constraints are increased exponentially, thus, it is hard to solve the problem even for a medium size network like Sioux Falls network. In this section, I adopt the genetic algorithm to solve the proposed problem to an acceptable solution within a reasonable time. For completeness, the major procedures are reported and explained.

**Representation and fitness function**: A 0-1 array is used to represent location decisions over the transportation network, in which each bit of the array is notated as $X_i$ to indicate whether or not to place a charging stations at a particular node of the network. For example $X_i = 1$ implies to build a charging station at node $i$.

**Parent selection, crossover, mutation and population replacement**: In this study, I adopt the binary tournament parent selection and the fitness-based crossover method used in [10]. In addition to generating child solution by crossover, another child solution could be constructed by selecting the worst solution from the population and flipping the value for one of the bits of the solution according to the relationship of a random value and the mutation rate $p$. For example, if the random value generated within 0 and 1 is less than $p$, then the value of the bit is flipped. After generating $M$ child solutions, the worst $M$ solutions in the population will be
replaced with the M child solutions, and through this process, I am expecting the quality of the population is improved and best solution of the population would also improve.

Finally, I include complete and detailed procedures of the heuristic as follows.

Step 1. Initialization: set values for parameters $M$, $N$, and $p$

Step 2. While $i \leq N$ do

(a) repeat following steps generate M child solutions

- parent selection
- crossover
- mutation (use $p$ as mutation rate)
- fix location decisions then solve the original problem by CPLEX to get the fitness value for the corresponding child solution

(b) Replace the worst $M$ solutions in the population with child solutions generated in the generation

Step 3. Return the best solution in the population as the final solution to the original problem

6.3 Results and discussions

In this section, I applied the proposed P-MPRLM on the Sioux Falls network [57] to demonstrate the applicability of the model. The Sioux Falls network, as shown in Figure 6.2 consists of 24 nodes and 76 links with distance labeled on the links. In this study, all nodes in the network are considered as origins and destinations of EV travel demand and candidate sites to locate charging stations. For each node, its
probability of becoming an EV adopter is randomly generated, as shown in Table 6.1. For each O-D pair, the first three shortest paths (K=3) are considered as candidate paths to be electrified to complete trips between that O-D pair. To simplify the tests, I assume that all vehicles are homogeneous and all vehicles are assumed to be fully charged at origins and the driving range of vehicles is set to be 100. Programmed in MATLAB, the heuristic is applied to solve the problem over 12 budget scenarios (from locating one charging station to 12 charging stations which provides full coverage for all cities). The heuristic solutions are compared with exact solutions obtained by CPLEX 12.6 running on a desktop with 8GB RAM and Intel Core i5-2500@3.30GHz processor under Windows 7 environment.

![The Sioux Falls network](image)

**Figure 6.2:** The Sioux Falls network

Figure 6.3 reports the optimality gap for 12 scenarios. For each budget scenario (number of stations), I run the heuristic 50 times and compute the average optimality gap between heuristic solutions and the optimal solution obtained by CPLEX. From Figure 6.3, I find that the heuristic works very well, since the largest average gap among the 12 scenarios is only 1.9%. In addition, for most of the scenarios the gap
Table 6.1: Probability of adopting EVs

<table>
<thead>
<tr>
<th>Node ID</th>
<th>Probability</th>
<th>Node ID</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.7689</td>
<td>13</td>
<td>0.7900</td>
</tr>
<tr>
<td>2</td>
<td>0.1673</td>
<td>14</td>
<td>0.3185</td>
</tr>
<tr>
<td>3</td>
<td>0.8620</td>
<td>15</td>
<td>0.5431</td>
</tr>
<tr>
<td>4</td>
<td>0.9899</td>
<td>16</td>
<td>0.0900</td>
</tr>
<tr>
<td>5</td>
<td>0.5144</td>
<td>17</td>
<td>0.1117</td>
</tr>
<tr>
<td>6</td>
<td>0.8843</td>
<td>18</td>
<td>0.1363</td>
</tr>
<tr>
<td>7</td>
<td>0.5880</td>
<td>19</td>
<td>0.6787</td>
</tr>
<tr>
<td>8</td>
<td>0.1548</td>
<td>20</td>
<td>0.4952</td>
</tr>
<tr>
<td>9</td>
<td>0.1999</td>
<td>21</td>
<td>0.1897</td>
</tr>
<tr>
<td>10</td>
<td>0.4070</td>
<td>22</td>
<td>0.4950</td>
</tr>
<tr>
<td>11</td>
<td>0.7487</td>
<td>23</td>
<td>0.1476</td>
</tr>
<tr>
<td>12</td>
<td>0.8256</td>
<td>24</td>
<td>0.0550</td>
</tr>
</tbody>
</table>

is close to or less than 1%. I also report the average solving time of the heuristic and compare with CPLEX solving time for each scenario in Figure 6.4b. Unlike CPLEX which observes exponential increase in solving time as number of stations increases, the heuristic consumes much less time than CPLEX to solve the same problems and the solving time by the heuristic is increased almost in a linear fashion. For example, to locate 12 charging stations which is the first time full coverage is observed for all cities when VR=100 and K=3, CPLEX took 11862 CPU seconds to solve the problem, while the heuristic on average would only take 380 CPU seconds (3.2% of what CPLEX needs).

![Figure 6.3: Optimalty gap](image)

I take the scenario of locating 7 charging stations as an example to present
the detailed location, route choice and charging schedule decisions on those selected paths. As shown in Figure 6.5, charging stations are placed at nodes highlighted in green. For O-D pair (1,21), instead of electrifying the shortest path $1 \rightarrow 3 \rightarrow 12 \rightarrow 13 \rightarrow 24 \rightarrow 21$, the second shortest path $1 \rightarrow 3 \rightarrow 12 \rightarrow 11 \rightarrow 14 \rightarrow 23 \rightarrow 24 \rightarrow 21$ is electrified to complete trips from node #1 to node #21 then back to node #1. Detailed charging schedule along the whole trip is illustrated in Figure 6.6, where the height of an orange bar represent remaining range when arriving at a particular node while the height of a green bar is the among of range charged at a node. And the total height of orange and green bar is the range available on board when a driver departs from a node. It is noted that multiple charging stops are needed to complete the trip from node #1 to node #21 and the round trip.

In further, I examined the impacts of vehicle range, initial battery status and
Figure 6.5: Detailed location and route choice decisions for VR=100 and K=3

Figure 6.6: Detailed charging schedule for O-D pair (1,21)

EV adoption probabilities. Figure 6.7 reports the objective values of three vehicle ranges, i.e., VR=100, 150 and 200 for all the budget scenarios. Given the vehicle range, the marginal benefit on improving the weighted total coverage is diminishing as number of stations increases. And this observation becomes more obvious for larger vehicle ranges. In addition, for each vehicle range, there exists a critical number of stations (7 for VR=150 and 200 and 11 for VR=100) that further increasing the number would result in trivial or even no coverage improvement. In other words, this might provide some insight for EV manufacturers to determine the best vehicle range if the number of stations are given and stations are rolled out in the network.
by adopting this proposed model.

In Figure 6.8 I attempt to analyze the impact of initial battery status by assuming that EVs are half charged at origins and solve for the 12 budget scenarios with VR=100 and K=3. Compared with the corresponding results of full charge at origins, the total weight coverage is lower by the difference is getting smaller as number of stations increases. For example, if 8 or more stations are located in the network, the relative difference of total weighted coverage between fully and half charged initial battery status would be less than 5.37%.

I also pick node #4 with the highest probability 0.9899 and node #24 with the lowest probability 0.0550 to demonstrate how coverage of individual cities evolve over different budget scenarios. In general, the coverage of each city would increase when more stations can be located, as shown in Figure 6.9. In the meantime, because of higher priority (probability) to attract resources, the coverage of node #4 is increased dramatically from 9% to 57% as number of stations increases from 1 to 3. In contrast, node #24 which has the lowest EV adoption probability sees no improves in coverage. However, coverage of node #24 is greater than node #4 when only one station is available in the network. This might be explained by the fact that more surrounding

![Figure 6.7: Impact of budget and vehicle range](image-url)
nodes are within the vehicle range 100 for node #24 than for node #4. In other words, given the vehicle range and station deployment, more O-D pairs originating from node #24 can be covered without the need of recharging. It is also noted that coverage of node #4 drops as number of stations improved from 7 to 8, while at the same time coverage of node #24 increases. This observation perfectly illustrate how optimization is working on a system level, i.e., in order to improve total system performance (objective value) some cities may sacrifice with compensation in improved coverage of other cities.

Finally, I also solve the problem with equal probabilities, i.e., $p_r = 1$, for all cities in the network for the scenario of 7 stations to demonstrate the effectiveness of incorporating EV adoption probability in siting charging stations. Figure 6.10 shows the detailed deployment of charging stations in the network and route choice for O-D pair (1,21), which are different from the counterparts in Figure 6.5. In Table 6.2 I compared coverage ($Z_r$) with and without differentiating effects of probability for cities with high ($\geq 0.80$) or low ($\leq 0.20$) EV adoption probability when $VR=100$ and $K=3$. Drop in coverage is observed for 6 out of the 9 cities with low probability when EV adoption probability is incorporated in the objective function. Except for node
#6, coverage for other cities with high probability are either improved or maintained. All these observations justified the incorporation of EV adoption probability of cities into location optimization of charging stations. As expected, with a given budget high probability cities are emphasized and experiencing improvement in coverage while low probability cities are de-emphasized and experiencing drop in coverage when probability information is integrated in optimization.

**Figure 6.9:** Coverage ($Z$) for node #4 and #24 when VR=100 and K=3

**Figure 6.10:** Detailed location and route choice decisions for VR=100 and K=3 with equal probability
**Table 6.2:** Comparison of coverage for cities with high ($\geq 0.80$) or low ($\leq 0.20$) probabilities

<table>
<thead>
<tr>
<th>Node ID</th>
<th>Probability of adopting EVs</th>
<th>Coverage probability</th>
<th>Coverage equally treated</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>0.0550</td>
<td>61%</td>
<td>78%</td>
</tr>
<tr>
<td>16</td>
<td>0.0900</td>
<td>87%</td>
<td>87%</td>
</tr>
<tr>
<td>17</td>
<td>0.1117</td>
<td>83%</td>
<td>87%</td>
</tr>
<tr>
<td>18</td>
<td>0.1363</td>
<td>83%</td>
<td>91%</td>
</tr>
<tr>
<td>23</td>
<td>0.1476</td>
<td>74%</td>
<td>78%</td>
</tr>
<tr>
<td>8</td>
<td>0.1548</td>
<td>87%</td>
<td>87%</td>
</tr>
<tr>
<td>2</td>
<td>0.1673</td>
<td>78%</td>
<td>83%</td>
</tr>
<tr>
<td>21</td>
<td>0.1897</td>
<td>74%</td>
<td>87%</td>
</tr>
<tr>
<td>9</td>
<td>0.1999</td>
<td>87%</td>
<td>87%</td>
</tr>
<tr>
<td>12</td>
<td>0.8256</td>
<td>96%</td>
<td>91%</td>
</tr>
<tr>
<td>3</td>
<td>0.8620</td>
<td>87%</td>
<td>87%</td>
</tr>
<tr>
<td>6</td>
<td>0.8843</td>
<td>83%</td>
<td>87%</td>
</tr>
<tr>
<td>4</td>
<td>0.9899</td>
<td>91%</td>
<td>87%</td>
</tr>
</tbody>
</table>
6.4 Summary

This chapter presents a probability-based multi-path refueling location model that incorporates EV adoption probabilities of demand cities, vehicle range, users deviation choice and round-trip between O-D pairs to maximized the total weighted coverage of all demand cities with a given budget. A genetic algorithm based heuristic is adopted to help efficiently solve the proposed problem. Numerical experiments are implemented the Sioux Falls network to demonstrate the model and heuristic, and sensitivities analysis are conducted to examine the impacts of vehicle range, initial batter status, and EV adoption probability. The results indicate that by integrating probability information in the model, cities with high and low potential of adopting EVs would experience improvement and decreases in coverage, which would provide insights to better utilize the limited resources.
Chapter 7

Conclusions

7.1 Summary of dissertation

This dissertation attempts to provide one such framework that is believed to offer potential in integrating vehicle range, deviation choice, charging schedule and topological dynamics of network with emerging EV markets to the flow-based location model for EV charging stations. With the proposed efficient solution methods, the framework is applicable to real world large scale cases as demonstrated on real world case studies of South Carolina. Extensive results indicates that the deployment plans by adopting the proposed multi-period model is not only cost effective to satisfy all the intercity O-D trips within a network, but also be able to provide high quality performance (coverage) under stochastic demand scenarios.

In Chapter 3, the multi-path refueling location model which innovatively integrates vehicle range and deviation choice into the process of determining the optimal deployment strategy for EV charging stations is presented and tested on two numerical examples. The results revealed the importance of vehicle range and the advantage of considering deviation paths for deciding the best locations of charging stations.
By allowing EV drivers to deviation from those shortest or pre-defined paths to get charged and complete O-D trips, less charging stations are needed to ensure all the O-D trips can be completed by EVs. In addition, critical vehicle range and locations for both test networks, and that may be interested to EV manufacturers and policy makers to trade-off between improving vehicle range and building more charging stations.

To solve the multi-path refueling location model effectively, two heuristics based on greedy adding algorithm are created in Chapter 4. Different node weighting strategies are explored to incorporate the impact of vehicle range and deviation choice on location decisions. Besides, a node substitution and a solution refining procedure are included to further improve solution quality. Convinced by results on the Sioux Falls network, the proposed heuristics are applied to a real world case study of South Carolina, and the results demonstrate the applicability of the multi-path model and heuristics for design charging station infrastructure networks on large transportation networks.

Chapter 5 extends the multi-path refueling location model to capture spatial expansion of EV markets over time stages. Comparison with the myopic solutions was conducted to show the advantage of taking into account spatial network dynamics of EV markets. A genetic algorithm based heuristic is developed to solve this multi-period multi-path refueling location model, and test results on both the Sioux Falls network and the South Carolina network shows great effectiveness and efficiency. The results indicate that the multi-period model helps reduce total cost of establishing the charging station network over time and yields a station deployment which has the potential to high quality performance (coverage) under stochastic demand scenarios.

Chapter 6 presents another extension of the multi-path refueling location model to consider the impact of each city’s potential of adopting EVs and the im-
pact of introducing round trip on charging station deployment strategies. A genetic algorithm based heuristic is adopted to solve the proposed model probability-based multi-path refueling location model. Tests on the Sioux Falls network justified the importance of integrating EV adoption probability information in optimally siting charging stations on a network.

7.2 Limitations of dissertation and future works

Although the proposed models are mathematically correct and justified by numerical experiments and case studies, it is still questionable whether the models can effectively represent the real world application, such as in practice what would be the actual cost of establishing the charging network and how well the deployed charging stations would service EV travel demand. This is mainly due to the lack of historical cases and data on designing EV charging station network. In the future, as EV market gets mature and more charging stations are located in practice, more real cases and data will be available to support the validation of the models.

Another limitation is that all vehicles and drivers are assumed to be homogeneous. Therefore, due to the difference in travel range of EVs and preference of EV drivers, it is possible that the proposed models may resulting deployments which might increase inconvenience for user of EVs with small travel range. In future, the proposed models can be extended to consider different groups of EVs and drivers to make it more realistic.

There are several other extensions that could enrich the context of this research. First, in this study the probability of each city becoming an EV adopter is assumed to be static and fixed throughout the planning horizon. A more realistic, time-dependent probability assessment of future demand would be essential for deter-
mining best possible phasing intervals. The results will offer rich insights on public policies, such as the critical timing to cease the tax incentives on EVs. Second, this study simplifies the modeling by assuming that the trajectory of expansion of future demand cities is known, based on the projected probabilities as a result of statistical analysis. In reality, the major issue be the stochasticity embedded in future demand. In other words, there are multiple possible trajectories depending on the multivariate socioeconomic data used. How to incorporate this stochasticity into modeling of infrastructure system expansion is a challenge in both modeling and solution and has been well noted in prior studies [73, 6, 7, 67, 2, 33, 28]. Depending on the nature, the problem may be formulated as a multistage stochastic program and solved by a nested decomposition method or reformulated as a dynamic programming problem and solved by approximate dynamic programing method [68]. Thirdly, when a massive adoption of EV is realized and EV flow on the roadway network can be readily predicted, the EV traffic flows can be incorporated in the future modeling practice and explicit station capacity design (e.g., number of chargers to be placed at each station) can be included in the model as an integer variable. A bi-level optimization framework can be used to incorporate traffic flow at lower level while the upper-level will be locational decisions. It leads to a mathematical problem with equilibrium constraints (MPEC) problem if traffic equilibrium is sought in the lower level. In terms of solution methods, I will investigate a new method to decompose the set of decisions to a set of charging station locations first, followed by the construction of routes between O-D pairs. Lastly, a charging station might get congested once EV market gets mature, and in this situation smart pricing on charging can be implemented to see how EV drives would response and further how traffic flow in the network would be affected.
Bibliography


109


