

RELIABLE TRANSPORTATION NETWORKS UTILIZING EMERGING
TECHNOLOGIES AND PRICING STRATEGIES

By

Seyede Fatemeh Fakhrmoosavi

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Civil Engineering—Doctor of Philosophy

2021

ABSTRACT

RELIABLE TRANSPORTATION NETWORKS UTILIZING EMERGING TECHNOLOGIES AND PRICING STRATEGIES

By

Seyede Fatemeh Fakhrmoosavi

Travel time reliability plays a pivotal role in the system efficiency and level of service of transportation networks. Transportation network users are heterogeneous, and they may value travel time reliability differently. The importance level of travel time reliability for different travelers depends upon many factors including the user's risk acceptance level and trip purpose and departure time. Thus, travelers tend to maximize travel time reliability, in addition to minimizing their travel times.

One of the main challenges in transportation planning is the high computational time of traffic simulation tools that consider heterogeneous users and their responses to travel time reliability. Path finding problem constitutes an essential problem in these traffic simulation tools. Therefore, this study presents two heuristic algorithms to improve the computational time of reliable path finding algorithms by reducing the network size of each specific origin and destination pair in stochastic time-dependent networks. The network contraction algorithms, presented in this study, are based on the comparison of optimistic and pessimistic solutions resulting from minimum and maximum travel time realizations of a Monte-Carlo simulation-based approach. The major contribution of the proposed approach is to improve computational efficiency of the stochastic path finding problem, considering travel time correlations and travelers' heterogeneity, in large-scale applications. Comparing the performance and accuracy of the approach with those of the approach without any network contraction for two large-scale networks demonstrates significant computational improvements and a high accuracy level.

Different traffic and demand management strategies have also been used to improve reliability of transportation networks. These strategies, including congestion pricing, have great impacts on users' reliable path choices. Considering a reliability measure in the travelers' path choices naturally impacts the congestion pattern, which in turn, affects the outcomes of pricing strategies. Furthermore, congestion pricing alters link travel time distributions in stochastic transportation networks. Therefore, this study finds an equitable pricing scheme that minimizes the total travel time of auto users in a general bimodal network considering heterogeneous users with different values of time and reliability. The main contribution of this proposed approach is accounting for travel time reliability in finding an equitable pricing scheme. This approach is successfully applied to a modified Sioux Falls network to explore the impacts of subsidization strategy, congestion level, and considering travel time reliability on the pricing strategy and its effectiveness.

Finally, emerging technologies, such as connected and autonomous vehicle technologies, have attracted the attention of transportation system planners in recent years, as an alternative to improve mobility and reliability of transportation networks. Having a traffic simulation tool that considers the presence of these technologies is essential to estimate their impacts on traffic congestion and travel time reliability. Therefore, this study presents a mesoscopic simulation tool to account for the presence of connected and autonomous vehicles at the network level by incorporating adaptive fundamental diagrams due to the non-uniform distribution of different vehicle types with heterogeneous drivers. This tool is then used to investigate the impacts of a mixed traffic of connected, autonomous, and human-driven vehicles on traffic flow and travel time reliability at the network level. The results show the superiority of connected and autonomous vehicles over regular vehicles in mitigating traffic congestion and improving travel time reliability.

Copyright by
SEYEDE FATEMEH FAKHRMOOSAVI
2021

*To my beloved husband, Vahid, and my wonderful parents and parents-in-law
for their unconditional support and love*

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor, Dr. Ali Zockaie, for his tremendous support, guidance, and patience throughout my study at Michigan State University. His unwavering enthusiasm to teaching has always motivated me to learn more. I would also like to thank my committee members, Dr. Mehrnaz Ghamami, Dr. Timothy Gates, Dr. Peter Savolainen, and Dr. Mohsen Zayernouri for their time, support, and guidance over the last five years. Their valuable insights and advice contributed greatly to my personal and professional development. I also extend my appreciation to Dr. Khaled Abdelghany and Dr. Alireza Talebpoor for their contributions in different aspects of my research. Also, a big thank you to my peers in the transportation group of Michigan State University and my dear friends who were always there for me. My special thanks go to my best friend and my husband, Vahid Morovati, for his endless love, support, and encouragement. Last but not least, I would like to thank my dear parents, parents-in-law, and siblings for always believing in me.

TABLE OF CONTENTS

LIST OF TABLES	ix
LIST OF FIGURES	x
CHAPTER 1 Introduction.....	1
1-1- Research Motivation	1
1-2- Problem Statement	4
1-3- Research Objectives and Contributions	6
1-3-1- Path Finding Problem Considering Heterogeneous Users.....	7
1-3-2- Travel Time Reliability and Congestion Pricing	7
1-3-3- Impacts of Connected and Autonomous Vehicles on Travel Time Reliability	8
1-4- Research Approach	8
1-5- Study Outline.....	11
CHAPTER 2 - Background Review	12
2-1- Overview	12
2-2- Travel Time Reliability	12
2-3- Reliable Path Finding Problem	13
2-4- Congestion Pricing Considering Travel Time Reliability.....	14
2-5- Traffic Impacts of Connected and Autonomous Vehicles	16
CHAPTER 3 - Network Contraction for the Path Finding Problem in Stochastic Networks Considering Heterogeneous Users	19
3-1- Overview	19
3-2- Problem Formulation.....	22
3-3- Methodology	24
3-3-1- Solution Algorithm	25
3-4- Numerical Experiments.....	28
3-4-1- Case Study Specifications.....	28
3-4-2- Numerical Results	30
3-5- Summary	38
CHAPTER 4 - An Iterative Learning Approach for Network Contraction	40
4-1- Overview	40
4-2- Problem Statement	43
4-3- Methodology Development.....	44
4-3-1- Data Intuition	46
4-3-2- Solution Algorithm	50
4-4- Numerical Experiments.....	57
4-5- Model Transferability.....	67
4-6- Summary	72
CHAPTER 5 - Travel Time Reliability and Congestion Pricing.....	74
5-1- Overview	74

5-2- Problem Formulation.....	77
5-3- Methodology	81
5-4- Numerical Experiments.....	88
5-5- Summary	95
CHAPTER 6 - Impacts of Connected and Autonomous Vehicles on the Network Travel Time	
Reliability.....	97
6-1- Overview	97
6-2- Microscopic Simulation Framework.....	100
6-2-1- Vehicles with No Communication Capability	100
6-2-2- Vehicles with Active Vehicle-To-Vehicle Communications	101
6-2-3- Autonomous Vehicles.....	102
6-3- Macroscopic Relations at the Facility Level	103
6-4- Mesoscopic Simulation Framework.....	105
6-4-1- Network Fundamental Diagram.....	109
6-5- Numerical Experiments.....	110
6-6- Summary	121
CHAPTER 7 - Conclusions and Future Work.....	
7-1- Summary and Concluding Remarks.....	123
7-2- Future Research Directions	127
REFERENCES	130

LIST OF TABLES

Table 4-1- The number of instances for different relative errors that the network contraction approach, using different data mining factors, has a worse objective function value than the full network of Zockaie et al. (4) for destination 84 using SPOTAR reliability rule	50
Table 4-2- The number of instances for different relative errors that network contraction approaches have a worse objective function value than the full network approach of Zockaie et al. (4) for destination 84 using SPOTAR reliability rule	62
Table 4-3- The number of instances that different learning approaches have a worse objective function, relative to the full network approach of Zockaie et al. (4), for different relative errors and destinations using a) MTTBP reliability rule; b) SPOTAR reliability rule	66
Table 4-4- The number of instances that different learning approaches have a worse objective function, relative to the full network approach of Zockaie et al. (4), for different relative errors, destinations, and MTTBP problem in the Salt Lake City network	71
Table 5-1- Definitions of parameters and variables.....	78
Table 5-2- Particle Swarm Optimization Algorithm Parameters and Variables.....	84
Table 5-3- Toll Distribution Strategies	91
Table 5-4- Optimum tolls and the resulting system and users' costs for different network scenarios and toll distribution strategies for different reliability relations	93
Table 5-5- Comparison of the results with and without considering travel time reliability for the congested network and the credit-based toll distribution strategy among the same OD pair	95

LIST OF FIGURES

Figure 3-1- Configuration of the Chicago downtown network	29
Figure 3-2- Mean vs. variance of travel time for different links of the network at different time intervals of the a) first, b) fifth, c) tenth, and d) fifteenth (last) 20-minute time interval (15)	30
Figure 3-3- Optimal path travel time distributions at different departure times with correlated and independent distributions and for both with and without network contraction (on top of each other) a) from origin 36 to destination 1; b) from origin 196 to destination 1 (Pessimistic and optimistic limits of 5 and 0.8, 100 stage 1 iterations, and 1000 stage 2 iterations).....	32
Figure 3-4- Computational performance for different origins to zone 1 a) execution times of the stochastic path finding algorithm with contraction; b) dynamic deterministic shortest path execution times; c) contraction ratio for different optimistic/pessimistic limits 35	35
Figure 3-5- Sub-network configurations for different optimistic/pessimistic coefficients a) Origin zone 100 to destination 87; b) Origin zone 20 to destination one	36
Figure 3-6- SPOTAR a) execution time and b) cumulative difference of objective function values relative to the stage one iterations of 1000 for the proposed method with contraction and parallelization and regular method without contraction (Destination one and optimistic/pessimistic coefficients of 0.8 and 5)	38
Figure 4-1- A schematic illustration of network contraction for the reliable path finding problem	45
Figure 4-2- a) Maximum b) Minimum time-dependent labels over 100 realizations from each node to destination 84 of the Chicago network, divided by FFTT, versus the FFTT from each node to the same destination	48
Figure 4-3- Average number of nodes in sub-networks using different data mining factors in the network contraction approach for all origins to the destination 84 of Chicago downtown network in 2-minute intervals of FFTTs	50
Figure 4-4- A brief illustration of the learning approach for efficiently solving reliable path finding problem.....	56
Figure 4-5- The relationship between the execution times and the number of nodes for the fixed learning approach (learning factor =1.3) to find the optimal path from different origins to destination 84	59
Figure 4-6- Number of nodes in the sub-network for all origins to the destination 84 in 2-minute intervals of FFTT for different network contraction approaches	61

Figure 4-7- The comparison of execution time for all origins to the destination 84 in 2-minute intervals of FFTT for different network contraction approaches	62
Figure 4-8- The comparison of number of nodes in the sub-network for all origins to the destination 1 in 2-minute intervals of FFTT for different network contraction approaches	64
Figure 4-9- The comparison of number of nodes in the sub-network for all origins to the destination 57 in 2-minute intervals of FFTT for different network contraction approaches	64
Figure 4-10- Sub-network configurations for the two learning approaches and the conservative approach from origin 141 to the destination 1 in the Chicago downtown network ..	67
Figure 4-11- Number of nodes in the sub-network in 2-minute intervals of FFTT for different network contraction approaches (calibrated based on the Chicago network) from all origins to destination a) 27, b) 80, c) 145, d) 55, and e) 113 of the Salt Lake City network.....	69
Figure 4-12- Sub-network configurations for the two approaches and the conservative approach from origin 1,178 to the destination 145 for the MTTBP problem applied to the Salt Lake City	72
Figure 5-1- Particle swarm optimization algorithm to find optimal toll values minimizing highway travel time and violation of Pareto-improving condition considering reliability	85
Figure 5-2- The reliability-based user equilibrium algorithm considering toll values and heterogeneous users.....	87
Figure 5-3- (a) Selected toll zones on the modified Sioux Falls network (x represents link flow and c denotes the capacity of the link) (b) the fitted concave function between standard deviation and mean travel time normalized to free flow travel time.....	90
Figure 5-4- Value of time (VOT) and value of reliability (VOR) (a) distributions and (b) classes	90
Figure 5-5- The search area for the credit-based toll distributed among the same OD pairs of the congested Sioux Falls network.....	92
Figure 6-1- a) Flow-density and b) speed-density calibrated relationships at the equilibrium state for a segment occupied only with certain driver classes of HDVs, CVs, or AVs...	105
Figure 6-2- Illustration of the proposed approach to update link speeds given a specific density distribution of different vehicle types with heterogeneous drivers	107
Figure 6-3- AM peak simulation demand for the Chicago downtown network.....	111
Figure 6-4- (a-d) NFD and (e) maximum density and area of hysteresis loop for different MPRs of CVs, and AVs.....	114

Figure 6-5- NFD results considering the impact of (a) only en-route users, (b) en-route users and adaptive traffic flow models on freeway links, (c) en-route users and freeway and arterial traffic flow models, and (d) en-route users, freeway and arterial traffic flow models, and adjusted intersection capacity 116

Figure 6-6- Values of maximum density and area of hysteresis loop considering the impacts of different factors on network dynamics 117

Figure 6-7- Standard deviation vs mean travel time normalized to free flow travel time for different links in a fully HDV network for the a) first, b) second, c) third, and d) fourth 100-minute time intervals 120

Figure 6-8- Standard deviation vs mean travel time normalized to free flow travel time for different links in a fully AV network for the a) first, b) second, c) third, and d) fourth 100-minute time intervals 121

CHAPTER 1 Introduction

1-1- Research Motivation

The performance of transportation systems is of significant importance in the growth of all nations. In addition, traffic congestion in urban areas has imposed many direct and indirect expenses on roadway users with a growing pattern. Transportation planners incorporate multiple strategies and technologies to improve mobility and safety and reduce emission across transportation systems. However, population growth and lack of available resources to develop a new transportation system or expand current road networks continue to cause serious challenges in mobility and safety, especially in large cities. Therefore, there have been extensive efforts in recent decades to take advantage of different strategies and technologies to improve mobility and safety to the highest level possible. To do so, the impacts of these strategies and technologies on transportation systems need to be investigated in order to make better plans for the future.

Travel time reliability is a critical measure indicating the performance level of transportation systems in addition to average travel time that has been used extensively as the main measure of service for transportation systems. Therefore, transportation analysts need to incorporate the travel time variability in addition to the average travel time in their models. Travel time uncertainty in transportation networks comes from either the demand side, e.g., special events or daily demand fluctuations, or the supply side, e.g., traffic incidents, natural hazards, work zones, and poor weather conditions (1). Consequently, different users tend to minimize not only their travel times but also the travel time variability caused by such uncertainties (2). In addition, different travelers have distinct decisions in response to travel time uncertainties that may vary by their preferences and risk tolerance levels. Therefore, travel time reliability and users'

heterogeneity are critical factors that should be considered in improving the mobility of transportation systems.

The path finding problem constitutes an extremely essential problem with numerous applications in different fields. This problem is the core element of many other problems in the transportation area and should be solved repeatedly to simulate traffic in transportation systems. Deterministic shortest paths, either dynamic or static, may lead users toward risky paths due to the uncertain nature of travel times in transportation networks. Consequently, different users tend to minimize not only their travel times but also the travel time variability caused by such uncertainties (2). In addition, as mentioned earlier, different travelers have distinct decisions in response to travel time uncertainties that may vary by their preferences and risk tolerance. The same traveler might frequently choose different paths for similar trips based on their departure time and purpose. Therefore, travelers' preferences are expected to be considered in the path selection procedure to find reliable routes for users, called stochastic or reliable path finding problem (3,4). As stochastic path finding problem should be solved multiple times for traffic simulations, it is significantly important to develop algorithms to solve this problem in a computationally reasonable time. In addition, the computational efficiency of stochastic path finding algorithms plays a critical role in predicting the performance of transportation systems under different situations to find the best strategies to improve mobility and safety.

In recent decades, transportation planners and policymakers have become increasingly interested in congestion pricing as a possible mechanism to mitigate traffic congestion in urban areas. Without congestion pricing, traffic is distributed between different routes in the network such that a user-equilibrium (UE) traffic route assignment pattern is obtained. This undesirable UE route assignment pattern results in an inefficient use of the network capacity and excessive delays.

However, pricing strategies shift the UE pattern towards a system optimal (SO) pattern, as a portion of the travelers modify their routes to avoid paying the tolls. Despite technological advancements in implementing pricing strategies, some serious impediments hinder the application of these strategies in practice. For instance, without a careful redistribution scheme of collected tolls, congestion pricing can inequitably affect travelers (providing benefit for some travelers and loss for others). Therefore, a comprehensive congestion pricing strategy is required not only to alleviate congestion, but also to neutralize equity issues.

In addition, given the growing utilization of congestion pricing strategies in recent years, it is crucial to account for the monetary cost of the tolls levied on users in addition to the travel time in stochastic path choice models. Doing so, heterogeneous users choose their path based on the least generalized cost path, which is the sum of out-of-pocket cost and the weighted value of travel time. The later value is mostly identified as the multiplication of users' Value of Time (VOT) by the paths average travel time in the literature. However, transportation policymakers need to incorporate the travel time variability in addition to the average travel time. Furthermore, considering a reliability measure in the travelers' path choice would naturally impact the congestion pattern in the network that in turn, affects the outcomes of pricing strategies. This calls for incorporating reliability measures in congestion pricing schemes and their resulting impacts on users' optimal paths. A few studies in the literature have considered applications of the variable travel time in congestion pricing strategies (5–9). Most of these studies either assumed simplified assumptions for travel time reliability or considered certain scenarios of pricing. Due to inevitable impacts of travel time variability on transportation networks, there is still the need to account for this realistic feature of travelers in congestion pricing schemes, which in turn impacts users' optimal path choice.

Connected and automated vehicle technologies are also expected to significantly contribute to improving mobility and safety. As connected and autonomous vehicles have not been used in practice at large-scale, there are still some uncertainties regarding their applications. Therefore, researchers utilize traffic simulation tools to model the presence of these vehicles. There are several studies on the impacts of vehicle connectivity and automation at the segment level. However, only a few studies have investigated these impacts on traffic flow at the network level. Most of these studies consider a uniform distribution of connected or autonomous vehicles over the network. They also fail to consider the interactions between heterogeneous drivers, with and without connectivity, and autonomous vehicles at the network level. Therefore, the impacts of such technologies on travel time reliability in large-scale networks should be investigated.

All in all, there is a gap in the body of literature to consider travel time reliability and its impacts on users' decisions and on improving or declining traffic mobility. Therefore, this research considers travel time uncertainty in different stages of traffic simulation and studies the impacts of different strategies and technologies on travel time reliability.

1-2- Problem Statement

This study is concerned with three aspects of considering travel time reliability in transportation studies; first, optimal path finding problem and its computational efficiency; second, considering travel time uncertainty and users' heterogeneity in demand control strategies, such as congestion pricing; and third, the impacts of different emerging technologies, such as connected and autonomous vehicle technologies, on travel time reliability. In stochastic networks, deterministic shortest path finding algorithms and heuristics are not applicable due to the non-additivity of link travel times over paths and the non-linear relation between link travel times and path travel times. On the other hand, proposed solution algorithms for this problem do not yield

optimal results within a computationally reasonable time, especially when the network size grows. Overcoming this computational burden requires developing innovative solution approaches, especially for large-scale applications.

The second challenge is to consider travel time reliability in pricing strategies, which in turn affects the paths choices of users. Roadway pricing and subsidizing public transit are common strategies that are used in practice to tackle the congestion problem. However, there are equity concerns associated with pricing strategies and financing issues for public transit subsidies. To overcome these concerns and issues, a self-funded and Pareto-improving pricing scheme should be explored incorporating heterogeneous users with multiple classes of VOT and Value of Reliability (VOR). For a pricing scheme to be self-funded in a bimodal network, the collected tolls from highway users should be distributed among transit users to increase the public transit utility, which is defined as transit-based strategy in this study. Also, the collected tolls can be used to distribute credits among all users (both highway network and transit users) in the network to compensate for any travel time loss, which is defined as credit-based toll distribution strategy. On the other hand, the utility of all network users should be improved after implementation of the pricing scheme to satisfy the Pareto-improving condition. In this study, the utility function of travelers, which determines the path assignment, is defined as a function of the mean travel time, monetary cost of travel (maintenance and fuel cost, fare, toll, etc), and travel time variability capturing the reliability valuation of heterogeneous users. Finding such pricing schemes is important to minimize the total travel time of the auto users, considering heterogeneous users with different VOT and VOR classes in a general bimodal network.

Finally, emerging technologies, such as connected and autonomous vehicle technologies, are expected to increase travel time reliability. The main challenge in quantifying the impacts of

these technologies on traffic congestion and travel time reliability is to have a traffic simulation tool that considers the presence of connected and autonomous vehicles in addition to conventional vehicles. There are several studies on the impacts of vehicle connectivity and automation at the segment level. However, it is challenging to consider the presence of these vehicles in large-scale networks. In addition, network level studies consider a uniform distribution of connected or autonomous vehicles over the network. They also fail to consider the interactions between heterogeneous drivers, with and without connectivity, and autonomous vehicles at the network level. Therefore, another main challenge studied in this manuscript is to consider a non-uniform distribution of connected or autonomous vehicles over the network and the interactions between heterogeneous drivers in a traffic simulation tool.

1-3- Research Objectives and Contributions

The main objective of this study is to consider travel time reliability in traffic simulation and its components, such as optimal path finding problem, to investigate the impacts of different strategies and technologies on travel time reliability and propose reliable strategies to improve mobility in large metropolitan areas. To this end, this study focuses on three different problems: improving computational efficiency of path finding problem in stochastic time-varying networks, incorporating travel time reliability and variability in proposing a reliable pricing strategy for large metropolitan areas, and investigating the impacts of emerging technologies, including connected and autonomous vehicle technologies, on travel time reliability of large-scale networks. The main contributions of this study are as follows:

1-3-1- Path Finding Problem Considering Heterogeneous Users

- Improving computational efficiency of the stochastic shortest path finding problem, considering spatial and temporal travel time correlations and travelers' heterogeneity
- Facilitating the use of stochastic path finding algorithms in cases of route guidance, especially in large-scale applications, by adopting an algorithm which is more adaptive to cases where only a specific Origin-Destination (OD) pair is of interest
- Developing a learning-based network contraction approach to reduce the network size throughout the iterations of the Monte-Carlo simulation-based (MCS) approach for optimal path finding

1-3-2- Travel Time Reliability and Congestion Pricing

- Introducing an algorithm to solve stochastic path finding problems considering the pricing strategies in addition to link travel time correlations for users with different values of time and reliability
- Accounting for the travel time reliability in finding an equitable pricing scheme
- Considering heterogeneity of users in response to travel time uncertainty using multiple classes for users' VOR defined based on the VOT distribution
- Incorporating link travel time correlations in the reliability-based user equilibrium problem for congestion pricing
- Exploring the impacts of different types of relations (i.e. linear, concave) between mean and standard deviation of link travel time on self-funded and Pareto-improving pricing schemes

1-3-3- Impacts of Connected and Autonomous Vehicles on Travel Time Reliability

Developing and incorporating a simulation model that captures the impacts of Connected Vehicles (CVs) and connected Autonomous Vehicles (AVs) on traffic flow and travel time reliability with the following unique features:

- Considering a mixed traffic of Regular Vehicles (HDVs), CVs, and AVs for large-scale applications
- Considering heterogeneous drivers for HDVs and CVs in terms of acceleration behavior
- Considering spatially and temporally varying distributions of HDVs, CVs, and AVs over the network
- Considering capacity variations at intersections in the presence of different shares of CVs and AVs
- Adjusting traffic flow models in arterials due to the presence of CVs and AVs
- Considering CVs and AVs as en-route or adaptive travelers

1-4- Research Approach

The path finding problem is an essential sub-problem of many other transportation problems including the reliability-based user equilibrium problem. A Monte Carlo Simulation-based approach, developed by Zockaie et al. (4,10,11) is used in this study to find the optimal paths for heterogeneous users in stochastic time-dependent networks with spatial and temporal travel time correlations. This approach consists of two stages; the first stage generates a set of candidate paths by applying a dynamic deterministic shortest path algorithm for different realizations of link travel times. This stage consists of a pre-specified number of iterations. In each iteration, realizations of link travel time are made by drawing random numbers from a joint link travel time

distribution. Based on these realizations, a candidate path is found, using a dynamic deterministic shortest path finding algorithm, and then added to the candidate paths set. After this iteration repeats for a certain number of times, the set of candidate paths can be used as the input for the second stage. For each iteration of this stage, a set of link travel time realizations is used to generate the set of candidate paths travel time distributions by summing the realized travel times of the links that exist in that path. Two approaches are proposed in this study to lower the computational time of the reliable path finding problem. The first approach improves the computational efficiency of the stochastic path finding algorithm by conservatively assigning minimum and maximum travel times to all network links and finding pessimistic and optimistic solutions for path travel times. The second approach takes advantage of the generated information during the travel time realizations to find optimistic and pessimistic travel times for each OD pair of interest and extracts a sub-network from the original large-scale network to lower the computational time regarding path finding problem.

In addition, a Particle Swarm Optimization (PSO) algorithm is developed to determine an equitable pricing scheme that minimizes the total travel time of auto users in a general bimodal network considering heterogeneous users with different values of time and reliability. A pricing scheme is considered equitable if it is self-funded and Pareto-improving. A Pareto-improving pricing scheme is developed such that it does not make any traveler worse off (relative to no pricing base case), and makes at least one traveler better off in terms of generalized costs. For self-funded congestion pricing schemes, revenues generated from the collected tolls could be utilized to improve public transportation services, subsidize the users of these services and/or compensate travelers who experience an increase in their generalized travel cost. A reliability-based user equilibrium (RBUE) algorithm is embedded into the PSO algorithm to assign travelers to the

equilibrated paths for different user classes given toll values. Users' heterogeneity in response to the reliability measures, their response to different toll values and toll distribution strategies, and link travel time correlations are considered in this RBUE algorithm.

Finally, this study develops a mesoscopic traffic simulation tool that considers a mixed fleet of vehicles, consists of HDVs, CVs, and AVs. To this end, this study incorporates different microscopic modeling frameworks for various vehicle types (i.e., HDV, CV, AV) and captures the collective effects of the interactions between them on traffic flow dynamics. The stochastic acceleration model of Hamdar et al. (12), which avoids (most) crashes by a perceived probability, is used to model the car-following behavior of HDVs. The acceleration behavior of CVs is modeled based on the Intelligent Driver Model (13), which is a deterministic model. Furthermore, the model of Talebpour and Mahmassani (14) is utilized for the car-following behavior of AVs. In order to translate traffic flow dynamics from a micro-scale to a meso-scale, the relationship between spacing and speed for each vehicle type is derived and used as an input to the mesoscopic model. The proportion of each vehicle type and driver class (if there is any) on each link is tracked in the traffic propagation process at each time interval. Using this proportion, a non-linear equation is solved to obtain the current speed of the link, satisfying the spacing values of all vehicles traversing it. This tool is then used to investigate the impacts of these emerging technologies on traffic flow and travel time reliability at the network level. To explore the network-wide impacts of CVs and AVs on traffic flow, the network-wide fundamental diagrams (NFDs) are generated and compared for different Market Penetration Rates (MPR) of these vehicles with heterogeneous drivers of HDVs and CVs. In addition, the hysteresis loop areas are compared in various scenarios to explore the stability of the system during unloading period. Furthermore, standard deviation of link travel times is used as the reliability measure to compare scenarios with HDVs and AVs in the network.

To produce link travel time variations, 86 scenarios that are different in traffic demand, weather conditions, number of crashes, and percentage of adaptive drivers are used.

1-5- Study Outline

This study is organized as follows. CHAPTER 2 presents a background review on travel time reliability measures, path finding problem, congestion pricing, and the impacts of connected and autonomous vehicles on traffic congestion and travel time reliability. CHAPTER 3 provides an algorithm to extract a sub-network from large-scale networks to lower the computation time of optimal path finding problems in stochastic time-dependent networks. CHAPTER 4 discusses a similar problem to CHAPTER 3 and proposes an iterative learning approach for network contraction. CHAPTER 5 presents an algorithm to determine self-funded and Pareto-improving pricing values minimizing the total travel time of auto users in a general bimodal network considering heterogeneous users with different values of time and reliability. This chapter also investigates the impacts of travel time reliability on toll values given different toll distribution strategies. CHAPTER 6 presents the mesoscopic traffic simulation tool of this study considering the presence of connected and autonomous vehicles on large-scale network. This traffic simulation tool is then used to investigate the impacts of connected and autonomous vehicles on traffic flow and travel time reliability. Finally, CHAPTER 7 summarizes the key findings of this study following by future steps.

CHAPTER 2 - Background Review

2-1- Overview

In this chapter, a comprehensive review of the previous studies on travel time reliability and its applications in different stages of traffic simulation is presented. First, travel time reliability is defined and different measures used in previous studies for this purpose are presented. Then, different reliability rules are presented with applications in the reliable path finding problem. Afterwards, studies on congestion pricing are reviewed with the focus on travel time reliability. Finally, the last sub-section presents multiple studies on the impacts of connected and autonomous vehicles on traffic flow.

2-2- Travel Time Reliability

Travel time uncertainty in transportation networks comes from either the demand side (e.g. special events, fluctuations in demand) or from the supply side (e.g. traffic incidents, work zones, weather conditions) (15). Different users have different responses to this uncertainty based on their personal preferences or risk acceptance levels. Therefore, travel time uncertainty affects the decisions of heterogeneous travelers in the network. In addition, our definition of travel time reliability significantly affects the complexity of the problem and the outcomes. According to Zockaie et al. (15), travel time reliability in transportation studies refers to two concepts; first, network reliability when there is a failure probability for each link of the network. The second one, which is the focus of the current study, is related to travel time uncertainty. In the later definition, link travel time follows a given probability distribution. There is no agreement on the measure of travel time variability in transportation networks (16). Standard deviation of travel time (17–19), the difference between the 90th and 10th percentile of travel time (20), the coefficient of variation

of travel time (21), and probability of being lower than a threshold (22) are common definitions used in the literature for travel time variability.

2-3- Reliable Path Finding Problem

A common way of considering travel time reliability in routing problem is to add a buffer index, representing the uncertainty, to the mean travel time (23,24). Mean and variance are the most important attributes of all random variables. Thus, this model can be considered as a typical mean-variance or a mean-standard deviation model. The mean-variance problem is easier to solve than the mean-standard deviation problem (25). However, both terms have the same units in a mean-standard deviation model; thus, standard deviation is a more intuitive measure of unreliability than variance (26). On the other hand, the standard deviation part is non-linear and non-additive in the path finding problem violating the Bellman's principle of optimality (24). This violation makes the reliable path finding problem difficult to solve especially for large-scale applications.

Many reliability rules are applied in the literature to generate optimal paths to be used for the reliable route guidance. While all reliability rules are consistent in their principal of reducing travel time variability, their optimality conditions may differ from one to another. One of these reliability rules is the Shortest Path problem with On Time Arrival Reliability (SPOTAR), which finds the optimal path with a minimum travel time budget that ensures arriving at the destination on-time with a desired probability. In this problem, for each feasible path and for any travel time budget, there is an associated arrival probability at the destination (11,27–32). Another reliability rule that has been used frequently in the literature is the Minimum Travel Time Budget Path problem (MTTBP), which finds the optimal path with a minimum travel time budget that consists of the mean path travel time and a scaled standard deviation of travel time (24,25,33–39). Several

studies are conducted in the literature to develop algorithms for solving SPOTAR and MTTBP problems (2–4,10,27,28,31,33,36). For example, an MCS approach is presented in (10,11) to approximately solve these path finding problems for static stochastic networks. Zockaie et al. (4) extended this approach to solve both problems for stochastic time-varying networks. The main advantage of this approach is its ability to deal with network-wide correlated link travel times through replications of joint link travel time distributions (10,24,38,39).

2-4- Congestion Pricing Considering Travel Time Reliability

Although road pricing brings about considerable benefits for the government while rectifying congestion externalities, this strategy still has acceptance issues among individual users who would change their route, departure time, or transportation mode due to toll implementation on their desired routes (40). Therefore, equity concerns are the main impediments in the way of road pricing. It is shown in previous studies that marginal cost pricing does not lead to equitable pricing values (41–43). Therefore, to make pricing favorable to the public, transportation researchers introduced equitable congestion pricing schemes that are not only system optimal, but also Pareto-improving (44). In 2010, Lawphongpanich and Yin (44) introduced an anonymous and Pareto-improving toll scheme for a general network. Lawphongpanich and Yin (44) also proved that such schemes may not always exist. A common way of achieving a Pareto-improving scheme is revenue redistribution in a way that offsets the losses of individuals. Therefore, Nie and Liu (43) derived conditions for the existence of a Pareto-improving and self-funded pricing scheme in a bimodal network. In the scheme used by Nie and Liu (43), the toll revenues are evenly distributed among transit users. Nie and Liu (43) also showed that this scheme exists when the users' VOT distribution function is concave. Guo and Yang (45) presented an OD-based Pareto-improving and revenue refunding congestion pricing scheme for a network with multi-class users (discrete

VOTs). However, in reality, OD trip information is not easy to collect. Also, Wu et al. (40) presented a Pareto-improving and revenue-neutral pricing scheme for a multi-modal network. Xiao and Zhang (46) implemented a link-based pricing scheme on a one-origin network which is Pareto-improving and Revenue-neutral. Many studies have also dedicated credits to all travelers. This credit can be used as transit fare, toll payment, or compensation for travel time increases (47–51). The outcome of pricing implementation varies significantly depending on the toll distribution strategy. Therefore, the current study compares different distribution strategies in terms of rectifying the equity issues as well as meeting the market goals.

While most of the studies in the literature on congestion pricing consider homogeneous users or discrete/continuous classes of VOT (41,49,52,53), Van den Berg and Verhoef (54) claimed that the distributional impacts of congestion pricing are controlled by something more than VOT classes. Van den Berg and Verhoef (54) considered a distribution of VOT and a value of schedule delay in a bottleneck model and stated that these two important factors should be considered in congestion pricing assessment studies. Therefore, in addition to value of time, the reliability valuation of users is a critical factor in indicating the welfare losses and gains of heterogeneous users due to congestion pricing. Carrion and Levinson (55) defined the reliability as an amount of money that individuals are willing to pay to reduce the variability of travel time. Tirachini et al. (16) used a mean-variance model to find the optimal pricing considering travel time variability in a bimodal network with transit and personal cars. The study by Tirachini et al. (16) formulates a mode choice model and optimizes the social welfare over the network. However, the study does not consider the user equilibrium as a part of the problem. Zheng et al. (56,57) proposed an adaptive dynamic pricing scheme utilizing the network-wide fundamental diagram to estimate the congestion level. Although predictable travel time variability is considered in the studies by

Zheng et al. (56,57), users are identified only by two VOT classes. Brent and Gross (58) investigated the response of heterogeneous users to high occupancy toll lanes. The study by Brent and Gross (58) highlights the importance of considering VOR in addition to VOT in studying welfare impacts of congestion pricing. Liu et al. (8) investigated the morning peak hour problem considering risk-averse and risk-prone users. Zhu et al. (59) found optimal pricing considering travel time reliability using values of schedule delay for different users. The study by Zhu et al. (59) utilizes the bottleneck in one link to find the optimal pricing incorporating VOR of travelers. All abovementioned studies have simplifying assumptions regarding heterogeneous users. There is also no study that integrates an RBUE, an equitable congestion pricing model, and a system optimal model in one modeling framework. Given the significant importance of travel time variability and reliability measures in the design of an equitable and efficient pricing scheme, there is still a need to account for VOR distributions and travel time variability, in addition to VOT and expected travel times, to fill the gap in the literature.

2-5- Traffic Impacts of Connected and Autonomous Vehicles

The studies on AVs are traced back to the late 20th century when the concept of Automated Highway System (AHS) was introduced. The AHS builds upon the assumption of a fully AV environment on a set of designated lanes (60). This concept was eventually extended to driver assistance systems, such as Adaptive Cruise Control (ACC) (61,62). Development of Vehicle-to-Vehicle (V2V) and Vehicle-to-Infrastructure (V2I) communication systems facilitated improvements of the driver assistance systems. Therefore, later versions of these systems, such as Cooperative Adaptive Cruise Control (CACC), were introduced by incorporating communication technologies. Van Arem et al. (63) were the first who modeled the car-following behavior of CACCs. Several other models are also presented to model the acceleration behavior of AVs and

analyze their impacts on traffic flow at the segment level (64,65). The effects of CV technology on network stability and throughput are also widely studied in the literature (66,67).

While these studies focus on the connectivity and automation of vehicles in a fully connected or automated environment, the transition period from HDVs towards completely connected and automated systems is gaining attention. This period is challenging since HDVs share road space with CVs and AVs, which affects the performance of these vehicles. Therefore, some studies considered the interactions between different types of vehicles in terms of connectivity and automation (68,69). To gain further insights into the future of transportation systems containing these new technologies, large-scale impacts of connectivity and automation at the network level should also be investigated. In one of the most recent studies, Mittal et al. (70) used a mesoscopic simulation approach to investigate flow-density relationship at different MPRs of CVs (traffic mixes with HDVs and only CVs). Mittal et al. (70) used a microsimulation framework developed in (71) to generate speed-density curves for different MPRs of CVs. Using the calibrated speed-density curves as the input of the mesoscopic simulation tool, they investigated the impacts of CV technology on travel time reliability and network throughput at the network level. In this study, Mittal et al. (70) assumed a uniform distribution of CVs throughout the network at any given link and time-interval. However, in reality, these vehicles are not distributed uniformly over all links of the network. In addition, this study does not consider the impacts of CVs on the capacity of intersections or traffic flow models of arterials. Furthermore, the presence of AVs are not considered in the study by Mittal et al. (70).

More recently, Chen et al. (72) conducted a theoretical study on the macroscopic capacity of equilibrated traffic in a mixed traffic condition, with the presence of both AVs and HDVs. In this study, Chen et al. (72) assumed fixed penetration rates of AVs throughout the system. This

assumption might be reasonable in free-flow traffic, while it will be violated when different links face various traffic regimes. Furthermore, Levin and Boyles (73) presented a cell transmission model for shared autonomous and human-driven vehicles that incorporates variations of capacity and backwards wave speed as well as the reaction times of different vehicle types. In this study, Levin and Boyles (73) investigated the impacts of different MPRs of AVs on intersection delay using a modified Dynamic Traffic Assignment (DTA) model. They used a conflict region algorithm to model the intersection movement in the presence of HDVs and AVs (74). While the study by Levin and Boyles (73) provides useful insights into the shared road networks, there is still the need to study a mixed traffic environment that incorporates the interactions between HDVs, CVs, and AVs, with heterogeneous drivers at the network level using real vehicle trajectories and calibrated microsimulation models. In addition, there is no study in the literature that investigates the impacts of CVs and AVs on travel time reliability for large-scale applications, which is one of the objectives of the current study.

CHAPTER 3 - Network Contraction for the Path Finding Problem in Stochastic Networks Considering Heterogeneous Users

3-1- Overview

The path finding problem constitutes an extremely essential problem because of its numerous applications in different fields. Deterministic shortest paths, either dynamic or static, may lead users toward risky paths due to the uncertain nature of travel times in transportation networks. This uncertainty can be caused by several reasons, including but not limited to traffic incidents, natural hazards, work zones, demand fluctuations, and inadequate base capacity (1). Consequently, different users tend to minimize not only their travel times, but also the travel time variability caused by such uncertainties (2). In addition, different travelers have distinct decisions in response to travel time uncertainties that may vary by their preferences and risk tolerance. The same traveler might frequently choose different paths for similar trips based on their departure time and purpose. Therefore, travelers' preferences are expected to be considered in the path selection procedure through different reliability rules (3,4).

Many reliability rules are applied in the literature to generate optimal paths to be used for reliable route guidance. While all reliability rules are consistent in their principal of reducing travel time variability, their optimality conditions may differ from one to another. Of particular interests to this study are two optimality conditions; the first is the SPOTAR, which finds the optimal path with a minimum travel time budget that ensures arriving at the destination on time with a desired probability. The second is the MTTBP, which finds the optimal path with a minimum travel time budget that consists of the mean path travel time and a scaled standard deviation of travel time. The primary objective of both mentioned reliability-based optimal path finding problems is to minimize the risk of arriving late, as well as the travel time budget for different classes of users.

In the SPOTAR problem, for each feasible path and for any travel time budget, there is an associated arrival probability at the destination (11,27–32). Thus, a user's preference leads to an optimal selection of a path and a required travel time budget that certify on-time arrival probability. In the MTTBP problem, the travel time budget for each path includes two main characteristics of that path's travel time distributions: the mean and the standard deviation of travel time. Users could weigh these two components differently, incorporating them into the travel time budget depending on their reliability valuation (24,25,33–39).

Several studies are conducted in the literature to develop algorithms for solving SPOTAR and MTTBP problems (2–4,10,27,28,31,33–36). For example, an MCS approach is presented in (11) and (10) to approximately solve these path finding problems for static stochastic networks. Zockaie et al. (4) extended this approach to solve both problems for stochastic time varying networks. The main advantage of this approach is its ability to deal with network-wide correlated link travel times through replications of joint link travel time distributions. The simulation-based approach, presented in (4), to solve the path finding problem for dynamic stochastic networks, with correlation for multiple classes of users, is considered in this chapter to be improved. The main idea in this approach is to draw a sample from a joint link travel time distribution and store the deterministic shortest path in a candidate set for each realization. Candidate paths are processed further to obtain their travel time distributions, from which the solution to the SPOTAR and MTTBP problems can be determined. The approach consists of two stages that include several realizations of link travel times. The computational burden of this approach is proportional to the sample size, which can be calibrated to achieve a desired solution quality. The major computational time of solving the dynamic path finding problem is attributed to the deterministic shortest path search in the first stage, where the set of candidate paths is found for different samples from the

joint link travel time distribution. Therefore, lowering the execution time of this stage reduces the computational effort of the dynamic stochastic path finding problem significantly. It also makes the algorithm more flexible, allowing the use of more sample sizes in both stages, so as to bring about a desired level of solution quality.

Many studies in the literature are focused on increasing the computational efficiency of the Shortest Path Problems (SPP). Ziliaskopoulos et al. (75) stated the computational efficiency of $nT\varrho^{1.4}$ for all-to-one shortest path problem in actual transportation networks, where n is the number of nodes in the network, T is the number of time intervals, and ϱ is the average in-degree of a node. An important extension of the basic shortest path problem is the All-pairs Shortest Paths Problem (APSPP). One of the most prevalent algorithms to solve the APSPP is the insertion point method, which is extended by (76,77) to aid programming loops by using non-existent links. The repeated single-source approach is another alternative method (75,78). Such an approach could be easily configured for parallel implementation where the computation for each origin/destination could be performed independently. Theoretically, sequential algorithms have a time bound value of $O(n^3)$ for networks with general topologies, where the number of incoming links to each node, in-degree of a node, is significantly high (79).

Despite many decomposition schemes that are devised to solve the all-pairs and all-to-one shortest path problems in parallel (79–83), there is still the need to have a decomposition approach that can be applied to the optimal path finding problem in stochastic networks, especially when the path finding problem needs to be solved only for specific OD pairs. For instance, in the route guidance applications considering the reliability in large-scale networks, computational efficiency plays a critical role. Following the approach proposed in (79), this chapter presents a novel algorithm for network decomposition for each relevant OD pair in stochastic networks. This

approach can also benefit from running shortest path algorithms for different OD pairs of interest in parallel. Taking advantage of the approach in (4) that considers a minimum and maximum value for each link travel time realization, a minimum and maximum bound can be considered for each OD pair travel time distribution, known as the optimistic and pessimistic solutions. To solve the stochastic optimal path finding problem, the novel methodology of this chapter incorporates these bounds to reduce the network size in order to improve computational efficiency. Hence, a sub-network can be extracted for each OD pair by comparing the optimistic and pessimistic solutions. Given that the computational efficiency of the stochastic shortest path algorithms is directly related to the number of nodes in the network (75), the proposed algorithm in this chapter significantly improves the efficiency of these algorithms, especially in large-scale networks. The content of this chapter is published by the author in (84).

3-2- Problem Formulation

A directed and connected network $G(N,A,P)$ is considered, where N is the set of nodes, A is the set of links, and P is the set of probability distributions for links that describes the statistics of link travel times (in case of correlated link travel time distributions, a joint link probability distribution should be considered in the set P). The travel times on the links (denoted as c_{at}) are assumed to be random variables, each of which has a marginal probability density function $p_{at}(\cdot)$. The travel time on path k^{odt} (which connects node o and destination d at departure time t) is denoted as c_k^{odt} . All paths that connect r and s at departure time t form the path set K^{odt} . Also, assume M as the set of different classes of users in terms of reliability valuation. Equations (3-1) to (3-4) formulate the SPOTAR problem as a mathematical optimization:

$$\text{Minimize} \left[\sum_{\forall k^{od} \in K^{od}} \sum_{\forall m \in M} \sum_{\forall \tau \in T} x_{k^{od}}^{\tau m} b_{k^{od}}^{\tau m} \right] \quad (3-1)$$

$$Pr(c_{k^{od}}^\tau \leq b_{k^{od}}^{\tau m}) \geq \alpha_m \quad \forall k^{od} \in K^{od}, \forall m \in M, \forall \tau \in T \quad (3-2)$$

$$\sum_{\forall k^{od} \in K^{od}} x_{k^{od}}^{\tau m} = 1 \quad \forall m \in M, \forall \tau \in T \quad (3-3)$$

$$x_{k^{od}}^{\tau m} = 0 \text{ or } 1 \quad \forall k^{od} \in K^{od}, \forall m \in M, \forall \tau \in T \quad (3-4)$$

where α_m is the required level of reliability for class m , $c_{k^{od}}^\tau$ is the path travel time random variable following a specific distribution (considering known link distributions), $b_{k^{od}}^{\tau m}$ is the minimum required budget for path k^{od} to ensure a reliability level of α_m at departure time interval τ , and $x_{k^{od}}^{\tau m}$ is a binary variable to select the optimal path among existing paths in the main path set (K^{od}) for user class m at departure time interval τ .

The objective function of MTTBP problem in path-based variables is also as follows.

$$\text{Minimize} \left[\sum_{\forall k^{od} \in K^{od}} \sum_{\forall m \in M} \sum_{\forall \tau \in T} (c_{k^{od}}^\tau + \lambda_m \sigma_{k^{od}}^\tau) x_{k^{od}}^{\tau m} \right] \quad (3-5)$$

$$\sum_{\forall k^{od} \in K^{od}} x_{k^{od}}^{\tau m} = 1 \quad \forall m \in M, \forall \tau \in T \quad (3-6)$$

$$x_{k^{od}}^{\tau m} = 0 \text{ or } 1 \quad \forall k^{od} \in K^{od}, \forall m \in M, \forall \tau \in T \quad (3-7)$$

where, λ_m is the weight of the standard deviation of travel time in the objective function for user class m . More details about the SPOTAR and MTTBP problems can be found in (4). This chapter aims to develop a decomposition methodology applicable to the dynamic path finding problem in stochastic networks that solves these two problems with the same accuracy and less execution time for specific OD pairs.

3-3- Methodology

In this sub-section, a decomposition approach is presented to extract a sub-network for each specific OD pair. This algorithm also benefits from executing the shortest path algorithm for different OD pairs in parallel. The modified MCS based approach of (4) for time-dependent stochastic networks with multiple classes of users is utilized in this chapter to solve the SPOTAR and MTTBP problems. The new algorithm takes advantage of the structure of this approach to find optimistic and pessimistic travel times for each OD pair of interest. For a certain OD pair, the minimum link travel time is considered for all network links and the travel times of both backward star and forward star shortest path problems are solved resulting in optimistic travel times from the origin to the destination passing through any node in the network. Next, all network link travel times are set to their maximum values, and the pessimistic travel time is calculated from the origin to the destination using the deterministic static shortest path algorithm. Comparing the optimistic travel time through any node to the pessimistic travel time for the OD pair specifies if the node should be retained in the sub-network or not. Then, realizations of link travel times are made by drawing random numbers from joint link travel time distributions for different iterations of the first stage. The deterministic dynamic shortest path algorithm finds a candidate path based on the realized link travel times for the sub-network. An all-to-one dynamic shortest path algorithm (75), which is based on the Bellman's principle of optimality, is used to solve the dynamic SPP in this stage. By repeating these iterations, the set of candidate paths can be prepared as the input for the second stage. The second stage also consists of a pre-specified number of iterations, including link travel time realizations, to generate path travel time distributions for the candidate paths. The steps of the algorithm are as follows.

3-3-1- Solution Algorithm

Step 1 Initialization

Form the joint time-dependent link travel time distribution, considering temporal and spatial correlations, and obtain the OD pair of interest as the main input.

Step 2 Find optimistic and pessimistic travel times

Step 2-1 Solve the forward star static shortest path problem from the origin to all other nodes using the minimum link travel times and store travel times as the optimistic travel times from the origin to all other nodes.

Step 2-2 Solve the backward star static shortest path problem from all nodes to the destination node using the minimum link travel times and store the optimistic travel time from all nodes to the destination.

Step 2-3 Solve the backward star static shortest path problem from all nodes (including the origin) to the destination using the maximum link travel times and store the pessimistic travel time from the origin to the destination.

Step 3 Extract a sub-network for the OD pair of interest by checking the condition in Step 3-2 for each network node

Step 3-1 Add the optimistic travel time from the origin to the node (from Step 2-1) to the optimistic travel time from the node to the destination (from Step 2-2), and store it as the optimistic travel time from the origin to the destination.

Step 3-2 If the optimistic travel time of the OD pair, (from Step 3-1) is larger than the pessimistic value (from Step 2-3) remove the node from the sub-network.

Step 4 Stage one of stochastic path finding

Draw a random sample of size L_1 from the joint time-dependent link travel time distribution. Set the current sample index as $l_1=1$.

Step 4-1 Set the link travel time on link a at time t as $c_{at}^{l_1}$, which is the random number in the l_1^{th} sample corresponding to link a at time interval t .

Step 4-2 Find the dynamic deterministic shortest paths from all nodes in the sub-network (Step 3) including the origin to the destination for all departure time intervals.

Step 4-3 For all departure time intervals, add the newly found path to the set of candidate paths of that departure time interval ($K^{od\tau}$) if it has not been found previously.

Step 4-4 Set l_1 as l_1+1 . If $l_1 \leq L_1$, then go to Step 4-1; else go to Step 5.

Step 5 Stage two of stochastic path finding

Draw a random sample of size L_2 from the joint time-dependent link travel time distribution. Set the current sample index as $l_2=1$.

Step 5-1 Set the link travel time on link a at time t as $c_{at}^{l_2}$, which is the random number in the l_2^{th} sample corresponding to link a at time interval t .

Step 5-2 For all departure time intervals such as τ , and all paths in the set of candidate paths of that departure time interval such as $k^{od\tau} \in K^{od\tau}$, find $\Pi_{k^{od\tau}} = \sum_{a \in k^{od\tau}} c_{at}^{l_2}$ as the path travel time, where t is the arrival time of link a based on the travel time of the previous link in the path $k^{od\tau}$. Set $\Pi_{k^{od\tau}}$ as $D(k^{od\tau}, l_2)$.

Step 5-3 Set l_2 as l_2+1 . If $l_2 \leq L_2$, then go to Step 5-1; else go to Step 6.

Step 6 Finding the optimal path for SPOTAR (for MTTBP go to Step 7)

Step 6-1 Sort D for each $k^{od\tau}$ in an increasing order and save the sorted set as $F_{k^{od\tau}}$, which is an estimation of the inverse of the cumulative distribution function and set $i=1$.

Step 6-2 For all departure time intervals τ , set

$$\Pi_{od\tau}^*(i) = \min_{k^{od\tau} \in K^{od\tau}} F_{k^{od\tau}}(i) \text{ and } k_{od\tau}^*(i) = \operatorname{argmin}_{k^{od\tau} \in K^{od\tau}} F_{k^{od\tau}}(i)$$

Step 6-3 Set i as $i+1$. If $i \leq L_2$, then go to Step 6-1; else go to Step 6-4.

Step 6-4 Report $\Pi_{od\tau}^*$ and $k_{od\tau}^*$ as the optimal path travel time and Pareto frontier path.

Step 6-5 For all user classes m , find the m^* based on the on-time arrival probability and store $U_{\tau c}^{*od} = \Pi_{od\tau}^*(m^*)$ and $p_{\tau c}^{*od} = k_{od\tau}^*(m^*)$.

Step 7 Finding the optimal path for MTTBP

For all departure time intervals such as τ do the following steps.

Step 7-1 Retrieve the first non-selected path $k^{od\tau}$ from $K^{od\tau}$ and remove it from that set.

Step 7-2 Calculate the path mean travel time as $C(k^{od\tau}) = \sum_{i=1}^{L_2} D(k^{od\tau}, i) / L_2$.

Step 7-3 Calculate the path standard deviation of travel time as

$$Std(k^{od\tau}) = \sqrt{\sum_{i=1}^{L_2} [C(k^{od\tau}) - D(k^{od\tau}, i)]^2 / L_2}.$$

Step 7-4 Calculate the objective function of the path for all user classes such as m with reliability valuation of $\lambda(m)$ as follows:

$$\Pi_{\tau m}^{od} = C(k^{od\tau}) + \lambda(m) \times Std(k^{od\tau}); \text{ If } \Pi_{\tau m}^{od} < U_{\tau m}^{*od}, \text{ set } U_{\tau m}^{*od} \text{ equal to } \Pi_{\tau m}^{od} \text{ and } p_{\tau m}^{*od} = k^{od\tau}.$$

Step 7-5 If $K^{od\tau}$ is empty, report $p_{\tau m}^{*od}$ and $U_{\tau m}^{*od}$ as the optimal solution and go back to Step 1; else go to Step 7-1.

In the next sub-section, this approach is implemented for the Chicago downtown network using a realistic joint time-dependent travel time distribution for the SPOTAR and MTTBP problems. If all OD pairs are of interest, step 1 through step 7 can be processed in parallel for different OD pairs. However, for only one OD pair, the approach is implemented without parallelization.

3-4- Numerical Experiments

A set of experiments is presented here to validate the accuracy and examine the computational performance of the algorithm developed in this chapter. The performance of the algorithm is compared with that of the algorithm presented in (4), which finds all-to-one shortest path trees for multiple destinations in each realization of the first stage without network contraction. An Intel® Xeon® CPU E5-2643 processor is used for all runs presented in this chapter, which includes 12 threads of 3.4 GHz and 128 GB of shared RAM. A maximum of ten threads are allocated for executing the algorithm.

3-4-1- Case Study Specifications

The Chicago downtown network, which is bound from West and East by O'Hare airport and Lake Michigan, respectively, is considered as the case study area of this chapter. This network includes downtown Chicago and some Western and Northern suburban cities of Chicago, and contains 1,578 nodes, 4,805 links, and 218 zones. To generate link travel time distributions and correlations of the AM peak period (5:00 AM to 10:00 AM), the network is simulated for different scenarios using the mesoscopic simulator of DYNASMART-P (85). This network is calibrated using the OD-estimation techniques presented in (86).

Based on real-world observations (i.e., total observed demand by loop detectors, weather conditions, and incidents), 86 scenarios are created and simulated in (15,87). Five hours of the simulation horizon are divided into 20-minute time intervals to create link travel time distributions for 15 departure time intervals. Aggregating travel time observations over one minute results into 1720 observations (86 scenarios and 20 time intervals) for each link and departure time interval. The mean and variance of travel time for each link are considered as the representatives of the link travel time distribution. Figure 3-1 depicts the configuration of the Chicago network. Figure 3-2

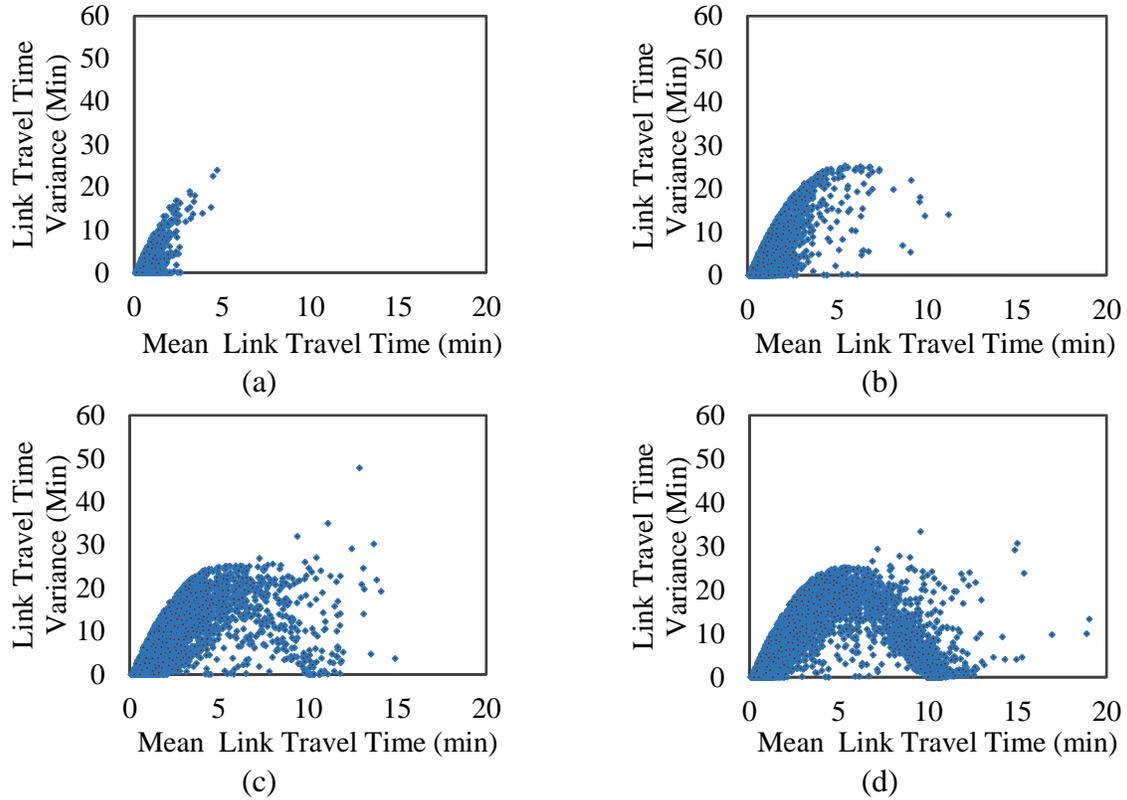


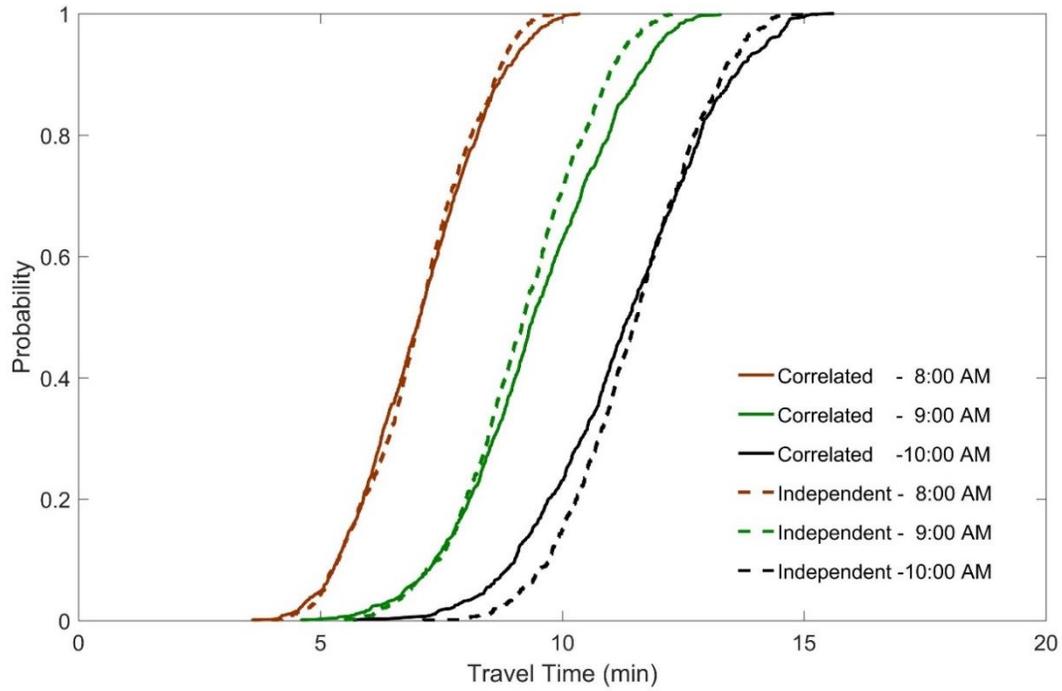
Figure 3-2- Mean vs. variance of travel time for different links of the network at different time intervals of the a) first, b) fifth, c) tenth, and d) fifteenth (last) 20-minute time interval (15)

3-4-2- Numerical Results

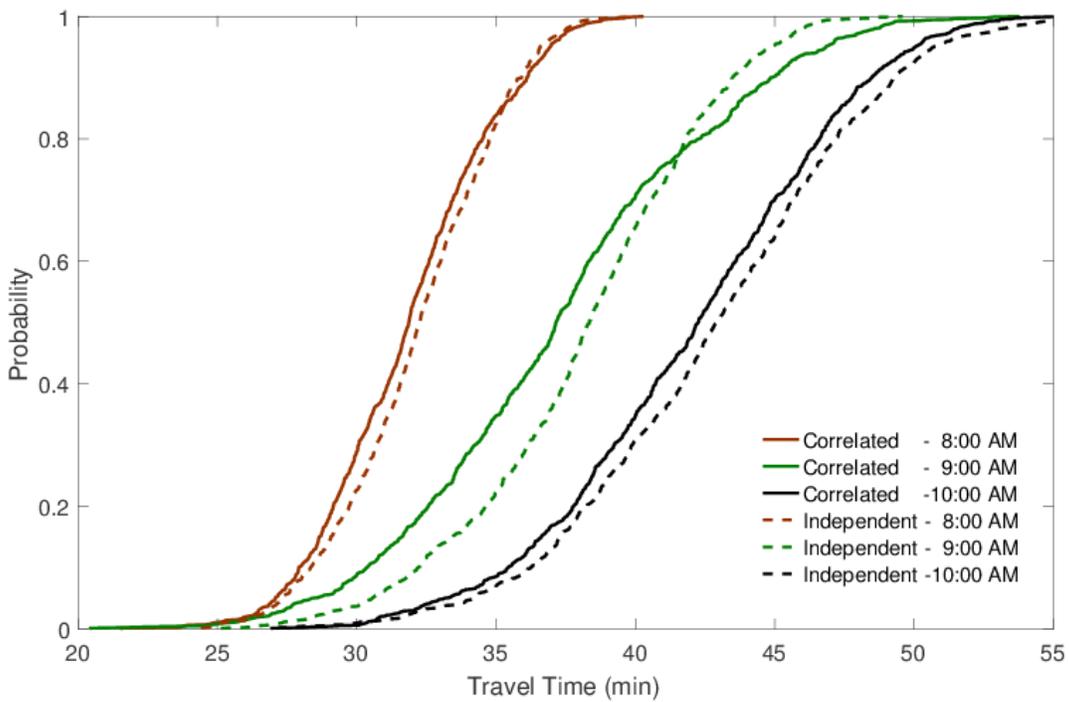
In order to benchmark the applicability of the proposed contraction method in this chapter for the stochastic path finding calculations, the algorithm is first applied to specific OD pairs. Different ranges for the upper bound (pessimistic value) and lower bound (optimistic value) of link travel times are considered. Then, the validity of the results is tested by comparing travel time distributions of the optimal paths using the proposed method with that of the algorithm in (4). The results show that optimal paths and their travel time distributions are exactly similar to those of the regular approach without contraction.

Figure 3-3 shows the travel time distributions for two random OD pairs, one with a small distance (Figure 3-3a), and one with a longer distance (Figure 3-3b), for the SPOTAR problem. Note that the distributions are found and compared both for correlated and independent (non-

correlated) link travel times. Moreover, for the temporal correlations, only consecutive time intervals are assumed to be correlated, and for the spatial correlations, the neighborhood order is defined as two. The neighborhood order of two means that each link is correlated to all other links that can be reached from that link by traversing at most two links. Given that the travel time distributions are the same for both proposed approach with contraction and the one without contraction, only one distribution is shown for each OD-time interval set and type of link travel time correlation in Figure 3-3. Medium reliability valuation (0.7 as the on-time arrival probability for SPOTAR) is considered for the distributions of Figure 3-3. As it can be seen from Figure 3-3a and Figure 3-3b, both the mean and standard deviation of the optimal path travel time are increasing during the simulation. This increase means that the network-wide traffic congestion evolves through the network during the simulation horizon. Furthermore, the travel time distributions of origin 196 to destination one at all departure times contain wider ranges of travel times than those of origin 36 to destination one, which is expected due to the longer distance between origin 196 and destination one than origin 36. It implies that the variability of travel times is more for longer distances during the AM peak hours. This trend is more obvious for the 9:00 AM departure time. Note that ignoring the link travel time correlations results in different optimal path distributions, which shows the importance of capturing this factor in the path finding algorithms.



(a)



(b)

Figure 3-3- Optimal path travel time distributions at different departure times with correlated and independent distributions and for both with and without network contraction (on top of each other) a) from origin 36 to destination 1; b) from origin 196 to destination 1 (Pessimistic and optimistic limits of 5 and 0.8, 100 stage 1 iterations, and 1000 stage 2 iterations)

The accuracy of the proposed method in terms of the optimal path travel time distributions is discussed above. Here, the computational performance of the proposed algorithm is compared with the developed approach in (4). The execution times of the stochastic dynamic path finding algorithm with contraction for SPOTAR and MTTBP problems are depicted in Figure 3-4a for different OD pairs and all departure time intervals. Note that the destination is fixed to node 1 in all of these OD pairs. An increasing trend can be observed as the OD pair has a longer distance since the sub-network size would be larger. The maximum value of the execution time for origin 179 is 162.31 seconds for the SPOTAR problem and 151.28 seconds for the MTTBP problem, which are considerably lower than the execution times observed in the regular method. The execution time of the stochastic path finding algorithm for all origin nodes to destination one, using the approach developed in (4) without considering network contraction, is equal to 5,049 seconds and 1,540 seconds for SPOTAR and MTTBP problems, respectively. This significant difference is due to the fact that the approach in (4) needs to find optimal stochastic paths for all origins, while the proposed method in this chapter only requires the calculation for one origin. It can also be observed in Figure 3-4a that the execution time of the nearest origin to the destination (origin 2) is approximately a quarter of the furthest origin (origin 179) in both problems (i.e., SPOTAR, MTTBP) after network contraction.

Figure 3-4b illustrates the average execution times for the deterministic shortest path calculation over stage one iterations for the same OD pairs as Figure 3-4a. The number of stage one iterations is set to 100 for all runs in this figure. The deterministic shortest path finding algorithm should be executed to the number of iterations assigned to stage one; hence, the execution time for this step is a proper measure in understanding the impacts of network contraction on the overall performance. The same increasing trend of Figure 3-4a can be observed

in this figure. However, the network contraction seems to have greater impacts on the dynamic deterministic shortest path finding, since the deterministic path finding algorithm is applied to the extracted sub-network at this stage. On average, the full network application of the approach developed in (4) requires 0.85 seconds of execution time for the deterministic shortest path algorithm. As can be seen in Figure 3-4b, the network contraction reduces this execution time for the nearest origin (node 2) to 0.01 seconds.

Optimistic (Opt) and pessimistic (Pess) bounds are critical factors in the performance of the proposed algorithm as they affect the number of nodes remaining in the sub-network. Therefore, different ranges of optimistic/pessimistic coefficients are tested in this chapter and the results are shown in Figure 3-4c. These coefficients should be multiplied by free flow travel times of OD pairs to generate the upper and lower bounds for the network contraction thresholds. Note that if the same thresholds are used as the minimum and maximum link travel times, the accuracy of the solution relative to the solution without contraction is guaranteed. However, smaller ranges might affect the accuracy of the algorithm while improving execution times. Figure 3-4c shows the impacts of the OD pair distance and pessimistic/optimistic coefficients on the size of the sub-network. The shorter distance and the smaller range of pessimistic/optimistic thresholds lead to a smaller sub-network size, which saves more execution time by implementing the proposed methodology. Note that the accuracy of the proposed method is not affected by reducing the pessimistic/optimistic bound.

Figure 3-5 shows sub-networks for two random OD pairs. The sub-networks contain a large number of nodes when the optimistic/pessimistic coefficients are 0.8 and 5 (same factors for minimum and maximum link travel times). The sub-network includes far fewer nodes when the distance between the OD pair is small or the optimistic/pessimistic coefficients are 1 and 2.5. It

can also be seen in Figure 3-5a that freeway nodes are included in the sub-network due to their lower travel times despite the greater distances.

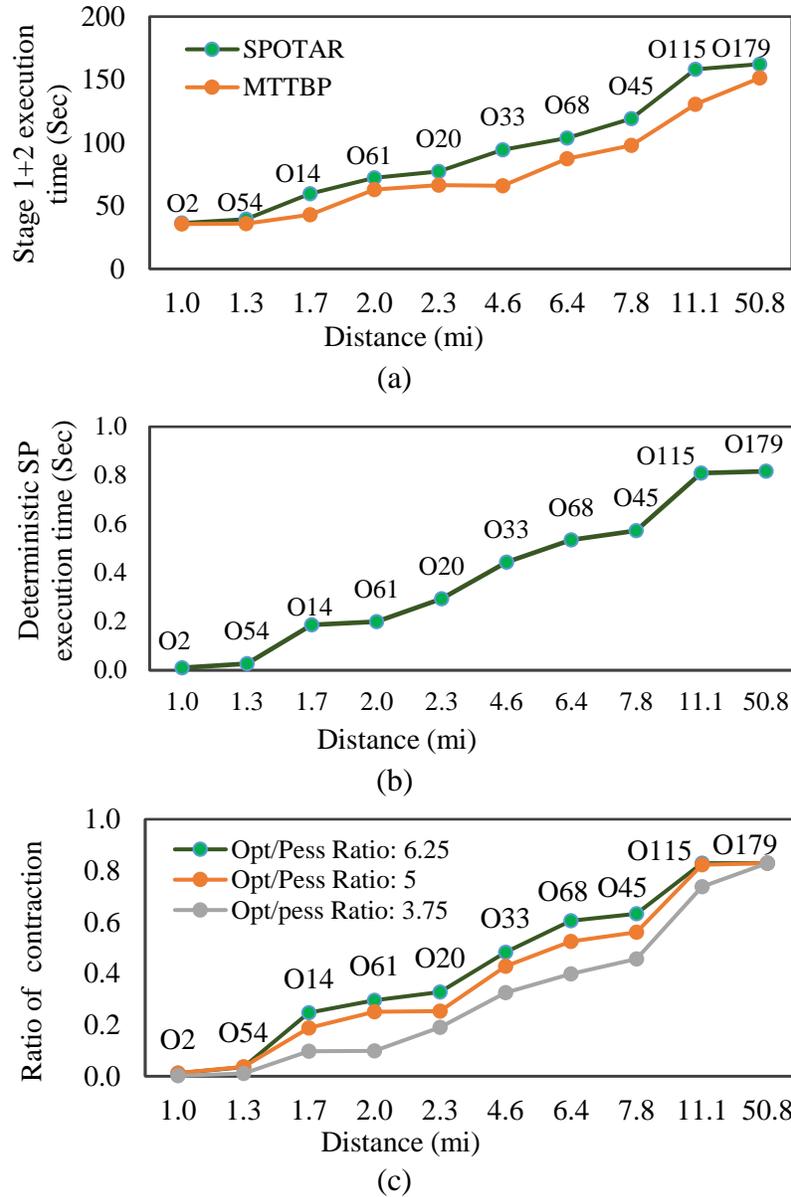


Figure 3-4- Computational performance for different origins to zone 1 a) execution times of the stochastic path finding algorithm with contraction; b) dynamic deterministic shortest path execution times; c) contraction ratio for different optimistic/pessimistic limits

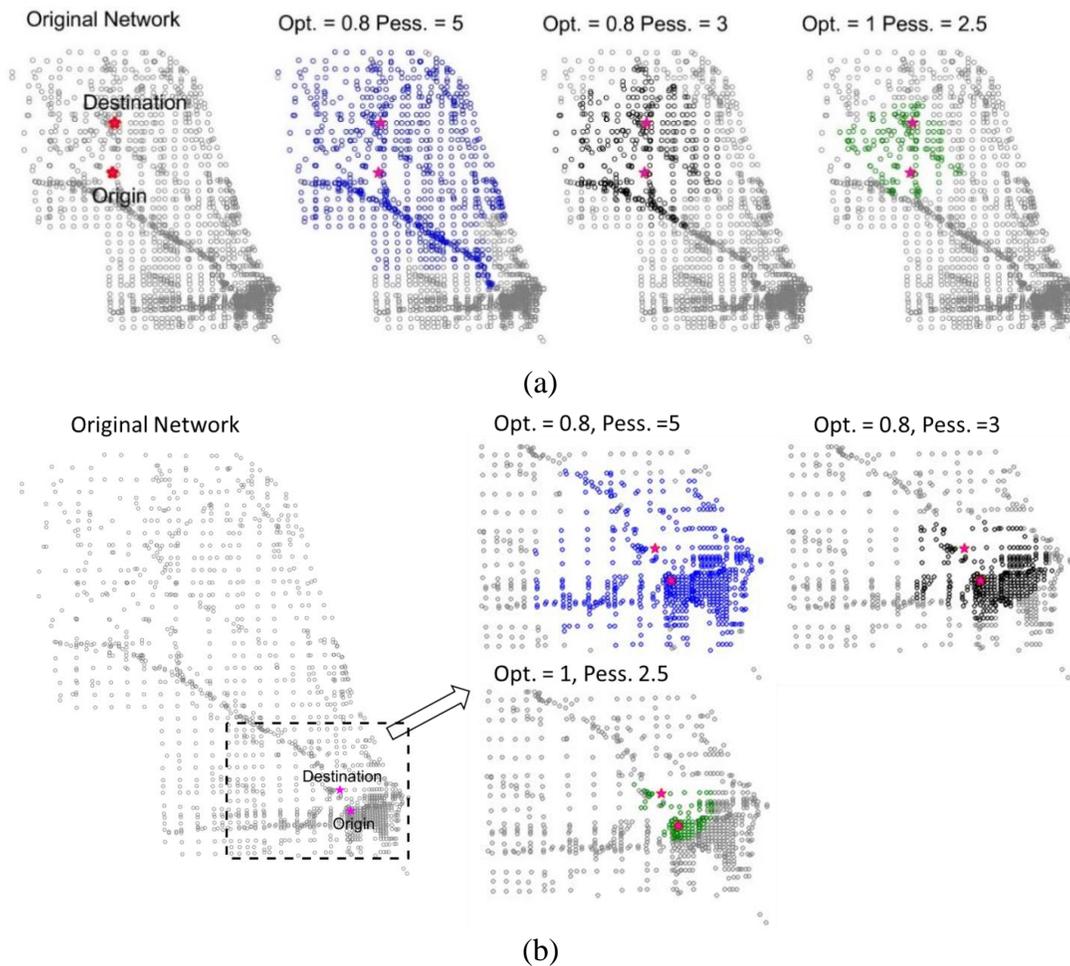
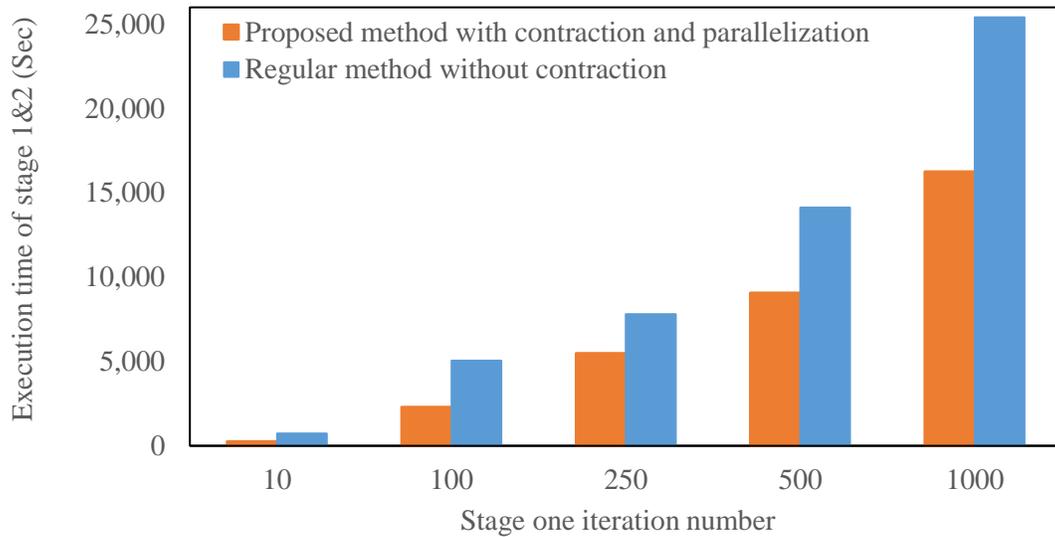


Figure 3-5- Sub-network configurations for different optimistic/pessimistic coefficients a) Origin zone 100 to destination 87; b) Origin zone 20 to destination one

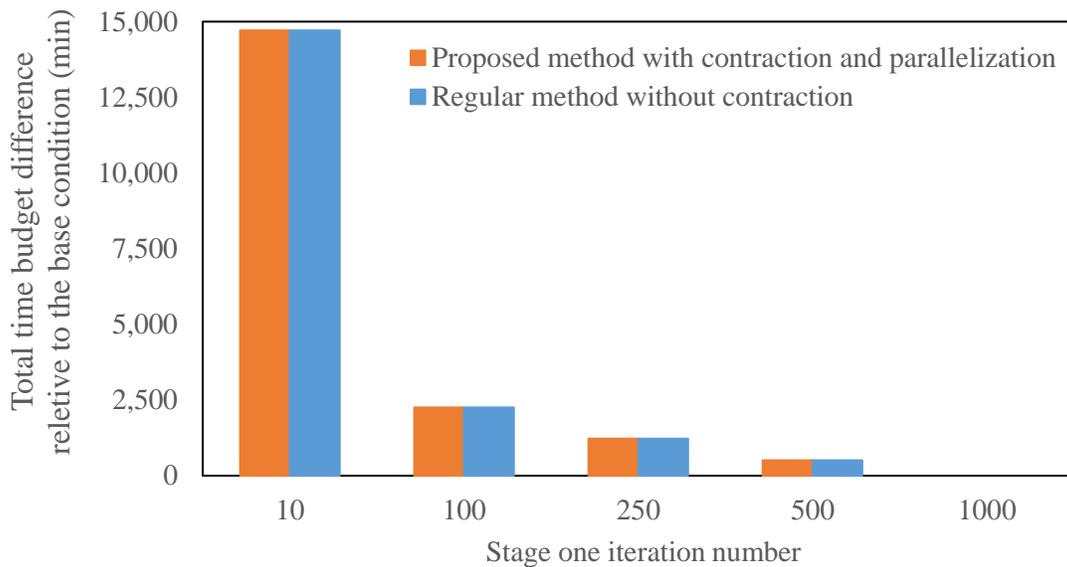
In addition to the improvement in the execution time associated with the network contraction, the decomposition of the path finding problems at the OD pair level facilitates parallelization and using multiple processors to gain execution time savings. To evaluate the performance of the proposed algorithm, the algorithm is applied for all origins to destination one in parallel. Comparing the nodes and travel time distributions of optimal paths for different OD pairs, departure times, value of reliability of different user classes, and various optimistic/pessimistic limits validates that the algorithm performs with a great accuracy even when smaller pessimistic/optimistic bounds are used. Comparing the execution times of the stochastic

dynamic path finding calculation for all origins to destination one shows that the execution time required for stochastic path finding of the proposed algorithm in this chapter is lower than that of the regular method for all optimistic/pessimistic limits. This improvement is much more significant for the ratio of 2.5.

In Figure 3-6, the number of iterations considered for stages one and two in all runs is set to 100 and 1000, respectively. Given that more link travel time realizations improve the accuracy of results, different numbers of stage one iterations are also proposed. In this figure, the results of the stochastic path finding algorithm implemented with contraction and parallelization are compared with the ones resulted from the proposed method in (4). Figure 3-6a illustrates the execution times of both methods for different iterations of stage one. As shown in this figure, the difference between execution times of both methods increases with an increase in the number of iterations of the first stage. The SPOTAR time budget, calculated in both the literature and proposed method, is also presented. The errors of the proposed algorithm in comparison with the literature method are almost zero over all OD pairs (with a fixed destination). In Figure 3-6b, it is assumed that the base condition is the run with 1000 stage one iterations. The cumulative relative differences of calculated time budgets in comparison with the base condition, for 218 origins to destination one at 15 time intervals and three different VOR user classes, are demonstrated in Figure 3-6b. Based on the results of this figure, large differences can be observed in time budgets of lower numbers of stage one iterations. Taking advantage of the computational improvements of the proposed method, illustrated in Figure 3-6a, a higher number of iterations can be used to improve the error term while keeping the execution time at the same level.



(a)



(b)

Figure 3-6- SPOTAR a) execution time and b) cumulative difference of objective function values relative to the stage one iterations of 1000 for the proposed method with contraction and parallelization and regular method without contraction (Destination one and optimistic/pessimistic coefficients of 0.8 and 5)

3-5- Summary

This chapter focused on improving the computational efficiency of the dynamic path finding algorithm in stochastic networks using a network contraction approach. Given that link travel time distributions have an upper and a lower bound, a pessimistic/optimistic range can be

defined for each OD pair. If the optimistic travel time for a specific OD pair is lower than the pessimistic value, the tested node remains in the sub-network; otherwise, it will be eliminated. Using this approach, the network is contracted for each OD pair of interest. Then, stage one and stage two of the MCS approach, presented in this chapter, are solved for the selected OD pair. Furthermore, stochastic path finding algorithms can be processed in parallel for different OD pairs in case all origins are of interest.

The proposed algorithm is applied to the real-world large-scale network of Chicago to evaluate its performance and analyze the sensitivity to different factors, such as distance and pessimistic/optimistic bounds. The results show that optimal paths and their travel time distributions are exactly similar to those of the approach without contraction. The computational gain of the algorithm has an inverse relation with the distance and range of the optimistic/pessimistic bounds. Smaller bounds relative to the minimum and maximum link travel times cannot guarantee the same accuracy. However, the results show that smaller bounds do not affect the accuracy and improve the execution time significantly. Finally, the proposed algorithm in this chapter leads to great improvements in execution times of dynamic stochastic path finding algorithms, especially when the algorithm should be applied to specific OD pairs. Even when all OD pairs are of interest, the proposed approach can be applied for different OD pairs in parallel and considerable gains are achieved. Hence, deterministic shortest path finding algorithm can be applied for more realizations of travel times in stage one of the dynamic stochastic path finding algorithm by utilizing the network contraction approach. A higher number of realizations leads to a more accurate and time-efficient stochastic path finding algorithm.

CHAPTER 4 - An Iterative Learning Approach for Network Contraction

4-1- Overview

Due to the recent developments of Intelligent Transportation Systems (ITS), introducing an efficient optimal path finding algorithm has attracted the attention of researchers and service providers. Finding optimal paths in a limited time is indeed a common requirement of these applications such as in-vehicle routing guidance systems and vehicle routing problems. As mentioned in the previous chapter, the computational burden of the path finding algorithms has a polynomial growth order with respect to the network size (number of nodes and link), which limits real-time applications of these algorithms in large-scale networks. This computational burden has led to conducting many studies to improve the efficiency of path finding algorithms using heuristic approaches based on artificial intelligence techniques (88,89). These approaches take advantage of available prior knowledge to reduce the search effort, which makes them superior to non-informative labeling algorithms. For example, traditional approaches search all intermediate nodes between origin and destination, regardless of their probability to be in the resulting optimal path. However, heuristic approaches use the probability of a node existing in the optimal path as prior information to limit the examined nodes in the search list.

There are multiple categories of heuristic shortest path finding algorithms in the literature. A* search, in which the scan eligible list order is prioritized, is used in many heuristic path finding algorithms (90,91). These algorithms estimate the probability of being in the shortest path by defining an evaluation function for each node. Alongside the algorithmic estimation, a merit is assigned to each node of the network, with current evaluated travel time from origin plus the estimated travel time to the destination (mostly approximated based on the Euclidean distance). The efficiency of these algorithms depends extremely on the quality of the estimation. The branch

pruning approach is another method to limit the search area for the shortest path calculation (92,93). The approach is quite similar to the A* algorithm with the difference being the low priority nodes in this method are pruned from the search list and will never be scanned. A major limitation of this approach is the potential for termination without finding the optimal path. Furthermore, several decomposition-based algorithms are suggested, in which large-scale networks are decomposed into several small sub-networks considering the polynomial order of the shortest path algorithms (79–81,94).

Despite the computational improvements associated with these heuristics, they are limited to deterministic problems, which ignore the uncertain nature of travel times in transportation networks. That in mind, the motivation of this chapter is to propose an approach readily applicable to stochastic time-dependent networks with a reasonable computational time. Given that the time required to solve the optimal path finding problem increases with the increase in the network size, the primary contribution of this chapter is to find an efficient sub-network for any given OD pair, extracted from the original network, in order to improve computational efficiency of the stochastic path finding problem. This chapter intends to use the existing information of stochastic dynamic path travel times to extract a sub-network for each OD pair without imposing extra computational costs. To this end, we incorporate the specifications of link travel time distributions to find such sub-networks based on the comparison of pessimistic and optimistic solutions. The condition for keeping a node in the sub-network is that the optimistic travel times through the node should be lower than the pessimistic travel time for the OD pair. The pessimistic and optimistic solutions for paths travel times can be obtained conservatively by assigning minimum and maximum travel times to all network links, which is explained in the previous chapter (84). Be that as it may, the assumption of all links in the network being at their worst/best operational state becomes less

probable and too conservative, especially as the distance between origin and destination increases. Hence, effort is motivated to deriving realistic and efficient bounds for the pessimistic and optimistic solutions that determine the sub-network size.

The proposed approach estimates the upper and lower bounds for the pessimistic and optimistic solutions used to determine if a node could be considered in the sub-network. These bounds are updated as more information on the travel time distributions becomes available with more iterations. Once these bounds are determined, they are strictly applied as nodes, with labels falling within these boundaries considered to be part of the obtained sub-network. It is worth mentioning that, even though these bounds are strictly applied in the solution algorithm, they are only heuristic estimates to improve computational efficiency.

In this context, this chapter presents a learning approach that incorporates link travel time realizations within initial iterations of the simulation-based procedures to find efficient optimistic and pessimistic bounds. At each iteration, link travel time realizations lead to a label realization for each node. The distributions of the realized labels can be used to estimate the efficient bounds that reduce the network size for deterministic shortest path calculations in the later iterations of the same procedure. The learning approach is developed and calibrated using realistic travel time distributions obtained for the large-scale network of the extended downtown area of Chicago. The transferability of the approach to other networks then comes under consideration using another large-scale network representing the roadway network of Salt-Lake City. The Chicago network contains approximately 1,600 nodes and 5,000 links; the Salt-Lake City network includes around 3,600 nodes and 8,300 links. The optimal path objective functions and the number of nodes in the obtained sub-network, as an indicator of the required execution time, are compared for the cases

with and without network contraction. The content of this chapter is published by the author in (95).

4-2- Problem Statement

In stochastic networks, deterministic shortest path finding algorithms and heuristics are not applicable due to the non-additivity of link travel times over paths and the non-linear relation between link travel times and path travel times. On the other hand, proposed solution algorithms for this problem do not yield optimal results within a computationally reasonable time, especially when the network size grows. In the path finding problems, only a part of the entire network is relevant to the optimal path between a certain origin and destination. Thus, this chapter aims to develop a network contraction approach to reduce the network size throughout the iterations of MCS approaches. The methodology aims to be applicable to dynamic path finding problems in stochastic networks considering spatial and temporal travel time correlations for heterogeneous travelers in terms of reliability valuation.

Similar to the previous chapter, this chapter considers a directed and connected network $G(N, A, P)$, where N is the set of nodes, A is the set of links, and P is the set of probability distributions for links that describe the statistics of link travel times. c_{at} is the travel time on link a at arrival time interval t , which is assumed to be a random variable. All paths that connect o and d at departure time τ form the paths set $K^{od\tau}$. The travel time on path $k^{od\tau}$ (which connects node o and destination d at departure time τ) is denoted as $c_{k^{od\tau}}^\tau$. Each origin and destination pair, od , has a maximum travel time $C_{pess}^{o \rightarrow d}$, which is the maximum travel time of all candidate paths up to the current point of simulation connecting origin o and destination d . $C_{opt}^{o \rightarrow i}$ and $C_{opt}^{i \rightarrow d}$ are also minimum travel times of all paths from origin, o , to any node, i , and any node, i , to destination, d . Furthermore, let $C_{opt}^{o \rightarrow i \rightarrow d}$ be the optimistic travel time from origin o , to destination d through node

i. Each internal node has also a free flow travel time (FFTT) from origin, $FFTT_{o \rightarrow i}$, and a free flow travel time to the destination, $FFTT_{i \rightarrow d}$. Also, assume M as the set of different classes of users in terms of reliability valuation.

4-3- Methodology Development

Given that in reliable path finding problems, each path has a minimum travel time and a maximum travel time, comparing the optimistic travel time through any node to the destination, with the pessimistic travel time for the OD pair specifies if the node should be retained in the sub-network or not. Figure 4-1 illustrates the schematic view of this optimistic/pessimistic travel time comparison. The optimistic travel time from the origin to the destination passing through any node can be calculated by considering the minimum travel time for each link in the network. To this end, both one-to-all and all-to-one deterministic static shortest path trees are found for all network nodes. Then for each node, the optimistic travel time can be measured as the sum of the travel time from the origin of interest to the node, and the travel time from the node to the desired destination. Next, all network link travel times are set to their maximum values, and the pessimistic travel time is calculated from the origin to the destination using the deterministic static shortest path algorithm. These two optimistic and pessimistic travel times are then compared to make a decision about each network node (84). However, the probability of all network links being at their maximum or minimum travel time at the same time is low, especially when there are many links in the optimal path. Furthermore, there is no other network contraction approach developed in the literature to be implemented efficiently on any general network, with different configurations and characteristics, for solving optimal path finding algorithm in stochastic dynamic networks. Therefore, this chapter presents a learning approach to derive efficient optimistic and pessimistic bounds. The learning approach makes use of the generated information within the initial iterations of the simulation-

based approach, as previously proposed in the literature (e.g., Zockaie et al. (4)) to solve the path finding problem in stochastic networks. The main purpose of the MCS approach is to produce a set of candidate paths by solving a dynamic deterministic shortest path problem. To ensure obtaining a comprehensive set of candidate paths, the process repeats for a pre-specified number of iterations, where each iteration uses a realization of travel time obtained from the joint link travel time distribution. To speed up this step, the approach of this chapter aims at reducing the network size in the later iterations based on information learned from the completed iterations. The generated labels for each node from the previous iterations are used to define new optimistic and pessimistic bounds that reduce the network size for later iterations.

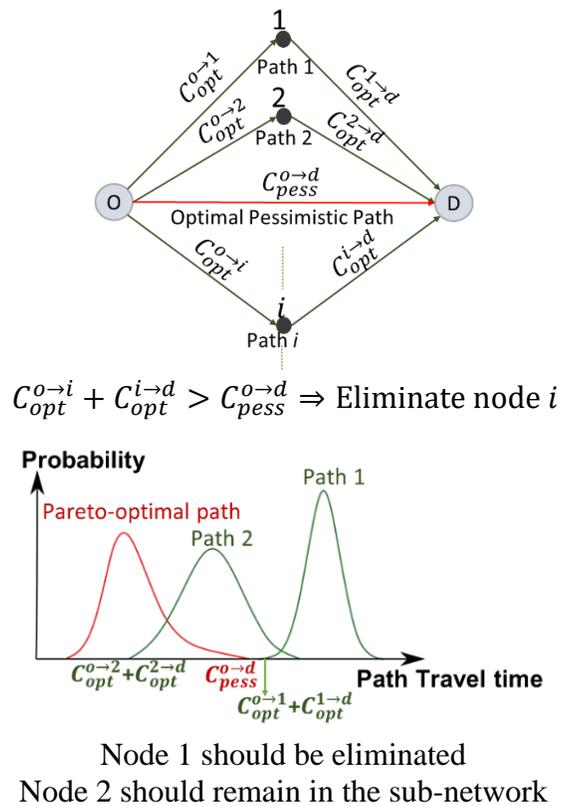


Figure 4-1- A schematic illustration of network contraction for the reliable path finding problem

4-3-1- Data Intuition

The Chicago downtown network, shown in Figure 3-1, is used in this chapter to infer a relationship between optimistic/pessimistic bounds and free flow travel times. To generate link travel time distributions and correlations for the AM peak period (5:00 AM to 10:00 AM), the network is simulated for different scenarios using the mesoscopic simulator of DYNASMART-P (85). An Intel® Xeon® CPU E5-2643 processor is used for all runs presented in this chapter, which includes 12 threads of 3.4 GHz and 128 GB of shared RAM.

Although the assumption of setting all link travel times to minimum/maximum possible values is expected to yield the same results as using the full network without any network contraction, it restricts the computational benefit of the network contraction approach. This assumption is highly conservative as the distance between origin and destination increases. Hence, the objective of this section is to study the realizations of link travel times and find a relation between minimum/maximum labels and FFTTs in order to propose realistic optimistic/pessimistic bounds for network contraction. The MCS approach presented by Zockaie et al. (4) for solving the SPOTAR and MTTBP problems is used in this chapter as the benchmark. The first stage of the stochastic path finding problem, including solving a deterministic shortest path problem at each iteration (with a link time realization), is executed for 100 iterations for a randomly selected destination in the time-dependent Chicago downtown network (destination 84). Note that a sensitivity analysis on the number of iterations was conducted in (96) and the number of iterations is set to 100 for the first stage as a value with satisfactory results in comparison to other approaches in the literature (38). However, any number of iterations more than 100 can be used without restricting the application of the learning algorithms. The computational efficiency of the learning approaches relative to the full network application improves further considering larger number of

iterations. The travel time labels from each node to the destination for different departure time intervals and iterations bring about the required insight towards the optimistic/pessimistic bounds. Thus, the minimum of realized time-dependent labels of each node over the 100 iterations is used to estimate the optimistic travel time from origin through that node to the destination. Also, the maximum realized time-dependent labels over the 100 iterations of each origin node is used to estimate the pessimistic travel time from that origin to the destination.

Figure 4-2 demonstrates the maximum and minimum of time-dependent labels over the 100 realizations obtained for any node located at a certain FFFT of the destination, divided by its FFFT. Each dot in Figure 4-2a and Figure 4-2b represents a node in the network. The horizontal axis represents the FFFT from each node to the destination, while the vertical axis represents the ratio of the minimum/maximum label to the FFFT. It can be observed from these figures that the ratio of optimistic bounds to FFFTs increases with the growing order of the FFFT from each node to the destination. Whereas, with the increase of the FFFT from nodes to the destination, the ratio of pessimistic bounds to their corresponding free flow travel time decreases. Thus, expectedly, one can devise a relationship between optimistic/pessimistic bounds and FFFTs. This relationship provides an intuition for obtaining efficient network contraction strategy. Using the observations of Figure 4-2 and according to the elimination condition presented in Figure 4-1, the probability of elimination from the sub-network would be larger for nodes that are located at further free flow travel times with respect to the origin and destination of interest. Therefore, adopting realistic optimistic/pessimistic bounds in lieu of fixed bounds helps improving the results through eliminating farther nodes to the destination, which have lower probability to be in the optimal path. Additionally, these realistic bounds result in smaller sub-networks for OD pairs with higher FFFTs. Once there are more links connecting an origin to a destination (OD pairs with higher

FFTTs), the possibility of having all links at their minimum or maximum bound at the same time would be less. Note that the FFTT is a better representative of how far the origin is located relative to the destination, than merely considering the distance between the origin and destination, since it considers the difference between various link types such as freeways and arterials.

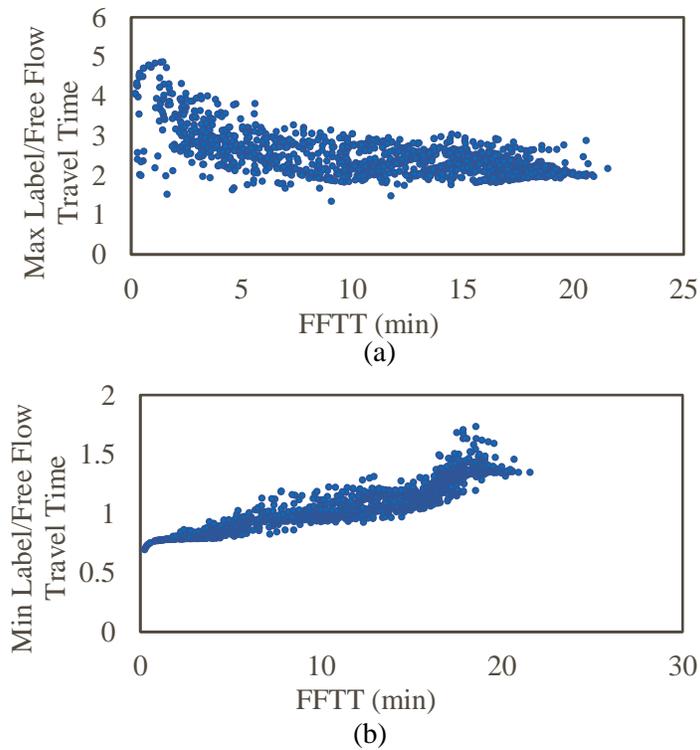


Figure 4-2- a) Maximum b) Minimum time-dependent labels over 100 realizations from each node to destination 84 of the Chicago network, divided by FFTT, versus the FFTT from each node to the same destination

Although considering the maximum and minimum realized labels results into improved optimistic/pessimistic bounds, there is still some potential to incorporate more aggressive bounds that may lead to smaller sub-networks. Thus, a data mining approach is used to estimate a factor that can be multiplied into the realized labels resulting in tighter bounds and smaller sub-networks. While there is an expectation that using tighter bounds improves the computational efficiency, this approach may result in sub-optimal solutions due to eliminating some nodes of the optimal path to form the sub-network. Therefore, each network requires calibrating the data mining factor.

The optimistic/pessimistic labels of destination 84 are used in order to conduct a sensitivity analysis on different data mining factors. The results are also compared with the conservative approach of CHAPTER 3 with a fixed pessimistic bound of 5 and a fixed optimistic bound of 0.8 (assumed maximum and minimum link travel times relative to link FFTTs). For the purpose of this sensitivity analysis, the maximum labels from all origins to the destination are first determined through optimal path calculations over 100 realizations. Different data mining factors are then multiplied by the minimum labels from any origin through the node to the destination. As mentioned earlier, comparing the resulting optimistic bound for each node and the pessimistic bound for each origin determines removal of the node from the associated sub-network to that origin. Reporting of the remaining number of nodes in the sub-network associated with each origin is a measure of computational efficiency.

Analyzing the results in Figure 4-3 and Table 4-1, one can derive tighter bounds by multiplying the minimum bound for each node by the data mining factor that a) achieves the highest reduction in the sub-network size and b) minimizes the chances for path calculation errors. In this illustrative example, a data mining factor of 1.3 reasonably satisfies these two conditions. As shown in Figure 4-3, it significantly reduces the sub-network size. In addition, based on the results in Table 4-1, the number of instances with more than 1% error in the objective function relative to the one reported by Zockaie et al. (4) is equal to zero. While increasing the learning factor to 1.35 is expected to reduce the sub-network size, this factor is associated with an instance with an error larger than 5%, which is relatively high. While we selected to only adjust the minimum bounds, the same sub-network could be obtained by tightening the maximum bounds, or tightening the minimum and maximum bounds simultaneously. However, one should expect different values for the tightening factors used in these cases.

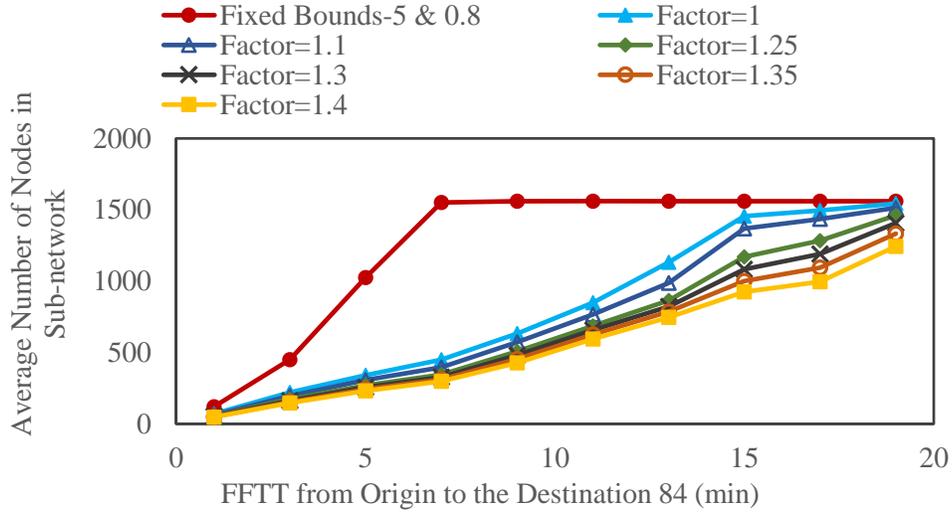


Figure 4-3- Average number of nodes in sub-networks using different data mining factors in the network contraction approach for all origins to the destination 84 of Chicago downtown network in 2-minute intervals of FFTTs

Table 4-1- The number of instances for different relative errors that the network contraction approach, using different data mining factors, has a worse objective function value than the full network of Zockaie et al. (4) for destination 84 using SPOTAR reliability rule

Data mining factors	0% < E* ≤ 1%	1% < E* ≤ 5%	5% < E* ≤ 10%	10% < E*	Total*
1.10	0	0	0	0	0
1.20	0	0	0	0	0
1.25	0	0	0	0	0
1.30	5 (0.05%)	0	0	0	5 (0.05%)
1.35	12 (0.12%)	9 (0.09%)	1 (0.01%)	0	22 (0.22%)
1.40	59 (0.6%)	54 (0.55%)	23 (0.23%)	4 (0.04%)	140 (1.43%)

*The number of error cases are out of 9810 cases which is for 218 origin zone, 15 time intervals, and 3 user classes

4-3-2- Solution Algorithm

This sub-section presents a learning approach to extract a sub-network for each specific OD pair, which is applicable to reliable path finding problems. A modified MCS approach for time-dependent stochastic networks with multiple classes of users is utilized in this study to solve the SPOTAR and MTTBP problems (4). However, utilizing any other MCS approach works for this purpose. As stated in the previous chapter, this approach consists of two stages; the first stage

generates a set of candidate paths by applying a dynamic deterministic shortest path algorithm for different realizations of link travel times. This stage consists of a pre-specified number of iterations. In each iteration, realizations of link travel time are made by drawing random numbers from a joint link travel time distribution. Based on these realizations, a candidate path is found, using a dynamic deterministic shortest path finding algorithm, and then added to the candidate paths set. After this iteration repeats for a certain number of times, the set of candidate paths can be used as the input for the second stage. Similar to the first stage, the second stage consists of a pre-specified number of iterations, which is greater than that used in the first stage. For each iteration, a set of link travel time realizations is used to generate the set of candidate paths travel time distributions by summing the realized travel times of the links that exist in that path.

The algorithm proposed in this chapter takes advantage of the produced information during the travel time realizations to find optimistic and pessimistic travel times for each OD pair of interest. For a certain OD pair, the minimum link travel time is first assigned to all network links and the travel times of both one-to-all and all-to-one deterministic static shortest path problems are solved, resulting in conservative optimistic travel times from origin to the destination passing through any node in the network. Next, all network link travel times are set to their maximum values, and the conservative pessimistic travel time is calculated from the origin to the destination using the deterministic static shortest path algorithm. Comparing the optimistic travel time through any node to the pessimistic travel time for the OD pair specifies if the node remains in the sub-network or not.

Next, realizations of link travel times are made by drawing random numbers from the joint link travel time distribution for the first n iterations of the first stage. An all-to-one dynamic shortest path algorithm (75) is used to solve the dynamic shortest path problem in this stage.

Performing these n iterations obtains a set of candidate paths and the realized maximum and minimum path travel time labels from each node to the destination. These maximum and minimum travel times can be used as realistic estimates of the pessimistic and optimistic bounds. Given that a destination-based algorithm is considered to solve the deterministic shortest path in the first stage, the optimistic bound from any node in the network is used as an estimate of the bound for the other part from the origin to the node. Utilizing optimistic/pessimistic bounds further contracts the sub-network to a smaller one. Repeatedly, all steps after extracting a new sub-network are executed for the next n iterations till reaching the maximum number of iterations considered for the first stage. Once the first stage is completed, the second stage then finds path travel time distributions. Here, n is defined as the updating phase of optimistic/pessimistic bounds.

Seeking the right balance between accuracy and efficiency, a learning factor is used to make the bounds smaller at each n iterations of the first stage. This factor is multiplied by the minimum path travel times in order to compensate for the conservative nature of the bounds. Two approaches are also proposed, called fixed learning and adaptive learning, for iterative network contraction. The former applies a fixed data mining factor, as described in the data intuition section, to the minimum realized label after the first n iterations. The latter starts with a small learning multiplier applied to the minimum realized labels after the first n iterations. Then, every n iteration, it increases the learning multiplier up to the point that it is less than or equal to the data mining factor. Since the minimum travel time labels decrease over the iterations and the maximum ones increase, the learning factors make the optimistic/pessimistic bounds less conservative. The fixed learning approach is more aggressive in reducing the network size as it starts with a relatively large data mining factor. However, the adaptive learning approach starts with a smaller learning

factor and ultimately reaches to a more informative data mining factor. Figure 4-4 provides an overview of the algorithm. A pseudo-code describing the steps of the algorithm is as follows.

Step 1 Initialization

- Form the joint time-dependent link travel time distribution considering temporal and spatial correlations
- Define the OD pair of interest
- Define the data mining factor (as described in Section 4-3-1)
- Define the number of stage one and stage two iterations (S_1 and S_2) and number of iterations for updating phases of optimistic/pessimistic bounds, n .

Step 2 Find conservative optimistic and pessimistic travel times by setting all links to their minimum and maximum possible values.

Step 2-1 Solve the one-to-all static shortest path problem from the origin to all other nodes using the minimum link travel times, and store travel times as the optimistic travel times from the origin to all other nodes ($C_{opt}^{o \rightarrow i}$).

Step 2-2 Solve the all-to-one static shortest path problem from all nodes to the destination using the minimum link travel times, and store the optimistic travel time from all nodes to the destination ($C_{opt}^{i \rightarrow d}$).

Step 2-3 Solve the all-to-one static shortest path problem from all nodes (including the origin) to the destination using the maximum link travel times and store the pessimistic travel time from the origin to the destination ($C_{pess}^{o \rightarrow d}$).

Step 3 Extract a sub-network for the OD pair of interest by checking the following condition for each node, i .

Step 3-1 Add the optimistic travel time from the origin to the node, $C_{opt}^{o \rightarrow i}$, (from step 2-1) to the optimistic travel time from the node to the destination, $C_{opt}^{i \rightarrow d}$, (from step 2-2), and store it as the optimistic travel time from the origin to the destination, $C_{opt}^{o \rightarrow i \rightarrow d}$.

Step 3-2 If the optimistic travel time of the OD pair, (from step 3-1) is smaller than the pessimistic value, $C_{pess}^{o \rightarrow d}$, (from step 2-3), keep the node in the sub-network; Otherwise, remove it.

Step 4 Draw a random sample of size S_1 for the sub-network from the joint time-dependent link travel time distribution. Set the current sample index as $s_1 = 1$.

Step 4-1 Set the link travel time on link a at time t as $c_{at}^{s_1}$, which is the random number in the s_1^{th} sample corresponding to link a at time interval t .

Step 4-2 Find the dynamic deterministic shortest paths from all nodes in the current sub-network including the origin to the destination for all departure time intervals.

Step 4-3 For all departure time intervals, add the newly found path to the set of candidate paths of that departure time interval ($K^{od\tau}$) if it has not been found previously.

Step 4-4 Set s_1 as $s_1 + 1$. If $s_1 \geq S_1$ go to Step 7; else if $s_1 = j \times n$, where j is any integer lower than $\frac{S_1}{n}$, go to Step 5; else go to Step 4-1.

Step 5 Determine intuitive optimistic/pessimistic travel times

Step 5-1 Determine the pessimistic travel time from all nodes (including the origin) to the destination from the candidate paths travel times of previous $j \times n$ iterations and update $C_{pess}^{o \rightarrow d}$

Step 5-2 Determine the minimum travel times of all candidate paths from all nodes to the destination, $C_{opt}^{i \rightarrow d}$.

In the ***adaptive learning approach***, update the optimistic travel times from all nodes to the destination by multiplying the minimum travel times by a learning factor L . This factor is less than or equal to the data mining factor defined in the Section 4-3-1. Update $C_{opt}^{i \rightarrow d}$ with the resulting value.

$$C_{opt}^{i \rightarrow d} = C_{opt}^{i \rightarrow d} \cdot L, \text{ where}$$

$$l = \text{Learning multiplier}^{(j-1)}$$

$$L = \begin{cases} l & , l < \text{Data mining factor} \\ \text{Data mining factor} & , l \geq \text{Data mining factor} \end{cases}$$

Therefore, the learning multiplier is constant over the iterations while learning factor (L) converges to the calibrated data mining factor.

In the ***fixed learning approach***, the minimum travel times are multiplied by a fixed learning factor equal to the data mining factor after the first n iterations.

Store the *optimistic bound* $= \frac{C_{opt}^{i \rightarrow d}}{FFTT_{i \rightarrow d}}$, where $FFTT_{i \rightarrow d}$ is the free flow travel time from i to d .

Step 5-3 Using the optimistic bound of Step 5-2, update the optimistic travel time from the origin to the destination, $C_{opt}^{o \rightarrow i \rightarrow d}$, as *optimistic bound* $\times (FFTT_{o \rightarrow i} + FFTT_{i \rightarrow d})$.

Step 6 Extract a new sub-network from the current sub-network using the information of the previous $j \times n$ iterations for the OD pair of interest.

If the optimistic travel time of the OD pair (from step 5-3), $C_{opt}^{o \rightarrow i \rightarrow d}$, is smaller than the pessimistic value (from step 5-1), $C_{pess}^{o \rightarrow d}$, keep the node in the sub-network; Otherwise, remove it. Go to Step 4-1.

Step 7 Draw a random sample of size S_2 from the joint time-dependent link travel time distribution. Set the current sample index as $s_2 = 1$.

Step 7-1 Set the link travel time on link a at time t as $c_{at}^{S_2}$, which is the random number in the s_2^{th} sample corresponding to link a at time interval t .

Step 7-2 For all departure time intervals such as τ , and all paths in the set of candidate paths of that departure time interval such as $k^{od\tau} \in K^{od\tau}$, find $\Pi_{k^{od\tau}} = \sum_{a \in k^{od\tau}} c_{at}^{S_2}$ as the path travel time, where t is the arrival time of link a based on the travel time of the previous link in the path $k^{od\tau}$.

Step 7-3 Set s_2 as $s_2 + 1$. If $s_2 \leq S_2$, then go to Step 7-1; else go to Step 8.

Step 8 Apply the reliability rule of interest to find the optimal path. Here, path travel time distributions are available for all candidate paths. Therefore, any reliability rule can be applied to find the optimal path.

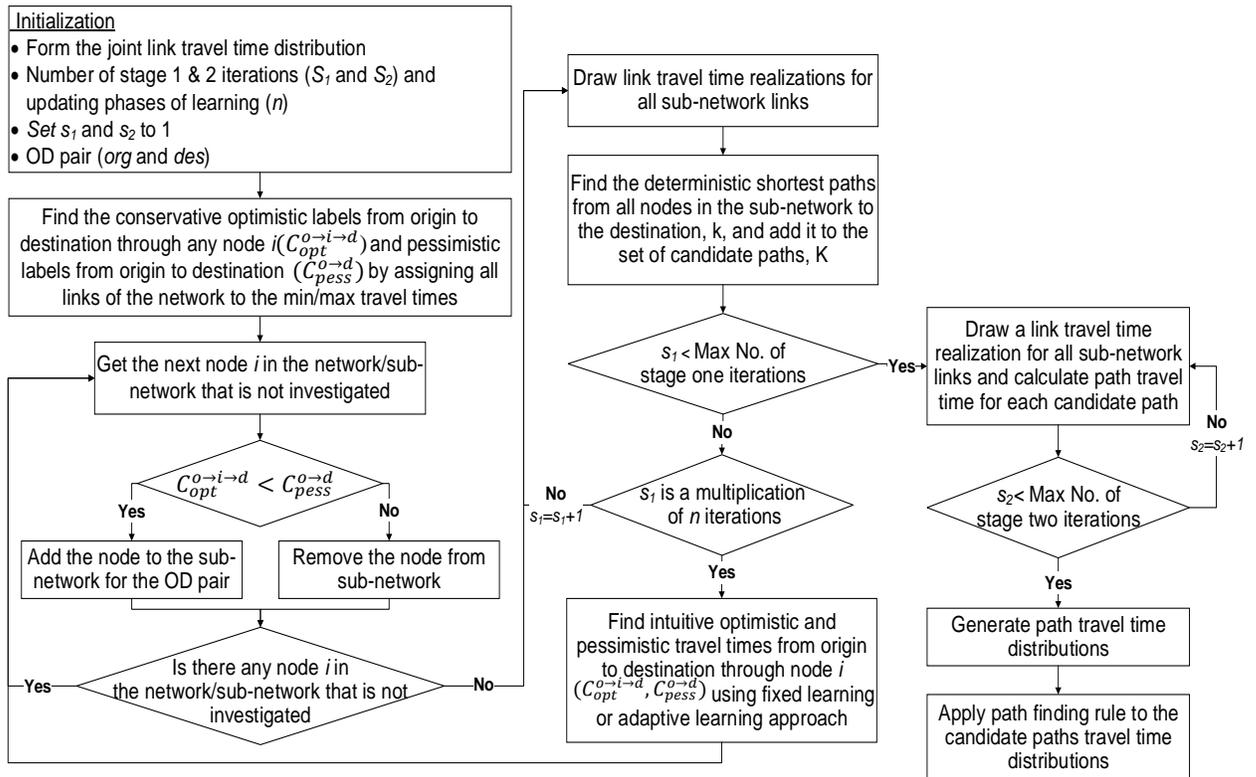


Figure 4-4- A brief illustration of the learning approach for efficiently solving reliable path finding problem

4-4- Numerical Experiments

In this sub-section, a set of experiments is presented to examine the computational performance and accuracy of the fixed learning and the adaptive learning approaches against those of the optimal path finding algorithm presented by Zockaie et al. (4). The algorithm presented by Zockaie et al. (4) considers a full original network in all iterations of the MCS approach. The approach of the mentioned study is considered as the benchmark in this chapter, since it accounts for link travel correlations and heterogeneous users in terms of reliability valuation. The optimality and solution time of the benchmark approach are already compared and validated (11,96) by comparing the solutions with the results of an outer approximation algorithm, which considers link travel time correlations (38), and a label correcting algorithm to solve SPOTAR problem (28). All approaches are implemented for the Chicago downtown network using a realistic joint time-dependent travel time distribution for the SPOTAR and MTTBP problems. Here, the algorithm is first applied to the same destination for which the data mining factor is calibrated (destination 84). Then, both learning approaches are applied to two randomly-selected destinations located at other parts of the network with different node and link configurations.

The number of iterations is assumed to be 100 for the stage one and 1000 for the stage two of the MCS approach. The number of iterations for the MCS approaches significantly affects the solution time and the accuracy of results. Thus, many studies in the literature provide different relations to estimate the optimum number (97,98). The desired margin of error, desired confidence level, and standard deviation of the sample are mentioned as effective parameters in the estimation of the optimum number of iterations (97). The study by Zockaie et al. (96) conducts a sensitivity analysis on the impacts of the number iterations in the first and second stages of the MCS approach of this study on the objective function value and the solution time. The study shows that 100 and

1000 as the number of iterations for the first and second stages of the MCS approach for the optimal path finding problem produce acceptable results with a significantly lower solution time relative to the outer approximation method, which directly solves the MTTBP problem. Thus, these two values are adopted as the number of iteration in the numerical experiments of this study.

As the execution time is usually sensitive to the computer configuration and could have fluctuations in multiple runs, the number of nodes is used in this study as a representative of the computational time. Nonetheless, to show the relation between the computational time and number of nodes, the total execution time and the time required for stage one of the algorithm are recorded with respect to the number of nodes, as demonstrated in Figure 4-5. In this figure, the computational time to solve the path finding problem from any given origin in the network to destination 84 is presented. The path finding problems are solved using the fixed learning approach, with a learning factor of 1.3. In this figure, the number of nodes in the sub-network associated with each origin are combined in 100-node intervals and the average execution time value is reported for each interval. In addition, the average execution time of five different runs are reported to avoid the fluctuations in the reported time of multiple executions. The results show that the major computational time of the current approach is devoted to executing the dynamic deterministic path finding algorithm for different realizations of link travel times in stage one. Improving the execution time of this stage, which grows approximately with the second order polynomial degree of the number of nodes, results in a significant reduction of the total execution time. A polynomial curve with the degree of two is also drawn for the total execution time of the fixed learning approach, which is shown to properly approximate the total execution times with respect to number of nodes. Finally, the execution time of the approach of Zockaie et al. (4) with

the full network is given in the Figure 4-5, which is significantly larger than the execution time for all of the OD pairs using the fixed learning approach.

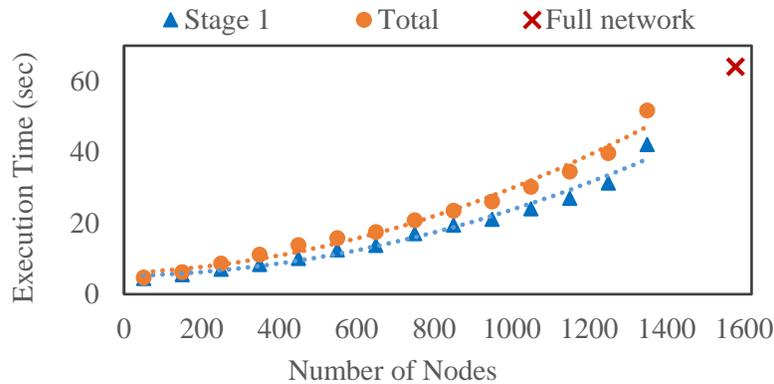


Figure 4-5- The relationship between the execution times and the number of nodes for the fixed learning approach (learning factor =1.3) to find the optimal path from different origins to destination 84

As the number of nodes in the sub-network of each OD pair varies over the iterations in the learning approaches, the average value of the number nodes in the sub-networks is used as an indicator of the algorithm computational requirement. Figure 4-6 compares the performance of the fixed learning approach with the adaptive learning approach considering two different learning multipliers. The figure also compares the performance of these two approaches with those of the pre-knowledge approach described in section 4-3-1 and the conservative approach of CHAPTER 3. In the fixed learning approach, the information of the first 10 iterations is used for network contraction. In the adaptive learning approach, the number of iterations used to update the optimistic/pessimistic bounds, n , is assumed to be 10. The updating phase should be small enough relative to the total number of iterations (100) to provide meaningful computational efficiency and large enough to include meaningful path travel time information to be used. Therefore, the value of 10 is assumed in this study. In Figure 4-6, the horizontal axis represents the FFTT from origin to destination 84. The origins are sorted based on the FFTTs between the origins and destination. The origins are then categorized based on their FFTTs to the destination in 2-minute intervals. The

average number of nodes in the sub-networks over all iterations of each OD pair is first calculated, and then the average number of remaining nodes in the sub-network for different origins of each 2-minute category is reported in the vertical axis.

Table 4-2 compares the accuracy of the mentioned scenarios in terms of the SPOTAR objective function against that of Zockaie et al. (4) in which the full network with constantly 1,578 nodes are used. The results of Table 4-2 and Figure 4-6 demonstrate that the adaptive learning approach with a learning multiplier of 1.05 has approximately the same performance and accuracy as the pre-knowledge approach with the data mining factor of 1.3. For instance, for an origin with the FFTT of 8 to 10 minutes, the fixed learning approach with the learning multiplier of 1.05 results in sub-networks with an average of 608 nodes during the iterations of the path finding algorithm, which is comparable to the conservative approach with an average of 481 nodes. Based on the results of Table 4-2, the probability of having error in the objective function is 0.06% for the adaptive learning approach in which the multiplier is equal to 1.05. Whereas, the fixed learning approach with a learning factor of 1.3 is prone to more errors in the objective function. The total number of errors in the objective function is recorded to be 126 cases (out of 9810 cases) for this approach (1.28%), which is still acceptable considering the savings in the execution time. For example, for an origin with FFTT of 8 to 10 minutes to the destination, sub-networks with an average of 360 nodes are obtained, which require an execution time of around 11.5 seconds. However, the full network and the conservative approach use 1,578 nodes for the same OD pairs, which require an execution time of 64.1 seconds.

Figure 4-7 compares the execution time of the fixed learning approach with the learning factor of 1.3 and the adaptive learning approach with the learning multiplier of 1.05 with the conservative approach of CHAPTER 3. This figure demonstrates significant computational gains

achieved by using the learning approaches relative to the conservative approach. Note that the execution time for the approach of Zockaie et al. (4) is 64.1 seconds for all FFTTs, which is significantly higher than the execution time for all OD pairs incorporating the learning approaches. Comparing the results of Figure 4-6 and Figure 4-7 suggests the number of nodes as a good indicator of the execution time. Therefore, hereafter, the results are presented for the number of nodes remaining in the sub-network to avoid fluctuations in the computational time for different executions of the algorithms.

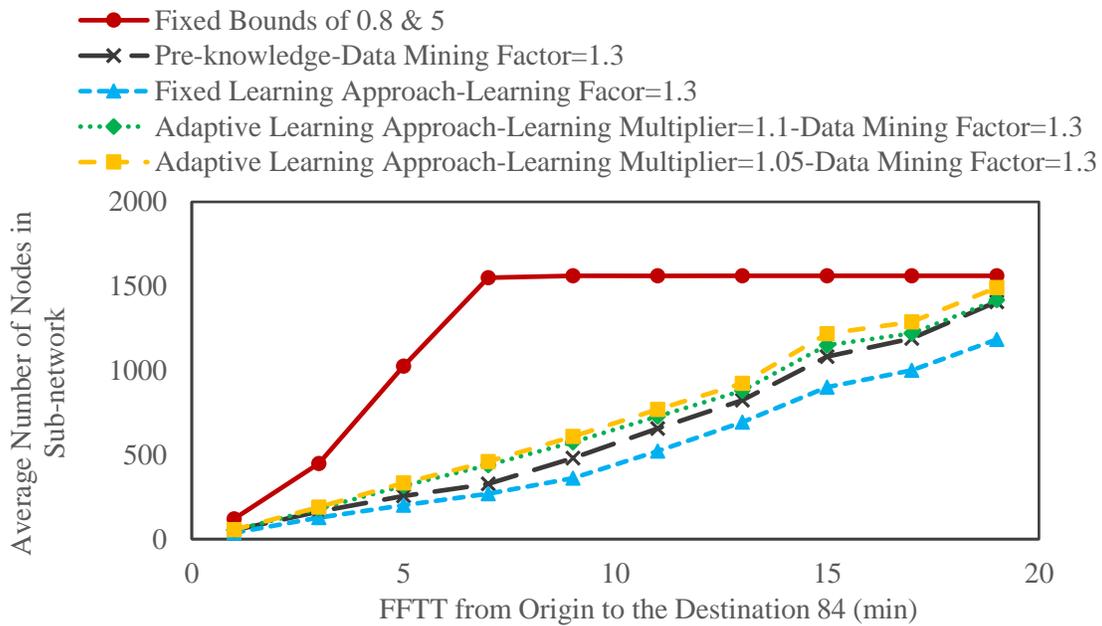


Figure 4-6- Number of nodes in the sub-network for all origins to the destination 84 in 2-minute intervals of FFTT for different network contraction approaches

Table 4-2- The number of instances for different relative errors that network contraction approaches have a worse objective function value than the full network approach of Zockaie et al. (4) for destination 84 using SPOTAR reliability rule

Scenarios	0% < E ≤ 1%*	1% < E ≤ 5%*	5% < E ≤ 10%*	10% < E*	Total*
Pre-knowledge data (Data mining factor=1.3)	5 (0.05%)	0	0	0	5 (0.05%)
Fixed bounds (0.8 and 5)	0	0	0	0	0
Fixed learning approach (Learning factor=1.3)	67 (0.68%)	44 (0.45%)	10 (0.10%)	5 (0.05%)	126 (1.28%)
Adaptive learning approach (Learning multiplier=1.1)	8 (0.08%)	5 (0.05%)	0	0	13 (0.13%)
Adaptive learning approach (Learning multiplier=1.05)	4 (0.04%)	2 (0.02%)	0	0	6 (0.06%)

*The number of error cases out of 9810 cases

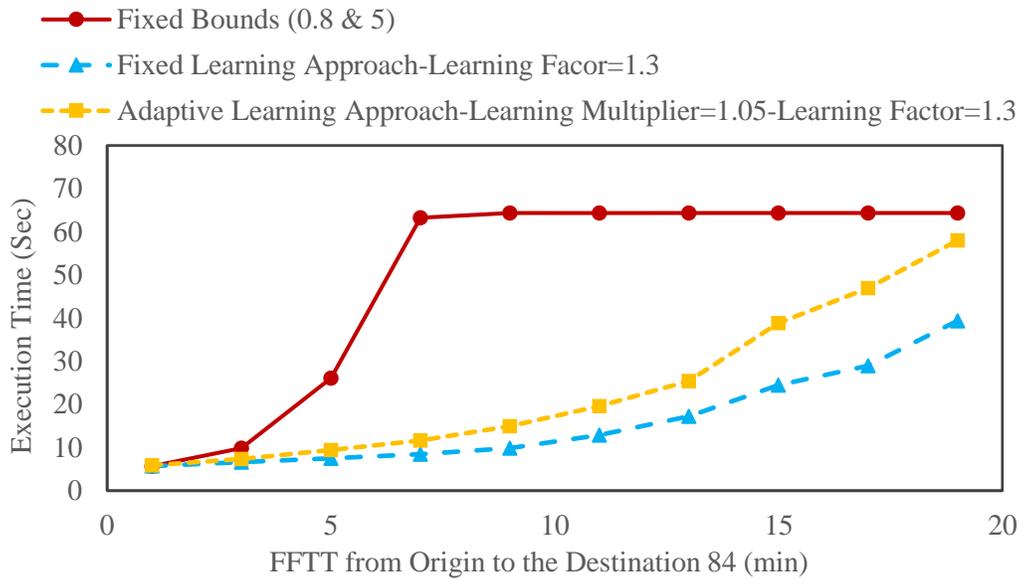


Figure 4-7- The comparison of execution time for all origins to the destination 84 in 2-minute intervals of FFTT for different network contraction approaches

The great improvements in the results of the two learning approaches relative to the conservative approach certify the application of the new network contraction algorithm in terms of the computational efficiency. A major limitation of MCS approaches is that they may lead to different results in multiple executions of the algorithm. However, increasing the number of realizations (iterations) solves this problem. Thus, the algorithm is executed five times for all origins of destination 84 in the Chicago network with the full network approach to study the

discrepancy of the results and verify if the selected number of iterations is adequate. The average relative difference between the objective functions of each two executions is 0.31%, which implies that the number of iterations is a proper setting. In addition, the adaptive learning approach of this chapter with a learning multiplier of 1.05 is applied for the same destination in 5 executions. The average relative difference between the objective functions of each two executions is 0.29%. Given the results of Figure 4-6 and Table 4-2, a learning multiplier of 1.05 has a comparable result with the pre-knowledge approach. Therefore, the fixed learning approach with a factor of 1.3 and the adaptive learning approach with a data mining factor of 1.3 and learning multiplier of 1.05 are selected to test the performance of the approach for other destinations. However, any other learning multiplier lower than the calibrated data mining factor can be used based on the desired trade-off between accuracy and efficiency. It is worth mentioning that the same updating phase (10) and the total number of iterations (100) should be used along with other calibrated factors.

Given that destination 84 is located in the northern part of the Chicago network, one destination is selected from the southern part with a dense topology of short links (destination 1) in addition to another random destination (destination 57). Application of the learning contraction algorithm to these two destinations verifies the accuracy and evaluates the performance of the approaches. Figure 4-8 and Figure 4-9, respectively, illustrate the computational benefits of both learning approaches in terms of the average number of nodes in the sub-networks for different OD pairs of destination 1 and destination 57. As shown in the figures, for both destinations, the learning approaches result in much smaller number of nodes in the sub-networks compared to that of the conservative approach and the full network. In addition, the results show that the fixed learning approach is superior to the adaptive learning approach in terms of the network size reduction for both destinations.

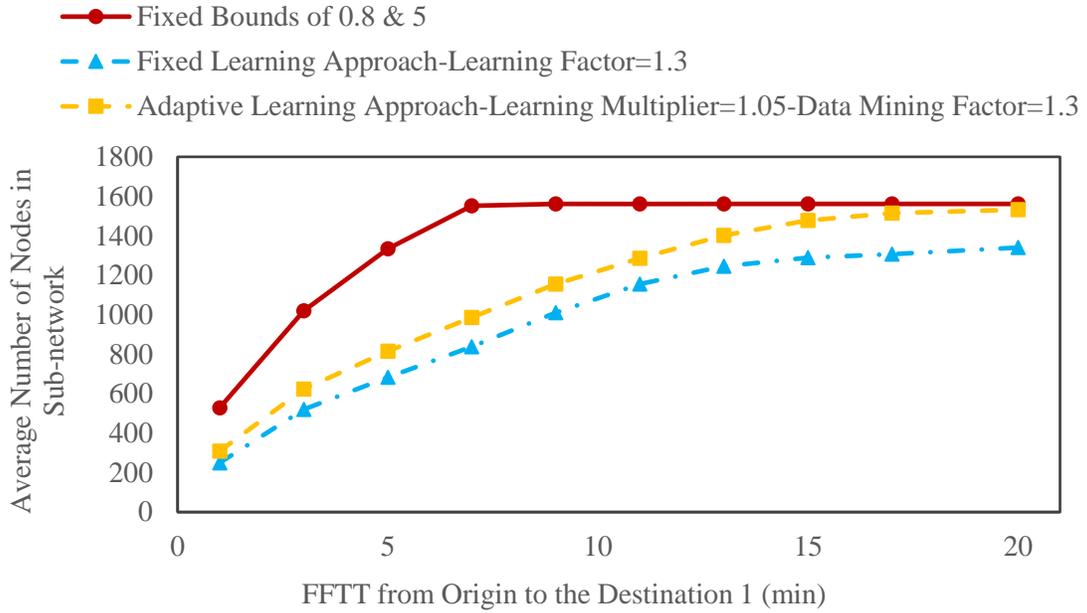


Figure 4-8- The comparison of number of nodes in the sub-network for all origins to the destination 1 in 2-minute intervals of FFFT for different network contraction approaches

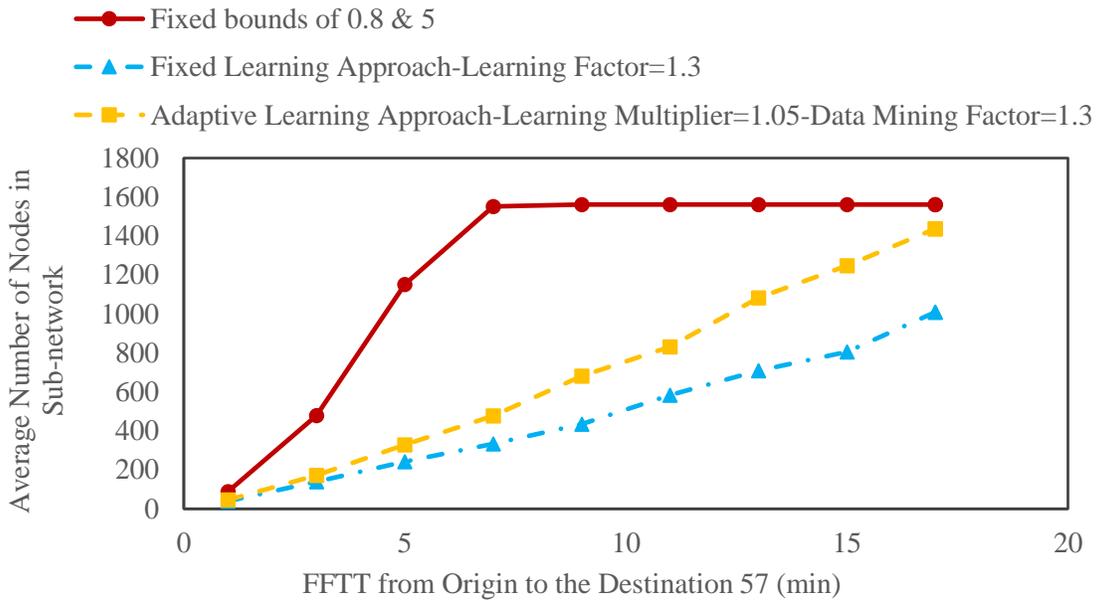


Figure 4-9- The comparison of number of nodes in the sub-network for all origins to the destination 57 in 2-minute intervals of FFFT for different network contraction approaches

Table 4-3 compares the performance of the learning approaches for the same destinations in terms of the accuracy of the estimated objective function for two different reliability rules: SPOTAR and MTTBP. The number of cases with different levels of error, with respect to the equivalent objective functions of Zockaie et al. (4), are presented in this table. For both reliability rules, the learning approaches have acceptable level in terms of number of errors. However, the adaptive learning approach with learning multiplier of 1.05 and data mining factor of 1.3 is superior to the fixed learning approach with the learning factor of 1.3 in terms of the objective function accuracy. Considering the results shown in Figure 4-8, Figure 4-9, and Figure 4-10, a trade-off between the desired accuracy level and the computational efficiency is important while selecting the proper learning approach. Overall, the adaptive learning approach shows satisfactory improvements in terms of computational efficiency with the same accuracy as that of the approach without any network contraction.

Figure 4-10 demonstrates the resulting sub-networks for a random-selected origin (i.e., node 141) to destination 1 using various learning approaches. As shown in the figure, the area of the retained nodes in the sub-network is significantly smaller in both learning approaches relative to the full network and that obtained using the conservative approach. This figure demonstrates the superiority of the learning approaches developed in this chapter relative to the approach of CHAPTER 3 and the existing approach in the literature. The full network, with 1,578 nodes, demonstrates the network used in the approach of Zockaie et al. (4) or any other algorithm that does not consider network contraction. The second network which includes 1,305 nodes is the sub-network resulted from applying the algorithm of CHAPTER 3. The other two networks, with 675 and 428 nodes, are the sub-networks resulted from the network contraction algorithm of this chapter.

Table 4-3- The number of instances that different learning approaches have a worse objective function, relative to the full network approach of Zockaie et al. (4), for different relative errors and destinations using a) MTTBP reliability rule; b) SPOTAR reliability rule

a)						
Destination	Approach	0% < E ≤ 1%*	1% < E ≤ 5%*	5% < E ≤ 10%*	10% < E*	Total*
84	Pre-knowledge	5 (0.05%)	0	0	0	5 (0.05%)
1	Fixed bounds (0.8 and 5)	0	0	0	0	0
1	Adaptive learning (Learning Multiplier=1.05)	0	0	0	0	0
1	Fixed learning (Learning Factor = 1.3)	37 (0.38)	86 (0.87%)	42 (0.43%)	7 (0.07%)	172 (1.75%)
57	Fixed bounds (0.8 and 5)	0	0	0	0	0
57	Adaptive learning (Learning Multiplier=1.05)	4 (0.04%)	0	0	12 (0.12%)	16 (0.16%)
57	Fixed learning (Learning Factor = 1.3)	104 (1.06%)	77 (0.78%)	1 (0.01%)	12 (0.12%)	194 (1.98%)
b)						
84	Pre-knowledge	6 (0.06%)	0	0	0	6 (0.06%)
1	Fixed bounds (0.8 and 5)	0	0	0	0	0
1	Adaptive learning (Learning multiplier=1.05)	0	0	0	0	0
1	Fixed learning (Learning Factor = 1.3)	44 (0.45%)	79 (0.81%)	45 (0.46%)	15 (0.15%)	181 (1.84%)
57	Fixed bounds (0.8 and 5)	0	0	0	0	0
57	Adaptive learning (Learning Multiplier = 1.05)	3 (0.03%)	0	0	0	3 (0.03%)
57	Fixed learning (Learning Factor = 1.3)	98 (0.99%)	88 (0.89%)	3 (0.03%)	46 (0.47%)	235 (2.39%)

*The number of error cases are out of 9810 cases

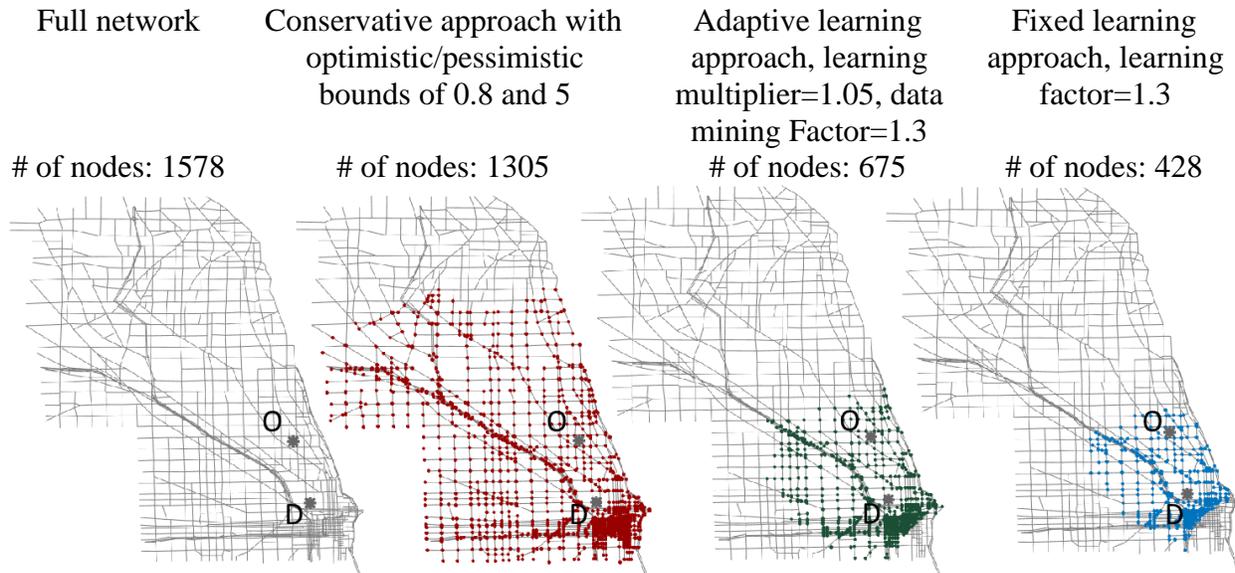


Figure 4-10- Sub-network configurations for the two learning approaches and the conservative approach from origin 141 to the destination 1 in the Chicago downtown network

4-5- Model Transferability

In the previous sub-section, the learning factor obtained based on the analysis conducted for destination 84 in the Chicago downtown network is used for two other destinations in that network. The level of time efficiency and accuracy of the network contraction approach depends on the network structure and link travel time distributions. Therefore, in this sub-section, the calibrated parameters using the Chicago downtown network are applied to another network with a different configuration, in order to check the transferability of the approach to other networks. To this end, the learning approaches are implemented for the Salt Lake City network bounded by the intersection of I-15 and State Route 89 on the North side and State Route 145 (Pioneer Crossing) on the South side (99). The network contains 3,626 nodes, 8,321 links, and 1,287 zones. To generate link travel time distributions and correlations, the network is simulated using DYNASMART-P for the AM peak period between 6 AM to 10 AM. This simulation horizon is divided into 13 time intervals to generate link travel time distributions based on the analytical approach of Zockaie et al. (100) to relate the mean and the variance of link travel times. In the

correlation structure, there is an assumption that the correlation exists spatially between adjacent links and temporally between the subsequent time intervals.

Five destinations are selected randomly in the Salt Lake City network to show the transferability of the learning approaches to different networks, using the calibrated factors for the Chicago downtown network. Three of the destinations are picked from three different parts of the network; a destination in the northern dense portion of the network (destination 27), a destination in the central part (destination 80), and a destination in the southern portion (destination 145). The other two are randomly selected over the entire network. The two learning approaches in addition to the conservative approach of CHAPTER 3 are applied for the network contraction of these destinations. Demonstration of the average number of nodes remaining in the sub-networks for the different origins to the selected destinations is available in Figure 4-11. As shown in this figure, the fixed learning approach and the adaptive learning approach provide significant improvements in terms of the average number of nodes in the sub-networks. For instance, for origins with the free flow travel time of 8 to 10 minutes to destination 27, the average numbers of nodes in the sub-networks during all iterations of the adaptive and fixed learning approaches are 2,350 and 1,517, respectively. However, the approach of Zockaie et al. (4) uses the full original network with constantly 3,626 nodes over all iterations and the conservative approach results in a sub-network with an average of 3,500 nodes. The improvements are even more noticeable for the destinations 80 and 145.

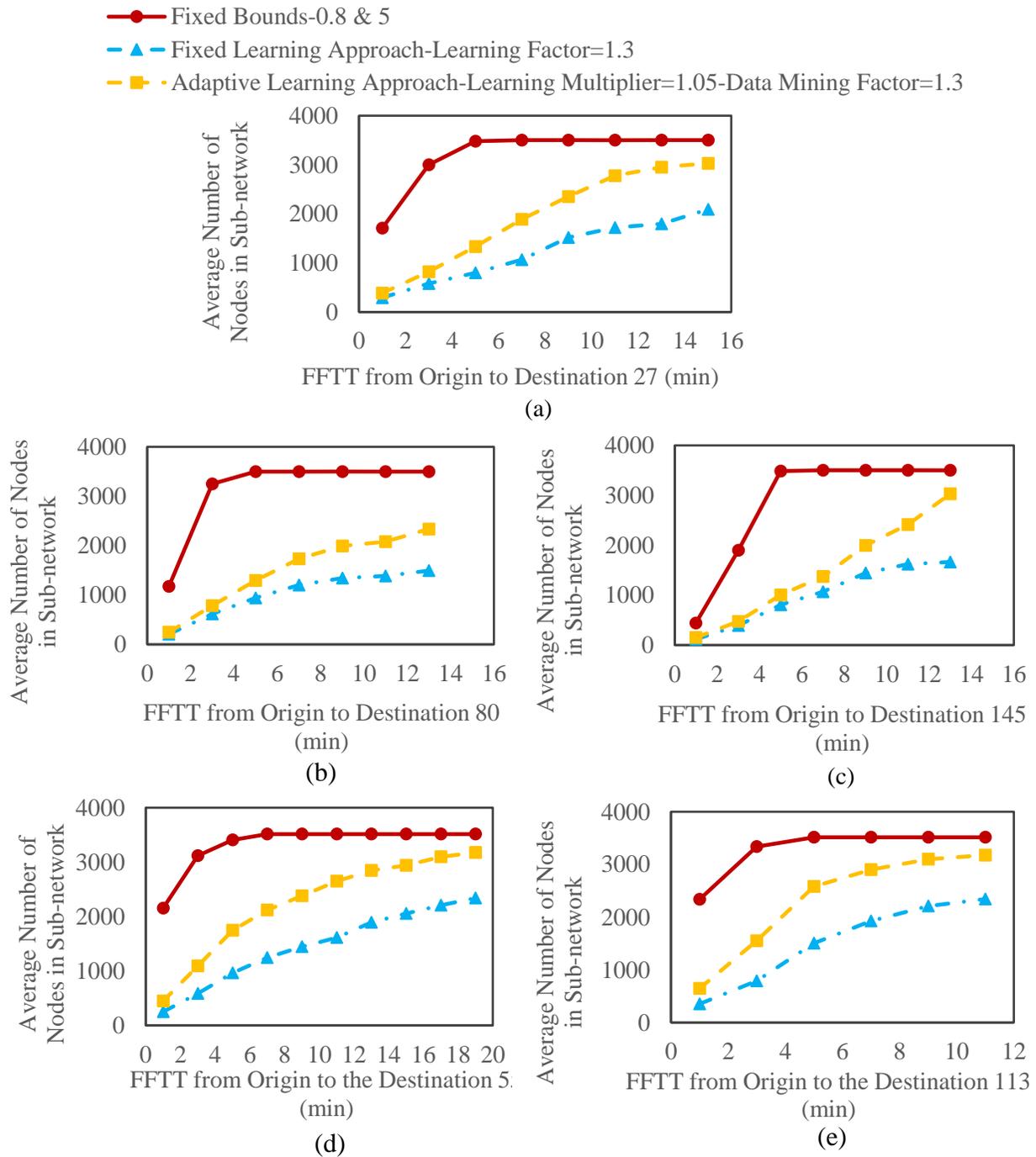


Figure 4-11- Number of nodes in the sub-network in 2-minute intervals of FFTT for different network contraction approaches (calibrated based on the Chicago network) from all origins to destination a) 27, b) 80, c) 145, d) 55, and e) 113 of the Salt Lake City network

Results related to the accuracy of the estimated objective functions are also obtained for the Salt Lake City network. Given the similar trend for both SPOTAR and MTTBP problems, the

results presented here are only for the MTTBP problem. The objective functions of the MTTBP problem for the considered OD pairs obtained using the conservative approach (fixed bounds of 0.8 and 5.0), the fixed learning approach, and the adaptive learning approach are compared with the objective functions obtained using the approach adopted by Zockaie et al. (4) in which no network contraction is considered (3,626 nodes for all iterations). These results are summarized in Table 4-4, which presents the different levels of error in the objective function relative to the approach of Zockaie et al. (4) for the three approaches. As expected, the conservative approach does not have any error. The adaptive learning approach has less than one percent error for all destinations (mostly less than 5% difference in the objective function values with and without network contraction). The fixed learning approach has variable accuracy for different destinations with errors that range from 1% to 10%. Thus, the adaptive learning approach has acceptable results in terms of both the average number of nodes in the sub-networks and the number of errors in the objective function. However, the fixed learning approach has a better computational efficiency at the cost of accuracy. Overall, both learning approaches calibrated based on the Chicago network produce good results for the Salt Lake City network, resulting in huge computational efficiency benefits with acceptable approximation in the objective function evaluation.

Finally, similar to the Chicago case study, the configuration of the Salt Lake City network in addition to the number of remaining nodes in the sub-networks of a selected OD pair (origin 1,187 to destination 145) using different network contraction approaches are available in Figure 4-12. Using the stochastic dynamic path finding algorithm of Zockaie et al. (4), all nodes in the network should be searched. However, applying the reliability rule and other steps of the stochastic path finding algorithm is necessary for much smaller sub-networks using the network contraction approaches.

Table 4-4- The number of instances that different learning approaches have a worse objective function, relative to the full network approach of Zockaie et al. (4), for different relative errors, destinations, and MTTBP problem in the Salt Lake City network

Destination	Approaches	0% < E ≤ 1%*	1% < E ≤ 5%*	5% < E ≤ 10%*	10% < E*	Total*
27	Fixed bounds (0.8 and 5)	0	0	0	0	0
	Adaptive learning	136 (0.27%)	60 (0.12%)	0	81 (0.16%)	275 (0.55%)
	Fixed learning	1128 (2.25%)	1206 (2.40%)	548 (1.09%)	2411 (4.8%)	5293 (10.54%)
80	Fixed bounds (0.8 and 5)	0	0	0	0	0
	Adaptive learning	154 (0.31%)	84 (0.17%)	11 (0.02%)	72 (0.14%)	321 (0.64%)
	Fixed learning	1087 (2.17%)	667 (1.33%)	741 (1.48%)	424 (0.84%)	2919 (5.81%)
145	Fixed bounds (0.8 and 5)	0	0	0	0	0
	Adaptive learning	209 (0.42)	130 (0.26%)	9 (0.02%)	57 (0.11%)	405 (0.81%)
	Fixed learning	644 (1.28)	492 (0.92%)	46 (0.09%)	291 (0.58%)	1143 (2.87%)
55	Fixed bounds (0.8 and 5)	0	0	0	0	0
	Adaptive learning	126 (0.25%)	41 (0.08%)	0	0	167 (0.33%)
	Fixed learning	290 (0.58%)	179 (0.36%)	59 (0.12%)	1723 (3.43%)	2251 (4.48%)
113	Fixed bounds (0.8 and 5)	0	0	0	0	0
	Adaptive learning	118 (0.23%)	23 (0.05%)	0	0	141 (0.28%)
	Fixed learning	121 (0.24%)	61 (0.12%)	17 (0.03%)	145 (0.29%)	344 (0.68%)

*The number of error cases are out of 50,193 cases

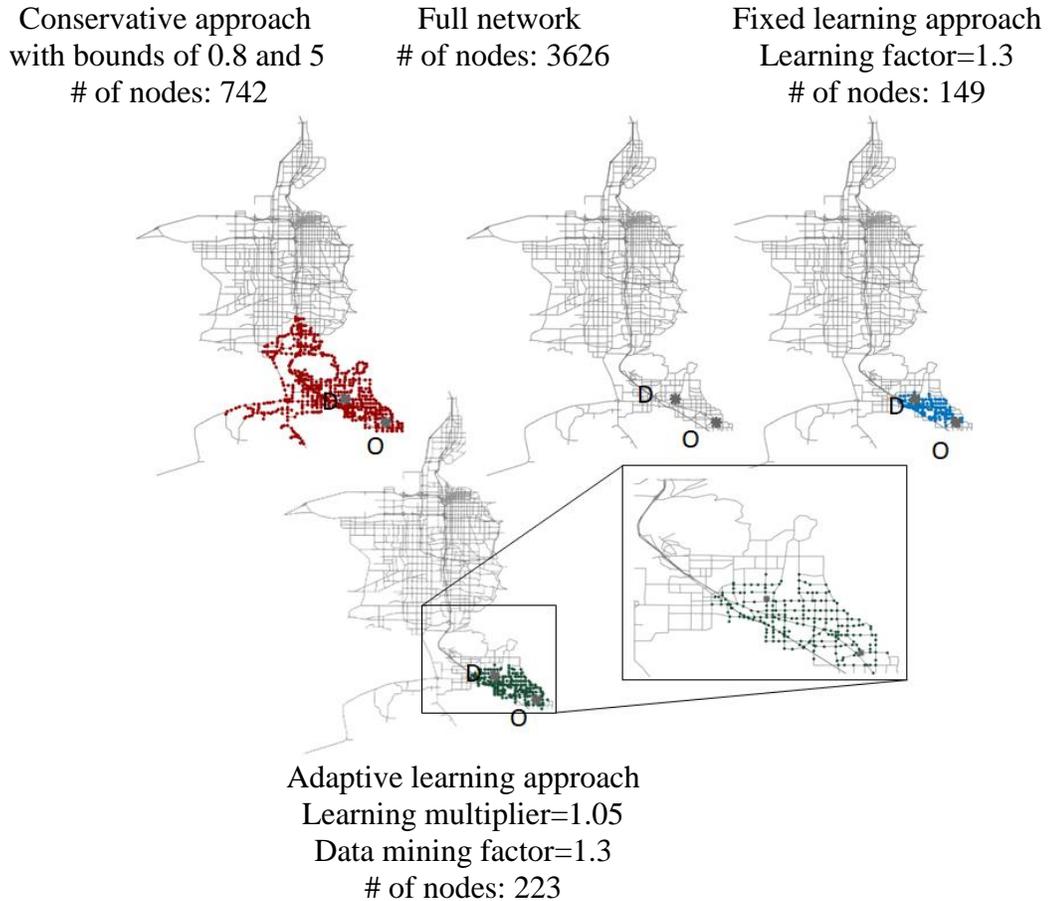


Figure 4-12- Sub-network configurations for the two approaches and the conservative approach from origin 1,178 to the destination 145 for the MTTBP problem applied to the Salt Lake City

4-6- Summary

The focus of this chapter was to develop a learning approach to decrease the network size for each specific OD pair within the initial iterations of the MCS approaches. The algorithm presented in this chapter takes advantage of the produced information during the travel time realizations of the initial iterations to find optimistic and pessimistic travel times for each OD pair of interest. First, conservative optimistic and pessimistic travel times for a certain OD pair are found. Comparing the optimistic travel time through any node to the pessimistic travel time for the OD pair specifies if the node should remain in the sub-network or not. Then, link travel time realizations are made by drawing random numbers from joint link travel time distributions for the

first n iterations of the first stage and the realized maximum and minimum path travel time labels from each node to the destination are defined. Use of the optimistic and pessimistic bounds further contract the sub-network. Seeking the right balance between accuracy and efficiency, two approaches, namely the fixed learning and the adaptive learning, are presented. The learning approaches are calibrated through the Chicago downtown network. The transferability of the approach to other networks is then checked using the network of Salt Lake City. The numerical results of the two large-scale applications lead to the following findings:

- The calibrated learning approaches for a particular large-scale network is used successfully for another large-scale application with a different structure of nodes and links. Although the errors in objective function values are larger in the Salt Lake City network relative to the calibration network, the level and amount of errors are still satisfactory.
- Significant reductions in the network size are observed using both learning approaches relative to the approach without any network contraction and the conservative approach.
- The learning approaches have an acceptable level and number of errors in terms of SPOTAR and MTTBP objective function estimations.
- The adaptive learning approach is superior to the fixed learning approach in terms of the accuracy of the objective function estimation, while the fixed learning approach is more aggressive in the network contraction. Therefore, a trade-off between a certain desired accuracy level and the computational efficiency is needed to select the proper learning approach.

CHAPTER 5 - Travel Time Reliability and Congestion Pricing

5-1- Overview

To make road pricing appealing to the public, many studies propose strategies for the distribution of toll revenues considering travelers' benefits and losses after pricing strategy implementation. Among the first in the line of research are Daganzo and Garcia (101) and Lawphongpanich and Yin (44) who proposed the Pareto-improving pricing strategy. Pareto-improving pricing refers to a scheme that does not make any traveler worse off, and makes at least one traveler better off in terms of generalized costs. Following this concept, several other studies propose Pareto-improving second-best pricing schemes for unimodal and bimodal networks (41,46). Redistributing toll revenues among travelers is also considered to avoid inequity problems associated with congestion pricing (43). In addition, in order to consider the variable impacts of congestion pricing on travelers with different socio-economic characteristics, trip purposes, and preferences, congestion pricing models that account for the variability of travelers' VOT have been developed. For example, Liu et al. (102) introduced a Pareto-improving pricing scheme, which is revenue-neutral for a bimodal network (i.e., transit and highway as different transportation modes). Their proposed scheme increases utility for all users and resolves the equity issue for travelers assuming a uniform VOT distribution. Nie and Liu (43) explored Pareto-improving schemes for different distributions of VOT and demonstrated that such a scheme always exists for concave VOT distributions. However, this is not the case for the realistic log-normal distribution. The impacts of congestion pricing on different traveler classes are also highly affected by the reliability valuation of network users. However, most studies have considered only different VOT classes, neglecting variations in the VOR.

As mentioned in the earlier chapters, travelers respond differently to travel time uncertainty, reflecting heterogeneity in their preferences and risk attitudes. Considering a reliability measure in travelers' path choice decisions would naturally impact the modeled congestion pattern in the network, which in turn, affects the outcomes of pricing strategies. Hence, incorporating measures regarding travel time reliability into congestion pricing schemes enhances the consistency of the schemes with travelers' route choice behavior. A few studies in the literature have considered applications of variable travel time in congestion pricing (7,8). For example, Li et al. (103) presented a bi-level reliability-based optimal toll design model in which travel time reliability is a network objective at a higher level. However, travel time reliability should also be considered as users' objective in route choice and traffic assignment. Boyles et al. (104) proposed an algorithm to find first-best pricing values for static transportation networks under daily capacity variations. They also defined the problem with travel time reliability and link travel time correlations. However, no solution algorithm was presented for this problem. In addition, the solution methodology considered only one single user class in terms of VOT and VOR. As such, two main limitations could be identified in existing models for equitable roadway pricing. First, most studies on congestion pricing either completely ignore travel time reliability or use simplified assumptions for its representation. Second, to our knowledge, there is no study that considers the RBUE in finding an equitable road pricing strategy for heterogeneous users with multiple VOT and VOR classes, which affects the fidelity of the traffic route assignment pattern in the network and hence the accuracy of the generated pricing schemes.

In this context, this chapter is motivated by the need to develop a modeling framework and efficient solution methodology for self-funded and Pareto-improving congestion pricing schemes. The framework ensures that all travelers are experiencing an improvement in their generalized

travel cost (utility) after deploying the pricing scheme. The framework explicitly captures the effect of travel time reliability on travelers' mode-route choice, considering heterogeneous travelers with multiple classes of VOT and VOR. In addition, for self-funded congestion pricing, revenues generated from the collected tolls could be utilized to improve public transportation services, subsidize the users of these services and/or compensate travelers who experience an increase in their generalized travel cost. However, there are concerns regarding the equity of the mechanism developed for toll revenues distribution. To overcome these concerns, the framework developed in this chapter is extended to address the self-funded congestion pricing problem. Two revenue distribution strategies are considered for self-funded pricing schemes in a bimodal network, namely the transit-based strategy and the credit-based strategy (43,48). For the transit-based strategy, the collected tolls from highway users are distributed only among transit users, reducing their travel cost and enhancing their regional accessibility. For the credit-based strategy, the collected tolls are distributed in the form of credits for all travelers (both private cars and transit users) to compensate for any increase in the travel cost.

The modeling framework developed in this chapter extends the second-best pricing optimization problem by integrating an RBUE algorithm. From the societal perspective, the objective of the pricing algorithm is to minimize total travel time of highway users given a revenue-neutral and Pareto-improving pricing scheme. Users' heterogeneity in response to the reliability measures, their response to different toll values and toll distribution strategies, and link travel time correlations are considered in the RBUE problem. A PSO algorithm is developed to determine the optimum toll values given the objectives of the current study and toll distribution strategies (e.g. credit-based, transit subsidy). The algorithm is applied to a modified Sioux Falls network with an area-based pricing strategy. The content of this chapter is published by the author in (105).

5-2- Problem Formulation

Consider a general bi-modal network, $G(N, \zeta)$, with a node set of N , a link set of ζ , and an OD set, rs . There is a fixed demand of q_{rs} for each OD pair, who opt their route among a set of highway paths, p_{rs} , and an alternative transit line, T_{rs} . Therefore, each origin, r , is connected to the destination, s , via multiple highway paths and a transit line which represents a slower but a more reliable path relative to other paths available for driving in uncongested traffic conditions. The travel time distribution of each highway link, ij , connecting node i to node j , is represented by the expected value of travel time, $E(t_{ij})$, and the standard deviation of travel time, $\sigma(t_{ij})$. Let $E(t_{rs}^T)$ and $\sigma(t_{rs}^T)$ denote the mean travel time and the standard deviation of travel time for transit line connecting origin, r , to destination, s . It is noted in the literature that transit travel time is affected by trip distance, the number of passengers, boarding and alighting procedures, and intersections with traffic signals (106,107). Thus, a smaller variation than the highway travel time variability is assumed for transit lines. Heterogeneous users are defined by m different VOT classes extracted from a VOT distribution. Each VOT class consists of k VOR classes that are selected from a VOT-dependent VOR distribution. α^m represents VOT of user class m , and β^{mk} represents VOR of the k^{th} class within the user class m with VOT of α^m . All notations and variables are summarized in the nomenclature table (Table 5-1).

Table 5-1- Definitions of parameters and variables

Sets	
$G(N, \zeta)$	A general bi-modal network with a node set of N , a link set of ζ
$(i, j) \in \zeta$	Highway link with upstream node i and downstream node j
Parameters	
C_o	Operational cost of auto per mile
C^A	Fixed cost of auto
λ	Pareto-improvement violation factor
η	A large constant
General Variables	
p_{rs}	Set of highway paths from origin r to destination s
T_{rs}	Transit line from origin r to destination s
α^m	Value of time of the user class m
β^{mk}	Value of reliability of the k th VOR class with VOT of α^m
$\delta_{ij}^{p_{rs}}$	A binary variable holding 1 if path p_{rs} passes link ij , and 0 otherwise
$E(t_{ij})$	Mean travel time of highway link ij
x_{ij}^A	Traffic flow on highway link ij
$\sigma(t_{ij})$	Standard deviation of travel time for highway link ij
$\tau_{p_{rs}}$	Charged toll on highway path p_{rs}
D_{ij}	Length of link ij
χ_{rs}	Credit distributed among users of each OD pair, rs
μ_{rs}^{mk}	Minimum travel cost from origin r to destination s for VOT class m and VOR class k at the user equilibrium state
$f_{p_{rs}}^{mk}$	Total number of users in VOT class m and VOR class k using highway path p_{rs} departing from origin r to destination s
$E(t_{rs}^T)$	Mean travel time of transit line between origin, r , and destination, s
$\sigma(t_{rs}^T)$	Standard deviation of travel time for transit line between origin, r , and destination, s
$C_{T_{rs}}$	Out of pocket cost of transit mode for the OD pair rs
$f_{T_{rs}}^{mk}$	Total number of users in VOT class m and VOR class k using transit
q_{rs}	Demand of all users between origin, r , and destination s
ΔG^{mk}	The change in generalized costs of a user in VOT class m and VOR class k before and after toll (after toll implementation minus before toll implementation)
$H(\Delta G^{mk})$	Heaviside step function of the generalized cost difference which holds 1 if $\Delta G^{mk} > 0$ and 0 otherwise
Δ_{ij}^z	A binary variable holding 1 if link ij is in zone z
$u_{ij}^{p_{rs}}$	A binary variable holding one for the link ij with the maximum toll along path p_{rs} and zero for other links
Decision Variables	
τ_{ij}^A	Charged toll on highway link ij
τ_z	Charged toll for zone z
τ_{rs}^T	Distributed subsidy among transit users of origin, r , and destination, s

Given these notations, the bimodal self-funded and Pareto-improving pricing problem considering travel time reliability for heterogeneous users is formulated as follows.

$$\min \sum_{(i,j) \in \zeta} x_{ij}^A E(t_{ij}) \quad (5-1)$$

Subject to

$$\begin{aligned} & \sum_{(i,j) \in \zeta} \alpha^m E(t_{ij}) \delta_{ij}^{p_{rs}} \\ & + \sqrt{\sum_{(i,j) \in \zeta} ((\beta^{mk})^2 \sigma^2(t_{ij}) \delta_{ij}^{p_{rs}}) + \sum_{(i,j) \in \zeta} \sum_{\substack{(l,k) \in \zeta \\ ij \neq lk}} ((\beta^{mk})^2 \sigma(t_{ij}, t_{lk}) \delta_{ij}^{p_{rs}})} \\ & + \sum_{(i,j) \in \zeta} C_o D_{ij} \delta_{ij}^{p_{rs}} + \tau_{p_{rs}} + C^A + \chi_{rs} = \mu_{rs}^{mk} \quad \text{if } f_{p_{rs}}^{mk} > 0, \forall m, k, p_{rs} \end{aligned} \quad (5-1-1)$$

$$\begin{aligned} & \sum_{(i,j) \in \zeta} \alpha^m E(t_{ij}) \delta_{ij}^{p_{rs}} \\ & + \sqrt{\sum_{(i,j) \in \zeta} ((\beta^{mk})^2 \sigma^2(t_{ij}) \delta_{ij}^{p_{rs}}) + \sum_{(i,j) \in \zeta} \sum_{\substack{(l,k) \in \zeta \\ ij \neq lk}} ((\beta^{mk})^2 \sigma(t_{ij}, t_{lk}) \delta_{ij}^{p_{rs}})} \\ & + \sum_{(i,j) \in \zeta} C_o D_{ij} \delta_{ij}^{p_{rs}} + \tau_{p_{rs}} + C^A + \chi_{rs} \geq \mu_{rs}^{mk} \quad \text{if } f_{p_{rs}}^{mk} = 0, \forall m, k, p_{rs} \end{aligned} \quad (5-1-2)$$

$$\alpha^m E(t_{rs}^T) + \beta^{mk} \sigma(t_{rs}^T) + \tau_{rs}^T + C_{T_{rs}} + \chi_{rs} = \mu_{rs}^{mk} \quad \text{if } f_{T_{rs}}^{mk} > 0, \forall m, k \quad (5-1-3)$$

$$\alpha^m E(t_{rs}^T) + \beta^{mk} \sigma(t_{rs}^T) + \tau_{rs}^T + C_{T_{rs}} + \chi_{rs} \geq \mu_{rs}^{mk} \quad \text{if } f_{T_{rs}}^{mk} = 0, \forall m, k \quad (5-1-4)$$

$$\sum_{p_{rs}} \sum_{m,k} f_{p_{rs}}^{mk} + \sum_{m,k} f_{T_{rs}}^{mk} = q_{rs} \quad \forall rs \quad (5-1-5)$$

$$\sum_{(r,s)} \sum_{p_{rs}} \sum_{m,k} f_{p_{rs}}^{mk} \tau_{p_{rs}} + \sum_{m,k} f_{T_{rs}}^{mk} \tau_{rs}^T + \sum_{(r,s)} q_{rs} \chi_{rs} \geq 0 \quad (5-1-6)$$

$$x_{ij}^A = \sum_{(r,s)} \sum_{p_{rs}} \sum_{m,k} f_{p_{rs}}^{mk} \delta_{ij}^{p_{rs}} \quad (5-1-7)$$

$$\Delta G^{mk} \leq 0 \quad \forall m, k \quad (5-1-8)$$

$$\tau_{p_{rs}} \geq 0 \quad \forall rs \quad \& \quad f_{p_{rs}}^{mk}, f_{T_{rs}}^{mk} \geq 0 \quad \forall m, k, rs \quad (5-1-9)$$

$$\begin{cases} \tau_{rs}^T = 0, \chi_{rs} \leq 0 & \text{Credit – based toll distribution strategy} \\ \tau_{rs}^T \leq 0, \chi_{rs} = 0 & \text{Transit – based toll distribution strategy} \end{cases} \quad \forall rs \quad (5-1-10)$$

The objective function, defined in Equation (5-1), minimizes the total travel time of highway users. This function is subject to the user equilibrium constraints, Constraints (5-1-1) to

(5-1-4), ensuring travelers choose the path with the least generalized cost. Constraints (5-1-1) and (5-1-2) describe the route assignment pattern for the private car users. Similarly, Constraints (5-1-3) and (5-1-4) describe the route assignment pattern for the transit users. The second term in Constraints (5-1-1) and (5-1-2) considers the correlation between the subsequent roadway links in the paths. As mentioned earlier, two approaches are considered in terms of subsidy distribution: transit-based and credit-based. In the transit-based strategy, the collected tolls from highway users are distributed only among transit users, τ_{rs}^T . In the credit-based strategy, a credit, χ_{rs} , is assigned to all users in the network. Constraint (5-1-5) expresses flow conservation in the network. Constraint (5-1-6) guarantees the self-funded (revenue neutrality) condition. Constraint (5-1-7) finds the highway link flows from paths flows. Constraint (5-1-8) ensures that all users are better off after toll implementation (Pareto-improving condition). Constraint (5-1-9) confirms that the variables associated with toll values, and path flows are non-negative continuous values. Finally, Constraint (5-1-10) defines the toll distribution strategy. Considering a transit-based toll distribution strategy, χ_{rs} is equal to zero. In case of a credit-based toll distribution strategy, τ_{rs}^T is equal to zero.

To satisfy Constraint (5-1-8), a scaled value of the rectified linear unit (ReLU) function of ΔG^{mk} is added with a large factor to the objective function as below.

$$\min \sum_{(i,j) \in \zeta} x_{ij}^A E(t_{ij}) + \lambda \sum_{(r,s)} \sum_{p_{rs}} \sum_{m,k} \frac{\Delta G^{mk} H(\Delta G^{mk})}{\alpha^m} \quad (5-2)$$

The second term in Equation (5-2) minimizes the users' loss due to toll implementation. In this equation, each user class's ReLU function of ΔG^{mk} is divided by the VOT value of the user class to make the units of the objective function terms consistent. The decision variables in this formulation are path toll, and transit user's subsidy or user's credit. Nonetheless, the proposed

formulation (and the solution methodology presented hereafter) is in general form and can be applied to any pricing strategy, such as link-based, corridor-based, cordon-based, and so on. For instance, in case a link-based pricing strategy is considered, the path toll is defined as follows.

$$\tau_{p_{rs}} = \sum_{(i,j) \in \zeta} \tau_{ij}^A \delta_{ij}^{p_{rs}} \quad (5-3)$$

In the numerical experiment section of this chapter, a zone-based pricing strategy is considered. To this end, the network is divided into multiple zones, z , and users are charged with the maximum toll value of the zones they pass. Therefore, the decision variable changes to the tolls on the links of each zone, τ_z . Given a zone-based toll implementation strategy, the following constraints should be considered to the mathematical program presented in (5-1).

$$\tau_{p_{rs}} \geq \sum_z \tau_z \Delta_{ij}^z \delta_{ij}^{p_{rs}} \quad (5-4-1)$$

$$\tau_{p_{rs}} \leq \sum_{(i,j) \in \zeta} \sum_z \tau_z \Delta_{ij}^z \delta_{ij}^{p_{rs}} + \eta(1 - u_{ij}^{p_{rs}}) \quad (5-4-2)$$

$$\sum_{(i,j) \in \zeta} \delta_{ij}^{p_{rs}} u_{ij}^{p_{rs}} = 1 \quad (5-4-3)$$

5-3- Methodology

The problem presented above is a non-convex mathematical problem with user equilibrium constraints, which is a class of problems difficult to solve, especially for large-scale networks. The objective function is not convex considering the ReLU function included to ensure the Pareto-improving condition of the selected pricing scheme. In addition, in this chapter a concave relation between the mean and standard deviation of the travel time is considered for each link to estimate its travel time variability. This problem is computationally demanding and its solution could be time-consuming without a proper approach, as the RBUE algorithm should be called multiple

times to find the optimal toll values. As gradient descent type algorithms are unable to solve this problem due to the non-convexity, we propose to use a metaheuristic algorithm based on PSO (108). Different metaheuristic algorithms are used in the literature to solve non-convex problems in transportation studies (109,110). Considering the distributed multi-agent search structure of the problem in this study, PSO has proved to be successful in a variety of problem domains, which motivated its use in this study. Different variants of the PSO algorithm are used for optimizing complex transportation problems (110–112). PSO is a population-based search algorithm and uses a swarm of particles to optimize the problem. Each particle has two features that identify its movements: position and velocity. The particle position represents a feasible solution for the problem's decision variables (i.e., toll values). The particle velocity defines the step size as the particles move within the feasible region in each iteration. This feasible region typically belongs to a multi-dimensional search space, where the particles move towards the best positions of individual particles (local best values) and the position of the entire swarm (global best value) (111,113). Several techniques are developed in the literature to represent the particles movement (114). In this study, we use the Clerc and Kennedy PSO approach to move particles within the search area (113).

For each particle, j , a position is first initialized (an initial solution) with a uniformly distributed random vector ($P_j \sim U(b_{low}, b_{up})$). The particle's best-known position is initialized to its initial position ($l_j := P_j$). The swarm's best-known position is also updated with the particle's position that have the lowest objective function value among all particles (g). While the maximum number of main iterations (I_{max}) is not reached, the velocity and position of each particle in each iteration, I , are updated using the following formulations.

$$v_j \leftarrow \omega v_j + \varphi_l r_l (l_j - p_j) + \varphi_g r_g (g - p_j) \quad (5-5)$$

$$P_j \leftarrow P_j + v_j \quad (5-6)$$

The first term in Equation (5-5) is the velocity of the particle in previous iteration (v_j), which is multiplied by an inertia weight factor, ω , to balance between the local and global positions. The second term moves the particle towards the local best position of the particle with the acceleration coefficient, φ_l , along with a random variate between zero and one, r_l , which prevents trapping in local optima. The last term moves the particle towards the global best position with the acceleration coefficient, φ_g , and random variate, r_g . In Clerc and Kennedy PSO approach, the acceleration coefficients are defined as follows.

$$\varphi_l = \psi \phi_1, \quad \varphi_g = \psi \phi_2 \quad (5-7)$$

where

$$\psi = \frac{2\kappa}{|2 - (\phi_1 + \phi_2) - \sqrt{(\phi_1 + \phi_2)^2 - 4(\phi_1 + \phi_2)}} \quad (5-8)$$

Here, κ , ϕ_1 , and ϕ_2 are constants to be defined. Once the position of the particle is updated using the velocity value and the previous position (Equation 5-6), the particle's local best position (l_j) and the swarm's best-known position (global best or g) are updated in case of finding an improved objective function value. The definitions of particle swarm optimization algorithm parameters and variables are summarized in Table 5-2.

An RBUE algorithm is embedded in the PSO algorithm to find the objective function value pertaining to each toll value (particle). The proposed path-based formulation cannot be applied to large-scale networks, since the number of paths increases exponentially with the network size. To address this issue a column generation approach is implemented to generate the set of used paths

by travelers. The details of the algorithm to solve the RBUE problem is described hereafter. To satisfy the self-funded condition, Constraint (5-1-6), the collected tolls are distributed among transit users or all users. As the current study can be applied to any toll distribution strategy, it should be specified as part of the input of the PSO algorithm. If a credit-based toll distribution strategy is selected, the credits are distributed among users at each iteration of the PSO algorithm. This credit can be used by users to pay toll or compensate the travel time loss. In the non-credit based or transit-based approach, the collected tolls are distributed at each iteration of the RBUE algorithm to ensure the self-funding condition. An illustration of the PSO algorithm to find the optimal toll values considering reliability and equity is shown in Figure 5-1.

Table 5-2- Particle Swarm Optimization Algorithm Parameters and Variables

Parameter /Variable	Definition
P_j	Position of particle j in the particle swarm optimization algorithm
b_{low}, b_{up}	Lower/Upper bound of toll values
l_j	Best-known position of particle j
g	Swarm's best-known position
I, I_{max}	Iteration/Maximum iteration number in the particle swarm optimization algorithm
v_j	Velocity of particle j in each iteration of the particle swarm optimization algorithm
φ_l, φ_g	Acceleration coefficients of the particle and swarm
r_l, r_g	Random variates between zero and one
ϕ_1, ϕ_2	Constants equal to 2.05 in Clerc and Kennedy approach
K	A constant in the particle swarm optimization algorithm
ω	Inertia weight factor
n, n^{max}	Iteration/Maximum iteration number in the reliability-based user equilibrium algorithm
$f(n)$	Step size at iteration n
$GC_{current}^{rs,mk}$	Generalized cost of the current path holding by VOT class m and VOR class k for each OD pair rs at each iteration of RBUE algorithm
$GC_{best}^{rs,mk}$	Minimum generalized cost for VOT class m and VOR class k for each OD pair rs at each iteration of RBUE algorithm
$\Gamma^{mk}(p_{rs})$	The gap between the generalized costs of the current path, p_{rs} , and the best path
Γ_{total}^n	The total gap of all user classes in the reliability-based user equilibrium algorithm at each iteration, n
ϵ	A small number that ensure convergence of the reliability-based user equilibrium algorithm

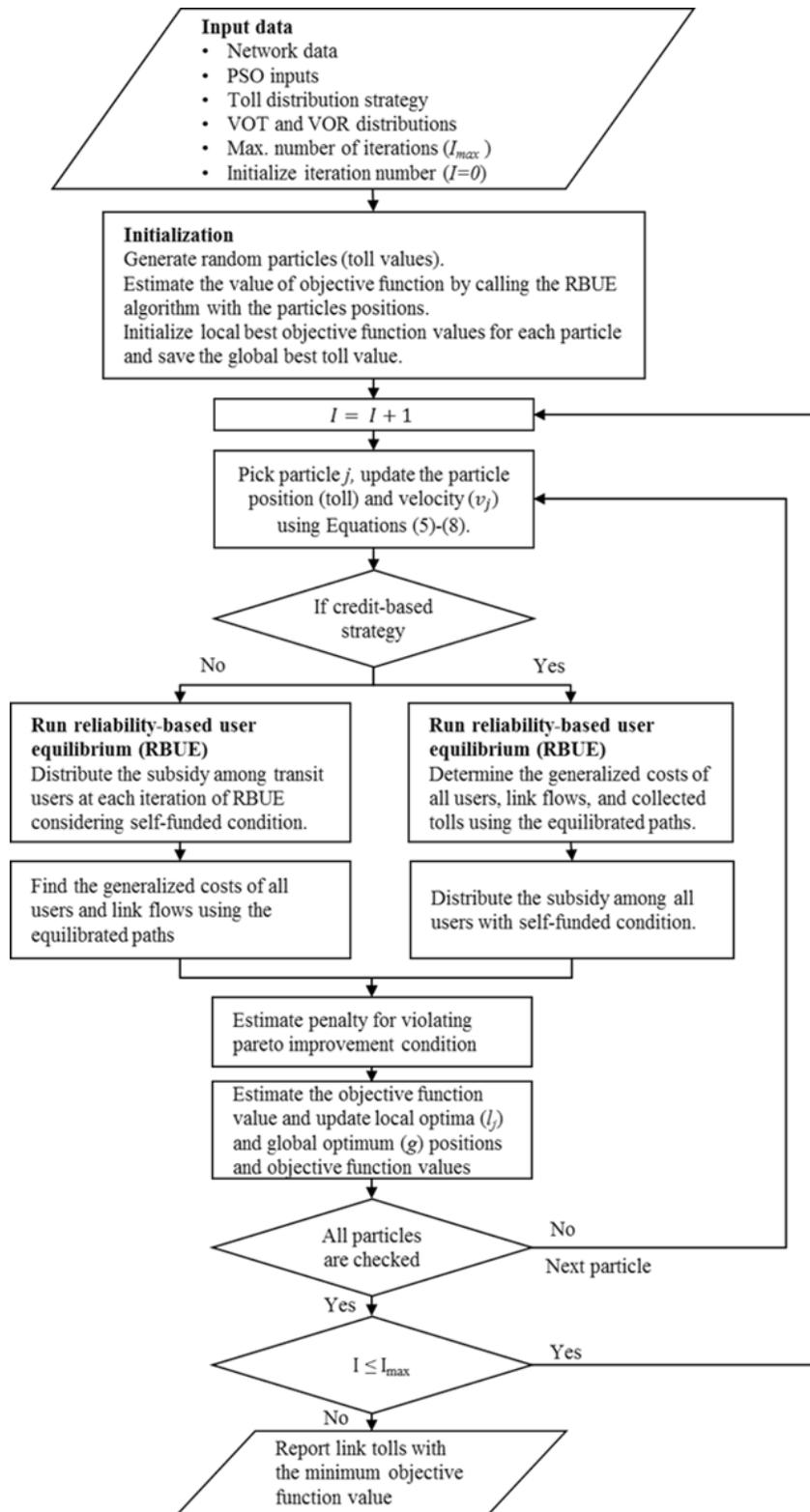


Figure 5-1- Particle swarm optimization algorithm to find optimal toll values minimizing highway travel time and violation of Pareto-improving condition considering reliability

As illustrated in Figure 5-1, the RBUE problem is solved for each particle's position (set of toll values) at each iteration of the PSO algorithm. The output of the RBUE algorithm includes the link flows, the collected tolls from highway users and the distributed subsidy among transit/all users. The output also includes the generalized costs of all users in the network, which are used to check the Pareto-improvement criterion. In this chapter, a heuristic approach is used to solve the RBUE problem in which each user class finds its least generalized cost path. The approach consists of three main components (1) a reliability-based optimal path finding procedure that can reflect heterogeneous users' preferences; (2) a column generation approach to generate the set of superior paths for each traveler; and (3) an algorithmic iterative process for redistributing user choice outcomes to achieve the desired equilibrium state.

For the path finding sub-problem, a Monte-Carlo simulation-based approach, adopted from the literature (86,115), is used. The uniqueness of this approach is to consider link travel time correlations and heterogeneous travelers in terms of reliability preferences. The methodology generates path travel time distributions for a set of candidate paths in a stochastic network based on a joint link travel time distribution at the network level. The optimal route choice of each user class is found using the least generalized cost path among the set of candidate paths. Furthermore, a variant of the method of successive averages (MSA) is used to redirect flow to the optimum path at each iteration of the RBUE algorithm (116). The likelihood of re-assigning a portion of user class from its current path to the optimal path, found for that user class in each iteration, depends on the iteration number in RBUE. Therefore, as the algorithm proceeds, a smaller portion of each user class is moved to the newly generated least generalized cost path. The algorithm converges when a pre-defined gap measure is smaller than a small threshold. The algorithm used for solving the RBUE problem is illustrated in Figure 5-2.

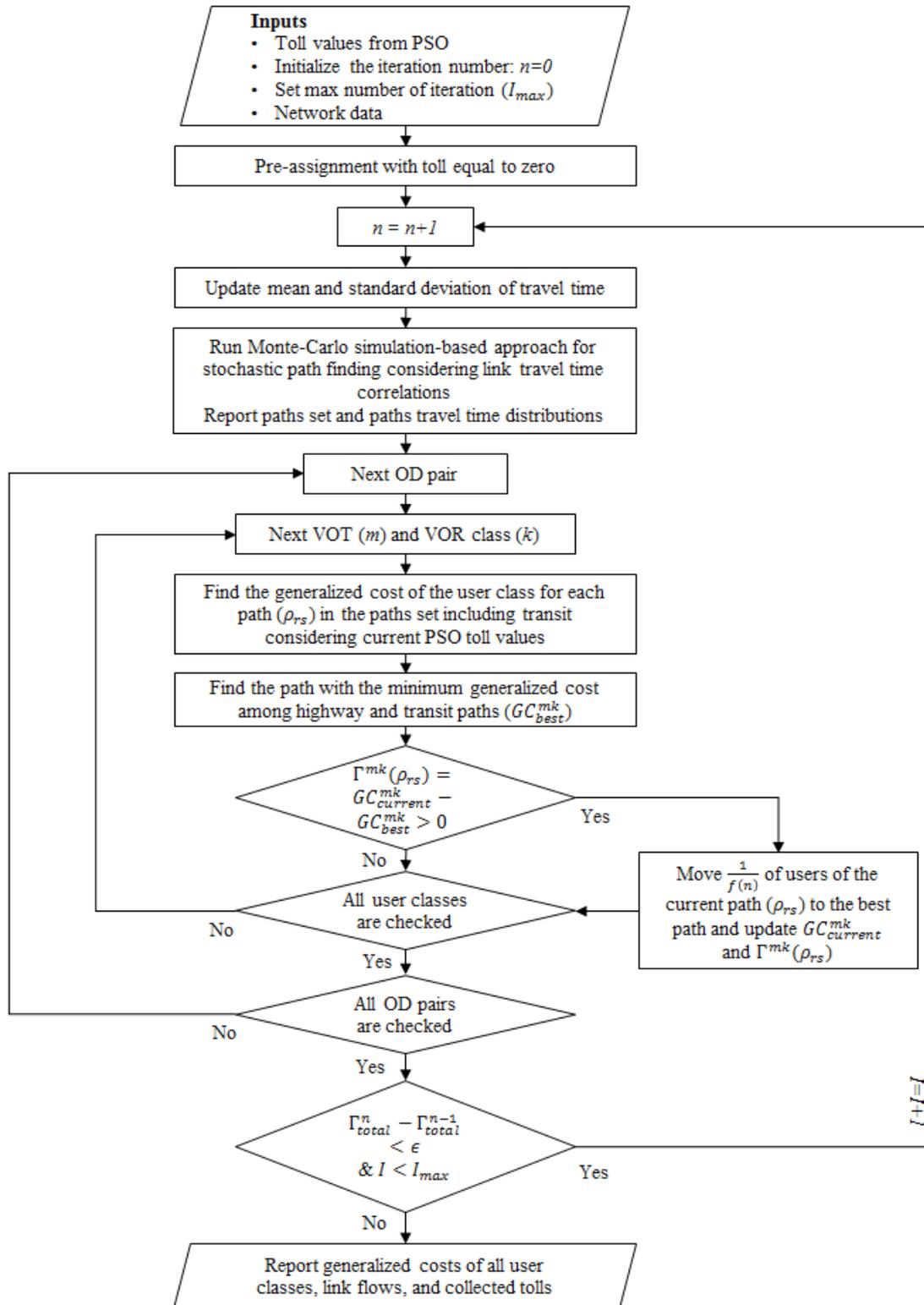


Figure 5-2- The reliability-based user equilibrium algorithm considering toll values and heterogeneous users

5-4- Numerical Experiments

The solution methodology described above is applied to a modified version of the Sioux Falls network, which consists of 24 nodes, 76 links, and 528 OD pairs (117). The network is modified in several ways. First, a transit line is assumed to exist between every OD pair in the original network to be able to examine the impact of implementing the pricing scheme on users' mode choices in a bimodal network. Second, considering the new bimodal network, the demand level for each OD pair is doubled to maintain a moderate level of congestion close to that in the original network (called "modified network" hereafter). In addition, a network scenario is considered with a higher demand level for highway links to investigate the impacts of the congestion level on pricing strategies. We refer to this network in the remaining of the chapter as the "congested network". Finally, two toll zones are assumed, where a specific toll value is applied to all links in the same zone. If a vehicle passes through both zones, the maximum toll between the two zones is charged. The schematic view of the modified Sioux Falls network, the new ratios of flow to capacity for each link in the non-tolled equilibrium state, and the two defined zones for applying tolls are shown in Figure 5-3a. To consider travel time variability, the standard deviation of the travel time is considered for each link, in addition to the mean travel time. There is no consensus in the literature on the relation between standard deviation of link travel time and its mean travel time. Mahmassani et al. (18) presented a linear relationship for this purpose. This linear relationship is more appropriate at the network level, and at the link level it is scattered. The proposed linear relation for the link level in their study is as follows.

$$\frac{\sigma(t_{ij})}{D_{ij}} = 0.99 \left(\frac{E(t_{ij})}{D_{ij}} \right) - 0.47 \quad (5-9)$$

In addition, many studies claim that standard deviation of travel time increases up to a certain point of mean travel time and then decreases with a concave relation (19). Adopting this

line of thoughts, the relationship between the standard deviation and the mean travel time values is obtained by fitting a second order polynomial function as given below using simulated travel time data for Chicago downtown network during 5-10 AM (15).

$$\frac{\sigma(t_{ij})}{t_f} = -0.16 \left(\frac{E(t_{ij})}{t_f} \right)^2 + 2.31 \left(\frac{E(t_{ij})}{t_f} \right) - 2.15 \quad (5-10)$$

where t_f is the free flow link travel time. The fitted curve to the simulation data with the R-square value of 0.93 is illustrated in Figure 5-3b. The algorithm of the current study is implemented using both relations to estimate standard deviation of travel time based on the estimated mean travel time (Equation (5-9) and Equation (5-10)), which itself is a function of link flow. The mean travel times of a transit line is arbitrarily assumed to be equal to four times of the deterministic path travel times between their relevant OD pairs assigning free flow travel times to all links. Due to lack of literature on the travel time variability of transit lines, they are assumed to be half of their mean travel time. In addition, a fixed travel time of six minutes, which is equivalent to the average walking distance of 0.25 mi, is considered for all transit users to consider extra time needed for walking to the station and getting on/off the transit line. A fixed fare of \$1 is assumed for all transit lines. The private car operation cost is assumed to be \$0.6 per mile (48).

As illustrated in Figure 5-4, the VOT distribution among travelers for each OD pair follows a log-normal distribution. This log-normal distribution is assumed to have an average value of \$21/hr., and a standard deviation of \$10.5/hr (5). The VOT distribution is discretized into ten classes from \$4/hr to \$60/hr as given in Figure 5-4b. Furthermore, each VOT class contains a uniform distribution of VOR ranging from 0.69 times to 1.12 times of the VOT (55). Four VOR classes are extracted for each VOT as illustrated in the figure.

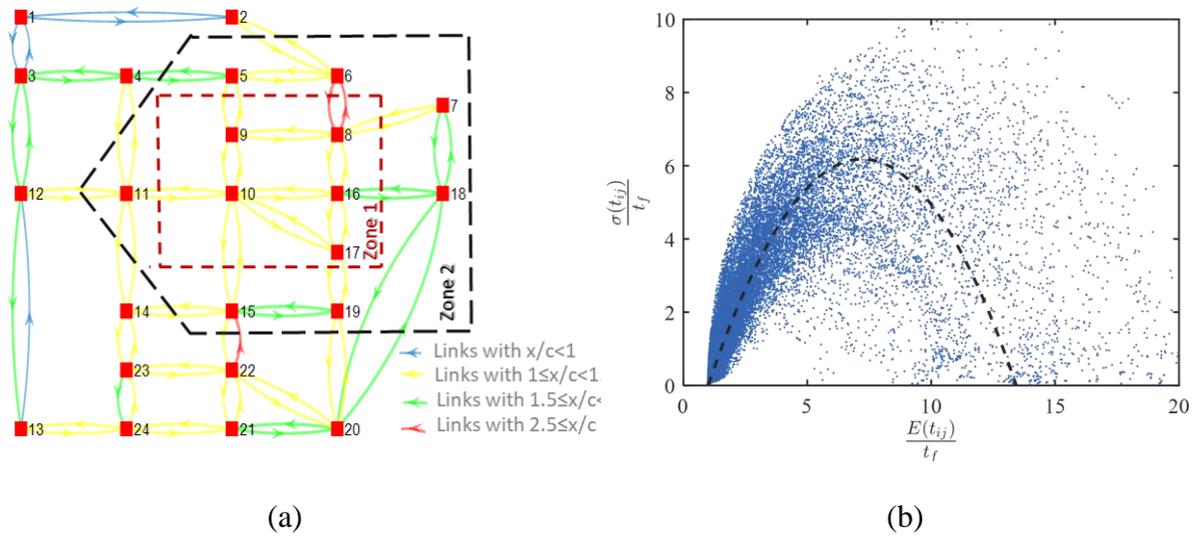


Figure 5-3- (a) Selected toll zones on the modified Sioux Falls network (x represents link flow and c denotes the capacity of the link) (b) the fitted concave function between standard deviation and mean travel time normalized to free flow travel time

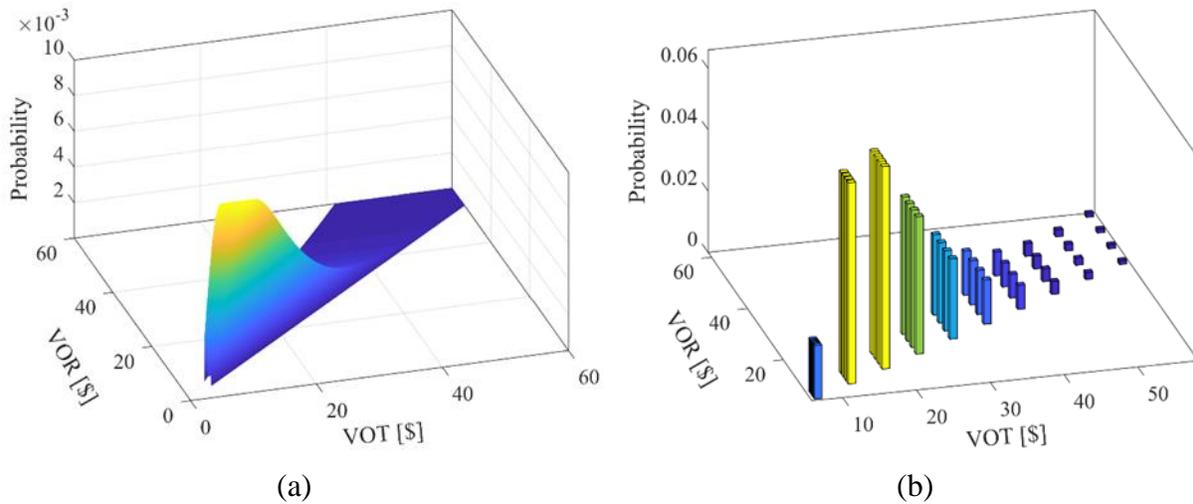


Figure 5-4- Value of time (VOT) and value of reliability (VOR) (a) distributions and (b) classes

Four toll distribution strategies are considered in this chapter: C_{All} , C_{OD} , T_{All} , and T_{OD} , which are defined in Table 5-3. Implementing the OD-based strategies is challenging, since it is not easy to differentiate travelers based on their OD pairs. However, the subsidy can be spent more efficiently on increasing transit utility for specific OD pairs or cluster travelers based on their home/work locations. Previous studies also have shown that finding a self-funded and Pareto-

improving pricing scheme that distributes tolls evenly among all network/transit users is not always achievable (48,49). The PSO algorithm is applied to the modified Sioux Falls network considering the two demand levels (i.e., modified network, congested network defined earlier) and the toll distribution strategies defined above. In the current implementation, the parameters κ , ϕ_1/ϕ_2 , ω , I_{max} , and η of the PSO algorithm are set at 1, 2.05, 0.99, 15, and 30, respectively. In addition, the maximum gap in the RBUE algorithm is assumed to be 0.0001. The lower and upper bounds of toll values (b_{low}/b_{up}) are also considered to be zero and five dollars, respectively.

Figure 5-5 shows the search area and the objective function values (Equation (5-2)) for a scenario in which the credit-based strategy is adopted for the high demand case. This figure certifies that the algorithm searches the entire feasible space and moves efficiently towards the optimum toll values for zone 1 and zone 2 (\$3.44 and \$1.18). The figure also illustrates the non-convexity of the objective function, which precludes the application of gradient-based search algorithms to adequately solve this problem.

Table 5-3- Toll Distribution Strategies

Toll Distribution Strategy	Definition
C _{All}	Toll distribution strategy, which allocates uniform credit to the entire network users
C _{OD}	Toll distribution strategy, which allocates OD specific credits with uniform distribution among users of each OD pair
T _{All}	Toll distribution strategy, which collects toll from highway users of all OD pairs and distributes evenly among all transit users
T _{OD}	Toll distribution strategy, which collects toll from highway users of each OD pair and distributes evenly among transit users of the same OD pair

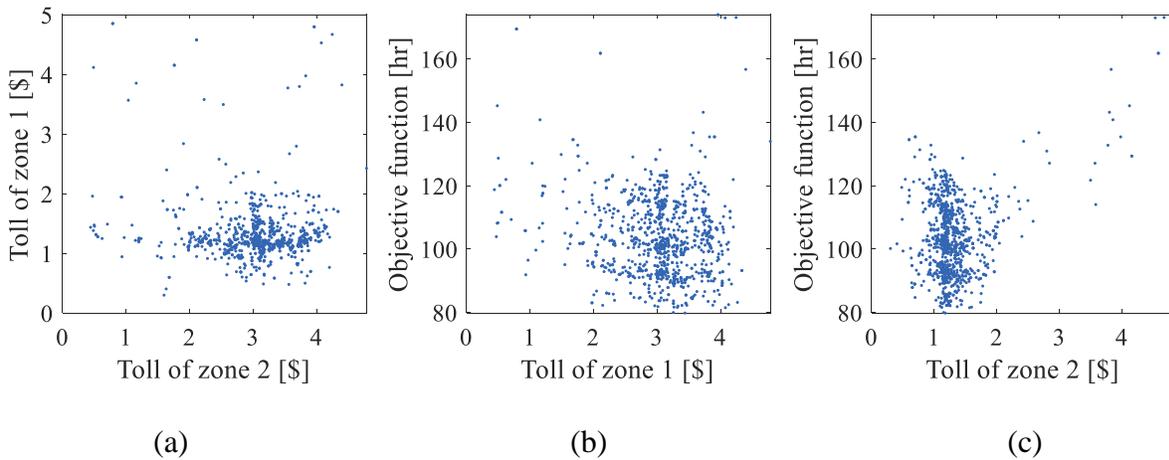


Figure 5-5- The search area for the credit-based toll distributed among the same OD pairs of the congested Sioux Falls network

Table 5-4 summarizes the results of applying the self-funded and Pareto-improving pricing scheme for the different toll distribution strategies using both linear and concave relationships between the standard deviation and mean of the links' travel times. The results are given for the congested and the modified networks. Several measures of performance are recorded for each case including the optimum tolls, the improvement in travel time, total generalized cost of users, and the average loss in the users' generalized cost. As illustrated in table, there is no self-funded and Pareto-improving toll value for the congested and modified networks, once the collected tolls are distributed evenly among all network/transit users. However, if collected tolls are distributed among travelers of the same OD pair, almost all scenarios find self-funded and Pareto-improving pricing values that can be applied to zone 1 (inner zone) and zone 2 (outer zone), respectively. In addition, credit-based toll distribution strategies impose higher tolls on links relative to that of the transit-based strategies. Accordingly, more improvements in the private car users' total travel time and all travelers' total generalized cost, considering their VOT and reliability valuation, is recorded for the credit-based strategy compared to the transit-based one. Furthermore, one can also observe the successful application of the penalty factor incorporated into the objective function to provide

Pareto-improving pricing schemes. The average loss in the travelers' generalized cost due to toll implementation is insignificant (less than 1 cent) and can be neglected.

Table 5-4- Optimum tolls and the resulting system and users' costs for different network scenarios and toll distribution strategies for different reliability relations

Network Scenario		Congested Network				Modified Network			
Toll Distribution Strategy	Reliability Relation	C _{OD}	C _{All}	T _{OD}	T _{All}	C _{OD}	C _{All}	T _{OD}	T _{All}
Toll values (Inner zone/Outer zone) (\$)	Linear	2.37/ 1.38	0/0	2.29/ 0.41	0/0	1.56/ 0.70	0/0	0.11/ 0	0/0
	Concave	3.44/ 1.18	0/0	3.04 /0.26	0/0	1.54/ 0.17	0/0	0/0	0/0
Total collected toll (\$)	Linear	570	0	358	0	340	0	18	0
	Concave	662	0	410	0	270	0	0	0
Users violating Pareto-improving condition (%)	Linear	0.90	0	2.23	0	0.23	0	0	0
	Concave	2.37	0	1.69	0	0.62	0	0	0
Average loss in the users' generalized cost (\$)	Linear	0.003	0	0.008	0	<0.00 1	0	0	0
	Concave	0.002	0	0.003	0	0.001	0	0	0
Total generalized cost before pricing (\$)	Linear	16,422	16,422	16,422	16,422	9,105	9,105	9,105	9,105
	Concave	16,343	16,343	16,343	16,343	8,968	8,968	8,968	8,968
Total generalized cost after pricing (\$)	Linear	15,749	16,422	15,644	16,422	8,762	9,105	9,012	9,105
	Concave	15,546	16,343	15,469	16,343	8,656	8,968	8,968	8,968
Total travel time of highway users before pricing (hr.)	Linear	96.00	96.00	96.00	96.00	68.69	68.69	68.69	68.69
	Concave	89.78	89.78	89.78	89.78	69.15	69.15	69.15	69.15
Total travel time of highway users after pricing (hr.)	Linear	80.75	96.00	75.55	96.00	60.68	68.69	66.61	68.69
	Concave	76.51	89.78	73.52	89.78	64.03	69.15	69.15	69.15

To highlight the importance of considering travel time reliability for developing self-funded Pareto-improving pricing strategy, the tolls are estimated without considering travel time

variability (fourth column of Table 5-5) for the scenario in which the credit-based toll distribution strategy is implemented for the congested network. The resulting tolls are then substituted into the network with travel time reliability (fifth column of Table 5-5). Note that for the results in the fourth column, no optimization is executed and the toll values are just used to run the RBUE algorithm. These results are then compared with the optimum results that consider travel time reliability (third column of Table 5-5). Although the tolls resulted from the scenario without considering travel time reliability are close to be Pareto-improving and self-funded, the results show that the optimum points with and without considering travel time reliability are significantly different. Furthermore, not considering travel time reliability overestimates the expected benefits. For example, using the concave relationship while ignoring travel time reliability, it is estimated to collect \$679 and have 20.86% improvement in travel time of private car users. However, substituting the toll values of this scenario into the realistic scenario, in which travel time reliability is considered, results in a total toll revenue of \$511 and 12.97% reduction in the total travel time. Therefore, it is important to consider travel time reliability in the design of the pricing scheme to obtain more accurate estimation of collected tolls and the improvement in the overall network performance.

Table 5-5- Comparison of the results with and without considering travel time reliability for the congested network and the credit-based toll distribution strategy among the same OD pair

Performance Measure	Reliability relation	With considering travel time reliability	Without considering travel time reliability	Substituting the results of without reliability in the realistic scenario
Toll values (Inner zone/Outer zone) (\$)	Linear	2.37/1.38	0/2.43	0/2.43
	Concave	3.44/1.18	0/2.43	0/2.43
Total collected toll (\$)	Linear	570	679	492
	Concave	662	679	511
Users violating Pareto-improving condition (%)	Linear	0.90	3.15	3.97
	Concave	2.37	3.15	5.80
Average loss in the users' generalized cost (\$)	Linear	0.003	0.011	0.015
	Concave	0.002	0.011	0.013
Improvement in generalized cost by toll implementation (%)	Linear	4.10	7.36	3.82
	Concave	4.88	7.36	3.86
Improvement in travel time of highway users by toll implementation (%)	Linear	15.88	20.86	16.76
	Concave	14.78	20.86	12.97

5-5- Summary

This chapter presented a modeling framework and solution methodology for finding revenue-neutral and Pareto-improving congestion pricing values considering travel time variability for heterogeneous users with different VOTs and VORs in a bi-modal network consisting of transit and personal cars. The framework extends the second-best pricing optimization problem by integrating an RBUE algorithm. The objective of the pricing algorithm minimizes the total travel time of highway users given a revenue-neutral and Pareto-improving pricing set. Users' heterogeneity in response to the reliability measures, their response to different toll values and toll distribution strategies, and link travel time correlations are considered in the

RBUE problem. A PSO algorithm is developed to determine the optimum toll values considering different toll distribution strategies (e.g. credit-based, transit subsidy).

The proposed algorithm is applied into a modified Sioux Falls network with different levels of congestion. In addition, two credit-based toll distribution strategies (i.e. OD-based versus among all users) and two transit-based toll distribution strategies are compared in the numerical experiments. This chapter also explored the impacts of different types of relationships (i.e. linear, concave) between mean and standard deviation of link travel time. The results show the importance of considering travel time reliability by comparing the optimum toll values, the changes in system cost, and the total travel time with and without considering travel time reliability. In addition, the results demonstrate that the PSO algorithm successfully searches the feasible region and efficiently finds the optimum toll values. Comparing the results of different toll distribution strategies also confirms that there is no self-funded and Pareto-improving pricing for the networks of this chapter, once the collected tolls are distributed evenly among all network/transit users. However, if the collected tolls from travelers of each OD pair are used to subsidize only travelers of the same OD pair, such pricing values are successfully found. This subsidization can be done through investments on the transit lines or other infrastructure of the OD pairs to avoid violating equity.

CHAPTER 6 - Impacts of Connected and Autonomous Vehicles on the Network Travel Time Reliability

6-1- Overview

Drivers' behavior plays a pivotal role in the development of new transportation technologies, such as connected and automated vehicle technology, and planning for future transportation systems. By the advent of CV technology, driver-vehicle behavior is expected to change significantly, since these vehicles can communicate with each other and also traffic management centers on a real-time basis. Travelers' short-term decisions would change remarkably through V2V communications. Moreover, V2I communications may affect the long-term decisions made by traffic management systems (69). In addition, the driving logic of an AV is different from the human-driven vehicles, as humans have higher reaction times compared to robots. Despite the fact that increasing levels of connectivity and automation of vehicles is expected to improve mobility and safety in transportation systems, there is a growing concern about the long-term impacts of these technologies on transportation planning due to the possible induced demand. As such systems have not been deployed in practice at large-scale, a traffic simulation tool that considers the impacts of connected and automated vehicle technology on travelers' behavior, transportation planning, and travel forecasting is the only available tool for transportation authorities. Therefore, from the modeling standpoint, the presence of connected and autonomous vehicles needs to be incorporated into traffic simulations.

Simulation-based DTA tools utilize mesoscopic traffic simulation to capture traffic dynamics at the network level. DTA frameworks incorporate drivers' response to different sources of information, so as to model traffic flows over a planning horizon in a transportation network. Time-dependent travel times provided by DTA models can be utilized in travel choice modeling

practices, and as a result, generate proactive strategies for systems based on predicted traffic conditions (118,119). Given that simulation-based DTA tools can be used as effective analysis tools to evaluate a broad range of operational and strategic improvements to transportation networks, having simulation models that are sensitive to the interactions between regular (human-driven vehicles without connectivity), connected (human-driven vehicles with connectivity), and autonomous vehicles plays a pivotal role in predicting future traffic dynamics under emerging vehicle technologies.

The effects of connectivity and automation of vehicles and their applications on reducing congestion and improving stability and throughput have been widely studied at the segment level (14,63,65,67). A complementary approach to gain further insights into the future of transportation systems is to investigate large-scale impacts of connectivity and automation at the network level. Most of the studies in this area assume a uniform distribution of CVs or AVs throughout the network. However, in reality, these vehicles are not distributed uniformly over all network links. For instance, the presence of en-route (adaptive) users results in various distribution of traffic flow in the network. These users are aware of the current traffic conditions at different regions of the network via access to real-time information. Thus, they may use alternate routes in the case of congestion or gridlock on initially selected routes (120,121). CVs and AVs are expected to have access to this real-time information through V2I information. Therefore, a portion of CVs and AVs that are assigned to the network, might use alternate paths, instead of their initial choice of route. In addition to the en-route users, drivers of CVs and HDV are heterogeneous and have different acceleration behavior depending on their level of risk acceptance and information availability. Therefore, different driving patterns can be considered even for a same vehicle type to account for the impacts of heterogeneous drivers in the traffic simulation model. This is mainly neglected in

most of the mesoscopic simulation tools for large-scale networks. Furthermore, there are only a few studies in the literature that consider the impacts of a mixed traffic condition on the capacity of intersections (at the network level, and not at a single intersection level), which is an important factor affecting the traffic flow of the network. Finally, it is established that traffic flow models are different in freeways and arterials at a macroscopic level. However, there is not much data available to fit microscopic models for arterials.

In this study, HDV refers to human-driven vehicles that may have access to en-route information via GPS or similar applications. CVs are human-driven vehicles that are connected to each other and have access to the information of their surrounding vehicles. Finally, AV refers to a self-driving vehicle, which is connected to other vehicles. In light of these definitions, this chapter aims to realistically observe the impacts of CV and AV technologies on traffic flow and travel time reliability at the network level. The simulation-based DTA tool of DYNASMART-P is updated and used as the base tool to simulate traffic dynamics at the network level (85). This chapter also incorporates different microscopic modeling frameworks for various vehicle types (i.e., HDV, CV, AV) and captures the collective effects of the interactions between them on traffic flow dynamics. The stochastic acceleration model of Hamdar et al. (12), which avoids (most) crashes by a perceived probability, is used to model the car-following behavior of HDVs. The acceleration behavior of CVs is modeled based on the Intelligent Driver Model (IDM) (13), which is a deterministic model. Furthermore, the model of Talebpour and Mahmassani (14) is utilized for the car-following behavior of AVs. In order to translate traffic flow dynamics from a micro-scale to a meso-scale, the relationship between spacing and speed for each vehicle type is derived and used as an input to the mesoscopic model. The proportion of each vehicle type and driver class (if there is any) on each link is tracked in the traffic propagation process at each time interval. Using

this proportion, a non-linear equation is solved to obtain the current speed of the link, satisfying the spacing values of all vehicles traversing it. Note that NGSIM data is used to establish different sets of parameters for the microscopic models of heterogeneous drivers (122) for HDVs and CVs. The content of this chapter is published by the author in (123).

6-2- Microscopic Simulation Framework

This study uses state-of-the-art microscopic simulation frameworks to model the acceleration behavior of different vehicle types with various levels of connectivity and automation. The details of these microscopic models for vehicles with no communication capability, CVs, and AVs are explained in the following sub-sections.

6-2-1- Vehicles with No Communication Capability

The stochastic acceleration model of Hamdar et al. (12) is used in this chapter to model the car following behavior of human-driven vehicles without any communication capability (HDVs). This model considers the acceleration decision of drivers based on the evaluation of potential gains and losses, from the acceleration and deceleration decisions with a perceived probability of being involved in a crash. This model is an extension of the sequential risk-taking model of Hamdar et al. (124) that generally avoids crashes by considering the behavioral mechanisms of the prospect theory (PT) (125). Accordingly, Hamdar et al. (12) introduced a value function for the subjective utilities of accelerations (i.e. gain in travel times), U_{PT} , as below.

$$U_{PT}(a_n) = \frac{\left[w_m + (1 - w_m) \left(\tanh\left(\frac{a_n}{a_0}\right) + 1 \right) \right]}{2} \left(\frac{a_n}{(1 + a_n)^{\frac{\gamma-1}{2}}} \right) \quad (6-1)$$

where $\gamma > 0$ and w_m are parameters to be estimated, a_n is the acceleration of vehicle n , and a_0 is used to normalize the acceleration. Users obtain the utility value defined in Equation (6-1) if they opt acceleration, a_n and do not get involved in a crash. To consider the probability of a

crash in the model, total utility of a given acceleration is estimated by considering the disutility of a crash occurrence with a given crash probability.

$$U(a_n) = (1 - p_{n,i})U_{PT}(a_n) - p_{n,i}w_c k(v, \Delta v) \quad (6-2)$$

where $p_{n,i}$, w_c , and $k(v, \Delta v)$ are the crash probability, crash weighting factor, and crash seriousness term, respectively. Finally, the logistic functional form, introduced by Hamdar et al. (124), is used to incorporate the stochastic response of drivers as below.

$$f(a_n) = \begin{cases} \frac{e^{\beta_{PT}U(a_n)}}{\int_{a_{min}}^{a_{max}} e^{\beta_{PT}U(a')} da'}, & a_{min} < a_n < a_{max} \\ 0, & otherwise \end{cases} \quad (6-3)$$

where β_{PT} denotes the sensitivity of a choice to the utility, which has higher values for experienced drivers.

6-2-2- Vehicles with Active Vehicle-To-Vehicle Communications

This chapter employs the IDM to replicate the acceleration behavior of CVs. As these vehicles have information of surroundings, including traffic conditions, crashes, and weather conditions, deterministic models, such as IDM, can properly model the car-following behavior of CV drivers. IDM identifies acceleration of the following vehicle based on its current speed (v_n), the ratio of current spacing (s_n) to the desired spacing (s^*), and its relative speed to the leading vehicle (Δv_n).

$$a_{IDM}^n(s_n, v_n, \Delta v_n) = a_n \left[1 - \left(\frac{v_n}{v_0^n} \right)^{\delta_n} - \left(\frac{s^*(v_n, \Delta v_n)}{s_n} \right)^2 \right] \quad (6-4-a)$$

$$s^*(v_n, \Delta v_n) = s_0^n + T_n v_n + \frac{v_n \Delta v_n}{2\sqrt{a_n b_n}} \quad (6-4-b)$$

where a_{IDM}^n is the acceleration of the following vehicle. In addition, v_0^n , T_n , s_0^n , a_n , b_n , and δ_n are parameters to be calibrated.

6-2-3- Autonomous Vehicles

Given that AVs are fully aware of their surrounding environment and are certain about other drivers' behaviors, deterministic acceleration models can properly work for these vehicles. The current study adopts the acceleration model presented by Talebpour et al. (14) to model the behavior of AVs. This model is superior to similar models in that it considers sensor characteristics in the modeling process and creating data for the acceleration model. The main assumption of the model presented by Talebpour et al. (14) is that the speed of an AV should be low enough to fully stop when there is a vehicle right out of its detection zone, which cannot be spotted by the sensors. Hence, the model assumes the maximum possible deceleration for the AV and its leader, and defines the maximum safe speed, v_{max} , as below.

$$\Delta x_n = (x_{n-1} - x_n - l_{n-1}) + v_n \tau + \frac{v_{n-1}^2}{2a_{n-1}^{decc}} \quad (6-5)$$

$$\Delta x = \min\{\text{Sensor Detection Range}, \Delta x_n\} \quad (6-6)$$

$$v_{max} = \sqrt{-2a_n^{decc} \Delta x} \quad (6-7)$$

where n and $n - 1$ are subscripts for the AV and its leader, respectively. x_n , l_n , v_n , and τ are the location, length, speed, and reaction time of vehicle n . a_{n-1}^{decc} is the maximum deceleration of vehicle $n - 1$. In addition to the safety constraint, Talebpour et al. (14) adopted the acceleration model of Van Arem et al. (63) to model the movement of AVs.

$$a_n^d = k_a a_{n-1}(t - \tau) + k_v (v_{n-1}(t - \tau) - v_n(t - \tau)) + k_d (s_n(t - \tau) - s_{ref}) \quad (6-8)$$

where a_n^d and s_n are the acceleration and spacing of the AV, n , and k_a , k_v , k_d are model parameters. s_{ref} is the maximum of three values; the minimum distance, which is assumed to be 2.0 meters in this study, the safe distance (s_{safe}), and the following distance based on the reaction time (s_{system}). s_{safe} and s_{system} are defined as below.

$$s_{safe} = \frac{v_{n-1}^2}{2} \left(\frac{1}{a_n^{decc}} - \frac{1}{a_{n-1}^{decc}} \right) \quad (6-9)$$

$$s_{system} = v_n \tau \quad (6-10)$$

Finally, one can estimate the acceleration of the AV by finding the minimum of a_n^d and $k(v_{max} - v_n(t))$ where k is a model parameter.

6-3- Macroscopic Relations at the Facility Level

In order to obtain the equilibrium relation between velocity, v , and spatial gap (spacing), s , all vehicles should acquire the same ($\Delta v = 0$) and constant ($\dot{v} = 0$) speed over each time interval. In addition, there is no variability in the behavior of each specific driver at the equilibrium state (126). Given the equilibrium conditions and the acceleration models presented in section 6-2, the spacing of each vehicle type (i.e., HDV, CV, AV) with its leading vehicle is estimated as below.

$$s_{RV}(v) = s_0 + \sqrt{2} \alpha v \tau \sqrt{\ln \left(\frac{\tau}{v} \right) + \ln \left(\frac{w_c}{\sqrt{2\pi} \alpha} \right)} \quad (6-11)$$

$$s_{CV}(v) = \frac{(s_0 + T_n v)}{\sqrt{1 - \left(\frac{v}{v_0} \right)^{\delta_n}}} \quad (6-12)$$

$$s_{AV}(v) = v \tau \quad (6-13)$$

where s_{RV} , s_{CV} , and s_{AV} are the spacing values of HDVs, CVs, and AVs, respectively. In addition, τ , w_c , α , and s_0 are parameters to be estimated. To derive Equation (6-11), the gradient of crash disutility with zero acceleration is obtained from the study by Hamdar et al. (12). Using the relation between spacing and density, the congested part of the macroscopic fundamental diagram can be obtained by the following relations.

$$k = \frac{1}{\sum_i \sum_j p_{ij} (l_{veh} + s_{ij}(v))} \quad (6-14)$$

$$q = kv \quad (6-15)$$

where k is the density, l_{veh} is the length of vehicle, and s_{ij} is the spatial gap of the vehicle type i (i.e., HDV, CV, AV) with driver class j with its leading vehicle. As AVs are driverless, j is considered 1 for this vehicle type. In addition, p_{ij} denotes the proportion of vehicles with type i and driver class j on the segment. q and v are also flow and speed values at the segment. Note that the speed of the uncongested part of the fundamental diagram is bounded by free-flow speed. Therefore, the maximum flow, which occurs when all vehicles are moving with the free-flow speed, v_f , is calculated by the following equation.

$$q_{max} = \frac{v_f}{\sum_i \sum_j p_{ij} (l_{veh} + s_{ij}(v_f))} \quad (6-16)$$

In this study, we calibrated the acceleration model of different vehicle types using Next Generation Simulation (NGSIM) data. The data includes trajectories of almost 6,000 vehicles collected as a part of the NGSIM project for Federal Highway Administration (FHWA) (122). In this chapter, we differentiate CVs and HDVs based on their different microscopic models. Therefore, the NGSIM data is used to calibrate the model presented in Equation (6-1) to Equation (6-3) for HDVs. In addition, this data is used to calibrate the IDM model presented in Equation (6-4) for CVs. The dataset is used to provide a realistic representation of vehicular movements along a highway segment and calibrate a set of parameters for HDVs and CVs. CV and HDV models were calibrated for a portion of NGSIM data and the parameters were clustered into different sets (driver class) for each vehicle type (i.e. HDV and CV). Therefore, each set of parameters represents the behavior of drivers who have the same car-following behavior (fall into one cluster). The acceleration model of HDVs is calibrated for the total of 10 HDV drivers. In addition, the

acceleration model of CVs is calibrated for 6 drivers. Each set of parameters shows the car-following behavior of one driver, which results in keeping a specific amount of spacing with the leading vehicle at the equilibrium conditions. A genetic algorithm calibration approach (127) is adopted to calibrate parameters of the models.

With an assumption that all vehicles on a segment are from the same vehicle type and are driven by one specific driver class (if there is any), the fundamental diagrams for a segment with different driver behaviors, calibrated for this study, are demonstrated in Figure 6-1. However, in reality, each vehicle type consists of different driver classes. In numerical examples of this chapter, it is assumed that driver classes, with fundamental diagrams shown in Figure 6-1, are uniformly distributed for each vehicle type. In this figure, following assumptions are considered: free-flow speed as 70 mi/hr, vehicle length as 20ft, and reaction time of AVs as one second.

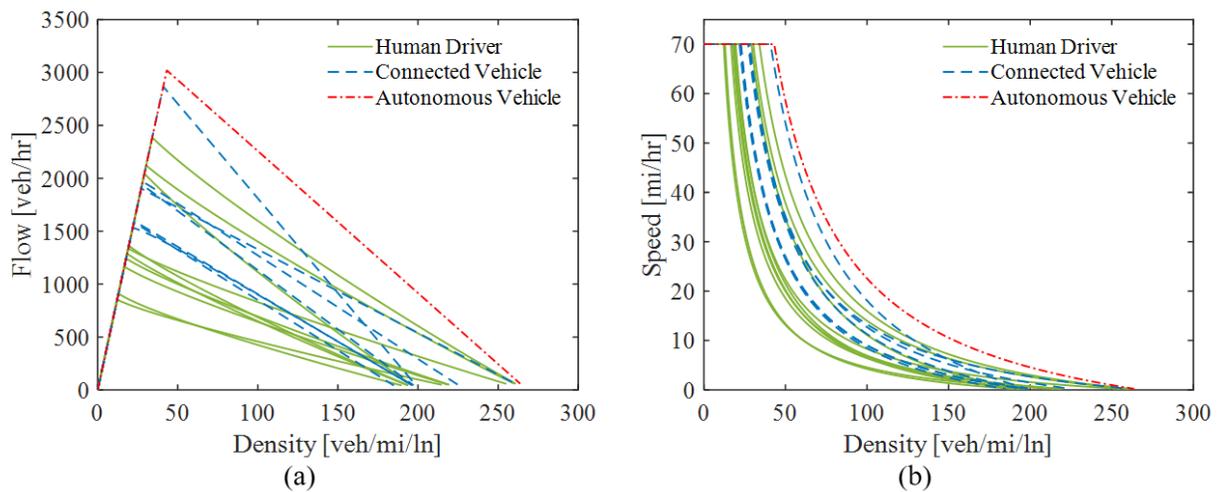


Figure 6-1- a) Flow-density and b) speed-density calibrated relationships at the equilibrium state for a segment occupied only with certain driver classes of HDVs, CVs, or AVs

6-4- Mesoscopic Simulation Framework

In this chapter, DYNASMART-P is used as the base mesoscopic simulation tool to simulate traffic dynamics at the network level. This tool considers en-route users, which is a specification of CVs and AVs. En-route users are aware of the current traffic conditions at different

regions of the network via real-time information. This information would influence drivers' route choice while traveling and may result in switching their paths. Given the unique features of this simulation tool, it is utilized as the base simulation tool for this study. In this research, this tool is updated to investigate the impacts of CVs with heterogeneous drivers and AVs on traffic flow and hysteresis phenomenon at the network level. To do so, realistic microsimulation frameworks for different vehicle types (i.e., HDV, CV, AV), presented in earlier sections, and the driving behavior of real vehicles are incorporated to develop adaptive macroscopic models at the facility level to be used in this mesoscopic tool. Note that the proposed approach is generic, and its application is not limited to this specific mesoscopic simulation tool. The process of updating DYNASMART-P is discussed as follows.

To perceive traffic flow dynamics in CV and AV environments from the micro-scale to the meso-scale, the spacing-density relationship presented in Equations 6-14 is used. For each link, m , in the network, the counts of different vehicle types such as i (e.g. HDV, CV, AV) with specific drivers such as j , $N_{ij}^{m,t}$ are estimated at each time interval t . Using the proportion of this count for each driver and vehicle type to the total number of drivers on the link, $N^{m,t}$, a nonlinear equation is solved to find a link speed that results into an average spacing between different vehicles (this average spacing is associated with the spacing values of all vehicles on the link). For each vehicle type and driver class, a desired spacing can be estimated based on the link speed and Equation (6-11) to Equation (6-13). We incorporate the golden section method to solve this nonlinear equation (128). Figure 6-2 provides a brief illustration of different steps to update speed values, $v^{m,t}$, based on the spatial and temporal varying distributions of mixed vehicle types. In this figure, l^m is the length of the link, V_u and V_d are the maximum and minimum speeds from which the search to find the speed of the link starts, and V_1/V_2 , K_1/K_2 , and \bar{s}_1/\bar{s}_2 are the intermediary speed, density, and

average spacing values that are used to reach the speed associated with the current density of vehicles at each link. In addition, ϵ is the stopping criteria for the golden section algorithm, demonstrating the accuracy of the algorithm.

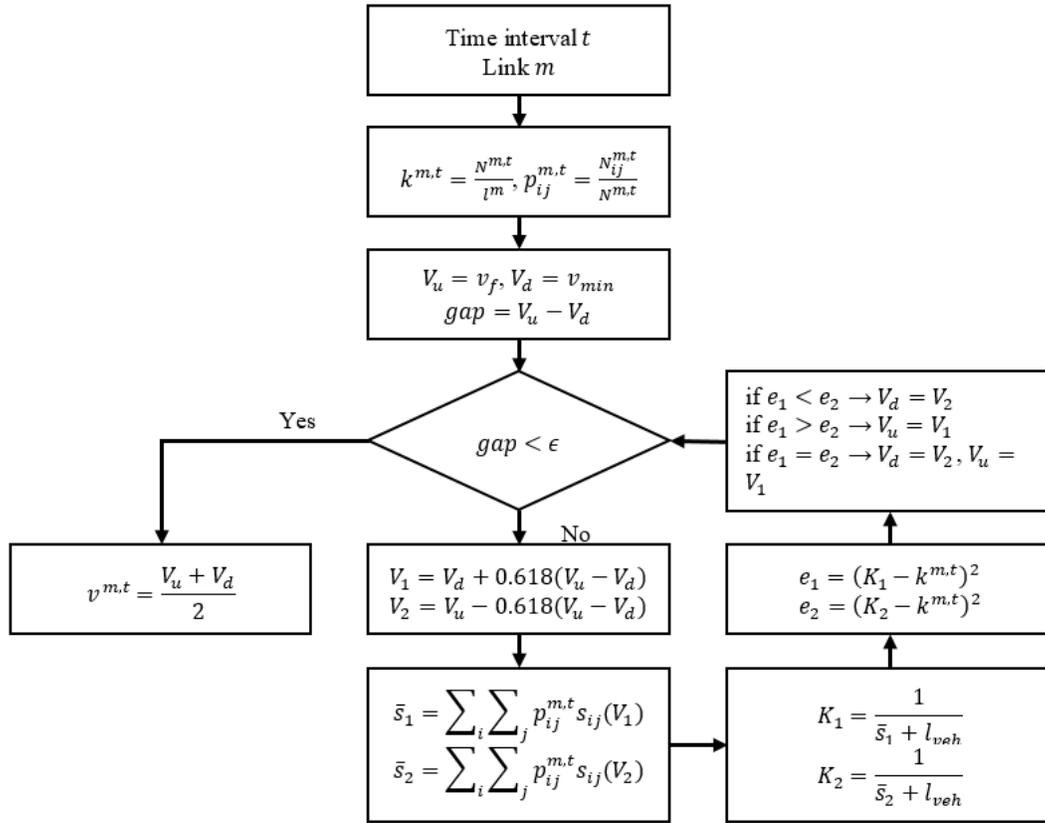


Figure 6-2- Illustration of the proposed approach to update link speeds given a specific density distribution of different vehicle types with heterogeneous drivers

As mentioned earlier, the dataset used in this chapter represents vehicular movements along a highway segment. Hence, the calibrated models are only used for freeway links. To account for the presence of a shared road network with heterogeneous drivers on all links of the network, including arterials, an adjustment factor is derived for each vehicle type in order to relate traffic models of CVs and AVs to HDVs. Note that there is no study on calibrating these models specifically for arterials to the authors' knowledge. Thus, this study assumes that CVs and AVs have the same impact on the traffic flow in freeways and arterials. Therefore, vehicles of each type

with different driver behaviors, calibrated by NGSIM data, are distributed uniformly across a segment with a unit length. The speed values relative to different density levels are then estimated for each vehicle type. The adjustment arterial factors are calculated for each density value as below.

$$f_i(k) = \frac{v_i(k)}{v_{RV}(k)} \quad (6-17)$$

where $v_i(k)$ and $v_{RV}(k)$ are the speed values when the density is equal to k and the segment is occupied by vehicles of type i (CV or AV) and HDVs, respectively. $f_i(k)$ is the adjustment factor, which is applied to the current fundamental diagram of arterials in DYNASMART, when CVs or AVs occupy a portion of arterial links.

In addition to the spacing-speed relationship, CVs and AVs influence the movement capacity at intersections. The number of vehicles, N , moving from link i to link j from a signalized intersection is bounded to three values.

$$N_{i,j} = \min\{\kappa_j, O_{i,j}, \zeta_i\} \quad (6-18)$$

where κ_j is the remained capacity of link j . $O_{i,j}$ represents the number of vehicles that are ready to move into link j from link i , and ζ_i is the maximum number of vehicles that can exit link i during simulation interval, Δt , according to its assigned green time, which is estimated as below.

$$\zeta_i = \Theta_i \left(\frac{g}{C} \right) \Delta t \quad (6-19)$$

where g is the green time phase for the movement of interest. In addition, Θ_i and C are saturation flow rate and cycle length, respectively. Given that different drivers and vehicle types do not have the same headways, the saturation flow rate at each intersection varies depending on the proportions of each vehicle type and driver class that pass the intersection. Therefore, the current study modifies the mesoscopic simulation tool to account for variable maximum flow rates

at intersections depending on the share of passing vehicles and their driver classes. Hence, an adjustment factor is proposed to increase or decrease the original capacity of movement based on the type of vehicles that cross the intersection and their drivers' acceleration behavior.

$$f_{int}^{ij} = \frac{q_{\max}^{ij}}{q_{\max}^b} \quad (6-20)$$

In this equation, q_{\max}^{ij} denotes the maximum flow rate, if all vehicles are type i with drivers of class j , which is the peak values shown in Figure 6-1a. q_{\max}^b is the maximum flow rate of the base case, which is assumed to be the average of HDV drivers. Finally, f_{int}^{ij} is the adjustment factor for the capacity of intersections. In the updated version of DYNASMART-P for the purpose of this study, vehicle type proportions are not known priory and intersection capacity cannot be modified. Thus, the remaining capacity of the intersection at any given time is adjusted based on the calculated factor associated with each vehicle, while it crosses the intersection.

6-4-1- Network Fundamental Diagram

The NFD, also known as macroscopic fundamental diagram (MFD), represents the aggregated traffic flow-density relationship at the network level. Simulation-based and empirical studies in the literature confirm the existence of a consistent and well-defined NFD (129–132). A relationship between the network-wide weighted average of traffic flow and density is used to define the NFD. These variables are calculated as the space-mean weighted averages of the link flows and densities, with link weights equal to the product of link length and number of lanes (130,133).

$$Q = \frac{\sum_{m=1}^M n_m l_m q_m}{\sum_{m=1}^M n_m l_m} \quad (6-21)$$

$$K = \frac{\sum_{m=1}^M n_m l_m k_m}{\sum_{m=1}^M n_m l_m} \quad (6-22)$$

Here, Q and K are the weighted average flow and density values respectively, M is the number of links in the network, l is the length of the link, and n is the number of lanes in each link. These equations yield ordered pairs of macroscopic flow and density at 5-minute time intervals that cumulatively form the simulation horizon. These values are plotted to represent NFD in this study.

The heterogeneous spatiotemporal distribution of congestion across real networks often creates hysteresis in the NFD (134,135). Hysteresis is a loop (complete or incomplete) in the NFD diagram that shows the degree to which the system is unstable during the unloading period. In large-scale networks with high levels of congestion, gridlock is formed (121), and the system cannot efficiently recover itself during the unloading phase, which causes the formation of a hysteresis loop in the NFD graph. Thus, the area of the hysteresis loop (Equation 6-23) is a good representative of the system instability.

$$A_{Hys} = \oint_L Q(k)dk \quad (6-23)$$

where, A_{Hys} is the area of a hysteresis loop in NFD, L is the hysteresis closed curve in NFD, and $Q(k)$ is the average network flow as a function of the average network density.

6-5- Numerical Experiments

In this section, the presented simulation frameworks in the earlier sections are applied to the Chicago downtown network, as a large-scale network, to investigate the potential impacts of a mixed traffic condition on traffic flow and travel time reliability at the network level. The Chicago downtown network, shown in Figure 3-1, is used to apply the traffic simulation tool. The simulation horizon is the AM peak period between 5:00 AM and 10:00 AM. The static hourly demand, provided by the Chicago Metropolitan Agency for Planning (CMAP), is transformed into

a time-dependent OD demand table with an OD estimation technique (86). The number of vehicles, simulated in the network during the AM peak period, is about 760,000. The demand profile is illustrated in Figure 6-3. Different MPRs of CVs and AVs are considered and the network is simulated with spatial and temporal varying distributions of different vehicle types with heterogeneous drivers over the network links. This study assumes that different driver classes, presented in Figure 6-1, are assigned uniformly among their vehicle types. While the vehicle trajectories in the network during the simulation horizon are available, the NFDs are plotted and the values associated with the area of hysteresis loops are estimated to analyze the impacts of connected and automated vehicle technology on traffic flow at the network level.

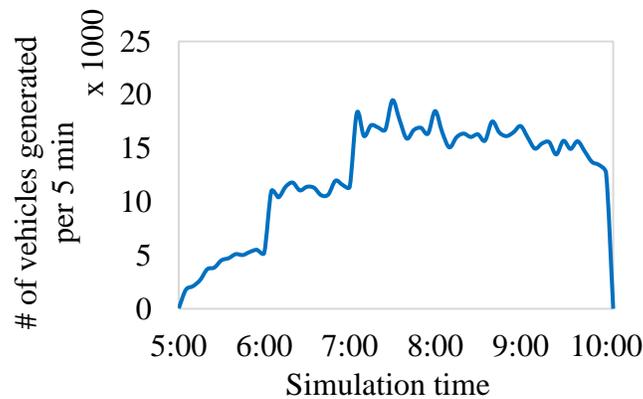


Figure 6-3- AM peak simulation demand for the Chicago downtown network

DYNASMART-P, which is used as the mesoscopic simulation tool of this study with modifications to incorporate a shared road network with heterogeneous drivers, is capable of considering en-route users. As it is mentioned earlier, en-route users can switch paths in the middle of their route due to congestion or gridlock. In the current study, all CVs and AVs are assumed to have this capability. In addition, to account for a realistic driving environment, 25% of the current vehicles on the road network, which are HDVs, are assumed to be en-route. Note that vehicles can have access to route information through GPS data even if they are not connected. In addition, it

should be mentioned that all AVs are assumed to have access to connected vehicle technology. In this study, the average vehicle length is assumed to be 20 ft. In addition, the reaction time of AVs is considered as one second.

Figure 6-4 illustrates the NFD graphs (Figure 6-4a to Figure 6-4d), and values of maximum density, and hysteresis loop area (Figure 6-4e) for different proportions of HDVs, CVs, and AVs in the network. As can be seen in Figure 6-4c and Figure 6-4d, the network cannot recover from congestion at high proportions of HDVs (100% HDV and 75% HDV). However, the congestion is resolved by replacing a portion of traffic stream with CVs or AVs, or both (Figure 6-4a and Figure 6-4b). Therefore, connectivity and automation of vehicles facilitate the network recovery from congestion. In addition, the network faces a reduction in the maximum density and the area of hysteresis loop by an increase in MPR of CVs or AVs, or both (Figure 6-4a-e). Therefore, a higher stability in the recovery phase is observed for the scenarios with higher CV or AV proportions. Based on the results of Figure 6-4e, replacing only 25% of the traffic demand, which contains only HDVs, by CVs and AVs decreases the area of hysteresis loop by 31% and 37%, respectively. Increasing this value to 50% MPR of CVs and AVs shrinks the hysteresis loop by 52% and 63%, respectively. This significant reduction in hysteresis loop area denotes that the gridlock dissipation and system recovery are accomplished faster for higher MPRs of CVs and AVs. In addition to the reduction of maximum density in higher MPRs of CVs and AVs, flow rates slightly increase at the same density values as the share of these vehicle types increases in the network. However, the variation of maximum observed throughput is not significant when we compare 100% CV scenario with 100% AV scenario (Figure 6-4d). This difference is small since in 100% AV scenario, network congestion recovers before reaching to its nominal capacity. The results of Figure 6-4 also show that the injection of AVs is more influential on traffic flow than

CVs, especially when the ratio of HDVs in the network is low. Figure 6-4b demonstrates a scenario with a mixed traffic consisting of all three vehicle types. As can be seen in this figure, the scenario with equal penetration rates of CVs and AVs (25%) falls between the two scenarios of 50% CV and 50% AV. This result highlights the superiority of CVs to the HDVs and AVs to the CVs in mitigating the traffic congestions in the network with a mixed traffic stream. Furthermore, it shows the applicability of the proposed methodology to consider a traffic mix including all three types of vehicles.

Figure 6-5 illustrates the impacts of CVs and AVs on traffic flow considering different aspects, including the access to real-time traffic information (being an en-route user), adaptive fundamental diagrams based on different vehicle types with heterogeneous drivers for each freeway link, the adjustment factor for traffic flow models on arterial links, and intersection movement capacity variations in the presence of CVs and AVs. Figure 6-6 demonstrates the maximum density value and the area of the hysteresis loop, as the measures of stability, for the same scenarios as Figure 6-5. Figure 6-5a compares the scenarios with 100% HDV, 100% CV, and 100% AV considering only the impacts of adaptive drivers. Note that only 25% of HDVs are assumed to be en-route, while all CVs and AVs receive the real-time information of routes. As can be seen in Figure 6-5a and Figure 6-6, having access to the en-route information significantly reduces the maximum density and the area of hysteresis loop, which means much faster recovery of network with the presence of CVs and AVs than only HDVs. Note that in this case, since the traffic flow models are not updated relative to HDVs, 100% CV and 100% AV scenarios have the same results.

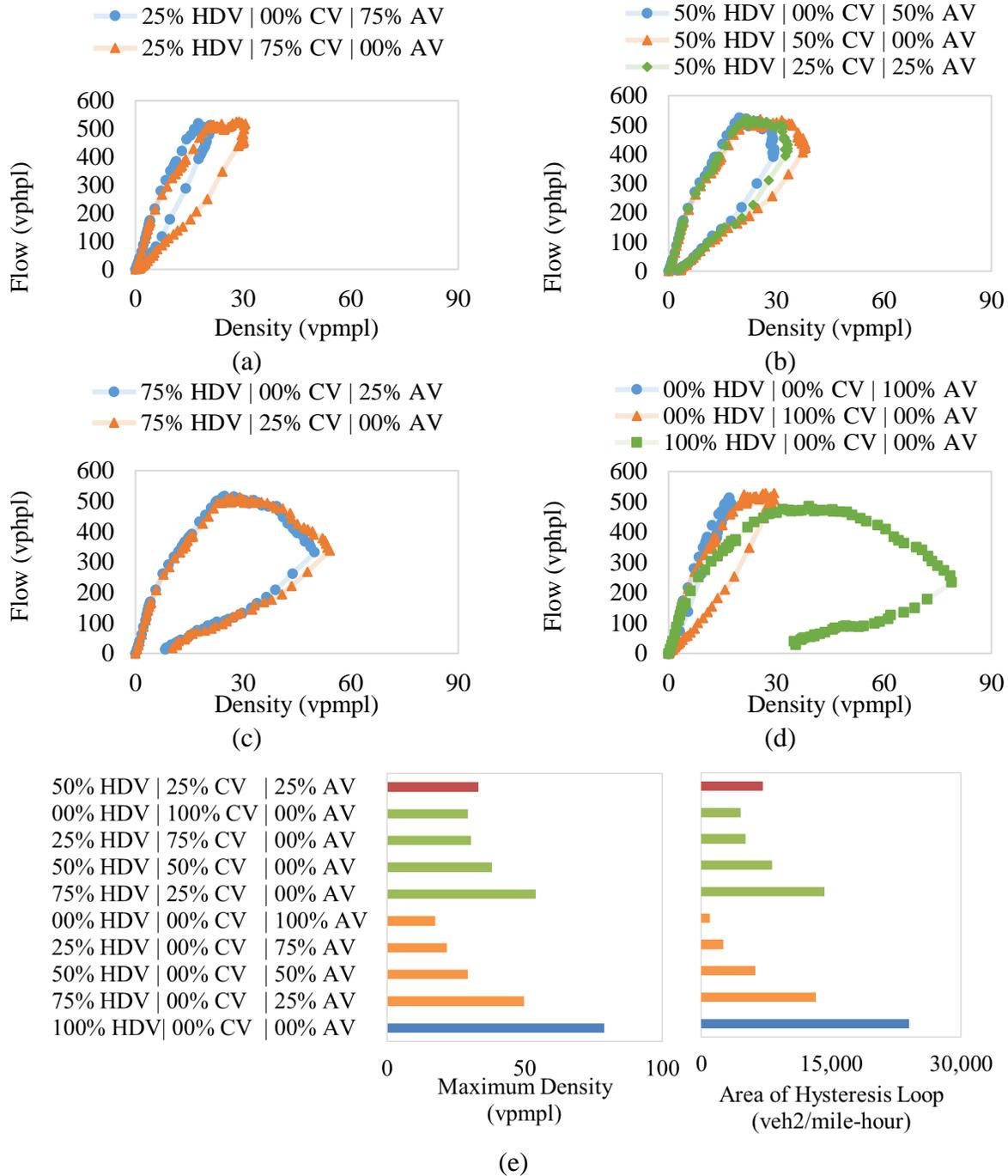


Figure 6-4- (a-d) NFD and (e) maximum density and area of hysteresis loop for different MPRs of CVs, and AVs

Figure 6-5b represents the impacts of traffic flow models for freeway links in the presence of CVs and AVs with different driver classes in addition to the impacts of en-route users. Based on the results of Figure 6-5b and Figure 6-6, freeway traffic flow models slightly improve the

performance of the network in congestion. This is due to the fact that freeway links are only a small proportion of the entire network, with only 11 percent of the total lane length of the network. The results of Figure 6-5c, which includes the impact of adjustment factors for arterial traffic flow models in addition to the ones considered in Figure 6-5b, show a rather same improvement as freeway traffic flow models. As the acceleration behavior of drivers is heterogeneous and a portion of HDV drivers have even a smaller spacing with the leading vehicle than CV drivers, these small improvements in traffic congestion due to the presence of different vehicle types on freeway and arterial links are justifiable. Finally, considering the effect of intersection movement capacity variations in the presence of CVs and AVs (Figure 6-5d) shows a significant enhancement in the network dynamics. This significant improvement highlights the importance of incorporating the impacts of a mixed traffic stream on intersection capacity variations.

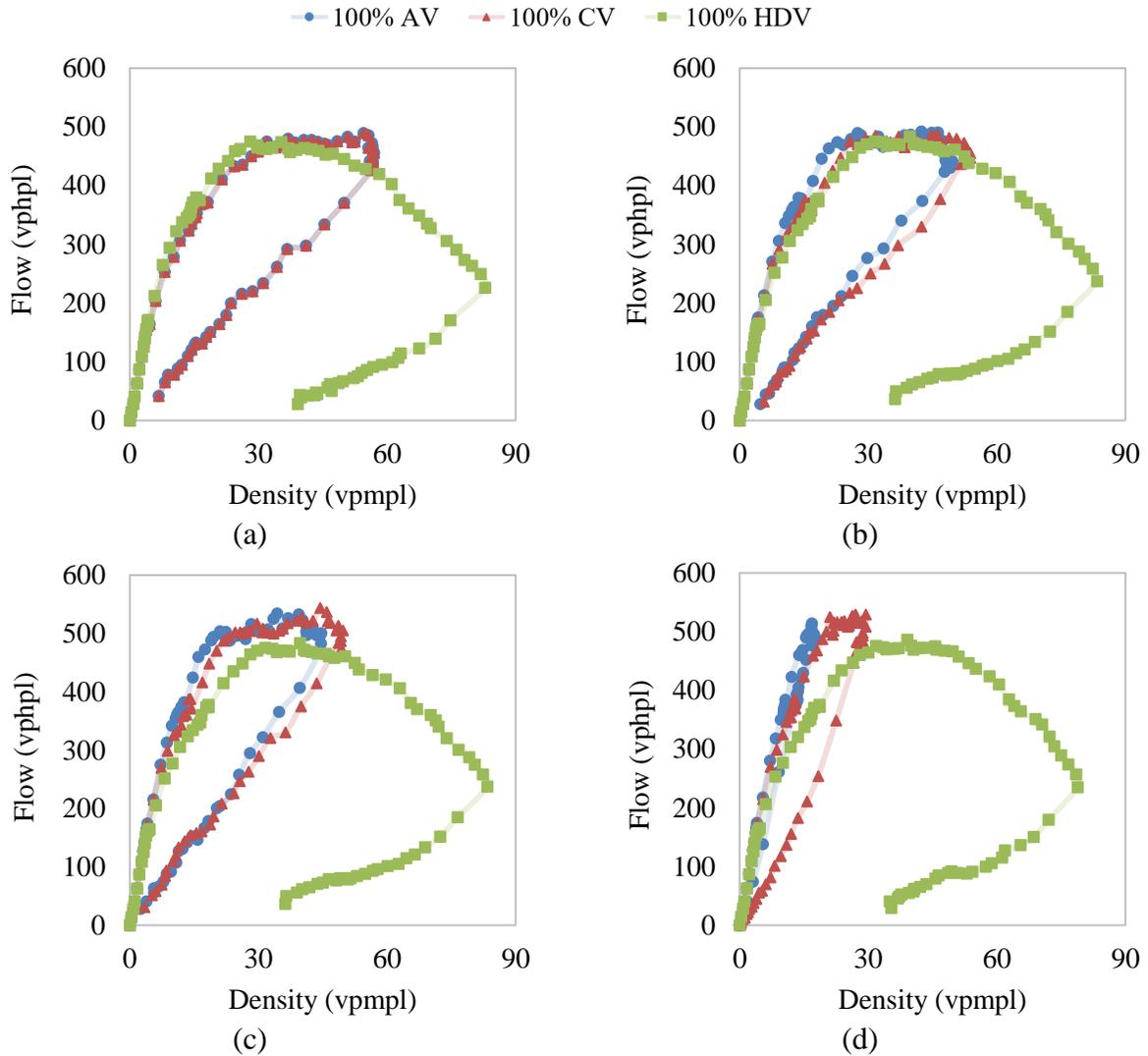


Figure 6-5- NFD results considering the impact of (a) only en-route users, (b) en-route users and adaptive traffic flow models on freeway links, (c) en-route users and freeway and arterial traffic flow models, and (d) en-route users, freeway and arterial traffic flow models, and adjusted intersection capacity

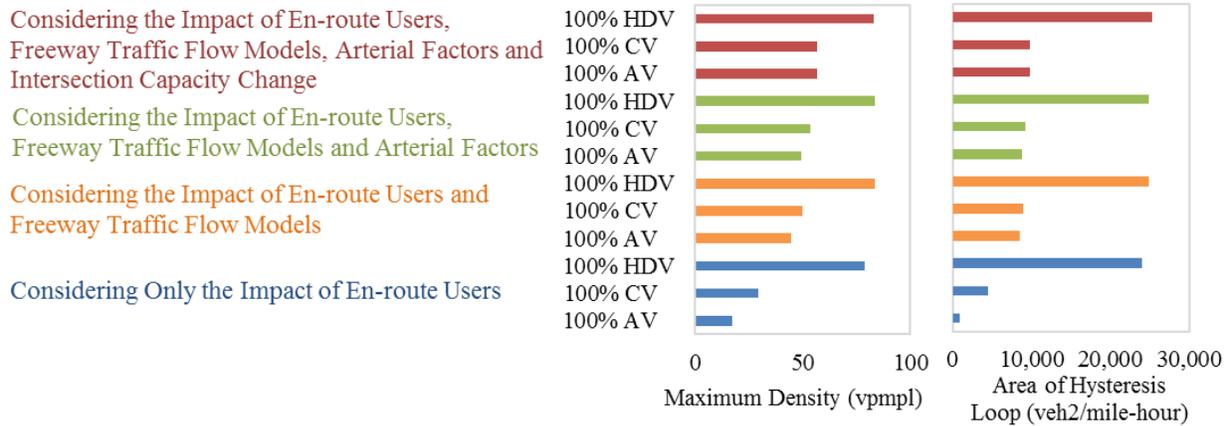


Figure 6-6- Values of maximum density and area of hysteresis loop considering the impacts of different factors on network dynamics

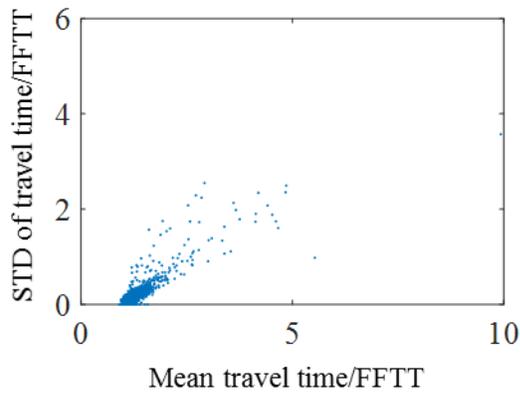
As mentioned throughout the manuscript, travel time reliability is an important measure to effectively evaluate the performance of road networks. In addition, emerging technologies, including connected and autonomous vehicle technologies, are expected to have promising benefits for transportation networks by enhancing efficiency, throughput, and reliability. There are different reliability measures that can be used to explore the impacts of connected and autonomous vehicles on travel time reliability. The mean and standard deviation of a travel time distribution is one of these measures. This measure should be normalized by distance or free flow travel time to control the impacts of trip distance and speed variations. Therefore, in this study, the mean and standard deviation of travel time for each link are estimated per free flow travel time. This study uses the simulation tool, developed and presented earlier in this chapter, to generate vehicle trajectories in a network to explore the relationship between standard deviation of travel time and mean travel time for a fully AV and a fully HDV networks. This approach extracts the travel time and travel distance for each set of consecutive links that are traveled by a certain vehicle. For link travel time distributions, 400 minutes of the simulation (5:00 AM to 10:00 AM plus unloading period) are divided in 100-minute time intervals to create 4 distributions for each link. Simulation

results are aggregated in one-minute observations, providing each link travel time distribution with $100 \times 86 = 8600$ observations.

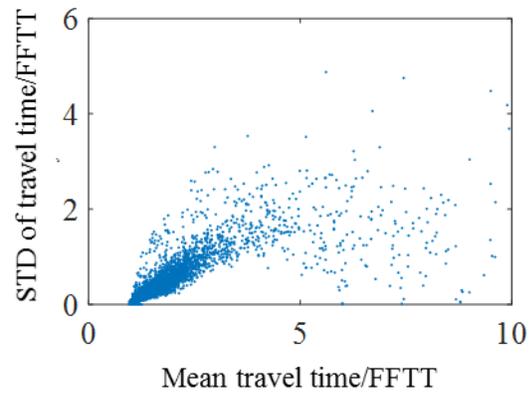
To generate link travel time variations, this study uses 86 scenarios defined by Zockaie et al. (15,87) based on real world observations. These scenarios are different in traffic demand, the percentage of adaptive drivers, weather conditions, and number of crashes. The scenarios are defined based on the data of the workdays in the first four months of 2010 for the Chicago downtown network. For different demand levels, the total throughput of all loop detectors across the network over AM peak period is found. Dividing this value for each scenario (i.e. one specific day) to an average of all scenarios results in a demand factor. The data of Automated Surface Observing System (ASOS) station at the Chicago Midway International Airport is also used to categorize different scenarios into clear weather, rainy day, and snowy day. Note that there is a weather adjustment factor in DYNASMART-P to consider weather conditions. In addition, Zockaie et al. (15) used the data provided by Illinois Department of Transportation to define crash locations, time, and the severity of crashes in terms of capacity drop in the location. Finally, a range between 40 to 50 percent en-route users are considered for HDVs in different scenarios. Note that this factor is constant for AVs as these vehicles have access to en-route information. Based on the above factors and real-world observations, 86 scenarios are simulated in DYNASMART-P for a fully AV and fully HDV networks. Using the simulation results, link travel time distributions are estimated and normalized standard deviation of travel time values are illustrated relative to the normalized mean travel time values in Figure 6-7 (fully HDV network) and Figure 6-8 (fully AV network). Each dot in these figures represents the normalized mean and standard deviation of travel time for one link in the Chicago downtown network that is simulated by 100% HDVs or 100% AVs. In addition, the results are categorized into different time intervals

and shown in Figure 6-7 and Figure 6-8. Figure 6-7a and Figure 6-8a illustrate the normalized standard deviation and mean travel times for the 1st 100-minute time intervals for the fully HDV and fully AV networks, respectively. Similarly, Figure 6-7b, Figure 6-7c, and Figure 6-7d plot the same performance measures for 2nd, 3rd, and 4th 100-minute time intervals.

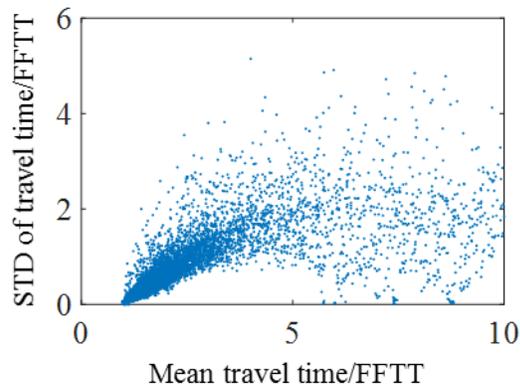
The comparison of standard deviation and mean travel times shows that expectedly, the mean travel times are generally lower in a fully AV network than a fully HDV network. As a result, the standard deviation of travel time is also lower in a fully AV network than a fully HDV one. In addition, the dispersion of normalized standard deviation values is generally low in a fully AV network. For example, in the fully HDV network, the results of last time interval reveal that the standard deviation of travel time is between 0.8 and 6 times of free flow travel time when the mean value is 5 times of free flow travel time. However, for the same ratio of mean to free flow travel time for a fully AV network, standard deviation values of different links are between 0.5 and 4 times of the mean travel time, which indicates that deploying AV technologies increases the reliability of network.



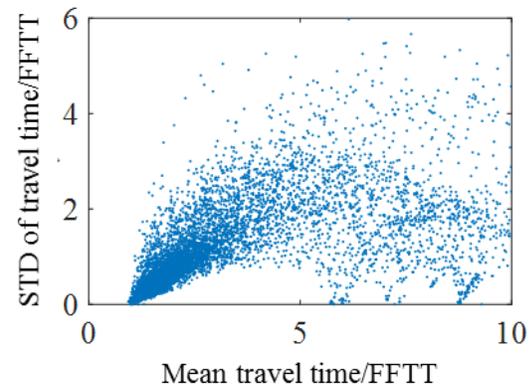
(a)



(b)



(c)



(d)

Figure 6-7- Standard deviation vs mean travel time normalized to free flow travel time for different links in a fully HDV network for the a) first, b) second, c) third, and d) fourth 100-minute time intervals

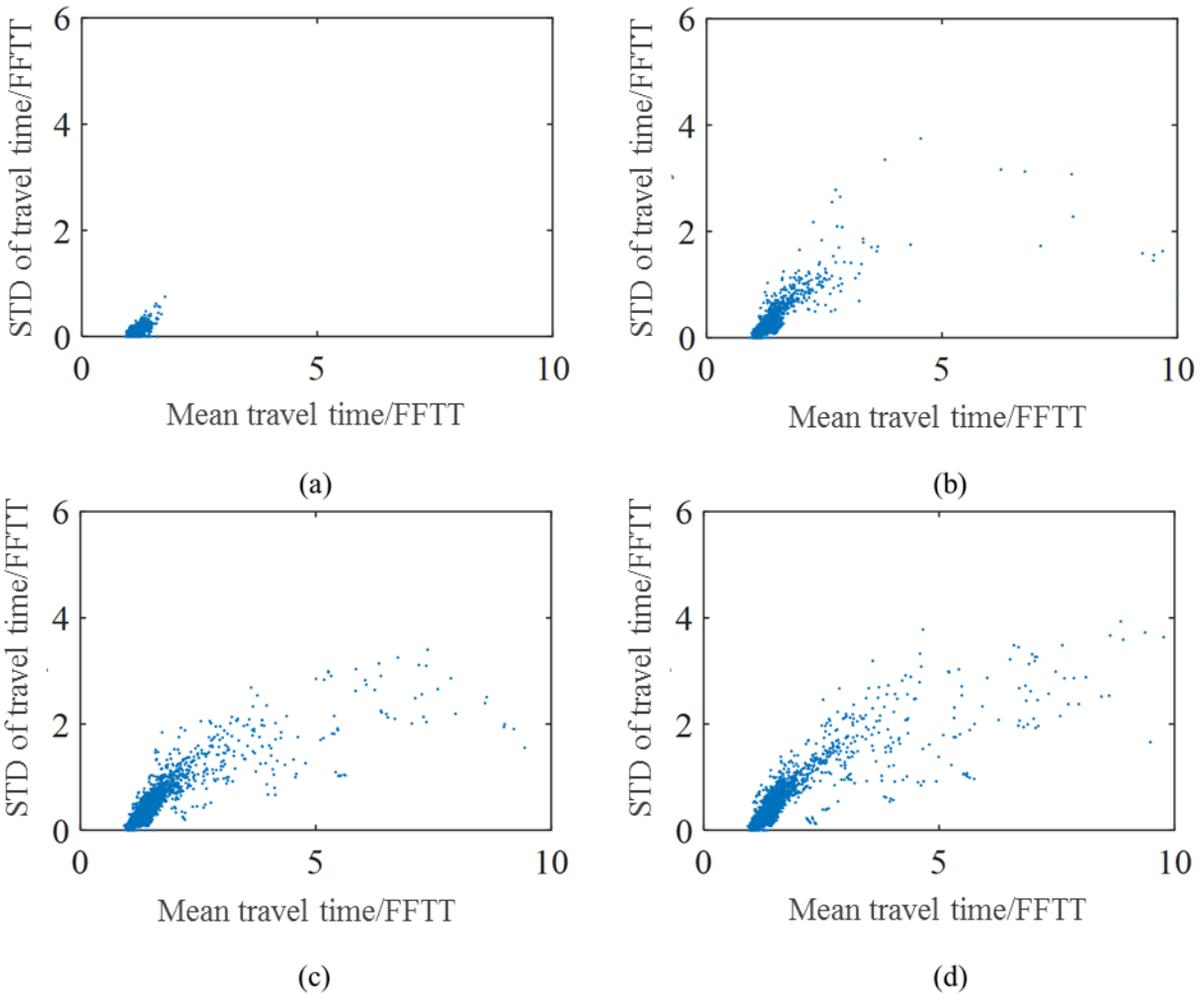


Figure 6-8- Standard deviation vs mean travel time normalized to free flow travel time for different links in a fully AV network for the a) first, b) second, c) third, and d) fourth 100-minute time intervals

6-6- Summary

This DTA simulation tool of DYNASMART-P was updated in this chapter to account for the presence of CVs and AVs at the network level by incorporating adaptive fundamental diagrams due to the spatial and temporal varying distribution of different vehicle types with heterogeneous drivers. The proposed approach is generic and can be applied to any mesoscopic simulation tool. DYNASMART-P is selected due to its availability to the authors, and its unique features to consider en-route users and signalized intersections in a mesoscopic simulation tool.

Microsimulation frameworks, presented in the literature, in addition to the trajectories of real vehicles are used to calibrate spacing-speed relationships for different vehicle types and heterogeneous drivers of HDVs and CVs. The relationships are then used in the mesoscopic simulation tool to update the speed of each link at any time interval in the presence of different distributions of vehicles and drivers. This chapter also considered the movement capacity variation at intersections due to a mixed fleet of CVs, AVs, and HDVs. In addition, adjustment factors were presented to modify the current fundamental diagram of arterials in DYNASMART-P, when CVs and AVs occupy a portion of arterial links. The proposed approach is also capable of considering a traffic mix of all three vehicle types (HDV, CV, and AV). The methodology of this study was applied to the Chicago downtown network. NFD and the area of hysteresis loop are used as representatives of the aggregated traffic flow-density relationship and the stability at the network level. In addition, standard deviation of travel time on different links normalized to the free flow travel time was used as a measure of travel time reliability. To generate travel time variation across the links, this study used 86 scenarios that are different in traffic demand, weather condition, traffic crashes, and number of adaptive drivers. The results confirmed the superiority of AVs and CVs in mitigating congestion and improving travel time reliability.

CHAPTER 7 - Conclusions and Future Work

7-1- Summary and Concluding Remarks

Travel time variability and reliability are primary determinants of travelers' route choice behavior and any developed demand management strategy such as congestion pricing. Despite the growing interests in the development of different strategies and technologies to improve mobility, there is still a gap in the current body of the literature to consider the reliability measures in developing equitable strategies and to use emerging technologies to improve the reliability of transportation networks. The computational burden is the main challenge of traffic simulation tools that consider travel time uncertainty and the heterogeneity of travelers. Therefore, the main goals of this study were threefold: first, improving computational time of optimal path finding algorithms in stochastic time-dependent networks with heterogeneous users; second, considering travel time reliability in finding equitable congestion pricing schemes; and third, presenting a traffic simulation tool that considers the presence of connected and autonomous vehicles in order to quantify the impacts of these vehicles on traffic congestion and travel time reliability at the network level.

In this study, two network contraction approaches are first developed to reduce the network size of each specific OD pair in stochastic time-dependent networks. The network contraction is based on the comparison of optimistic and pessimistic solutions resulting from minimum and maximum travel time realizations of an MCS approach. In this respect, the pessimistic and optimistic solutions for path travel times are first obtained conservatively by assigning minimum and maximum travel times to all network links in CHAPTER 3. This conservative assumption is then relaxed in CHAPTER 4 and two learning approaches are presented to utilize the information of the realizations in the initial iterations of the MCS approach. The learning approaches are

calibrated through the Chicago downtown network. The transferability of the approach to other networks is then checked using the network of Salt Lake City. The numerical results of the two large-scale applications lead to the following findings:

- The calibrated learning approaches for a particular large-scale network is used successfully for another large-scale application with a different structure of nodes and links. Although the errors in objective function values are larger in the Salt Lake City network relative to the calibration network, the level and amount of errors are still satisfactory.
- Significant reductions in the network size are observed using both learning approaches relative to the approach without any network contraction and the conservative approach.
- The learning approaches have an acceptable level and number of errors in terms of SPOTAR/MTTBP objective function estimations.
- The adaptive learning approach is superior to the fixed learning approach in terms of the accuracy of the objective function, while the fixed learning approach is more aggressive in the network contraction. Therefore, a trade-off between a certain desired accuracy level and the computational efficiency is needed to select the proper learning approach.

Overall, the learning approaches developed for the network contraction in the reliable path-finding problems are successfully applied to two large-scale applications with satisfactory approximations of the SPOTAR and the MTTBP objective functions and considerable reductions in networks sizes, using optimistic/pessimistic travel times as hints. As long as the network structure and link costs (distribution) are available, the methodology of this study for network contraction is applicable to any other MCS approach in different fields. However, more conservative, or less conservative approaches (different learning factors and multipliers) might be utilized depending on the application of interest. The networks used for calibrating and showing

the transferability of network contraction approach are both flat networks. In case of hierarchical networks, the presented algorithm can be applied to each layer (level) of the network with expected computational benefits due to the advantages of network contraction that is transferred from one layer to another. However, one might expect an error, in terms of finding the optimal path, in one layer to propagate to subsequent layers. Such error propagations across the layers call for setting parameters differently for each layer to certify accuracy of the solution algorithm.

A modeling framework and solution methodology are also presented in this study for finding revenue-neutral and Pareto-improving congestion pricing values considering travel time variability for heterogeneous users with different VOTs and VORs in a bi-modal network consisting of transit and personal cars. The presented framework extends the second-best pricing optimization problem by integrating an RBUE algorithm. The objective function of the pricing algorithm minimizes the total travel time of highway users given a revenue-neutral and Pareto-improving pricing set. Users' heterogeneity in response to the reliability measures, their response to different toll values and toll distribution strategies, and link travel time correlations are considered in the RBUE problem. A PSO algorithm is also developed to determine the optimum toll values considering different toll distribution strategies (e.g. credit-based, transit subsidy). The proposed approach is successfully applied to a modified Sioux Falls network to explore the impacts of subsidization strategy, congestion level, and considering travel time reliability on the pricing strategy and its effectiveness. The results show the significant importance of considering travel time reliability by comparing the optimum toll values, the changes in system cost, and the total travel time with and without considering travel time reliability.

Finally, the impacts of CV and AV technologies on traffic flow and travel time reliability were observed at the network level. For this purpose, the simulation-based DTA tool of

DYNASMART-P is updated to capture the collective effects of the interactions between different vehicle types (i.e., AVs, CVs, HDVs) on traffic flow dynamics. In order to translate traffic flow dynamics from a micro-scale to a meso-scale, the relationship between spacing and speed for each vehicle type is derived and used as an input to the mesoscopic model. The proportion of each vehicle type and driver class on each link is tracked in the traffic propagation process at each time interval. Using this proportion, a non-linear equation is solved to obtain the current speed of the link, satisfying the spacing values of all vehicles traversing it. The methodology of this study is applied to the Chicago downtown network. NFD and the area of hysteresis loop are used as representatives of the aggregated traffic flow-density relationship and the stability at the network level. The normalized values of standard deviation of travel time for different links in the network are also used as an indicator of travel time reliability. The numerical experiments of the large-scale application lead to the following findings:

- With a constant proportion of HDVs in the network, higher MPRs of CVs and AVs lead to a lower maximum density and smaller hysteresis loop area, meaning a faster network recovery from congestion. In addition, higher proportions of CVs and AVs relative to HDVs result in a lower maximum density, slightly higher maximum flow, and a more stable network.
- Being en-route is the most effective feature of CVs and AVs, which results in significant improvements in traffic flow and a much faster network recovery. In addition, the presence of CVs and AVs significantly improves the intersection capacity.
- As the acceleration behavior of drivers is heterogeneous and a portion of HDV drivers have even a smaller spacing with the leading vehicle than CV drivers, the impacts of the presence of CVs and AVs on freeways and arterial links are small compared to the impacts of real-

time route choice and intersection capacity. However, this effect is still significant and should be considered to differentiate the impacts of different vehicle types on traffic flow.

- A mixed traffic condition with CVs, AVs, and HDVs has a different impact on NFD and hysteresis loop area from other scenarios with only two vehicle types. The results also indicate that the injection of AVs in the network is more effective on traffic flow conditions than CVs. Therefore, it is necessary to consider different shares of CVs and AVs in addition to HDVs to investigate a realistic impact of these vehicles with heterogeneous drivers on traffic flow at the network level.
- AVs significantly improve the reliability of network relative to the scenario with only HDVs in the network. First, the standard deviation of link travel times generally decreases by replacing HDVs with AVs because of having much lower congestion and mean travel times in the network. In addition, the coefficient of variation is also decreased significantly by replacing HDVs with AVs.
- It can also be concluded from above observations that despite the undoubtable impacts of automation on mobility, route choice and an even congestion distribution over the network have much more significant impacts on traffic flow at the network level than other effective factors.

7-2- Future Research Directions

The proposed methodologies in this study to improve the computational time of optimal path finding algorithms benefit from a significant reduction in computational times and negligible impacts on the accuracy. However, there are still some limitations to consider for future studies. For instance, the network reduction logic adopted in this study can be extended to allow the fusion of some extra information/hints, in addition to the optimistic/pessimistic travel times, in the

learning process to determine nodes eligible for elimination, such as a link or sub-path that is known to be part of the optimal paths from a certain set of origins to a given destination. Furthermore, the learning approaches for network contraction could extend to incorporate a reinforcement learning methodology that adjusts the learning factor in the successive iterations based on feedbacks received from previous iterations.

Furthermore, the methodology of this study to consider travel time reliability in finding equitable pricing schemes provides important insights regarding the impacts of travel time variability and reliability on finding an efficient and equitable pricing strategy. However, in this study, users are not allowed to have multi-modal trips. In addition, all OD pairs are assumed to have alternative transit lines, which might not be the case in real-world transportation networks. Previous studies show that an alternative route to the tolled route is necessary in having an equitable pricing strategy. Therefore, it is suggested to update the algorithm of this study to consider multi-modal trips, consistent with the vision of integrated corridor management, for the future research. Also, a further step of this research is to consider travel time reliability in dynamic road pricing. Extending the research to a state-dependent congestion pricing, in which the toll values are updated based on the real-time monitoring of traffic conditions is another future research suggestion. Furthermore, using real-world data to define travel time uncertainty on highway links and transit lines would improve the fidelity of the obtained pricing schemes.

Finally, this study provides useful insights into the impacts of CVs and AVs on traffic flow and travel time reliability at the network level. However, the tool, presented in this study to simulate traffic consisting of CVs, AVs, and HDVs, still has some limitations that should be addressed for future research. For instance, the signal timings used in this study are not optimized for a mixed traffic stream, consisting of HDVs, CVs, and AVs. Having an adaptive signal timing

may significantly influence the traffic flow of the network. In addition, more data collection efforts are needed to have a more realistic estimation of the adjustment factors for arterial traffic flow models and intersection capacity in the presence of CVs and AVs. In this study, we also assumed that the drivers of connected vehicles are certain about other drivers' behavior due to the flow of information in a V2V/V2I communications network. Therefore, their behavior is more deterministic compared to human-driven vehicles. Utilizing a dataset containing the trajectories of connected vehicles to calibrate the traffic flow models could result in more realistic parameters for the acceleration models of connected vehicles. Finally, in this study, we investigated the impacts of CVs and AVs purely on traffic flow and travel time reliability at the network level. However, the secondary impacts of CVs and AVs on travel demand need to be considered as well.

REFERENCES

REFERENCES

1. Mahmassani HS, Kim J, Chen Y, Stogios Y, Brijmohan A, Vovsha P. Incorporating Reliability Performance Measures into Operations and Planning Modeling Tools. Transportation Research Board; 2014.
2. Ji Z, Kim YS, Chen A. Multi-objective α -reliable path finding in stochastic networks with correlated link costs: A simulation-based multi-objective genetic algorithm approach (SMOGA). *Expert Syst Appl*. 2011;38(3):1515–28.
3. Fan YY, Kalaba RE, Moore II JE. Shortest paths in stochastic networks with correlated link costs. *Comput Math with Appl*. 2005;49(9–10):1549–64.
4. Zockaie A, Mahmassani HS, Nie Y. Path finding in stochastic time varying networks with spatial and temporal correlations for heterogeneous travelers. *Transp Res Rec*. 2016;2567(1):105–13.
5. Lam TC, Small KA. The value of time and reliability: measurement from a value pricing experiment. *Transp Res Part E Logist Transp Rev*. 2001;37(2):231–51.
6. Gardner LM, Boyles SD, Waller ST. Quantifying the benefit of responsive pricing and travel information in the stochastic congestion pricing problem. *Transp Res Part A Policy Pract*. 2011;45(3):204–18.
7. Jiang L, Mahmassani H, Zhang K. Congestion pricing, heterogeneous users, and travel time reliability: multicriterion dynamic user equilibrium model and efficient implementation for large-scale networks. *Transp Res Rec J Transp Res Board*. 2011;(2254):58–67.
8. Liu Y, Li Y, Hu L. Departure time and route choices in bottleneck equilibrium under risk and ambiguity. *Transp Res Part B Methodol*. 2018;117:774–93.
9. Di X, Liu HX, Ban XJ. Second best toll pricing within the framework of bounded rationality. *Transp Res Part B Methodol*. 2016;83:74–90.
10. Zockaie A, Nie Y, Mahmassani H. Simulation-based method for finding minimum travel time budget paths in stochastic networks with correlated link times. *Transp Res Rec J Transp Res Board*. 2014;(2467):140–8.
11. Zockaie A, Nie Y, Wu X, Mahmassani H. Impacts of correlations on reliable shortest path finding: a simulation-based study. *Transp Res Rec J Transp Res Board*. 2013;(2334):1–9.
12. Hamdar SH, Mahmassani HS, Treiber M. From behavioral psychology to acceleration modeling: Calibration, validation, and exploration of drivers cognitive and safety parameters in a risk-taking environment. *Transp Res Part B Methodol*. 2015;78:32–53.
13. Kesting A, Treiber M, Helbing D. Enhanced intelligent driver model to access the impact of driving strategies on traffic capacity. *Philos Trans R Soc A Math Phys Eng Sci*.

- 2010;368(1928):4585–605.
14. Talebpour A, Mahmassani HS. Influence of connected and autonomous vehicles on traffic flow stability and throughput. *Transp Res Part C Emerg Technol.* 2016;71:143–63.
 15. Zockaie A, Mahmassani HS, Kim J. Network-wide Time-dependent Link Travel Time Distributions with Temporal and Spatial Correlations. In: *Transportation Research Board 95th Annual Meeting* Transportation Research Board. 2016.
 16. Tirachini A, Hensher DA, Bliemer MCJ. Accounting for travel time variability in the optimal pricing of cars and buses. *Transportation (Amst).* 2014;41(5):947–71.
 17. Hellinga B, van Lint H, Hofman F. The use of exogenously defined standard deviation versus mean travel time relationships for estimating the impact of policy measures on reliability. 2012.
 18. Mahmassani HS, Hou T, Dong J. Characterizing travel time variability in vehicular traffic networks: deriving a robust relation for reliability analysis. *Transp Res Rec.* 2012;2315(1):141–52.
 19. Peer S, Koopmans CC, Verhoef ET. Prediction of travel time variability for cost-benefit analysis. *Transp Res Part A Policy Pract.* 2012;46(1):79–90.
 20. Tu H. Monitoring travel time reliability on freeways. Netherlands TRAIL Research School; 2008.
 21. Eliasson J. Forecasting travel time variability. In: *European Transport Conference.* 2006.
 22. Asakura Y. Reliability measures of an origin and destination pair in a deteriorated road network with variable flows. In: *Transportation Networks: Recent Methodological Advances Selected Proceedings of the 4th EURO Transportation Meeting* Association of European Operational Research Societies. 1999.
 23. Shao H, Lam WHK, Meng Q, Tam ML. Demand-driven traffic assignment problem based on travel time reliability. *Transp Res Rec.* 2006;1985(1):220–30.
 24. Xing T, Zhou X. Finding the most reliable path with and without link travel time correlation: A Lagrangian substitution based approach. *Transp Res Part B Methodol.* 2011;45(10):1660–79.
 25. Boyles SD, Waller ST. A mean-variance model for the minimum cost flow problem with stochastic arc costs. *Networks.* 2010;56(3):215–27.
 26. Khani A, Boyles SD. An exact algorithm for the mean--standard deviation shortest path problem. *Transp Res Part B Methodol.* 2015;81:252–66.
 27. Nie YM, Wu X, Dillenburg JF, Nelson PC. Reliable route guidance: A case study from Chicago. *Transp Res Part A Policy Pract.* 2012;46(2):403–19.

28. Nie YM, Wu X. Shortest path problem considering on-time arrival probability. *Transp Res Part B Methodol.* 2009;43(6):597–613.
29. Levy H. Stochastic dominance and expected utility: survey and analysis. *Manage Sci.* 1992;38(4):555–93.
30. Fan YY, Kalaba RE, Moore JE. Arriving on time. *J Optim Theory Appl.* 2005;127(3):497–513.
31. Wu X, Nie Y. Implementation issues for the reliable a priori shortest path problem. *Transp Res Rec J Transp Res Board.* 2009;(2091):51–60.
32. Wu X, Nie YM. Modeling heterogeneous risk-taking behavior in route choice: A stochastic dominance approach. *Transp Res part A policy Pract.* 2011;45(9):896–915.
33. Hall RW. Travel outcome and performance: the effect of uncertainty on accessibility. *Transp Res Part B Methodol.* 1983;17(4):275–90.
34. Sivakumar RA, Batta R. The variance-constrained shortest path problem. *Transp Sci.* 1994;28(4):309–16.
35. Sen S, Pillai R, Joshi S, Rathi AK. A mean-variance model for route guidance in advanced traveler information systems. *Transp Sci.* 2001;35(1):37–49.
36. Lo HK, Luo XW, Siu BWY. Degradable transport network: travel time budget of travelers with heterogeneous risk aversion. *Transp Res Part B Methodol.* 2006;40(9):792–806.
37. Dong J, Mahmassani H. Flow breakdown and travel time reliability. *Transp Res Rec J Transp Res Board.* 2009;(2124):203–12.
38. Shahabi M, Unnikrishnan A, Boyles SD. An outer approximation algorithm for the robust shortest path problem. *Transp Res Part E Logist Transp Rev.* 2013;58:52–66.
39. Miller-Hooks E, Mahmassani H. Path comparisons for a priori and time-adaptive decisions in stochastic, time-varying networks. *Eur J Oper Res.* 2003;146(1):67–82.
40. Wu D, Yin Y, Lawphongpanich S. Pareto-improving congestion pricing on multimodal transportation networks. *Eur J Oper Res.* 2011;210(3):660–9.
41. Chu Z, Chen H, Cheng L, Zhu S, Sun C. A Pareto-improving hybrid rationing and pricing policy with multiclass network equilibria. *Transp Plan Technol.* 2018;41(2):211–28.
42. Hau TD. Economic fundamentals of road pricing: a diagrammatic analysis, Part I Fundamentals. *Transportmetrica.* 2005;1(2):81–117.
43. Nie YM, Liu Y. Existence of self-financing and Pareto-improving congestion pricing: Impact of value of time distribution. *Transp Res Part A Policy Pract.* 2010;44(1):39–51.
44. Lawphongpanich S, Yin Y. Solving the Pareto-improving toll problem via manifold

- suboptimization. *Transp Res Part C Emerg Technol.* 2010;18(2):234–46.
45. Guo X, Yang H. Pareto-improving congestion pricing and revenue refunding with multiple user classes. *Transp Res Part B Methodol.* 2010;44(8–9):972–82.
 46. Xiao F, Zhang HM. Pareto-improving toll and subsidy scheme on transportation networks. *Eur J Transp Infrastruct Res.* 2014;14(1).
 47. Zhu D-L, Yang H, Li C-M, Wang X-L. Properties of the multiclass traffic network equilibria under a tradable credit scheme. *Transp Sci.* 2015;49(3):519–34.
 48. Liu Y, Nie YM. A credit-based congestion management scheme in general two-mode networks with multiclass users. *Networks Spat Econ.* 2017;17(3):681–711.
 49. Xiao F, Long J, Li L, Kou G, Nie Y. Promoting social equity with cyclic tradable credits. *Transp Res Part B Methodol.* 2019;121:56–73.
 50. Kockelman KM, Kalmanje S. Credit-based congestion pricing: a policy proposal and the public response. *Transp Res Part A Policy Pract.* 2005;39(7–9):671–90.
 51. Li W, Kockelman KM, Huang Y. Traffic and Welfare Impacts of Credit-Based Congestion Pricing Applications: An Austin Case Study. *Transp Res Rec.* 2021;2675(1):10–24.
 52. Song Z, Yin Y, Lawphongpanich S, Yang H. A Pareto-improving hybrid policy for transportation networks. *J Adv Transp.* 2014;48(3):272–86.
 53. Zheng N, Geroliminis N. Area-Based Equitable Pricing Strategies for Multimodal Urban Networks with Heterogeneous Users. *Transp Res Part A Policy Pract.* 2020;
 54. van den Berg VAC, Verhoef ET. Congestion pricing in a road and rail network with heterogeneous values of time and schedule delay. *Transp A Transp Sci.* 2014;10(5):377–400.
 55. Carrion C, Levinson D. Value of travel time reliability: A review of current evidence. *Transp Res part A policy Pract.* 2012;46(4):720–41.
 56. Zheng N, Rérat G, Geroliminis N. Area-Based Pricing Scheme for Multimodal Systems and Heterogeneous Users in an Agent-based Environment. 2015.
 57. Zheng N, Rérat G, Geroliminis N. Time-dependent area-based pricing for multimodal systems with heterogeneous users in an agent-based environment. *Transp Res Part C Emerg Technol.* 2016;62:133–48.
 58. Brent DA, Gross A. Dynamic road pricing and the value of time and reliability. *J Reg Sci.* 2018;58(2):330–49.
 59. Zhu S, Jiang G, Lo HK. Capturing value of reliability through road pricing in congested traffic under uncertainty. *Transp Res Part C Emerg Technol.* 2018;94:236–49.

60. Varaiya P, Shladover SE. Sketch of an IVHS systems architecture. In: Vehicle Navigation and Information Systems Conference, 1991. 1991. p. 909–22.
61. Chien C-C, Zhang Y, Ioannou PA. Traffic density control for automated highway systems. *Automatica*. 1997;33(7):1273–85.
62. Broucke M, Varaiya P. The automated highway system: A transportation technology for the 21st century. *Control Eng Pract*. 1997;5(11):1583–90.
63. Van Arem B, Van Driel CJG, Visser R. The impact of cooperative adaptive cruise control on traffic-flow characteristics. *IEEE Trans Intell Transp Syst*. 2006;7(4):429–36.
64. Wang M, Treiber M, Daamen W, Hoogendoorn SP, van Arem B. Modelling supported driving as an optimal control cycle: Framework and model characteristics. *Procedia-Social Behav Sci*. 2013;80:491–511.
65. Shladover SE, Su D, Lu X-Y. Impacts of cooperative adaptive cruise control on freeway traffic flow. *Transp Res Rec*. 2012;2324(1):63–70.
66. Zhao L, Sun J. Simulation framework for vehicle platooning and car-following behaviors under connected-vehicle environment. *Procedia-Social Behav Sci*. 2013;96:914–24.
67. Zeng X, Balke K, Songchitruksa P. Potential connected vehicle applications to enhance mobility, safety, and environmental security. Southwest Region University Transportation Center (US); 2012.
68. Talebpour A, Mahmassani HS. Modeling acceleration behavior in a connected environment. In: Midyear Meetings and Symposium Celebrating 50 Years of Traffic Flow Theory. 2014. p. 87–91.
69. Talebpour A, Mahmassani HS. Influence of autonomous and connected vehicles on stability of traffic flow. 2015.
70. Mittal A, Mahmassani HS, Talebpour A. Network flow relations and travel time reliability in a connected environment. *Transp Res Rec*. 2017;2622(1):24–37.
71. Talebpour A, Mahmassani HS, Bustamante FE. Modeling driver behavior in a connected environment: Integrated microscopic simulation of traffic and mobile wireless telecommunication systems. *Transp Res Rec*. 2016;2560(1):75–86.
72. Chen D, Ahn S, Chitturi M, Noyce DA. Towards vehicle automation: Roadway capacity formulation for traffic mixed with regular and automated vehicles. *Transp Res part B Methodol*. 2017;100:196–221.
73. Levin MW, Boyles SD. A multiclass cell transmission model for shared human and autonomous vehicle roads. *Transp Res Part C Emerg Technol*. 2016;62:103–16.
74. Levin MW, Boyles SD. Intersection auctions and reservation-based control in dynamic

- traffic assignment. *Transp Res Rec.* 2015;2497(1):35–44.
75. Ziliaskopoulos A, Mahmassani H. Time-dependent, shortest-path algorithm for real-time intelligent vehicle highway system applications. 1993;
 76. Carré BA. An algebra for network routing problems. *IMA J Appl Math.* 1971;7(3):273–94.
 77. Tabourier Y. All shortest distances in a graph. An improvement to Dantzig’s inductive algorithm. *Discrete Math.* 1973;4(1):83–7.
 78. Fredman ML. New bounds on the complexity of the shortest path problem. *SIAM J Comput.* 1976;5(1):83–9.
 79. Abdelghany K, Hashemi H, Alnawaiseh A. Parallel all-pairs shortest path algorithm: Network decomposition approach. *Transp Res Rec.* 2016;2567(1):95–104.
 80. Habbal MB, Koutsopoulos HN, Lerman SR. A decomposition algorithm for the all-pairs shortest path problem on massively parallel computer architectures. *Transp Sci.* 1994;28(4):292–308.
 81. Hribar MR, Taylor VE, Boyce DE. Implementing parallel shortest path for parallel transportation applications. *Parallel Comput.* 2001;27(12):1537–68.
 82. Song Q, Wang X. Partitioning graphs to speed up point-to-point shortest path computations. In: 2011 50th IEEE Conference on Decision and Control and European Control Conference. 2011. p. 5299–304.
 83. Song Q, Wang X. Efficient routing on large road networks using hierarchical communities. *IEEE Trans Intell Transp Syst.* 2010;12(1):132–40.
 84. Fakhrmoosavi F, Zockaie A, Abdelghany K, Hashemi H. Decomposition of Stochastic Time-Varying Networks for the Path-Finding Problem Considering Travel Time Correlations and Heterogeneity of Users. In: Transportation Research Board 97th Annual Meeting. 2018.
 85. Mahmassani HS, Dong J, Kim J. DYNASMART-P 1.5 User’s Guide and Programmer’s Guide. 2009;
 86. Zockaie A, Chen Y, Mahmassani HS. Adaptive drivers and time-dependent origin-destination demand estimation: methodology and application to large-scale network. 2014.
 87. Zockaie A, Mahmassani HS, Fakhrmoosavi F. Reliability-based user equilibrium in dynamic stochastic networks: A scenario approach considering travel time correlations and heterogeneous users. In: Transportation Research Board 98th Annual Meeting. 2019.
 88. Fu L, Rilett LR. Expected shortest paths in dynamic and stochastic traffic networks. *Transp Res Part B Methodol.* 1998;32(7):499–516.

89. Xu XA, Fakhrmoosavi F, Zockaie A, Mahmassani HS. Estimating Path Travel Costs for Heterogeneous Users on Large-Scale Networks: Heuristic Approach to Integrated Activity-Based Model–Dynamic Traffic Assignment Models. *Transp Res Rec J Transp Res Board*. 2017;2667(1):119–30.
90. Nicosia G, Oriolo G. An approximate A* algorithm and its application to the SCS problem. *Theor Comput Sci*. 2003;290(3):2021–9.
91. Karimi HA, Sutovsky P, Durcik M. Accuracy and performance assessment of a window-based heuristic algorithm for real-time routing in map-based mobile applications. In: *Map-based Mobile Services*. Springer; 2008. p. 248–66.
92. Lysgaard J. A two-phase shortest path algorithm for networks with node coordinates. *Eur J Oper Res*. 1995;87(2):368–74.
93. Karimi HA. Real-time optimal-route computation: a heuristic approach. *J Intell Transp Syst*. 1996;3(2):111–27.
94. Bagloee SA, Sarvi M, Patriksson M. A hybrid branch-and-bound and benders decomposition algorithm for the network design problem. *Comput Civ Infrastruct Eng*. 2017;32(4):319–43.
95. Fakhrmoosavi F, Zockaie A, Abdelghany K, Hashemi H. An iterative learning approach for network contraction: Path finding problem in stochastic time-varying networks. *Comput Civ Infrastruct Eng*. 2019;34(10):859–76.
96. Zockaie A, Nie YM, Mahmassani HS. Simulation-based method for finding minimum travel time budget paths in stochastic networks with correlated link times. *Transp Res Rec*. 2014;2467(1):140–8.
97. Hahn GJ. Sample sizes for Monte Carlo simulation. *IEEE Trans Syst Man, Cybern Syst*. 1972;
98. Oakley JE, Brennan A, Tappenden P, Chilcott J. Simulation sample sizes for Monte Carlo partial EVPI calculations. *J Health Econ*. 2010;29(3):468–77.
99. Kim J, Mahmassani H, Alfelor R, Chen Y, Hou T, Jiang L, et al. Implementation and evaluation of weather-responsive traffic management strategies: insight from different networks. *Transp Res Rec J Transp Res Board*. 2013;(2396):93–106.
100. Zockaie A, Mahmassani HS, Nie Y, Fakhrmoosavi F. Reliability-Based User Equilibrium Formulation in Dynamic Stochastic Networks with Correlated Travel Times and Heterogeneous User Preferences. In: *Transportation Research Board 97th Annual Meeting*. 2018.
101. Daganzo CF, Garcia RC. A Pareto improving strategy for the time-dependent morning commute problem. *Transp Sci*. 2000;34(3):303–11.

102. Liu Y, Guo X, Yang H. Pareto-improving and revenue-neutral congestion pricing schemes in two-mode traffic networks. *Netnomics*. 2009;10(1):123–40.
103. Li H, Bliemer MCJ, Bovy PHL. Network reliability-based optimal toll design. *J Adv Transp*. 2008;42(3):311–32.
104. Boyles SD, Kockelman KM, Waller ST. Congestion pricing under operational, supply-side uncertainty. *Transp Res Part C Emerg Technol*. 2010;18(4):519–35.
105. Fakhrmoosavi F, Zockaie A, Abdelghany K. Incorporating Travel Time Reliability in Equitable Congestion Pricing Schemes for Heterogeneous Users and Bimodal Networks. *Transp Res Rec [Internet]*. 0(0):03611981211019737. Available from: <https://doi.org/10.1177/03611981211019737>
106. Abkowitz MD, Engelstein I. Factors affecting running time on transit routes. *Transp Res Part A Gen*. 1983;17(2):107–13.
107. Yetiskul E, Senbil M. Public bus transit travel-time variability in Ankara (Turkey). *Transp Policy*. 2012;23:50–9.
108. Clerc M. Standard Particle Swarm Optimisation [Internet]. 2012. Available from: <https://hal.archives-ouvertes.fr/hal-00764996>
109. Fakhrmoosavi F, Kavianipour M, Shojaei M, Zockaie A, Ghamami M, Wang J, et al. Electric Vehicle Charger Placement Optimization in Michigan Considering Monthly Traffic Demand and Battery Performance Variations. *Transp Res Rec*. 2020;0361198120981958.
110. Sadeghi J, Sadeghi S, Niaki STA. Optimizing a hybrid vendor-managed inventory and transportation problem with fuzzy demand: an improved particle swarm optimization algorithm. *Inf Sci (Ny)*. 2014;272:126–44.
111. Do Chung B, Yao T, Friesz TL, Liu H. Dynamic congestion pricing with demand uncertainty: A robust optimization approach. *Transp Res Part B Methodol*. 2012;46(10):1504–18.
112. Zheng Y-J, Chen S-Y. Cooperative particle swarm optimization for multiobjective transportation planning. *Appl Intell*. 2013;39(1):202–16.
113. Clerc M, Kennedy J. The particle swarm-explosion, stability, and convergence in a multidimensional complex space. *IEEE Trans Evol Comput*. 2002;6(1):58–73.
114. Lindfield G, Penny J. Introduction to nature-inspired optimization. Academic Press; 2017.
115. Fakhrmoosavi F, Zockaie A, Abdelghany K, Hashemi H. A Network Contraction Algorithm Using an Iterative Learning Approach for Path Finding Problem in Stochastic Time-Varying Networks considering Travel Time Correlations. In: *Transportation Research Board 98th Annual Meeting*. 2019.

116. Leclercq L, Verchier A, Krug J, Menendez M. Investigating the performances of the method of successive averages for determining dynamic user equilibrium and system optimum in Manhattan networks. In: DTA2016, 6th International Symposium on Dynamic Traffic Assignment. 2016.
117. Stabler B. TransportationNetworks/SiouxFalls [Internet]. 2020. Available from: <https://github.com/bstabler/TransportationNetworks/tree/master/SiouxFalls>
118. Balakrishna R, Morgan D, Yang Q, Slavin H. Comparison of simulation-based dynamic traffic assignment approaches for planning and operations management. In: Vortrag: 91st Annual Meeting of the Transportation Research Board, Washington, DC. 2012.
119. Zockaie A, Mahmassani HS, Saberi M, Verbas Ö. Dynamics of urban network traffic flow during a large-scale evacuation. *Transp Res Rec.* 2014;2422(1):21–33.
120. Mahmassani HS, Chen PS-T. Comparative assessment of origin-based and en route real-time information under alternative user behavior rules. *Transp Res Rec.* 1991;1306:69.
121. Mahmassani HS, Saberi M, Zockaie A. Urban network gridlock: Theory, characteristics, and dynamics. *Transp Res Part C Emerg Technol.* 2013;36:480–97.
122. Federal Highway Administration (FHWA). Next Generation Simulation: US-101 Highway Dataset. 2006.
123. Fakhrmoosavi F, Saedi R, Zockaie A, Talebpour A. Impacts of Connected and Autonomous Vehicles on Traffic Flow with Heterogeneous Drivers Spatially Distributed over Large-Scale Networks. *Transp Res Rec.* 2020;
124. Hamdar SH, Treiber M, Mahmassani HS, Kesting A. Modeling driver behavior as sequential risk-taking task. *Transp Res Rec.* 2008;2088(1):208–17.
125. Kahneman D, Tversky A. Prospect theory: An analysis of decision under risk. In: *Handbook of the fundamentals of financial decision making: Part I.* World Scientific; 2013. p. 99–127.
126. Hamdar SH. Modeling driver behavior as a stochastic hazard-based risk-taking process, civil and environmental engineering. PhD Diss Georg Washingt Univ Washington, USA. 2009;
127. Hamdar SH, Treiber M, Mahmassani HS. Calibration of a Stochastic Car-Following Model Using Trajectory Data: Exploration and Model Properties. In: *Transportation Research Board 88th Annual Meeting.* 2009.
128. Sheffi Y. *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods.* 1984.
129. Godfrey JW. The mechanism of a road network. *Traffic Eng Control.* 1969;8(8).
130. Mahmassani HS, Williams JC, Herman R. Investigation of network-level traffic flow

- relationships: some simulation results. *Transp Res Rec.* 1984;971:121–30.
131. Geroliminis N, Daganzo CF. Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. *Transp Res Part B Methodol.* 2008;42(9):759–70.
 132. Kavianipour M, Saedi R, Zockaie A, Saberi M. Traffic State Estimation in Heterogeneous Networks with Stochastic Demand and Supply: Mixed Lagrangian--Eulerian Approach. *Transp Res Rec.* 2019;0361198119850472.
 133. Saberi M, Mahmassani H, Hou T, Zockaie A. Estimating network fundamental diagram using three-dimensional vehicle trajectories: extending edie's definitions of traffic flow variables to networks. *Transp Res Rec J Transp Res Board.* 2014;(2422):12–20.
 134. Geroliminis N, Sun J. Properties of a well-defined macroscopic fundamental diagram for urban traffic. *Transp Res Part B Methodol.* 2011;45(3):605–17.
 135. Saedi R, Saeedmanesh M, Zockaie A, Saberi M, Geroliminis N, Mahmassani HS. Estimating network travel time reliability with network partitioning. *Transp Res Part C Emerg Technol.* 2020;112:46–61.