Infrastructure Design for Electric and Autonomous Vehicles

by

Ömer Burak Kınay

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Management Sciences

Waterloo, Ontario, Canada, 2022

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner:	Stefan Bock Professor, Schumpeter School of Business and Economics University of Wuppertal
Supervisors:	Sibel Alumur Alev Associate Professor, Department of Management Sciences University of Waterloo
	Fatma Gzara Professor, Department of Management Sciences University of Waterloo
Internal Members:	Fatih Safa Erenay Associate Professor, Department of Management Sciences University of Waterloo
	Hossein Abouee Mehrizi Associate Professor, Department of Management Sciences University of Waterloo
Internal-External Member:	Nasser Lashgarian Azad Associate Professor, Department of Systems Design Engineering University of Waterloo

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

Chapter 2 of this thesis is based on the publication Kınay et al. (2021). Chapter 3 is a submitted paper under review in Kınay et al. (2022). Both of these chapters have been co-authored by myself and my supervisors, Dr. Fatma Gzara and Dr. Sibel Alumur Alev.

I am the sole author of the rest of this thesis.

Abstract

This thesis focuses on infrastructure design for the disruptive transportation technologies of electric vehicles (EVs) and autonomous vehicles (AVs) to enable their adoption at large scale. Particularly, two EV-related problem frameworks concerning the spatial distribution of charging stations and their respective capacity levels are studied, and a new problem is introduced to determine the optimal deployment of AV lanes and staging facilities to enable shared autonomous transportation in urban areas.

The first problem is centered around determining optimal locations of fast-charging stations to enable long-distance transportation with EVs. A new mathematical model is developed to address this problem. This model not only determines optimal facility locations but also finds optimal routes for every origin-destination (OD) trip which follows the path that leads to the minimum total en route recharging. Through computational experiments, this model is shown to outperform the widely used maximum and set cover problem settings in the literature in terms of several routing-related performance measures. A Benders decomposition algorithm is developed to solve large-scale instances of the problem. Within this algorithm, a novel subproblem solution methodology is developed to accelerate the performance of the classical Benders implementation. Computational experiments on real-world transportation networks demonstrate the value of this methodology as it turns out to speed the classical Benders up to 900 times and allows solving instances with up to 1397 nodes.

The second problem extends the previous one by seeking to determine EV charging station locations and capacities under stochastic vehicle flows and charging times. It also considers the route choice behavior of EV users by means of a bilevel optimization model. This model incorporates a probabilistic service requirement on the waiting time to charge, and it is studied under a framework where charging stations operate as M/M/c queuing systems. A decomposition-based solution methodology, that uses a logic-based Benders algorithm for the location-only problem, is developed to solve the proposed bilevel model. This methodology is designed to be versatile enough to be tailored for the cooperative or uncooperative EV user behavior. Computational experiments are conducted on real-life highway networks to evaluate how service level requirements, deviation tolerance levels, and route choice behavior affect the location and sizing decisions of charging stations.

The third problem entails the staging facility location and AV lane deployment problem for shared autonomous transportation. The proposed problem aims to find the optimal locations of staging facilities utilizing a bi-objective model that minimizes total travel distance and the total AV travel not occurring on AV lanes with respect to a given AV lane deployment budget and a number of staging facilities to locate. A Benders decomposition algorithm with Pareto-optimal cuts is developed and the trade-offs with optimal solutions on benchmark instances are evaluated. Computational experiments are performed to analyze the effects of AV lane budget, staging facility count, and the objective preferences of decision makers on optimal solutions.

Acknowledgements

This thesis would not have been possible without the inspiration and support of many wonderful individuals — my thanks to all of them for being part of this journey.

First of all, I owe my deepest gratitude to my advisors Dr. Sibel Alumur Alev and Dr. Fatma Gzara for their continuous guidance and patience. Their valuable insights and wisdom have always pointed me to the right direction at any juncture I faced during my studies. They provided me the room to be creative, yet, they knew how to put me back on track when I lost my way. I could not have imagined having better mentors for my doctoral journey, I am thankful that our paths crossed.

I thank the examining committee members for their interest, time and effort spent in reading this thesis, and providing helpful feedback as well as the reviewers of the two manuscripts submitted to journals for their constructive advice on improving the quality of the papers and this thesis.

To my many friends — especially Mustafa, Alp, Deniz, Merve, Semih, Emrecan, Zeynep, Onur, Basak, Betul, Esma and Mesut, as well as my bike and Waterloo Cycling Club; thank you for keeping me sane and being there when I needed you the most. You should know that your presence means the world to me.

To my mom and dad; words fail to express my feelings of gratitude for you. You have been my teachers in life and my compass of righteousness. No matter how much physical distance we have between us, feeling your unconditional love and trust gives me the strength to keep going. Thank you for selflessly providing me the opportunities that made me who I am today.

Finally, to my lovely wife, Sinem; I never would have made it without your loving support and understanding. Thank you for believing in me even when I didn't believe in myself. Thank you for being the joy of my life, my proofreader, editor, psychologist and even debugger. But above of all, thank you for being my best friend. I will eternally be indebted to you. Now, it is time to celebrate; besides your own, you also earned this degree right along with me.

Dedication

This is dedicated to my parents.

No matter where I go in this life and beyond,
I know that where you are will always be home.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Increased awareness of extensive use of nonrenewable energy resources as well as the globally recognized impacts of climate change provoked worldwide interest in environmental sustainability in the 21^{st} century. In order to achieve a cleaner and more sustainable future, development of zero-emission vehicles (ZEVs) became vital since passenger and freight road vehicles account for over 17% of global direct CO_2 emissions (IEA, 2021).

The Electric Vehicles Initiative, which is established by fifteen countries including Canada, China, Germany, and the United Kingdom, has set an aspirational goal of reaching 30% sales share of electric vehicles (EVs) by 2030. The United States declared their goal of achieving 50% of all new passenger cars and light trucks sold in 2030 to be ZEVs (The White House, 2021). The majority of the EU countries and the United Kingdom aim 2025-2035, and Canada aims 2040 to reach 100% ZEV sales (IEA, 2021). As for the private sector commitments towards electric mobility, the Climate Group's EV100 Initiative brings together over 100 companies dedicated to making electric transport the new normal by 2030 (The Climate Group, 2021a). Many corporations are aiming to speed up the transition to electric mobility by converting their fleets to EVs and designing private networks by locating charging stations. Despite these global targets, initiatives and endeavors, significant adoption barriers for EVs proliferation remain. In particular, the EV100 members

reported the lack of charging infrastructure as their top concern (The Climate Group, 2021b). Many developing countries are yet to build a network of fast-charging stations to enable long-distance transportation with EVs. Recent technological advancements also put heavy duty EVs under the spotlight for long-haul deliveries and the success of such a shift in freight transportation is particularly dependent on the availability of effective charging infrastructure.

In addition to the pursuit of promoting electric mobility, the development of enabler technologies such as the internet of things (IoT), big data, and artificial intelligence have been paving the way towards another disruptive technology in transportation: autonomous travel. Autonomous vehicles (AVs) are contended to revolutionize how we think of mobility by mitigating the urbanization-related trends that exacerbate congestion in cities. They also have the potential to drastically reduce fatalities by improving traffic safety, promote environmental sustainability by decreasing emissions via reducing private vehicle ownership and vehicle-miles-traveled, and offer convenience by reducing the drivers' value of time (Alcorn and Kockelman, 2021; Jones and Leibowicz, 2019; Brownell and Kornhauser, 2014). Besides these aspects, this synergy may reshape the need for parking spaces, which could transform the land use in cities (Fagnant and Kockelman, 2015; Bagloee et al., 2016). Experts hypothesize that AVs are more likely to be initially introduced to the urban transportation market by providing shared mobility services since they are expected to be significantly costlier than human-driven vehicles (HV) in the early adoption phases (Bansal and Kockelman, 2017; Shaheen and Cohen, 2019). Widespread shared mobility services utilizing AVs could help use transportation resources more effectively and provide a dramatic reduction in the number of privately owned passenger vehicles (Golbabaei et al., 2021). Successful and cost-effective integration of shared AV fleets in urban transportation depends on strategic infrastructure deployment decisions. Particularly, the deployment of AV lanes to exploit the benefits of infrastructure-supported autonomy and locating staging

facilities that would help reduce cruising congestion are two key infrastructural decisions to be addressed.

In this thesis, motivated by these aspects, infrastructure design for electric and autonomous vehicles is studied. The aim is to develop analytical models and methods that would determine optimal infrastructure decisions to promote the appeal of these vehicles by improving the convenience of their users.

Chapter 2 presents a new full cover modeling framework to design charging station infrastructure, where the focus is on locating fast-charging stations for battery EVs to enable long-distance transportation. A mathematical model is introduced to determine the optimal locations of these charging stations so that every origin-destination trip on a given transportation network is covered with respect to vehicle range. This full cover model allows deviations from the shortest paths and also determines an optimal route for each trip that requires the minimum total en route recharging. Two variants of this model are proposed: one that minimizes the total cost of locating charging stations and total en route recharging, and another that determines the locations of a predetermined number of stations to minimize the total en route recharging. Computational experiments performed on benchmark data sets validate that the proposed full cover models perform better than the maximum or set cover problem settings in the literature in terms of routing-related measures, such as total trip distance and maximum deviation from the shortest paths. A Benders decomposition algorithm is developed to optimally solve real-life instances of the problem. The Benders subproblem is identified as a many-to-many shortest path problem with an additional constraint that restricts the nodes that can be used to open facilities that are determined by the master problem. A new algorithmic methodology is developed to construct the dual solution for this subproblem and to generate non-dominated optimality cuts and strong valid inequalities for feasibility cuts. This novel algorithm is shown to accelerate the performance of the Benders algorithm up to 900 times over the tested large-

size instances. The main contributions of this work are *(i)* a new objective function that has not been considered by similar studies in the field, *(ii)* a new modeling approach to find optimal routes of OD trips which does not require an exogenous shortest path deviation tolerance, *(iii)* an efficient Benders decomposition algorithm to solve large-scale instances, and *(iv)* a novel methodology to solve the Benders subproblem.

Chapter 3 addresses the problem of determining the strategic locations of charging stations and their capacity levels under stochastic EV flows and charging times while taking the route choice behavior of users into account. The problem is modeled using bilevel optimization where the network planner or leader minimizes the total infrastructure cost of locating and sizing charging stations while ensuring a probabilistic service requirement on the waiting time to charge. EV users or followers, on the other hand, minimize route length and may be cooperative or non-cooperative. Their choice of route in turn determines the charging demand and waiting times at the charging stations and hence the need to account for their decisions by the leader. The bilevel problem reduces to a single-level mixed-integer model using the optimality conditions of the follower's problem when charging stations operate as M/M/c queues and followers are cooperative. To solve the bilevel model, a decomposition-based solution methodology is developed which uses a new logic-based Benders algorithm for the location-only problem. Computational experiments are performed on benchmark and real-life highway networks, including a new Eastern U.S. network, and the impact of route choice behavior, service requirements, and the deviation tolerance on the location and sizing decisions are analyzed. The analysis demonstrates that more stringent service requirements increase the capacity levels at open charging stations rather than their number and that higher deviation tolerance solutions are less costly. Moreover, the difference between solutions under cooperative and uncooperative route choice is more significant when the deviation tolerance is lower. Core contributions of this work are *(i)* merging the bilevel nature of the problem with realistic queuing models to

incorporate stochastic travel demand, *(ii)* an exact solution algorithm for the uncapacitated version of the problem, *(iii)* and a novel decomposition based algorithm that is capable of rapidly finding high-quality solutions and also able to handle both cooperative (optimistic) and uncooperative (pessimistic) responses of EV users.

Chapter 4 investigates determining optimal deployment of AV lanes and AV staging facilities in urban transportation networks to enable shared autonomous mobility. Simultaneously considering these strategic infrastructure deployment decisions is ideal. This is because staging facility locations, which are smart mobility hubs required by shared AV operators to idle, will have a significant impact on roadways that AVs will utilise. Consequently, this will have a direct influence on AV lane deployment decisions. Defining this problem using a demand framework that originates from OD trip requests of service seekers, we propose a bi-objective optimization model that minimizes total travel distance and the total AV travel not occurring on AV lanes while a given AV lane deployment budget is met and a given number of staging facilities are located. A Benders decomposition algorithm coupled with a non-dominated cut generation scheme is proposed to solve this problem. Through computational experiments on a benchmark network, effects of AV lane budget, staging facility count, and objective preferences of decision maker(s) on optimal solutions are evaluated. The main contributions of this study are *(i)* a novel problem definition for infrastructure design of shared autonomous vehicles, *(ii)* a bi-objective mathematical formulation for this problem, *(iii)* a decomposition algorithm to solve large-scale instances, and *(iv)* extensive managerial insights analyzing trade-offs between AV lane deployment and staging facility location decisions as well as optimal designs.

The motivations and detailed definitions for each problem studied in this thesis are provided within their corresponding chapters along with the related literature reviews. Finally, Chapter 5 summarizes the key points of each chapter in this thesis, includes concluding remarks, and provides a brief discussion on directions of future research.

Chapter 2

Full Cover Charging Station Location Problem with Routing

Even though EVs seem to be a remedy for energy inefficiencies, sustainability and resiliency in road transportation, certain adoption barriers have been preventing their proliferation. The most prominent adoption barriers include their relatively high upfront ownership cost, shorter range, longer recharging time, and limited availability of charging infrastructure. Despite the recent technological developments and governmental incentives to increase their appeal, market penetration is still lower than the desired levels. A recent automotive consumer survey revealed that 83% of respondents from United States expressed the availability of fast-charging options on EV journeys as a major concern (Capgemini Invent Global, 2018). Many developing countries are yet to build a network of fast-charging stations to enable long-distance transportation with EVs. Recent technological advancements also put heavy duty EVs under the spotlight for long-haul deliveries and the success of such a shift in freight transportation is particularly dependent on the availability of an effective charging infrastructure.

In summary, the count and spatial distribution of refueling/recharging facilities still pose a major challenge for a widespread adoption of ZEVs, specifically when the focus is on long-distance travel or long-haul transportation with EVs. The EV trips of commuters or shorter distance travelers can readily rely only on the existing slow-charging options (home charging, work charging, shopping mall charging etc.). However, the success and convenience of long-distance trips will substantially depend on the en route fast-charging station availability and their spatial distribution. Therefore, strategically positioning these stations is one of the key aspects to enable convenient long-distance travel and to reach the next level in market penetration of these vehicles for a cleaner and greener future. In this work, we address the charging station location problem with a new full cover approach. Particularly, we propose a mathematical formulation that determines the locations of charging stations so that it is possible to complete every trip on a given transportation network without running out of energy. Simultaneously, the proposed model determines the optimal routes that require the least amount of en route recharging energy for each origin-destination (OD) trip. Two variants of this model are proposed: one that minimizes the total cost of locating charging stations and total en route recharging energy, and another that determines the locations of a predetermined number of stations to minimize the total en route recharging energy. We filter out shorter OD trips to fully focus on a long-distance problem setting, which is known to generate more challenging instances to solve.

The main contributions of the work presented in this chapter are *(i)* a new objective function that minimizes the total en route recharging, which has not been considered by similar studies in the field, *(ii)* a new modeling approach to find optimal routes of OD trips that eliminates the need to introduce an exogenous shortest path deviation tolerance, *(iii)* development of an efficient Benders decomposition algorithm to solve large-scale instances, and *(iv)* a novel methodology to solve the Benders subproblem. This novel methodology

constructs the dual solution for a feasible subproblem and generates the corresponding non-dominated optimality cuts. It also introduces a feasibility cut generation scheme, where strong inequalities are derived when the subproblem is infeasible.

Computational experimentation shows that the novel subproblem solution algorithm accelerates the performance of Benders decomposition up to 900 times compared with the case when subproblems are solved by CPLEX optimizer. In particular, this methodology solves all instances from the California road network with 339 nodes and 1275 OD pairs to optimality with an average computation time that is less than 20s. We also stress-test the solution algorithm and evaluate the limits of its computational performance on larger instances using two variants of the Germany road network with 818 and 1397 nodes.

In the next section of this chapter, we provide a review of the related literature and highlight the significant contributions to this research field. In Section 2.2, we discuss the modeling aspects of the full cover approach and provide two new mathematical formulations. In Section 2.3, we present the Benders reformulation and introduce the subproblem solution methodology that is developed to improve the computational performance of the Benders algorithm. In Section 2.4, we report the results of extensive computational experiments to evaluate the effectiveness of the full cover approach against the existing maximum or set cover approaches. In addition, we evaluate the computational performance of the proposed solution methodology. Finally, Section 2.5 provides some concluding remarks.

2.1 Literature Review

Earlier works on locating refueling service stations for gasoline retailing date back to late 1980s and use location-allocation models (Goodchild and Noronha, 1987). Research on locating these stations for alternative fuel vehicles started to receive attention after early

2000s (Ko et al., 2017; Alumur and Bektaş, 2019). There are two main streams of research that are characterized by the way the demand is modeled. The location-routing problem with intra-route facilities studies the case where node-based customers are the source of demand, and the need for refueling results from the vehicle tours. It is notably suitable for applications that consider short-haul distribution logistics (see *e.g.,* Schiffer and Walther (2017); Schiffer et al. (2019)). The other stream, which focuses solely on OD refueling demand, is geared towards longer range of travel or mid to long-haul transportation. A recent review paper by Shen et al. (2019) provides a synthesis of the research perspectives in this field.

The focus of this literature review will be on the latter stream that models demand between pairs of OD nodes. The *flow refueling location model* (FRLM) proposed by Kuby and Lim (2005) is the pioneering work in the field, where flow demand is assigned to use a shortest path for each OD pair. For such given paths, the objective is to maximize the number of feasible round-trips, or covered demand, by selecting a set of nodes to locate the refueling stations. Using this maximum covering location approach (max cover), they develop a mixed integer programming formulation that locates p facilities. The main assumptions of the generic FRLM are: uncapacitated stations, deterministic demand, identical vehicles, *a priori* known locations by drivers, and no deviation from the shortest path. An OD pair is covered if the trip can be completed in both directions. Kuby and Lim (2005) also introduce an assumption in order to ensure round trip feasibility based on the presence of a facility on the OD nodes: vehicles leave the origin with 50% fuel level if there is no facility located there and with 100% otherwise. Similarly, vehicles arrive at the destination with at least 50% fuel level in the absence of a facility located at this node.

The extensions to the FRLM include locating facilities on arcs (Kuby and Lim, 2007), maximizing total distance traveled by covered demand (Kuby et al., 2009), maximizing a threshold coverage where a node is covered if it exceeds a threshold percentage of its

overall outbound trips (Hong and Kuby, 2016), and a capacitated framework that limits the number of trips that can be refueled at an open facility (Upchurch et al., 2009).

The FRLM takes all feasible facility location combinations with respect to vehicle range for every shortest OD path as an input. Such a preprocessing task is computationally expensive and is even impractical for large-size problems. Therefore, heuristic algorithms are developed in the literature (Lim and Kuby, 2010; Tran et al., 2018). Capar and Kuby (2012), Capar et al. (2013), MirHassani and Ebrazi (2013) propose new reformulations to eliminate the pregeneration step and increase computational efficiency. MirHassani and Ebrazi (2013) point out that a vehicle will always use the shortest path segment between two consecutive stops for refueling. Based on this key observation, they propose a network transformation and a new formulation to the FRLM referred to as the *refueling station location problem* (RSLP). Additional improvements on the size of the solvable instances and solution times are obtained by Arslan and Karaşan (2016) by applying Benders decomposition on the formulation by Capar et al. (2013). The former work is under the context of plug-in hybrid vehicles and their objective is to maximize the distance traveled using electricity.

Kim and Kuby (2012) introduce the *deviation flow refueling location model* (DFRLM), that allows OD demand to deviate from its shortest path to be able to complete the trip. Like the generic FRLM, this study computes all feasible facility location combinations for every OD pair as input to the model. In a subsequent work, Kim and Kuby (2013) propose a network transformation heuristic in order to solve large-size instances of the DFRLM. The former study defines deviation paths of OD pairs exogenously whereas the latter uses a greedy substitution algorithm to generate deviation paths on the fly. Since deviation from the shortest path implies routing choices in the model, the same problem is also referred to as the *RSLP with routing* (RSLP-R) by Yıldız et al. (2016) and Arslan et al. (2019).

Yıldız et al. (2016) is the first study that allows deviation from the shortest path without pregenerating the routes. Instead, they use *deviation tolerance* to sustain a tractable problem size. They develop a path-segment formulation and solve it using a branch-and-price algorithm, which improves the solution times reported by Kim and Kuby (2012) considerably. Arslan et al. (2019) propose a tight formulation to the RSLP-R based on a notion called *q-node cuts* and solve it using a branch-and-cut algorithm. This methodology is shown to improve the computational performance as well as the size of the solvable instances compared to Yıldız et al. (2016). Göpfert and Bock (2019) propose a similar branch-and-cut algorithm to solve large scale instances. They evaluate the effects of filtering out shorter distance OD trips and show that this substantially raises the complexity of the instances to be solved.

There is an alternative research direction for the FRLM where a location set covering (set cover) approach is adopted to ensure that all OD pairs complete their trips while minimizing the number of located facilities. Wang and Lin (2009) satisfy the feasibility of an OD trip by tracking fuel level at every node of the route. This approach helps eliminate the path pregeneration process at the expense of defining excessively many decision variables. Wang and Wang (2010) propose a bi-objective extension to Wang and Lin (2009) by considering the minimization of total facility location costs simultaneously with the maximization of coverage of nodal demands. Li and Huang (2014) and Huang et al. (2015) extend the set cover model by allowing shortest path deviations, where the deviation paths are exogenously determined and fuel level is still tracked on every node. Another study with the set cover perspective is by Lin and Lin (2018), which defines the *p-center* RSLP with the objective to minimize the maximum deviation percentage of OD pairs.

There are several studies that incorporate demand or range uncertainty within the FRLM framework (*e.g.,* Hosseini and MirHassani (2015) de Vries and Duijzer (2017), Lee and Han (2017), Boujelben and Gicquel (2019)). None of these studies, however, consider

the shortest path deviations. On the other hand, Yıldız et al. (2019) take stochastic demand and capacitated facilities into account besides allowing deviations via pregeneration of the set of all possible stops for an OD pair considering an exogenous deviation tolerance.

Leitner et al. (2019) point out the relevance of the network design problem with relays in the context of refueling station applications. On a transportation network, relays correspond to the refueling stations extending the reach of the ZEVs; on telecommunication networks, relays are regenerators extending the reach of the signals. On both types of networks, an aim is to find the spatial distribution and the count of the facility locations. Leitner et al. (2019) also discuss that some solution methodologies can be used interchangeably. As a matter of fact, the formulations by Yıldız and Karaşan (2017) and Yıldız et al. (2018) for relay network design that consider regenerator location, routing, and bandwidth allocation have similarities with the RSLP-R formulations. Arslan et al. (2019) highlight that the drivers in the RSLP-R applications will not be willing to deviate more than a certain distance from their shortest path unlike the regenerator location applications, where there is no bound on the length that a signal deviates from its shortest path.

The widely used max cover approach in the literature may lead to building ineffective networks that leave many OD pairs uncovered (Arslan et al., 2019). However, the complete coverage of every possible OD trip on a given transportation network is crucial to ensure that the availability of charging station infrastructure is no longer a barrier against proliferation of EVs. In addition, a framework that allows deviation from the shortest paths implies the importance of routing decisions. Yet, none of the existing studies take the optimal routes into account while determining the facility locations. Incorporating routing based performance measures into the problem setting will yield better solutions. Besides, when shortest path deviations are allowed, both the max cover and set cover approaches proposed in the literature are ineffective beyond 100% coverage. However, it is important to determine the best possible routes even when the number of located facilities is suffi-

cient to ensure 100% coverage. These issues are elaborated while describing the modeling approach in the upcoming section.

2.2 Problem Definition and Mathematical Formulation

In this section, we first discuss the drawbacks of using the existing max cover approach in the RSLP-R models and present the basics of the full cover approach. The formulations for the full cover RSLP-R, which use an expanded network as a basis, are presented afterwards.

2.2.1 Full Cover Approach

In contrast with no shortest path deviation assumption of Kuby and Lim (2005), scarcity of refueling facilities on a transportation network may force drivers to deviate from their shortest paths to be able to complete a trip. Therefore, taking shortest path deviations into account in this problem setting is vital as is the case in the DFRLM/RSLP-R. The RSLP-R (Yıldız et al., 2016) finds the locations of p facilities to maximize the total flow of covered OD pairs. It enforces hard constraints on the allowable deviation tolerance from the shortest path lengths. This problem is referred to as the max cover problem (MCP) hereafter.

Among the uncovered OD pairs of a MCP solution, there are those that would have been covered if the deviation tolerance is relaxed. However, there may be others that would remain uncovered due to vehicle range. The results of Arslan et al. (2019) reveal that an OD demand is more likely to be uncovered due to vehicle range than due to the deviation tolerance when using the RSLP-R. It is important to ensure that there are no uncovered

OD pairs due to range limitations as full connectivity is essential for the proliferation of EVs. Consumer confidence in these vehicles is bound to increase if they know that the spatial distribution of charging stations allows them to complete every possible trip on a given transportation network.

Although the MCP allows for deviations, it ignores the cost of flows, *i.e.,* the significance of routing decisions. To highlight this issue, consider alternative optimal solutions with the same number of open facilities. One solution may dominate others with respect to routing based performance measures such as total trip distance, and the maximum deviation length. This is particularly important when p is sufficiently high to provide 100% coverage. In this case, both the MCP and the set cover approach fail to differentiate between alternative optimal solutions with respect to such routing based performance measures.

The full cover approach that we introduce ensures that every OD pair on a given transportation network is able to complete its trip, while simultaneously minimizing the total cost of locating charging stations and the total en route recharging energy. This problem is referred to as the *full cover problem* (FCP) hereafter. This setting may easily be modified to include the minimization of only the total recharging amount while p facilities are located (to be referred to as the *p-FCP*).

Unlike other studies addressing the DFRLM/RSLP-R, we do not pregenerate the deviation paths or include any constraints to restrict the deviation tolerance, but rather minimize the total en route recharging required to complete each trip. Not having a predetermined deviation tolerance makes it more challenging to solve large-scale instances in reasonable computational time. If one considers the deviation tolerance as an important aspect of the problem, proposed formulations can easily be modified to incorporate this.

Considering that the recharging times required for EVs as well as the associated cost of en route recharging can be significant, we believe that always forcing fully recharging at

each station visit may not be desirable since a driver's urge to reach their destination as fast as possible would encourage them to charge only up to an amount that is sufficient to reach the subsequent stop. This partial recharging policy, to be referred hereafter as the charge-just-enough policy, not only minimizes the total energy required to complete an OD trip, but it may also find a better solution in terms of the total trip time. Figures 2.1 and 2.2 demonstrate why total recharging required to complete a trip should be considered in addition to trip distance. Figure 2.1 shows an OD path from node A to node D. The arc labels represent the percentage of vehicle range (R) required to traverse each arc. Given the assumption on charge level at origin and destination nodes, at least two facilities should be established in order to make both directions of this trip possible. Figure 2.2 illustrates two feasible solutions on the path shown in Figure 2.1. Circle-shaped nodes represent the nodes visited along the path and square-shaped nodes represent the locations of the stations. Percentage values above each node show the battery level of the vehicle. In case of recharging, entering and leaving battery levels are shown above the node and are separated by a right arrow ($\rightarrow$). The recharged amount is indicated under the node. For Figure 2.2(a), the trip from A to D includes two en route recharging stops at nodes B and C with a total of 140% of R. For the same trip, this value is only 90% for the solution presented on Figure 2.2(b), which locates recharging facilities at nodes B and D. Although the trip length in distance is the same for both, the one on Figure 2.2(b) will possibly be faster as it needs less recharge to complete the trip. Therefore, it outperforms the solution on Figure 2.2(a) with respect to these measures. Hence, it is important to come up with a framework that distinguishes such solutions.

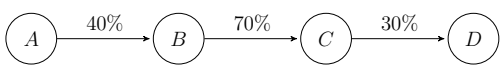
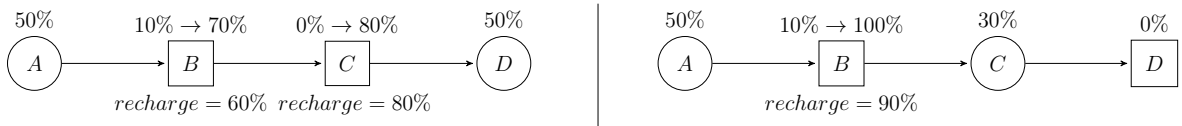


Figure 2.1: An example path with four nodes.



(a) Feasible Solution 1 (stations at B and C)

(b) Feasible Solution 2 (stations at B and D)

Figure 2.2: Analysis of feasible solutions with two charging stations.

As a result of the charge-just-enough policy, the total recharging required to complete a trip may be calculated as a function of the trip distance and battery level at the origin and destination nodes. It is not necessary to keep track of the battery level at intermediary nodes of an OD route. The total en route recharging may be calculated by summing up the total distance and battery level at the destination and subtracting the battery level at the origin. The distance and battery level are defined as a percentage of R. For the path on the left in Figure 2.2, total recharging required during the trip is $140 + 50 - 50 = 140\%$, and for the path on the right, it is $140 + 0 - 50 = 90\%$. In this work, we treat the forward and backward directions of a round trip independently, i.e., they do not happen in tandem. Long-distance trips tend to include overnight stays at destination where recharging will not have an impact on the actual travel duration. In that case, total en route recharging is the same in both directions of a round trip. Next, we present the mathematical formulation of the FCP.

2.2.2 Mathematical Model of the Full Cover Problem

Consider $G^o = (N^o, A^o)$, a transportation network defined by the set of nodes N^o and set of directed arcs A^o. These represent road/highway intersections and segments on this network, respectively. K is the set of OD pairs, where O_k and D_k denote the origin and destination nodes for $k \in K$.

We use an expanded network where an arc represents a feasible shortest path in terms

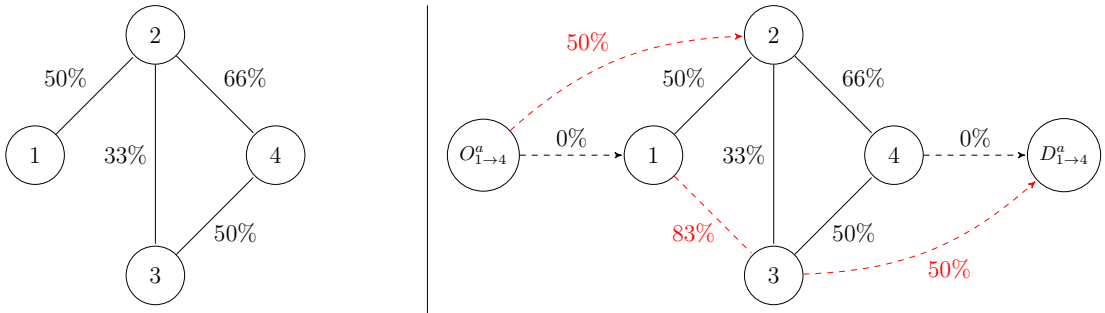
of R between any two nodes on the original network (MirHassani and Ebrazi, 2013). This network transformation is originally proposed for the fixed path setting in the RSLP to eliminate the need to pregenerate all feasible facility combinations and to ensure round trip feasibility. Nevertheless, it can also be used when shortest path deviations are allowed and this is described next.

Let $G = (N, A)$ be the expanded network constructed from the original network G^o, where N denotes the set of its nodes and A denotes the set of its arcs. The length of an arc (i, j) is denoted by ℓ_{ij}. The steps of this transformation is detailed below.

- For each OD pair $k \in K$, two artificial OD nodes O_k^a and D_k^a are added so that
  - $N = N^o \cup \left(\bigcup_{k \in K} \{O_k^a, D_k^a\}\right)$,
  - $A_1 = \bigcup_{k \in K} (O_k^a, O_k)$, $\ell_{O_k^a O_k} = 0$, and
  - $A_2 = \bigcup_{k \in K} (D_k, D_k^a)$, $\ell_{D_k D_k^a} = 0$.
- Let π_{ij}^* denote the shortest path length between two node pairs $i, j \in N$. The additional set of arcs of the expanded network are defined as:
  - $A_3 = \bigcup_{i,j \in N^o,\, i \neq j} \left\{(i, j) \mid \pi_{ij}^* \leq R\right\}$, $\ell_{ij} = \pi_{ij}^*$,
  - $A_4 = \bigcup_{j \in N^o} \left\{(O_k^a, j) \mid \pi_{O_k^a j}^* \leq R/2, \right\}$, $\ell_{O_k^a j} = \pi_{O_k^a j}^*$,
  - $A_5 = \bigcup_{i \in N^o} \left\{(i, D_k^a) \mid \pi_{i D_k^a}^* \leq R/2\right\}$, $\ell_{i D_k^a} = \pi_{i D_k^a}^*$, and
  - $A = A^o \bigcup_{i=1}^{5} A_i$.

If the shortest path between any two nodes may be completed without recharging, a stop for a recharge is not needed at an intermediary node on that shortest path. Hence, a path from a source node to a sink node in G only requires locating a charging station at the head of each used arc, except D_k^a, to be feasible. Artificial nodes O_k^a and D_k^a are included to ensure round trip feasibility as they represent leaving the origin and entering

the destination nodes with 50% battery level. Figure 2.3 demonstrates the construction of the expanded network using an example graph with 4 nodes.



(a) Sample network with 4 nodes **(b) Expanded network for OD pair 1→4**

Figure 2.3: An illustration of expanded network generation.

This graph transformation is generic enough to be implemented for a heterogeneous vehicle setting by incorporating R_{mk} instead of R for each vehicle type m on OD pair k. We define the length of arcs in the transportation network in terms of their energy consumption as a percentage of R. We believe this is a generic way to represent energy consumption of a trip as factors other than the total trip distance, such as average speed or weight of a vehicle and road gradients, may be incorporated into calculations. Before presenting the mathematical model, we introduce the notation to be used hereafter.

Parameters:

R: range of an EV.
p: number of facilities to be located.
ℓ_{ij}: total energy required to traverse arc $(i,j) \in A$, as % of R.
d_k: demand of OD pair $k \in K$.
f_i: fixed cost of locating a charging station at $i \in N^o$.

Decision variables:

$$y_i = \begin{cases} 1 & \text{if a charging station is located at node } i, \\ 0 & \text{otherwise.} \end{cases} \qquad i \in N^o$$

$$x_{ijk} = \begin{cases} 1 & \text{if arc } (i,j) \text{ is traversed by OD pair } k, \\ 0 & \text{otherwise.} \end{cases} \qquad (i,j) \in A, k \in K$$

t_k : charging energy required for OD pair $k \in K$ to complete its trip as % of R.

r_i : battery level at origin/destination node $i \in \bigcup_{k \in K} \{O_k, D_k\}$.

$$(\mathscr{P}_0) \text{ minimize} \quad \sum_{k \in K} d_k t_k + \sum_{i \in N^o} f_i y_i \tag{2.1}$$

$$\text{s.t.} \quad \sum_{j:(i,j) \in A} x_{ijk} - \sum_{j:(j,i) \in A} x_{jik} = \begin{cases} 1, & \text{if } i = O_k^a, \\ -1, & \text{if } i = D_k^a, \\ 0, & otherwise. \end{cases} \quad i \in N, \quad k \in K \tag{2.2}$$

$$\sum_{i:(i,j) \in A} x_{ijk} \leq y_j \qquad j \in N^o, \quad k \in K \tag{2.3}$$

$$t_k = \sum_{(i,j) \in A} \ell_{ij} x_{ijk} + r_{D_k} - r_{O_k} \qquad k \in K \tag{2.4}$$

$$r_i = \frac{1}{2}(1 + y_i) \qquad i{=}O_k, \quad k \in K \tag{2.5}$$

$$r_i \geq \frac{1}{2}(1 - y_i) \qquad i{=}D_k, \quad k \in K \tag{2.6}$$

$$t_k \geq 0 \qquad k \in K \tag{2.7}$$

$$r_i \geq 0 \qquad i \in \{O_k, D_k\}, \quad k \in K \tag{2.8}$$

$$x_{ijk} \in \{0,1\} \qquad (i,j) \in A, \quad k \in K \tag{2.9}$$

$$y_i \in \{0,1\} \qquad i \in N^o \tag{2.10}$$

The objective function (2.1) has two terms, the first term minimizes the total en route recharging to complete all OD trips, implicitly minimizing the deviation from the shortest paths, and the second term minimizes the total fixed cost of locating charging facilities. Constraints (2.2) are flow balance constraints and ensure that there is a feasible path for every OD pair. Constraint (2.3) ensures that a charging station is located at the head of each used arc except for the artificial destination nodes. Constraint (2.4) defines the

decision variable t_k and computes the total recharging needed on a trip. Depending on the availability of a charging facility, constraints (2.5) and (2.6) make sure that the battery level assumption on the origin and destination nodes of a trip holds. Constraint (2.5) ensures that the battery level at the origin is 100% if there is a facility located there and equals to 50% otherwise. Constraint (2.6) defines the lower bound on the battery level at a destination: if there is a facility at the destination node, there is no requirement to satisfy; otherwise, it should at least be 50%. Constraints (2.7)-(2.10) determine the domains of the decision variables.

We model a similar setting of the well-known location set covering problem (Toregas et al., 1971); however, we additionally minimize the total recharging energy of all trips. Hence, we refer to this model as the *Full Cover Problem* (FCP).

The objective function (2.1) comprises of terms that are different in units, *i.e.,* the first term is the total en route recharging amount (as a function of R) and the latter is total cost. The scale of the parameters d_k and f_i would play a critical role in creating a trade-off between these objective function components. This trade-off can be avoided by multiplying one of the terms, typically the total cost term, with an adjustment factor, similar to an objective weight for linear weighted scalarization of a multi-objective optimization problem. This way the decision maker ensures that the secondary objective (total en route recharging energy) is minimized among the solutions which have the best value of the main objective (total cost), i.e., (2.1) is interpreted as a hierarchical objective function.

Another way to handle this nature of the problem is treating one of the objectives as a constraint. As a result, the FCP may easily be modified to locate p facilities rather than minimizing total location cost. In order to make such a change, the objective function term $\sum_{i \in N^o} f_i y_i$ is dropped and the constraint $\sum_{i \in N^o} y_i \leq p$ is added. This modification has an underlying assumption that f_i values are the same for all $i \in N^o$. If this is not

the case, then the right-hand side should include the total budget and the left-hand side should have the total cost term. We refer to this variant as the p-FCP formulation and introduce its mathematical model below:

$$(p - \mathscr{P}_0) \text{ minimize} \quad \sum_{k \in K} d_k t_k \tag{2.11}$$

$$\text{s.t.} \quad \sum_{j \in N^o} y_j \leq p \tag{2.12}$$

$$(2.2) \; - \; (2.10)$$

The p-FCP may be infeasible since, unlike the MCP, it requires finding a feasible path for every OD pair. However, the MCP does not guarantee to find a better solution in terms of routing decisions when p is increased. On the other hand, increasing the number of open facilities is expected to improve the routing-related performance measures when the p-FCP formulation is used. The value of using the p-FCP would be significant especially when the FCP finds solutions with relatively higher deviations since it does not enforce a deviation tolerance. By increasing p gradually, a decision maker would be able to evaluate the trade-off between the number of located facilities and routing-related performance metrics.

In the context of the FCP or p-FCP, we focus on a long-distance transportation setting that only includes OD pairs whose shortest path length exceeds R since only these will have to use the charging stations on every trip. This long-distance problem setting is particularly suitable for locating fast charging stations for EVs as the trips shorter than R can readily be supported by home recharging. On another note, as pointed out by Göpfert and Bock (2019), filtering out shorter trips substantially increases the complexity of the instances to be solved since many of the remaining trips are required to be recharged multiple times along the way.

By taking only the long distance travel demand into account, we can simplify the

formulation. The right-hand side of constraint (2.5) is readily substitutable in (2.4). Then, we change the sign of constraint (2.6) to an equality due to the charge-just-enough policy and substitute r_{D_k} variables in constraint (2.4). As all OD pairs would always satisfy (2.7), we can substitute the right hand side of (2.4) directly in the objective function. With these changes, we drop constraints (2.4)-(2.8) and present $\mathscr{P}_1$ as the reformulation of the FCP by relaxing the domain of the flow variables.

$$(\mathscr{P}_1) \text{ minimize} \quad \sum_{k\in K}\sum_{(i,j)\in A} d_k\left(\ell_{ij}x_{ijk} - \frac{y_{O_k}+y_{D_k}}{2}\right) + \sum_{i\in N^o} f_i y_i \tag{2.13}$$

$$\text{s.t.} \quad (2.2),\ (2.3)$$

$$x_{ijk} \geq 0 \qquad (i,j)\in A,\ k\in K \tag{2.9'}$$

$$y_i \in \{0,1\} \qquad i \in N^o \tag{2.10}$$

The p-FCP formulation $p-\mathscr{P}_0$ can also be simplified in the same way. To achieve that, constraint (2.12) will be added to the constraints of $\mathscr{P}_1$ and the term $\sum_{i\in N^o} f_i y_i$ will be dropped from (2.13).

Next, we prove that there always exists an optimal integral solution to $\mathscr{P}_1$.

Proposition 2.1. *$\mathscr{P}_1$ always has an optimal solution with integral x_{ijk} values for given binary y_i values.*

Proof: We show that, for given y_i values, the constraint matrix of $\mathscr{P}_1$, denoted by $\mathcal{A}$, is totally unimodular (TU). We use the theorem by Ghouila-Houri (1962) which asserts that a matrix is TU if and only if any subset of its rows can be partitioned into two disjoint sets $\mathscr{R}_1$ and $\mathscr{R}_2$ such that the vector obtained by $\sum_{i\in\mathscr{R}_1} a_{ij} - \sum_{i\in\mathscr{R}_2} a_{ij}$ consists only of $0, \pm 1$. We now provide such a partitioning for $\mathcal{A}$ given binary values for y_i.

The structure of $\mathcal{A}$ for each $k \in K$ is shown in Figure 2.4. $\mathcal{A}$ decomposes by $k \in K$, let $\mathcal{A}_k$ denote the constraint matrix for OD pair $k \in K$. Each such submatrix $\mathcal{A}_k$ has

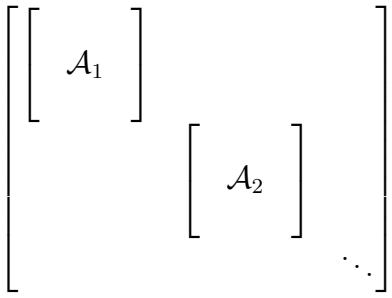


Figure 2.4: The structure of the constraint matrix $\mathcal{A}$.

$|N| + |N^o|$ rows and $|A|$ columns. The column corresponding to arc (i, j) has $+1$ in row i and -1 in row j, corresponding to flow balance constraints (2.2). This column also has $+1$ in row $|N| + j$ corresponding to constraint (2.3). The rest of its entries are zero.

Given a subset of rows $\mathscr{R}$ of $\mathcal{A}_k$, $\sum_{i \in \mathscr{R}} a_{i(i,j)k}$ may not be equal to 0 or ± 1 only if row $j + |N|$ is in $\mathscr{R}$ and row j is not in $\mathscr{R}$. In that case, it is sufficient to remove row $j + |N|$ from this subset of rows. We construct a partitioning $\mathscr{R}_1$ and $\mathscr{R}_2$ of $\mathscr{R}$ such that the vector obtained by $\sum_{i \in \mathscr{R}_1} a_{i(i,j)k} - \sum_{i \in \mathscr{R}_2} a_{i(i,j)k} = \{0, \pm 1\}$ as follows:

- For $i \leq |N|$, place row i in $\mathscr{R}_1$.
- For $|N| + 1 \leq i \leq |N| + |N^o|$,
  - if, row $i - |N| \in \mathscr{R}$, place row i in $\mathscr{R}_1$.
  - Otherwise, place row i in $\mathscr{R}_2$.

Hence, we can claim that each $\mathcal{A}_k$, $k \in K$, and consequently $\mathcal{A}$, is TU and $\mathscr{P}_1$ always has an optimal solution with integral x_{ijk} values for given binary y_i values. We note here that this proof can be extended for the p-FCP. $\square$

The preliminary computational analysis with $\mathscr{P}_1$ shows that off-the-shelf solvers are not able to handle the size of the problem for large networks and go out of memory. In order to solve $\mathscr{P}_1$ for real-life instances, we develop a Benders decomposition algorithm as discussed in the next section.

2.3 Solution Methodology

In this section, first the Benders reformulation of $\mathscr{P}_1$ is presented. Then a novel subproblem solution algorithm is introduced, which is developed to improve the computational performance of the solution methodology.

2.3.1 The Benders Reformulation

Given the structure of $\mathscr{P}_1$, there are two types of decisions to make; namely, location and flow, where flow decisions are based on the selection of charging station locations. This structure may be exploited so that the location decisions are determined through the master problem and optimal paths between OD pairs are constructed within the subproblem.

Let $\overline{y}_i$ denote the given values for the location variables. The subproblem ($\mathcal{SP}$) is defined as:

$$(\mathcal{SP})\ q(Y) = \text{minimize} \quad \sum_{(i,j)\in A}\sum_{k\in K} d_k \ell_{ij} x_{ijk} \tag{2.14}$$

$$\text{s.t.} \quad (2.2),$$

$$\sum_{i:(i,j)\in A} x_{ijk} \leq \overline{y}_j \quad j \in N^o, \quad k \in K \tag{2.3}$$

$$x_{ijk} \geq 0 \qquad (i,j) \in A, \quad k \in K \tag{2.9'}$$

$\mathcal{SP}$ is a many-to-many shortest path problem with an additional constraint that restricts the nodes that can be used to the set of open facilities. Let λ_{jk} and μ_{jk} be the dual variables associated with constraints (2.2) and (2.3), respectively. The dual of $\mathcal{SP}$ is:

$$(\mathcal{DSP})\ \text{maximize} \quad \sum_{k\in K} \lambda_{O_k^a k} - \sum_{k\in K} \lambda_{D_k^a k} + \sum_{j\in N^o}\sum_{k\in K} \overline{y}_j \mu_{jk} \tag{2.15}$$

$$\text{s.t.} \quad \lambda_{ik} - \lambda_{jk} \leq d_k \ell_{ij} \qquad j = D_k^a, \quad (i,j) \in A, \quad k \in K \tag{2.16}$$

$$\lambda_{ik} - \lambda_{jk} + \mu_{jk} \leq d_k \ell_{ij} \qquad j \in N^o, \quad (i,j) \in A, \quad k \in K \tag{2.17}$$

$$\mu_{jk} \leq 0 \qquad j \in N^o, \quad k \in K \tag{2.18}$$

Let $q(Y)$ denote the objective value of $\mathcal{SP}$ given $Y = [y_i]$, $i \in N^o$. $\mathscr{P}_1$ is reformulated as follows:

$$(\mathscr{P}_1) \text{ minimize} \quad \sum_{k \in K} -d_k \left(\frac{y_{O_k} + y_{D_k}}{2} \right) + \sum_{i \in N^o} f_i y_i + q(Y) \tag{2.19}$$

$$\text{s.t.} \quad y_i \in \{0,1\} \qquad i \in N^o \tag{2.10}$$

It is trivial to show that $\mathcal{SP}$ is bounded for $d_k \geq 0$, $k \in K$ and $\ell_{ij} \geq 0$, $(i,j) \in A$. Hence, when feasible, $\mathscr{P}_1$ has a finite optimal solution for $f_i \geq 0$, $i \in N^o$. It follows that, either $\mathcal{DSP}$ and $\mathcal{SP}$ are both feasible, or $\mathcal{DSP}$ is unbounded and $\mathcal{SP}$ is infeasible. Let $\mathcal{R}$ be a set of extreme rays and $\mathcal{H}$ be a set of extreme points of the polyhedron defined by (2.16)-(2.18). The Benders master problem $\mathcal{BMP}$ is:

$$(\mathcal{BMP}) \quad \text{minimize} \quad \sum_{k \in K} -d_k \left(\frac{y_{O_k} + y_{D_k}}{2} \right) + \sum_{i \in N^o} f_i y_i + \mathbf{z} \tag{2.20}$$

$$\text{s.t.} \quad \sum_{k \in K} \lambda^r_{O_k^a k} - \sum_{k \in K} \lambda^r_{D_k^a k} + \sum_{j \in N^o} \sum_{k \in K} y_j \mu^r_{jk} \leq 0 \quad r \in \mathcal{R} \tag{2.21}$$

$$\mathbf{z} \geq \sum_{k \in K} \lambda^h_{O_k^a k} - \sum_{k \in K} \lambda^h_{D_k^a k} + \sum_{j \in N^o} \sum_{k \in K} y_j \mu^h_{jk} \quad h \in \mathcal{H} \tag{2.22}$$

$$y_i \in \{0,1\} \qquad i \in N^o \tag{2.10}$$

When defined on complete sets of extreme rays $\mathcal{R}$ and extreme points $\mathcal{H}$, $\mathcal{BMP}$ is a

complete reformulation of $\mathscr{P}_1$. However, it is not practical to generate all $r \in \mathcal{R}$ and $h \in \mathcal{H}$ beforehand. A Benders algorithm is used instead to generate feasibility and optimality cuts as needed. $\mathcal{BMP}$ can be tailored for the p-FCP by adding constraint (2.12) to it and dropping the term $\sum_{i \in N^o} f_i y_i$ from (2.20). Since the subproblem $\mathcal{SP}$ decomposes by $k \in K$, the Benders algorithm has aggregated and disaggregated implementations depending on whether a single aggregated cut is generated for all $k \in K$ or disaggregated $|K|$ cuts are generated, one for each $k \in K$. We use the former approach in implementations.

Benders algorithm can be implemented to work on a single branch-and-bound search tree for the master problem rather than solving it from scratch each time a new cut is added. We adopt this methodology which is referred to as the branch-and-Benders-cut (Rahmaniani et al., 2017) where branching is performed by the commercial solvers. Based on computational experience with $\mathscr{P}_1$ using CPLEX optimizer, we observed that over 95% of the computational time is spent on solving the subproblems. Therefore, we develop a solution algorithm to solve the subproblem and generate Benders cuts in significantly less time.

2.3.2 Solution of the Subproblem

In this section, we first discuss how to construct an optimal solution of $\mathcal{DSP}$ given an optimal solution to $\mathcal{SP}$ in order to generate optimality cuts. Then, we introduce valid inequalities to be substituted for feasibility cuts when $\mathcal{SP}$ is infeasible. Ultimately, we present the algorithm itself. For ease of exposition and without loss of generality, we work with a unit demand for each OD pair, *i.e.,* $d_k = 1$, $k \in K$.

Both $\mathcal{SP}$ and $\mathcal{DSP}$ are decomposable by $k \in K$. Let $\mathcal{SP}_k$ and $\mathcal{DSP}_k$ denote these decomposed problems for OD pair k. $\mathcal{SP}_k$ and consequently $\mathcal{SP}$, may be solved as a shortest path problem using a labeling algorithm by embedding constraint (2.3) *a priori*

on the network. In order to embed constraint (2.3) on a reduced network $G_k = (N_k, A_k)$, we set $N_k = O_k \cup D_k \cup \{i \in N^o \mid \overline{y}_i = 1\}$ and $A_k = \{(i,j) \in A \mid i,j \in N_k\}$, *i.e.*, eliminating redundant nodes and arcs of G.

The shortest path problem on G_k is equivalent to $\mathcal{SP}_k$ and a solution for $\mathcal{SP}_k$ may be obtained by a labeling algorithm assuming that it is feasible. However, in order to derive Benders cuts, one needs to construct the complete dual solution. Besides, when there are multiple optimal solutions to $\mathcal{DSP}_k$, we would like to find one that leads to a non-dominated Benders cut. In the next section, we introduce a novel approach to construct such dual solutions and present an efficient algorithm to calculate them.

Dual Subproblem Solution and Optimality Cuts:

Assume that $\mathcal{SP}$ and $\mathcal{DSP}$ are both feasible and let $\overline{x}_k$ be an optimal solution to $\mathcal{SP}_k$ with corresponding dual optimal solution $\overline{\lambda}_k$ when solved on G_k. While determining an optimal solution to $\mathcal{SP}$ from $\overline{x}_k$ as well as the values of the dual variables for $i \in N_k$ is trivial, constructing a complete optimal solution to $\mathcal{DSP}$ is a challenging task. The values of the dual variables for the remaining nodes cannot be calculated in the same trivial manner.

Given an optimal solution $\overline{x}_k$ to $\mathcal{SP}_k$, $k \in K$, let I_1^k denote the set of nodes on the shortest path for $\mathcal{SP}_k$; *i.e.*, $I_1^k = \{i \in N_k \mid \overline{x}_{ijk} = 1,\ j \in N_k\}$. Let the objective value of $\mathcal{SP}_k$ be denoted as $\mathbf{z_k}$. A primal-dual optimal solution pair $\left(x^*_{ijk}, (\lambda^*_{ik}, \mu^*_{ik})\right)$ to $\mathcal{SP}$ can be partially determined as follows:

$$x^*_{ijk} = \overline{x}_{ijk} \qquad (i,j) \in A_k, k \in K \qquad \text{(i)}$$
$$x^*_{ijk} = 0 \qquad (i,j) \notin A_k, k \in K \qquad \text{(ii)}$$
$$\lambda^*_{ik} = \overline{\lambda}_{ik} \qquad i \in I_1^k, k \in K \qquad \text{(iii)}$$
$$\mu^*_{ik} = 0 \qquad i \in I_1^k, k \in K \qquad \text{(iv)}$$

Conditions (i) and (ii) are trivial since $\mathcal{SP}_k$ is solved on G_k. The dual λ_{ik} values determined by a labeling algorithm would satisfy constraint (2.16) as an equality for the arcs that are on the shortest path and entering D_k^a. Similarly, for the rest of the shortest path arcs, the corresponding λ_{ik} values would also satisfy (2.17) as an equality by the shortest path property and complementary slackness. Thus, setting μ_{ik} values of the shortest path nodes as zero satisfies constraint (2.17).

Determining the rest of the optimal dual variables μ_{ik}^*, $i \notin I_1^k$ corresponding to constraint (2.3) is non-trivial since that constraint is embedded in G_k and not explicitly considered when solving $\mathcal{SP}_k$. Based on constraints (2.16) and (2.17) of $\mathcal{SP}_k$, we derive the following inequality for an arbitrary OD path $\mathbf{p}$ on G:

$$\sum_{j \in I_p} \mu_{jk} \leq \mathbf{z_{pk}} - \mathbf{z_k}$$

where $\mathbf{z_{pk}}$ is the length of this arbitrary path $\mathbf{p}$ and I_p denotes its nodes. Since path $\mathbf{p}$ is an arbitrary path on G, it may not necessarily use only the open facilities but any node $i \in N^o$. Recall that $\mathbf{z_k}$ is the current shortest path length with respect to $\overline{y}_j$. If $\mathbf{z_k} \leq \mathbf{z_{pk}}$, a redundant case is observed as the right hand side of this inequality is nonnegative and $\mu_{jk} \leq 0$ due to (2.18). On the other hand, if $\mathbf{z_k} > \mathbf{z_{pk}}$, $\sum_{j \in I_p} \mu_{jk}$ may be interpreted as the improvement achieved over the current shortest path length by locating additional facilities at nodes $j \in I_p$. Every such path $\mathbf{p}$ would yield some information based on the μ_{ik} values and one can determine the tightest μ_{ik} values by considering all possible OD paths on G. However, this is computationally inefficient. At this point, we propose an algorithm that determines optimal μ_{ik}^* values to generate a non-dominated optimality cut.

We define two subgraphs of G to be used hereafter. Let $G_k^* = (N, A_k^*)$ where $A_k^* = A \setminus \{(i, O_k), (D_k, i) | i \in N^o\}$, *i.e.,* arcs entering the original origin and leaving the original destination nodes are eliminated from the expanded network arc set A. Let $G_k^f = (N, A_k^f)$

where $A_k^f = \{(i,j) \in A \mid i \in N_k,\ j \in N^o \cup D_k\}$, *i.e.,* only the arcs leaving the nodes in N_k are considered. Let w_{ik} be the shortest path length from O_k to $i \in N^o \cup D_k^a$ on G_k^*; u_{ik} be the shortest path length from O_k^a to node $i \in N^o \cup D_k^a$ on G_k^f, and v_{ik} be the shortest path length from D_k^a to node $i \in N^o \cup O_k^a$ on the reverse graph of G_k^f. While u_{ik} and v_{ik} depend on the set of open facilities in N_k, w_{ik} does not. Based on these definitions we rewrite (iii) as:

$$\lambda_{ik}^* = -u_{ik} \qquad i \in I_1^k, k \in K \tag{iii}$$

The remaining nodes are partitioned as I_2^k and I_3^k such that $I_2^k = \left\{i \in N^o \backslash I_1^k \mid w_{ik} + v_{ik} \geq \mathbf{z_k}\right\}$ and $I_3^k = \left\{i \in N^o \backslash I_1^k \mid w_{ik} + v_{ik} < \mathbf{z_k}\right\}$ for $k \in K$, where the condition differentiates the nodes based on their potential to improve the current shortest path length. Consider the following for I_2^k and I_3^k:

$$\begin{aligned}
&\lambda_{ik}^* = -w_{ik} && i \in I_2^k,\ k \in K && \text{(v)}\\
&\mu_{ik}^* = 0 && i \in I_2^k,\ k \in K && \text{(vi)}\\
&\lambda_{ik}^* = -\mathbf{z_k} + v_{ik} && i \in I_3^k,\ k \in K && \text{(vii)}\\
&\mu_{ik}^* = 0, && i \in I_3^k\ \overline{y}_i = 1,\ k \in K && \text{(viii)}
\end{aligned}$$

At this point, all remaining μ_{jk}^* can be calculated based on λ_{ik}^* determined by (iii),(v) and (vii):

$$\mu_{jk}^* = \min\left\{0, \min_{(i,j)\in A}\left\{\ell_{ij} - \lambda_{ik}^* + \lambda_{jk}^*\right\}\right\} \qquad j \in I_3^k,\ \overline{y}_j = 0,\ k \in K \tag{ix}$$

Proposition 2.2. *The primal-dual solution* $\left(x_{ijk}^*, (\lambda_{ik}^*, \mu_{ik}^*)\right)$ *defined by* (i) – (ix) *is optimal to* $\mathcal{SP}$ *and* $\mathcal{DSP}$*, respectively.*

Proof: It is trivial to show that x^* determined by (i) and (ii) is feasible to $\mathcal{SP}$. We

now show that $(\lambda^*_{ik}, \mu^*_{ik})$ is dual feasible and strong duality holds.

Recall the dual feasibility conditions (2.16)-(2.18). For nodes $i \in I_2^k$, constraint (2.16) reduces to $w_{ik} + v_{ik} \geq \mathbf{z_k}$, since $\ell_{iD_k^a} = v_{ik}$ under this case, and this is already satisfied by the very definition of I_2^k. For $j \in I_2^k$, constraints (2.17) and (2.18) imply $\mu_{jk} \leq \min\left\{\left(0, \min_{(i,j)\in A}\left\{\ell_{ij} - \lambda_{ik}\lambda_{jk}\right\}\right)\right\}$. Due to (vi), it is sufficient to show that $\ell_{ij} - \lambda_{ik} + \lambda_{jk} \geq 0$, $(i,j) \in A$ and $j \in I_2^k$ for dual feasibility to hold. Note that $w_{ik} + \ell_{ij} \geq w_{jk}$ holds for $(i,j) \in A$ by the shortest path property. Let us evaluate $\ell_{ij} - \lambda_{ik} + \lambda_{jk} \geq 0$ under 3 cases:

- For $i \in I_1^k$, by (iii) $\ell_{ij} + u_{ik} - w_{jk} \geq \ell_{ij} + w_{ik} - w_{jk} \geq 0$ as $u_{ik} \geq w_{ik}$ by definition.
- For $i \in I_2^k$, by (v) $\ell_{ij} + w_{ik} - w_{jk} \geq 0$.
- For $i \in I_3^k$, $w_{ik} + v_{ik} < \mathbf{z_k}$; by (vii), $\ell_{ij} + \mathbf{z_k} - v_{ik} - w_{jk} > \ell_{ij} + w_{ik} - w_{jk} \geq 0$.

All three cases hold by the shortest path property; thus, dual feasibility is satisfied for $j \in I_2^k$.

For $i \in I_3^k$ and $\overline{y}_i = 1$, constraint (2.16) reduces to $-\mathbf{z_k} + v_{ik} + \mathbf{z_k} \leq v_{ik}$ which trivially holds. For $j \in I_3^k$ in constraints (2.17) and (2.18), it is sufficient to show that $\ell_{ij} - \lambda_{ik} + \lambda_{jk} \geq 0$, $(i,j) \in A$. This expression is also evaluated under 3 cases:

- For $i \in I_1^k$, by (iii), $\ell_{ij} + u_{ik} \geq \mathbf{z_k} - v_{jk} \implies \ell_{ij} + u_{ik} + v_{jk} \geq \mathbf{z_k}$. This is always true since $i \in I_1^k$ and $y_j = 1$. Otherwise, j would have been in I_1^k, *i.e.,* a node on the shortest path.
- For $i \in I_2^k$, due to its set definition and (v), we obtain $\ell_{ij} + w_{ik} - \mathbf{z_k} + v_{jk} \geq \ell_{ij} + \mathbf{z_k} - v_{ik} - \mathbf{z_k} + v_{jk} \geq 0$. This reduces to $\ell_{ij} + v_{jk} \geq v_{ik}$ which holds by the shortest path property on the reverse graph of G_k^f.
- For $i \in I_3^k$, by (vii), $\ell_{ij} + \mathbf{z_k} - v_{ik} - \mathbf{z_k} + v_{jk} \geq 0$. This also reduces to $\ell_{ij} + v_{jk} \geq v_{ik}$ which holds by the shortest path property on the reverse graph of G_k^f.

Thus, dual feasibility is satisfied for $j \in I_3^k, \overline{y}_j = 1$.

For $j \in I_3^k$ and $\overline{y}_j = 0$, constraint (2.16) reduces to $w_{ik} + v_{ik} \geq \mathbf{z_k}$ and the arguments follow as before. Constraints (2.17) and (2.18) reduce to $\mu_{jk} \leq \min\left\{\left(0, \min_{(i,j)\in A}\left\{\ell_{ij} - \lambda_{ik} + \lambda_{jk}\right\}\right)\right\}$. Due to (ix), letting μ_{jk} equal to the RHS of this inequality will always satisfy the dual feasibility conditions in the tightest way. Therefore, $(\lambda^*_{ik}, \mu^*_{ik})$ is dual feasible.

Recall that $\mathbf{z_k}$ is the primal optimal solution to $\mathcal{SP}_k$ and by (iii) $\lambda^*_{O^a_k k} = 0$, $\lambda^*_{D^a_k k} = -\mathbf{z_k}$. As $\mu^*_{jk} = 0$ for $\{j \in N^o \mid \overline{y_j} = 1\}$ by (iv), (vi) and (viii), the optimal objective value of $\mathcal{DSP}_k$ is $\mathbf{z_k}$. This holds for all feasible $k \in K$; hence, by strong duality, $\left(x^*_{ijk}, (\lambda^*_{ik}, \mu^*_{ik})\right)$ defined by (i) – (ix) is optimal to $\mathcal{SP}$ and $\mathcal{DSP}$, respectively.□

The optimal dual solution constructed by (i) – (ix) is proven to generate a non-dominated optimality cut for $\mathcal{BMP}$ next.

Proposition 2.3. *A $\mathcal{DSP}$ solution $(\lambda^*_{jk}, \mu^*_{jk})$ constructed by (i) – (ix) generates a non-dominated optimality cut.*

Proof: For solution $(\lambda^*_{jk}, \mu^*_{jk})$, $\lambda^*_{O^a_k k} = -u_{O^a_k} = 0$ and $\lambda^*_{D^a_k k} = -u_{D^a_k} = -\mathbf{z_k}$ by (iii). Hence, $\lambda^*_{O^a_k k} - \lambda^*_{D^a_k k} = \mathbf{z_k}$ and optimality cut (2.22) may be written as:

$$\mathbf{z} \;\geq\; \sum_{k\in K} \mathbf{z_k} + \sum_{j\in N^o}\sum_{k\in K} y_j \mu^*_{jk}$$

Moreover for any alternative optimal solution $(\overline{\lambda}_{jk}, \overline{\mu}_{jk})$, $\overline{\lambda}_{O^a_k k} - \overline{\lambda}_{D^a_k k} = \mathbf{z_k}$ by strong duality. Solution $(\lambda^*_{jk}, \mu^*_{jk})$ defines a non-dominated cut if there does not exists an alternative optimal solution $(\overline{\lambda}_{jk}, \overline{\mu}_{jk})$ such that $\overline{\mu}_{jk} \geq \mu^*_{jk}$, for which a strict inequality holds for at least one $j \in N^o$, $k \in K$ (Magnanti and Wong, 1981). Recall that $\mathcal{DSP}$ constraints (2.17) and (2.18) reduce to $\mu_{jk} \leq \min\left\{0, \min_{(i,j)\in A}\left\{\ell_{ij} - \lambda_{ik} + \lambda_{jk}\right\}\right\}$. For OD pair $k \in K$:

- When $\mu^*_{jk} = 0$ (i.e., by (iv), (vi) or (viii)), $\overline{\mu}_{jk} \leq \mu^*_{jk}$ by constraint (2.18).

- When $\mu^*_{jk} < 0$, $j \in I^k_3$ by (ix) and $v_{jk} \neq \infty$. Moreover, (ix) ensures that constraint (2.17) is satisfied at equality for at least one $i \in N^o$ such that $(i,j) \in A$. This also holds for any $(i,\hat{j}) \in A$ such that $\hat{j} \in I^k_1 \cup I^k_2$ by the shortest path property since $\mu^*_{\hat{j}k} = 0$ by (iv) and (vi).

  Assume for some $j \in I^k_3$, $\overline{\mu}_{jk} > \mu^*_{jk}$. It follows that $-\overline{\lambda}_{ik} + \overline{\lambda}_{jk} > -\lambda^*_{ik} + \lambda^*_{jk}$ for some arc $(i,j) \in A$ for which constraint (2.17) is binding. Assuming without loss of generality $\overline{\lambda}_{O^a_k k} = 0$, which also implies $\overline{\lambda}_{ik} = \lambda^*_{ik}$ for $i \in I^k_1$. This follows that:

  - $\overline{\lambda}_{jk} \leq \lambda^*_{jk}$ by (2.16) and (2.17) since $v_{jk} \neq \infty$ implies that there exists an arc $(j,m) \in A$ such that either $\overline{y}_m = 1$ or $m = D^a_k$; and $\overline{\lambda}_{O^a_k k} - \overline{\lambda}_{D^a_k k} = \mathbf{z_k}$ and $\overline{\mu}_{mk} = 0$ (if $m \neq D^a_k$) due to complementary slackness in any alternative optimal $\mathcal{DSP}$ solution.
  - If $\overline{\lambda}_{ik} < \lambda^*_{ik}$ and constraint (2.17) is binding for (i,j), $\overline{\mu}_{jk} > \mu^*_{jk}$. To maintain feasibility of constraint (2.17) for some arc $(m,i) \in A$ for which this constraint is also binding, *a)* either $\overline{\mu}_{ik} < \mu^*_{ik}$ (when $\overline{y}_i = 0$) and $(\lambda^*_{jk}, \mu^*_{jk})$ is non-dominated, *b)* or $\overline{\lambda}_{mk} < \lambda^*_{mk}$. If $m \in I^k_1$, this contradicts with $\overline{\lambda}_{mk} = \lambda^*_{mk}$, and $(\lambda^*_{jk}, \mu^*_{jk})$ is non-dominated. Else, identify an arc for which (2.17) is binding and either conclude that $(\lambda^*_{jk}, \mu^*_{jk})$ is non-dominated or continue until arc (O^a_k, O_k) is reached. In that case $\overline{\lambda}_{O^a_k k} < \lambda^*_{O^a_k k}$ contradicts with $\overline{\lambda}_{O^a_k k} = \lambda^*_{O^a_k k} = 0$ and $(\lambda^*_{jk}, \mu^*_{jk})$ is non-dominated.

Hence, $\mathcal{DSP}$ solution $(\lambda^*_{jk}, \mu^*_{jk})$ generates a non-dominated optimality cut. □

Feasibility Cuts:

Infeasibility of $\mathcal{SP}$ means that there exists at least one OD pair $k \in K$ whose origin and destination nodes are disconnected given the set of open facilities $N^o_y = \{j \in N^o \mid \overline{y}_j = 1\}$,

i.e., active nodes.In this case, the corresponding $\mathcal{DSP}_k$ is unbounded and a Benders feasibility cut (2.21) is generated. Alternatively, a trivial way to cut out this infeasible solution is to add a constraint to $\mathcal{BMP}$ that ensures locating at least one additional facility:

$$\sum_{j \in N^o \backslash N_y^o} y_j \geq 1 \tag{2.23}$$

Cut (2.23) may be strengthened as follows. Let $\hat{k}$ denote an infeasible OD pair. Given the set of open facilities N_y^o, we determine the set of non-facility nodes that are reachable from $\mathcal{O}_{\hat{k}}^a$ on $G_{\hat{k}}^f$. These are the nodes that have a potential to connect $\mathcal{O}_{\hat{k}}^a$ to the rest of the graph. This is repeated from the other end of the network starting from $\mathcal{D}_{\hat{k}}^a$ to determine the set of reachable inactive nodes from $\mathcal{D}_{\hat{k}}^a$ on the reverse graph of $G_{\hat{k}}^f$. Using these sets of nodes, we derive two feasibility cuts for OD pair $\hat{k}$:

$$\sum_{i : \left\{ i \in N^o \backslash N_y^o \,|\, u_{i\hat{k}} \neq \infty \right\}} y_i \geq 1 \tag{2.24}$$

$$\sum_{i : \left\{ i \in N^o \backslash N_y^o \,|\, v_{i\hat{k}} \neq \infty \right\}} y_i \geq 1 \tag{2.25}$$

The cardinality of $\left\{ i \in N^o \backslash N_y^o \,|\, u_{i\hat{k}} \neq \infty \right\}$ and $\left\{ i \in N^o \backslash N_y^o \,|\, v_{i\hat{k}} \neq \infty \right\}$ is always less than or equal to the cardinality of $N^o \backslash N_y^o$. In addition, a solution that satisfies (2.24) and (2.25) also satisfies (2.23) but the opposite is not always true.

Cut (2.23) may also be strengthened as

$$\sum_{j \in N^o \backslash N_y^o} y_j \geq \Gamma_{\hat{k}}, \tag{2.26}$$

where $\Gamma_{\hat{k}}$ denotes the minimum number of additional facilities required to connect $\mathcal{O}_{\hat{k}}^a$ to $\mathcal{D}_{\hat{k}}^a$ on G. We determine $\Gamma_{\hat{k}}$ using breadth-first-search starting from the active nodes that

are reachable from $\mathcal{O}^a_{\hat{k}}$ on G^f_k. $\Gamma_{\hat{k}}$ is equal to the number of layers of G searched until the first active node that is reachable from $\mathcal{D}^a_{\hat{k}}$ on the reverse graph of G^f_k is encountered. Preliminary numerical analysis revealed this to be computationally more expensive than generating (2.24) and (2.25). Besides, $\Gamma_{\hat{k}}$ typically took values that are less than or equal to two which undermines the significance of (2.26). Therefore, we do not include (2.26) in the implementation.

We now summarize the subproblem solution algorithm. After solving $\mathcal{BMP}$ to obtain $\overline{y}_j$, $j \in N^o$, $\mathcal{SP}$ is decomposed into $\mathcal{SP}_k$, $k \in K$, and Algorithm 2.1 is used to solve $\mathcal{SP}$. It processes each $\mathcal{SP}_k$ independently and eventually adds the non-dominated optimality cut (2.22) to $\mathcal{BMP}$ using the dual solutions constructed for each $\mathcal{DSP}_k$ by (i) – (ix) if $\mathcal{SP}$ is feasible. If it identifies an infeasible $\mathcal{SP}_k$, it concludes that $\mathcal{SP}$ is infeasible. At this point, we have two options before adding a feasibility cut. Either we process the remaining OD pairs, identify the infeasible ones, and generate additional feasibility cuts for each of them. Alternatively, one can stop solving $\mathcal{SP}$ as soon as an infeasible OD pair for $\mathcal{SP}_k$ is identified and add a single set of feasibility cuts to $\mathcal{BMP}$. The former multi-cut approach throughout the numerical experiments.

Algorithm 2.1 Subproblem solution and Benders cut generation algorithm

$I_1^k, I_2^k, I_3^k := \varnothing$
Generate G_k^f
Solve a shortest path problem on G_k^f and the reverse graph of it
if $u_{D_k k} == \infty$ **then**
  $\mathcal{SP}_k$ and $\mathcal{SP}$ are infeasible
  Generate feasibility cuts (2.24) and (2.25)
else if $u_{D_k k} \neq \infty$ **then**
  $\mathcal{SP}_k$ is feasible
  $\lambda_{O_k^a} = -u_{O_k^a k} = 0$ and $\lambda_{D_k^a} = -u_{D_k^a k} = \mathbf{z_k}$
  $I_1^k := \{i \in N \mid u_{ik} = v_{ik} \neq \infty\}$
  $I_2^k := \left\{i \in N^o \backslash I_1^k \mid w_{ik} + v_{ik} \geq \mathbf{z_k}\right\}$
  $I_3^k := \left\{i \in N^o \backslash I_1^k \mid w_{ik} + v_{ik} < \mathbf{z_k}\right\}$
  for $i = 1 : N^o$ **do**
    if $i \in I_1^k$ **then**
      $\lambda_{ik} = -u_{ik}, \ \mu_{ik} = 0$
    else if $i \in I_2^k$ **then**
      $\lambda_{ik} = -w_{ik}, \ \mu_{ik} = 0$
    else if $i \in I_3^k$ **then**
      $\lambda_{ik} = -\mathbf{z_k} + v_{ik}$
      if $\overline{y}_i == 1$ **then**
        $\mu_{ik} = 0$
      else
        $\mu_{ik} = \min\left\{\left(0, \min_{(j,i)\in A}\left\{\ell_{(j,i)} - \lambda_{jk} + \lambda_{ik}\right\}\right)\right\}$
      end if
    end if
  end for
end if

2.4 Computational Experiments

In this section, we present a comprehensive computational study to evaluate the effectiveness of using the FCP is presented. Besides the computational performance of the novel solution algorithm is evaluated on two real world road networks. The experiments are executed using CPLEX Optimizer v12.9.0 in the Java API with Concert Technology on a computer with Intel i7-9700K 3.60GHz processor and 16GB of RAM. The Benders algorithm framework is implemented using lazy constraint callback feature of CPLEX and a predetermined time limit of one hour is set for all implementations.

2.4.1 Data Sets

We use three benchmark transportation networks from the literature. The first one with 25 nodes is introduced by Simchi-Levi and Berman (1988), and it is one of the most widely used benchmark networks for the FRLM. This network will be referred to as *N25* henceforth. The second one, with 339 nodes, is introduced more recently by Arslan et al. (2014) and will be referred to as *CA339*. It is a representation of the state of California highway network and was used previously by Arslan et al. (2019) and Yıldız et al. (2016) for the RSLP-R. The potential OD nodes are selected from the population centers that have more than 50,000 inhabitants. The third network, introduced by Göpfert and Bock (2019), involves 5642 nodes and 16,034 arcs that constitute the road network of Germany and will be referred to as *GER*. The detailed specifications of these transportation networks are summarized in Table 2.1.

Table 2.1: Specifications of data sets used for numerical experiments.

	N25	*CA339*	*GER*
Number of nodes	25	339	5,642
Number of potential OD nodes	25	51	430
Number of arcs	84	1234	16,034
Mean arc length (km)	4.60	18.96	11.07
Standard deviation of arc lengths	1.81	12.96	12.65

For *N25* and *CA339*, we work with 3 different values of R as commonly done in the literature (10, 12 and 15 for *N25*; 100, 150 and 200 for *CA339*). Each R leads to a different problem setting since the set of arcs of the expanded network depends on this value. For each setting, we identify the long-distance OD pairs by filtering out the trips with shortest path distances that are less than R. These OD pairs represent all possible trips on the network, where the vehicles have to recharge en route at least once in order to complete the trip. The details of these settings in terms of the number of OD pairs, expanded

network arcs and distance related measures are depicted in Table 2.2. Throughout the computational experiments with these networks, f_i, $i \in N^o$ is set to 10^6, and a unit demand between OD pairs is considered. In this setting, the magnitude of f_i is high enough to ensure that total recharging is minimized among the solutions which have the minimum total cost of opening facilities. In these computational studies, we assume that the energy consumption of an arc is directly proportional to its distance.

Table 2.2: Specifications of *N25* and *CA339* problem settings.

	R	**# OD**	**# Expanded arcs**	**Mean OD distance**	**Max OD distance**
N25	10	211	373	17.68	38
	12	181	466	18.87	38
	15	133	585	20.97	38
CA339	100	806	24,358	193.64	463.5
	150	542	44,524	227.43	463.5
	200	335	63,325	258.91	463.5

GER network has two variants with 818 and 1397 nodes introduced by (Göpfert and Bock, 2019), which include different subsets of its 5642 nodes. We test both of these variants and label them as *GER818* and *GER1397*. It is known that the majority of the EVs have already started to offer driving ranges greater than 300km. The number of the expanded network arcs increases substantially for higher R values so we believe that investigating how the proposed solution algorithm would behave for higher R values is important. For these reasons, we set R to 300, 350 and 400 km for both *GER818* and *GER1397*. For each R, we test different cardinalities of OD pair sets ($|K| = 50, 100, 200, 500$) to evaluate its effect on the computational performance. The OD demands are based on the traffic data from the year 2010 (Göpfert and Bock, 2019) and these sets are created by selecting the most popular $|K|$ long-distance OD pairs of the networks. The details of the problem settings, representatively for $|K| = 500$, in terms of expanded network size, number of OD

nodes, distance related measures (mean and maximum OD distances) and the percentage of flow volume included with respect to the total long-distance flow in each problem setting are shown in Table 2.3.

Table 2.3: Specifications of the *GER818* and *GER1397* problem settings for $|K| = 500$.

	R	# Expanded arcs	Mean OD distance	Max OD distance	# OD nodes	Total flow volume included (%)
GER818	300	122,450	448.1	747.2	132	86.9
	350	160,155	518.2	873.3	129	92.6
	400	202,012	586.8	994.8	135	98.6
GER1397	300	348,010	432.9	747.2	170	64.3
	350	456,312	500.7	874.8	174	71.6
	400	573,252	574.4	997.9	167	79.1

2.4.2 Effectiveness of the FCP and p-FCP Solutions

We first solve all three *N25* problem settings using the FCP formulation and record the number of used facilities, referred to as $\overline{p}$. Then, for comparison, we solve the MCP for the same setting using the number of facilities set to $\overline{p}$ at four deviation tolerance values (0%, 20%, 50% and 100%). Note that these tolerances represent the maximum allowed deviation distance as a percentage of the corresponding shortest path length. For each instance, we repeat solving the MCP at these deviation tolerances by increasing $\overline{p}$ one at a time until all OD pairs are covered to identify the minimum p value that ensures 100% coverage. Such solutions are equivalent to the set cover solutions, *i.e.,* the solutions provide 100% coverage with the minimum number of located facilities under the MCP setting. We also report the p-FCP solutions that uses the minimum p values for the set cover solutions. The first set of MCP solutions enables comparing the output of both models for the same number of open facilities. The second set of solutions (set cover), on the other hand, allows comparing both models for the case of 100% coverage.

Table 2.4 summarizes the results with *N25*. The solutions are compared on three main groups of performance measures: number of stations located, coverage based attributes, and distance based attributes. For each R, the first set of rows present the solutions of the FCP and p-FCP and the four subsequent rows show the MCP solutions corresponding to different deviation tolerances. Similarly, the last four rows correspond to the solutions of the set cover problem for various deviation tolerances and corresponding minimum p values.

When $R = 10$ and $p = 8$, neither of the four MCP solutions can provide full coverage; in fact, to cover 100%, 13 charging stations are required when a maximum of 50% deviation from the shortest path is allowed, whereas this number increases to 17 when the deviation tolerance is set to 0% or 20%. The FCP solution dominates both of the MCP solutions with 100% deviation tolerance over all performance measures. Even though the lower deviation tolerance settings for $p = 8$ might seem to provide better metrics for the last three columns, this is evidently biased since this solution leaves significantly many OD pairs uncovered.

When $R = 12$, the FCP solution look more effective. It dominates the MCP solutions at both 50% and 100% deviation tolerance levels when p is set to be the same as in the FCP solution. Unlike for $R = 10$, the MCP with 100% deviation tolerance at the same p can provide full coverage. In order to ensure complete coverage considering the deviation tolerance (set cover setting), the MCP needs to locate one, six and eight additional facilities for 50%, 20% and 0% deviation tolerance settings, respectively, which is still not enough to make 50% setting solution better than the FCP solution on any other measure.

When $R = 15$, the FCP solution looks worse on distance based attributes listed on the last three columns simply because the coverage of the MCP decreases substantially. Despite this, the FCP always provides a better coverage than the MCP when $p = 5$ as none of the deviation tolerance settings is sufficient to have 100% coverage for the FCP.

Table 2.4: Optimal FCP and p-FCP solutions versus optimal MCP and set cover solutions for *N25*.

	Deviation Tolerance	# Located Facilities (p)	Coverage	# Uncovered OD pairs	Avg Distance	Avg Deviation	Max Deviation
$R = 10$							
FCP	**N/A**	**8**	**100%**	**0**	**22.51**	**4.83**	**22**
p-FCP	**N/A**	**10**	**100%**	**0**	**19.25**	**1.56**	**12**
	N/A	**13**	**100%**	**0**	**18.10**	**0.42**	**4**
	N/A	**17**	**100%**	**0**	**18.04**	**0.36**	**4**
MCP	0%	8	52.6%	100	17.06	0	0
	20%	8	63.98%	76	18.30	0.71	5
	50%	8	82.46%	37	19.89	2.99	11
	100%	8	96.6%	7	24.09	6.21	38
Set Cover	0%	17	100%	0	17.68	0	0
	20%	17	100%	0	18.06	0.38	6
	50%	13	100%	0	21.82	4.14	16
	100%	10	100%	0	24.44	6.75	26
$R = 12$							
FCP	**N/A**	**7**	**100%**	**0**	**21.99**	**3.12**	**17**
p-FCP	**N/A**	**8**	**100%**	**0**	**20.20**	**1.32**	**7**
	N/A	**13**	**100%**	**0**	**19.46**	**0.59**	**6**
	N/A	**15**	**100%**	**0**	**19.24**	**0.37**	**6**
MCP	0%	7	55.8%	79	17.21	0	0
	20%	7	77.9%	40	19.22	1.32	4
	50%	7	94.47%	10	22.67	3.71	15
	100%	7	100%	0	25.02	6.14	29
Set Cover	0%	15	100%	0	18.87	0	0
	20%	13	100%	0	19.35	0.47	5
	50%	8	100%	0	25.07	6.20	17
	100%	7	100%	0	25.02	6.14	29
$R = 15$							
FCP	**N/A**	**5**	**100%**	**0**	**30.85**	**9.88**	**26**
p-FCP	**N/A**	**6**	**100%**	**0**	**23.58**	**2.60**	**15**
	N/A	**7**	**100%**	**0**	**22.05**	**1.08**	**8**
	N/A	**9**	**100%**	**0**	21.32	**0.35**	**6**
	N/A	**12**	**100%**	**0**	**21.13**	**0.16**	**4**
MCP	0%	5	48.12%	69	19.76	0	0
	20%	5	65.41%	46	20.55	0.78	3
	50%	5	79.69%	27	24.41	4.68	15
	100%	5	90.97%	12	28.91	7.71	31
Set Cover	0%	12	100%	0	20.97	0	0
	20%	9	100%	0	21.71	0.75	5
	50%	7	100%	0	29.15	8.18	17
	100%	6	100%	0	30.68	9.71	32

The p-FCP solutions show the strength of the proposed formulation as p increases, since the distance based attributes get significantly better at each increment. Specifically for $R = 15$, the improvement compared to the FCP solution by locating a single additional facility is substantial. Hence, such p-FCP solutions can easily dominate the corresponding set cover solutions. However, when p is high enough to provide a feasible set cover solution for 0% deviation tolerance, we observe that the distance based attributes of set cover solutions might be slightly better than the p-FCP solutions. This is due to the fact that the distance based attributes are not directly considered in the p-FCP objective function.

Figure 2.5 illustrates three histograms of the shortest path deviation distributions. These figures compare the FCP solutions of each setting with the corresponding MCP solution for the same p value and 50% deviation tolerance. The horizontal axis represents deviation intervals and the vertical axis shows the number of OD pairs belonging to each interval. The MCP solution for $R = 10$ leaves 37 OD pairs uncovered with $p = 8$ and 50% deviation tolerance, and it renders a maximum deviation of 11. As shown in Figure 4(a), the FCP solution has only 25 OD pairs that are over the deviation tolerance of 11 units, which is less than those 37 not covered by the MCP solution. These histograms illustrate that the distribution of deviations for both approaches are similar for the first three bins. The number of OD pairs covered in the first two bins is 11 higher for the FCP on $R = 10$. So, these results reveal that the FCP solution dominates the MCP solution in terms of deviation distributions. Figure 4(b) supports these findings for $R = 12$ since the number of OD pairs in the first bin is much higher for the FCP solution. On the other hand, Figure 4(c) shows that deviation distances of the FCP solution are longer compared to 79% coverage solution of the MCP, which leaves 27 OD pairs uncovered. This is expected since leaving that many OD pairs allows the MCP to cover only the nodes in concentrated zones.

An additional set of experiments with the p-FCP are performed. Using p-FCP helps find

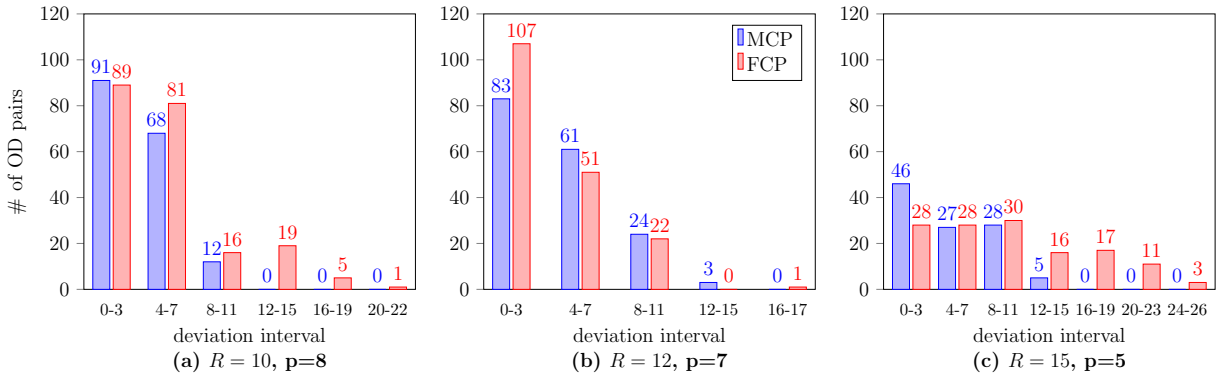


Figure 2.5: Deviation histograms for the FCP and MCP, deviation tolerance = 50%.

better solutions in terms of routing measures when p is sufficient to ensure 100% coverage. Table 2.5 shows the results of the p-FCP and MCP at varying levels of p. The lowest p value selected for each setting is the lowest possible value to ensure complete coverage with the FCP formulation, *i.e.,* any lower p will result in infeasibility of this model. The MCP solutions are obtained at 50% deviation tolerance as we observed that any lower value results in significantly lower coverage rates, which leads to solutions that are not insightful to compare.

For $R = 10$, each additional facility improves the FCP solutions in terms of routing-related metrics. However, the MCP does not guarantee to find a better solution in terms of these metrics when p is increased. Thus, it may yield fluctuating values of them that are always worse than the FCP solution, except for $p = 8$. Nevertheless, increasing p values allows the MCP to improve coverage. Similarly, all $R = 12$ solutions of the FCP dominate the MCP solutions by a large margin except for $p = 7$. Solutions for $R = 15$ demonstrate the strength of using the FCP as its maximum deviation metric is even comparable to the average deviation of the MCP solutions when $p \geq 7$. Opening one more facility than the lowest p for the FCP provides much better solutions in terms of routing-related performance measures as depicted in Figure 2.6. These results highlight the dominance of the FCP formulation over the max cover/set cover on routing-related metrics especially

when the p-FCP is used and the number of open facilities is the same. Figure 5 illustrates that these metrics of the FCP solutions are superior than those of the MCP solutions in addition to ensuring complete coverage of all the OD pairs. It also shows the fluctuation in the MCP solutions more clearly.

Table 2.5: The p-FCP and MCP solutions for *N25*.

$R = 10$		$p = 8$	$p = 9$	$p = 10$	$p = 11$	$p = 12$	$p = 13$
p-FCP	Avg Recharge(R)	1.61	1.42	1.24	1.18	1.12	1.06
	Avg Dev	4.83	3.12	1.56	0.99	0.50	0.42
	Max Dev	22	17	12	8	6	4
MCP	Coverage (%)	82.46	90.04	95.26	98.57	99.53	100
	Avg Dev	2.99	6.35	7.31	7.39	4.88	4.14
	Max Dev	11	16	19	19	16	16
$R = 12$		$p = 7$	$p = 8$	$p = 9$	$p = 10$	$p = 11$	$p = 12$
p-FCP	Avg Recharge(R)	1.22	1.04	0.96	0.91	0.86	0.81
	Avg Dev	3.12	1.32	0.81	0.64	0.59	0.72
	Max Dev	17	7	6	6	6	6
MCP	Coverage (%)	94.47	100	100	100	100	100
	Avg Dev	3.71	6.20	3.97	4.73	5.44	5.70
	Max Dev	15	17	15	19	18	19
$R = 15$		$p = 5$	$p = 6$	$p = 7$	$p = 8$	$p = 9$	$p = 10$
p-FCP	Avg Recharge(R)	1.48	0.98	0.87	0.79	0.72	0.66
	Avg Dev	9.88	2.60	1.08	0.51	0.35	0.35
	Max Dev	26	15	8	6	6	6
MCP	Coverage (%)	79.69	95.49	100	100	100	100
	Avg Dev	4.68	4.93	8.18	6.21	4.09	6.06
	Max Dev	15	15	17	19	15	13

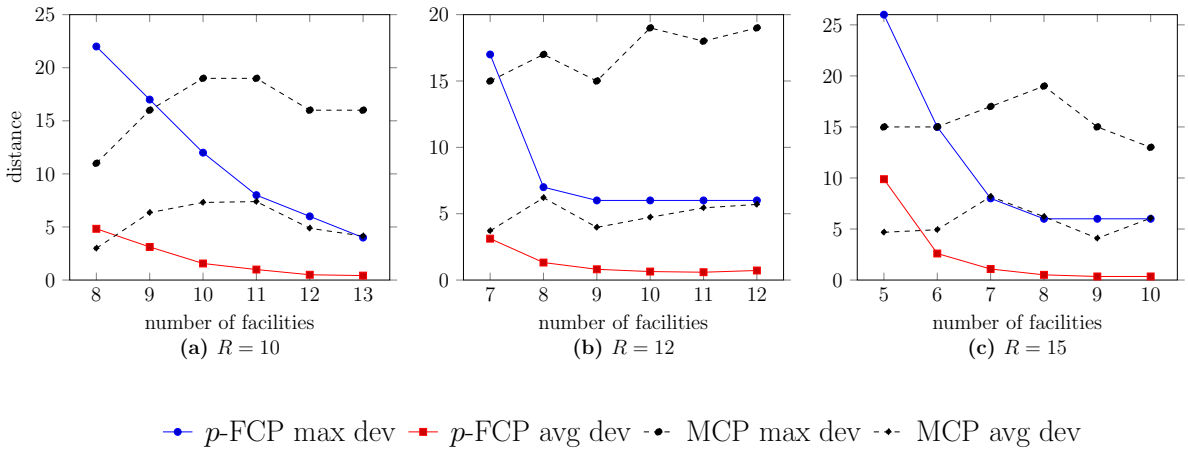


Figure 2.6: Comparison of routing metrics on *N25*.

The optimal solutions of the FCP and p-FCP for *CA339* instances are presented next. The results are depicted in Table 2.6. The number of open facilities in each instance are listed in the second column and the routing-related performance measures are presented on the subsequent columns. Total en route recharging is presented as a factor of the vehicle range R, therefore, the unit of measurement of average recharge metric shown on the third column of this table is R. In other words, since R is an indicator of the total battery capacity, if a vehicle with the range of R completes 1.50R of total en route recharging, the vehicle needs to have at least 150% of its battery to be filled up along the trip before reaching its destination.

Table 2.6: The FCP and p-FCP solutions for *CA339*.

	p	**Avg Recharge (R)**	**Avg Trip Distance (km)**	**Avg Deviation (km)**	**Max Deviation (km)**
$R = 100$	8 (FCP)	2.52	262.27	68.63	232.53
	9	1.99	212.47	18.83	123.24
	10	1.89	206.95	13.31	123.24
	11	1.83	201.75	8.11	84.01
	12	1.78	200.04	6.40	84.01
	13	1.75	198.63	4.99	84.01
$R = 150$	5 (FCP)	1.92	295.23	67.80	192.90
	6	1.52	240.75	13.32	81.78
	7	1.42	235.18	7.03	74.34
	8	1.38	235.18	7.75	74.34
	9	1.34	233.21	5.78	74.89
	10	1.31	233.81	6.38	68.03
$R = 200$	4 (FCP)	1.36	289.84	30.93	192.90
	5	1.21	267.54	8.62	113.64
	6	1.15	260.91	2.00	37.68
	7	1.12	263.87	4.95	55.74
	8	1.08	263.42	4.50	55.74
	9	1.04	263.42	4.50	55.74

The first row of each R setting depicts the FCP solution and the subsequent ones with incremented p values are the p-FCP solutions. The improvement on routing measures is substantial for the p-FCP even after opening one additional facility. These results validate the effectiveness of using the p-FCP. Each additional p improves the coverage of the MCP

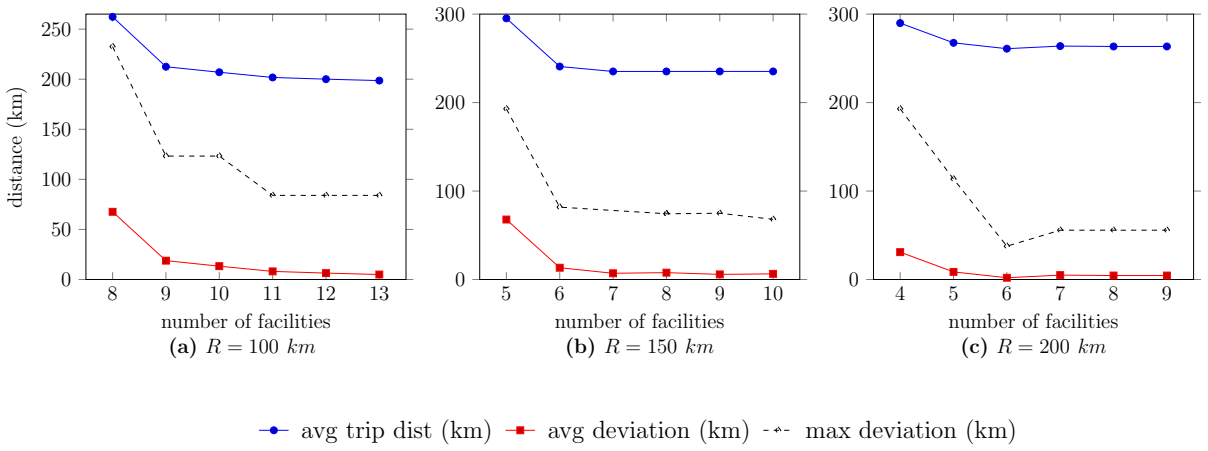


Figure 2.7: Comparison of routing metrics for *CA339*.

but is not always helpful to determine a solution with better routing metrics. However, these metrics improve as p increases in the FCP solutions. While p increases, the maximum deviation metric of the p-FCP solutions gets nearly as good as average deviation of the MCP solutions. In addition to superior coverage, this is the other main benefit of using the FCP rather than the MCP to obtain higher quality solutions for the RSLP-R applications.

Figure 2.7 presents a visual representation of the distance based attributes. Even though some FCP solutions might result in high deviations for certain OD trips, by solving the p-FCP iteratively, the trade-off between p and the maximum deviation can be evaluated by the decision makers. The significant improvements in distance based attributes in the first few increments in p can be seen in Figure 2.7. Such managerial results are insightful for the decision makers to elucidate the marginal benefits of locating additional facilities.

2.4.3 Performance of Benders Decomposition and Algorithm 2.1

In this section, we first evaluate the computational perfomance of the proposed solution methodology on the *CA339* dataset. The size of the expanded network grows significantly with the size of the original graph, for example, the California highway network results in

more than 20 million decision variables for each problem setting ($R = 100, 150$ and 200). State of the art solvers or computers cannot handle a problem of this size; yet using the proposed Benders decomposition framework coupled with the new subproblem solution algorithm, we can solve all instances to optimality in less than a minute for *CA339*.

Table 2.7: Performance comparison of the Benders algorithm for *CA339*.

		CPLEX Subproblems				**Algorithm 2.1**		
	p	**# Feas Cuts**	**# Opt Cuts**	**Sub-p Time (%)**	**Solution Time (s)**	**# Feas Cuts**[†]	**# Opt Cuts**	**Solution Time (s)**
$R = 100$	8	54	15	88.9	819	264 (157)	55	27.4
	9	109	180	95.0	4,607	175 (77)	141	31.7
	10	94	220	96.1	5,426	170 (88)	134	35.8
	11	73	224	95.8	4,615	107 (65)	163	37.2
	12	65	218	94.2	3,486	35 (86)	234	51.3
	13	87	252	95.7	5,922	96 (57)	213	50.9
$R = 150$	5	26	31	94.9	816	44 (38)	49	9.3
	6	43	73	97.6	2,674	57 (9)	45	7.1
	7	29	127	96.5	3,517	58 (17)	61	11.0
	8	35	181	97.6	5,159	53 (29)	79	16.3
	9	22	198	96.9	4,413	55 (26)	67	15.1
	10	26	198	97.5	6,756	55 (31)	96	20.1
$R = 200$	4	27	57	95.5	772	42 (4)	22	2.9
	5	23	122	97.5	2,712	28 (12)	52	6.8
	6	19	122	96.9	2,583	36 (11)	17	2.8
	7	17	114	96.1	1,969	41 (15)	44	6.7
	8	20	86	97.4	2,485	31 (17)	57	8.3
	9	12	83	97.1	2,511	28 (19)	52	8.7
Average:		43.3	138.9	95.9	3,402	76.4 (42.1)	86.7	19.4

[†] For a fair comparison, the number of iterations when $\mathcal{SP}$ is infeasible is presented in parenthesis next to the number of added feasibility cuts as Algorithm 2.1 adds multiple cuts each time $\mathcal{SP}$ is infeasible.

We present the computational performance comparison of the subproblem solution algorithm against CPLEX in Table 2.7 for each instance on *CA339*. On the left hand side of this table, we list the number of feasibility and optimality cuts added along with the total percentage of solution time spent on solving the subproblems by CPLEX and the total solution time in seconds. All problem settings are solved to optimality within reasonable computational times. The shortest solution time is 772s ($R = 200$, $p = 4$), whereas the longest and the mean time to reach optimal solutions are 6756s (1.88hr) and 3402s (0.95hr), respectively. Over 95% of the solution time is spent on solving the subproblems.

On the right hand side of Table 2.7, we list the number of added feasibility and optimality cuts, and the total solution time for each *CA339* instance when the subproblem solution methodology is employed within the Benders algorithm. For $R = 100, 150$ and 200, the algorithm is 106, 288 and 411 times faster than CPLEX, respectively. Average solution time decreases to 19.4s from 3402s. The shortest and longest solution times turn out to be 2.8s ($R = 200$, $p = 6$) and 51.3s ($R = 100$, $p = 12$), respectively, where for the instance with $R = 200$ and $p = 6$, the algorithm is 922 times faster than CPLEX. The comparison also reveals that the number of optimality cuts that are added by the solution algorithm is less than that of CPLEX, except for three instances: $R = 100$, $p = 8, 12$, and $R = 150$, $p = 5$. The number of added feasibility cuts tend to be higher since the algorithm adds multiple cuts everytime an infeasible solution is encountered. On the other hand, the number of iterations where an infeasible $\mathcal{SP}$ is encountered is comparable with the number of feasibility cuts added by CPLEX.

Next, we explore the limits of the developed solution methodology and evaluate its computational performance with respect to varying number of nodes and expanded network arcs using larger-sized instances on *GER818* and *GER1397* networks. We test instances with three different range values ($R = 300, 350, 400$), four different sets of OD pairs ($|K| = 50, 100, 200, 500$), and varying number of facilities to open (p). These combinations yield to over 110 instances using *GER818* and *GER1397*. The computational performance of the algorithm on these instances are depicted in Table 2.8.

We observe that the computational effort to solve these instances are more than those of the *CA339* since the computation times are significantly higher and their variance is also larger. The solution algorithm is still able to find the optimal solutions for the majority of instances within the predetermined time limit. The average computation time of optimally solved instances is 556.8s, whereas the maximum is 2615.2s. For *GER818*, 20 out of 57 instances remain unsolved within the time limit. The average optimality gap is 1.68% for

Table 2.8: Performance of the Benders algorithm for *GER818* and *GER1397*.

		GER818			***GER1397***		
	$\|K\|$	**p**	**Solution Time (s)**	**Gap (%)**	**p**	**Solution Time (s)**	**Gap (%)**
$R = 300$	50	11 (FCP)	22.3	0	14 (FCP)	3,600	15.36
		15	3,600	9.67	15	3,600	14.03
		20	20.8	0	20	247.6	0
		25	69	0	25	1,344.4	0
		30	64.6	0	30	3,600	0.62
	100	15 (FCP)	168.1	0	15 (FCP)	3,600	19.89
		20	2,050	0	20	3,600	8.81
		25	3,600	1.45	25	3,600	3.09
		30	3,600	0.98	30	3,600	1.72
	200	16 (FCP)	98.3	0	19 (FCP)	3,600	14.07
		20	1,177.7	0	20	3,600	5.33
		25	3,600	0.52	25	3,600	8.08
		30	3,600	2.37	30	3,600	3.41
		35	3,600	1.24	35	3,600	2.18
	500	18 (FCP)	239.9	0	-	-	-
		20	905	0	20 (FCP)	722.5	0
		25	3,600	6.02	25	3,600	9.42
		30	3,600	5.85	30	3,600	7.89
		35	3,600	3.82	35	3,600	4.97
			Average:	**1.68**		**Average:**	**6.60**
$R = 350$	50	9 (FCP)	135.3	0	-	-	-
		10	288.0	0	10 (FCP)	162.3	0
		15	46.6	0	15	851.1	0
		20	38.3	0	20	3,600	1.18
		25	23.6	0	25	645.7	0
	100	11 (FCP)	108.9	0	12 (FCP)	3,600	1.15
		15	334.1	0	15	778.4	0
		20	51.2	0	20	1,965.6	0
		25	80.5	0	25	1,399.0	0
		30	74.4	0	30	383.7	0
	200	14 (FCP)	814.9	0	14 (FCP)	245.9	0
		15	2615.2	0	15	1,078.1	0
		20	224.3	0	20	3,600	1.48
		25	3,600	0.17	25	3,600	0.31
		30	3,600	0.22	30	1,281.4	0
	500	15 (FCP)	1,361.9	0	15 (FCP)	637.7	0
		20	3,600	1.41	20	3,600	5.39
		25	3,600	3.72	25	3,600	3.20
		30	3,600	2.87	30	3,600	1.60
			Average:	**0.44**		**Average:**	**0.81**
$R = 400$	50	8 (FCP)	34.4	0	8 (FCP)	604.5	0
		10	124.1	0	10	149.1	0
		15	41.7	0	15	60.4	0
		20	26.2	0	20	43.6	0
		25	20.5	0	25	47.1	0
	100	9 (FCP)	90.1	0	-	-	-
		10	104.3	0	10 (FCP)	410.1	0
		15	2,087.2	0	15	552.5	0
		20	310.0	0	20	3,600	0.37
		25	3,600	0.32	25	385.2	0
	200	10 (FCP)	29.5	0	10 (FCP)	642.1	0
		15	255.8	0	15	1,805.4	0
		20	3,600	0.50	20	3,600	1.95
		25	3,600	0.58	25	3,600	1.35
	500	13 (FCP)	1,810.1	0	13 (FCP)	3,600	5.43
		15	1,586.6	0	15	3,600	2.19
		20	3,600	1.38	20	3,600	2.21
		25	3,600	2.07	25	3,600	2.74
		30	3,600	2.72	30	3,600	2.53
			Average:	**0.40**		**Average:**	**1.04**

$R = 300$ instances whereas it decreases to 0.44% for $R = 350$, and to 0.40% for $R = 400$. On the other hand, the majority of the instances (30 out of 54) of *GER1397* remain with an optimality gap after an hour. The average percentages of gaps are more than doubled compared to *GER818*: 6.60%, 0.81% and 1.04% for $R = 300, 350$ and 400, respectively. Interestingly, considering fewer OD pairs does not always require less computational effort, particularly for $R = 300$ and lower p values, where the recorded optimality gaps are the largest on both networks. On the other hand, the instances of $R = 350$ and $R = 400$ with higher number of OD pairs typically require higher computational effort to solve than the lower $|K|$ ones.

2.5 Conclusions

This chapter introduces the first study in the EV charging station location problem literature that allows shortest path deviations without pregenerating the routes or using a deviation tolerance. Since the widely used max cover approach and deviation tolerance may lead to building ineffective networks that leave many OD pairs uncovered, we propose a new full cover modeling approach to design a charging station infrastructure that enables long-distance EV travel on a transportation network. Not having a predetermined deviation tolerance is observed to significantly increase the computational effort to solve the problem. Hence, in order to solve large-size instances, we developed a Benders decomposition algorithm. Computational experiments revealed that almost 95% of the overall computational effort is spent on solving the Benders subproblems. Thus, we introduced a novel method to solve these subproblems. To this end, as a major methodological contribution, we construct the optimal dual solution to generate non-dominated optimality cuts when the subproblem is feasible; else, we derive valid inequalities based on the infeasible subgraph structures and add these to the master problem as feasibility cuts.

Numerical results support the effectiveness of the new modeling approach over the existing models in the literature and demonstrate its strengths. FCP models can find solutions that dominate the MCP solutions in terms of routing-related performance metrics, when the same number of stations are open and the coverage of the MCP is closer to 100%. While testing the p-FCP formulation on highway network of California, we observe that opening additional stations offer substantial improvements over the routing measures compared with the FCP solution which minimizes the number of stations as a priority. Even though these measures are not explicitly considered while minimizing total en route recharging, computational results reveal that they are still captured by this objective function.

When the performance of the subproblem solution algorithm is compared with CPLEX, the results reveal that the proposed algorithm can speed-up the solution up to 900 times on the tested large-size instances. We explore the computational limits of the solution algorithm using over 100 instances of the Germany road network and determine that the computational performance is proportional to the graph size and the number of nodes and inversely proportional to the vehicle range.

We believe that the core problem of charging infrastructure design for EVs shows much potential in terms of variety of directions that the researchers can look into. In this work, we assume that the charging stations are uncapacitated. Since the adoption rate of EVs is still considerably low, it is reasonable to assume that these stations will not be fully utilized in the near future. However, once EVs become the popular choice of transportation, it will be crucial to take the station capacities, waiting times in the queue, and vehicle heterogeneity into account. In connection with these aspects, it would also be important to consider nonlinearity in charging times for EVs alongside the effects of traffic congestion to conduct a comprehensive research that focuses on trip time rather than distance based measures.

Chapter 3

Charging Station Location and Sizing for Electric Vehicles Considering User Behavior and Congestion

The majority of shorter distance trips using EVs typically rely on the private slow-chargers readily available at homes and workplaces. However, private slow-charging has to be complemented by publicly accessible en route fast-charging stations to support EV charging on the go for long-distance trips. Furthermore, recent advancements in battery and charging technologies for passenger EVs are paving the way towards long-haul freight distribution with heavy-duty EVs, i.e., electric trucks (ETs). Lately, manufacturers such as Tesla, Daimler AG, and Volvo, have unveiled their plans for mass production of ETs within this decade. International Energy Agency states that planning needs to start now for mega chargers to enable long-distance electric trucking as there is currently no infrastructure to support it (IEA, 2021). Therefore, the feasibility and convenience of long-distance trips with electric vehicles depend on the spatial distribution of adequately capacitated DC fast-charging stations.

The capacity of a DC fast-charging station is defined as the number of charging stalls, also known as outlets, chargers, or EVSE. Considering the relatively long recharging times for EVs even using fast-chargers, waiting times may become a major inconvenience. Moreover, the demand for charging at a charging station is hardly known in advance. Hence, the stochasticity of demand and service times should not be neglected when determining the locations and capacities of the charging stations.

In this chapter, we seek to develop mathematical models to find the optimal locations of DC fast-charging stations and their respective number of chargers to serve a given set of origin-destination (OD) EV trips on a transportation network. To ensure quality of service considering stochasticity in charging demand and service times, probabilistic service level requirements are defined that make sure that the expected waiting times at a charging station do not exceed inconvenience limits. To provide cost-effective solutions, the framework of this problem allows the possibility of detours from shortest paths within a threshold, referred to as the *deviation tolerance* (Arslan et al., 2019).

Given the possibility of detours, when there exist alternative routing options, the decision makers need to consider the choice of charging stations to be visited by rational service-seekers who act with self-interest. This necessitates the route choice behavior of EV users to be taken into account when the location and sizing decisions are made. We use a bilevel structure where the *leader* is a regulatory body that seeks to establish the charging infrastructure by determining the strategic locations and sizes of the charging stations (e.g. governments or private EV network operators), and the *follower* is the set of EV users who seek to complete their long-distance EV trips.

The leader's problem minimizes the total cost of locating and sizing charging stations while ensuring a convenient service level taking follower's response into account. The follower's problem, on the other hand, is defined to incorporate the route and charging

station choice of EV users while minimizing the length of their trips. The choice of facilities by the EV users, i.e., the allocation of recharging demand to charging stations, is crucial for the leader since the congestion level at stations will directly impact the required capacity.

To the extent of our knowledge, this study is the first to model a chance-constrained bilevel optimization framework and that makes charging station sizing decisions based on stochastic queuing models focusing on long-distance transportation with EVs. The main contributions of this work are (i) a new bilevel mathematical model that allows shortest path deviations without path pregeneration or enumeration of alternative routes, (ii) merging the bilevel nature of the problem with realistic queuing models to incorporate stochastic travel demand, (iii) an exact solution algorithm for the uncapacitated version of the problem, (iv) and a novel decomposition based algorithm that is capable of rapidly finding high-quality solutions and also able to handle both cooperative (optimistic) and uncooperative (pessimistic) responses of EV users. Through computational experimentation, we evaluate the performance of the solution algorithm and show how the follower's position, deviation tolerance, and service level parameters affect the obtained solutions on benchmark instances and real life California and Eastern US highway networks.

The next section of the chapter provides the literature review of the capacitated facility location problems regarding charging stations while emphasizing the contributions. Section 3.2 gives the problem statement, defines the underlying expanded network and its reduction procedure, and presents the bilevel modeling framework. Section 3.3 provides the analysis of the bilevel optimization model under cooperative and uncooperative EV user response, and M/M/c queuing systems. Section 3.4 presents the logic-based Benders decomposition algorithm developed for the uncapacitated version of the problem and a decomposition based algorithm that can account for the cooperative or uncooperative user behavior for the original problem. Section 3.5 introduces the data sets and test instances and presents the results of extensive computational experiments performed to derive man-

agerial insights as well as to assess the performance of the algorithm. Lastly, Section 3.6 includes a summary of the insights, concluding remarks, and key takeaways.

3.1 Literature Review

Many studies in the literature adopt a focus that is solely based on the spatial dimension of the charging infrastructure problem within an uncapacitated framework. The corresponding literature have already been extensively reviewed in Section 2.1. Here, the scope of the review is on the studies that use a congestion-aware setting for the EV charging station problems.

The location-specific studies that ignore capacity decisions at charging stations consequently neglect EV users' route choice response to facility locations or congestion. On the other hand, some existing studies which model the capacitated extensions of the charging station location problem (e.g., Upchurch et al., 2009; Jiang et al., 2012; Hosseini et al., 2017) adopt a rather simplistic approach that imposes hard limits on the number of vehicles that can be served by each open station instead of assuming that all flow passing through a station can be served regardless of their volume. Additionally, there are studies that consider capacity level expansions under a multi-period framework. Among these, Zhang et al. (2017) extends the capacitated FRLM considering the demand dynamics (increasing EV market share) throughout the planning horizon. Anjos et al. (2020) uses a hybrid approach that takes both node-based (urban) and flow-based (long-distance) travel demands into account.

There are stochastic modeling approaches that use queuing theory to determine station capacity levels. Xie et al. (2018) and Xie and Lin (2021) use M/M/c queuing systems to model facility congestion under the context of multi-period capacity expansion and charg-

ing station location problem. The former proposes a stochastic optimization model that minimizes total system cost defined as the sum of the capital cost of locating and expanding charging stations and the penalty cost of not satisfying (covering) an OD trip. The deviations from the shortest paths are taken into account by means of pregenerating a number of alternative paths for each OD trip. Even though deviations are considered, all alternative paths are assumed to be equivalent for the EV users; i.e., EV users' response to charging station locations are neglected. The resulting mathematical model is solved by a genetic algorithm under the assumption of last-minute charging which may cause suboptimal demand allocation to charging stations. The computational experiments are based on a case study in California. Xie and Lin (2021) extends this framework by introducing an inconvenience cost function of travellers in their system cost definition. This model is also solved by a genetic algorithm but it no longer allows deviations from the shortest paths.

There are several studies which take the route choice behaviour of EV users into account using bilevel location models under an uncapacitated framework (e.g., Jing et al., 2017; Zheng et al., 2017; He et al., 2018; Guo et al., 2018; Tran et al., 2021). These studies are more suitable for an urban setting as they consider traffic congestion using stochastic or deterministic user equilibrium. Hence, the case studies and numerical tests featured in these works cover smaller geographical regions. These mathematical models rely on path pregeneration or enumeration, except for Zheng et al. (2017) in which the state of charge is tracked in the model to ensure path feasibility. Bilevel optimization problems are known to be intrinsically hard to solve. Even the models with both linear leader and follower's problems, which are generally the simplest to solve, are shown to be strongly NP-hard (Labbé and Marcotte, 2021). Typically, solution methods used for these studies are metaheuristics, such as a genetic algorithm or large-scale neighbourhood search, or a single-level reformulation of the bilevel problems is proposed to be able to solve the models using commercial solvers.

To the best of our knowledge, there are only two studies to date which make charging station location and sizing decisions while taking EV users' response into account. Huang and Kockelman (2020) use a bilevel profit maximization problem for multi-period charging station location and sizing. Their model allows for incorporating charging price elasticity and station congestion into a discrete set of route choices for EV users under elastic demand. However, station capacity deployment is still modeled in a deterministic way. The numerical experiments are performed on an urban network with 74 nodes. A genetic algorithm is used to find solutions to the proposed model. This algorithm uses a queuing averse logit choice model that determines users' allocations to charging stations. Makhlouf et al. (2019) also propose a bilevel problem where the upper level problem is a max-cover type station location and sizing problem and the lower level problem represents EV-user behaviour in terms of making the minimum number of stops to reach their destination. The solution approach is based on a single-level reduction assuming cooperative response and they solve the resulting reformulation using commercial solvers on randomly generated network instances of 100 nodes.

To sum up, none of the above studies consider a bilevel framework that simultaneously takes the infrastructure cost of locating and sizing charging stations, route choice response of EV users, and stochastic waiting times at facilities into account. Moreover, the existing studies that use a bilevel model assume only a cooperative user behavior, whereas we develop a solution methodology that can be tailored for both the cooperative and uncooperative responses of the follower. This algorithm is capable of solving large-scale transportation networks with up to 420 nodes, which is the largest network size that has been solved within a bilevel framework in the literature.

3.2 The Bilevel Charging Station Location and Sizing Problem

Let the underlying transportation network be defined as $G^o = (N, A^o)$, where N denotes the set of nodes, e.g., highway intersections, population centres, and A^o denotes the set of directed arcs, e.g., highway segments. Consider K as the set of OD pairs, where O_k and D_k denote the origin and destination nodes for OD pair $k \in K$. Each OD pair represents a stochastic stream of EV flow between two population centres for which the travel distance exceeds the EV battery range R, and thus, recharging en route is necessary.

A charging station may contain up to C charging stalls and each stall offers stochastic service times. As a result of limited capacity, stochastic arrivals, and stochastic service times, charging stations may experience congestion. Consequently, EV users may have to enter a service queue when the system is busy, i.e., when all charging stalls are occupied.

The waiting time at a charging station is a function of its capacity given by the number of charging stalls and the total allocated EV charging demand. In order to ensure service convenience, a service level requirement is introduced to guarantee that the probability of waiting at most α minutes is greater than or equal to $\beta\%$ at each open station. For example, a service level requirement is for the waiting time to be less than or equal to $\alpha = 10$ minutes, $\beta = 90\%$ of the time.

A deviation tolerance is defined to model EV users' willingness to deviate from their shortest path to charge. This deviation tolerance τ is expressed in terms of the percentage difference of the length of the path taken and the length of a shortest path (e.g., Kim and Kuby, 2012; Yıldız et al., 2016; Arslan et al., 2019; Göpfert and Bock, 2019).

Let $f(\cdot)$ denote the fixed cost function of establishing a charging station and $g(\cdot)$ denote the variable cost function of capacity installation at the stations. The former is typically a

function of economic characteristics of the land of each candidate station location, whereas the latter is a non-decreasing function of the number of charging stalls installed at a station.

The bilevel optimization model minimizes the total infrastructure cost of locating charging stations and installing charging stalls, while ensuring feasible OD routes and probabilistic service level requirements. Feasible OD routes visit en route charging stations to satisfy range limitations of EVs and are within the respective deviation tolerance.

In order to model OD trip feasibility without tracking the state of charge of the vehicles or pregenerating all the feasible stopping combinations, the idea of an expanded network introduced by MirHassani and Ebrazi (2013), which was previously discussed in Section 2.2.2 of the previous chapter. Recall that the expanded network includes additional arcs on G^o between nodes $i, j \in N$, given that the shortest travel distance between them is less than or equal to the EV range. This allows embedding the EV range information in feasible network arcs, since if the shortest path between any two nodes can be completed without recharging, a stop for a recharge is not required at intermediary nodes.

Let $G = (N, A)$ be the expanded network of G^o for a given range R. We now show how to further reduce the expanded network by defining a permissible arc for OD pair $k \in K$ as follows:

Definition 3.1. *An expanded network arc $(i, j) \in A$ is called a permissible arc for OD $k \in K$ if it connects O_k to D_k within the allowed deviation tolerance. The set of permissible arcs A_k, $k \in K$ is given by:*

$$A_k = \{(i, j) : (i, j) \in A \text{ and } \delta_{O_k i} + \delta_{ij} + \delta_{jD_k} \leq (1 + \tau)\delta_{O_k D_k}\}, \qquad k \in K$$

where δ_{ij} is the shortest path length from $i \in N$ to $j \in N$.

We then construct $G_k = (N_k, A_k)$ as the permissible arc expanded network for $k \in K$,

where N_k consists of the nodes that are incident to the arcs in the set A_k. This reduced set A_k allows to define the mathematical model concisely by removing redundant constraints due to arcs that cannot be in a feasible path within the deviation tolerance. However, modeling the problem on $\cup_{k\in K} G_k$ rather than G does not ensure that the deviation tolerance requirement is met.

To formulate the bilevel charging station location and sizing problem under congestion, we introduce the following decision variables:

$$y_j = \begin{cases} 1, & \text{if candidate location } j \in N \text{ has a charging station,} \\ 0, & \text{otherwise.} \end{cases}$$

$z_j =$ number of charging stalls installed at charging station $j \in N$.

$$x_{ijk} = \begin{cases} 1, & \text{if expanded arc } (i,j) \in A_k \text{ is traversed by OD pair } k \in K, \\ 0, & \text{otherwise.} \end{cases}$$

$W_j =$ waiting time at charging station $j \in N$.

The bilevel model BLP is then formulated as follows:

$$\text{[BLP] “minimize”} \quad \sum_{j\in N} \left(f(y_j) + g(z_j)\right) \tag{3.1}$$

$$\text{s.t.} \quad z_j \le C y_j \qquad j \in N \tag{3.2}$$

$$\mathbb{P}\big(W_j \le \alpha\big) \ge \beta \qquad j \in N \tag{3.3}$$

$$y_j \in \{0,1\} \qquad j \in N \tag{3.4}$$

$$z_j \ge 0 \text{ and integer} \qquad j \in N \tag{3.5}$$

$$W_j \ge 0 \qquad j \in N \tag{3.6}$$

$$x_{ijk} \in \operatorname{argmin} \sum_{k\in K} \sum_{(i,j)\in A_k} \delta_{ij} x_{ijk} \tag{3.7}$$

$$\text{s.t.}$$

$$\sum_{j:(i,j)\in A_k} x_{ijk} - \sum_{j:(j,i)\in A_k} x_{jik} = \begin{cases} 1, & \text{if } i = O_k, \\ -1, & \text{if } i = D_k, \\ 0, & otherwise. \end{cases} \quad i \in N_k, \quad k \in K \tag{3.8}$$

$$\sum_{i:(i,j)\in A_k} x_{ijk} \leq y_j \qquad j \in N_k, \quad j \neq D_k, \quad k \in K \tag{3.9}$$

$$\sum_{(i,j)\in A_k} \delta_{ij} x_{ijk} \leq (1+\tau)\delta_{O_k D_k} \qquad k \in K \tag{3.10}$$

$$x_{ijk} \in \{0,1\} \qquad (i,j) \in A_k, \quad k \in K \tag{3.11}$$

Bilevel optimization problems are often interpreted as Stackelberg games that are structured to have an upper-level (leader's) and a lower-level (follower's) problem. In this context, the leader makes a decision and the follower responds. The framework of such a problem is based on the fact that the leader anticipates the follower's optimal reaction to their decisions.

The objective function of the leader (3.1) minimizes total infrastructure cost comprised of the cost of locating charging stations and installing charging stalls. Constraint (3.2) ensures that station capacity may be non-zero only for open charging stations. Chance constraints associated with service level requirements are defined by (3.3). For each charging station, this constraint ensures that the probability of waiting less than or equal to α minutes is greater than or equal to β. Constraints (3.4)-(3.6) define the domains of the leader's decisions.

The objective of the follower's problem is to minimize the total travel distance. Constraint (3.7) defines the feasible solution space of the follower's routing decisions in response to the leader's charging station location decisions, which in itself is an optimization problem. Constraints (3.8) are flow balance constraints for OD $k \in K$. Constraint (3.9) ensures that there exists a charging station, located by the leader, at the head of each used arc, ex-

cept the final arc of each OD path. By the definition of the expanded network, this ensures vehicle range feasibility besides determining which charging stations to use on each OD path. Constraint (3.10) makes sure that each OD path is within the deviation tolerance τ. Lastly, constraint (3.11) defines the domain of the flow variables x_{ijk}.

The quotation marks are used in the definition of (3.1) as there is a lack of clarity in case there are alternative optimal solutions for the follower's problem for a given solution vector of the leader's problem. This can be resolved by defining the notions of cooperative and uncooperative responses of the follower. In a cooperative response, the leader anticipates the follower to choose from among the optimal solutions of the lower level problem the one that yields the best objective value for the leader's problem; i.e., a cooperation between the leader and follower exists. In the latter, the leader assumes a reaction of the follower that generates the worst possible outcome for the leader's objective function.

In the next section, we mathematically define both positions of the follower and then, in Section 3.3.1, show how BLP may be reformulated as a single level problem under the cooperative response. In Section 3.3.2, we introduce the characterization of a DC fast-charging station network as an open multi-server Jackson network, and using the Erlang-C function and queuing system stability constraints, we present the deterministic equivalents of the probabilistic service level constraints and modify the single-level reformulation of BLP under cooperative response as a mixed-integer linear optimization problem (MILP).

3.3 Analysis of the Bilevel Modeling Framework

Let $\Psi(\mathbf{y})$ define the reaction set of the follower based on the leader's decisions $\mathbf{y} = \{y_j, j \in N\}$ as:

$$\Psi(\mathbf{y}) = \left\{ \mathbf{x} = \{x_{ijk},\ (i,j) \in A_k,\ k \in K\} : \mathbf{x} \in \text{argmin} \left\{ \sum_{k \in K} \sum_{(i,j) \in A_k} \delta_{ij} x_{ijk} : (3.8) - (3.11) \right\} \right\}$$

The optimal solution of BLP under cooperative (optimistic) response $\mathbf{x}^o$ is:

$$\mathbf{x}^o = \text{argmin} \left\{ \sum_{j \in N} (f(y_j) + g(z_j)) : \mathbf{x}^o \in \Psi(\mathbf{y}) \right\}$$

The uncooperative position reflects the case where the leader protects themselves against the worst possible outcome of the follower's response. Based on the definition of the follower's alternative reaction set $\Psi(\mathbf{y})$, the optimal uncooperative (pessimistic) response solutions of BLP are defined as follows:

$$\mathbf{x}^p = \text{argmax} \left\{ \sum_{j \in N} (f(y_j) + g(z_j)) : \mathbf{x}^p \in \Psi(\mathbf{y}) \right\}$$

For the case of cooperative response and in the presence of convex follower's problems, it is possible to reformulate a bilevel optimization model into a single level one using optimality conditions of the follower's problem (Bard, 2013).

Uncooperative response is known to be less tractable compared to the cooperative response even in the case of a convex follower's problem. In fact, pessimistic bilevel optimization problems, so far, do not have any computational methods that are generally applicable (Zeng, 2020). Commonly, solving these problems require processing all alternative optimal solutions of the follower's problem to determine the solution that is the worst

for the leader's objective function, which is deemed not tractable (Sinha et al., 2017).

We study both the cooperative and uncooperative response in this study by developing the single-level reformulation of BLP under the cooperative response case as well as providing a solution algorithm to solve BLP for the uncooperative response case and M/M/c queues.

3.3.1 Single-Level Reduction of BLP under Cooperative Response

In this subsection, we show that the follower's problem always has an integer optimal solution when feasible, and then, reformulate BLP into a MILP taking advantage of the integrality property of the follower's problem under cooperative response.

Proposition 3.1. *Follower's problem of BLP always has an integer optimal solution when feasible.*

Proof: Assume that there exists an OD pair $k \in K$ with fractional x_{ijk} in an optimal solution. Without loss of generality, let p_1 and p_2 be two, not necessarily disjoint, paths with distances z_1 and z_2, respectively. Let χ_1 and χ_2 denote the fractional flow on these paths, such that $\chi_1 + \chi_2 = 1$ by constraint (3.8). By the optimality of this solution, $z_1 = z_2$. Otherwise, one may construct a solution that only uses path $p^* = \text{argmin}\{z_1, z_2\}$ with objective function value $\min\{z_1, z_2\} \leq z_1\chi_1 + z_2\chi_2$. Since $z_1 = z_2$, an integral solution may be obtained by selecting one of the equal length paths arbitrarily and setting the corresponding x_{ijk} values to 1 on this path, and to 0 on the rest. □

Proposition 3.1 means that the domain constraint (3.11) of the follower's problem may be dropped and its optimal solution is characterized by the optimality conditions. Let θ_{jk}, γ_{jk}, and ϕ_k be the dual variables associated with (3.8), (3.9), and (3.10), respectively. The single-level reformulation of BLP, denoted by SLP, is:

[SLP] minimize (3.1)

s.t. (3.2) − (3.6), (3.8) − (3.11)

$$\sum_{(i,j)\in A_k} \delta_{ij} x_{ijk} = (\theta_{O_k k} - \theta_{D_k k}) + \sum_{j\in N_k} y_j \gamma_{jk} + \delta_{O_k D_k}(1+\tau)\phi_k \qquad k \in K \tag{3.12}$$

$$\theta_{ik} - \theta_{D_k k} + \delta_{iD_k}\phi_k \le \delta_{iD_k} \qquad k \in K,\ (i, D_k) \in A_k \tag{3.13}$$

$$\theta_{ik} - \theta_{jk} + \gamma_{jk} + \delta_{ij}\phi_k \le \delta_{ij} \qquad k \in K,\ (i,j) \in A_k,\ j \neq D_k \tag{3.14}$$

$$\theta_{jk} \text{ free} \qquad j \in N_k,\ k \in K \tag{3.15}$$

$$\gamma_{jk} \le 0 \qquad j \in N_k,\ k \in K \tag{3.16}$$

$$\phi_k \le 0 \qquad k \in K \tag{3.17}$$

In this reformulation, (3.12) is the strong duality condition of the follower's problem and its dual feasibility requrirements are defined by constraints (3.13) − (3.17). Note that (3.12) is written for $k \in K$ since the dual problem is decomposable for each OD pair. This constraint includes a quadratic term that can be linearized by introducing auxiliary decision variables $q_{jk} = y_j \gamma_{jk}$, $j \in N_k$, $k \in K$. Accordingly, constraint (3.12) is linearized by replacing it with (3.18)-(3.21):

$$\sum_{(i,j)\in A_k} \delta_{ij} x_{ijk} = (\theta_{O_k k} - \theta_{D_k k}) + \sum_{j\in N_k} q_{jk} + \delta_{O_k D_k}(1+\tau)\phi_k \qquad k \in K \tag{3.18}$$

$$q_{jk} \ge \overline{\gamma}_{jk} y_j \qquad j \in N_k, k \in K \tag{3.19}$$

$$q_{jk} \ge \gamma_{jk} \qquad j \in N_k, k \in K \tag{3.20}$$

$$q_{jk} \le \gamma_{jk} - (1 - y_j)\overline{\gamma}_{jk} \qquad j \in N_k, k \in K \tag{3.21}$$

$$q_{jk} \le 0 \qquad j \in N_k, k \in K \tag{3.22}$$

where $\overline{\gamma}_{jk}$ denotes the lower bound for the non-positive dual variable γ_{jk}. Constraint (3.22) defines the domain of $\overline{\gamma}_{jk}$ as non-positive and (3.19) ensures that $\gamma_{jk} = 0$, when $y_j = 0$. On the other hand, constraints (3.20) and (3.21) ensure that $q_{jk} = \gamma_{jk}$, when $y_j = 1$.

Both BLP and SLP include variable W_j, which is a function of x_{ijk} and z_j as well as demand and service rates. The characterization of W_j and the associated probabilistic service level constraints (3.3) depend on the underlying queuing system at the charging stations. Consequently, the solvability of SLP still depends on the ability to characterize chance constraints (3.3). Next, we discuss the case of Poisson arrivals and exponential service times, which allows the reformulation of (3.3) using mixed-integer optimization.

3.3.2 Case with Poisson Arrivals and Exponential Service Times: M/M/c Queuing System

DC fast charging stations for EVs typically include multiple charging stalls, that can deliver 80-400kW at 50-1000V (IEA, 2021). Each charging station provides simultaneous service to multiple EVs where each stall acts as a server utilized by service-seekers on a first-come-first-serve basis. Moreover, common courtesy dictates that there is a single queue for each charging station in case of concurrent utilization of the servers. This single queue – multiple parallel server structure constitutes as the main framework of the queuing system models of DC fast charging stations.

Existing public charging stations are often located near discretionary service facilities (e.g., restaurants, stores, plazas, or shopping centres) in order to mitigate the relatively longer recharging times for EVs. As a result, not every customer unplugs their EV right after getting up to their desired state of charge (SoC). This is one of the main reasons for existing studies to consider stochastic, and typically exponential, service times (Xie et al., 2018; Xie and Lin, 2021). The stochastic flow rate of an OD is assumed to have Poisson distribution. The rate of the Poisson process may be determined as a function of the population of two cities, their respective EV adoption rates and the travel distance between them. For long-distance trucking, for example, it may be a function of the number

of ETs dispatched per day from an origin to a destination.

Under a single queue, multiple parallel server, Poisson arrivals and exponential service time system, each charging station may be modeled as an M/M/c queue. However, a long distance OD route might require multiple charging stops and ODs may have varying flow rates resulting in a queuing network.

Burke's theorem states that the departure rate of a stationary M/M/1 queue is identical to its arrival rate as long as the queuing system is stable (Burke, 1956). This concept of Poisson flow conservation is also generalizable to M/M/c queues, which allows the characterization of a queuing network that comprises of several interconnected M/M/c queues (Bose, 2013). Using these results, an arbitrary transportation network with M/M/c charging stations, is an open multi-server Jackson network for which the routing probabilities are set by the route choices of EV users.

We define λ_k to denote the Poisson flow rate of OD pair $k \in K$. Let Λ_j denote the flow rate faced by open station $j \in N$, determined as the total Poisson OD demand.

$$\Lambda_j = \sum_{i:(i,j)\in A} \sum_{k\in K} \lambda_k x_{ijk} \qquad j \in N \tag{3.23}$$

Recall that the expanded network ensures that $x_{ijk} = 1$ of OD $k \in K$ charges at $j \in N$, Moreover, the underlying open multi-server Jackson network ensures that the arrival rate of an OD to a station is equal to their departure rate as long as all queues on the network are stable. The queue stability requirement is defined as:

$$\Lambda_j \leq \mu z_j - \epsilon \qquad j \in N \tag{3.24}$$

where μ is the exponential service rate of each server of a station and ϵ is an infinitesimal constant. This constraint ensures that the Poisson flow rate faced by a station is strictly

less than its processing rate in order to have a finite queue length.

We define the average load of a stall at a charging station by $\rho_j = {}^{\Lambda_j}/_{\mu z_j}$, and rewrite the left-hand-side of constraint (3.3) using a modification of the Erlang-C formula (Chromy et al., 2012; Xie et al., 2018).

$$\mathbb{P}\big(W_j \leq \alpha\big) = 1 - \frac{\frac{(z_j\rho_j)^{z_j}}{z_j!(1-\rho_j)}}{\frac{(z_j\rho_j)^{z_j}}{z_j!(1-\rho_j)} + \sum_{t=1}^{z_j-1} \frac{z_j\rho_j{}^t}{t!}} e^{-\mu(z_j-\rho_j z_j)\alpha} \qquad j \in N \tag{3.25}$$

Although this substitution in (3.3) involves a quadratic function, it may be rewritten using a step function on integer z_j:

$$z_j = \begin{cases} 0, & \Lambda_j = 0 \\ 1, & \epsilon \leq \Lambda_j < b_1 \\ 2, & b_1 \leq \Lambda_j < b_2 \\ \dots & \dots \\ C, & b_{C-1} \leq \Lambda_j < b_C. \end{cases} \qquad \text{if } \mathbb{P}\big(W_j \leq \alpha\big) \geq \beta \tag{3.26}$$

Service level parameters α and β of constraint (3.3) determine the values of the breakpoints $(b_1, b_2, \dots, b_C)$ of this step function by a preprocessing procedure based on (3.25). Such a function can be easily incorporated within any MILP model using special ordered set of type 1 (SOS1) variables or using the following set of additional constraints to substitute (3.3).

$$\sum_{m=1}^{C} s_{jm} = 1 \qquad j \in N \tag{3.27}$$

$$\Lambda_j \leq \sum_{m=1}^{C} b_m s_{jm} \qquad j \in N \tag{3.28}$$

$$\Lambda_j \geq \sum_{m=2}^{C} b_{m-1} s_{jm} \qquad j \in N \tag{3.29}$$

$$z_j \geq \sum_{m=1}^{C} m s_{jm} \qquad j \in N \tag{3.30}$$

$$s_{jm} \in \{0,1\} \qquad j \in N,\ m \in \{1,2,...,C\} \tag{3.31}$$

Constraints (3.28)-(3.29) determine the two consecutive breakpoints of the step function that Λ_j value lies between. Constraints (3.30) exploit the fact that both m and z_j are integer and determine the minimum z_j value in conjunction with the minimization objective function (3.1) of BLP. Implementing these constraints rather than using SOS1 variables is observed to be more computationally effective during preliminary computational analysis.

The number of s_{jm} variables increases with C, so choosing the right value of C is important. In this study, we determine this value so that it is large enough to keep constraint (3.2) non-binding for the optimal solution but small enough so that the number of binary s_{jm} variables is not excessively many.

Substituting a step function instead of the chance constraint (3.3) by pregenerating its breakpoints based on the Erlang-C formula and linearizing this step function with the above constraints, the single level reduction of the bilevel model is updated as below:

[DLP] minimize (3.1)

s.t. (3.2), (3.4) − (3.6), (3.8) − (3.11), (3.13) − (3.24), (3.27) − (3.31)

To summarize, when OD flows are Poisson distributed and service rate at a stall is exponential, SLP is reformulated as a MILP which is denoted as DLP. Even though DLP may be solved using off-the-shelf solvers, the preliminary computational experiments revealed that these solvers are not able to handle the problem complexity for large-sized networks. In order to solve DLP for real-life instances, we develop a decomposition based algorithm that exploits the problem structure as described in the next section.

3.4 Solution Methodology

There are three decisions to make in this problem setting: the locations of the charging stations, the number of stalls at open stations, and the OD routes. We propose a solution method that decomposes these decisions and iteratively solves two single-level problems: a location problem and a sizing problem. As the location problem turns out to be computationally demanding, we develop a tailored logic-based Benders algorithm that solves all instances within seconds.

The proposed decomposition structure first solves a location problem taking into account OD flows to obtain the charging station location vector $\overline{\mathbf{y}}$, and the corresponding optimal reaction set of the follower defined on $\overline{\mathbf{y}}$:

$$\Psi(\overline{\mathbf{y}}) = \left\{ \overline{\mathbf{x}} : \overline{\mathbf{x}} \in \operatorname{argmin} \left\{ \sum_{k \in K} \sum_{(i,j) \in A_k} \delta_{ij} x_{ijk} : (3.8), (3.10), (3.11), (3.32) \right\} \right\}.$$

$$\text{where,} \quad \sum_{i:(i,j) \in A_k} x_{ijk} \leq \overline{y}_j \qquad j \in N_k, \quad j \neq D_k, \quad k \in K \tag{3.32}$$

For given $\overline{\mathbf{y}}$ and $\Psi(\overline{\mathbf{y}})$, the sizing problem reduces to determining the number of stalls at open stations ($\overline{\mathbf{z}}$) and allocating charging demand of OD flows to open stations ($\overline{\mathbf{x}}$). The sizing problem accounts for route choice behavior by considering all alternative optimal paths on the network defined by $\overline{\mathbf{y}}$ and by modelling both cooperative and uncooperative response cases. The sizing problem determine a partial solution $(\overline{\mathbf{z}}, \overline{\mathbf{x}})$ that is optimal for given $\overline{\mathbf{y}}$, and a complete solution $(\overline{\mathbf{y}}, \overline{\mathbf{z}}, \overline{\mathbf{x}})$ that is bilevel feasible. The flowchart of the solution methodology is shown in Figure 3.1. At each iteration, an optimality-type cut is generated after solving the sizing problem and is fed back into the location problem.

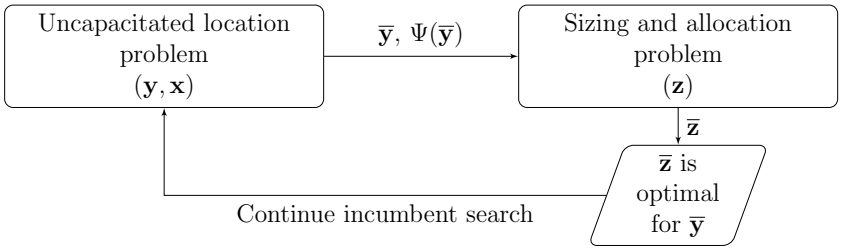


Figure 3.1: The flowchart of the decomposition based solution method for BLP.

The rest of the section is organized as follows. Section 3.4.1 describes the location problem and the suggested logic-based Benders algorithm. Section 3.4.2 models the sizing problem under M/M/c and provides the complete decomposition based algorithm for the cooperative response as well as an adaptation of the algorithm for the uncooperative response case in Section 3.4.2.

3.4.1 Solution of the Location Problem

When the sizing decisions and service level constraint (3.3) are dropped, the follower's choice of route will not affect the leader's objective. Consequently, the remaining problem reduces to a single level location problem and determines the minimum cost charging station locations subject to flow balance and deviation tolerance constraints. The location problem P_L is:

$$\begin{aligned} [P_L] \text{ minimize} \quad & \sum_{j \in N} f(y_j) \\ \text{s.t.} \quad & (3.4), (3.8) - (3.11) \end{aligned} \tag{3.33}$$

P_L may be solved directly by a solver or using Benders decomposition, where the master problem determines charging station locations and the subproblems check path feasibility for each OD. Recall that the subproblems are linear due to Proposition 3.1.

Despite being a relaxation of BLP, preliminary testing showed that P_L is still challenging to solve in reasonable times both using the solver and Benders decomposition. Since P_L is solved repeatedly within the solution method of BLP, we develop a logic-based Benders algorithm that is able to solve P_L optimally within a few seconds for the majority of the large-scale instances. The logic-based Benders master problem is given by:

$$\begin{aligned} &\text{[MP] minimize} && \sum_{j \in N} f(y_j) && && (3.34)\\ &\text{s.t.} && y_j \in \{0,1\} && j \in N && (3.4)\\ & && + \text{logic-based Benders cuts} \end{aligned}$$

and the subproblem is a path feasibility problem with deviation constraints on the network defined by $\overline{\mathbf{y}}$, the solution of MP.

We further modify the subproblems in two ways. First, we drop the deviation tolerance constraints and deal with them using feasibility cuts. Second, we keep the objective function of the follower and solve a minimum cost path problem instead of a feasibility problem. It is redundant to say that this does not impact the correctness of the logic-based Benders algorithm even though it may lead to an alternative optimal solution for P_L. However, the objective function of the follower becomes critical within the solution methodology of BLP where it is necessary to account for their minimum length path choice. The subproblem SP_k for OD $k \in K$ is given by:

$$[SP_k] \text{ minimize} \quad \sum_{(i,j) \in A_k} \delta_{ij} x_{ijk} \tag{3.35}$$

$$\text{s.t.} \sum_{j:(i,j) \in A_k} x_{ijk} - \sum_{j:(j,i) \in A_k} x_{jik} = \begin{cases} 1, & \text{if } i = O_k, \\ -1, & \text{if } i = D_k, \\ 0, & otherwise. \end{cases} \quad i \in N_k \tag{3.36}$$

$$\sum_{i:(i,j)\in A_k} x_{ijk} \leq \overline{y}_j \qquad j \in N_k, \quad j \neq D_k \tag{3.37}$$

$$x_{ijk} \in \{0,1\} \qquad (i,j) \in A_k \tag{3.38}$$

When a solution $\overline{\mathbf{y}}$ of MP is infeasible to SP_k, the aim is to derive valid cut(s) that *(i)* remove $\overline{\mathbf{y}}$ from the feasible region of MP, and *(ii)* do not eliminate any feasible solution of P_L. The valid cuts are added to MP and the latter is solved repeatedly to find a new set of station locations. When the valid cut(s) satisfy the two conditions above, this approach is guaranteed to converge to an optimal solution (Hooker and Ottosson, 2003). We next discuss the derivation of the valid cuts and prove the correctness of the algorithm.

Once a solution $\overline{\mathbf{y}}$ is obtained, the node restrictions enforced by constraint (3.37) are used to form a reduced graph $G_k^r = (N_k^r, A_k^r)$ for each OD pair $k \in K$ such that $N_k^r = \left\{j \in N_k \mid \overline{y}_j = 1\right\} \cup \{O_k, D_k\}$ and $A_k^r = \{(i,j) \in A_k \mid i,j \in N_k^r\}$. Subproblems SP_k are solved on their respective reduced graphs G_k^r using a labeling algorithm. A subproblem maybe feasible in which case the optimal length may not satisfy the deviation tolerance constraints. Alternatively, a subproblem may be infeasible if there is no path that connects OD in the reduced graph. In both cases, feasibility cuts are generated and added to MP. We first discuss the case where SP_k is infeasible for some OD $\hat{k} \in K$. In this case, the corresponding graph $G_{\hat{k}}^r$ has at least two connected components that separately include $O_{\hat{k}}$ and $D_{\hat{k}}$. Let $G_{\hat{k}}^o$ be the connected component of $G_{\hat{k}}^r$ that includes $O_{\hat{k}}$, and $G_{\hat{k}}^d$ be the reverse graph of the component of $G_{\hat{k}}^r$ that includes $D_{\hat{k}}$. Let A_k^{rev} denote the reverse A_k. We define $\mathcal{N}_f$ as the set of nodes that connect $G_{\hat{k}}^o$ to $G_{\hat{k}}^d$, and $\mathcal{N}_b$ as the set of nodes that connect $G_{\hat{k}}^d$ to $G_{\hat{k}}^o$:

$$\mathcal{N}_f = \left\{j \in N_{\hat{k}} \backslash N_{\hat{k}}^r : \ (i,j) \in A_{\hat{k}}, i \in N_{\hat{k}}^o\right\} \tag{3.39}$$

$$\mathcal{N}_b = \left\{j \in N_{\hat{k}} \backslash N_{\hat{k}}^r : \ (i,j) \in A_{\hat{k}}^{rev}, i \in N_{\hat{k}}^d\right\} \tag{3.40}$$

Using $\mathcal{N}_f$ and $\mathcal{N}_b$, we introduce the following feasibility cuts for an infeasible $SP_{\hat{k}}$:

$$\sum_{i \in \mathcal{N}_f} y_i \geq 1 \tag{3.41}$$

$$\sum_{i \in \mathcal{N}_b} y_i \geq 1 \tag{3.42}$$

Cuts (3.41) and (3.42) are added to MP to cut the current infeasible solution. Note that adding either of them is sufficient to cut the current infeasible solution.

For the case when SP_k is feasible and deviation tolerance is exceeded for some OD $\hat{k} \in K$, i.e., $\sum_{(i,j) \in A_{\hat{k}}} \delta_{ij} x_{ij\hat{k}} > (1+\tau)\delta_{O_{\hat{k}}D_{\hat{k}}}$, we introduce the following cut:

$$\sum_{i \in N_{\hat{k}} \setminus N^r_{\hat{k}}} y_i \geq 1 \tag{3.43}$$

Cut (3.43) is added to MP to eliminate the current solution that is infeasible with respect to deviation tolerance. This cut ensures that at least one charging station from $N_{\hat{k}} \setminus N^r_{\hat{k}}$ is open since opening facilities at $i \in N^r_{\hat{k}}$ is infeasible.

The logic-based Benders decomposition algorithm solves the relaxed master problem MP and subproblem SP_k iteratively, and adds feasibility cuts (3.41), (3.42), and (3.43) to MP as needed. We present the pseudo-code of the logic-based Benders algorithm (LBBA) in Algorithm 3.1 and prove its correctness in Proposition 2.

Note that P_L is feasible as long as there exist $\mathbf{y}$ that is feasible for the follower's problem.

Proposition 3.2. *LBBA determines an optimal solution to P_L given that it is feasible.*

Proof: (i) It is trivial that cuts (3.41), (3.42), and (3.43) remove a solution $\overline{\mathbf{y}}$ infeasible with respect to (3.8) and (3.43) from the feasible region of the master problem MP.

(ii) To show that cuts (3.41) and (3.42) do not remove any feasible solution to P_L,

Algorithm 3.1 The exact LBBA for the location problem P_L

Initialization → SP_k is infeasible
while SP_k is infeasible **do**
  SOLVE MP to obtain $\overline{\mathbf{y}}$
  for $k \in K$ **do**
    SOLVE SP_k by a labeling algorithm on G_k^r to obtain $\overline{\mathbf{x}}$
    if SP_k is infeasible due to constraint (3.36) **then**
      SET $\mathcal{N}_f$ and $\mathcal{N}_b$
      ADD (3.41) and (3.42) to MP
    else if SP_k is feasible but constraint (3.10) is violated **then**
      ADD (3.43) to MP
    end if
  end for
  if all SP_k are feasible and (3.10) is satisfied **then**
    RETURN optimal solution as $(\overline{\mathbf{y}}, \overline{\mathbf{x}})$
    BREAK while
  end if
end while

assume to the contrary that there exists a feasible solution $\hat{\mathbf{y}}$ that does not satisfy (3.41) or (3.42) for an OD pair $\hat{k} \in K$. Since this solution is feasible, there exists at least one open charging station at node $\hat{i}$ that is connected to $O_{\hat{k}}$. Therefore, $\hat{i} \in \mathcal{N}_f$ and $\hat{\mathbf{y}}$ satisfies (3.41). This is a contradiction since $\hat{\mathbf{y}}$ satisfies cut (3.41) as this specific node is already in the corresponding $\mathcal{N}_f$. The same argument follows for $D_{\hat{k}}$, $\mathcal{N}_b$, and (3.42). Hence, valid cuts (3.41) and (3.42) do not eliminate any feasible solutions.

To show that cuts (3.43) do not remove any feasible solution to P_L, assume to the contrary, there exists a feasible solution which does not satisfy this cut. This means that none of the facilities in $N_k \backslash N_k^r$ are open. This creates a contradiction, since opening any subset of facilities in N_k^r cannot improve path lengths and the solution remains infeasible. Hence, cuts (3.43) do not eliminate any feasible solutions. □

The solution of the location problem opens charging stations and determines OD flow paths within the deviation tolerance. It remains to determine the number of charging stalls at each open location to satisfy the service requirement, which we refer to as the sizing problem. These two problems are solved iteratively to generate solutions for DLP. Next, we present the sizing problem and the complete solution for BLP.

3.4.2 The Sizing Problem and DLP Solution

When the charging station locations are fixed, one can always equip each station with sufficient number of stalls to satisfy the service level constraints (3.3), provided stall capacity C is sufficiently large. To determine a solution that minimizes the total cost of capacity installation, one has to account for the amount of flow through each station, which in turn depends on the shortest OD paths and the stopping behavior of EV users when there are alternative stopping options on their shortest path(s). Consider the example in Figure 3.2.

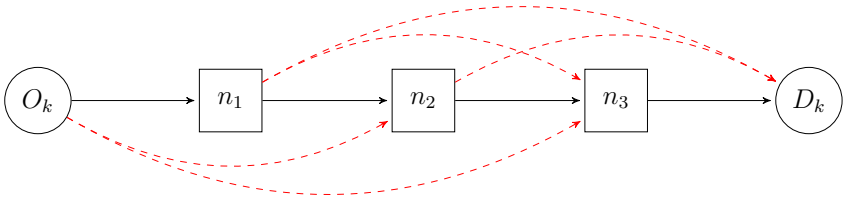


Figure 3.2: An example OD path with alternative stopping options.

This path has three open charging stations, represented with square nodes. Artificial expanded network arcs, represented with dashed red lines, indicate that recharging at any of the three open stations results in a feasible trip. There are seven alternative shortest paths for this OD, each corresponding to different feasible stopping combinations. Each shortest path may lead to different total flows through the stations and, consequently, different requirements in the number of stalls. Hence, to model the sizing problem correctly, we have to consider all possible stopping combinations on the network defined by $\overline{\mathbf{y}}$.

The set of alternative solutions for OD pair $k \in K$ determine its optimal reaction set $\Psi_k(\overline{\mathbf{y}})$, where $\Psi_k(\overline{\mathbf{y}})$ is populated using a shortest path algorithm on G_k^r. We next discuss how to formulate the sizing problem and present the decomposition based algorithm under cooperative and uncooperative responses of the follower.

The Cooperative Response:

Under the cooperative response of the follower, the sizing problem determines the number of stalls in open charging stations, and allocates charging demand to the stations so that the service level requirement is satisfied while minimizing total stall installation cost. Let a_{jrk} be a binary parameter that takes the value of 1 if an open station $j \in N_k$ is used in path $r \in \Psi_k(\overline{\mathbf{y}})$ of OD pair $k \in K$; and u_{rk} be a binary decision variable that takes the value of 1 if OD pair $k \in K$ is assigned to path $r \in \Psi_k(\overline{\mathbf{y}})$. The sizing problem P_S under the cooperative response and M/M/c case is:

$$[P_S] \text{ minimize} \quad \sum_{j \in N_k \,|\, \overline{y}_j = 1} g(z_j) \tag{3.44}$$

$$\text{s.t.} \quad \sum_{r \in \Psi_k(\overline{\mathbf{y}})} u_{rk} = 1 \qquad k \in K \tag{3.45}$$

$$\Lambda_j = \sum_{k \in K} \sum_{i:(i,j) \in A_k} \sum_{r \in \Psi_k(\overline{\mathbf{y}})} \lambda_k a_{jrk} u_{rk} \qquad j \in N_k \,|\, \overline{y}_j = 1 \tag{3.46}$$

$$(3.5), (3.27) - (3.31)$$

$$u_{rk} \in \{0, 1\} \qquad r \in \Psi_k(\overline{\mathbf{y}}),\ k \in K \tag{3.47}$$

The objective function (3.44) minimizes the total cost of installing charging stalls at open stations. Constraint (3.45) ensures that each OD pair is assigned to exactly one of the alternative paths in its optimal reaction set. Constraint (3.46) calculates the total allocated demand to a charging station. Constraints (3.27) - (3.31) ensure that the required capacity is calculated based on the linearization of the Erlang-C function as discussed in Section 3.3.2. Finally, constraints (3.5) and (3.47) define the domains of the capacity variables z_j and the allocation variables u_{rk}, respectively. Although not decomposable by $k \in K$, preliminary computational experiments showed that P_S may be solved within seconds using the commercial solver Gurobi for sufficiently large instances.

Upon solving P_S, we update the solution $\overline{\mathbf{x}}$ to reflect the selected paths determined by optimal u_{rk}. The partial solution $(\overline{\mathbf{x}}, \overline{\mathbf{z}})$ is optimal for a given $\overline{\mathbf{y}}$, and solution $(\overline{\mathbf{y}}, \overline{\mathbf{x}}, \overline{\mathbf{z}})$ is feasible to DLP. The algorithm to solve DLP proceeds by generating cuts (3.48) to cut the current solution and generate new ones.

$$\sum_{j:\left\{j\in N \mid \overline{y}_j \neq 1\right\}} y_j \geq 1 \tag{3.48}$$

When appended to MP, this optimality-type cut ensures that the current set of open facilities is never generated again. The complete algorithm for DLP is depicted in Algorithm 3.2. Step 3 utilizes LBBA to determine a minimal cost, feasible set of charging station locations. Steps $4-6$ are the solution procedure for P_S and $7-10$ are for incumbent tracking. Finally, Step 11 adds cuts (3.48) to MP to eliminate the current solution from the feasible region of P_L. In the computational experiments, an iteration of the algorithm refers to Steps $3-11$ and the algorithm terminates using time or iteration limits. We refer to this algorithm as decomposition based algorithm (DA).

Algorithm 3.2 Decomposition based algorithm (DA) for DLP

1: $Incumbent \leftarrow +\infty$
2: **while** time or iteration limit is not reached **do**
3: **OBTAIN** $\overline{\mathbf{y}}$ using LBBA
4: **SET** $\Psi_k(\overline{\mathbf{y}})$ for each $k \in K$ by a k-shortest path algorithm on G_k^r
5: **SOLVE** P_S to obtain $\overline{\mathbf{u}}$ and $\overline{\mathbf{z}}$
6: **SET** $\overline{\mathbf{x}}$ based on $\overline{\mathbf{u}}$
7: $tempObjective \leftarrow$ (3.1) for $\overline{\mathbf{y}}$ and $\overline{\mathbf{z}}$
8: **if** $tempObjective < Incumbent$ **then**
9: $Incumbent \leftarrow tempObjective$, $\mathbf{y}^* := \overline{\mathbf{y}}$, $\mathbf{x}^* := \overline{\mathbf{x}}$ and $\mathbf{z}^* := \overline{\mathbf{z}}$
10: **end if**
11: **ADD** (3.48) to MP
12: **end while**
13: **RETURN** $Incumbent$ and $(\mathbf{y}^*, \mathbf{x}^*, \mathbf{z}^*)$

The Uncooperative Response.:

The cooperative response assumes that EV drivers will use a shortest path that minimizes the leader's objective. In practice, however, it is difficult to enforce such a choice.

Consequently, the leader may be interested in investigating the uncooperative response or the worst possible outcome of the follower's choices. To account for the uncooperative response, the formulation of the sizing problem P_S is modified as follows. First, the objective function is converted to a maximization to reflect the outcome of the uncooperative EV user behavior defined by $\mathbf{x^p}$ in Section 3.3. Next, in order the prevent unboundedness, the sign of constraint (3.30) is changed to equality. The rest of P_S remains the same and algorithm DA can be implemented accordingly.

Upon preliminary computational experiments, it is observed that this definition of $\mathbf{x^p}$ generally leads to solutions where the worst case path is one that uses all charging stations regardless of their proximity and the need to charge. This is unlikely to happen in reality since EV users would generally prefer fewer stops. As a result, an uncooperative response solution should exclude unjustifiable charging stops. This behavior may be reflected in the BLP formulation under uncooperative response by changing $\Psi_k(\overline{\mathbf{y}})$. For each OD pair $k \in K$, we modify the selection of OD paths in the optimal reaction set $\Psi_k(\overline{\mathbf{y}})$ to minimal paths where such a path is one that becomes infeasible if one charging station is removed. We use this definition of the uncooperative response in the computational testing.

3.5 Computational Experiments

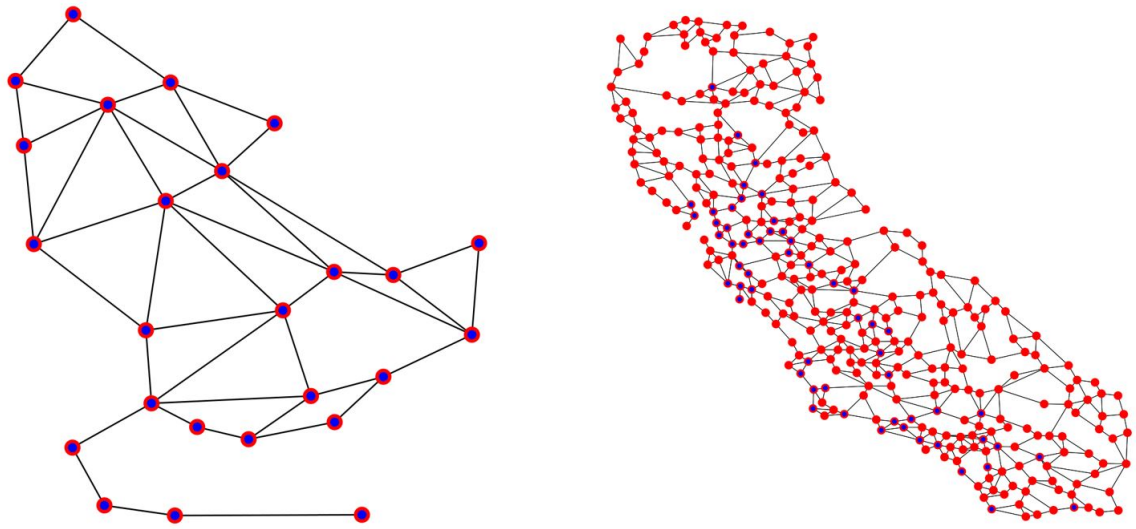
In this section, we present extensive computational analysis to assess the performance of the proposed solution algorithm as well as to derive insights from effects of the follower's position, deviation tolerance, and service level parameters on optimal solutions. We compare the quality of the solutions obtained by DA with optimal solutions found by solving DLP using a commercial solver on smaller-size instances. We also derive several managerial insights from optimal solutions. Additionally, we carry out experiments on two real world highway networks to assess the proposed solution methodology by evaluating its compu-

tational performance on larger instances. Experiments are performed using Gurobi v9.1.1 with Python API and *NetworkX* package v2.5.1 (Hagberg et al., 2008) on a computer with Intel i7-9700K 3.60GHz processor and 32GB of RAM.

3.5.1 Data Sets and Instance Generation Specifics

We use three test networks of different sizes, one of which is a network that we introduce to the literature based on the highway network of the United States. The other two are from the literature; namely the Simchi-Levi and Berman (1988)'s 25-node benchmark network and the California road network, which are already introduced earlier in Chapter 2. These networks will be continued to be as *N25* and *CA339*.

The *N25* and *CA339* networks are depicted in Figure 3.3. The blue nodes on these networks highlight the potential OD nodes, which correspond to population centers.



(a) *N25* Network (b) *CA339* Network

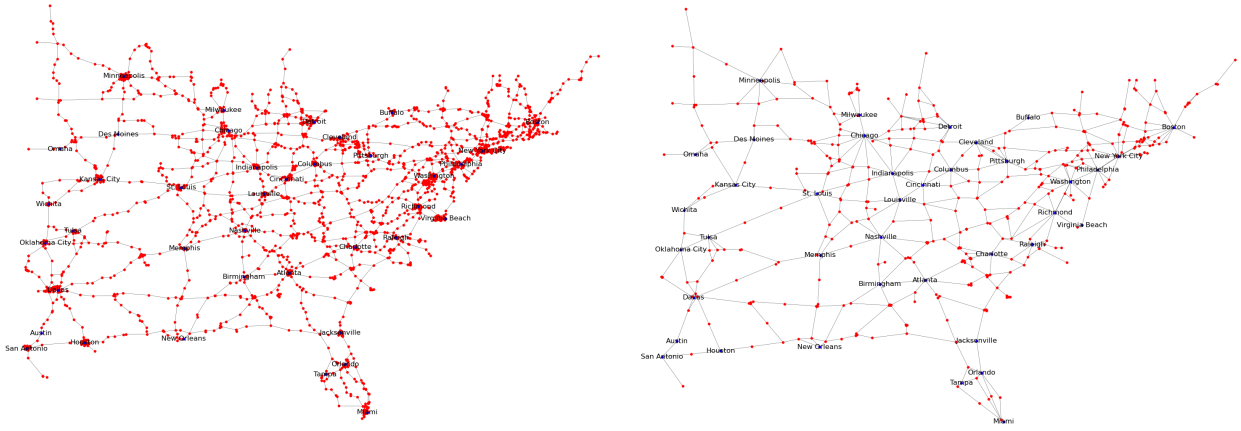
Figure 3.3: Small benchmark network *N25* and California road network *CA339*.

We introduce a new transportation network to the literature, referred to as *US-E*, based on the highway network of the United States. This new dataset is formed based on the GIS

shapefiles provided by the US Census Bureau - Department of Commerce, which include all primary roads that cover the continental United States. *US-E* contains the Eastern half of this output and has 2,575 nodes and 6,736 arcs. We simplify the raw GIS data by removing some node clusters that represent highway intersections, and many of the mid-arc nodes that are used to represent the geographical direction changes on highway sections. After these reductions, the resulting network, referred to as *US-E420*, contains 420 nodes and 1276 arcs. The visual representations of *US-E* and *US-E420* are shown in Figure 3.4. The specifications of the three networks are summarized in Table 3.1.

Table 3.1: Specifications of the three networks used in computational experiments.

	N25	*CA339*	*US-E420*
Number of nodes	25	339	420
Number of potential OD nodes	25	51	39
Number of arcs	86	1234	1276
Mean arc length (km)	4.60	18.96	91.35
Standard deviation of arc lengths	1.81	12.96	67.25



(a) Raw *US-E* Network

(b) *US-E420* Network

Figure 3.4: Eastern United States highway networks: raw *US-E* and simplified *US-E420*.

For this computational analysis, we assume that every EV user starts their journey with 80% of the available battery range, and they may charge up to 80% SoC level at a charging station visit. The reason for this assumption is threefold. First, it takes the range anxiety and energy consumption variability into account by leaving a 20% reserve that concerned

individuals may utilize. Second, battery charging process at public charging stations gets exponentially slower after 80% SoC, which may cause additional congestion. In practice, such a limit is currently enforced at busy Tesla Superchargers in the United States (Electrek, 2019). Third, this assumption regards EV user battery degradation concerns by continuously maintaining a SoC above 80%, which is shown to significantly accelerate degradation (Pelletier et al., 2017).

We test three R values for *N25* (15, 20, and 25), two for *CA339* (200 and 250 km), and two for the larger scale *US-E420* (400 and 600 km) to represent the state of the art EV and electric truck ranges. We further filter all possible long-distance trips between any potential OD nodes for which the EVs need to stop for recharging at least once based on the selection of R. Each R setting forms a new problem instance as the number of arcs on the expanded network changes with this parameter. Typically, when R increases, the size of the expanded network also increases.

For the deviation tolerance τ, we test 0%, 10%, 25%, and 50%. Moreover, we test 1, 5, 10 and 30 minutes for service level waiting threshold α, and 95%, 90%, and 80% for service level probability β. As we do not have access to a real world OD flow data for the transportation networks used in this study, we assume that the Poisson OD flow rate λ_k is randomly distributed between $(0, 2)$ vehicles per hour for every trip. Considering that the potential OD nodes are typically higher degree nodes of the underlying graphs and that we take all possible long-distance OD pairs into account, the aggregate EV flow rate out of an origin node is observed to be realistic using this scale of individual random rates. We take the exponential service time μ as 30 minutes (Xie et al., 2018; Xie and Lin, 2021). Moreover, throughout the computational analysis, we set the fixed cost of building a charging station as \$150,000 and the cost of deploying one DC charging stall as \$75,000 (Nicholas, 2019).

3.5.2 Computational Performance of the Decomposition Based Algorithm

In this section, we first compare the optimal solutions under cooperative response found by the solver with the solutions of the decomposition based algorithm (DA) on the benchmark *N25* network. Then, we test the computational performance of this algorithm on the large-scale *CA339* and *US-E420* networks.

For *N25*, we generate 48 instances for each R value we test, which adds up to a total of 144 test instances. For each instance, the number of all possible long-distance OD pairs, i.e., the cardinality of set K, is 167, 108, and 57 for $R = 15, 20$, and 25, respectively. Each instance is solved both with the optimization model DLP and DA, and their solutions are compared to assess the performance of the latter. The stopping condition of the algorithm is set to 60 iterations. The algorithm is able to find the optimal solutions for all 144 instances. The results of these experiments are detailed in Table B.2 of the Appendix. We further visualize the average solution time comparisons in Figure 3.5.

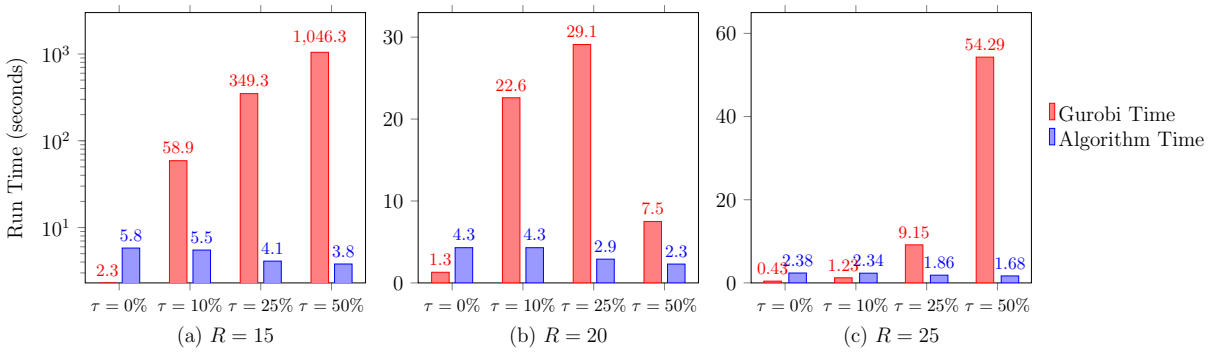


Figure 3.5: Average solution time comparison for *N25* instances.

The solver finds the optimal solutions quicker only when the deviation tolerance is zero as the size of the expanded network is smaller due to the introduction of the set of permissable arcs. Instances with higher R values are also relatively quicker to solve by the solver as they include fewer long-distance OD pairs. On the other hand, time savings by

the algorithm are drastically higher for $R = 15$ instances with deviation tolerance greater than 0%. Particularly, the average solution time obtained with Gurobi for 50% deviation tolerance instances of $R = 15$ is over 1,000 seconds, whereas the algorithm is capable of terminating with an optimal solution in less than 4 seconds.

Now, we focus on the large-sized *CA339* and *US-420E* instances to further evaluate the computational performance of the algorithm. We generate 32 instances on the *CA339* network at two different R values, 200 and 250 km, with a fixed β of 90%. The $R = 200$ and $R = 250$ instances include 502 and 335 long-distance OD pairs, respectively. The same 32 configurations as with *CA339* are used to create the *US-E420* instances, with higher values of R (400 and 600 km). For these instances, we include a trip distance threshold of 1,600 km as those longer than this threshold would require overnight stays. This results in a total of 469 and 429 long-distance OD pairs defined on the *US-E420* network for $R = 400$ and $R = 600$ instances, respectively.

We first attempted to solve each instance using Gurobi and were unable to obtain even an incumbent solution for deviation tolerance levels greater than 0% within a 2-hour time limit. When this time limit is increased, Gurobi terminates the solution procedure with an out of memory error. This indicates that the state-of-the-art solvers are not able to handle problem instances on real-world networks. Consequently, we utilize DA to solve these instances with a stopping condition of 1-hour. The results of these extensive experiments are summarized in Tables B.1 and B.3 and presented in the Appendix.

The performance of DA is evaluated in terms of the number of logic-based Benders cuts added; i.e., feasibility cuts (3.41) and (3.42) and deviation tolerance cuts (3.43), and the number of iterations completed within the time limit. The average of these performance metrics are depicted in Figures 3.6 and 3.7, for *CA339* and *US-E420*, respectively.

It is observed from these results that the number of iterations DA can complete in

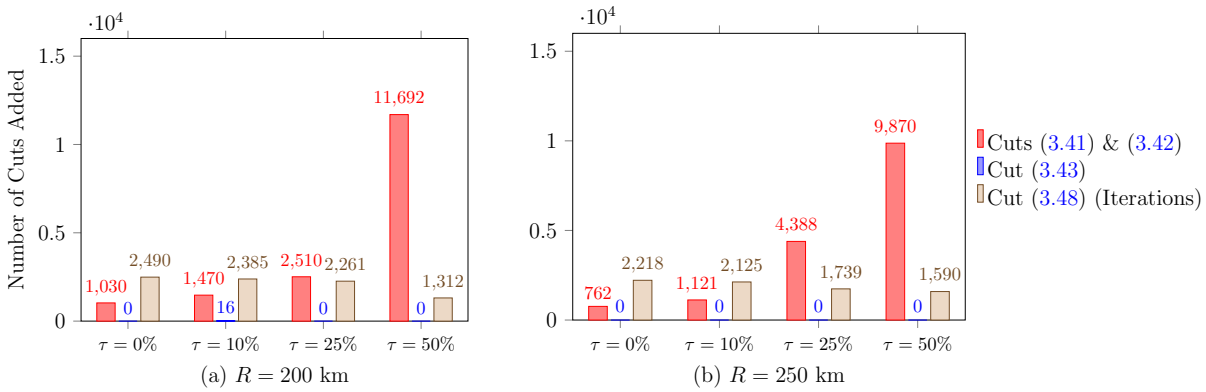


Figure 3.6: Algorithm performance metrics for *CA339* instances.

1-hour time limit tend to decrease slightly for *CA339* instances as deviation tolerance increases. This is due to the decreasing number of required feasibility cuts. Figure 3.6 (a) and (b) also show that the deviation tolerance cuts (3.43) are utilized just 16 times throughout the experiments. This is a clear indication that introducing permissible arcs is effective for generating solutions that are deviation tolerance feasible.

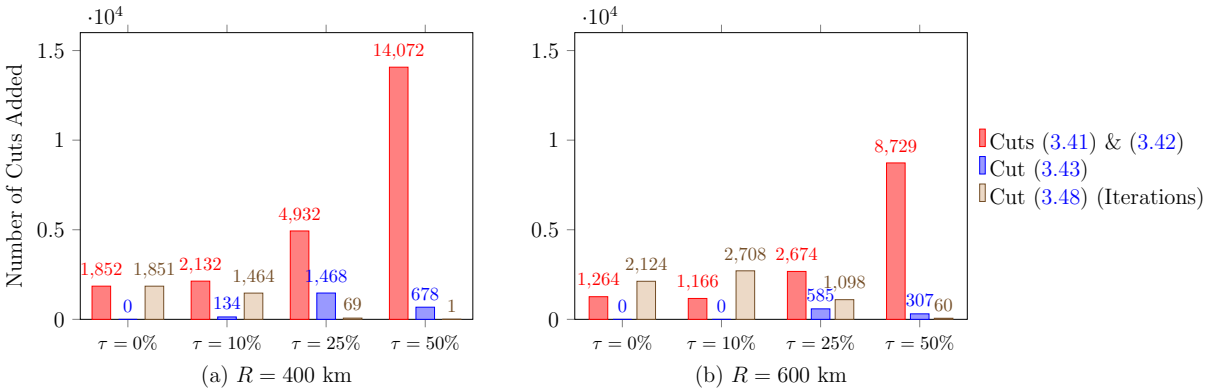


Figure 3.7: Algorithm performance metrics for *US-E420* instances.

The expanded networks of *US-E420* include over 700,000 arcs for $R = 400$ and 1 million arcs for $R = 600$ at high levels of deviation tolerances. This results in tens of millions of decision variables and constraints for the bilevel optimization model. When R is lower and the number of OD pairs is higher, the likelihood of obtaining an infeasible set of station locations in the algorithm increases. Consequently, it is inevitable to add thousands of

feasibility cuts even with the efficient approach that we are proposing. In particular, observe from Figure 3.7 (a) that when $R = 400$ and the deviation tolerance is 50%, the algorithm adds over 14,000 feasibility cuts, whereas it can only complete one iteration. Moreover, the number of times deviation tolerance cuts are utilized is significantly higher for *US-E420* compared to *CA339* instances.

3.5.3 Insights from Bilevel Solutions

In this section, we derive insights from the solutions by evaluating the effects of deviation tolerance, vehicle range, and service level parameters under the cooperative response. We first analyze the *N25* solutions and then move on to the large-scale instances.

For all R configurations of the small-sized *N25* network, 0% and 10% deviation tolerance solutions turn out to be identical. Moreover, the number of located stations does not change with varying α or β values for a fixed deviation tolerance. This indicates that the service level only affects the capacities of the stations for the tested configurations. In all instances, stricter α and β result in installing more capacity, and hence, a higher objective function value. For example, when $R = 15$ and the deviation tolerance is 25%, $\alpha = 1$ min and $\beta = 95\%$ combination opens 31% more charging stalls compared to the case where $\alpha = 30$ min and $\beta = 80\%$. Higher deviation tolerance solutions require less stations and/or less stalls and result in a lower objective function value, e.g., when $R = 15$, 0% deviation tolerance solutions utilize eight stations and an average of 136 charging stalls in total, whereas 25% deviation tolerance solutions require six stations and 132 charging stalls on average. The value of using a bilevel framework is higher when solving higher deviation tolerance instances since the solution space of the follower's route choice response grows larger. Within the bilevel framework, the follower picks the shortest available route; and, the average deviation from the shortest paths never exceeds 5.65% among the optimal

solutions of all the 50% deviation tolerance instances.

Next, we analyze *CA339* and *US-E420* solutions obtained by DA to derive insights from sensitivity analysis on the problem parameters. For *CA339* solutions, similar to the *N25* results, varying α only changes the capacity levels deployed but not the number of located stations. As the geographical shape of the *CA339* network is narrow and potential OD nodes are not scattered to the boundaries, between two and seven stations are observed to be able to sufficiently serve various parameter settings that we tested on this network.

In Figure 3.8, we further analyze the difference in the objective values based on the waiting time threshold and deviation tolerance of *CA339* solutions. These charts demonstrate how the objective value decreases with higher deviation tolerance (τ) and increases with stricter waiting time threshold (α). The most significant cost-savings is achieved when 0% deviation tolerance is increased to 10%.

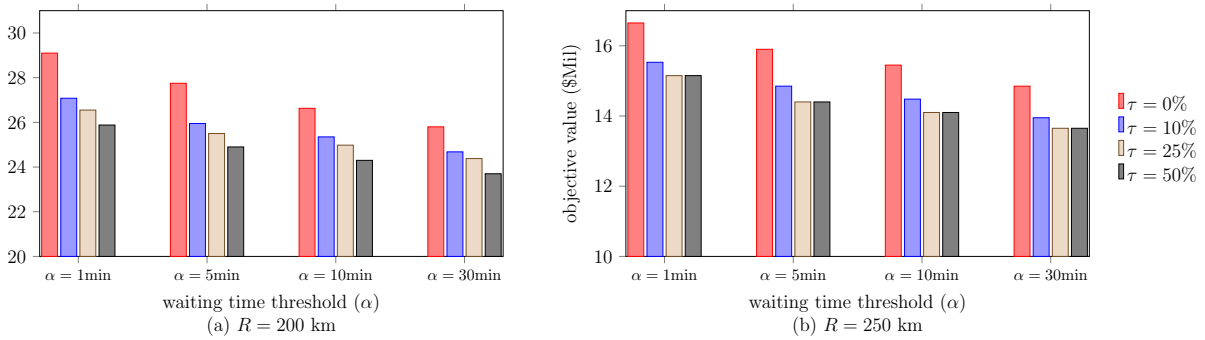


Figure 3.8: Objective value comparison of CA339 solutions, $\beta = 90\%$.

Figure 3.9 show the average and maximum deviation metrics at 10%, 25%, and 50% deviation tolerance levels for $\alpha = 10$ min instances of *CA339*. For $R = 250$, average deviation does not exceed 0.88% even when the deviation tolerance is set to 50%. However, this performance metric can be as high as 7.07% at 50% deviation tolerance for $R = 200$. Maximum deviation metric follows the deviation tolerance limit closely for this range level, whereas it has nearly 30% slack for $R = 250$ instances at 50% deviation tolerance.

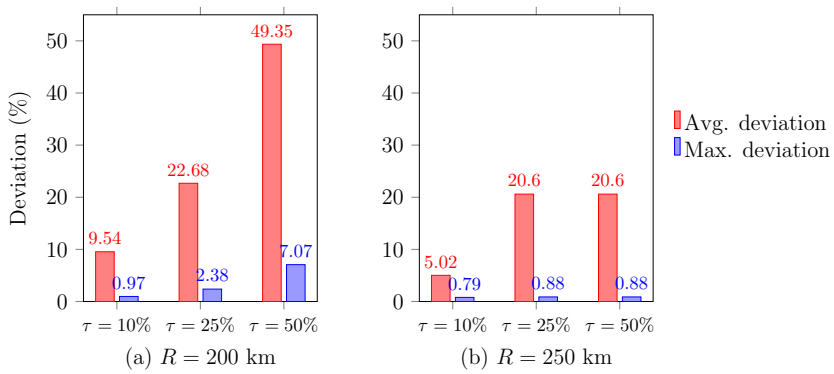


Figure 3.9: Average and maximum deviation comparison of *CA339* solutions, $\alpha = 10$, $\beta = 90\%$.

The *US-E420* instances require significantly more facilities than the other networks that we tested. The highest number of stations located is 97 for $R = 400$, whereas the lowest is 25 for $R = 600$ instances. We also observe that $R = 400$ instances require almost twice as many stations as their $R = 600$ counterparts although the number of OD trips is only 9% more for the latter. As it is the case with the other two networks, lower values of α might require up to 20% additional capacity at the established stations whereas it does not change the station count in any of the experiments. The station locations are observed to be insensitive to the changes in α in the majority of the instances.

Figure 3.10 shows the sensitivity of the objective values based on the tested waiting time threshold and deviation tolerance values for *US-E420* solutions. Similar to the *CA339* results, the difference between the objective values of the solutions with 0% and 10% deviation tolerances is the highest. On the other hand, the objective function value differences between 25% and 50% deviation tolerance instances are relatively smaller particularly for higher R. For both $R = 400$ and $R = 600$ instances, the $\alpha = 1$ min solutions are approximately 5.5%, 10.2%, and 20% more costly than $\alpha = 5, 10, 30$ min solutions, respectively.

The average and maximum deviation comparison of *US-E420* solutions for $\alpha = 10$ instances are illustrated in Figure 3.11. For both $R = 400$ and $R = 600$ solutions, the

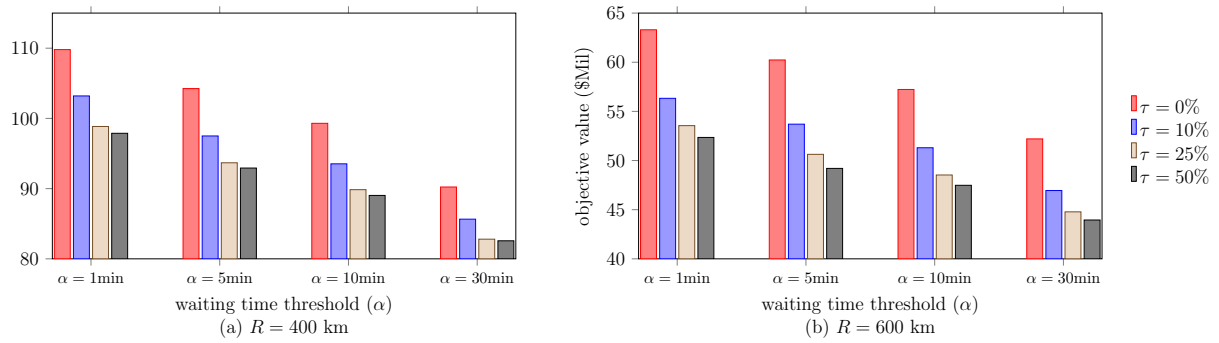


Figure 3.10: Objective value comparison of US-E420 solutions, $\beta = 90\%$.

value of maximum deviation closely follows the deviation threshold. On the other hand, the average deviation is only as high as 13% and as low as 1.42% over all solutions.

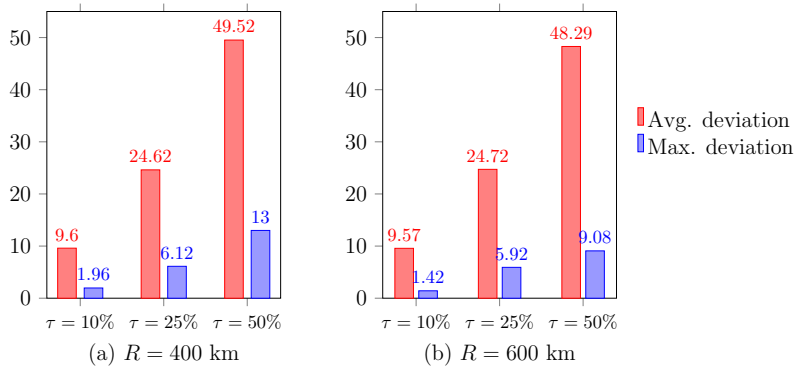
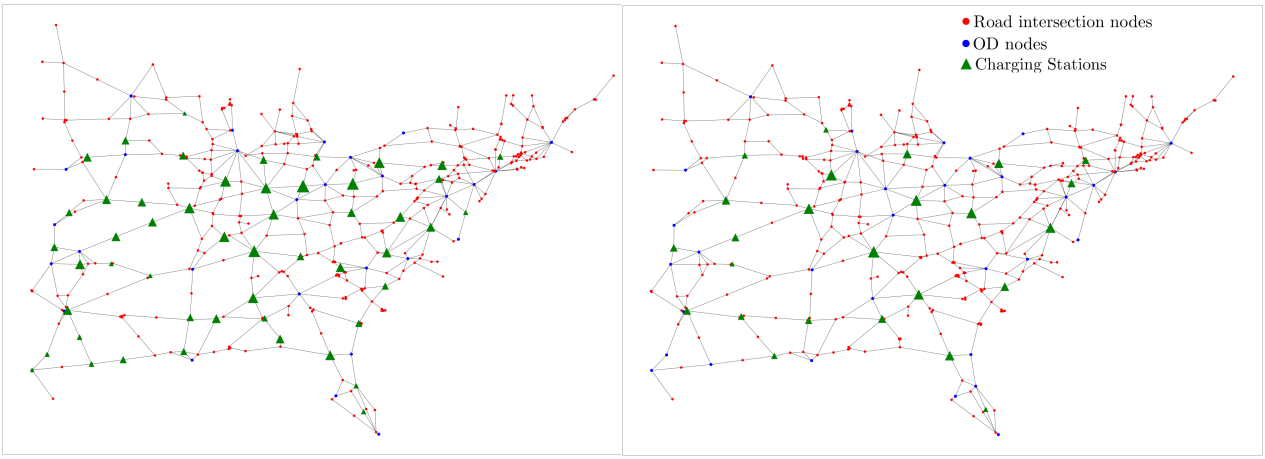


Figure 3.11: Average and maximum deviation comparison of *US-E420* solutions, α=10 min, $\beta = 90\%$.

Figure 3.12 presents the spatial comparison of charging station locations and sizes on two representative solutions with two different deviation tolerances for $R = 600$ km, $\alpha = 10$ min, and $\beta = 90\%$ on *US-E420*. The sizes of the charging stations are depicted using different sizes for the green triangles. When the deviation tolerance is 10% (Figure 3.12a), there are 39 charging stations. The largest stations are located near the center of the network, whereas numerous smaller stations are scattered throughout the region. On the other hand, when the deviation tolerance is 50% (Figure 3.12b), there are only 25 charging stations and the majority are of larger sizes that are located closer to the center; the

number of smaller charging stations is significantly fewer.



(a) $\tau = 10\%$ (b) $\tau = 50\%$

Figure 3.12: *US-E420* $\alpha = 10$ min, $\beta = 90\%$ solutions for $\tau = 10\%$ and $\tau = 50\%$.

For a more extensive analysis on how the deviation tolerance and vehicle range affect the size of the facilities, we provide the histograms in Figure 3.13 that compare the solutions with fixed α and β values at 10 min and 90%, respectively. For both $R = 600$ and $R = 400$, instances with smaller deviation tolerances require significantly more stations of smaller sizes. Higher deviation tolerances allow to aggregate the charging demand at fewer stations that are of larger size. This aggregation not only locates fewer facilities, but it also requires less capacity in total (see Table B.3 in Appendix). This indicates the importance of determining the right deviation tolerance for the right applications.

3.5.4 Analyzing EV User Response Behavior

In this section, we compare optimal cooperative response solutions with that of the uncooperative response solutions on select *N25* instances. Recall that the uncooperative response solutions are obtained by the adaptation of the algorithm with the modification of P_S as described in Section 3.4.2. The optimality of the uncooperative response solutions can only be claimed by exhausting all possible alternative solutions of the problem, which is not

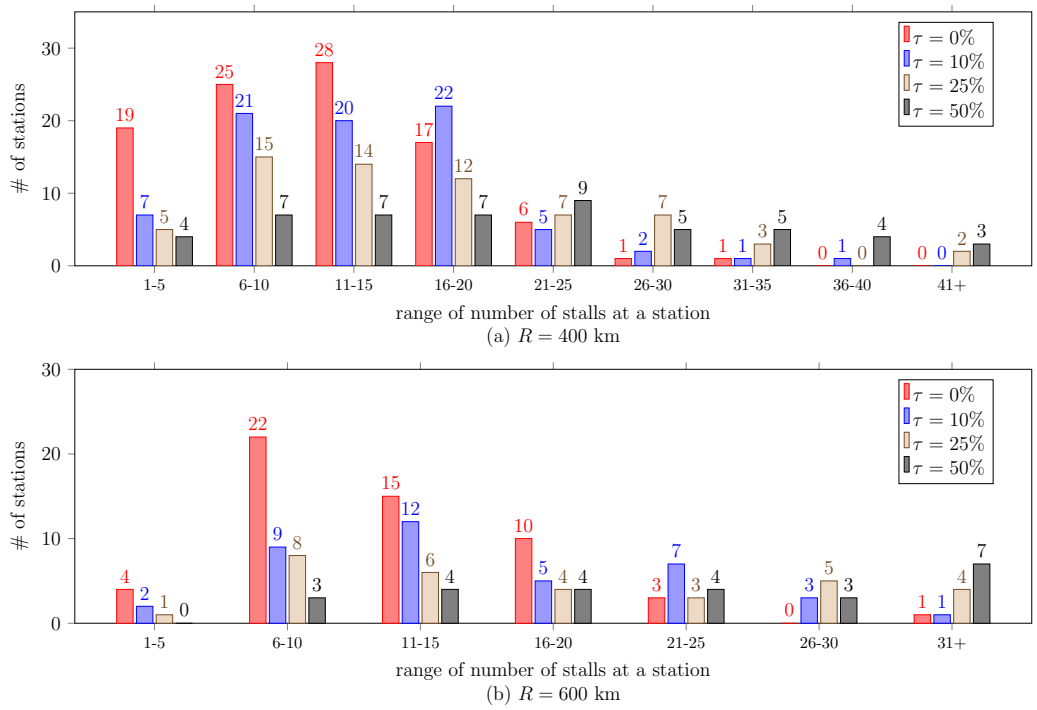


Figure 3.13: Deviation histograms for the number of charging stalls installed, $\alpha = 10$ min, $\beta = 90\%$, (US-E420).

tractable in practice. Hence, we use the same stopping condition for the uncooperative response version of the algorithm as with the cooperative response.

Since the follower's problem ensures that each OD pair follows their minimum length path, the average and maximum deviation metrics stay the same when exactly the same set of stations are open in the cooperative and uncooperative response solutions. In particular, for the solutions of $R = 20$ and $R = 15$ instances, the set of charging station locations stays the same, whereas the uncooperative response solutions result in different required capacity levels. On the other hand, the set of open stations are different for all $R = 25$ instances when the deviation tolerance is greater than 0%. This indicates that the locations of the charging stations may change based on the position determined for the follower. The complete results of this comparison are provided in Table B.4 in the Appendix.

Interestingly, smaller differences in the objective function values are observed when

the deviation tolerance is higher. For that matter, the cooperative and uncooperative response solutions turn out to be equivalent when the deviation tolerance is 50% for both $R = 25$ and $R = 20$. This is an insightful result for the decision makers since it shows that having a higher deviation tolerance mitigates the difference that might result from an uncooperative response of the EV users. In other words, higher deviation tolerances for the cooperative response case plays an important role to provide more robust solutions for the cases in which the follower may not behave as desired. On the other hand, lower deviation tolerance levels tend to yield higher differences in the objective function values between the cooperative and uncooperative response solutions. Intuitively, one might think that the bilevel model does not make a difference when the deviation tolerance is set to 0% as the route choices are no longer relevant. However, the computational results reveal that the objective function difference between the cooperative and uncooperative response solutions is usually higher for the 0% deviation tolerance instances. This indicates that using lower deviation tolerances, especially 0%, might generate solutions that may easily be overridden by uncooperative responses of the EV users in terms of choosing at which station(s) to recharge.

3.6 Conclusion

In this chapter, we present a bilevel optimization model to determine the strategic locations and capacities of DC fast-charging stations under stochastic demand to enable long-distance travel with EVs. The model allows the possibility of detours from shortest paths within a deviation tolerance without endogenously determining the deviation paths. Through the bilevel framework, the model takes into account the route choice behavior of EV users, as well as their cooperative or uncooperative charging station selection responses, when alternative solutions exist. Besides providing a general formulation of the problem,

we model the M/M/c queuing system characterization of the DC fast-charging stations for the case of Poisson arrivals and exponential service times. We derive the MILP equivalent of the chance-constrained stochastic bilevel optimization formulation by modeling the underlying charging station network as an open multi-server Jackson network.

We develop a decomposition based algorithm to solve the proposed bilevel optimization model under both cooperative and uncooperative follower responses. The algorithm utilizes a new exact logic-based Benders methodology for the uncapacitated version of the problem as a subroutine. The solution quality and the computational performance of the algorithm are verified on smaller-sized network instances, where it finds optimal solutions for all of the 144 test instances. The computational performance comparison with a commercial solver reveals that the algorithm is much more efficient on instances with higher deviation tolerances and lower range values. Allowing for higher deviation tolerances (up to 50%) can especially be relevant for autonomous long-distance travel with EVs.

The results demonstrate that the number of located charging stations is insensitive to the changes in the service level requirements. Typically, more stringent service level thresholds require more capacity to be installed at the charging stations. The solutions with higher deviation tolerance are observed to require less infrastructure cost and are not prone to generate alternative stopping options for the EV users. This leads to robust solutions that are equivalent for the cooperative and uncooperative responses of the EV users and protect the decision-maker against the worst possible outcome. On the other hand, the solutions allowing for lower deviation tolerance tend to increase the number of required stops and result in higher differences in the optimal objective function values obtained under the cooperative and uncooperative responses.

Future work may focus on different types of queuing system characterizations for charging stations. For example, an extension with M/D/c systems may be a viable option for

centralized systems with autonomous vehicles or for commercial applications that can enforce deterministic service times to the users of a charging station. Another research dimension may consider a multi-period framework to determine infrastructure expansion plans considering the growth in charging demand due to the increasing adoption rate of EVs.

Chapter 4

Infrastructure Design for Shared Autonomous Transportation

Autonomous vehicles (AVs) have the potential to improve traffic safety, reduce congestion, private vehicle ownership and vehicle-miles-traveled, promote environmental sustainability by decreasing emissions, and offer convenience by reducing the drivers' value of time (Alcorn and Kockelman, 2021; Jones and Leibowicz, 2019; Brownell and Kornhauser, 2014). Extensive use of AVs may reshape the need for parking spaces, which could transform the land use in cities (Fagnant and Kockelman, 2015; Bagloee et al., 2016).

In the early adoption phases, AVs are likely to be introduced to the urban transportation market by providing shared mobility services since they are expected to be significantly costlier than human-driven vehicles (Bansal and Kockelman, 2017; Shaheen and Cohen, 2019). Widespread shared mobility services with AVs could help better utilize transportation resources and provide a dramatic reduction in the number of privately-owned passenger vehicles (Golbabaei et al., 2021).

To provide enhanced safety and versatility for AVs, autonomous travel may be sup-

ported by smart infrastructure equipped with auxiliary ambient and obstacle perception technologies known as the *vehicle-to everything technology* (V2X). This technology paves the way towards a single, centralized and globally-optimal control system that would further promote mobility-as-a-service systems. The deployment of smart mobility corridors of AV lanes with infrastructure-supported or infrastructure-enabled autonomous driving would play an important role for a successful transition into high automation (Level 4) transportation systems in cities for several reasons: (i) when dedicated, AV lanes limit the risks associated with human driver-AV interactions in the initial phases of the implementation (Chen et al., 2017), (ii) they improve navigation and maintain safety during non-ideal driving/road/infrastructure conditions or on-board sensor failures (Van Brummelen et al., 2018), (iii) coupled with vehicle-to-vehicle systems, they harness the connectivity benefits of AV technology capabilities so that AVs can form platoons, hence further improve vehicle flows and congestion in urban areas (Ghiasi et al., 2017). Given these benefits of employing AV lanes, it is important to determine optimal deployment strategies of such lanes within an urban transportation network to maximize infrastructure efficiency.

To avoid congestion and reduce the need for curbside parking in urban areas, shared autonomous mobility service providers require facilities where AVs can idle before picking up and after dropping off passengers. These sites are often referred to as staging facilities (Duvall et al., 2019). These futuristic facilities, which act as smart mobility hubs, would offer automated centralized parking, auxiliary AV support services that include charging and contribute to a cleaner and resilient transportation future by utilizing solar or kinetic energy harvesting. AV technology coupled with centralized parking would allow parking vehicles very close to each other, saving up to 62% more space compared to conventional parking lots (Nourinejad et al., 2018). Moreover, the reduced need for curbside parking in densely populated urban areas not only creates a pleasant and safer environment for pedestrians but also improves vehicle flows by enabling all parts of streets to be effectively

used for transportation. An important problem is then to determine the best locations for these staging facilities that may encourage wider adoption of shared AVs and efficient vehicle flows in condensed urban areas (Golbabaei et al., 2021).

In this study, we work on determining optimal deployment of AV lanes to exploit the cost-effectiveness and safety benefits of infrastructure-supported autonomy, while simultaneously determining optimal locations of staging facilities that would help reduce cruising congestion besides providing essential services, such as recharging, to fleets of shared AVs. Concurrent consideration of these strategic investment decisions is key to achieving better designs since the selection of staging facility locations will heavily impact the roads that AVs are going to use, and this will directly impact the AV lane deployment decisions. To the best of our knowledge, this is the first study to optimize the spatial distribution of AV lanes and staging facilities at the same time.

We adopt a multi-objective optimization approach that allows decision-makers to evaluate the trade-offs. In particular, we seek to find the optimal locations of staging facilities utilizing a bi-objective model that minimizes total travel distance and the total AV travel not occurring on AV lanes with respect to a given AV lane deployment budget and a number of staging facilities to locate. For each origin-destination (OD) trip, an AV leaves from a staging facility and travels from the origin to the destination node of the trip, which are the designated pick-up and drop-off locations, and returns to a staging facility while using the dedicated AV lanes en-route. Each trip's origin and destination are assigned to their respective nearest staging facility. We formulate the problem with a mixed-integer model and also develop a Benders decomposition algorithm with a non-dominated cut generation scheme to be able to solve large-size instances. Through extensive computational experiments, we evaluate the optimal solutions, network designs, and the effects of AV lane budget and staging facility count on the benchmark Sioux Falls transportation network.

The contributions of this study are *(i)* a novel problem definition for infrastructure design of shared autonomous vehicles, *(ii)* a bi-objective mathematical formulation for this problem, *(iii)* a decomposition algorithm to solve large-scale instances, and *(iv)* extensive managerial insights analyzing the trade-offs between AV lane deployment and staging facility location decisions.

The remainder of this chapter is organized as follows. Section 4.1 provides the review of the literature focusing on AV-related infrastructure planning problems and highlights the research gaps. In Section 4.2, the staging facility location and AV lane deployment problem is defined and a multi-objective mathematical model is presented. Section 4.3 introduces the Benders reformulation of the model and the developed subproblem solution methodology that leads to non-dominated Benders optimality cuts. In Section 4.4, we provide and analyze the results of comprehensive computational experiments conducted to evaluate the effects of problem parameters on optimal solutions and derive managerial insights. Finally, Section 4.5 concludes this study with some remarks and reflections as well as future research opportunities.

4.1 Literature Review

In this section, among the AV-related optimization problems, we will review the literature related to AV operations and infrastructure planning, in particular, studies on AV traffic assignment and lane deployment, AV parking policy and management, and shared AV system design and analysis problems.

Many researchers focus on a problem definition that only includes privately owned AVs. In this context, there are numerous publications that study traffic assignment problems, and investigate AVs and HVs route-choice equilibrium conditions. Wang et al. (2021a) and

Wu et al. (2020) work on the toll-setting problem assuming AV lanes are given. Particularly, Wang et al. (2021a) develops a multi-class traffic assignment problem with elastic demand and considers dedicated AV and AV toll (AVT) lanes. AVT lanes are assumed to grant free access to AVs while allowing HVs to access these lanes by paying a toll. The proposed formulation determines optimal toll rates to improve network performance and an exact solution algorithm, that is based on linearizing the variable terms in the toll-setting upper-level problem by leveraging the sensitivity analysis results of the lower-level equilibrium problem, is developed. Wu et al. (2020) considers a mixed network with AV lanes and non-autonomous local streets, and models AV route choice decisions under both user equilibrium and system optimum, where cordon-based tolling schemes are used to characterize system optimum. Liu et al. (2021) stands out in this literature as the authors explore an infrastructure-enabled autonomous driving system and develop a network user equilibrium model that considers a mixed-autonomy traffic, i.e., AVs that may or may not use AV lanes, on a given set of such lanes. Inconvenience costs due to transitions between autonomous and manual driving are explicitly modeled. Combining traditional route-choice and driving-mode-choice equilibrium conditions, they propose a nonlinear formulation and solve it using a route-swapping based solution algorithm.

There are other studies that relax the "given AV lanes assumption", and decide on AV lane deployment while still working with traffic equilibrium conditions. Chen et al. (2016) describes the flow distribution of AVs and HVs on the transportation network via a multi-class network equilibrium model initially for a given set of AV lanes. Using these results, the authors use a diffusion model to forecast the market penetration of autonomous vehicles, and with this new information, they propose a model that optimizes the deployment plan of AV lanes, particularly determining when, where and how many lanes to be deployed to minimize total social cost of the system. Chen et al. (2017) focuses on determining certain areas of road networks that are dedicated to AVs, referred as AV zones. These zones

consist of set of links that are tailored to AVs and it is assumed that AVs follow given routes in these zones. The network equilibrium model captures mixed-routing behaviors where HVs need to avoid AV zones and AVs decide whether to access the zones when making their routing choices. Given the equilibrium model results, the authors propose a bilevel optimization model where the lower-level problem is the equilibrium model and the upper level model determines AV lane deployment. The objective is to minimize total social cost of the system. This model is solved using heuristic algorithms. Liu and Song (2019) considers AVTs and toll rate optimization to regulate AV traffic and congestion. Similar to Chen et al. (2016) and Chen et al. (2017), the authors first propose a multi-class user equilibrium model to describe the flow distribution. A nonlinear robust optimization model is then introduced to find the optimal AV and AVT lane deployment strategy and toll rates that minimize the worst-case total system travel time.

There are also studies that focus on the parking aspect for privately-owned AVs. Levin et al. (2020) proposes traffic assignment models with destination choice for AV parking and considers empty repositioning at varying market penetration levels. Additionally, it introduces a bilevel optimization problem of adjusting zone-specific parking fees to manage repositioning and cruising congestion. Wang et al. (2021b) proposes a continuous-time stochastic dynamic model for the parking management of connected AVs, incorporating interactions among parking garages by regulating parking rates. A solution algorithm based on Pontryagin's minimum principle is introduced and the effectiveness of the proposed approach is shown via Monte Carlo simulations. Bahrami et al. (2021) studies the parking behavior of privately owned AVs in downtown areas considering cruising and home-routing. They propose an equilibrium model for AV parking choices and discuss that time-based congestion toll may be necessary to manage AVs parking in the downtown and discourage intentional slowing down in cruising. Bahrami and Roorda (2022) uses simulation to study the impacts different parking policies for privately-owned AVs. The policies tested include

routing AVs to park at home after downtown drop-offs, dynamic parking rates to mitigate AV cruising congestion, and assigning the same parking rate across all the parking lots. Liu et al. (2021); Liu (2018) and Zhang et al. (2019) provide models for an equilibrium analysis of traffic assignment and commuter parking in the era of AVs. Particularly, Liu et al. (2021) aims to present findings to provide a better understanding of the interactions between AV travel and urban characteristics. The results show that adopting frequency-based AVs, i.e., the AVs that have higher passenger capacity and are dispatched on a timetable, alleviates traffic congestion and may reduce the cost for suburban residents. All of these studies assume that existing parking lots and curbside parking zones will be used for AV parking.

There are several studies that explore the impact of shared AVs on urban parking demand via simulation (Okeke, 2020; Kumakoshi et al., 2021; Zhang et al., 2015). Results of Zhang et al. (2015) indicate that it may be possible to eliminate up to 90% of parking demand for system at a market penetration rate of 2%. Kumakoshi et al. (2021) points out that cruising shared AVs may increase congestion in the urban zones where the reduction of parking demand is higher. The only study that considers parking infrastructure while focusing on a shared AV system design is by Li et al. (2021). They introduce a time-dependent shared AV system design problem and formulate it as a MILP. They jointly consider AV fleet size, parking infrastructure deployment, and AV routes while minimizing total cost of the system. They develop a heuristic algorithm to provide an upper bound for the proposed model. This study does not incorporate infrastructure-enabled or supported AV operations, i.e., the deployment of AV lanes are left out of scope.

What emerges clearly from this literature review is that the two infrastructure-related problems, i.e., parking facility location and AV lane deployment, have never been modeled simultaneously although these problems are very much intertwined.

4.2 The Staging Facility Location and AV Lane Deployment Problem

In this section, we first define the staging facility location and AV lane deployment problem and then present a mathematical formulation.

4.2.1 Problem Definition

In this problem setting, the source of demand is defined in terms of OD flows such that each OD pair represents an AV trip request of a user between two points, i.e., between a pick-up and a drop-off location. Upon receiving a trip request, an AV is dispatched from the staging facility that is closest to the pick-up location to pick-up the user from this origin node. The AV then travels from the origin to the destination of this trip on its shortest path and drops off the passengers at the destination. Finally, the empty AV returns to the nearest staging facility after dropping the passengers off at the desired destination. Accordingly, for each OD pair, the AV trip has three legs: from the nearest staging facility to the origin, from the origin to the destination, and from the destination to the nearest staging facility. AVs take the shortest paths along each leg of a trip.

Figure 4.1 depicts the problem setting on an illustrating example. This example includes two OD trips ($O_1 \rightarrow D_1$ and $O_2 \rightarrow D_2$) on a service region which includes three staging facilities that are represented with red squares. As a consequence of staging facility locations and the allocations based on nearest-assignment requirement, the shortest paths that AVs will follow are shown with black and blue arrows for OD_1 and OD_2, respectively. Along these paths, AVs are using the road segments with the deployed AV lanes that are highlighted with red connections.

As depicted in Figure 4.1, we consider a mixed-autonomy setting where AV travel

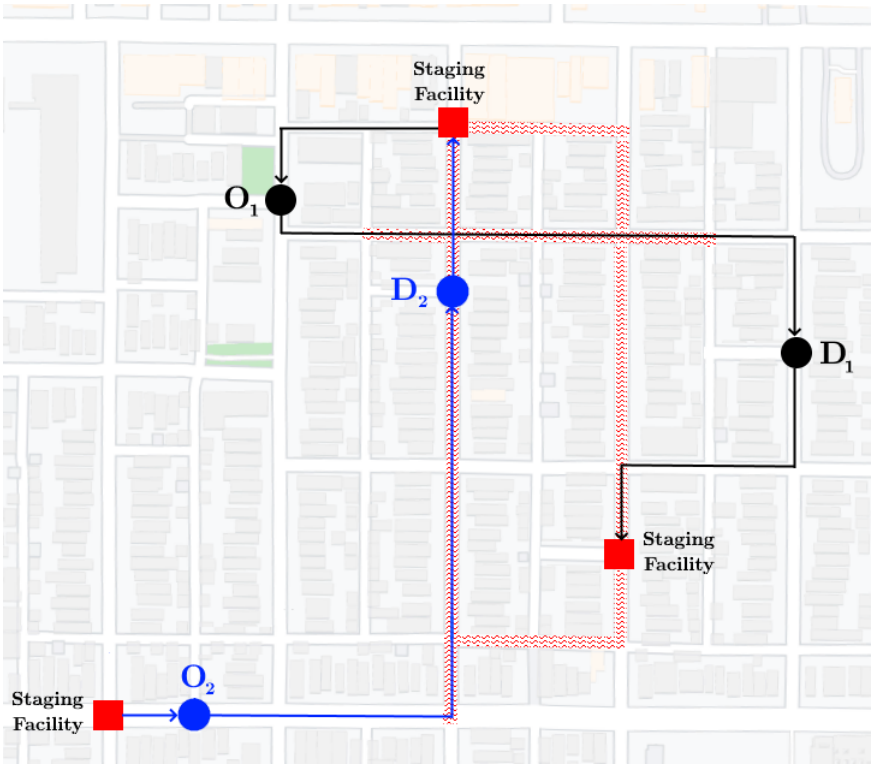


Figure 4.1: An illustrating example.

outside AV lanes is allowed, but it is discouraged to elevate the benefits of infrastructure-supported autonomy (Liu et al., 2021). Ideally, AVs would only travel along the roads that have AV lanes but since such a large-scale infrastructural investment may be very costly to implement in the early adoption phases of shared AV mobility in urban transportation, we consider the setting where there is a given budget allocated to the deployment of AV lanes.

Given the frequency of OD flows and a limited AV lane deployment budget, the problem is to determine the optimal locations of a given number of staging facilities and the optimal deployment of AV lanes on this network. It is essential to optimize the location and AV lane deployment decisions simultaneously since both components of the problem are intertwined. In other words, the locations of the staging facilities determine the paths to be taken by the AVs and the total travel distance; consequently, this will directly impact the

AV lane deployment decisions to reduce AV travel outside the AV lanes. This intertwined structure leads to a bi-objective problem setting by nature where a decision-maker would like to minimize the total AV travel distance to limit cruising congestion caused by the AVs, and also minimize the total AV travel occurring outside the AV lanes to exploit the benefits of utilizing infrastructure-supported autonomy. Accordingly, we model a bi-objective problem where the first objective is to minimize the total AV travel distance and the second objective is to minimize the total AV travel occurring outside the AV lanes. The next section presents the notation and a mathematical formulation of this problem.

4.2.2 Mathematical Formulation

Let the underlying urban transportation network be defined as $G = (N, A)$, where N denotes the set of nodes, e.g., road intersections, and A denotes the set of directed arcs, e.g., road segments of the transportation network. Additionally, A_{ij} denotes the set of shortest path arcs from node $i \in N$ to $j \in N$, and ℓ^*_{ij} denotes the shortest path length between nodes $i, j \in N$. Consider O_k and D_k as the origin and destination nodes of the trip associated with OD pair $k \in K$, where K denotes the set of all OD pairs. Let N_k denote $\{O_k, D_k\}$ for $k \in K$, and the set of all pick-up and drop-off locations is defined as $\mathcal{N} = \bigcup_{k \in K} N_k$. The frequency factor of each OD trip is denoted by f_k, $k \in K$.

The total budget allocated for the deployment of AV lanes is represented with B. The cost of AV lane deployment on an arc of the network is denoted by c_a, and the length of each arc is denoted by ℓ_a for $a \in A$. Let p denote the maximum number of staging facilities to be located by the decision maker. Furthermore, let α and β denote the weights (importance factors) of the two objective function components perceived by the decision maker, namely, minimizing total AV travel between pick-up/drop-off locations and staging facilities and minimizing total AV travel not occurring on AV lanes, respectively.

In order to model the nearest-assignment relationship, let the r-th closest candidate staging facility location to node $j \in N$ be denoted by j_r where $r = 1, \ldots, |N|$. The decision variables of the model are defined as follows:

$$y_i = \begin{cases} 1, & \text{if staging facility } i \in N \text{ is open,} \\ 0, & \text{otherwise.} \end{cases}$$
$$z_a = \begin{cases} 1, & \text{if arc } a \in A \text{ has an AV lane,} \\ 0, & \text{otherwise.} \end{cases}$$
$$w_{ij} = \begin{cases} 1, & \text{if node } j \in \mathcal{N} \text{ is assigned to staging facility } i \in N, \\ 0, & \text{otherwise.} \end{cases}$$

Using the aforementioned parameters and decision variables, the staging facility location and AV lane deployment problem is formulated as follows:

$$[\mathcal{P}] \quad \min \quad \alpha\left(\sum_{i\in N}\sum_{k\in K}\sum_{j\in N_k} f_k \ell^*_{ij} w_{ij}\right) + \beta\left(\sum_{k\in K}\sum_{i\in N}\sum_{j\in N_k}\sum_{a\in A_{ij}} f_k \ell_a w_{ij}(1-z_a) - \sum_{k\in K}\sum_{a\in A_{O_kD_k}} f_k \ell_a z_a\right) \tag{4.1}$$

$$\text{s.t.} \quad \sum_{i\in N} y_i = p \tag{4.2}$$

$$\sum_{i\in N} w_{ij} = 1 \qquad j \in \mathcal{N} \tag{4.3}$$

$$w_{ij} \le y_i \qquad i \in N, \quad j \in \mathcal{N} \tag{4.4}$$

$$\sum_{s=r+1}^{|N|} w_{j_s,j} + y_{j_r} \le 1 \qquad j \in \mathcal{N}, \quad r = 1, \ldots, |N|-1 \tag{4.5}$$

$$\sum_{i\in N}\sum_{a\in A} c_a z_a \le B \tag{4.6}$$

$$y_i, z_a, w_{ij} \in \{0,1\} \qquad i \in N,\ a \in A,\ j \in \mathcal{N} \tag{4.7}$$

$\mathcal{P}$ is presented as a multi-objective optimization problem that combines two objectives into one composite term defined by (4.1) using the weighted sum method. The first component of (4.1) defines the total AV travel distance to/from OD nodes from/to assigned staging facilities, whereas the second component defines the total distance of AV travel not

occurring on AV lanes. Constraint (4.2) states that p staging facilities are to be located. Constraint (4.3) ensures that every OD node is assigned to exactly one staging facility and constraint (4.4) ensures that this assignment is only possible to open staging facilities. The set of closest assignment constraints are defined by (4.5), and these assign each OD node to its nearest open staging facility. Constraint (4.6) sets a maximum budget of B to deploy AV lanes on the network. Lastly, constraint (4.7) defines the binary domains of the decision variables.

Although there are several versions of closest assignment constraints that serve the same purpose, constraint (4.5), that is first proposed by Wagner and Falkson (1975), is shown to be computationally superior than the others (Espejo et al., 2012). After preliminary computational experiments with other types of closest assignment constraints, we also observed that (4.5) yields faster solution times for $\mathcal{P}$. Thus, we choose to use Wagner and Falkson (1975)'s version in the formulation.

The second component in the objective function (4.1) is non-linear due to multiplication of binary variables w_{ij} and z_a. It may be linearized by introducing a new binary variable u_{aij}, substituting $w_{ij}z_a$ with it in (4.1), and including the following constraints in the formulation:

$$u_{aij} \geq w_{ij} + z_a - 1 \qquad i \in N, \quad j \in \mathcal{N}, \quad a \in A_{ij} \tag{4.8}$$

$$u_{aij} \leq w_{ij} \qquad i \in N, \quad j \in \mathcal{N}, \quad a \in A_{ij} \tag{4.9}$$

$$u_{aij} \leq z_a \qquad i \in N, \quad j \in \mathcal{N}, \quad a \in A_{ij} \tag{4.10}$$

$$u_{aij} \in \{0, 1\} \qquad i \in N, \quad j \in \mathcal{N}, \quad a \in A_{ij} \tag{4.11}$$

Since the cardinality of set A_{ij} is much smaller than cardinality of A itself, the number of u_{aij} variables as well as the number of linearization constraints (4.8)-(4.11) can be kept

relatively low. After substituting u_{aij} in (4.1), it may be rewritten as:

$$\min\ \alpha\Bigg(\sum_{i\in N}\sum_{k\in K}\sum_{j\in N_k} f_k \ell^*_{ij} w_{ij}\Bigg) + \beta\Bigg(\sum_{k\in K}\sum_{i\in N}\sum_{j\in N_k}\sum_{a\in A_{ij}} f_k \ell_a (w_{ij} - u_{aij}) - \sum_{k\in K}\sum_{a\in A_{O_k D_k}} f_k \ell_a z_a\Bigg) \tag{4.12}$$

With these changes, we present $\mathcal{P}_r$ as a reformulation of the $\mathcal{P}$:

$[\mathcal{P}_r]$ min (4.12)

s.t. (4.2) – (4.6), (4.8) – (4.10)

$$y_i, z_a \in \{0,1\} \qquad i \in N,\ a \in A \tag{4.13}$$

$$w_{ij}, u_{aij} \geq 0 \qquad i \in N,\ j \in \mathcal{N},\ a \in A_{ij} \tag{4.14}$$

Proposition 4.1. *$\mathcal{P}_r$ always has an optimal solution with integral w_{ij} and u_{aij} values.*

Proof: Without loss of generality, assume that there exists some $w_{i_1\hat{\jmath}}$ and $w_{i_2\hat{\jmath}}$ that take fractional values in an optimal solution of $\mathcal{P}_r$ such that $w_{i_1\hat{\jmath}} + w_{i_2\hat{\jmath}} = 1$ by constraint (4.3), and $y_{i_1} = y_{i_2} = 1$ by constraints (4.4) and (4.13). Without loss of generality, assume that i_1 is closer to $\hat{\jmath}$ than i_2. By the closest assignment constraints (4.5), when $\hat{\jmath}_r = i_1$, then $w_{i_2\hat{\jmath}} = 0$, which leads to a contradiction. In addition, $w_{i_1\hat{\jmath}} = 1$ by constraint (4.3). Therefore, every w_{ij} takes the value of 0 or 1 in an optimal solution of $\mathcal{P}_r$.

When both z_a and w_{ij} are binary, it is trivial to show that $u_{aij} = 1$ if both corresponding $z_a = 1$ and $w_{ij} = 1$ by constraint (4.8). If either of them are zero, then $u_{aij} \leq 0$ implied either by (4.9) or (4.10). Hence, $u_{aij} = 0$ by constraint (4.14). Therefore, every u_{aij} takes the value of 0 or 1 in an optimal solution of $\mathcal{P}_r$. □

The bi-objective structure of $\mathcal{P}_r$ allows the decision maker to analyze the trade-off

between the objectives of the model by running extensive tests with varying α, β, B, and p values. Thus, investigating the Pareto-frontier of each setting will provide useful insights before making significant investments on the infrastructure for shared AVs in urban areas. In the next section, we develop a Benders decomposition algorithm to solve $\mathcal{P}_r$.

4.3 Solution Methodology: Benders Decomposition

In this section, we first present the Benders reformulation of $\mathcal{P}_r$ and then introduce the Pareto-optimal cut generation scheme using the Magnanti-Wong problem to improve the computational performance.

4.3.1 The Benders Reformulation

Given the structure of $\mathcal{P}_r$ and considering that u_{aij} is an auxiliary variable, there two types of decisions to make: infrastructure decisions (staging facility locations and AV lane positions) and assignment decisions. This structure may be used to define a master problem that determines the infrastructure decisions, and a subproblem that finds the optimal OD node to staging facility assignments.

Let $\overline{y}_i$ and $\overline{z}_a$ denote the given values for the infrastructure variables. The subproblem ($\mathcal{SP}$) is then defined as:

$$[\mathcal{SP}] \quad \min \ \alpha\Bigg(\sum_{i\in N}\sum_{k\in K}\sum_{j\in N_k} f_k \ell^*_{ij} w_{ij}\Bigg) + \beta\Bigg(\sum_{k\in K}\sum_{i\in N}\sum_{j\in N_k}\sum_{a\in A_{ij}} f_k \ell_a (w_{ij} - u_{aij})\Bigg) \tag{4.15}$$

s.t. (4.3), (4.9)

$$w_{ij} \le \overline{y}_i \qquad i \in N, \quad j \in \mathcal{N} \tag{4.16}$$

$$\sum_{s=r+1}^{|N|} w_{j_s,j} \le 1 - \overline{y}_{j_r} \qquad j \in \mathcal{N}, \quad r = 1, \ldots, |N| - 1 \tag{4.17}$$

$$u_{aij} - w_{ij} \geq \overline{z}_a - 1 \qquad i \in N, \quad j \in \mathcal{N}, \quad a \in A_{ij} \tag{4.18}$$

$$u_{aij} \leq \overline{z}_a \qquad i \in N, \quad j \in \mathcal{N}, \quad a \in A_{ij} \tag{4.19}$$

$$w_{ij}, u_{aij} \geq 0 \qquad i \in N,\ j \in \mathcal{N},\ a \in A_{ij} \tag{4.20}$$

$\mathcal{SP}$ is an easy problem to solve. The optimal values of w_{ij} may be obtained by assigning each OD node $j \in \mathcal{N}$ to its closest open staging facility. These optimal w_{ij} values, and the given $\overline{z}_a$ values may then be used to find the optimal u_{aij} such that $u_{aij} = 1$ only if both the corresponding w_{ij} and $\overline{z}_a$ are 1, and $u_{aij} = 0$ otherwise.

Let ϕ_j, γ_{aij}, μ_{ij}, θ_{jr}, λ_{aij}, and ψ_{aij} be the dual variables associated with constraints (4.3), (4.9), (4.16)-(4.19), respectively. In order to define the dual of $\mathcal{SP}$, we need to define additional notation. Let s_{ij} denote the order of proximity of node $i \in N$ to $j \in \mathcal{N}$, e.g. $s_{ij} = 5$ means that staging facility j is fifth closest candidate staging facility to node i. The dual of $\mathcal{SP}$, denoted as $\mathcal{DSP}$, is then formulated as:

$$[\mathcal{DSP}] \quad \max \sum_{j \in \mathcal{N}} \phi_j + \sum_{i \in N} \sum_{j \in \mathcal{N}} \overline{y}_i \mu_{ij} + \sum_{j \in \mathcal{N}} \sum_{r=1}^{|N|-1} (1 - \overline{y}_{j_r}) \theta_{jr} + \sum_{i \in N} \sum_{j \in \mathcal{N}} \sum_{a \in A_{ij}} (\overline{z}_a - 1) \lambda_{aij} + \overline{z}_a \psi_{aij} \tag{4.21}$$

$$\text{s.t. } \phi_j - \sum_{a \in A_{ij}} (\gamma_{aij} + \lambda_{aij}) + \mu_{ij} + \sum_{r=1}^{s_{ij}-1} \theta_{jr} \leq \sum_{k \in K: j \in N_k} \left(\alpha f_k \ell^*_{ij} + \sum_{a \in A_{ij}} \beta f_k \ell_a \right) \ i \in N, j \in \mathcal{N}, i \neq j \tag{4.22}$$

$$\phi_j - \sum_{a \in A_{ij}} (\gamma_{aij} + \lambda_{aij}) + \mu_{ij} \leq \sum_{k \in K: j \in N_k} \left(\alpha f_k \ell^*_{ij} + \sum_{a \in A_{ij}} \beta f_k \ell_a \right) \qquad i \in N, j \in \mathcal{N}, i = j \tag{4.23}$$

$$\gamma_{aij} + \lambda_{aij} + \psi_{aij} \leq -\beta \sum_{k \in K: j \in N_k} f_k \ell_a \qquad i \in N,\ j \in \mathcal{N},\ a \in A_{ij} \tag{4.24}$$

$$\phi_j \text{ free}, \quad \gamma_{aij}, \mu_{ij}, \theta_{jr}, \psi_{aij} \leq 0, \quad \lambda_{aij} \geq 0 \qquad i \in N,\ j \in \mathcal{N},\ a \in A_{ij},\ r = 1, ..., |N| - 1 \tag{4.25}$$

Let $q(\mathbf{y}, \mathbf{z})$ denote the objective function function value of $\mathcal{SP}$ given that $\mathbf{y} = [y_i]$, $i \in N$ and $\mathbf{z} = [z_a]$, $a \in A$. Then, $\mathcal{P}_r$ can be rewritten as:

$$[\mathcal{P}_r] \quad \min \quad \beta\Bigg(-\sum_{k\in K}\sum_{a\in A_{O_k D_k}} f_k \ell_a z_a\Bigg) + q(\mathbf{y}, \mathbf{z}) \tag{4.26}$$

$$\text{s.t.} \quad (4.2), (4.6)$$

$$y_i, z_a \in \{0, 1\} \qquad i \in N,\ a \in A \tag{4.13}$$

It is trivial to show that $\mathcal{SP}$ is always feasible and bounded when $\alpha, \beta \geq 0$, $f_k \geq 0$ for $k \in K$, $\ell_a \geq 0$ for $a \in A$ and the solution vector $\mathbf{y}$ is not empty. This follows that, $\mathcal{DSP}$ is also always feasible and bounded. Thus, $\mathcal{P}_r$ always has a finite optimal solution when $p > 0$.

Based on $\mathcal{DSP}$ and above deductions, the Benders master problem $\mathcal{MP}$ is:

$$[\mathcal{MP}] \quad \text{minimize} \quad \beta\Bigg(-\sum_{k\in K}\sum_{a\in A_{O_k D_k}} f_k \ell_a z_a\Bigg) + \mathbf{Z} \tag{4.27}$$

$$\text{s.t.} \sum_{i\in N} y_i = p \tag{4.2}$$

$$\sum_{a\in A} c_a z_a \leq B \tag{4.6}$$

$$\mathbf{Z} \geq \sum_{j\in\mathcal{N}} \phi_j^h + \sum_{i\in N}\sum_{j\in\mathcal{N}} y_i \mu_{ij}^h + \sum_{j\in\mathcal{N}}\sum_{r=1}^{|N|-1} (1 - y_{j_r})\theta_{jr}^h + \sum_{i\in N}\sum_{j\in\mathcal{N}}\sum_{a\in A_{ij}} (z_a - 1)\lambda_{aij}^h + z_a \psi_{aij}^h \quad h \in \mathcal{H} \tag{4.28}$$

$$y_i, z_a \in \{0, 1\} \qquad i \in N, a \in A \tag{4.7}$$

where $\mathcal{H}$ denotes the set of extreme points of the polyhedron defined by (4.22)-(4.25). When defined on the complete set of $\mathcal{H}$, $\mathcal{MP}$ is the Benders reformulation of $\mathcal{P}_r$. Since it is impractical to pregenerate all $h \in \mathcal{H}$, a Benders algorithm is used to generate optimality cuts (4.28) as required. It is shown that such an algorithm either finds an optimal solution or yields that the problem is infeasible after a finite number of iterations (Benders, 1962).

This Benders algorithm may be implemented to work on the same branch-and-bound search tree, instead of solving $\mathcal{MP}$ from scratch and creating a new search tree, each time a new optimality cut is appended to $\mathcal{MP}$. We adopt this methodology known as the branch-and-Benders-cut method (Rahmaniani et al., 2017), which have been previously described in Section 2.3 .

The preliminary computational experience with $\mathcal{P}_r$ and the Benders algorithm implemented using the Gurobi solver revealed that the lower-bound convergence, even in smaller instances, may be slow. In order to accelerate the solution process and enhance the performance of the Benders algorithm, we propose identifying and adding non-dominated optimality cuts using a technique proposed by Magnanti and Wong (1981).

4.3.2 The Magnanti-Wong Problem and Pareto-Optimal Benders Cut Generation

When a unique optimal solution to $\mathcal{DSP}$ does not exist, there are numerous cuts (4.28) that may be added to $\mathcal{MP}$. Consider the following term to define the right-hand-side of cut (4.28) for a given $h \in \mathcal{H}$:

$$\mathcal{D}(\phi,\gamma,\mu,\theta,\lambda,\psi,\mathbf{y},\mathbf{z}) = \sum_{j\in\mathcal{N}}\phi_j + \sum_{i\in N}\sum_{j\in\mathcal{N}} y_i\mu_{ij} + \sum_{j\in\mathcal{N}}\sum_{r=1}^{|N|-1}(1-y_{j_r})\theta_{jr} + \sum_{i\in N}\sum_{j\in\mathcal{N}}\sum_{a\in A^*_{ij}}(z_a-1)\lambda_{aij} + z_a\psi_{aij}$$

A cut (4.28) corresponding to $\hat{\phi},\hat{\gamma},\hat{\mu},\hat{\theta},\hat{\lambda},\hat{\psi}$ is said to dominate another cut corresponding to $\tilde{\phi},\tilde{\gamma},\tilde{\mu},\tilde{\theta},\tilde{\lambda},\tilde{\psi}$ if and only if $\mathcal{D}(\hat{\phi},\hat{\gamma},\hat{\mu},\hat{\theta},\hat{\lambda},\hat{\psi},\mathbf{y},\mathbf{z}) \geq \mathcal{D}(\tilde{\phi},\tilde{\gamma},\tilde{\mu},\tilde{\theta},\tilde{\lambda},\tilde{\psi},\mathbf{y},\mathbf{z})$ for all $(\mathbf{y},\mathbf{z}) \in \mathcal{Y}$, with a strict inequality for at least one point; where $\mathcal{Y}$ denote the feasible set of values for the master problem variables (Papadakos, 2008). Consequently, a Benders cut is pareto-optimal if it is not dominated by any other cut.

Magnanti and Wong (1981) proposes solving a new problem, often referred to as the Magnanti-Wong problem, to generate pareto-optimal Benders cuts. This problem uses the notion of core points where $(\mathbf{y}, \mathbf{z})$ is a core point of $\mathcal{Y}$ if it is in the relative interior of the convex hull of set $\mathcal{Y}$. Note that a point is said to belong to the relative interior of a set if there exists a ball centered around that point which is contained entirely within the set. Given $(\mathbf{y}^0, \mathbf{z}^0)$ as a core point of $\mathcal{Y}$, the Magnanti-Wong problem is formulated as follows:

$$[\mathcal{MWP}] \quad \max \ \mathcal{D}(\phi, \gamma, \mu, \theta, \lambda, \psi, \mathbf{y}^0, \mathbf{z}^0) \tag{4.29}$$

$$\text{s.t.} \quad \mathcal{D}(\phi, \gamma, \mu, \theta, \lambda, \psi, \overline{\mathbf{y}}, \overline{\mathbf{z}}) \ = \ q(\overline{\mathbf{y}}, \overline{\mathbf{z}}) \tag{4.30}$$

$$(4.22) - (4.25)$$

Upon solving this problem and obtaining optimal variable values, a pareto-optimal cut (4.28) may be generated. In order to determine a new core point at each Benders algorithm iteration, for each $i \in N$, we set $y_i^0 = 0.85$ when $\overline{y}_i = 1$, whereas when $\overline{y}_i = 0$, we determine an identical-valued y_i^0 so that (4.2) is satisfied. Similarly, we set $z_a^0 = 0.85$ for $\overline{z}_a = 1$, for $a \in A$. In order to calculate z_a^0 values for $\overline{z}_a = 0$, we first calculate the excess budget given z_a^0 values of deployed lanes. We then distribute the excess budget to the remaining z_a^0 that correspond to the non-AV lane arcs, directly proportional to the respective ℓ_a, and subtract ϵ from the resulting value, where ϵ is an infinitesimal constant, so that the $\mathbf{z}^0$ does not make constraint (4.6) binding.

4.4 Computational Experiments

In this section, we present the computational studies we conducted to derive insights from effects of the decision maker's objective preferences, number of staging facilities to locate, and budget on the optimal solutions. We also evaluate the performance of the proposed

Benders algorithm on benchmark transportation networks.

These computational experiments are executed using Gurobi v9.5.0 in the Python API on a computer with Intel i7-9700K 3.60GHz processor and 32GB of RAM. The Benders algorithms are implemented using lazy constraint callbacks of Gurobi and a predetermined time limit of one hour is set for all tests. The *NetworkX* package v2.6.2 (Hagberg et al., 2008) for Python is also utilized in the implementations.

4.4.1 Data Set

We use an urban road network commonly used in transportation research that is obtained from Bar-Gera et al. (2021). This benchmark network includes 24 nodes and 76 arcs, and it is introduced by Leblanc (1975). It is an aggregated road network of the city of Sioux Falls, South Dakota, and it is one of the most widely used test-beds for traffic equilibrium and network design problems. We will refer to this network as *SF24*. The visual representation of *SF24* is depicted in Figure 4.2. We use the network parameters (e.g., arc lengths, free flow travel times) as presented in Bar-Gera et al. (2021).

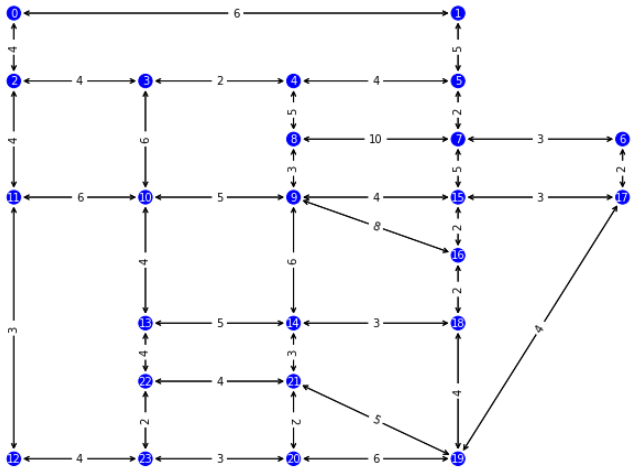


Figure 4.2: Visual representation of the benchmark Sioux Falls (*SF24*) network.

We assume that the cost of AV lane deployment on an arc of *SF24*, $\$c_a$ for $a \in A$, is

directly proportional to its length ℓ_a. To create the *SF24* problem instances, we work with randomly selected 50 OD pairs and set a unit popularity factor f_k for $k \in K$. Throughout the computational experiments, the selection of values used for α and β is made within the interval $(0, 1)$ such that $\alpha + \beta = 1$, and a range of values for p and B are used to perform an extensive analysis on how these parameters may affect the optimal decisions.

4.4.2 Analysis of Optimal Designs

In this section, we the present the results of the computational experiments aiming to derive insights about the effects of p, B, α, and β on optimal solutions and designs.

We generate instances where B ranges between $60 - 120$ with increments of 10, and p ranges between $2 - 5$. For each B and p setting, we use a total of 19 different $\alpha - \beta$ combinations, where α ranges between 0.05 and 0.95 with 0.05 increments and $\beta = 1 - \alpha$. This results in a total of 532 test instances for which we use a randomly generated set of OD pairs K such that the cardinality of set K is 50. The results of these instances are depicted in Figures 4.3–4.6.

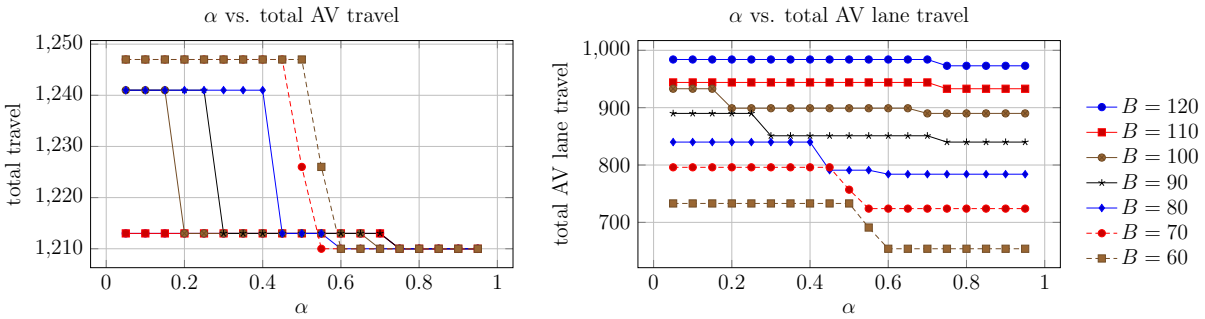


Figure 4.3: *SF24* instances: $p = 2$ solutions.

Since total AV travel is only impacted by the choice of facility locations, two solutions that share the same value for this metric will have the same set of facilities open unless there exists alternative optimal solutions. Hence, the plots on the left of each figure also

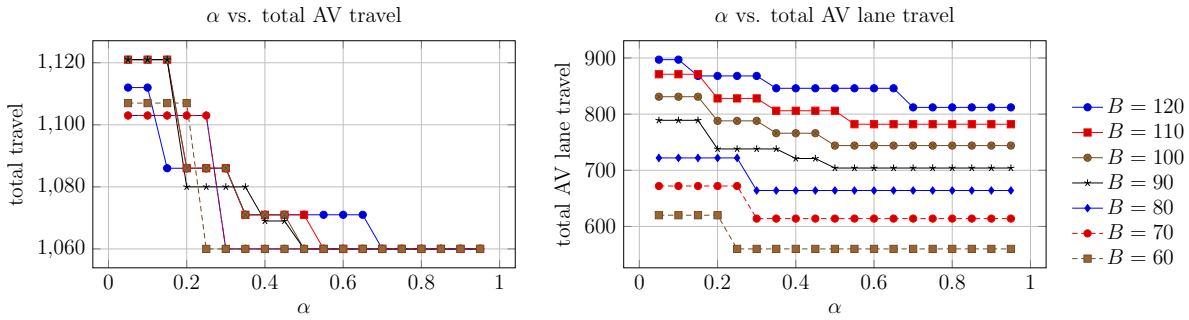


Figure 4.4: *SF24* instances: $p = 3$ solutions.

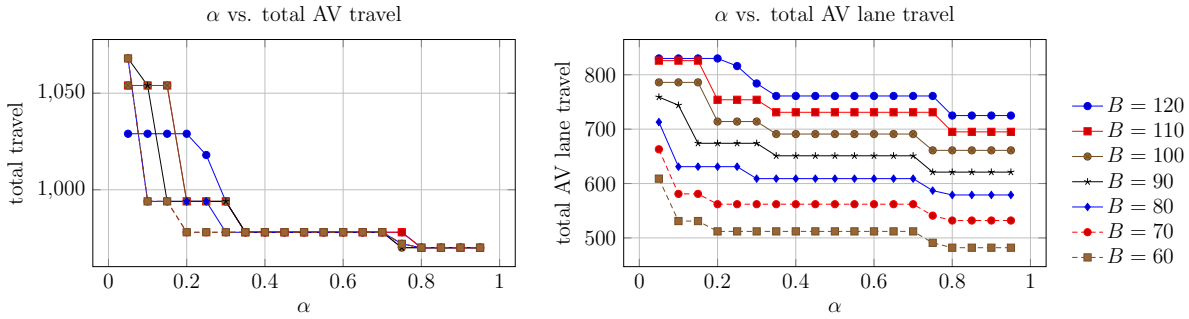


Figure 4.5: *SF24* instances: $p = 4$ solutions.

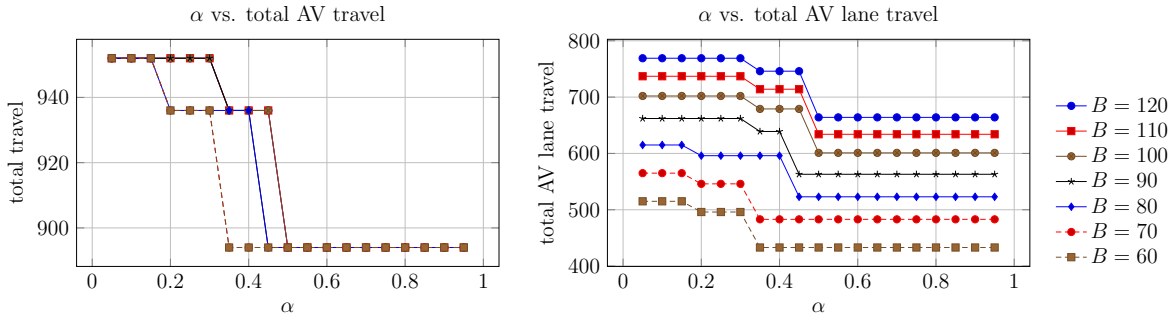


Figure 4.6: *SF24* instances: $p = 5$ solutions.

give an idea about the set of open facilities for changing B and α besides presenting the analysis on how total travel changes with different values of B and α. On the other hand, the plots on the right of each figure show the relationship of α and total AV lane travel which is a concurrent consequence of facility location and AV lane deployment decisions.

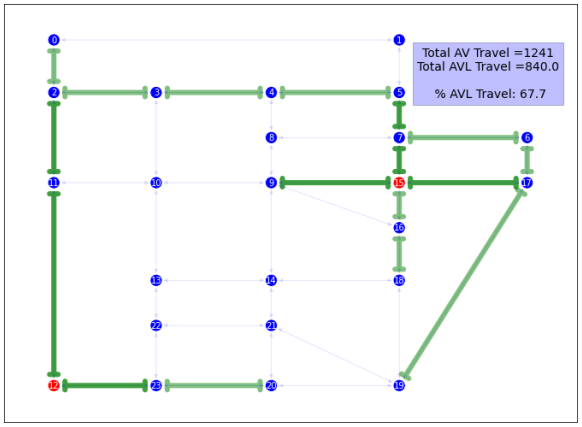
All these figures demonstrate the trade-off between the two objective components: when α is lower (β is higher), total AV lane travel increases at the expense of increasing total travel; and a high α not only decreases the total travel but also the total AV lane travel. Among all the tested p levels, we observe numerous overlaps of the total travel values between different B levels. In addition, this metric tends to converge to the same value for varying B values when α increases. These indicate that different levels of B do not have a significant effect on the facility location decisions, and especially when α is much higher than β, they do not have an effect at all. Naturally, total AV lane travel increases with higher B. However, there are a few solutions, such as $B = 70$ and $B = 80$ instances for $p = 2$ and $\alpha = 0.45$, which may yield higher AV lane travel with slightly lower budget as a result of the trade-off between the objective function components. For $p = 4$ and $p = 5$, the α breakpoints of plateaus regions of the plots are similar. For example, for $p = 5$, both objective function components tend to have a sharp decrease when α is between 0.4 and 0.5, particularly for B ranging between 120 and 80. On the other hand, such a pattern is not observed for lower p solutions, i.e. $p = 2$ and $p = 3$. When $p = 2$, the number of different solutions obtained for the tested $\alpha - \beta$ combinations is observed to be lower than those of higher p values. Total AV lane travel tend to be stay plateaus when α increases, particularly for $B = 120$ and 110. Over 19 tested $\alpha - \beta$ combinations, majority of the instances yield 3-4 different solutions in terms of total AV lane travel for a fixed B level. The maximum such value is observed to be 5 for all $p = 4$ solutions except $B = 100$ and $B = 110$.

Figure 4.7 illustrates three optimal solutions of $B = 80$, $\alpha = 0.3$ instances for $p = 2, 3$,

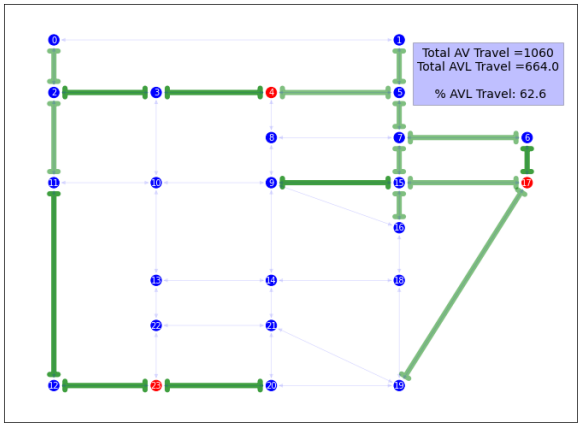
and 4. The arcs that have AV lanes deployed are highlighted as green and dark green arcs imply bidirectional AV lanes between two nodes. The optimal staging facility locations are depicted with red nodes. When $p = 2$, staging facilities at nodes 12 and 15 are open, total AV travel is 1241, and total AV lane travel is 840. When p increases to 3, facilities at nodes 4, 17 and 23 are open, i.e., the optimal facility locations of the $p = 2$ solution are no longer used. Increasing p decreases total AV travel by 14.5%, while AV lane travel also decreases by approximately 21%. When $p = 4$, both of the facilities that are open for $p = 2$ are also open including two additional facilities located at nodes 2 and 5. At this level of p, total AV travel is 7.7% less than that of $p = 3$. Whereas total AV travel decreases by about 8% compared to $p = 3$. An important general observation here is that lower p helps to use the given budget more effectively to utilize AV lanes, i.e., increasing p decreases the percentage of travel happening on AV lanes. This pattern is noticed in all *SF24* solutions.

Figure 4.8 depicts how the selection of α and β affects the optimal solutions that correspond to $p = 2$ and $B = 80$ instances. Figure 4.8a, that is for $\alpha = 0.3$ solution, is equivalent to the one shown in Figure 4.7a. When $\alpha = 0.5$, facility at node 12 is no longer open; instead, node 11 has a facility located. Although the positions of AV lanes are similar, percentage of travel happening on AV lanes decreases from 67.7% to 65.2% while total AV travel decreases from 1241 to 1213. When α is increased to 0.7, facility at node 11 is still used alongside with a new facility at node 17. This change coupled with the differences in the AV lane positions results is offering in a slight decrease in total travel whereas it also decreases the percentage of travel happening on AV lanes.

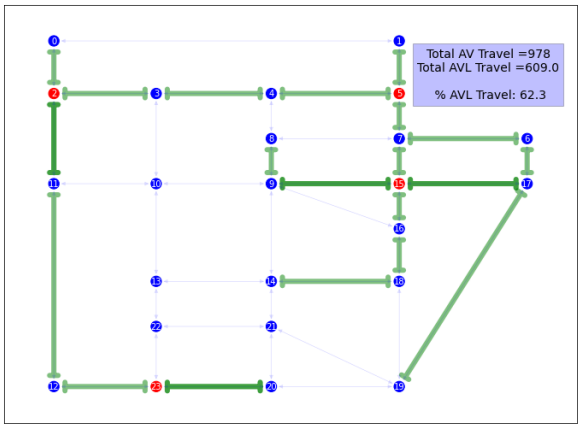
It is also important to point out that the AV lane networks of all solutions presented in Figures 4.7 and 4.8 are connected despite that there is no specific constraint enforcing connectivity. These optimal solutions with connected AV lane networks may particularly exploit the potential connectivity benefits of the AV lane infrastructure.



(a) $p = 2$



(b) $p = 3$

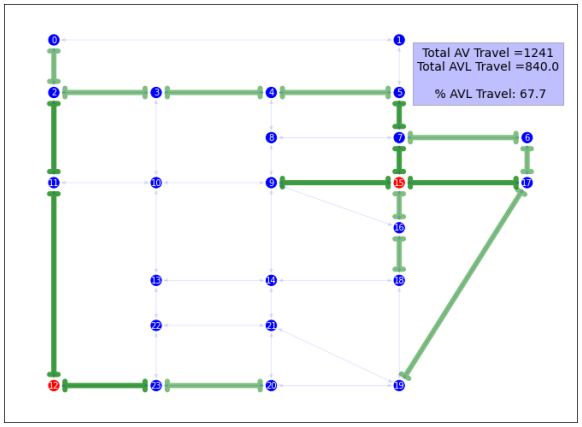


(c) $p = 4$

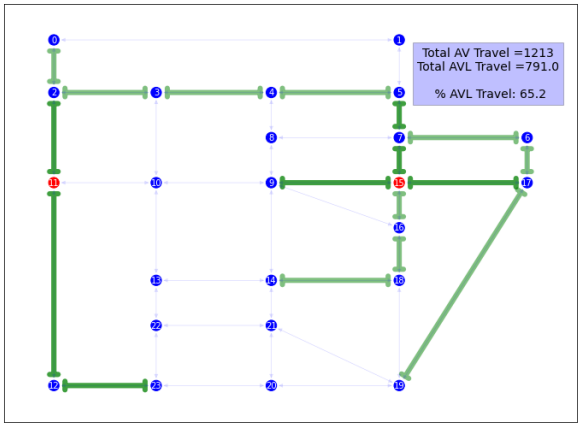
Figure 4.7: Visual representations of $\alpha = 0.3, \beta = 0.7$ solutions given $B = 80$ for *SF24*.

4.4.3 The Performance of the Benders Algorithm

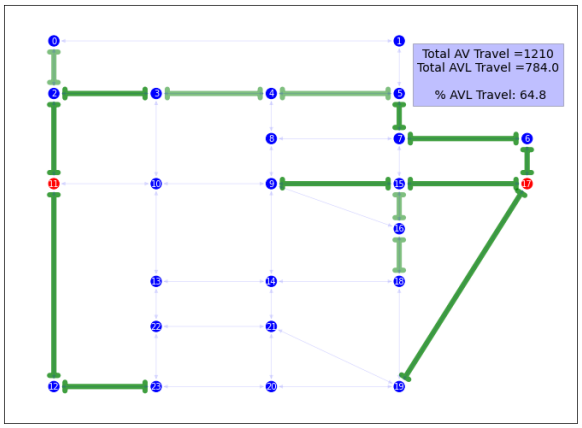
In this section, we evaluate the computational performance of the pareto-cut generation scheme of the Benders decomposition algorithm. Particularly, we compare the performance of the classical Benders decomposition (without pareto optimal cuts) with the version that adds pareto-optimal cuts over select *SF24* instances. In order to understand the behavior better, the instance selection features low-mid-high combinations of α values (0.3, 0.5, 0.7) and three budget levels B (120, 90, 60). Each combination is tested over p values ranging between $2-5$. The results are depicted in Table 4.1 which provides the information on the number of cuts added and solution times with both approaches. When an optimal solution



(a) $\alpha = 0.3,\ \beta = 0.7$



(b) $\alpha = 0.5,\ \beta = 0.5$



(c) $\alpha = 0.7,\ \beta = 0.3$

Figure 4.8: Visual representations of $p = 2$ solutions given $B = 80$ for *SF24*.

is not obtained within the 1-hour time limit, the remaining MIP gap value is also provided.

These results reveal that the pareto-cut generation scheme is not only trivially better than the classical Benders decomposition in terms of computational time, but also it is essential to be able to solve the problem. Among 36 test instances, regular Benders failed to find the optimal solution within the 1-hour time limit in 17 of them, whereas the pareto-cut Benders algorithm is able to solve all instances to optimality within an average solution time of 82 seconds. The number of cuts added by the pareto-cut Benders algorithm is always less than those of the regular Benders.

For the Benders algorithm, it is observed that lower α instances are significantly harder to solve than high α ones. Similarly, low budget tends to make the instances more challenging to solve which leads to higher computation times or higher MIP gap percentages. Increasing p also has a similar effect on computational performance, i.e., higher p instances tend to require more cuts and higher solution times.

On the other hand, as expected, Gurobi's built-in branch-and-bound solution methodology is able to outperform the Benders algorithm with pareto-cuts for these small instances. The current testing is preliminary and extensive testing on larger and more realistic in-

Table 4.1: Computational performance of the Benders algorithm with and without pareto-optimal cuts for *SF24*.

			Benders w/o Pareto Cuts			**Benders with Pareto-Cuts**		**Gurobi**
p	α	B	**# Cuts**	**Time(s)**	**Gap(%)**	**# Cuts**	**Time(s)**	**Time(s)**
2	0.3	120	2193	422.5	-	410	19.9	5.0
		90	4286	3600	2.41	407	18.9	13.1
		60	6302	3600	12.26	666	32.2	8.5
	0.5	120	1115	101.3	-	188	9.7	1.6
		90	2708	556.9	-	214	10.8	5.4
		60	3020	3600	1.58	329	16.6	9.3
	0.7	120	354	29.5	-	128	5.6	4.0
		90	663	54.8	-	110	5.3	1.9
		60	1275	140.5	-	124	6.3	3.5
3	0.3	120	6121	3600	6.85	1096	54.6	6.6
		90	8424	3600	20.27	1623	86.2	13.8
		60	8948	3600	36.29	1400	80.1	7.1
	0.5	120	2485	546.2	-	526	24.2	4.4
		90	4394	3600	1.17	631	29.8	3.6
		60	5444	3600	2.63	679	33.9	3.9
	0.7	120	550	51.4	-	208	9.9	3.3
		90	700	96.1	-	211	9.7	5.8
		60	1502	237.1	-	190	10.1	2.7
4	0.3	120	10148	3600	17.67	2616	166.8	5.1
		90	7950	3600	19.25	2508	153.5	4.5
		60	7479	3600	33.64	4997	395.4	5.8
	0.5	120	2336	494.6	-	877	42.4	3.5
		90	3493	3600	1.00	901	44.6	2.4
		60	5113	3600	4.19	1225	63.9	2.3
	0.7	120	601	51.87	-	228	10.6	4.8
		90	923	98.95	-	250	11.9	4.3
		60	1288	185.85	-	309	14.8	3.6
5	0.3	120	6030	3600	7.90	2926	195.6	2.0
		90	8971	3600	26.27	4933	408.3	2.8
		60	7379	3600	31.21	5901	606.3	5.2
	0.5	120	2259	318.1	-	527	25.93	3.0
		90	4115	2005.7	-	941	48.0	2.1
		60	2783	3600	4.04	1281	69.8	1.4
	0.7	120	145	185	-	141	6.9	5.0
		90	177	249	-	165	8.2	5.1
		60	274	319	-	266	13.4	4.9

stances is required to assess the performance of Benders algorithm against commercial solvers.

4.5 Conclusion

In this chapter, we study optimizing the deployment of AV lanes and locations of staging facilities and propose the first study in the literature to optimize the spatial distribution of AV lanes simultaneously with staging facilities. We present a bi-objective optimization model that minimizes total travel distance and total AV travel not occurring on AV lanes with respect to a given AV lane deployment budget and a number of staging facilities to locate. We formulate this problem using a mixed-integer optimization model and prove an integrality property concerning the allocation decision variables. This property allows us to use a Benders decomposition solution methodology, which is developed together with a non-dominated cut generation scheme.

We evaluate the effects of AV lane budget, staging facility count, and objective prioritization of decision makers on optimal solutions using the benchmark Sioux Falls network, and derive managerial insights analyzing the trade-offs between AV lane deployment and staging facility location decisions. When minimizing total AV travel not occurring on AV lanes objective is prioritized, travel on AV lanes is observed to increase at the expense of higher total travel. The numerical analysis reveals that AV lane travel increases with a higher budget, whereas varying budget levels are not observed to have a significant effect on the facility location decisions.

By evaluating optimal designs, we analyze how the selection of α and β values, i.e., objective prioritization parameters, affect the optimal solutions. It is observed that the selection of optimal staging facility locations are sensitive to changes in these values whereas

the AV lane deployment decisions are less sensitive. We analyze how varying p values change the optimal designs and observe that total AV lane travel tends to be higher when fewer staging facilities are located while the budget, α, and β are fixed.

Infrastructure design for shared autonomous travel has a lot of research opportunities and this work opens the door for a variety of future research directions. For example, an extension that relaxes the shortest path travel assumption may lead to using AV lanes more effectively at the expense of having detours. Another research direction may consider capacitated staging facilities, charging times, and congestion, where capacity decisions under realistic dynamic demand and routing scenarios can be evaluated using simulation.

Chapter 5

Conclusions and Future Research Directions

The emerging disruptive technologies in transportation, such as EVs and AVs, have been aiming to promote environmental sustainability and provide a cost-effective, cleaner and safer transportation future. In this thesis, analytical models and methods are developed to support the proliferation of these technologies by strategically making infrastructure deployment decisions.

Chapter 2 presents a full cover modeling approach to design charging station infrastructure that enables long-distance EV travel on a transportation network. We incorporated a novel objective function, which determines the optimal OD routes so that the total en route recharging is minimal for each OD trip. We introduced an optimal recharging policy referred to as charge-just-enough which helps compute the total recharging required to complete an OD trip without tracking the battery level on every node of a route. The results of computational experiments highlight the effectiveness of the new modeling approach and bring out its strengths over the existing models in the literature. The solutions of this

new model dominate the max-cover or set-cover solutions in terms of routing related measures. Since objective of the proposed model can effectively distinguish alternative optimal solutions with 100% coverage in terms of these routing-related measures, we observe that opening additional stations offer substantial improvements over these measures. A Benders decomposition algorithm is developed to solve the problem on large-scale transportation networks. Within this algorithm, we developed an efficient subroutine to construct the optimal dual solution to generate non-dominated optimality cuts when the Benders subproblem is feasible; and derive two-sided feasibility cuts based on the infeasible subgraph structures when the Benders subproblem is infeasible. This solution methodology is shown to promote computational performance significantly. Two main limitations of this work may be listed as assuming uncapacitated facilities and neglecting congestion. These aspects may become particularly important when the adoption of EVs increase at a rate to saturate existing capacities of charging stations.

Next, an extension of this problem is presented in Chapter 3. A new framework that additionally considers congestion and sizing decisions at charging stations is proposed. Particularly, a bilevel optimization model, that allows detours from shortest paths, is developed to determine the strategic locations and capacities of DC fast-charging stations under stochastic long-distance EV flows. The modeling approach considers the route choice behavior and cooperative/uncooperative charging station selection response of EV users. For the case where DC fast-charging stations are characterized as M/M/c queuing systems, the MILP equivalent of a chance-constrained stochastic bilevel optimization formulation is presented. In order to solve this formulation, a decomposition based algorithm is developed which can address cooperative or uncooperative responses. The solution quality and the computational performance of the algorithm are validated on smaller-sized benchmark networks, and it is shown that the algorithm finds optimal solutions over all the test instances. The results of the computational experiments show that the number of located

charging stations is not sensitive to the changes in the service level requirements, and the existing charging stations start to require additional capacity when more stringent service level thresholds are used. The solutions with higher deviation tolerance are observed to require a less total cost of infrastructure deployment, and these instances are also observed to be not prone to generate much difference between cooperative and uncooperative EV user responses. This leads to robust solutions that can protect the decision-maker against worst-case scenarios. The main limitation of this work is the underlying assumption regarding homogeneous vehicles that share a similar range and energy consumption. These aspects of the problem may become increasingly important if there are significant feature differences for heterogeneous EVs, e.g., range and charging technology compatibility.

Chapter 4 focuses on the domain of AVs and enabling shared autonomous transportation in urban areas by strategically deploying infrastructure. Defined in this context, the novel staging facility location and AV lane deployment problem is introduced. This problem is modeled mathematically to find the optimal locations of staging facilities utilizing a bi-objective model that minimizes total travel distance and the total AV travel not occurring on AV lanes with respect to a given AV lane deployment budget and a number of staging facilities to locate. A Benders decomposition algorithm with Pareto-optimal cuts is developed and computational analyses are conducted to evaluate the trade-offs with optimal solutions on benchmark instances. This analysis demonstrates the trade-off between the two objective components: when the first objective is prioritized, total travel and total AV lane travel both decreases. On the other hand, when the latter objective is deemed more important by the decision maker, total AV lane travel increases at the expense of increasing total travel. It is further observed that total AV lane travel increases with the budget; however, different budget levels do not have significant effects on the facility location decisions. Finally, there do not exist many Pareto-optimal solutions for a given test instance, so the problem is observed to be insensitive to the small changes in the weights

of the two objective function components. The main limitation of this work is regarding the assumption that does not allow detours from shortest paths and assuming that staging facilities are uncapacitated.

Although we discussed several potential extensions and future research directions to these works in their corresponding chapters, there are many other exciting research opportunities in the broader area of infrastructure deployment for disruptive technologies in transportation. For instance, it is known that EVs and autonomous travel create a powerful synergy, and this has a high potential to be the future of freight transportation. In this context, enabling sustainable and centrally dispatchable long-haul autonomous freight transportation is dependent on the synergy of different types of infrastructural decisions such as charging stations, wireless charging lanes, AV lanes, platooning hubs, and cross-docking facilities. Hence, determining a centralized deployment plan for these components would be a huge step in supporting the shift towards autonomous and electric mobility.

Extension of the work presented in Chapter 3 may consider heterogeneous EVs, which would be another promising future research avenue. In reality, many EV models exist and they may have significantly different ranges and charging technology compatibilities. Besides vehicle flow, the congestion faced at charging stations would be dependent on these key features, hence, it could be important to take them into account while making sizing decisions. However, modeling this complex aspect of the problem mathematically would be a challenge. Each feature that changes the charging time of an EV necessitates the use of queuing systems with multiple customer classes. Even in the case that arrival times are assumed to be Poisson distributed and service times are assumed as exponentially distributed for each class, the queuing network can no longer be represented as an open multi-server Jackson network. In order to tackle such a complex problem setting, other aspects of the problem may need to be simplified, such as not allowing detours from the shortest paths. As a result, such prospective studies would require completely new

modeling approaches, and, consequently, novel solution methods.

Another interesting direction is exploiting shared autonomous travel with EVs to promote the renewable energy resilience of transportation systems. Some micro-grids may consist of locally generated renewable energy (e.g., hydro, wind, or solar) that could effectively be repositioned using the battery packs of shared autonomous EVs to promote renewable energy integration in the energy mix. Under such a problem setting, locating facilities that support renewable energy sharing as well as managing cost-effective charging and discharging operations of AV fleets could definitely be an interesting opportunity of research.

References

Alcorn, L. and Kockelman, K. (2021). Automated vehicles and vehicles of the future. In *The Routledge Handbook of Public Transport*, pages 535–550. Routledge.

Alumur, S. A. and Bektaş, T. (2019). Green location problems. In *Location Science*, pages 591–610. Springer.

Anjos, M. F., Gendron, B., and Joyce-Moniz, M. (2020). Increasing electric vehicle adoption through the optimal deployment of fast-charging stations for local and long-distance travel. *European Journal of Operational Research*, 285(1):263–278.

Arslan, O. and Karaşan, O. E. (2016). A Benders decomposition approach for the charging station location problem with plug-in hybrid electric vehicles. *Transportation Research Part B: Methodological*, 93:670–695.

Arslan, O., Karaşan, O. E., Mahjoub, A. R., and Yaman, H. (2019). A branch-and-cut algorithm for the alternative fuel refueling station location problem with routing. *Transportation Science*, 53(4):1107–1125.

Arslan, O., Yıldız, B., and Karaşan, O. E. (2014). Impacts of battery characteristics, driver preferences and road network features on travel costs of a plug-in hybrid electric vehicle for long-distance trips. *Energy Policy*, 74:168–178.

Bagloee, S. A., Tavana, M., Asadi, M., and Oliver, T. (2016). Autonomous vehicles: challenges, opportunities, and future implications for transportation policies. *Journal of Modern Transportation*, 24(4):284–303.

Bahrami, S. and Roorda, M. (2022). Autonomous vehicle parking policies: A case study of the City of Toronto. *Transportation Research Part A: Policy and Practice*, 155:283–296.

Bahrami, S., Vignon, D., Yin, Y., and Laberteaux, K. (2021). Parking management of automated vehicles in downtown areas. *Transportation Research Part C: Emerging Technologies*, 126:103001.

Bansal, P. and Kockelman, K. M. (2017). Forecasting Americans' long-term adoption of connected and autonomous vehicle technologies. *Transportation Research Part A: Policy and Practice*, 95:49–63.

Bar-Gera, H., Stabler, B., and Sall, E. (2021). Transportation networks for research core team. https://github.com/bstabler/TransportationNetworks, accessed 2021-11-10.

Bard, J. F. (2013). *Practical bilevel optimization: algorithms and applications*, volume 30. Springer Science & Business Media.

Benders, J. F. (1962). Partitioning procedures for solving mixed-variables programming problems. *Numerische mathematik*, 4(1):238–252.

Bose, S. K. (2013). *An introduction to queueing systems*. Springer Science & Business Media.

Boujelben, M. K. and Gicquel, C. (2019). Efficient solution approaches for locating electric vehicle fast charging stations under driving range uncertainty. *Computers & Operations Research*, 109:288–299.

Brownell, C. and Kornhauser, A. (2014). A driverless alternative: fleet size and cost requirements for a statewide autonomous taxi network in New Jersey. *Transportation Research Record*, 2416(1):73–81.

Burke, P. J. (1956). The output of a queuing system. *Operations Research*, 4(6):699–704.

Capar, I. and Kuby, M. (2012). An efficient formulation of the flow refueling location model for alternative-fuel stations. *IIE Transactions*, 44(8):622–636.

Capar, I., Kuby, M., Leon, V. J., and Tsai, Y.-J. (2013). An arc cover–path-cover formulation and strategic analysis of alternative-fuel station locations. *European Journal of Operational Research*, 227(1):142–151.

Capgemini Invent Global (2018). Electric cars: At the tipping point? Technical report.

Chen, Z., He, F., Yin, Y., and Du, Y. (2017). Optimal design of autonomous vehicle zones in transportation networks. *Transportation Research Part B: Methodological*, 99:44–61.

Chen, Z., He, F., Zhang, L., and Yin, Y. (2016). Optimal deployment of autonomous vehicle lanes with endogenous market penetration. *Transportation Research Part C: Emerging Technologies*, 72:143–156.

Chromy, E., Misuth, T., and Weber, A. (2012). Application of Erlang formulae in next generation networks. *International Journal of Computer Network and Information Security*, 4(1):59.

de Vries, H. and Duijzer, E. (2017). Incorporating driving range variability in network design for refueling facilities. *Omega*, 69:102–114.

Duvall, T., Hannon, E., Katseff, J., Safran, B., and Wallace, T. (2019). A new look at autonomous-vehicle infrastructure. *McKinsey & Company: Washington, DC, USA*.

Electrek (2019). Tesla starts limiting charge to 80% at busy superchargers to reduce wait times. https://electrek.co/2019/05/24/tesla-limiting-supercharger-busy/, accessed 2021-06-10.

Espejo, I., Marín, A., and Rodríguez-Chía, A. (2012). Closest assignment constraints in discrete location problems. *European Journal of Operational Research*, 219:49–58.

Fagnant, D. J. and Kockelman, K. (2015). Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations. *Transportation Research Part A: Policy and Practice*, 77:167–181.

Ghiasi, A., Hussain, O., Qian, Z. S., and Li, X. (2017). A mixed traffic capacity analysis and lane management model for connected automated vehicles: A markov chain method. *Transportation Research Part B: Methodological*, 106:266–292.

Ghouila-Houri, A. (1962). Caractérisation des matrices totalement unimodulaires. *Comptes Redus Hebdomadaires des Séances de l'Académie des Sciences (Paris)*, 254:1192–1194.

Golbabaei, F., Yigitcanlar, T., and Bunker, J. (2021). The role of shared autonomous vehicle systems in delivering smart urban mobility: A systematic review of the literature. *International Journal of Sustainable Transportation*, 15(10):731–748.

Goodchild, M. F. and Noronha, V. T. (1987). Location-allocation and impulsive shopping: the case of gasoline retailing. In *Spatial Analysis and Location-Allocation Models*, pages 121–136. van Nostrand Reinhold New York.

Göpfert, P. and Bock, S. (2019). A branch&cut approach to recharging and refueling infrastructure planning. *European Journal of Operational Research*, 279(3):808–823.

Guo, F., Yang, J., and Lu, J. (2018). The battery charging station location problem: Impact of users' range anxiety and distance convenience. *Transportation Research Part E: Logistics and Transportation Review*, 114:1–18.

Hagberg, A., Swart, P., and S Chult, D. (2008). Exploring network structure, dynamics, and function using NetworkX. Technical report, Los Alamos National Lab.(LANL), Los Alamos, NM (United States).

He, J., Yang, H., Tang, T.-Q., and Huang, H.-J. (2018). An optimal charging station location model with the consideration of electric vehicle's driving range. *Transportation Research Part C: Emerging Technologies*, 86:641–654.

Hong, S. and Kuby, M. (2016). A threshold covering flow-based location model to build a critical mass of alternative-fuel stations. *Journal of Transport Geography*, 56:128–137.

Hooker, J. N. and Ottosson, G. (2003). Logic-based Benders decomposition. *Mathematical Programming*, 96(1):33–60.

Hosseini, M. and MirHassani, S. (2015). Refueling station location problem under uncertainty. *Transportation Research Part E: Logistics and Transportation Review*, 84:101–116.

Hosseini, M., MirHassani, S. A., and Hooshmand, F. (2017). Deviation-flow refueling location problem with capacitated facilities: Model and algorithm. *Transportation Research Part D: Transport and Environment*, 54:269–281.

Huang, Y. and Kockelman, K. M. (2020). Electric vehicle charging station locations: Elastic demand, station congestion, and network equilibrium. *Transportation Research Part D: Transport and Environment*, 78:102179.

Huang, Y., Li, S., and Qian, Z. S. (2015). Optimal deployment of alternative fueling stations on transportation networks considering deviation paths. *Networks and Spatial Economics*, 15(1):183–204.

IEA (2021). *Global EV Outlook 2021: Accelerating Ambitions Despite the Pandemic*.

IEA (2021). Improving the sustainability of passenger and freight transport. *Tracking Transport 2020*. https://www.iea.org/topics/transport, accessed 2022-02-14.

Jiang, Y., Zhang, Y., Zhang, C., and Fan, J. (2012). Capacitated deviation-flow fueling location model for sitting battery charging stations. In *CICTP 2012: Multimodal Transportation Systems—Convenient, Safe, Cost-Effective, Efficient*, pages 2771–2778.

Jing, W., An, K., Ramezani, M., and Kim, I. (2017). Location design of electric vehicle charging facilities: A path-distance constrained stochastic user equilibrium approach. *Journal of Advanced Transportation*, 2017.

Jones, E. C. and Leibowicz, B. D. (2019). Contributions of shared autonomous vehicles to climate change mitigation. *Transportation Research Part D: Transport and Environment*, 72:279–298.

Kim, J.-G. and Kuby, M. (2012). The deviation-flow refueling location model for optimizing a network of refueling stations. *International Journal of Hydrogen Energy*, 37(6):5406–5420.

Kim, J.-G. and Kuby, M. (2013). A network transformation heuristic approach for the deviation flow refueling location model. *Computers & Operations Research*, 40(4):1122–1131.

Kınay, Ö. B., Gzara, F., and Alumur, S. A. (2021). Full cover charging station location problem with routing. *Transportation Research Part B: Methodological*, 144:1–22.

Kınay, Ö. B., Gzara, F., and Alumur, S. A. (2022). Charging station location and sizing for electric vehicles considering user behavior and congestion. *Transportation Science*, Under Review.

Ko, J., Gim, T.-H. T., and Guensler, R. (2017). Locating refuelling stations for alternative fuel vehicles: a review on models and applications. *Transport Reviews*, 37(5):551–570.

Kuby, M. and Lim, S. (2005). The flow-refueling location problem for alternative-fuel vehicles. *Socio-Economic Planning Sciences*, 39(2):125–145.

Kuby, M. and Lim, S. (2007). Location of alternative-fuel stations using the flow-refueling location model and dispersion of candidate sites on arcs. *Networks and Spatial Economics*, 7(2):129–152.

Kuby, M., Lines, L., Schultz, R., Xie, Z., Kim, J.-G., and Lim, S. (2009). Optimization of hydrogen stations in Florida using the flow-refueling location model. *International Journal of Hydrogen Energy*, 34(15):6045–6064.

Kumakoshi, Y., Hanabusa, H., and Oguchi, T. (2021). Impacts of shared autonomous vehicles: Tradeoff between parking demand reduction and congestion increase. *Transportation Research Interdisciplinary Perspectives*, 12:100482.

Labbé, M. and Marcotte, P. (2021). Bilevel network design. In *Network Design with Applications to Transportation and Logistics*, pages 255–281. Springer.

Leblanc, L. J. (1975). An algorithm for the discrete network design problem. *Transportation Science*, 9(3):183–199.

Lee, C. and Han, J. (2017). Benders-and-price approach for electric vehicle charging station location problem under probabilistic travel range. *Transportation Research Part B: Methodological*, 106:130–152.

Leitner, M., Ljubić, I., Riedler, M., and Ruthmair, M. (2019). Exact approaches for network design problems with relays. *INFORMS Journal on Computing*, 31(1):171–192.

Levin, M. W., Wong, E., Nault-Maurer, B., and Khani, A. (2020). Parking infrastructure design for repositioning autonomous vehicles. *Transportation Research Part C: Emerging Technologies*, 120:102838.

Li, S. and Huang, Y. (2014). Heuristic approaches for the flow-based set covering problem with deviation paths. *Transportation Research Part E: Logistics and Transportation Review*, 72:144–158.

Li, Y., Long, J., and Yu, M. (2021). A time-dependent shared autonomous vehicle system design problem. *Transportation Research Part C: Emerging Technologies*, 124:102956.

Lim, S. and Kuby, M. (2010). Heuristic algorithms for siting alternative-fuel stations using the flow-refueling location model. *European Journal of Operational Research*, 204(1):51–61.

Lin, C.-C. and Lin, C.-C. (2018). The p-center flow-refueling facility location problem. *Transportation Research Part B: Methodological*, 118:124–142.

Liu, W. (2018). An equilibrium analysis of commuter parking in the era of autonomous vehicles. *Transportation Research Part C: Emerging Technologies*, 92:191–207.

Liu, Z., Chen, Z., He, Y., and Song, Z. (2021). Network user equilibrium problems with infrastructure-enabled autonomy. *Transportation Research Part B: Methodological*, 154:207–241.

Liu, Z. and Song, Z. (2019). Strategic planning of dedicated autonomous vehicle lanes and autonomous vehicle/toll lanes in transportation networks. *Transportation Research Part C: Emerging Technologies*, 106:381–403.

Magnanti, T. L. and Wong, R. T. (1981). Accelerating Benders decomposition: Algorithmic enhancement and model selection criteria. *Operations Research*, 29(3):464–484.

Makhlouf, W., Kchaou-Boujelben, M., and Gicquel, C. (2019). A bi-level programming approach to locate capacitated electric vehicle charging stations. In *2019 6th International Conference on Control, Decision and Information Technologies (CoDIT)*, pages 133–138. IEEE.

MirHassani, S. and Ebrazi, R. (2013). A flexible reformulation of the refueling station location problem. *Transportation Science*, 47(4):617–628.

Nicholas, M. (2019). Estimating electric vehicle charging infrastructure costs across major us metropolitan areas. *URL: https://theicct.org/sites/default/files/publications/ICCT_EV_Charging_Cost_20190813. pdf*.

Nourinejad, M., Bahrami, S., and Roorda, M. J. (2018). Designing parking facilities for autonomous vehicles. *Transportation Research Part B: Methodological*, 109:110–127.

Okeke, O. B. (2020). The impacts of shared autonomous vehicles on car parking space. *Case Studies on Transport Policy*, 8(4):1307–1318.

Papadakos, N. (2008). Practical enhancements to the Magnanti–Wong method. *Operations Research Letters*, 36(4):444–449.

Pelletier, S., Jabali, O., Laporte, G., and Veneroni, M. (2017). Battery degradation and behaviour for electric vehicles: Review and numerical analyses of several models. *Transportation Research Part B: Methodological*, 103:158–187.

Rahmaniani, R., Crainic, T. G., Gendreau, M., and Rei, W. (2017). The Benders decomposition algorithm: A literature review. *European Journal of Operational Research*, 259(3):801–817.

Schiffer, M., Schneider, M., Walther, G., and Laporte, G. (2019). Vehicle routing and location routing with intermediate stops: A review. *Transportation Science*, 53(2):319–343.

Schiffer, M. and Walther, G. (2017). The electric location routing problem with time windows and partial recharging. *European Journal of Operational Research*, 260(3):995–1013.

Shaheen, S. and Cohen, A. (2019). Shared ride services in North America: definitions, impacts, and the future of pooling. *Transport Reviews*, 39(4):427–442.

Shen, Z.-J. M., Feng, B., Mao, C., and Ran, L. (2019). Optimization models for electric vehicle service operations: A literature review. *Transportation Research Part B: Methodological*, 128:462–477.

Simchi-Levi, D. and Berman, O. (1988). A heuristic algorithm for the traveling salesman location problem on networks. *Operations Research*, 36(3):478–484.

Sinha, A., Malo, P., and Deb, K. (2017). A review on bilevel optimization: from classical to evolutionary approaches and applications. *IEEE Transactions on Evolutionary Computation*, 22(2):276–295.

The Climate Group (2021a). EV100 initiative. https://www.theclimategroup.org/ev100, accessed 2021-09-23.

The Climate Group (2021b). EV100 progress and insights report 2021. https://www.theclimategroup.org/media/7941/download, accessed 2021-09-23.

The White House (2021). Executive order on strengthening American leadership in clean cars and trucks. https://www.whitehouse.gov/briefing-room/presidential-actions/2021/08/05/

executive-order-on-strengthening-american-leadership-in-clean-cars-and-trucks/, accessed 2021-09-24.

Toregas, C., Swain, R., ReVelle, C., and Bergman, L. (1971). The location of emergency service facilities. *Operations Research*, 19(6):1363–1373.

Tran, C. Q., Ngoduy, D., Keyvan-Ekbatani, M., and Watling, D. (2021). A user equilibrium-based fast-charging location model considering heterogeneous vehicles in urban networks. *Transportmetrica A: Transport Science*, 17(4):439–461.

Tran, T. H., Nagy, G., Nguyen, T. B. T., and Wassan, N. A. (2018). An efficient heuristic algorithm for the alternative-fuel station location problem. *European Journal of Operational Research*, 269(1):159–170.

Upchurch, C., Kuby, M., and Lim, S. (2009). A model for location of capacitated alternative-fuel stations. *Geographical Analysis*, 41(1):85–106.

Van Brummelen, J., O'Brien, M., Gruyer, D., and Najjaran, H. (2018). Autonomous vehicle perception: The technology of today and tomorrow. *Transportation Research Part C: Emerging Technologies*, 89:384–406.

Wagner, J. L. and Falkson, L. M. (1975). The optimal nodal location of public facilities with price-sensitive demand. *Geographical Analysis*, 7:69–83.

Wang, J., Lu, L., Peeta, S., and He, Z. (2021a). Optimal toll design problems under mixed traffic flow of human-driven vehicles and connected and autonomous vehicles. *Transportation Research Part C: Emerging Technologies*, 125:102952.

Wang, S., Levin, M. W., and Caverly, R. J. (2021b). Optimal parking management of connected autonomous vehicles: A control-theoretic approach. *Transportation Research Part C: Emerging Technologies*, 124:102924.

Wang, Y.-W. and Lin, C.-C. (2009). Locating road-vehicle refueling stations. *Transportation Research Part E: Logistics and Transportation Review*, 45(5):821–829.

Wang, Y.-W. and Wang, C.-R. (2010). Locating passenger vehicle refueling stations. *Transportation Research Part E: Logistics and Transportation Review*, 46(5):791–801.

Wu, W., Zhang, F., Liu, W., and Lodewijks, G. (2020). Modelling the traffic in a mixed network with autonomous-driving expressways and non-autonomous local streets. *Transportation Research Part E: Logistics and Transportation Review*, 134:101855.

Xie, F. and Lin, Z. (2021). Integrated US nationwide corridor charging infrastructure planning for mass electrification of inter-city trips. *Applied Energy*, 298:117142.

Xie, F., Liu, C., Li, S., Lin, Z., and Huang, Y. (2018). Long-term strategic planning of inter-city fast charging infrastructure for battery electric vehicles. *Transportation Research Part E: Logistics and Transportation Review*, 109:261–276.

Yıldız, B., Arslan, O., and Karaşan, O. E. (2016). A branch and price approach for routing and refueling station location model. *European Journal of Operational Research*, 248(3):815–826.

Yıldız, B. and Karaşan, O. E. (2017). Regenerator location problem in flexible optical networks. *Operations Research*, 65(3):595–620.

Yıldız, B., Karaşan, O. E., and Yaman, H. (2018). Branch-and-price approaches for the network design problem with relays. *Computers & Operations Research*, 92:155–169.

Yıldız, B., Olcaytu, E., and Şen, A. (2019). The urban recharging infrastructure design problem with stochastic demands and capacitated charging stations. *Transportation Research Part B: Methodological*, 119:22–44.

Zeng, B. (2020). A practical scheme to compute the pessimistic bilevel optimization problem. *INFORMS Journal on Computing*, 32(4):1128–1142.

Zhang, A., Kang, J. E., and Kwon, C. (2017). Incorporating demand dynamics in multi-period capacitated fast-charging location planning for electric vehicles. *Transportation Research Part B: Methodological*, 103:5–29.

Zhang, W., Guhathakurta, S., Fang, J., and Zhang, G. (2015). Exploring the impact of shared autonomous vehicles on urban parking demand: An agent-based simulation approach. *Sustainable Cities and Society*, 19:34–45.

Zhang, X., Liu, W., Waller, S. T., and Yin, Y. (2019). Modelling and managing the integrated morning-evening commuting and parking patterns under the fully autonomous vehicle environment. *Transportation Research Part B: Methodological*, 128:380–407.

Zheng, H., He, X., Li, Y., and Peeta, S. (2017). Traffic equilibrium and charging facility locations for electric vehicles. *Networks and Spatial Economics*, 17(2):435–457.

Appendix A

Appendix of Chapter 2

A.1 The Max Cover Problem (MCP) Formulation

Additional Notation:

Λ_k: maximum deviation path length of OD pair $k \in K$

$$\theta_k = \begin{cases} 1 & \text{if OD pair } k \text{ is covered,} \\ 0 & \text{otherwise.} \end{cases} \qquad k \in K$$

Formulation:

$$\text{(MCP) maximize} \quad \sum_{k \in K} d_k \theta_k \tag{A.1}$$

$$\text{s.t.} \sum_{j:(i,j)\in A} x_{ijk} - \sum_{j:(j,i)\in A} x_{jik} = \begin{cases} \theta_k, & \text{if } i = O_k^a, \\ -\theta_k, & \text{if } i = D_k^a, \\ 0, & \textit{otherwise.} \end{cases} \qquad i \in N, \quad k \in K \tag{A.2}$$

$$\sum_{(i,j)\in A} \ell_{ij} x_{ijk} \le \Lambda_k \theta_k \qquad k \in K \tag{A.3}$$

(2.3), (2.12), (2.9), (2.10)

$$\theta_k \in \{0, 1\} \qquad k \in K \tag{A.4}$$

A.2 The Set Cover Problem Formulation

$$\begin{aligned} \text{(SCP)} \quad \text{minimize} \quad & \sum_{j \in N^0} y_j \\ \text{s.t.} \quad & \text{(2.2)},\ \text{(2.3)},\ \text{(A.3)},\ \text{(2.9)},\ \text{(2.10)} \end{aligned} \tag{A.5}$$

Appendix B

Appendix of Chapter 3

B.1 Detailed Computational Results

Table B.1: Algorithm results on *CA339* instances under cooperative EV user response.

Configuration										
R	**Dev Tol**	**α (min)**	$\sum y$	$\sum z$	**Obj ($Mil)**	**Avg Dev(%)**	**Max Dev(%)**	**# Cuts (3.41), (3.42)**	**# Cuts (3.43)**	**Iterations -# Cuts (3.48)-**
250	0%	30	4	190	14.85	-	-	762	0	2175
		10	4	198	15.45	-	-	762	0	2229
		5	4	204	15.90	-	-	762	0	2233
		1	4	214	16.65	-	-	762	0	2233
	10%	30	3	180	13.95	0.79	5.02	1132	0	2167
		10	3	187	14.475	0.79	5.02	1132	0	2167
		5	3	192	14.85	0.66	9.22	1130	0	2132
		1	3	201	15.525	0.79	5.02	1090	0	2032
	25%	30	2	178	13.65	0.88	20.60	4388	0	1709
		10	2	184	14.10	0.88	20.60	4388	0	1767
		5	2	188	14.40	0.88	20.60	4388	0	1770
		1	2	198	15.15	0.77	23.43	4388	0	1710
	50%	30	2	178	13.65	0.88	20.60	9870	0	1593
		10	2	184	14.10	0.88	20.60	9870	0	1592
		5	2	188	14.40	0.88	20.60	9870	0	1587
		1	2	198	15.15	0.88	20.60	9870	0	1588
200	0%	30	7	330	25.80	-	-	1030	0	2491
		10	7	345	26.925	-	-	1030	0	2490
		5	7	356	27.75	-	-	1030	0	2488
		1	7	374	29.10	-	-	1030	0	2492
	10%	30	5	319	24.675	1.01	9.54	1470	16	2446
		10	5	328	25.35	0.97	9.54	1470	16	2434
		5	5	336	25.95	1.01	9.54	1470	16	2389
		1	5	351	27.075	1.01	9.54	1470	16	2270
	25%	30	3	319	24.375	2.38	22.68	2510	0	2309
		10	3	327	24.975	2.38	22.68	2510	0	2231
		5	3	334	25.50	1.77	22.29	2510	0	2247
		1	3	348	26.55	1.90	22.29	2510	0	2257
	50%	30	3	310	23.70	3.35	40.94	11,692	0	1309
		10	3	318	24.30	7.07	49.35	11,692	0	1342
		5	3	326	24.90	7.78	49.35	11,692	0	1215
		1	3	339	25.875	7.07	49.35	11,692	0	1380

Table B.2: *N25* results under cooperative EV user response.

Configuration			$R = 15$							$R = 20$							$R = 25$						
Dev Tol	**β**	**α (min)**	$\sum y$	$\sum z$	**Obj ($1M)**	**Avg Dev(%)**	**Max Dev(%)**	**Gurobi Time(s)**	**Algo Sol Time(s)**	$\sum y$	$\sum z$	**Obj ($1M)**	**Avg Dev(%)**	**Max Dev(%)**	**Gurobi Time(s)**	**Algo Sol Time(s)**	$\sum y$	$\sum z$	**Obj ($1M)**	**Avg Dev(%)**	**Max Dev(%)**	**Gurobi Time(s)**	**Algo Sol Time(s)**
0%	95%	30	8	125	10.575	-	-	1.80	6.03	5	67	5.775	-	-	0.56	4.68	3	32	2.85	-	-	0.47	2.44
		10	8	141	11.775	-	-	4.48	5.48	5	76	6.45	-	-	1.21	4.08	3	36	3.15	-	-	0.53	2.44
		5	8	149	12.375	-	-	2.09	5.49	5	80	6.75	-	-	1.47	4.16	3	38	3.30	-	-	0.29	2.37
		1	8	159	13.125	-	-	2.10	5.41	5	86	7.20	-	-	1.21	4.27	3	40	3.45	-	-	0.43	2.41
	90%	30	8	122	10.35	-	-	2.62	5.86	5	65	5.625	-	-	1.61	4.66	3	31	2.775	-	-	0.27	2.36
		10	8	134	11.25	-	-	2.91	6.54	5	72	6.15	-	-	0.90	4.04	3	34	3.00	-	-	0.55	2.32
		5	8	141	11.775	-	-	1.86	5.90	5	76	6.45	-	-	2.19	4.08	3	36	3.15	-	-	0.32	2.35
		1	8	150	12.45	-	-	1.57	5.78	5	80	6.75	-	-	1.43	4.24	3	38	3.30	-	-	0.39	2.36
	80%	30	8	118	10.05	-	-	2.62	6.23	5	62	5.40	-	-	1.09	4.97	3	30	2.70	-	-	0.55	2.36
		10	8	127	10.725	-	-	1.92	5.95	5	68	5.85	-	-	0.86	4.09	3	32	2.85	-	-	0.47	2.34
		5	8	132	11.10	-	-	1.76	5.37	5	71	6.075	-	-	1.62	4.12	3	34	3.00	-	-	0.64	2.48
		1	8	139	11.625	-	-	1.90	5.53	5	74	6.30	-	-	1.66	4.30	3	35	3.075	-	-	0.29	2.35
10%	95%	30	8	125	10.575	0	0	37.63	5.73	5	67	5.775	0	0	3.32	5.08	3	32	2.85	0	0	1.00	2.35
		10	8	141	11.775	0	0	79.01	5.21	5	76	6.45	0	0	13.12	4.03	3	36	3.15	0	0	1.71	2.36
		5	8	149	12.375	0	0	64.78	5.31	5	80	6.75	0	0	23.86	4.12	3	38	3.30	0	0	1.02	2.36
		1	8	159	13.125	0	0	57.49	5.17	5	86	7.20	0	0	50.26	4.20	3	40	3.45	0	0	0.87	2.35
	90%	30	8	122	10.35	0	0	70.88	5.56	5	65	5.625	0	0	4.47	4.51	3	31	2.775	0	0	0.90	2.33
		10	8	134	11.25	0	0	54.89	5.85	5	72	6.15	0	0	8.46	4.13	3	34	3.00	0	0	1.34	2.33
		5	8	141	11.775	0	0	53.86	5.60	5	76	6.45	0	0	39.63	4.01	3	36	3.15	0	0	1.22	2.35
		1	8	150	12.45	0	0	96.37	5.48	5	80	6.75	0	0	43.98	4.16	3	38	3.30	0	0	1.66	2.34
	80%	30	8	118	10.05	0	0	30.36	5.64	5	62	5.40	0	0	5.92	5.24	3	30	2.70	0	0	0.79	2.31
		10	8	127	10.725	0	0	54.77	5.42	5	68	5.85	0	0	31.09	3.96	3	32	2.85	0	0	1.46	2.30
		5	8	132	11.10	0	0	60.02	5.10	5	71	6.075	0	0	7.84	4.10	3	34	3.00	0	0	1.59	2.39
		1	8	139	11.625	0	0	46.81	5.34	5	74	6.30	0	0	39.12	4.28	3	35	3.075	0	0	1.18	2.33
25%	95%	30	6	123	10.125	1.74	21.43	260.32	4.07	3	63	5.175	2.44	22.22	42.43	2.80	2	31	2.625	0.6	13.04	5.21	1.88
		10	6	135	11.025	1.74	21.43	432.03	4.03	3	70	5.70	1.49	16.67	10.94	2.83	2	34	2.85	0.6	13.04	7.55	1.85
		5	6	143	11.625	1.74	21.43	519.97	4.06	3	74	6.00	1.49	16.67	50.22	2.82	2	36	3.00	1.78	22.73	6.31	1.87
		1	6	152	12.30	1.74	21.43	437.18	4.12	3	79	6.375	1.49	16.67	24.17	2.83	2	38	3.15	0.45	13.04	37.06	1.86
	90%	30	6	119	9.825	1.74	21.43	186.63	4.05	3	62	5.10	2.44	22.22	11.05	2.84	2	30	2.55	0.46	13.04	6.99	1.86
		10	6	130	10.65	1.74	21.43	312.39	4.05	3	67	5.475	2.44	22.22	19.67	2.90	2	32	2.70	0.45	13.04	5.60	1.84
		5	6	136	11.10	1.74	21.43	340.21	4.03	3	70	5.70	2.44	22.22	34.66	3.00	2	34	2.85	0.45	13.04	6.12	1.86
		1	6	145	11.775	1.74	21.43	757.12	4.12	3	75	6.075	2.44	22.22	106.45	2.82	2	36	3.00	0.46	13.04	5.22	1.90
	80%	30	6	116	9.60	1.74	21.43	119.71	4.13	3	60	4.95	2.44	22.22	10.91	2.85	2	29	2.475	0.45	13.04	9.76	1.88
		10	6	124	10.20	1.74	21.43	242.59	4.09	3	64	5.25	1.49	16.67	12.98	2.85	2	31	2.625	0.23	13.04	6.00	1.85
		5	6	129	10.575	1.74	21.43	295.89	4.14	3	67	5.475	2.44	22.22	14.63	2.81	2	32	2.70	0.45	13.04	7.34	1.87
		1	6	135	11.025	1.74	21.43	287.59	4.00	3	70	5.70	1.49	22.22	11.18	2.85	2	34	2.85	0.45	13.04	6.67	1.85
50%	95%	30	5	121	9.825	4.12	38.46	490.50	3.76	2	63	5.025	5.67	44.44	6.66	2.34	2	30	2.55	3.85	39.13	19.49	1.67
		10	5	133	10.725	4.12	38.46	837.50	3.75	2	69	5.475	5.67	44.44	7.58	2.30	2	34	2.85	0.46	13.04	15.01	1.67
		5	5	140	11.25	4.12	38.46	1109.32	3.72	2	72	5.70	5.67	44.44	5.66	2.29	2	36	3.00	0.46	13.04	26.24	1.67
		1	5	149	11.925	4.12	38.46	1346.00	3.79	2	77	6.075	5.67	44.44	6.59	2.29	2	38	3.15	2.91	30.43	59.52	1.68
	90%	30	5	119	9.675	4.12	38.46	1487.53	3.72	2	62	4.95	5.67	44.44	4.59	2.28	2	30	2.55	1.78	22.73	21.67	1.67
		10	5	128	10.35	4.12	38.46	693.80	3.75	2	67	5.325	5.67	44.44	5.35	2.32	2	32	2.70	2.91	30.43	25.28	1.71
		5	5	134	10.80	4.12	38.46	1426.74	3.73	2	70	5.55	5.67	44.44	10.20	2.33	2	34	2.85	0.23	13.04	70.60	1.68
		1	5	142	11.40	4.12	38.46	1275.94	3.77	2	73	5.775	5.67	44.44	13.89	2.31	2	36	3.00	1.87	26.09	72.84	1.68
	80%	30	5	116	9.45	4.12	38.46	519.05	3.72	2	61	4.875	5.67	44.44	5.23	2.31	2	29	2.475	0.46	13.04	181.38	1.67
		10	5	123	9.975	5.58	46.15	1630.67	3.78	2	65	5.175	5.67	44.44	10.22	2.30	2	31	2.625	0.46	13.04	16.77	1.68
		5	5	127	10.275	4.12	38.46	802.30	3.73	2	67	5.325	5.67	44.44	5.99	2.32	2	32	2.70	0.46	13.04	81.06	1.67
		1	5	133	10.725	4.12	38.46	935.82	3.73	2	70	5.55	5.67	44.44	7.53	2.31	2	34	2.85	0.45	13.04	61.61	1.67

Table B.3: Algorithm results on *US-E420* instances under cooperative EV user response.

Configuration			Cooperative							
R	**Dev Tol**	**α (min)**	$\sum y$	$\sum z$	**Obj ($Mil)**	**Avg Dev(%)**	**Max Dev(%)**	**# Cuts (3.41), (3.42)**	**# Cuts (3.43)**	**Iterations -# Cuts (3.48)-**
600	0%	30	55	586	52.20	-	-	1264	0	2126
		10	55	653	57.225	-	-	1264	0	2119
		5	55	693	60.225	-	-	1264	0	2129
		1	55	734	63.30	-	-	1264	0	2120
	10%	30	39	548	46.95	1.41	9.57	1166	0	2710
		10	39	606	51.30	1.42	9.57	1166	0	2705
		5	39	638	53.70	1.42	9.57	1166	0	2712
		1	39	673	56.325	1.41	9.57	1166	0	2705
	25%	30	31	535	44.775	5.89	24.72	2674	585	1117
		10	31	585	48.525	5.92	24.72	2674	585	1116
		5	31	613	50.625	5.80	24.31	2674	585	1117
		1	31	652	53.55	5.80	24.72	2674	585	1043
	50%	30	25	536	43.95	9.08	48.29	8732	315	62
		10	25	583	47.475	9.08	48.29	8720	304	58
		5	25	606	49.20	9.08	48.29	8732	304	60
		1	25	648	52.35	9.08	48.29	8732	304	59
400	0%	30	97	1009	90.225	-	-	1852	0	1826
		10	97	1130	99.30	-	-	1852	0	1760
		5	97	1196	104.25	-	-	1852	0	1916
		1	97	1270	109.80	-	-	1852	0	1903
	10%	30	79	984	85.65	1.86	9.56	2132	134	1435
		10	79	1089	93.525	1.96	9.60	2132	134	1450
		5	79	1142	97.50	2.07	9.85	2132	134	1458
		1	79	1218	103.20	1.86	9.56	2132	134	1513
	25%	30	65	974	82.80	6.12	24.62	4934	1488	71
		10	65	1068	89.85	6.12	24.62	4930	1444	66
		5	65	1119	93.675	6.12	24.62	4934	1488	71
		1	65	1188	98.85	6.12	24.62	4930	1451	67
	50%	30	51	999	82.575	13.0	49.52	14,072	678	1
		10	51	1085	89.025	13.0	49.52	14,072	678	1
		5	51	1137	92.925	13.0	49.52	14,072	678	1
		1	51	1203	97.875	13.0	49.52	14,072	678	1

Table B.4: *N25* results under cooperative and uncooperative EV user response.

Configuration				Cooperative				Uncooperative				Difference			
R	**Dev Tol**	**β**	**α (min)**	$\sum y$	$\sum z$	**Obj ($Mil)**	**Avg Dev(%)**	$\sum y$	$\sum z$	**Obj ($Mil)**	**Avg Dev(%)**	$\sum y$	$\sum z$	**Obj ($Mil)**	**Avg Dev(%)**
15	0%	90%	30	8	122	10.35	-	8	128	10.80	-	0	+6	4.35	-
			10	8	134	11.25	-	8	141	11.775	-	0	+7	4.67	-
			1	8	150	12.45	-	8	157	12.975	-	0	+7	4.22	-
		80%	30	8	118	10.05	-	8	124	10.50	-	0	+6	4.48	-
			10	8	127	10.725	-	8	133	11.175	-	0	+6	4.20	-
			1	8	139	11.625	-	8	146	12.15	-	0	+7	4.52	-
	10%	90%	30	8	122	10.35	0.00	8	128	10.80	0.00	0	+6	4.35	0.00
			10	8	134	11.25	0.00	8	141	11.775	0.00	0	+7	4.67	0.00
			1	8	150	12.45	0.00	8	157	12.975	0.00	0	+7	4.22	0.00
		80%	30	8	118	10.05	0.00	8	124	10.50	0.00	0	+6	4.48	0.00
			10	8	127	10.725	0.00	8	133	11.175	0.00	0	+6	4.20	0.00
			1	8	139	11.325	0.00	8	146	12.15	0.00	0	+7	4.52	0.00
	25%	90%	30	6	119	9.825	1.74	6	123	10.125	1.74	0	+4	3.05	0.00
			10	6	130	10.65	1.74	6	134	10.95	1.74	0	+4	2.82	0.00
			1	6	145	11.775	1.74	6	150	12.15	1.74	0	+5	3.18	0.00
		80%	30	6	116	9.60	1.74	6	120	9.90	1.74	0	+4	3.13	0.00
			10	6	124	10.20	1.74	6	128	10.50	1.74	0	+4	2.94	0.00
			1	6	135	11.025	1.74	6	140	11.40	1.74	0	+5	3.40	0.00
	50%	90%	30	5	119	9.675	5.58	5	121	9.825	4.12	0	+2	1.55	-1.46
			10	5	128	10.35	5.58	5	131	10.575	4.12	0	+3	2.17	-1.46
			1	5	142	11.40	5.58	5	145	11.625	4.12	0	+3	1.97	-1.46
		80%	30	5	116	9.45	5.58	5	119	9.675	3.97	0	+3	2.38	-1.61
			10	5	123	9.975	5.58	5	125	10.125	4.12	0	+2	1.50	-1.46
			1	5	133	10.725	4.12	5	137	11.025	4.12	0	+4	2.80	0.00
20	0%	90%	30	5	65	5.625	-	5	69	5.925	-	0	+4	5.33	-
			10	5	72	6.15	-	5	77	6.525	-	0	+5	6.10	-
			1	5	80	6.75	-	5	86	7.20	-	0	+6	6.67	-
		80%	30	5	62	5.40	-	5	67	5.775	-	0	+5	6.94	-
			10	5	68	5.85	-	5	72	6.15	-	0	+4	5.13	-
			1	5	74	6.30	-	5	79	6.675	-	0	+5	5.95	-
	10%	90%	30	5	65	5.625	0.00	5	69	5.925	0.00	0	+4	5.33	0.00
			10	5	72	6.15	0.00	5	77	6.525	0.00	0	+5	6.10	0.00
			1	5	80	6.75	0.00	5	85	7.125	0.00	0	+5	5.56	+0.30
		80%	30	5	62	5.40	0.00	5	67	5.775	0.00	0	+5	6.94	0.00
			10	5	68	5.85	0.00	5	72	6.15	0.00	0	+4	5.13	0.00
			1	5	74	6.30	0.00	5	79	6.675	0.00	0	+5	5.95	0.00
	25%	90%	30	3	62	5.10	1.49	3	64	5.25	1.49	0	+2	2.94	0.00
			10	3	67	5.475	1.49	3	69	5.625	1.49	0	+2	2.74	0.00
			1	3	75	6.075	1.49	3	77	6.225	1.49	0	+2	2.47	0.00
		80%	30	3	60	4.95	1.49	3	62	5.10	1.49	0	+2	3.03	0.00
			10	3	64	5.25	1.49	3	66	5.40	1.49	0	+2	2.86	0.00
			1	3	70	5.70	1.49	3	72	5.85	1.49	0	+2	2.63	0.00
	50%	90%	30	2	62	4.95	5.67	2	62	4.95	5.67	0	0	0.00	0.00
			10	2	67	5.325	5.67	2	67	5.325	5.67	0	0	0.00	0.00
			1	2	73	5.775	5.67	2	73	5.775	5.67	0	0	0.00	0.00
		80%	30	2	61	4.875	5.67	2	61	4.875	5.67	0	0	0.00	0.00
			10	2	65	5.175	5.67	2	65	5.175	5.67	0	0	0.00	0.00
			1	2	70	5.55	5.67	2	70	5.55	5.67	0	0	0.00	0.00
25	0%	90%	30	3	31	2.775	-	3	32	2.85	-	0	+1	2.70	-
			10	3	34	3.00	-	3	36	3.15	-	0	+2	5.00	-
			1	3	38	3.30	-	3	39	3.375	-	0	+1	2.27	-
		80%	30	3	30	2.70	-	3	31	2.775	-	0	+1	2.78	-
			10	3	32	2.85	-	3	33	2.925	-	0	+1	2.63	-
			1	3	35	3.075	-	3	37	3.225	-	0	+2	4.88	-
	10%	90%	30	3	31	2.775	0.15	3	32	2.85	0.00	0	+1	2.70	-0.15
			10	3	34	3.00	0.15	3	36	3.15	0.00	0	+2	5.00	-0.15
			1	3	38	3.30	0.15	3	39	3.375	0.00	0	+1	2.27	-0.15
		80%	30	3	30	2.70	0.15	3	31	2.775	0.00	0	+1	2.78	-0.15
			10	3	32	2.85	0.15	3	33	2.925	0.00	0	+1	2.63	-0.15
			1	3	35	3.075	0.15	3	37	3.225	0.00	0	+2	4.88	-0.15
	25%	90%	30	2	30	2.55	0.46	2	30	2.55	1.78	0	0	0.00	+1.32
			10	2	32	2.70	0.45	2	32	2.70	1.78	0	0	0.00	+1.33
			1	2	36	3.00	0.45	2	36	3.00	1.78	0	0	0.00	+1.33
		80%	30	2	29	2.475	0.46	2	29	2.475	0.45	0	0	0.00	-0.01
			10	2	31	2.625	0.46	2	31	2.625	1.78	0	0	0.00	+1.32
			1	2	34	2.85	0.46	2	34	2.85	1.78	0	0	0.00	+1.32
	50%	90%	30	2	30	2.55	5.78	2	30	2.55	1.78	0	0	0.00	-4.00
			10	2	32	2.70	2.91	2	32	2.70	1.78	0	0	0.00	-1.13
			1	2	36	3.00	1.87	2	36	3.00	1.78	0	0	0.00	-0.09
		80%	30	2	29	2.475	5.78	2	29	2.475	0.45	0	0	0.00	-5.33
			10	2	31	2.625	5.78	2	31	2.625	1.78	0	0	0.00	-4.00
			1	2	34	2.85	1.87	2	34	2.85	1.78	0	0	0.00	-0.09