Passenger-focused Scheduled Transportation Systems: from Increased Observability to Shared Mobility

by

Jiangtao Liu

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved March 2018 by the
Graduate Supervisory Committee:

Xuesong Zhou, Chair
Pitu Mirchandani
Ram Pendyala
Yingyan Lou

ARIZONA STATE UNIVERSITY

May 2018

ABSTRACT

Recently, automation, shared use, and electrification are proposed and viewed as the "three revolutions" in the future transportation sector to significantly relieve traffic congestion, reduce pollutant emissions, and increase transportation system sustainability. Motivated by the three revolutions, this research targets on the passenger-focused scheduled transportation systems, where (i) the public transit systems provide high-quality ridesharing schedules/services and (ii) the upcoming optimal activity planning systems offer the best vehicle routing and assignment for household daily scheduled activities.

The high quality of system observability is the fundamental guarantee for accurately predicting and controlling the system. The rich information from the emerging heterogeneous data sources is making it possible. This research proposes a modeling framework to systemically account for the multi-source sensor information in urban transit systems to quantify the estimated state uncertainty. A system of linear equations and inequalities is proposed to generate the information space. Also, the observation errors are further considered by a least square model. Then, a number of projection functions are introduced to match the relation between the unique information space and different system states, and its corresponding state estimate uncertainties are further quantified by calculating its maximum state range.

In addition to optimizing daily operations, the continuing advances in information technology provide precious individual travel behavior data and trip information for operational planning in transit systems. This research also proposes a new alternative modeling framework to systemically account for boundedly rational decision rules of travelers in a dynamic transit service network with tight capacity constraints. An agent-

based single-level integer linear formulation is proposed and can be effectively by the Lagrangian decomposition.

The recently emerging trend of self-driving vehicles and information sharing technologies starts creating a revolutionary paradigm shift for traveler mobility applications. By considering a deterministic traveler decision making framework, this research addresses the challenges of how to optimally schedule household members' daily scheduled activities under the complex household-level activity constraints by proposing a set of integer linear programming models. Meanwhile, in the microscopic car-following level, the trajectory optimization of autonomous vehicles is also studied by proposing a binary integer programming model.

DEDICATION

To

My beloved wife, Yongzhao

and

My beloved twin daughters, Yuqian (Amy) and Yumo (Mia)

ACKNOWLEDGMENTS

Life is not easy. When you feel it not hard as you imagine, surely someone is bearing that hard for you. Looking back to my five-year doctoral study, there are so many people who deserve my special acknowledgement.

First, I wish to express my gratitude to my advisor, Dr. Xuesong Zhou. It is impossible to finish this dissertation without his insightful guidance. As an advisor, his hard work and insistence inspired me to accept any challenges and never give up. As a friend, his sharing and understanding gave me enough support and freedom to handle many difficulties and reflect on myself.

I would like to thank my committee members. Dr. Pitu Mirchandani leaded me to the optimization and network flow theory area in his courses and gave me confidence to learn more by myself. Dr. Ram Pendyala provided the opportunity to learn the integration of Activity-based Model (ABM) and Dynamic Traffic Assignment (DTA) and gave a strong recommendation for my internship. Dr. Yingyan Lou taught me transportation operations and the course project about traffic state estimation made me learn a lot. I also thank them for taking time to serve my committee and asking questions to create insights. In addition, I appreciate the collaborations with Dr. Jee Eun Kang from the University at Buffalo about household activity pattern problem and with Dr. Chung-Cheng Lu from National Chiao Tung University about eco-system optimal dynamic traffic assignment problem.

I also greatly appreciate the unforgettable and enjoyable experiences with my friends. Specifically, thanks are to my colleagues Peiheng Li, Dr. Pengfei Li, Jun Xiao, Xiushuang Li, Sravani Vadlamani, Monireh Mahmoudi, Nana Zhu, Alisa Bonz, Golnoosh Miri, Venu

Garikapati, Daehyun You, and Hossein Jalali. Thanks are also to those visiting scholars and students in our research team, including Dr. Yuguang Wei, Dr. Jianrui Miao, Dr. Junhua Chen, Dr. Yunchao Qu, Dr. Jinjin Tang, Tie Shi, Lu Tong, Ran Chen, Cafer Avcı, Pan Shang and Yongxiang Zhang. I also want to thank Xin Wang for all his help and Baloka Belezamo for his friendship.

Last, but not the least, I would like to deeply thank my family. My parents and my parents-in-law always provide their most selfless help; my wife and my twin daughters always give me the strongest support; and my brother and my brother-in-law always offer their most sincerely care.

TABLE OF CONTENTS

Page

CHAPTER Page

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

.

LIST OF TABLES

Table Page

Table Page

Table Page

CHAPTER 1

INTRODUCTION

1.1 Motivation and Problem Definition

The eternal principle of transportation systems is to serve passengers or freights with a timely and reliable journey from origin to destination. The continuous increase of travel demand and the mismatching low capability of service supply is resulting in traffic congestion, air quality and sustainability issues. Various approaches responding to them are being proposed and studied, including active transportation demand management, public transportation-oriented development, transportation network expansion plan, and intelligent vehicle highway system deployment. In addition, the recently emerging advanced sensing, telecommunications and vehicular technologies, are driving a new wave of rich information (e.g., GPS and Smartphone trajectory data, connected and autonomous vehicles test data, video image processed data and etc.,) and providing a great opportunity to completely unveil the inner mechanism of transportation system operations, while creating unprecedentedly unclear impacts on human activity- travel behavior, system operations and planning, and land use.

The hidden principle of those feasible solutions and possible uncertainties above for system performance improvements can be summarized as how to best schedule/allocate the limited resources to reach the goal of optimizing pre-oriented performance measures. In other words, best serving passenger or goods from origin to destination is still a typical scheduling problem in transportation systems. Specifically, the scheduled transportation system usually refers to train or bus services that can only be accessed at certain times

(according to the timetable) and certain locations (Nuzzolo and Crisalli, 2009). The main goal of public transit systems is to provide a reliable scheduled service with high passenger-carrying capacity and low environmental impacts to satisfy the time-dependent characteristics of travel demands. In addition, from the perspective of scheduling, the current available transportation management decisions are directly or indirectly making schedules for system performance improvements: (1) active travel demand management aims to schedule passengers' departure time; (2) advanced traveler information system tries to schedule passengers' path trajectory; (3) signal control and ramp metering intends to schedule the supply services to those incoming vehicles; (4) network construction and expansion can be viewed as offering continuous service/schedule on the new roads for vehicles. Actually, going back to focus on the origin of travel demand, the daily activities generated by each person in each household are generally also scheduled with preferred arrival and departure time windows at candidate locations to achieve mandatory or optional goals. Further, the upcoming autonomous vehicles and currently developing information sharing technologies starts creating a revolutionary paradigm shift in the coming years for traveler mobility applications. The optimal schedule of vehicles on household-level scheduled activity chains is critical and more valuable than single-demand responsive service providing.

Recently, automation, shared use, and electrification are proposed as the "three revolutions" in the future transportation sector to significantly relieve traffic congestion, reduce pollutant emissions, and increase transportation system sustainability. Exactly, the traditional scheduled public transit system satisfy the three requirements and should be enhanced and integrated in future with the flexible autonomous vehicle applications where

the shared use of autonomously electric vehicles is also scheduled to perform household members' scheduled activity chain. Of course, it is admitted that there are much flexible thinking about the future transportation system, such as, whether autonomous vehicles should be owned by person or managed by the agency, how to seamlessly integrate the transit systems with autonomous vehicles, whether people are really accepting the ridesharing with just a couple of thoroughly strange persons in a space-limited vehicle. This research conducted in this dissertation with careful assumptions is just one starting point.

As a short summary, by considering the current and future transportation system characteristics, the scheduled transportation system studied in this dissertation is (i) a passenger-focused public transit system for the public and (ii) an information-shared and rideshared vehicle systems for household members' scheduled activity performing.

There is nothing more important than understanding how the real-world transportation system works and how to improve it. The rich information from the emerging heterogeneous data sources is making it possible, including loop detector data, automatic vehicle identification data, GPS and Smartphone semi-continuous trajectory data, outside sensing data from moving connected and autonomous vehicles, video image processed data. Compared to the previously low response rates from travel surveys for transit data collection, the current big data provides more valuable information to better observe time-varying transit conditions and accordingly propose adaptive travel demand management and supply (capacity) control strategies. On the other hand, in face of the overwhelming data, the leverage of data on supporting system operational decisions should be

meticulously examined, and the increased system observability should be analytically quantified to measure the value of different information and proposed daily countermeasures.

In addition to optimizing daily operations, the continuing advances in information technology provide precious individual travel behavior data and trip information for operational planning in transit systems. Taking the transit service network design problem as an application, the individual traveler agents, tight time-varying vehicle capacity constraints, and individual traveler's behavior are systematically addressed for designing transit service networks, such as, constructing a new transit line with detailed schedules, adding train or bus schedules, or dynamically selecting different types of vehicles.

The traditional scheduled transportation system above is currently playing a major part in shared mobility and will be enhanced rather than weakened in future to improve the transportation system efficiency. Meanwhile, the advanced information-sharing technologies are making it possible to provide the optimal vehicle routes and activity selection choices to each household member to finish his/her scheduled daily activities from the household level or from the transportation system level. In addition, with the upcoming autonomous vehicles, the micro-level vehicle trajectory planning and the macro-level vehicle route guidance for shared mobility are also meaningful and will be addressed in this dissertation.

1.2 Objectives and Challenges

This dissertation aims to improve the scheduled dynamic transportation system performance from its daily operations to the medium-term operational planning based on

the currently available multi-source sensor data, and further make the best vehicle scheduling for performing those scheduled daily activities of each household member with the help of the emerging information-shared technologies and upcoming autonomous vehicles.

The specific objectives and challenges for each problem above are explained as follows.

(1) Observe the urban transit system in terms of flexible system state definitions based on the currently available multi-source sensor data. Specifically, the following crucial questions are still not clear and need to be deeply explored: (i) how to mathematically represent the available multi-source information in a unified way so that different system states can be estimated based on the well-developed information representation, (ii) what is the exact inner relation between the information and different focused system states, (iii) how to quantify the observability or the uncertainty of estimated states on the basis of available multi-source information with result consistency, and (iv) how to develop a theoretically elegant model and design a correspondingly efficient algorithm to solve the model. So finally the system observability quantification can be used for further system state prediction and daily optimal operational control.

(2) Perform operational planning by service network design in dynamic transit systems based on the general individual traveler trip information and realistic travel behavior assumption analyzed from the multi-source sensor data above. Specifically, the following challenging questions need to be carefully addressed: (i) what is the generally realistic travel behavior and the stable system condition under tight vehicle capacity constraints, (ii) how to model the interaction of individual traveler in the dynamic transit systems, (iii) how

to build an elegant mathematical model for this classical and challenging network design problem, and (iv) how to design an efficient algorithm to solve the proposed model for large-scale networks.

(3) Schedule and assign the available vehicles to perform the scheduled different daily activities of each household member under the complex household activity constraints and interactions. Specifically, we aim to study the following scenarios and questions under the condition that all available human-driven or autonomous vehicles are owned by households rather than the private companies or the government: (i) there is no ridesharing, so each household member will drive alone or choose to stay at home; (ii) there is a ridesharing among the household members, and the driver is known or the autonomous vehicle will always stay with someone rather to serve other households; (iii) how to capture the road congestion endogenously formed by those activities' realization; (iv) how to model those complex requirements and solve it efficiently for the large-scale networks; and (v) how to plan autonomous vehicles' trajectories or schedule their each moving step from the micro-level car-following perspective. Of course, we admit that autonomous vehicles probably will be managed by companies or government, and the correspondingly open questions are numerous and absolutely worthy in hard future research, such as, how many autonomous vehicles and depots are really needed, how to locate those depots, how to protect the privacy in the ridesharing, how to build an optimal multi-modal transportation system and so on.

1.3 Research Overview

This dissertation first develops a modeling framework in the time-discretized space-time networks or space-time-state networks to systematically understand and improve the passenger-focused scheduled dynamic transportation system, ranging from the near-term system observability, the medium-term operational planning, to the upcoming optimal vehicle scheduling and assignment for household-level scheduled activity chain rather than single activity responsive transport. Our specific conducted research are explained as follows.

(1) Accurately estimating what the system is happening based on available information/observations is the foundation to deeply understand the system working mechanism and further optimally control the system. The theoretical relation among sensor data, states, and observability (or the uncertainty of state estimates) in urban dynamic transit systems is explored. Specifically, (i) the information space is generated by a system of linear equations and inequality constraints as the multi-source information representation; (ii) The information errors are considered by a least square model to correct the directly observed measurements, such as, trip time of grouped passengers from smart card, aggregated passenger count from video systems, path choice information from cell phone trajectory data; (iii) Different projection functions are proposed and used to map the unique information space with our focused different system states (e.g., passenger density on the station platform, in the vehicle, or in the transfer corridor) for further state uncertainty quantifications; (iv) Our developed models are finally solved as a simplified linear programming model by using Frank-Wolfe algorithm and Dantzig-Wolfe

decomposition algorithm, which improves the computational efficiency of solving our models.

(2) From the perspective of operational planning, the aim of the transit service network design problem is to provide better service to users and to increase operating efficiency. Specifically, (i) Focusing on a dynamic transit service network with tight capacity, we utilize the property of constant travel times on space-time arcs to formulate the boundedly rational travel behavior of each traveler through a set of integer linear inequalities; (ii) By exogenously listing a set of viable space-time path alternatives for each agent, we offer a single-level 0-1 integer linear programming model to study the complex discrete network design problem under a set of quite specific but realistic assumptions. With the aim of minimizing the total transit system travel time, this new reformulation avoids the use of possible non-convex flow-based models where its link travel time is usually represented by a kind of nonlinear functions. (iii) Although the proposed agent-based and time-dependent formulation introduces additional dimensions and a large number of binary variables, it is further shown that, after dualizing hard constraints (i.e., capacity and rational decision constraints), the original problem can be decomposed into two sub-problems. These sub-problems have computationally efficient algorithms available on large scale networks, namely a time-dependent least cost path problem and a knapsack problem.

(3) With the upcoming information-sharing and autonomous vehicle technologies, it become possible to implement the optimal vehicle schedule and assignment to each household member to finish his/her scheduled daily activities with different purposes, while creating a large number of uncertainties on travelers' mobility and the future

transportation system pattern. Based on our careful assumptions as a starting point to discuss the future potential changes, two special cases are considered and further the road congestion incurred by finishing those activities is also embedded in our models.

Specifically, (i) we consider Case A as a multi-vehicle and multi-person vehicle routing problem with mandatory and discretionary activities. (ii) With the given ride-sharing options for each household, we model our Case B as a multi-vehicle and multi-person ridesharing problem with mandatory and discretionary activities. (iii) These two problems are formulated as 0-1 integer linear programming models in a space-time network and a space-time-state network, respectively. (iv) The road capacity constraint is considered to model the network congestion and resulting activity pattern change. (v) Through dualizing the capacity constraints to the objective function by Lagrangian relaxation, our proposed model can be further solved through time-dependent state-dependent least cost path-finding algorithms, which permits the use of fast computational algorithms on large-scale high-fidelity transportation networks. (vi) An integer programming models is presented for scheduling autonomous vehicles' longitudinal trajectories under the safety requirement and various vehicle communication characteristics.

1.4 Organization of the Dissertation

As shown in Figure 1-1, this research targets on the passenger-focused scheduled transportation system including the public urban transit system and the household vehicle scheduling system. For the public transit system, the system observability problem in daily operation level and the service network design problem in operational planning level are well studied. In addition, the vehicle routing and assignment for household daily scheduled

activities in the route level and the trajectory planning of autonomous vehicles in car-following level are also carefully investigated. Since the system observability is the foundation of optimizing any transportation systems rather than just public transit system, we put the research on observability quantification at Chapter 5 rather than Chapter 3 to summarize the insights of how to use the overwhelming multi-source sensor data in transportation systems. However, the seamlessly integrated future multi-modal scheduled transportation system is still beyond the scope of this dissertation. The summary of each chapter is explained as follows.

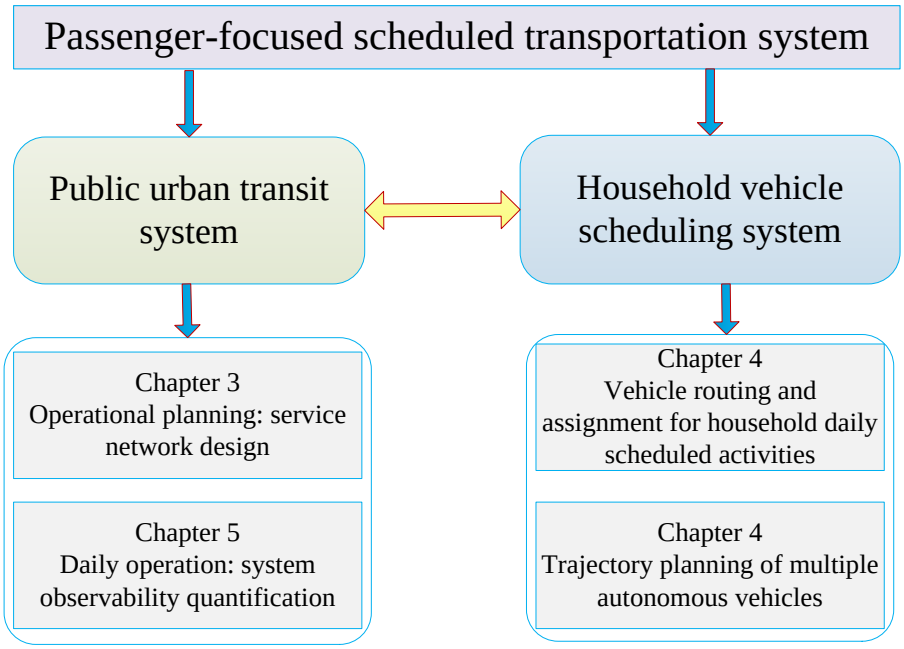


Figure 1-1 Study Flowchart of This Dissertation

Chapter 2 provides a comprehensive review on passenger-focused transportation state estimation and system observability, transportation service network design, and vehicle routing and assignment for travelers' scheduled activities.

At the operational planning level, Chapter 3 offers a modelling framework for the dynamic transit service network design problem to consider individual traveler and his/her realistic travel behaviors under oversaturated transit conditions. First, a simple network is used to show the final traffic conditions and modeling challenges based on different travel behavior assumptions under tight link capacity constraints. Then we propose a single-level integer linear programming model to capture the bounded rational travel behavior. The model can be further decomposed by Lagrangian relaxation as a time-dependent shortest path problem and a knapsack problem, which can be efficiently solved by currently classical algorithms. Numerical experiments are also performed to demonstrate our methodology and algorithms.

Chapter 4 directs to the household-level scheduled activities, which could be optimally performed by each individual with the upcoming information-sharing and autonomous vehicle technologies. Two special cases, solo driving and ride sharing, are considered by making vehicle scheduling and assignment for each household member, and the road congestion is further incorporated in the proposed models, which can be efficiently solved to obtain a lower bound by Lagrangian relaxation under the rapid development of hardware of computers in memory and fast computation speed. In addition, the trajectory planning of multiple autonomous vehicles is studied based on Newell's simplified car-following model.

Chapter 5 first presents a conceptual illustration about the relationship among sensor, system states, and system observability in a general transportation network rather than in transit systems, and then proposes a modeling framework for our scheduled urban transit system to quantify the uncertainty of system state estimate (system observability) based on

the different available multi-source sensor data in time-discretized space-time networks. The proposed models finally can be transformed as a linear programming problem by Frank-Wolfe algorithm and Dantzig-Wolfe decomposition algorithm. Numerical experiments are conducted following our proposed methodology and algorithms.

Finally, Chapter 6 summarizes the research of this dissertation by conclusions and contributions, and also discuss the future research directions under consideration.

CHAPTER 2 LITERATURE REVIEW

This chapter reviews several topics relevant to passenger-focused scheduled transportation systems. Section 2.1 presents a general overview of observability quantification or uncertainty quantification of system state estimates in traffic networks and urban transit systems. In traffic systems with heterogamous sensor sources, the Origin-Destination matrix estimation problem, the sensor network design problem, and the traffic states (e.g. path/link travel time, queuing delay) estimation problem are related with each other. In urban transit systems, the smart card data are mainly used to estimate passengers' trip destination in bus transit system and study passengers' travel behavior or route choice in rail transit system. Section 2.2 provides a comprehensive review on traffic and transit network design problems, where the travel behavior assumptions, physical resource constraints, modeling approach, and algorithms designed for the bi-level programming problem are key elements. In Section 2.3, how the household daily scheduled activities can be realized is reviewed. Meanwhile, vehicle routing and assignment for single on-demand activity request and the emerging new transit service are also discussed.

2.1 System Observability Quantification in Transportation Systems

2.1.1 State Estimation and Sensor Network Design in Traffic Systems

Observability is a concept introduced by Kalman (1959) for linear dynamic systems in control theory. It is a measure for how well internal states of a system can be inferred by knowledge of its external outputs. In other words, it aims to quantify or measure the uncertainty of estimated internal states based on the available external observations under a given sensor environment with sampling errors, sensor error and model errors. A

comprehensive literature review can be found at the paper (Castillo et al., 2015). As for evaluating the estimation uncertainty or accuracy, origin-destination (OD) trip matrix estimation is a widely studied classical problem due to its under-determination attribute, which means that there is an infinite number of OD trips that can generate link flows consistent with the observations. Yang et al. (1991) first introduced the concept of Maximum Possible Relative Error (MPRE) to theoretically investigate the estimation uncertainty and reliability of the OD estimated trips obtained by the entropy model. Bianco et al. (2001) further explored the accuracy of estimated OD matrix bound under different sensor location strategies. In addition, Bierlaire (2002) proposed the concept of total demand scale as a new measure to examine the quality of estimated OD trip tables from link counts, by maximizing/minimizing the total travel demand satisfying all observations. In the general transportation observability problems, a number of research (Castillo et al., 2007; Castillo et al., 2008; Gentili and Mirchandani, 2012;) modeled the problems as a system of linear equations and/or inequalities and then determine whether the system or one unknown variable is observable or not by analyzing the properties of its coefficient matrix. Meanwhile, in the system of linear inequalities, a general bound of unknown variables can be derived through the dual cone approach. In general, the observability problem more cares about the list of variables to be observed rather than the specific system states uncertainty ranges.

In order to increase the estimation quality, the integration of ubiquitous sensor network design and state estimation has received a growing attention in the past few years. Yang and Zhou (1998) proposed integer linear programming models and four sensor location rules to determine the optimal number and location of point sensors for origin-destination

matrix estimation. Based on the trace of the a posteriori covariance matrix produced in a Kalman filtering model, Zhou and List (2010) offered an information-theoretic framework for locating fixed sensors in the traffic OD demand estimation problem. In addition, based on the observability problem definition, the optimal count number and location of active sensors (Gentili and Mirchandani, 2005) and counting and scanning sensors (Castillo et al., 2012) for estimating path/link flows are studied by analyzing a set of linear equations. Hu et al. (2009) proposed one "basis link" method to find the smallest subset of links in a network to locate sensors so that the traffic flow of all links can be accurately estimated under steady-state conditions, so finally the OD information and route choice behavior are not necessary. Further, Ng (2012) introduced a new solution approach (''synergistic sensor location'') to avoid possible path enumeration under the assumed steady-state conditions. Xu et al. (2016) proposed a robust network sensor design to completely observe link flows whiling accounting for the accumulation of observation errors. For other traffic states, link travel time estimation errors are selected as the optimization criterion for point sensor location problems (Ban et al., 2009; Danczyk and Liu, 2011), and a reliable sensor location method is proposed to consider probabilistic sensor failures (Li and Ouyang, 2011). Based on a Kalman filtering structure, Xing et al. (2013) developed measurement and uncertainty quantification models to explicitly consider several important sources of errors in path travel time estimation/prediction. In the real-time traffic conditions, Eisenman et al. (2006) conducted a sensitivity analysis of estimation and prediction accuracy under different sensor locations and coverage scenarios based on a real-time dynamic simulation system, DYNASMART-X. Boylse and Waller (2011) studied the optimal location selection for providing the real-time traffic information to drivers with the adaptive travel behavior by

proposing heuristic algorithms, and Ban et al. (2011) studied the real-time queue length estimation at signalized intersections by focusing on queuing delay patterns and queue length changes based on travel times from mobile traffic sensors.

2.1.2 State Estimation and Smart Card Use in Urban Transit Systems

In urban transit systems, usually, the automatic fare collection system (AFC) or smart card usually records both the time and station for entry and exit for each passenger in rail transit systems, but only the boarding time and stop and route number normally can be reported in bus transit systems. A comprehensive literature review about smart card data use can be found in the papers (Pelletier et al., 2011; Ma et al., 2013). Obviously, the unknown destination information greatly increase the state uncertainty of bus transit system. Trépanier et al. (2007) estimate the alighting point for each passenger based on the smallest distance to the boarding stop of his/her next route from individually continuous riding records in smart card. Seaborn et al. (2009) proposed maximum elapsed time thresholds to identify transfers for bus-to-underground, underground-to-bus, and bus-to-bus to identify and assess multi-modal trips in London. Meanwhile, Munizaga and Palma (2012) estimated a multimodal transport OD matrix from smartcard and GPS data whiling consider unobserved trips by expansion factors in Santiago, Chile. Yuan et al. (2013) proposed a space alignment approach by considering the aligning the monetary space and geospatial space with the temporal space to infer each passenger's trajectory and the results improve the detection of uses' home and work places. Nassir et al. (2015) applied the smart card data to detect activity and identify transfers to estimate the true origins and destinations. Nunes et al. (2016) further proposed four endogenous spatial validation rules to enhance

the accuracy of estimated passenger destination choice. Alsger et al. (2016) evaluated and improved existing OD estimation method according to available OD information and assessed the previous last destination assumptions in bus transit systems. Under the situation that passenger's boarding stop information is not record in smart cards, Ma et al. (2012) developed a Markov chain based Bayesian decision tree algorithm to estimate the sequential stops on the bus route and then match those stops with the recorded boarding time to infer passengers' origin. Further, Ma et al. (2015) improved their previous algorithms to increase the estimation accuracy and computation efficiency.

Depending on the available OD travel information from smart card in urban rail transit systems, a number of studies focus on the route choices and transfer patterns, which can be viewed as different system states necessary for estimation. Kusakabe et al. (2010) focused on the passengers' train choice behavior by assuming that each passenger aims to minimize the total waiting time at the departure station, loss time at the arrival station, and the transfer frequency. Usually, in the logit discrete choice models the tight side constraints (e.g. strict vehicle capacity constraint) are still hard to include. Sun and Xu (2012) estimated the path choice based on the observed overall probability density of journey time and the derived distribution of individual path travel time from the rail transit smart card. In addition, Zhou and Xu (2012) used a matching degree function value to assign the trip to the most likely path based on derived boarding plan of path. Remarked that the verification of those assigned path flows above has not been soundly performed due to the limit of observed data and complicated route choice behaviors, such as passenger's particular travel preference. Kusakabe and Asakura (2014) proposed a data fusion methodology to consider both the smart card data and person trip survey data by Bayes probabilistic model to

estimate behavioral attributes of trips in the smart card data. Based on passenger OD matrix information and vehicle stop time and location data, Zhu et al. (2017a; 2017b) proposed probabilistic models to estimate the individual train loads, left behind probabilities, time-dependent crowding levels at stations and etc., under tight vehicle capacity considerations.

2.2 Transportation Network Design Problem

2.2.1 Traffic Network Design Problem

The discrete traffic network design problem has been traditionally formulated as a bi-level programming model, where the upper-level problem decides where and how many links should be built, and the lower-level problem aims to predict travelers' response to changes in the network conditions, by assuming certain user behavior rules such as Wardrop or Nash user equilibrium. A number of comprehensive reviews on the discrete traffic network design problem have been offered by the classical papers by Magnanti and Wong (1984), Yang and Bell (1998), as well as a recent study by Farahani et al. (2013). Typical solution algorithms in the discrete cases include branch and bound (Leblanc, 1975), support functions in bender decomposition (Gao et al., 2005), meta-heuristics (e.g. Xiong and Schneider, 1992; Drezner and Wesolowsky, 2003; Poorzahedy and Rouhani, 2007).

Another active research line for considering user equilibrium (UE) conditions is to construct a single-level programming model with constraints corresponding to the UE principle. The early work by Bard and Moore (1990) provided a reformulation based on the Karush–Kuhn–Tucker (KKT) conditions with a branch and bound solution scheme. Recently, Farvaresh and Sepehri (2011) presented a single-level mixed integer linear problem (MILP) by representing the UE condition as KKT constraints and linearized those

non-linear terms by introducing binary auxiliary variables. In addition, Lou et al. (2009) studied a robust network design approach and formulated the problem as a mathematical program with complementarity constraints, where a cutting-plane scheme was proposed for solving this problem with demand uncertainties. Wang and Lo (2010) recast the lower-level model by complementarity constraints, and then introduced a set of binary variables to transform the "if-then" conditions into an equivalent set of linear constraints, which can be further approximated with a MILP framework to search for global optimal solutions. Luathep et al. (2011) formulated the user equilibrium condition as a variational inequality problem, and also adopted the linear approximation technique to transform the original problem as an MILP model, which is solvable based on a cutting constraint method. Wang et al. (2013) developed two types of global optimization methods, and one of them involves using the system optimal (SO) traffic assignment principle to construct effective lower bounds. Recently, Wang et al. (2015) considered both which links should be built and how many capacities should be assigned to those built links simultaneously, and then proposed a global optimization method incorporating linearization, outer approximation and range reduction to solve this problem.

Those previous studies, based on either bi-level programming or transformed single-level models, have made great contributions to a deep understanding and numerical efficient algorithm development for different classes of discrete traffic network design problems. To the best of our knowledge, still very few studies have completely considered the tight transportation capacity and (reasonable and realistic) user equilibrium in the context of bi-level or single-level network design framework, especially for time-dependent transit service networks.

2.2.2 Transit Service Network Design Problem

This research takes particular interests in the dynamic transit service network design, as a representative example of discrete transportation service network design in a space-time network. There are a wide range of studies on transit network design, represented by excellent survey papers such as Guihaire and Hao (2008), Kepaptsoglou and Karlaftis (2009), Farahani et al. (2013), and Ibarra-Rojas et al. (2015). Table 2-1 lists a number of studies closely related to the problem under consideration in this chapter, with a special focus on topics such as travelers' behavior in terms of deterministic user equilibrium or Nash equilibrium, time-dependent travel demand, space-time network representation, and general transit assignment problems which is the lower level of the transit service network design problem.

As listed in Table 2-1, in some early studies (Nguyen and Pallottino, 1988; Spiess and Florian, 1989; De Cea and Fernandez, 1993; Nuzzolo et al., 2001; Gao et al., 2003), the tight capacity constraints are not considered for user equilibrium in transit assignment/network design problems; in some succeeding studies, the tight capacity constraints are generally addressed in the following two ways: (i) extend or modify existing link travel time functions to penalize the generalized cost values when the assigned flow is above the capacity (Lam et al., 1999; Nguyen et al., 2001; Cepeda et al., 2006; Szeto and Jiang, 2014a), (ii) explicitly consider strict capacity constraints through an inequality where the assigned flow is strictly equal to or less than the given capacity, implemented by

Table 2-1 Related Studies on User Equilibrium, Capacity Constraint, or Time-dependent Travel Demand in Transit Systems

Paper	Problem type	User equilibrium vs. System Optimal	Capacity constraint	Demand	Space-time network	Solution method
Nguyen and Pallottino (1988)	Transit assignment	Strategy-based UE represented by variational inequality	No	Static	No	Shortest hyperpath algorithm
Spiess and Florian (1989)	Transit assignment	Strategy-based UE represented by one convex cost differentiable problem	No	Static	No	Linear approximation based algorithm
De Cea and Fernandez (1993)	Transit assignment	Volume-delay function-based UE represented by variational inequality	No	Static	No	Diagonalization algorithm
Lam et al. (1999)	Transit assignment	Generalized Stochastic UE	Modified link cost function	Static	No	Lagrangian algorithm
Nguyen et al. (2001)	Transit assignment	Nash equilibrium represented by variational inequality	Modified link cost function	Time-dependent	No	Column generation scheme
Nuzzolo et al. (2001)	Transit assignment	Dynamic Stochastic user equilibrium	No	Responsive time-dependent	Yes	Iterative algorithm with network loading
Gao et al. (2004)	Transit network design	Volume-delay function-based UE represented by variational inequality	No	Static	No	Heuristic algorithm based on sensitivity analysis
Poon et al. (2004)	Transit assignment	Dynamic user equilibrium	Tight	Time-dependent	No	Iterative algorithm with network loading
Hamdouch et al. (2004)	Transit assignment	Strategy–based UE represented by variational inequality	Tight	Static	No	Iterative algorithm with network loading
Cepeda et al. (2006)	Transit assignment	Strategy–based UE represented by gap function	Modified link cost function	Static	No	Method of successive average (MSA)
Hamdouch and Lawphongpanich (2008)	Transit assignment	Strategy–based Dynamic UE represented by variational inequality	Tight	Time-dependent	Yes	Method of successive average (MSA) with network loading
Nuzzolo et al. (2012)	Transit assignment	Dynamic Stochastic user equilibrium	Tight	Responsive time-dependent	Yes	Iterative algorithm with network loading
Niu and Zhou (2013)	Transit service network design (Timetable optimization)	System-optimal	Tight	Time-dependent	Yes	Genetic algorithm
Hamdouch et al. (2014)	Transit assginment	Strategy–based Stochastic UE	Tight	Time-dependent	Yes	Iterative algorithm with network loading
Szeto and Jiang (2014a)	Transit assignment	Approach-based UE represented by variational inequality	Modified link cost function	Static	No	Extragradient method
Szeto and Jiang (2014b)	Transit network design	System-optimal, Bi-level model	Tight	Static	No	Hybrid artificial bee colony algorithm
Niu et al. (2015)	Transit service network design (Timetable optimization)	System-optimal	Tight	Time-dependent	Yes	Optimization solvers
Verbas and Mahmassani (2015)	Transit service network design (Frequency allocation)	Dynamic user equilibrium	Tight	Responsive time-dependent	No	Heuristic method and simulation via bi-level programming
Liu and Zhou (2016)	Transit service network design	Capacitated network equilibrium with boundedly rational agents	Tight	Time-dependent	Yes	Lagrangian decomposition

simulation-typed network loading for equilibrium condition (Poon et al., 2004; Hamdouch et al., 2004; Hamdouch and Lawphongpanich, 2008; Nuzzolo et al., 2012; Hamdouch et al., 2014; Verbas and Mahmassani, 2015) or analytically mathematic models for system optimum (Niu and Zhou, 2013; Szeto and Jiang, 2014b; Niu et al., 2015).

Further, to find one equilibrium condition in the schedule-based transit assignment problem with tight capacity constraints, most of previous studies adopted an iterative procedure with simulation-type network loading, where both the best path finding for dynamic travel demand and the network loading for path cost calculation are performed iteration by iteration until reaching convergence. Generally, there are three types of approaches to find the best paths, (i) the approach by Poon et al. (2004) finds the least-generalized cost path by the specific time-dependent optimal path algorithm (Tong and Richardson, 1984); (ii) Hamdouch and Lawphongpanich (2008) proposed the optimal strategy (a set of paths with least expected travel cost) for different passenger groups by dynamic programming; (iii) Nuzzolo et al. (2012) selected a set of paths with assignment probability by a route choice model while considering the real-time transit information. In this chapter, the path selection mechanism we adopt is based on the boundedly rational travel behavior that the path cost of each agent should be within the sum of the agent's least path cost and indifference band. Since non-atomic game and atomic game belong to two different modeling frameworks, in order to avoid the confusion with the traditional BRUE with nonatomic players, the agent-based transit assignment result in this chapter can be treated as a kind of network equilibrium with boundedly rational agents.

2.2.3 Wardrop's User Equilibrium and Bounded Rationality Behavior

A number of questions should be systematically addressed in order to analytically represent practically important modeling aspects, such as how to consider tight capacity constraints (e.g. vehicle carrying capacity and station spatial capacity) within UE-oriented behavioral assumptions. In an early study, Hearn (1980) stated that a real user equilibrium proposed by Wardrop may not exist in the traffic assignment problem with tight link capacity constraints. Adding the tight link capacity constraint to the Beckmann-McGuire-Winsten (BMW) model (Beckmann et al., 1956), Larsson and Patriksson (1995) proposed an augmented Lagrangean dual algorithm for a modified BMW model and showed that there exists a generalized UE where Lagrangean multipliers of link capacity constraints can be treated as link tolls. Further, Larsson and Patriksson (1995), Larsson and Patriksson (1998), Nie et al. (2004), and Marcotte et al. (2004) noted that those Lagrangian multipliers in generalized user equilibrium is not unique. In order to obtain a specific toll scheme (Lagrangian multipliers) for implementation, it usually requires modelers to consider a secondary goal (Larsson and Patriksson, 1998), such as, minimizing the total system tolling charges, minimizing the maximum toll on any individual arc, or minimizing the number of toll booths based on the tolling implementation conditions (Florian and Hearn, 1999). The other way to address tight capacity constraints is to use the marginal cost as link toll to achieve a generalized user equilibrium so that travelers can choose the routes (the solution) of system optimum (Florian and Hearn, 1999). In addition, the classical work by Correa et al. (2004) proves that the solution of the modified BMW model belongs to one of Nash equilibriums where no travelers can reduce their travel cost by unilaterally changing routes in a capacitated network.

Typically, once both the tight capacity and self-disutility minimization behavior (Wardrop's first principle) are considered, the problem can be modeled as a generalized Nash equilibrium problem (GNEP) (Facchinei and Kanzow, 2010). Each player has one disutility function and aims to choose a strategy that minimizes his/her disutility. Meanwhile, the player's strategy also depends on the rival players' strategies due to limited resources. However, it is still challenging to find a widely accepted solution approach to this complex problem, and more modeling details are systematically examined in the excellent survey paper by Facchinei and Kanzow (2010).

A theoretically important study by Szeto (2003) points out that dynamic user equilibrium (DUE) may not exist when considering the physical queue in dynamic traffic systems. Han et al. (2015a) first presented a rigorous continuity result for the path delay operator, which is the fundamental importance to the existence of DUE conditions, based on the Lighthill–Whitham–Richards (LWR) network model capable of capturing physical queues and spillback, by assuming that the network supply is bounded away from zero or is within desired boundedness. In addition, Han et al. (2015b) proposed a bi-level model for traffic network signal control problems where a continuum signal model is employed to ensure the existence of DUE in the lower level problem. As a remark, the dynamic user equilibrium studied in the last two decades is usually analytically expressed of Nash-like equilibrium condition (Han et al., 2015a).

In order to address the non-existence issue of flow-based dynamic user equilibrium, Szeto (2003) developed a tolerance-based dynamic traffic assignment (DTA) model, where the tolerance can also be treated as an indifference band, which can be viewed as an extension

or adaption from the boundedly rational behavioral model first introduced through the seminar work by Mahmassani and Chang (1987). In those metropolitan's transit system under oversaturated conditions, the tight vehicle capacity constraints lead to possible discontinuity of path travel time and unfortunately force some travelers fail to board on the preferred line, which could result in travelers' boundedly rational behavior based on their day-to-day (possibly stochastic) travel experiences. Meanwhile, many empirical studies using GPS trajectory data (e.g., Morikawa et al., 2005; Zhu, 2011) have shown that travelers do not always choose the shortest paths in reality

2.3 Vehicle Routing and Assignment for Household Daily Scheduled Activities

The activity-based modeling approach has been widely studied in the area of transportation planning and operations to better capture various facets of travel behavior and decision making. How to recognize complex resource constraints, multi-agent interactions, and consistency through trip chains of different individuals is an important concern for accurate activity-based modelling and analysis at the household level. The currently most adopted approach is to develop a probabilistic micro-simulation-based utility maximization models by Bhat et al. (2004), Pendyala et al. (2005), Pribyl and Goulias (2005), Miller and Roorda (2003), and Arentze and Timmermans (2004), mainly based on land-use, sociodemographic, activity system, and transportation level-of-service attributes.

Currently, the emerging mobile apps with multi-modal traveler information and personal activity schedules enable travelers to intelligently schedule their activities and share their trip requests. In addition, transportation network companies such as Uber and Lyft and the forthcoming autonomous vehicle system would allow and encourage a fully optimized

planning process for mapping household activities and travel requests (to be met by personal or shared vehicles) rather than just single on-demand request for optimally allocating available vehicle resources and minimizing the household-level or system-level total travel cost. Actually, the household activity pattern problem (HAPP) that is first systematically formulated by Recker (1995) fully satisfy the requirements above for the emerging and upcoming new transportation mobility pattern. The HAPP aims to find the optimal path of household members for completing their prescribed activities based on the available number of vehicles, scheduled activity participation, and ride-sharing options within a long period as the unit of analysis.

Typically, based on a conventional mixed integer linear programming model for the pickup and delivery problem with time windows (PDPTW), many typical cases in HAPP, e.g., five cases in a classical paper by Recker (1995), Recker (2001), Recker et al. (2001), and Gan and Recker (2008), require a very large number of linear and integer constraints to capture the complex rules in real-world household-level activity scheduling progress. Recently, several algorithms had been proposed to address more realistic side constraints and large-sized examples, to name a few, Chow and Recker (2012) and Kang and Recker (2013). In addition, Liao et al. (2013a, 2013b) presented a new set of super-network models for various person-level activity scheduling problems, where the multi-dimensional network construct contains travel links, state transition links and activity transaction links. To formulate HAPP as a mathematically rigorous model, how to fully consider complex coupling constraints among three layers, namely household members, vehicles and mandatory/optional activities, is extremely challenging, especially for large-scale multi-

modal transportation network with flexible ride-sharing and household member activity-coordination options

To consider the traffic congestion and feedback loops associated with complex trip interactions, there are a wide range of studies aiming to combine Activity-based Model (ABM) and Dynamic Traffic Assignment (DTA) to better capture the interplay between human activity-travel decisions and underlying congested networks with tight road capacity constraints, as shown in Figure 2-1. Iteration by iteration, the generated activities and time-dependent transportation condition will finally converge to a stable point or equilibrium, and then the individual activity-travel pattern will be formed and the DTA module will also output how those activities are finished and the resulting time-dependent transportation states.

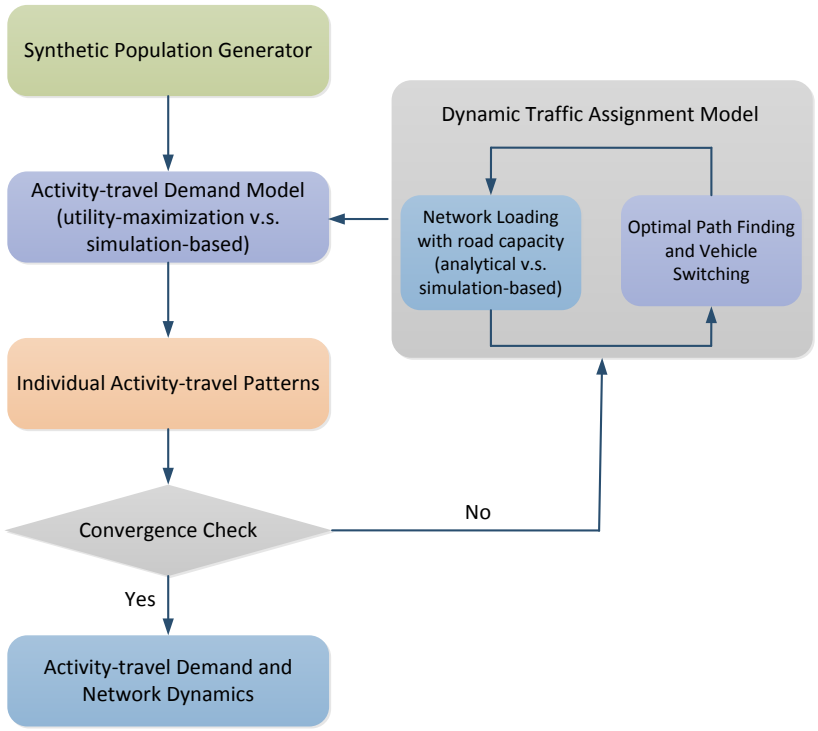


Figure 2-1 Existing Framework of Integration between ABM and DTA

For example, Lin et al. (2008) proposed a conceptual framework and explored the model integration of activity-based model (CEMDAP) and dynamic traffic assignment model (VISTA). Pendyala et al. (2012) further integrated activity-travel demand models (OpenAMOS), DTA tools with the long-term land use modeling layer (UrbanSim). To further study the impacts of dynamic traffic management strategies and real-time traveler information provision, Pendyala et al. (2017) proposed a tightly integrated modeling framework for representing activity-travel demand and traffic dynamics in an on-line environment.

Based on mathematical programs of HAPP, Kang et al. (2013) studied the network design problem considering the interaction between the household-level activity pattern and infrastructure changes. Chow and Djavadian (2015) proposed a new market equilibrium model to capture the interaction of traveler activity schedules in a capacitated system with a macroscopic flow restriction on a link or node facility. Abdul Aziz and Ukkusuri (2013) examined capacitated vehicle routing problems with the time-dependent congestion costs, which are determined by a network-wide cell transmission model. In a recent study by Fu et al. (2016), the intra-household interactions are considered through Markov decision processes and the road congestion effect is reflected by the static travel time function.

However, facing the upcoming autonomous vehicle (AV) application, it is possible that those AVs are owned by some private companies rather than just owned by each household assumed in the above, so it will be an important research topic to explore the new household activity pattern and the optimal planning for those household daily scheduled activities and AVs from some specific depots. The current research more focus on the vehicle routing

problem for individual on-demand travel request with pick-up and delivery services with time windows. A recent literature review can be found in the paper (Mahmoudi and Zhou, 2016) where a space-time-state network representation is able to comprehensively enumerate possible vehicles' carrying states at any given time along vehicle space–time paths, and further allows a forward dynamic programming solution algorithm to solve the single vehicle VRPPDTW problem in large-scale networks by Lagrangian relaxation.

In addition, the information-sharing technology is also changing the existing transit system. One example is the emerging customized bus services for regular commuters in congested metropolitan areas. A comprehensive literature review was conducted in the paper (Tong et al., 2017) where the paper developed a multi-commodity network flow-based model to optimize the utilization of the vehicle capacity while satisfying individual demand requests in a space-time network.

Actually, the research in this dissertation for the current public transit system and household-level activity completion optimization is just a beginning. In future, it is more possible to build a seamless multimodal scheduled transportation system, in which autonomous vehicles provide the optimal services for household scheduled daily mandatory and optional activities while connecting with the new public transit services, which can be fully observable, controllable and optimized based on more heterogeneous sources information.

CHAPTER 3 CAPACITATED TRANSIT SERVICE NETWORK DESIGN

3.1 Introduction

In general, the transportation network design problem aims to minimize the total transportation system disutility by optimizing the location of capacity enhancement strategies or various transportation service plans. Specifically, the aim of the transit service network design problem is to provide better service to users and to increase operating efficiency, involving key decisions such as constructing a new transit line, adding train or bus schedules, or dynamically selecting different types of vehicles to meet time-dependent transit demand.

While a large number of existing studies have been devoted to network design problems with static origin-destination (OD) demand input, this paper intends to study a class of practically important problems for designing discrete transit service networks with (i) individual traveler agents (corresponding to time-dependent demand matrices) and (ii) tight time-varying capacity constraints in terms of the number of passengers a transit vehicle or a station can carry. In addition, our study aims to address a number of theoretically challenging questions for realistically capturing and possibly affecting individual traveler's behavior in an oversaturated transit system. Under possibly extremely heavy congested conditions in transit systems (e.g. in Beijing and Tokyo), for example, each traveler wants to minimize his/her disutility within his/her rational/tolerance bands and preferred arrival times, but finally they need to select a close-to-user-optimal path from a limited number of capacity-feasible routing options, which could have very different path travel times and path-dependent prices.

3.2 Conceptual Illustration

Table 3-1 lists general indices, sets, parameters and variables in optimization models appeared in this sub-section.

Since Wardrop's user equilibrium may not exist due to strict capacity constraints, the gap of route travel time among travelers with same origin, destination, and departure time could force travelers to accept an indifference band through day-to-day travel experiences and finally form boundedly rational travel behavior (Liu and Zhou, 2016). For each agent, his/her perceived travel cost on the selected route is constrained within the respective indifference bands given by:

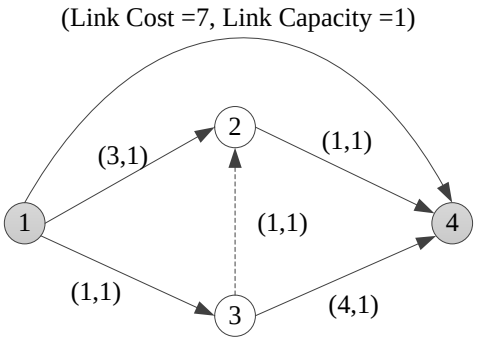
$$C_w^a \leq \pi_w + \varepsilon(a), \forall a \tag{3.1}$$

It should be noted that, the traditional flow-based BRUE condition is described by the following "if-then" condition: if $h_{w,p} > 0$, then $C_{w,p} \leq \pi_w + \varepsilon_w$ (Lou et al., 2010; Guo and Liu, 2011; Di et al., 2013; Di et al., 2014; Han et al., 2015c; Di et al., 2016). In our proposed case, for any path selected by one or more agents, the corresponding path flow $h_{w,p}$ is equal to or more than 1. As a result, the path cost should be constrained within the predefined bound $\pi_w + \varepsilon_w$ for an OD pair, or equivalently $\pi_w + \varepsilon(a)$ which reflects an individual's indifference band.

For illustrative purposes, one simple network is created that was adapted from the paper by Correa et al. (2004), with 4 nodes and 6 links along one OD pair $(1,4)$ shown in Figure 3-1. The link cost and link capacity are sequentially displayed in parenthesis. The demand from node 1 to node 4 has 2 agents.

Table 3-1 Indices, Sets, Parameters and Variables

Indices	Definition
i, j, j'	Index of nodes, $i, j, j' \in N$
(i,j)	Index of physical link between two adjacent nodes, $(i,j) \in L$
a	Index of agents, $a \in A$, defined based on each (time-dependent) OD pair
w	Index of OD pairs, $w \in W$
p	Index of path set P_w
k	Index of a path
$o(a)$	Index of origin node of agent a
$d(a)$	Index of destination node of agent a
t, s	Index of time intervals in the space-time network
Sets	
N	Set of nodes in the physical transportation network
L	Set of links in the physical transportation network
W	Set of Origin-Destination (OD) pairs
A	Set of agents
P_w	Set of paths of OD pair w
L_C	Set of current links in the physical transportation network
L_B	Set of potential built links in the physical transportation network
$\Phi(w(a), k)$	Set of links in the k^{th} path of OD pair $w(a)$ of agent a
$\Omega(w(a))$	Set of possible paths of OD pair $w(a)$ of agent a
V	Set of vertices in the space-time network
E	Set of edges/arcs in the space-time network
Parameters	
d^w	Travel demand of OD pair w
$\delta_{(i,j)}^{w,p}$	Path-link incidence index of route p of OD pair w on link (i, j)
$Cap_{i,j}$	Capacity of link (i,j)
$Cap_{i,j,t,s}$	Capacity of traveling arc (i,j,t,s) in the space-time network
$\varepsilon(a)$	The indifference band or tolerance value of agent a
ε_w	The indifference band of OD pair w
$b_{i,j}$	Construction cost of link (i,j)
B	Total financial budget
M	An assumed large value as an auxiliary parameter
DT^a	The departure time of agent a
AT^a	The assumed arrival time of agent a
$c_{i,j,t,s}$	Travel cost of traveling arc (i,j,t,s) in the space-time network
T	The time horizon in the space-time network
$BT_{w,DT}$	Budget time of all agents of OD pair w at departure time DT
$h_{w,p}$	Path flow of path p of OD pair w
Variables	
$c_{i,j}$	Travel cost of link (i,j)
$\bar{c}_{i,j}$	Generalized travel cost of link (i,j)
C_w^a	Path cost of agent a which departs from OD pair w
π_w	The minimal path travel cost of OD pair w
$C_{w(a),p}$	Path cost of path p of OD pair w of agent a
$x_{i,j}^a$	$= 1$ if agent a is assigned on link (i,j); $= 0$ otherwise
$f_{i,j}$	Traveler flow on link (i,j)
$y_{i,j}$	$= 1$ if physical link (i,j) is decided to be constructed in the physical transportation network; $=$ 0 otherwise
$x_{i,j,t,s}^a$	$= 1$, if Agent a is assigned on traveling/waiting arc (i,j,t,s) in the space-time network; $= 0$ otherwise
$y_{i,j,t,s}$	$= 1$ if service arc (i,j,t,s) is decided to be operated; $= 0$ otherwise



Path ID	Node Sequence	Path Cost	Path Trajectory
Path 1	1→2→4	4	
Path 2	1→3→4	5	
Path 3	1→4	7	
Path 4	1→3→2→4	3	

(link cost, link capacity)

Potential link to be built

Figure 3-1 One Simple Illustrative Network Modified from Correa et al. (2004)

The results are then examined under different traveling behavioral assumptions, including (i) generalized UE in the modified BMW model, (ii) system optimal, (iii) self-disutility minimization behavior (Wardrop's first principle), and (iv) CNEBRA for our network design problem.

(i) The modified BMW model:

$$min \sum_{(i,j)\in L} \int_0^{f_{i,j}} c_{i,j}(u)du \tag{3.2}$$

Subject to,

$$\sum_{p\in P_w} h_{w,p} = d_w, \forall w \in W \tag{3.3}$$

$$\sum_{w\in W} \sum_{p\in P_w} (\delta_{(i,j)}^{w,p} \times h_{w,p}) = f_{i,j}, \forall (i,j) \in L \tag{3.4}$$

$$f_{i,j} \le Cap_{i,j}, \forall (i,j) \in L \tag{3.5}$$

$$h_{w,p} \ge 0, \forall\, p \in P_w, \forall w \in W \tag{3.6}$$

The formulation is the standard BMW model with adding tight link capacity constraints. Eq. (3.3) is the path flow conversation constraint. In Eq. (3.4) the link flow is equal to the path flow multiplied by the path-link incidence value. Inequality (3.5) is the tight link capacity constraint, and inequality (3.6) defines the path flow as a nonnegative continuous variable. As mentioned in Section 2.2.3, the generalized UE (GUE) includes the Lagrangean multipliers $\mu_{i,j}$ of link capacity constraints (3.5), and the GUE condition is

derived from the KKT conditions shown in Appendix A. The generalized link cost can be represented as

$$\bar{c}_{i,j} = c_{i,j} + \mu_{i,j},, \forall (i,j) \in L \tag{3.7}$$

(ii) System optimal:

$$min \sum_{(i,j)\in L} (c_{i,j}(f_{i,j}) \times f_{i,j}) \tag{3.8}$$

Subject to constraints (3.3)-(3.6).

(iii) Self-disutility minimization behavior (Wardrop's first principle): since it is a generalized Nash equilibrium problem and the model is difficult to solve, the solution is enumerated in this simple example.

(iv) CNEBRA: the boundedly rational travel behavior has been formulated by inequality (1), and all possible solutions are also enumerated.

Table 3-2 lists comparison results for all cases.

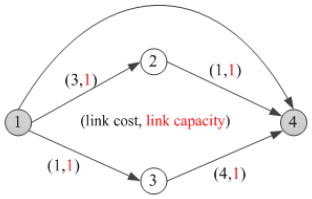
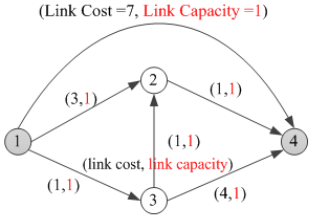
What we can observe from Table 3-2 is summarized as follows.

(1) Since the cost of all links is constant, the modified BMW model is the same as the system optimal model. In addition, no agent can reduce his/her travel cost by unilaterally changing routes, due to link capacity constraints in case 1 and case 2. This means that the two solutions belong to Nash equilibriums.

(2) In case 3, all travelers are assumed to be inclined to minimize their own disutility when selecting routes, which is consistent with Wardrop's first principle. Interestingly, one can identify the Braess paradox when building a new link. By enumeration, (i) if link $3 \rightarrow 2$ is not built, one agent will choose path 1 and the other will choose path 2, so the total system

cost is 9 (money units); (ii) if link 3 → 2 is built, one agent will choose the new shortest path, path 4, and the other has to accept the only one available path, path 3, due to tight link capacity constraints, so the total system cost is 10.

Table 3-2 Comparison Results of Different Cases for the Network Design Problem

Cases	(Link Cost =7, Link Capacity =1)  Without link 3 → 2						(Link Cost =7, Link Capacity =1)  With link 3 → 2						
	Path 1	Path 2	Path 3	Total Cost	π	Gap	Path 1	Path 2	Path 3	Path 4	Total Cost	π	Gap
1. Modified BMW	√	√	×	4 + 5 = 9	4	1	√	√	×	×	4 + 5 = 9	3	3
2. System Optimal	√	√	×	4 + 5 = 9	4	1	√	√	×	×	4 + 5 = 9	3	3
3. Self-disutility minimization	√	√	×	4 + 5 = 9	4	1	×	×	√	√	3 + 7 = 10	3	4
4. CNEBRA (0,0)	×	×	×	N/A	4	N/A	×	×	×	×	N/A	3	N/A
5. CNEBRA (0,1)	√	√	×	4 + 5 = 9	4	1	×	×	×	×	N/A	3	N/A
6. CNEBRA (0,4)	√	√	×	4 + 5 = 9	4	1	×	×	√	√	3 + 7 = 10	3	4
	√	×	√	4 + 7 = 11	4	3							
7. CNEBRA (2,4)	√	√	×	4 + 5 = 9	4	1	√	√	×	×	4 + 5 = 9	3	3
	√	×	√	4 + 7 = 11	4	3	×	×	√	√	3 + 7 = 10	3	4
	×	√	√	5 + 7 = 12	4	4	√	×	√	×	4 + 7 = 11	3	5
							×	√	√	×	5 + 7 = 12	3	6

CNEBRA (0,1): one agent's indifference band value is 0, and the other one's is 1;

√: One agent will choose the corresponding path; ×: No agent will choose the corresponding path;

N/A: Not applicable. There is no feasible solution.

(3) If link 3 → 2 is built, different solutions of cases 1, 2 and 3 satisfy the definition of Nash equilibrium, but the solution of case 3 is the only result obeying Wardrop's first principle in a capacitated network, that is, representing travelers' self-disutility minimization behavior with strict link capacity constraints. Further discussions along this line can be found in the paper by Correa et al. (2004). It is observed that the shortest path (path 4) is not used in cases 1 and 2, even the solution of the two cases meets the Nash equilibrium condition. It is consistent with the statement of Larsson and Patriksson (1995) about the result of generalized user equilibrium.

(4) The gap function we consider is $\sum_a (C_w^a - \pi_w)$. If link $3 \rightarrow 2$ is built, the gap for the first two cases is 3. With a larger gap value of 4, case 3 (applying Wardrop's first principle) does not lead to a smaller gap value in a capacitated network. The reason is that the gap function assumes that all travelers should be at the minimum cost paths after assignment (Hearn, 1982), but the tight capacity constraints cause some travelers have to accept a longer path in the final solution.

(5) Given the KKT condition (described in Appendix A) in case 1 with link $3 \rightarrow 2$, the Lagrange multipliers of links $(3,2)$ and link $(1,4)$ are 0 because the link flow is less than its link capacity. For the remaining links, those multipliers are not unique as long as the generalized path travel cost of path 1 and path 2 is equal and not more than that of the longest path (path 3). To some extent, this behavior is related to the argument of Nie et al. (2004) and Marcotte et al. (2004) that it may be questionable to treat those multipliers as queuing delays because of the non-uniqueness of the multiplier values.

(6) Now consider different cases for CNEBRA with specific indifference bands for two individuals. In case 4, when the indifference band of two agents is 0, it is not surprising to see there is no solution for agents without tolerance. For a particular example where link $3 \rightarrow 2$ is not built, the specific set of inequalities for boundedly rational behavior can be listed as follows:

For agent 1: $C^1 \leq 4 + 0$

For agent 2: $C^2 \leq 4 + 0$.

As shown in Figure.3-1, the path cost of the shortest path (path 1) is 4, but its capacity is only 1, so it is impossible for two agents to use one feasible path at the same time.

(7) In case 5, with the increased indifference band of 1 for the second agent, the feasible solution exists for the do-nothing case; but no CNEBRA solution can be found if link $3 \rightarrow 2$ is built. In reality, when facing the change of network conditions, travelers may need to change their own indifference band in order to adapt to the new situations. Interested readers are referred to a recent paper by Lo (2013), which interprets similar behavior by highlighting that bounds on rationality are determined by physiological and environmental constraints. However, how to quantify the changing bounds on rationality is still a very complex research topic in its own right and beyond the scope of this paper.

(8) In case 6, (i) if link $3 \rightarrow 2$ is not built, there are two feasible solutions with different system cost values of 9 and 11, which can be viewed as the best case vs worst case of CNEBRA solutions. (ii) if link $3 \rightarrow 2$ is built, there is only one feasible solution and the total cost is 10. As a result, if we consider the best case of CNEBRA as the selected solution for the before and after scenarios, it would lead to the Braess paradox. However, the Braess paradox can be avoided in this particular case if the worst case of CNEBRA is assumed. Therefore, selection of feasible solution(s) of CNEBRA would affect the final network design decision, as are also clearly identified by a recent important paper by Lou et al. (2010). Additionally, in case 7, there are multiple solutions whether or not link $3 \rightarrow 2$ is built, so it is difficult to select a single solution from the before or after solution set to support the network design decision.

3.3 Modelling on a Discrete Network Design Problem with Boundedly Rational Agents

In order to clearly illustrate the forthcoming dynamic transit service network design problem in Section 3.4, this sub-section will focus on the static case as a starting point.

Consider a transportation network as a directed graph with N as the set of nodes and L as the set of links. Let W denote the set of OD pairs connected by the set of feasible paths, P_w. Each agent a is defined based on its origin $o(a)$ and destination $d(a)$ and has its own tolerance value $\varepsilon(a)$. Each link (i,j) has its travel cost $c_{i,j}$ and capacity $Cap_{i,j}$. In order to improve the total transportation system efficiency, a total financial budget B is planned and the construction cost of new line (i,j) is $b_{i,j}$.

The static discrete transportation network design with bounded-rational agents can be formulated as follows:

Objective function:

$$\min \sum_a \sum_{(i,j)\in L} (c_{i,j} \times x_{i,j}^a) \tag{3.9}$$

Subject to:

Budget constraint:

$$\sum_{(i,j)\in L} b_{i,j} \times y_{i,j} \leq B, \forall (i,j) \in L, b_{i,j} = 0 \; for \; (i,j) \in L_C \tag{3.10}$$

Capacity constraint:

$$\sum_a x_{i,j}^a \leq y_{i,j} \times Cap_{i,j}, \; \forall (i,j) \in L, y_{i,j} = 1 \; for \; (i,j) \in L_C \tag{3.11}$$

Flow balance constraint:

$$\sum_{i:(i,j)\in L} x_{i,j}^a - \sum_{i:(j,i)\in L} x_{j,i}^a = \begin{cases} -1 & \forall a, j = o(a) \\ 1 & \forall a, j = d(a) \\ 0 & otherwise \end{cases} \tag{3.12}$$

Boundedly rational travel decision rule:

$$\sum_{(i,j)\in L}\{c_{i,j} \times x_{i,j}^{a}\} \le \pi_{w(a)} + \varepsilon(a), \forall a \tag{3.13}$$

Definition of the shortest path:

$$\pi_{w(a)} = \min\{C_{w(a),1}, C_{w(a),2}, \dots, C_{w(a),k}\}, \forall k \in \Omega(w(a)) \tag{3.14}$$

Binary variables: $x_{i,j}^{a} = \{0,1\}$ and $y_{i,j} = \{0,1\}$

The objective function of this network design problem is to minimize the total transportation system travel cost of all agents. Inequality (3.10) represents the budget constraint, where the construction cost of existing links is 0. Inequality (3.11) is the link capacity constraint, where $y_{i,j} = 1$ for existing links. Eq. (3.12) is the standard agent-based flow balance constraint. Inequality (3.13) and Eq. (3.14) represent boundedly rational travel decision rule and the shortest path discussed above, respectively. They can be combined into the following set of inequalities (3.15):

$$\begin{cases} \sum_{(i,j)\in L}\{c_{i,j} \times x_{i,j}^{a}\} \le C_{w(a),1} + \varepsilon(a) \\ \sum_{(i,j)\in L}\{c_{i,j} \times x_{i,j}^{a}\} \le C_{w(a),2} + \varepsilon(a) \\ \dots \\ \sum_{(i,j)\in L}\{c_{i,j} \times x_{i,j}^{a}\} \le C_{w(a),k} + \varepsilon(a) \end{cases}, \forall a \tag{3.15}$$

Meanwhile, the path cost of feasible path k of OD pair w_a can be formulated as Eq. (3.16), where an auxiliary large value M is introduced to represent link cost as $c_{i,j} + (1 - y_{i,j}) \times M$. If link (i,j) is built, its cost is $c_{i,j}$; otherwise, its large cost value prevents any agent from selecting that link.

$$C_{w(a),k} = \sum_{(i,j)\in\Phi(w_a,k)}\{c_{i,j} + (1 - y_{i,j}) \times M\} \tag{3.16}$$

Therefore, the boundedly rational travel decision rule can be reformulated as follows:

$$\sum_{(i,j)\in L}\{c_{i,j} \times x_{i,j}^{a}\} \le \sum_{(i,j)\in\Phi(w_a,k)}\{c_{i,j} + (1 - y_{i,j}) \times M\} + \varepsilon(a), \forall a, \forall k \in \Omega(w(a)) \tag{3.17}$$

A simple study case is created for illustrating the transformation above in Figure 3-2. One agent (agent a) departs from origin node 1 to destination 5 with indifference band of 1 time unit. Link $(3,5)$ and link $(4,5)$ are potentially built.

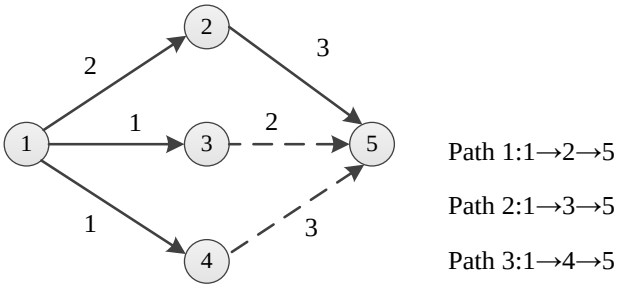


Figure 3-2 Simple Study Case for the Illustration Above

The boundedly rational travel decision rule can be represented as

$$\sum_{(i,j)\in L}\{c_{i,j} \times x_{i,j}^{a}\} \le C_{w(a),1} + \varepsilon(a) = 5 + 1 \quad (3.17.a)$$

$$\sum_{(i,j)\in L}\{c_{i,j} \times x_{i,j}^{a}\} \le C_{w(a),2} + \varepsilon(a) = 1 + 2 + (1 - y_{3,5}) \times M + 1 \quad (3.17.b)$$

$$\sum_{(i,j)\in L}\{c_{i,j} \times x_{i,j}^{a}\} \le C_{w(a),3} + \varepsilon(a) = 1 + 3 + (1 - y_{4,5}) \times M + 1 \quad (3.17.c)$$

The possible chosen path set of agent a is listed in Table 3-3. The different assumed network design decisions could lead to different shortest paths for one specific OD pair, so the tightest and active constraint of boundedly rational travel decision rule could be different, which bounds the feasible path set to each agent.

Table 3-3 Possible Chosen Path Set of Agent *a* under Different Network Design Decisions

Network design decision	The least path cost π	Rational bound $\pi + \varepsilon(a)$	The tightest constraint of boundedly rational travel decision rule	Feasible path set of agent a
$y_{3,5} = 0, y_{4,5} = 0$	5	6	Inequality (17.a)	Path 1
$y_{3,5} = 1, y_{4,5} = 0$	3	4	Inequality (17.b)	Path 2
$y_{3,5} = 0, y_{4,5} = 1$	4	5	Inequality (17.c)	Path 3 and path 1
$y_{3,5} = 1, y_{4,5} = 1$	3	4	Inequality (17.b)	Path 2 and path 3

As a result, the critical variable π_w for the boundedly rational travel decision rule does not appear in the model and it is implicitly defined through a set of path cost inequalities (3.17). This requires to exogenously enumerate a number of paths for the OD pair of each agent, including the shortest path in the existing network and those paths that have potential built links and are shorter that the shortest path in the existing network. Typical route set generation algorithms can be found in the dissertation by Ramming (2002) for large-scale networks. For simplicity, the K-shortest path algorithm (Yen, 1971; Xu, et al., 2012) can also be adopted to seek the potential shortest path π_w for each OD pair.

In the typical bi-level programming structure, the lower level UE problem only considers the links to be built decided from the upper level problem (and existing links). In our proposed single-level model, the boundedly rational travel decision rule is represented by considering the complete set of all possible paths that embed the decision variable $y_{i,j}$ covering two cases: to be built or not to be built in the final optimal solution. More specifically, if $y_{i,j} = 0$ for the link not to be built, the right-hand-side path cost is infinity, allowing inequality (3.17) to hold in any case. If $y_{i,j} = 1$ for the link to be built in the final solution, (i) capacity inequality (3.11) allows the corresponding link flow to be positive; (ii) a feasible path in the final solution will have a path cost $C_{w,p} < \infty$. Furthermore, if there is a feasible solution for all agents' boundedly rational behavior with $\pi_w + \varepsilon(a) < \infty$, then each agent must be able to select a capacity-feasible path p from the path set $\Omega(w(a))$ satisfying $C_{w,p} \leq \pi_w + \varepsilon(a) < \infty$.

The non-uniqueness of traditional BRUE solutions has been studied or discussed in several papers (Lou et al., 2010; Guo and Liu, 2011; Di et al., 2013; Di et al., 2014; Han et al.,

2015c; Di et al., 2016). More specifically, most researchers focus on the best case (risk-averse), the worst case (risk-prone), and the neutral case (risk-neutral) for the unique solution of BRUE. In this chapter, since the objective function of our proposed single-level programming model (presented herein), $min_{y,x(y)} f(y,x)$, is to minimize the total travel time, it indicates that the best case of the CNEBRA solution is assumed as travelers' responses. Usually, $min_y min_{x(y)} f(y,x)$ is one representation for the objective function of network design problems with the best case of traditional BRUE conditions (Di et al., 2016) from the bi-level programming perspective. In order to handle a very complex relationship between x and y represented through the lower level optimization problem in the bi-level formulation, the most straightforward way to find the optimal solution is to enumerate all feasible y, solve the corresponding lower level model $min_{x(y)} f(y,x)$, and then compare all $min_{x(y)} f(y,x)$ to find the minimum of $min_{x(y)} f(y,x)$, which is $min_y min_{x(y)} f(y,x)$. It is important to note that, within the same solution space of x and y, in our proposed single-level programming model, the relation among y and x has been clearly defined in constraints, and the objective function $min_{y,x(y)} f(y,x)$ can also be viewed to find the minimum of $min_{x(y)} f(y,x)$ based on all possible y. Therefore, we can state that only the best case is considered in our model. In the future, we will conduct further studies to examine different assumptions about travelers' boundedly rational decision behavior in our agent-based framework.

Typically, the discussions on the existence and uniqueness of traditional flow-BRUE are built on nonlinear and convex programming techniques, while the proposed CNEBRA formulation in this chapter is an integer linear programming model in nature. In many cases,

there might be multiple path solutions even for the simplest single OD pair shortest path problem in a grid network. Thus, the properties (solution existence and uniqueness) of our proposed model will be dependent on one specific integer linear programming model on different time-discretized networks, and the further discussions of those solution properties are beyond the scope of this paper.

3.4 Dynamic Discrete Transit Service Network Design with Boundedly Rational Agents

In order to represent (i) the transit service network based on the given train/bus schedule, (ii) transit vehicle capacity, and (iii) transit station/platform storage capacity, we extend the physical transit network into a space-time network from the given timetable. Consider a physical transit network with a set of nodes (stations) N and a set of links L. Each link can be denoted as a directed link (i, j) from upstream node i to downstream node j, with one deterministic scheduled travel time. We then construct a space-time network, where V is the set of vertices and E is the set of edges/arcs. Node i is extended to a set of vertices (i, t) at each time interval t in the study horizon, $t = 1,2, \dots, T$. Each agent a, where $a \in A$, is assumed to have a planned departure time DT^a at origin node $o(a)$ to its destination node $d(a)$. At each destination node, there is one assumed large arrival time T for all agents. Meanwhile, set the travel cost of waiting arcs on the destination node as 0 in the space-time network to represent the end-to-end trip time. There are two following types of arcs.

(1) Link traveling arcs are extended from a link (i, j) and each arc traverses from vertex (i, t) to vertex (j, s) based on the given timetable, where $(\mathrm{s} - \mathrm{t})$ is the scheduled link

travel time and should be integer multipliers of one time interval and capacity is defined as the transit vehicle capacity.

(2) Waiting arcs from (i, t) to $(i, t + 1)$ at any node i have waiting time as one time interval and capacity is defined as the station/platform storage capacity. The waiting cost is set as 0 for the destination stations.

In addition, to address the common line issue in transit systems that there are multiple same schedules (travelling arcs) directly joining two vertexes in the space-time network, we can separate one vertex as multiple vertexes, each of which corresponds to one transit schedule. That method is similar to the approach used by Poon et al. (2004) and Hamdouch and Lawphongpanich (2008).

In order to find what impacts travelers' route choice most, one very interesting travel survey was performed in the Chicago Metropolitan area in 2010 (Nie et al., 2010), which reveals that over 80% of the responders choose travel time as their most important concern in their route choice. The remaining factors sorted by importance are reliability, cost, comfort and convenience, safety, and emission and energy conversation, respectively. Therefore, our first modeling priority regarding the general travel cost is still on travel time, including in-vehicle travel time, at-station/stop waiting time and transfer time. Nuzzolo et al. (2001) and Nuzzolo et al. (2012) considered a flow-based path travel cost composed of travel time (in-vehicle time, waiting time and transfer time), number of transfers, and in-vehicle discomfort cost. Hamdouch and Lawphongpanich (2008) considered travel time, transit fare, penalty for early/late arrival and early departure, and in-vehicle discomfort cost. In those papers, discomfort cost is represented by a flow-based link cost function, and all

other factors are weighted by different given parameters. With our proposed space-time network structure, one can further incorporate various weights of in-vehicle travel time, waiting time, and arc-based transit fare as constant costs of different types of arcs. The modelling issue related to OD-based dynamic transit fare and penalty for early/late arrival, as part of general path cost, will be discussed in sub-sections 3.6.1 and 3.6.3, respectively.

For illustrative purposes, Figure 3-3(a) depicts a simple transit physical network with three nodes and two links. The demand has three agents departing at time 0 from node 1 to node 3, whose tolerance values/indifference bands are 0, 2, and 2, respectively. The station capacity at each node is assumed to be 2. A potential open new line is from node 1 to node 3. Table 3-4 lists the vehicle capacity parameters and scheduled travel time, corresponding to the space-time network presented in Figure 3-3(b).

Table 3-4 Vehicle Capacity and Its Scheduled Travel Time

Transit line with its vehicle		Departure Time	Vehicle Capacity	Schedule Travel Time		
				Link (1,2)	Link (2,3)	Link (1,3)
Existing Line($1 \rightarrow 2 \rightarrow 3$)	Vehicle 1	0	1	1	2	×
	Vehicle 2	2	2	1	2	×
Potential Line($1 \rightarrow 3$)	Vehicle 3	1	1	×	×	3

In the space-time network, the origin becomes the vertex $(1,0)$ and the destination is the vertex $(3,5)$, which is a single origin to single destination problem (one-to-one network).

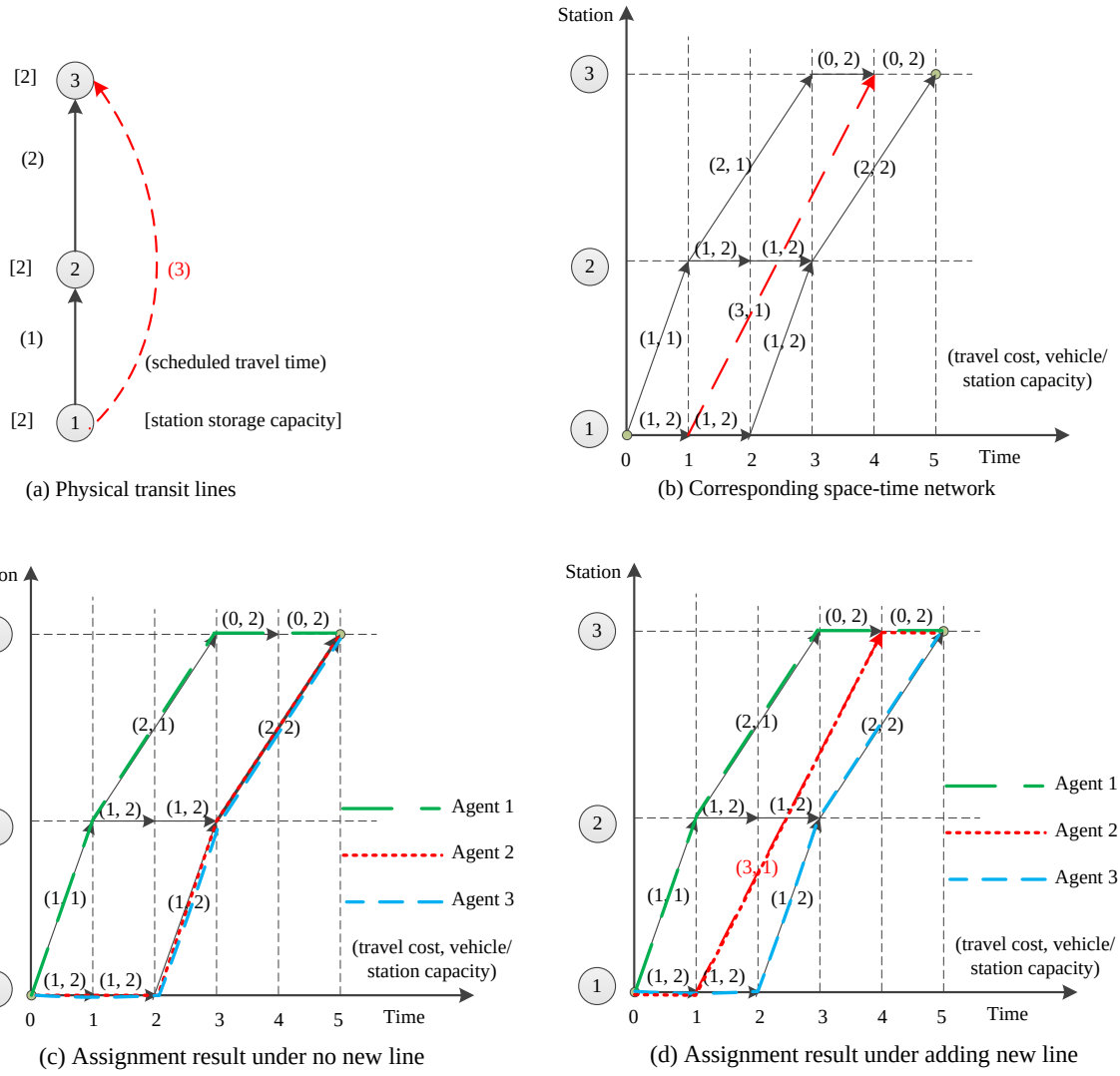


Figure 3-3 Illustration of Physical Transit Network, Corresponding Space-time Network and Assignment Results

Based on the tolerance values of each agent and boundedly rational travel decision rule, $C_w^a \leq \pi_w + \varepsilon(a)$, the set of inequalities for each agent can be written as: $\pi_w = 3$, $\varepsilon(1) = 0$, $\varepsilon(2) = 2$, $\varepsilon(3) = 2$, so $C^1 \leq 3 + 0$, $C^2 \leq 3 + 2$, $C^3 \leq 3 + 2$. If new transit line (1,3) is not built, the single solution satisfying the set of inequalities is shown in Figure 3-3(c). Agent 1 will choose path $(1,0) \rightarrow (2,1) \rightarrow (3,3)$. Due to the vehicle capacity constraint,

both agent 2 and agent 3 have to wait at station 1 for the next available vehicle at time 2, so the selected path will be $(1,0) \rightarrow (1,1) \rightarrow (1,2) \rightarrow (2,3) \rightarrow (3,5)$ with a total system-wide travel time of 13. After the new transit line is open, the best case solution is that agents 1 and 3 still choose the previous path and agent 2 will choose path $(1,0) \rightarrow (1,1) \rightarrow (3,4) \rightarrow (3,5)$, shown in Figure 3-3(d) with the total system-wide travel time as 12. Table 3-5 provides a more systematic comparison between different scenarios and it should be noted that the agent number is only used for representing one traveler. It doesn't mean that agent 1 has the priority to choose a shorter path compared with other agents.

Table 3-5 Assignment result comparison of scenarios with and without adding new line

Scenarios	Agent ID	Agent 1	Agent 2	Agent 3
	Tolerance band	0	2	2
No new line	Inequality of boundedly rational travel decision rule	$C^1 \leq 3+0$	$C^2 \leq 3+2$	$C^3 \leq 3+2$
	Path node sequence	(1,0)→(2,1) →(3,3)	(1,0)→(1,1) →(1,2) →(2,3) →(3,5)	(1,0)→(1,1) →(1,2) →(2,3) →(3,5)
	Total cost		13	
Adding new line	Inequality of boundedly rational travel decision rule	$C^1 \leq 3+0$	$C^2 \leq 3+2$	$C^3 \leq 3+2$
	Path node sequence	(1,0)→(2,1) →(3,3)	(1,0)→(1,1) →(3,4) →(3,5)	(1,0)→(1,1) →(1,2) →(2,3) →(3,5)
	Total cost		12	

Prior to introducing the proposed model, the following key assumptions are presented as careful response to many potential modeling issues discussed in the aforementioned illustrative examples.

(1) The travel cost of each agent just considers the travel time on travelling arcs and waiting arcs on the extended space-time network.

(2) The time-dependent transit demand is given and deterministic.

(3) The tolerance value of each agent is given and does not change with respect to changes in the transit service network.

(4) Regarding the possible multiple solutions of CNEBRA, the best case is chosen as the selected solution.

(5) For simplicity, the first-in-first-out (FIFO) rule is not considered in these dynamic transit systems, because the non-FIFO phenomenon still exists under certain conditions. For example, when travelers are waiting at a platform for the next desirable transit vehicle, it may not actually be the next available vehicle. Also, when travelers transfer to another station or board a transit vehicle, it is possible for them not to entirely obey the FIFO rule in reality.

The following presents the objective function and constraints applied in the proposed model.

Objective function:

$$\min \sum_a \sum_{(i,j,t,s)\in E} (c_{i,j,t,s} \times x^a_{i,j,t,s}) \tag{3.18}$$

Subject to,

Budget constraint:

$$\sum_{(i,j)\in L} b_{i,j} \times y_{i,j} \leq B, \forall (i,j) \in L, b_{i,j} = 0 \; for \; (i,j) \in L_C \tag{3.19}$$

Capacity constraint:

$$\sum_a x^a_{i,j,t,s} \leq y_{i,j} \times Cap_{i,j,t,s}, \; \forall (i,j) \in L, y_{i,j} = 1 \; for \; (i,j) \in L_C \tag{3.20}$$

Flow balance constraint:

$$\sum_{i,t:(i,j,t,s)\in E} x^a_{i,j,t,s} - \sum_{i,t:(j,i,s,t)\in E} x^a_{j,i,s,t} = \begin{cases} -1 & \forall a, j = o(a), s = DT^a \\ 1 & \forall a, j = d(a), s = T \\ 0 & otherwise \end{cases} \quad (3.21)$$

Boundedly rational travel decision rule:

$$\sum_{(i,j,t,s)\in E}\{c_{i,j,t,s} \times x^a_{i,j,t,s}\} \le \sum_{(i,j,t,s)\in \Phi(w_a,k)}\{c_{i,j,t,s} + (1 - y_{i,j}) \times M\} + \varepsilon(a), \forall a, \forall k \in \Omega(w(a)) \quad (3.22)$$

Binary variables:

$$x^a_{i,j,t,s} = \{0,1\} \quad (3.23)$$

$$y_{i,j} = \{0,1\} \quad (3.24)$$

Similar to the objective function and side constraints of the static case presented in Section 3.3, this formulation is extended for the dynamic case in the space-time network based on the given schedule of existing transit lines and potential built or open service lines. Each agent chooses a set of arcs in the space-time network, so the decision variable $x^a_{i,j,t,s}$ is a binary variable. We need to recognize that once those binary variables on each arc are required to be aggregated for flow-based arc/link attributes, such as, for in-vehicle discomfort cost function, it will be a challenge to solve this problem. This is because the objective function (3.18) becomes a nonlinear function with binary variables. In addition, $y_{i,j}$ in constraints (3.19), (3.20) and (3.24) could also be extended as $y_{i,j,t,s}$ when considering whether or not to add or close specific service arcs defined by the time-dependent schedule in the space-time network. In short, after problem decomposition in Section 3.5, the unique feature of our agent-based formulation with constant arc costs allows us to handle much simpler time-dependent shortest path subproblems, compared to general multi-commodity flow-based formulations.

3.5 Lagrangian Decomposition Based Solution Procedure

3.5.1 Problem Decomposition

Within a commonly used Lagrangian relaxation framework, a set of nonnegative capacity constraint multipliers $\mu_{i,j,t,s}$ and tolerance bound constraint multipliers λ_k^a are defined to dualize capacity constraints (3.20) and boundedly rational travel decision rule (3.22), respectively, onto the objective function (3.18) presented in Section 3.4 to generate lower bounds. The objective function is now transformed to

$$\min \sum_a \sum_{(i,j,t,s)\in E}\left(c_{i,j,t,s} \times x_{i,j,t,s}^a\right) + \sum_{(i,j,t,s)\in E}[\mu_{i,j,t,s} \times \left(\sum_a x_{i,j,t,s}^a - y_{i,j} \times Cap_{i,j,t,s}\right)] + \sum_a \sum_k \lambda_k^a \times \{\sum_{(i,j,t,s)\in E}\left(c_{i,j,t,s} \times x_{i,j,t,s}^a\right) - \sum_{(i,j,t,s)\in \Phi(w_a,k)}\left[c_{i,j,t,s} + \left(1 - y_{i,j}\right) \times M\right] - \varepsilon(a)\} \quad (3.25)$$

Subject to constraints (3.19), (3.21), (3.23), and (3.24).

Based on the decision variables $x_{i,j,t,s}^a$ and $y_{i,j}$, the dualized problem above can be decomposed into two sub-problems P_x and P_y.

Subproblem P_x for finding time-dependent shortest path for each agent:

$$\min \sum_a \sum_{(i,j,t,s)\in E}(c_{i,j,t,s} + \mu_{i,j,t,s} + \sum_k \lambda_a^k \times c_{i,j,t,s}) \times x_{i,j,t,s}^a \quad (3.26)$$

Subject to (3.21) and (3.23).

Subproblem P_y as a binary knapsack problem for service arc selections:

$$\max \sum_{(i,j,t,s)\in E} \mu_{i,j,t,s} \times y_{i,j} \times Cap_{i,j,t,s} + \sum_a \sum_k \lambda_a^k \times \{\sum_{(i,j,t,s)\in \Phi(w_a,k)}\left[c_{i,j,t,s} + \left(1 - y_{i,j}\right) \times M\right] + \varepsilon(a)\} \quad (3.27)$$

Subject to (3.19) and (3.24).

3.5.2 Lagrangian Relaxation Based Algorithm Design

The general procedure of the Lagrangian relaxation-based algorithm is designed as shown in Figure 3-4, and one can apply dynamic transit simulation to find upper bound feasible solutions based on the network design decisions from the lower bound model (i.e. the relaxed problem).

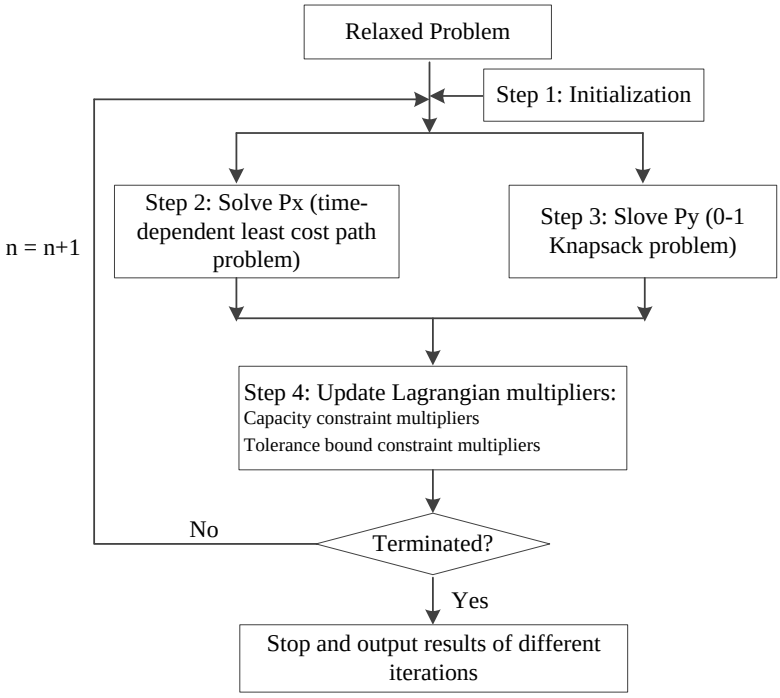


Figure 3-4 Flowchart of Solving the Relaxed Problem under Lagrangian Relaxation Framework

The specific procedure is described as follows:

Step 1: Initialization

Initialize iteration number $n = 0$;

Construct the space-time network based on the given transit physical network and time schedule;

Initialize the set of Lagrangian multipliers $\mu_{i,j,t,s}$ and λ_k^a as positive values.

Step 2: Solve subproblem P_x for $x_{i,j,t,s}^a$.

Step 3: Solve subproblem P_y for $y_{i,j}$.

Step 4: Update Lagrangian multipliers:

4.1 Calculate the subgradients:

Subgradient of capacity constraint: $\nabla\mu_{i,j,t,s} = \sum_a x_{i,j,t,s}^a - y_{i,j} \times Cap_{i,j,t,s}$

Subgradient of tolerance bound constraint: $\nabla\lambda_k^a = \sum_{(i,j,t,s)\in E}\left(c_{i,j,t,s} \times x_{i,j,t,s}^a\right) - \sum_{(i,j,t,s)\in\Phi(w_a,k)}\left[c_{i,j,t,s} + \left(1 - y_{i,j}\right) \times M\right] - \varepsilon(a)$

4.2 Update Lagrangian multipliers:

Multiplier of capacity constraint: $\mu_{i,j,t,s}^{n+1} = \max\{0, \mu_{i,j,t,s}^n + \alpha^n \times \nabla\mu_{i,j,t,s}\}$

Multiplier of tolerance bound constraint: $\lambda_k^a(n+1) = \max\{0, \lambda_k^a(n) + \alpha^n \times \nabla\lambda_k^a\}$ Where, α^n is the step size, and $\alpha^n = 1/(n+1)$.

Step 5: Termination condition test

If n is equal to the predetermined maximum iteration number N, terminate the algorithm; otherwise, $n = n + 1$ and go back to Step 2 with updated $\mu_{i,j,t,s}$ and λ_k^a values.

At step 2, all arcs have constant travel cost in the space-time network, and the Lagrangian multipliers are given as constant values at each iteration. The subproblem P_x becomes a standard time-dependent least cost path problem. At step 3, the arc travel cost, arc capacity, Lagrangian multipliers, big M, and tolerance value of each agent are given and constant. The subproblem P_y becomes a standard 0-1 knapsack problem. Ziliaskopoulos and Mahmassani (1993) and Pallottino and Scutellà (1998) provided more details about the space-time network construction and time-dependent least cost path finding algorithms, which does not require the addition of a dummy node for the corresponding destination node in many-origin-to-many-destination networks. For the 0-1 knapsack problem, there are several available computationally efficient algorithms based on dynamic programming, branch and bound, or the hybridizations of the both (Martello and Toth, 1990).

3.6 Discussions

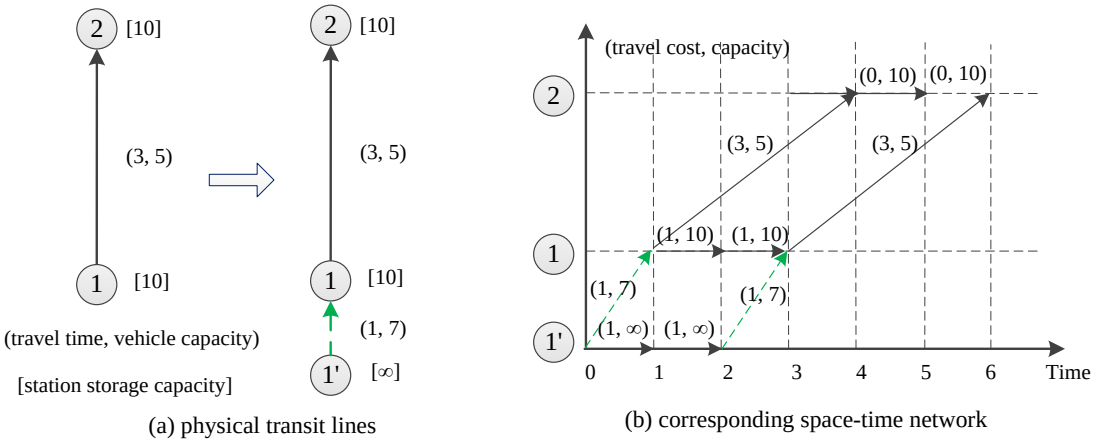
3.6.1 Transit Pricing

Transit pricing is one important factor affecting final decisions on the transit service network design. To consider time-dependent and personalized agent-based or OD based (given) pricing parameters, one can extend the space time arc cost $c_{i,j,t,s}$ to $c^{a}_{i,j,t,s}$, and add dummy starting arcs to represent OD-specific or agent-specific price for related travel distance or/and traveling time periods. That is, for each OD or agent, there is one dummy link with a specific cost value. Considering the price or incentive as part of the decision variables in the network design problem, different pricing/incentive strategies can be listed using different dummy arcs. Then add a restriction that only one of the dummy arc sets for

each time-dependent OD pair or agents can be selected for the final strategy to be used in the final solution.

3.6.2 Passenger Inflow Rate Control to Transit Stations

One approach for addressing the safety concern in a totally oversaturated condition is to limit the number of passengers in the station. This practice was implemented in actual subway operations during peak hours in Beijing (Xu et al., 2014). To model this inflow passenger volume gating strategy, a virtual node needs to be added to the related origin node in the physical network. A simple network with two nodes and one link is used to illustrate our method as shown in Figure 3-5(a).



(a) physical transit lines

(b) corresponding space-time network

Figure 3-5 Original and Modified Space-time Networks for Modeling

There are two vehicles with capacity of 5 and scheduled travel time of 3, departing at time 1 and 3, respectively. The station/platform capacity of node 1 and 2 is 10. Now assume that the permitted inflow rate is 7. The change in the physical network is shown in Figure 3-5(a). Link $(1', 1)$ is treated as the entry point to the station and its capacity represents the permitted passenger inflow rate. The corresponding space-time network is shown in Figure

3-5(b). If the number of agents at node $1'$ at time 0 is more than 7 at first, some agents will need to wait at node $1'$ until time 2 to enter the station for the second vehicle.

3.6.3 Model the Penalty for Early/Late Arrival

In addition to the indifferent/tolerance value, each traveler may have one specific preferred arrival time for his or her trip within the bound. The penalty for early and later arrival had been considered in some traffic and transit studies (such as, Zhou et al., 2008; Hamdouch and Lawphongpanich, 2008). In this sub-section, the issue is considered through updating the space-time network, where each destination node will be correspondingly given a virtual node as a super destination. The travelling arcs from the real destination node to the virtual node can be viewed as penalty arcs, the cost of which is defined in advance for each agent as part of his or her own path travel cost.

The network without walk arcs in the paper (Hamdouch and Lawphongpanich, 2008) is chosen as our illustrative example. Based on the given schedule of three lines, the corresponding space-time network is built in Fig. 3-6. Assume that agent 1 departs from node 1 to node 4 at time 1 with preferred arrival time of 5 and agent 2 has the same OD pair of agent 1 with different departure time and preferred arrival time, which are time 3 and time 7, respectively. The penalties $c^a_{i,j,t,s}$ for the early and late arrival of agent 1 is predefined as $c^1_{4,4',4,11} = 1$, $c^1_{4,4',5,11} = 0$, $c^1_{4,4',7,11} = 4$, $c^1_{4,4',9,11} = 8$ and $c^1_{4,4',10,11} = 10$. Also, the early or late arrival penalties of agent 2 can be defined in advance in the space-time network. As a result, the arc cost in the space-time network should be represented by $c^a_{i,j,t,s}$ instead of $c_{i,j,t,s}$ in the objective function and the boundedly rational travel decision rule. Except for those penalty arcs, the remaining arc cost of each agent $c^a_{i,j,t,s}$ is still equal

to $c_{i,j,t,s}$. Also, it is applicable by changing every arc cost if users impose different weights or parameters on different types of arcs, such as, in-vehicle travel arcs, waiting and transfer arcs, and penalty arcs for early and late arrival.

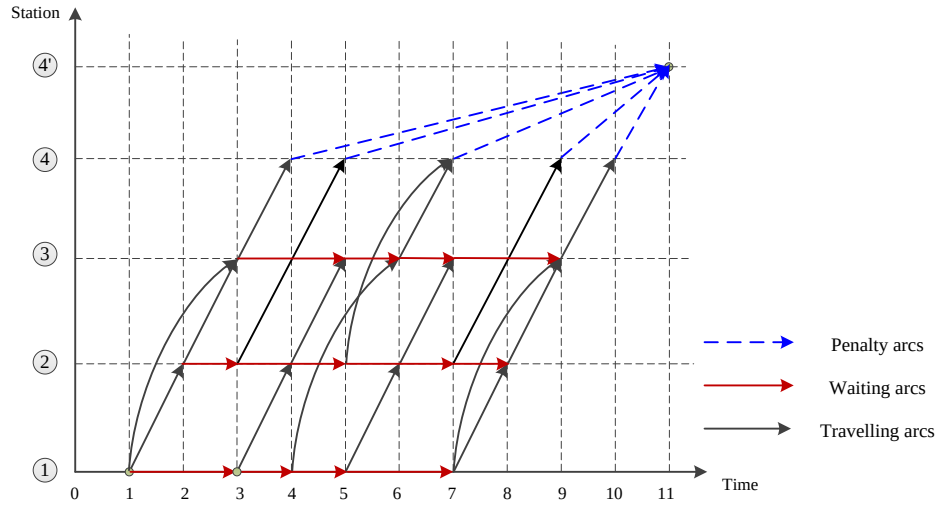


Figure 3-6 Modified Space-time Network with Early and Later Arrival Penalty Arcs

3.6.4 Adding Travel Time Budget to Simplify Path Enumeration on Large-scale Transit Networks

The model proposed in section 3.4.2 requires to enumerate all possible paths for each agent. For a large-scale transit network, it will be extremely burdensome to perform the space-time path enumeration task. For simplification, the acceptable bound for each agent shown in inequality (3.22) can be assumed to be a constant travel time budget $BT^{a}_{w,DT}$ based on each OD pair at different departure times. In this approach, only a limited number of paths satisfying the following constraint are required.

$$\sum_{(i,j,t,s)\in E}\{c_{i,j,t,s} \times x^{a}_{i,j,t,s}\} \leq BT^{a}_{w,DT} \qquad (3.28)$$

Usually, the travel time budget can be obtained through a travel survey. A real-world survey including travel time budget was performed in Chicago Metropolitan area (Nie et al., 2010). The detailed discussion can be found in this review paper (Mokhtarian and Chen, 2004), which states that individuals' travel time expenditures do show patterns and are strongly related to individual and household characteristics, attributes of activities at the destination, and characteristics of residential areas.

In addition, constraint (3.28) can be further modeled as virtual arcs to represent possible inaccessibility for this OD pair under congested condition. As a result, inequality (3.28) will be eliminated in our model and is simply coded in the space-time network as an approximation. The related application can be found at accessibility-based network design problem (Tong et al., 2015). Meanwhile, this virtual arc-based network modeling method could greatly reduce the computational complexity for large-scale applications, because it does not need to consider a set of constraints (constraint 3.28) in the primal problem or a set of dualized constraints in the Lagrangian relaxation procedure.

3.7 Numerical Examples

This section examines different aspects in a dynamic discrete transit service network design problem, where (i) the scheduled travel time is constant and (ii) the capacity on travelling arcs and waiting arcs can represent the vehicle capacity and the station/platform capacity, respectively. The integer linear programming model and the proposed Lagrangian relaxation procedure are demonstrated by the general purpose optimization package GAMS (Rosenthal, 2015) in two small transit networks, and further be tested by our time-dependent shortest path algorithm in C++, by enhancing an open-source mesoscopic

dynamic traffic assignment model namely DTALite (Zhou and Taylor, 2014), for large-scale applications on one workstation with 20 CPU cores (40 threads) with 192GB RAM. The related source codes can be downloaded at the website: https://www.researchgate.net/publication/295968354_Case_1_Transit_network_design_LR and https://www.researchgate.net/publication/295968184_Case_2_Demand_Level_1_Transit_network_design_LR

3.7.1 A Simple Case

This case examines the simple transit network in section 3.4.1 with an opening cost of 10 units and a total budget of 15 units for the potential open line. All Lagrangian multipliers are initially assumed to be 0.1, and the big $M = 6$ is syffucuebt because the time horizon is just 5. In order to not enumerate all possible paths of OD pair $(1,3)$, at first, the shortest path is found in the existing space-time network and then all possible paths that contain the potential open lines are compared. As a result, $k = 2$ for OD pair $(1,3)$. Path 1 is $(1,2,0,1) \rightarrow (2,3,1,3) \rightarrow (3,3,3,4) \rightarrow (3,3,4,5)$ and path 2 is $(1,1,0,1) \rightarrow (1,3,1,4) \rightarrow (3,3,4,5)$.

Based on the procedure of Lagrangian relaxation proposed in section 3.5.2, the objective values of P_x, P_y, and the lower bound are obtained and listed in Table 3-6 after 10 iterations. Figure 3-7 demonstrates a comparison between the lower bound and the optimal value with a final solution gap of 0.83%.

Table 3-6 The Objective Values of P_x, P_y, and the Lower Bound during 10 Iterations

Iterations	1	2	3	4	5	6	7	8	9	10
P_x	12	15.5	18.5	21.93	21.45	22.07	21.95	22.36	22.04	22.35
P_y	7.7	8.9	10.9	10.47	9.72	10.32	10.08	10.51	10.14	10.47
Z_{lb}	4.3	6.6	7.6	11.47	11.73	11.75	11.87	11.85	11.90	11.88

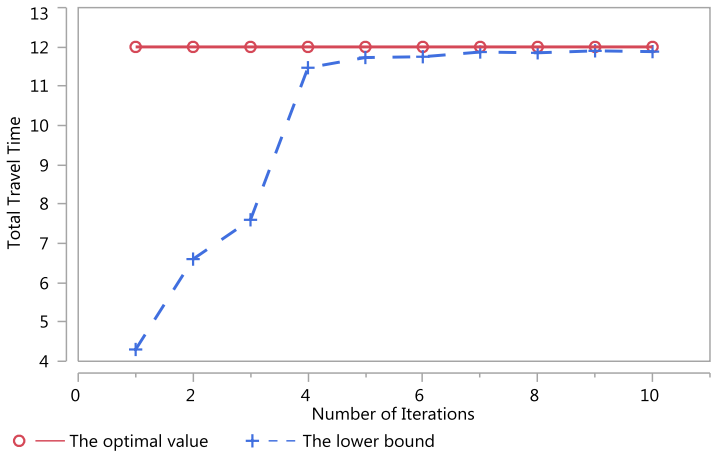


Figure 3-7 Comparison Between the Lower Bound and the Optimal

3.7.2 Transit Service Network Design Based on the Simplified Sioux-fall Physical Network

The proposed algorithm will be also tested in the following network with four hypothetic transit lines, shown in Figure 3-8.

Table 3-7 lists the existing transit service arcs based on given the timetable of the four transit lines. The potential service arcs to be opened and related station storage capacity are listed in Table 3-8 and Table 3-9, respectively. All serviced arcs are displayed in the space-time network shown in Figure 3-9, where the virtual waiting cost at the destination 8 and 9 is 0.

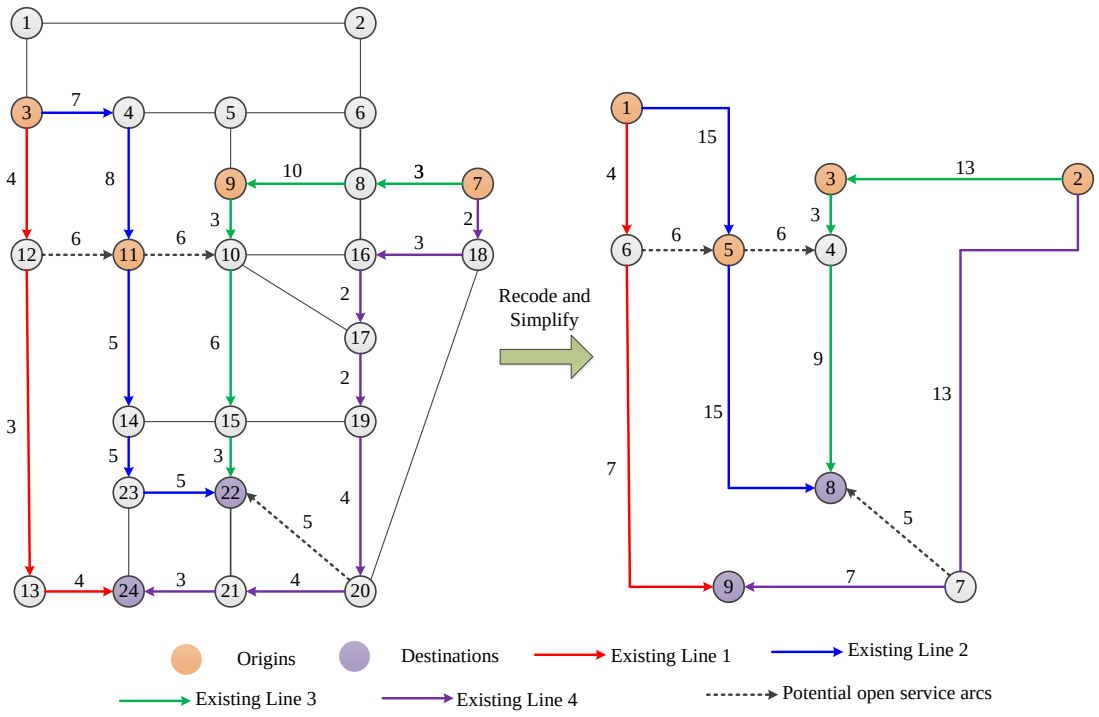


Figure 3-8 Hypothetic Sioux-Fall Transit Network

3.7.2.1 Different Levels of Time-dependent Transit OD Demand

The time-dependent OD demand for all OD pairs is listed in Table 3-10. Initially, three levels of demand are provided to observe its impact on the final service network design result.

In addition, all vehicles' capacity is assumed to be 20. For simplicity, the tolerance value of all agents is assumed to be same as 15 (min or any time unit). For each demand level, three cases are compared: (i) based on the existing transit service network, solved by the dynamic transit assignment problem with bounded rational user equilibrium conditions by GAMS; (ii) considering with potential new service arcs, optimal solution directly solved by GAMS for our proposed model in section 3.4.2; (iii) with potential new service arcs, solved using the proposed Lagrangian relaxation method implemented in GAMS.

Comparison of the three cases under three demand levels are shown in Figure 3-10(a), (b) and (c), respectively.

Table 3-7 Hypothetic Existing Transit Service Arcs

Service Arc	Start Time	End Time	Service Arc	Start Time	End Time
(1,6)	0	4	(4,8)	16	25
(6,9)	4	11	(2,3)	15	28
(1,6)	15	19	(3,4)	28	31
(6,9)	19	26	(4,8)	31	40
(1,6)	30	34	(2,3)	30	43
(6,9)	34	41	(3,4)	43	46
(1,5)	0	15	(4,8)	46	55
(5,8)	15	30	(2,7)	0	13
(1,5)	15	30	(7,9)	13	20
(5,8)	30	45	(2,7)	15	28
(1,5)	30	45	(7,9)	28	35
(5,8)	45	60	(2,7)	30	43
(2,3)	0	13	(7,9)	43	50
(3,4)	13	16			

Table 3-8 Potential Open Transit Service Arcs

Service Arc	Start Time	End Time	Service Arc	Start Time	End Time	Service Arc	Start Time	End Time
(6,5)	4	10	(5,4)	10	16	(7,8)	13	18
(6,5)	19	25	(5,4)	25	31	(7,8)	28	33
(6,5)	34	40	(5,4)	40	46	(7,8)	43	48

Table 3-9 Station Storage Capacity

Station	1	2	3	4	5	6	7	8	9
storage capacity	40	40	30	30	30	30	30	M	M

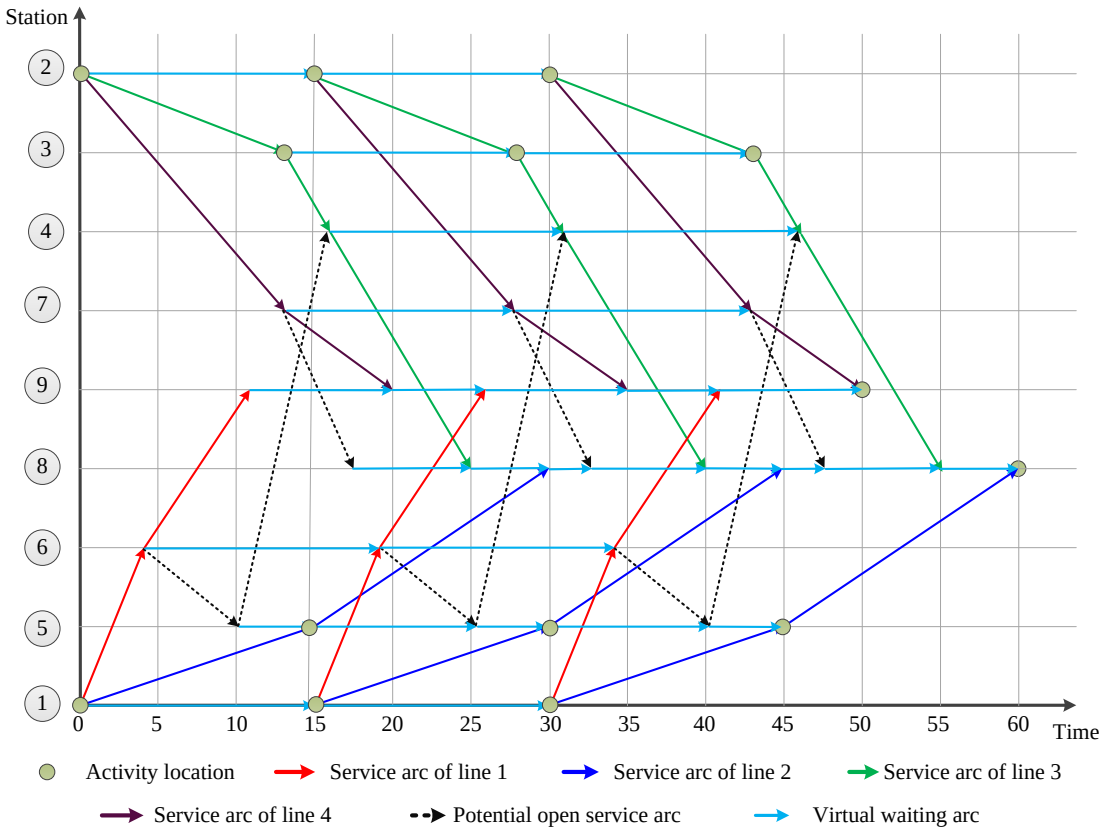


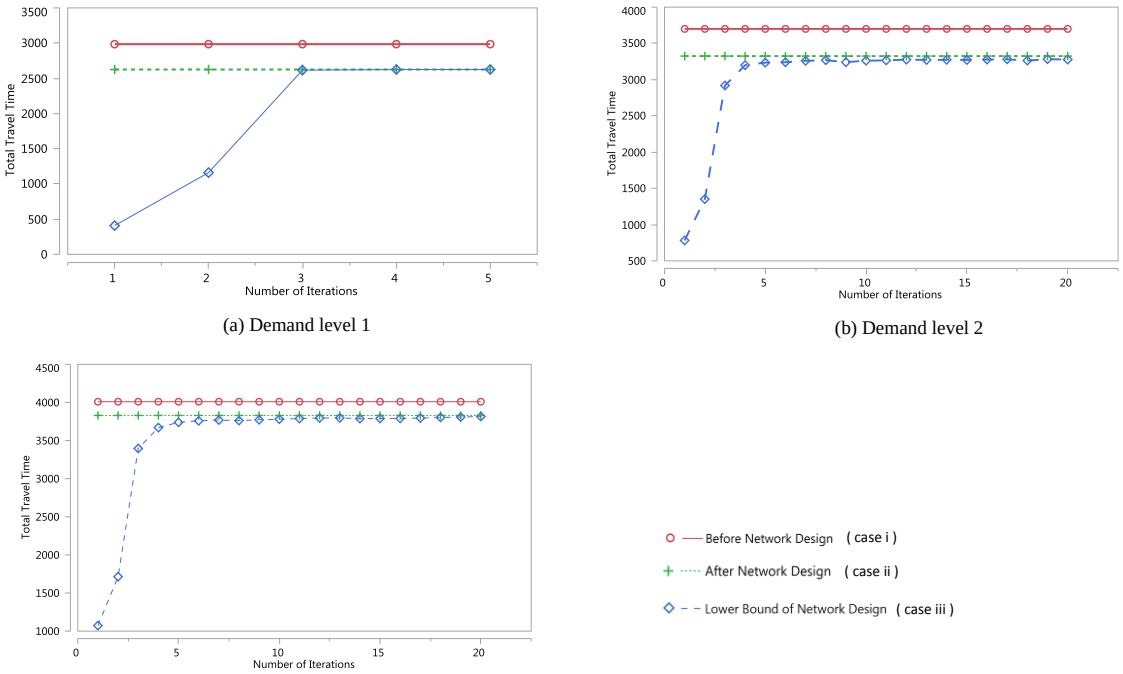
Figure 3-9 The Corresponding Space-time Transit Service Network

Table 3-10 Time-dependent OD Demand

OD	Departure Time 1	Demand (agents)			Departure Time 2	Demand(agents)			Departure Time 3	Demand(agents)		
		L1	L2	L3		L1	L2	L3		L1	L2	L3
1→8	0	10	15	15	15	10	10	10	30	10	10	10
1→9	0	10	15	20	15	10	10	15	30	10	10	10
5→8	15	5	10	10	30	5	5	5	45	5	5	5
2→8	0	10	15	15	15	10	10	10	30	10	10	10
2→9	0	10	15	20	15	10	10	15	30	10	10	10
3→8	13	5	10	10	28	5	5	5	43	5	5	5

Result presented in Figure 3-10 indicates that the total system-wide travel time acquires reduced after transit service network optimization and our proposed Lagrangian relaxation procedure is able to achieve a quick convergence to the optimal solution.

In addition, in order to check the computation efficiency of GAMS solver and the proposed Lagrangain relaxation method, two high level OD demand inputs (demand levels 4 and 5) are designed and presented in Table 3-11.



(a) Demand level 1

(b) Demand level 2

(c) Demand level 3

Figure 3-10 Comparison of the System-wide Travel Time under Three Demand Levels

Table 3-11 Two High Level Time-dependent OD Demand

OD	Departure time 1	Demand(agents)		Departure time 2	Demand(agents)		Departure time 3	Demand(agents)	
		L4	L5		L4	L5		L4	L5
1→8	0	60	75	15	40	50	30	40	50
1→9	0	80	100	15	60	75	30	40	50
5→8	15	40	50	30	20	25	45	20	25
2→8	0	60	75	15	40	50	30	40	50
2→9	0	80	100	15	60	75	30	40	50
3→8	13	40	50	28	20	25	43	20	25

The comparison result on computation efficiency under five different levels of OD demand is displayed in Figure 3-11 .The total number of agents of demand level 1 to 5 is 150, 180, 200, 800, and 1000, respectively. It is obvious that the computation CPU time of the GAMS solver for optimal solutions is significantly increased under demand level 4 and 5, up to more than 40 mins. In comparison, the computation CPU time of the proposed Lagrangian relaxation method always remains low for all 5 demand levels considered. Even under demand level 5, its computation CPU time is less than 3 mins through 50 iterations, about only 6% of that required by the GAMS solver.

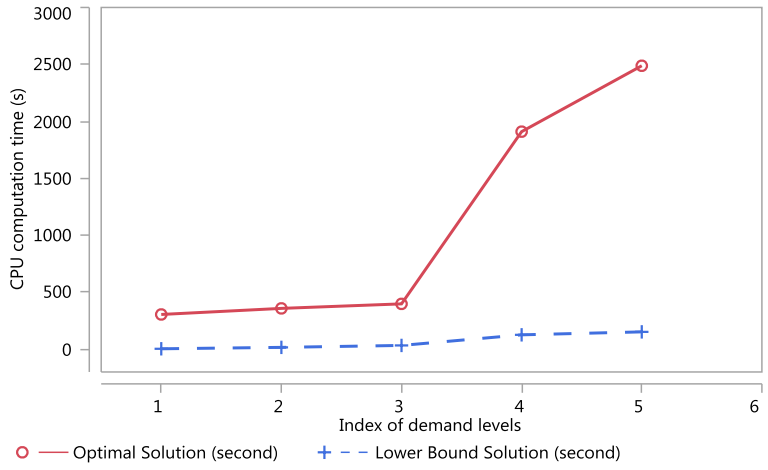


Figure 3-11 Comparison on CPU Computation Time under Different Demand Levels

3.7.2.2. Considering Different Levels of Transit Vehicle Capacity

In order to perform the sensitivity analysis on the vehicle capacity, demand level 2 is adopted as the time-dependent OD demand input. The different vehicle capacities considered are 18, 19, 20, 21, and 22, respectively. The optimal system-wide travel time under different cases are shown in Figure 3-12. It shows that based on the optimal solutions in the network design problem, the transit system efficiency improves with increased

capacity of all vehicles. That is rational in that the larger vehicle capacity increase the service capacity of the whole network, including the shortest paths among OD pairs.

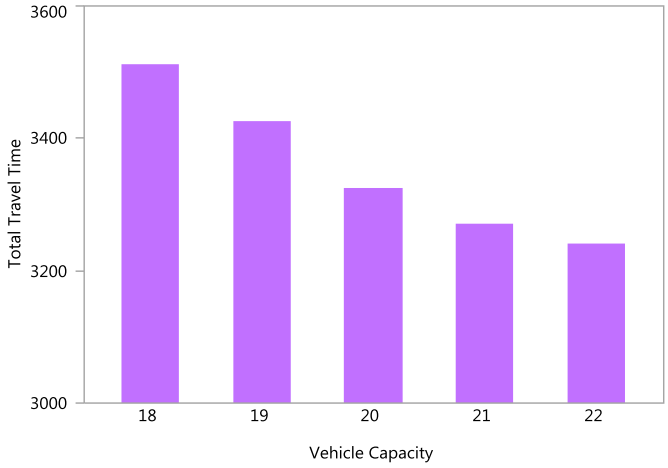


Figure 3-12 Comparison on System-wide Travel Time under Different Vehicle Capacities

3.7.2 3. Different Levels of Agents' Tolerance Value

The agent's tolerance value can also influence the final optimal solutions, so this sub-section examines six scenarios to observe their impacts on the transit service network design. Scenario 1 to 4 have the same tolerance value for all agents, which is 7, 8, 9, and 10, respectively. For scenario 5, different agents could have different specific tolerance values as shown in Table 3-12. For example, for OD pair (1→8), there are 5 agents departing at time 0 with tolerance value of 0, 5 agents with tolerance value of 5, and 5 agents with tolerance value of 10.

For scenario 6, the boundedly rational travel decision rule is ignored to obtain one system optimal solution for the transit service network design problem. The system performance results directly solved in GAMS are listed in Table 3-13.

The results yield the following observations:

(1) If all agents' tolerance value is 7, no feasible solution exists, which means that some agents will cancel their trips or the tolerance value is not accurate and needs further calibration

Table 3-12 Number of Agents with a Specific Tolerance Value for Scenario 5

OD	Departure time 1	Tolerance value			Departure time 2	Tolerance value			Departure time 3	Tolerance value		
		0	5	10		0	5	10		0	5	10
1→8	0	5	5	5	15	5	0	5	30	5	0	5
1→9	0	5	5	5	15	5	0	5	30	5	0	5
5→8	15	5	0	5	30	0	0	5	45	0	0	5
2→8	0	5	5	5	15	5	0	5	30	5	0	5
2→9	0	5	5	5	15	5	0	5	30	5	0	5
3→8	13	5	0	5	28	0	0	5	43	0	0	5

Table 3-13 System Performance of Different Scenarios

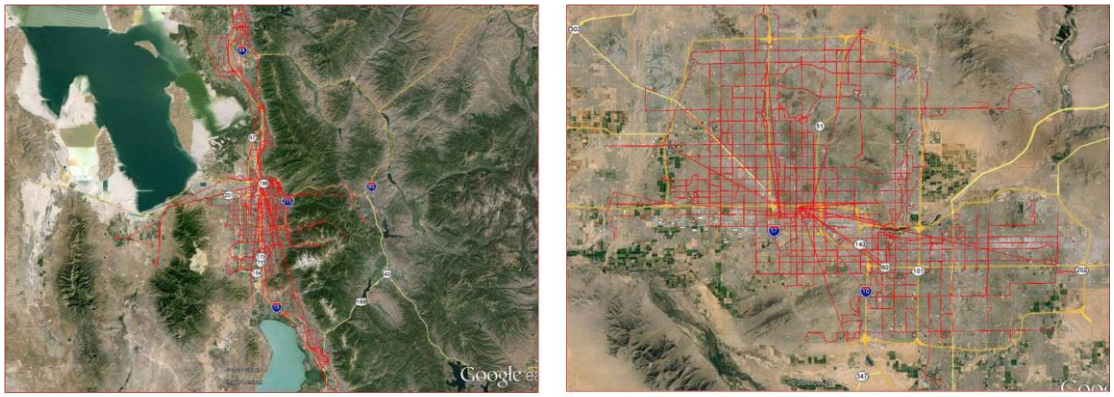
Scenario No.	Tolerance value	Total system travel time
1	7	Infeasible
2	8	3500
3	9	3500
4	10	3325
5	Specified	3400
6	System Optimal	3325

(2) As the tolerance values of all agents increase, the system cost is closer to the system optimal value, because more agents can tolerance a longer path to allow some agents to select a shorter path

(3) When the specific tolerance value of each agent is different, the transit system will have a more complicated route choice set.

3.7.3 Large-scale Experiments

In order to demonstrate the computational efficiency of our proposed Lagrangian relaxation framework, this sub-section tests our proposed algorithm on two regional transit networks, Salt Lake City regional transit network and Phoenix regional transit network, respectively. The transit network data are based on the public GTFS (**General Transit Feed Specification**) feed. Through extracting the transit stop information in *stops.txt* and stop sequences of each trip in *stop_times.txt*, the two transit networks are created and then are visualized in Google Earth as shown in Figure 3-13. It can be observed that there are some unrealistically long straight transit links in the two networks. The reason is that the geographical shape information of each transit line in *shapes.txt* is not read. We just utilize the longitude and latitude coordinates of each stop of each transit line in *stops.txt*, and then one straight link is correspondingly created to connect the two joint stops.



(a) Salt Lake City regional transit network visualized in Google Earth

(b) Phoenix regional transit network visualized in Google Earth

Figure 3-13 Salt Lake City and Phoenix Regional Transit Feed Data Visualized in Google Earth, Respectively

In order to obtain the dynamic transit travel OD demand data, two scenarios are assumed for the tests, (i) 10% of regional traffic demands as the transit OD demand input and (ii) 5%

of regional traffic demands as the transit OD demand input. The details are listed in Table 3-14.

Table 3-14 Transit Travel OD Demand Input of Two Scenarios

Networks	Demand time period	Total traffic demand	Transit demand (10%)	Transit demand (5%)
Salt Lake City	15pm-18pm	1.35 million agents	135,000 agents	67,500 agents
Phoenix	all day (24 hours)	10.5 million agents	1.05 million agents	525,000 agents

In addition, the activity location of each zone in the regional traffic network is matched with the nearest transit stop, so the final transit network information is listed in Table 3-15.

Table 3-15 Transit Network Information

Transit networks	Number of stops	Number of links connecting two joint stops at each transit line	Number of zones
Salt Lake City	6393	7219	2302
Phoenix	6788	7015	3022

Based on the simplification method in section 3.6.4 and the Lagrangian relaxation framework in Section 3.5, the relaxed problem can be decomposed as two subproblems at each iteration: one is time-dependent least cost path problem for each agent, which is the most time-consuming part in computation, and the other is the knapsack problem for service arc selection. In our proposed algorithm implemented in C++ code, Open Multi-Processing (OpenMP) is used as the application programming interface (API) for parallel computing. The tasks of path finding for each agent on the transit network are assigned into different available CPU threads and performed in parallel. The two regional transit networks under two different scenarios are tested on our Dell Precision T7610 Workstation with 20 CPU cores and 192G RAM. The general CPU computation time of each iteration in the Lagrangian relaxation approach are listed in Table 3-16.

Table 3-16 CPU Computation Time of One Iteration in the Lagrangian Relaxation Approach

Transit networks	Scenario 1: Transit demand (10%)	Scenario 2: Transit demand (5%)
Salt Lake City	2 min 11 seconds	1 min 22 seconds
Phoenix	12 min 30 seconds	6 min 45 seconds

The CPU computation time could be affected by the number of optimized agents, the number of zones, the optimized time period, the topology of transit networks, the time budget value, etc. Looking back upon the study cases solved in GAMS in section 3.7.2, our proposed algorithm coded by C++ with the parallel computing method has an obvious advantage at computational efficiency. In future, based on the service arc selection decisions from the lower bound, the dynamic transit simulation with bounded rational agents will be developed for the upper bound. Finally, the gap between the lower bound and the upper bound can be used to check the quality of our solutions.

CHAPTER 4

HOSUEHOLD DAILY SCHEDULED ACTIVITIES: DRIVING ALONE AND RIDE SHARING

4.1 Introduction

Different with Chapter 3 which focus on public scheduled transit systems, this chapter will turn to the household daily scheduled activities (Liu et al., 2017). The advanced technologies is creating opportunities to optimally plan those household activities for reduce the household-level or system-level total travel cost. In this research, available vehicles are assumed to be owned by each household. We admit that the upcoming autonomous vehicles could be owned by the companies and stayed together at depots, which will be our future study. Therefore, the problems we aim to solve is similar with the Household Activity Pattern Problem (HAPP) proposed by Recker (1995).

This chapter will reformulates two special cases of HAPPs as system-optimal multi-household activity scheduling and then consider the tight road capacity to capture the impacts of traffic congestion on activity generation and scheduling. Compared to the traditional formulation, the space-time-state network is built to precisely represent and translate side constraints in the state dimension, which could eliminate activity time window and vehicle selection constraints in the resulting optimization model. Through dualizing the capacity constraints to the objective function by Lagrangian relaxation, our proposed model can be further solved through time-dependent state-dependent least cost path-finding algorithms, which permits the use of fast computational algorithms on large-scale high-fidelity transportation networks.

In addition, in a micro-level care-following perspective, jointly optimizing multi-vehicle trajectories is a critical task in the next-generation transportation system with autonomous and connected vehicles (Wei et al., 2017). An integer programming model will be presented for scheduling longitudinal trajectories, where the goal is to consider both system-wide safety and throughput requirements under supports of various communication technologies. Finally, a simple study case is conducted to illustrate the optimal trajectory control of multiple autonomous vehicles under different communication conditions

4.2 Problem Statement

Table 4-1lists the notations used for the household daily scheduled activities problems.

This research focuses on three particular cases (namely A, B, C) and further extends to one more general case D for optimally planning those household daily scheduled activities. The general given input includes the population used for activity generation, a physical transportation network, a set of different types of activities (mandatory, semi-mandatory, optional activities) with specific time windows and utility values, a set of vehicles, as well as the activity/vehicle assignment set to each household member. By adapting the classical assumption/definition from Recker (1995), we present the following problem statements.

(1) Case A is a multi-vehicle and multi-person vehicle routing problem with mandatory and discretionary activities, which is similar to Case IV in the paper by Recker (1995). (i) Members of the household share a set of vehicles; a subset of vehicles may be available for use by any member of the household, and the remainder may be reserved for use by certain members; (ii) A subset of activities can be performed by any member of the household,

and the remaining activities must be performed by certain members; (iii) Certain members can have specific mandatory activities or optional activities;

Table 4-1 Notations Used in the Household Daily Scheduled Activities Problems

Notations	Definition
N	Set of nodes in the physical network, including necessary virtual nodes
N_v	Set of vehicle nodes for vehicle selection
L	Set of links in the physical network, including necessary virtual links
P	Set of household members
P_m	Set of household members who have mandatory activities
P_n	Set of household members who chooses one mandatory activity from multiple candidates
P_q	Set of household members who have discretionary activities
V	Set of available vehicles
A	Set of activities
$A_n(p)$	Set of household member p's candidate activities for one kind of mandatory activity
A_v	Set of mandatory activities of vehicle v's driver
R	Set of vertices in the space-time/space-time-state network
E	Set of edges/arcs in the space-time/space-time-state network
W	Set of cumulative vehicle activity-performing state
$E(p, a_m)$	Set of edges/arcs of household member p's mandatory activity a_m
$E(p, a_n)$	Set of edges/arcs of household member p's candidate activity a_n for one kind of mandatory activity
$E(p, a_q)$	Set of edges/arcs of household member p's discretionary activity a_q
$E(v, a_m)$	Set of edges/arcs of mandatory activity a_m of vehicle v's driver
i, j	Index of node set N
(i, j)	Index of link set L
t, s	Index of time intervals in the space-time-state network
w, w'	Index of state in the space-time-state network
(i, t)	Index of vertex in the space-time network
(i, j, t, s)	Index of edges/arcs in the space-time network
(i, t, w)	Index of vertex in the space-time-state network
(i, j, t, s)	Index of edges/arcs in the space-time-state network
p	Index of household member set P
a	Index of activity set A
$t(i, j)$	Travel time of link (i, j)
$c^p_{i,j,t,s}$	Travel cost of arc (i, j, t, s) of person p in the space-time network
$c^v_{i,j,t,s,w,w'}$	Travel cost of arc (i, j, t, s, w, w') of vehicle v in the space-time-state network
$[a_k, b_k]$	The time window of event k, such as, activity starting time window, activity ending time window
$TD(p)/TD(v)$	Earliest departure time of household member p/ vehicle v
$O(p)/O(v)$	Origin node of household member p/ vehicle v
$D(p)/D(v)$	Destination node of household member p/ vehicle v
T	The time horizon in the space-time network/space-time-state network
$Cap_{i,j,t,s}$	Capacity of arc (i, j, t, s)
$Cap_{i,j,t,s,w,w'}$	Capacity of arc (i, j, t, s, w, w')
$x^p_{i,j,t,s}$	Binary variable, $= 1$, if household member p visits the traveling/waiting arc (i, j, t, s) in the space-time network; $= 0$ otherwise
$x^v_{i,j,t,s,w,w'}$	Binary variable, $= 1$, if vehicle v visits the traveling/waiting arc (i, j, t, s, w, w') in the space-time-state network; $= 0$ otherwise

(iv) Some members may perform no activities; some vehicles may not be used.

(2) Case B is a multi-vehicle and multi-person ridesharing problem with mandatory and discretionary activities, which can be treated as a special sub-problem of Case V in the paper by Recker (1995). The specific definition is: (i) the ride-sharing pattern that which household members will share one vehicle and which one is the driver has been given; (ii) A subset of activities can be performed by any member of the household, and the remaining activities must be performed by certain members; (iii) Certain members can have specific mandatory activities or optional activities.

(3) Case C is an extension of cases A and B, which considers tight road capacity constraints to capture the underlying congestion in physical transportation networks, so that the influence of time-dependent link travel time on household activity patterns can be observed. As a result, this case is a system optimal multi-household activity scheduling problem under time-varying traffic conditions.

(4) Case D is a dynamic household-level equilibrium problem where each household is inclined to choose the optimal activity pattern, which considers vehicle selection, mode choice and ride-sharing options simultaneously. As studied in a recent paper by Liu and Zhou (2016), when there is no link capacity constraint, each agent (e.g., passenger, vehicle, or household) can choose the best/shortest path without affecting each other. Once the limited resource constraint is strictly considered, some agents may have to accept a longer path in order to finish their own travel and this kind of decision mechanism could invoke bounded rationality to those agents.

In addition, compared to the existing integration of ABM and DTA, the mathematical program oriented modelling framework shown in Figure 4-1 aims to optimize the household activity decisions with system optimal goals under household-level activity requirements and network capacity constraints. In our future study, Case D will be further examined to study possible dynamic household-level equilibriums with household activity interactions.

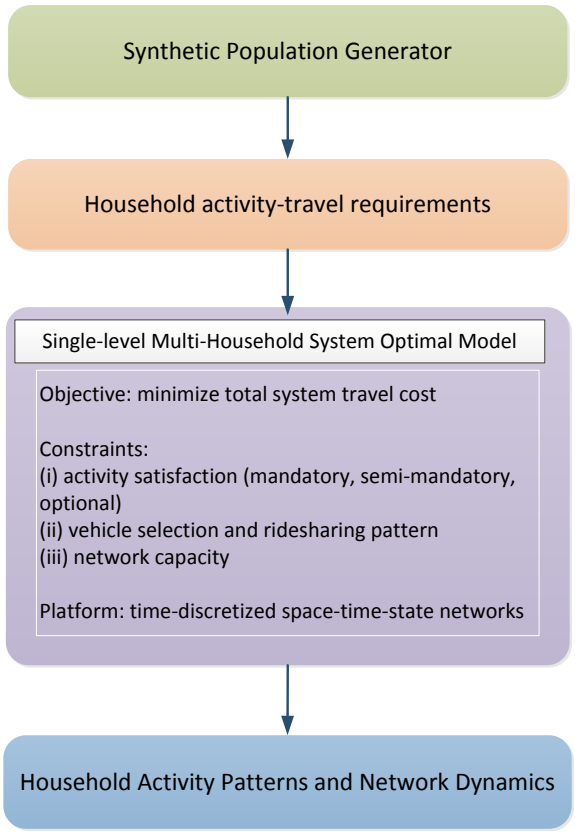


Figure 4-1 Proposed Modelling Framework of Case C

4.2.1 Network Construction and Conceptual Illustration of Case A

For illustrative purposes, a hypothetic three-node network shown in Figure 4-2(a) is used to explain the problem addressed by Case A. There are two household members (p_1 and

p_2), two available vehicles (v_1 and v_2), and two activities (a_1 and a_2). The available vehicle set of household members p_1 and p_2 is $\{v_1, v_2\}$ and $\{v_2\}$. The available activity set of household members p_1 and p_2 is $\{a_1, a_2\}$ and $\{a_1\}$, respectively, and the both activities belong to the mandatory activity and should be finished finally.

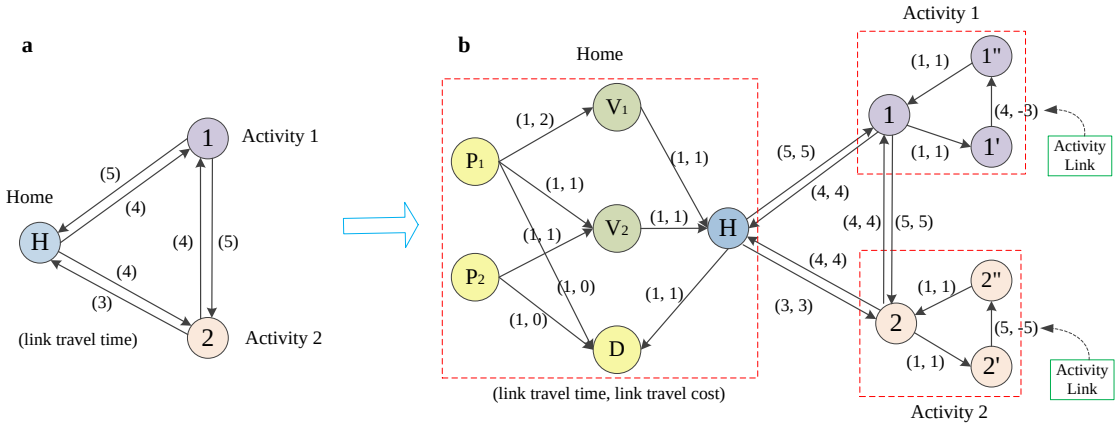


Figure 4-2 (a) Physical network; (b) Corresponding modified network

In order to model those requirements above, the physical network is modified as shown in Figure 4-2(b), where the previous home node and activity nodes are split as several nodes. The detailed explanations are as follows. First, the home node is extended as six nodes, where (i) each household member has his/her dedicated node as his/her origin node, (ii) two vehicle nodes are created and one can view the links between household member origin node and vehicle nodes as vehicle selection links, while each vehicle node can only be visited less than or equal to once by all passengers, and (iii) node D serves as the super destination node. Moreover, to follow an activity-on-the-link representation scheme, the extended network on the right has now activity starting node 1' and ending node 1" corresponding to the activity node 1 on the left-hand side, and the link between the two nodes can be used to represent the required activity time duration.

In order to consider passenger-to-vehicle preference, the travel costs on those vehicle selection links can be passenger-specific. For example, the travel cost from passenger p_1's origin to the super destination is 0, which indicates that when passenger p_1 stays at home as one particular vehicle selection, there is no travel cost. In addition, the travel cost on link (p_1, v_1) is higher than that on link (p_1, v_2), indicating passenger p_1's higher preferences toward vehicle 2 compared to vehicle 1.

Each activity in HAPP typically has one specific time window, then we assume that the beginning time windows for (i) passengers p_1 and p_2 and (ii) activities a_1 and a_2 are [1,3], [1,4], [9,10], and [20, 21], respectively, along the total time horizon of 32 time units. Furthermore, the waiting cost of each time interval at origin nodes and destination node is assumed to be 0, and the waiting at activities nodes has a cost of 1 at each time interval. Within a deterministic disutility minimization framework, we assume negative cost values on activity links shown in Figure 4-2.

A standard time-discretized space-time network can be constructed through the procedure proposed in the papers (Tong et al., 2015; Li et al., 2015; Liu and Zhou, 2016; Lu et al., 2016), and the feasible space-time prism can be greatly reduced as illustrated in Figure 4-3. As a result, the problem becomes how to find passengers' trajectory satisfying all time windows and activity requirements in the space-time network so as to minimize the total travel cost of all household members.

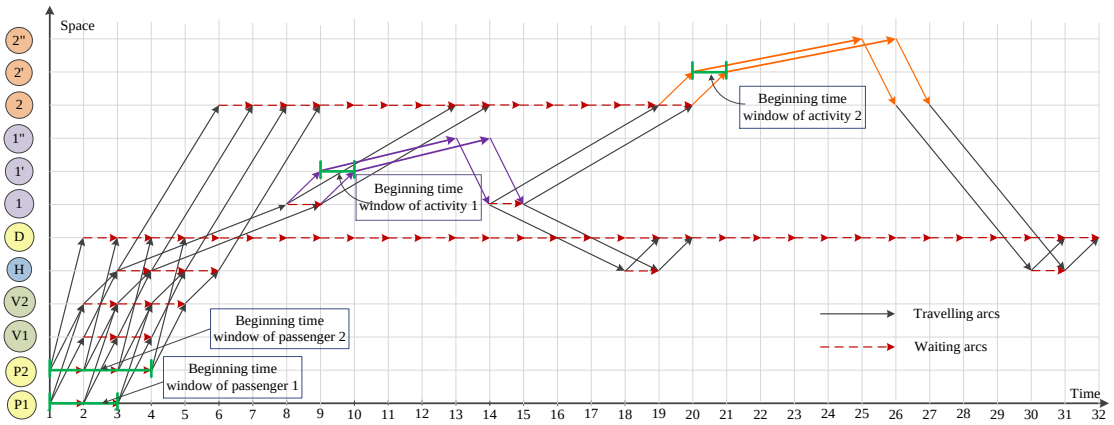


Figure 4-3 Feasible Searching Region in the Space-time Network

As a remark, the (time-dependent) travel time on each travelling arc could be given in advance to reflect the congestion due to complex travel route choice interactions in the real-world traffic network. However, in the following Case C, we directly consider tight link/arc capacity inside the model to compute the resulting congestion effect explicitly. When the number of inflow vehicles exceeds the capacity of traveling arc, some vehicles have to wait at the waiting arc for available travelling arc capacity at next time interval. The detail about how tight capacity constraint is considered in the time-discretized space-time networks can be found in recent papers by Lu et al. (2016) and Liu and Zhou (2016). Their agent-based approach does not use the traditional flow-based nonlinear link/path cost function, and the travel cost of each agent is the result of the interaction among different agents in space-time networks.

4.2.2 Network Construction and Conceptual Illustration of Case B

The most difficult issue in modeling the household-level ridesharing problem is how to recognize the complex coordination among different household members, pertaining to the following questions such as who is the driver and where/when the driver should drop off

and pick up passengers. Considering offline planning applications, our Case B assumes that the set of possible ridesharing patterns is pre-specified with the given drop-off and pick-up locations with time windows to choose.

An illustrative example is given in Figure 4-4(a), where there are two household members (p_1 and p_2), one available vehicle, and three activities (a_1, a_2 and a_3). The given ride-sharing pattern requires that driver p_1 needs to drop off the passenger p_2 to his/her own activities within given beginning time windows, and then this driver needs to pick up p_2 from the activity locations within given activity ending time windows. The driver p_1 could accompany passengers to perform their activity, and also can leave to conduct his/her own mandatory activities. In this example with a quite busy household activity agenda, the driver has to perform the mandatory (deriver as D) activity a_1 while the passenger needs to finish the mandatory (passenger as P) activities a_2 and a_3.

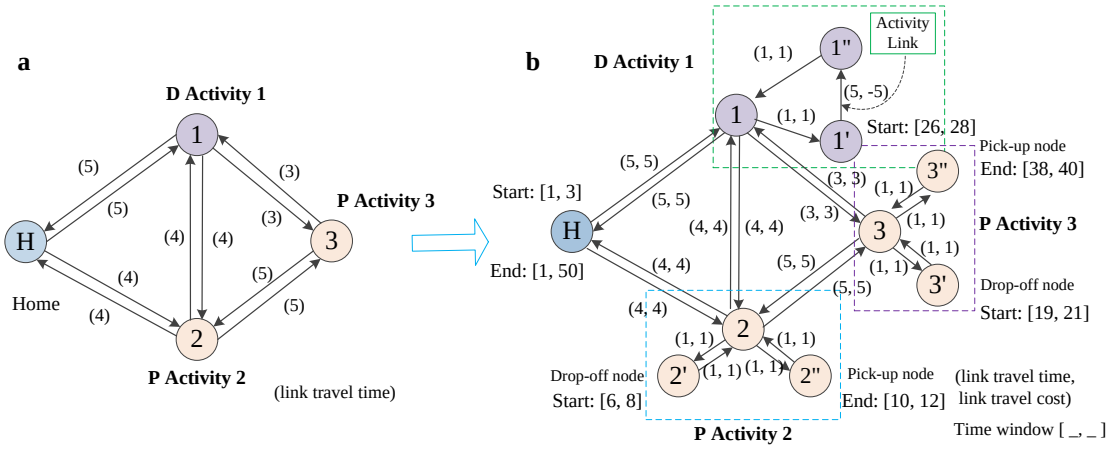


Figure 4-4 (a) Physical Network; (b) Corresponding Modified Network

Accordingly, we construct a drop-off node and a pick-up node for each passenger at the activity location in Figure 4-4(b). It should be remarked that, in a typical case, one is

dropped off and picked up at the same activity location, but our formulation also makes it possible that one passenger is dropped off at one activity and then picked up at another location if he/she can take other travel modes (walking, transit or taxi) to the (spatially different) pick-up location. The starting node and ending node of the passenger activity (shown as P activity 3' and 3") is considered a special drop-off node and pick-up node in drivers' network.

In addition to using the two dimensions (space and time) to depict vehicles' travel trajectory, this section will introduce one more state dimension to model ride-sharing status. More precisely, the state code covers each traveler's service status, including the driver and all passengers. Through adding one more dimension and exogenously listing the possible relation of location, time, and vehicle state, a set of hard activity-performing constraints for the driver and passengers in each vehicle could be embedded in advance in the space-time-state network, which will greatly reduce the set of side constraints and make our proposed mathematical model tractable for network flow optimization algorithms.

To solve the single-vehicle routing problem with pickup and delivery service with time windows (VRPPDTW), Psaraftis (1983) proposed a cumulative service state {1,2,3} to record the service status of each passenger, where 3 means that the passenger has not been picked up, 2 means that the passenger has been picked up but not been delivered, and 1 means that the passenger has been successfully delivered. In this chapter, we adopt the cumulative state representation as {0,1,2}: 0 means that the activity has not been performed, 1 means that the activity is being performed or the passenger has been dropped off at the activity location but not been picked up, and 2 means that the activity has been performed

or the passenger has been picked up. While Mahmoudi and Zhou (2016) firstly proposed a third dimension as vehicle carrying state to solve the VRPPDTW, our third dimension of household-oriented state (with a rich representation of different household members, driver, passenger and associated activities) and the process of state transition are systematically different with those of Mahmoudi and Zhou (2016). The general comparison is listed in Table 4-2.

Table 4-2 Comparison of Model Building between Mahmoudi and Zhou (2016) and Our Case B

	Vehicle-oriented state based VRPPDTW (Mahmoudi and Zhou, 2016)	Household-oriented state in proposed HAPP Case B
State representation	Vehicle carrying state that indicates how many passengers are carried subject to its carrying capacity.	Vehicle, designated driver (as a household member) multiple passengers, activity-execution states
Multiple passenger activity task	Not modelled	one passenger could conduct multiple mandatory and optional activities
The third dimension (state)	0: passenger is not carried by the vehicle; 1: passenger is being carried by the vehicle.	0: the activity of one passenger has not been performed; 1: the activity is being performed or the passenger has been dropped off at the activity location but not been picked up; 2: the activity has been performed or the passenger has been picked up
State transition logic between driver and passenger	Not modelled	one passenger could have several activities, and when he/she is dropped off by the driver at one activity location, it is impossible for him/her to be dropped off at other activity location before he/she is picked up at the previous activity location

Since one activity could have 3 different states, if there are n activities for all passengers in one vehicle, it would require 3^n variables to represent all possible states. The total number of states is shown in Table 4-3 depending on the number of activities. However, if one passenger has multiple activities, the possible states could be reduced because one passenger cannot perform multiple activities simultaneously. Also, the tight time window

and transition preference for each activity can greatly reduce the number of possible states reasonable within a feasible space-time prism. In addition, the rapid development of hardware of computers could provide more memory and faster computation speed to address those large number of state search decisions.

Table 4-3 The Maximum Number of Possible States Corresponds With the Number of Activities in One Vehicle

Number of activities	1	2	3	4	5	6	7	8
Maximum number of possible states	3	9	27	81	243	729	2187	6561

We now use the example above to illustrate our cumulative activity-performing state and the state transition at different activity locations and times. There are one vehicle with two household members and three activities, so the vehicle's activity-performing state can be $[a_1, a_2, a_3]$, or more generically denoted as $[\_, \_, \_]$, where the first slot represents the driver's activity-performing state of activity a_1 and the second slot and the third one represent passenger p_2 's two activity-performing states of activities a_2 and a_3, respectively. To reduce the number of states in this combinatorial optimization problem, one can also implement the activity-performing requirement as constraints on the activity link $(1' \rightarrow 1")$for the driver, so the resulting reduced state vector is $[a_2, a_3]$.

Since activities a_2 and a_3 are mandatory for passenger p_2, all possible vehicle's state could be $[a_2 = 0, a_3 = 0]$, [1,0], [2,0], [0,1], [0,2], [2,1], [1,2], and [2,2] by enumeration. It is noticed that $[a_2 = 1, a_3 = 1]$ is not included because it is impossible that passenger p_2 is dropped off at two locations simultaneously. Figure 4-5(a) illustrates a graph of possible state transitions for the example above. In addition, if there is the same

type of multiple activities, such as, shopping at location 2 vs. at location 3, passenger 2 may just need to choose one of the two locations, so the resulting possible state transition will be that shown in Figure 4-5(b).

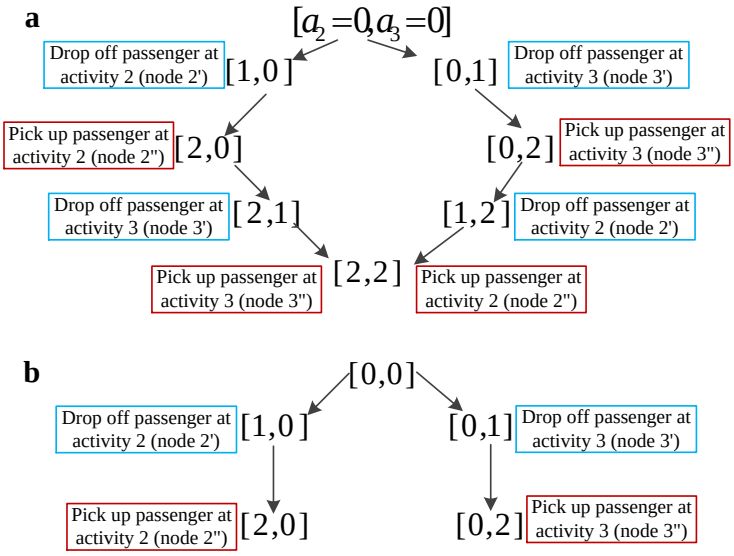


Figure 4-5 (a) Both Activities Need to Be Performed; (b) Exact One of Two Activities Should Be Performed

There are three types of mutually exclusive multidimensional arcs in the space-time-state network:

Travelling arcs $(i, j, t, s, w, w' = w)$ with a time-dependent cost on link (i, j) departing at time t, with the same state w as transportation services do not change activity performing states.

Waiting arcs $(i, i, t, t + 1, w, w')$ with a unit of waiting costs at location i from time t to time $t + 1$. A special Case is that, the waiting cost should be zero at the super home origin and destination nodes.

State transition/service arcs (i, i, t, t, w, w') with a utility (i.e. negative travel cost) when performing their activities at the drop-off location. As shown in Figure 4-6, at node $i = 2'$, time $t = 6$, 7 or 8 within a given time window, we have a number of possible state changes, for example, $w = [0,0]$ with a possible transition to $w' = [1,0]$, or $w = [0,2]$ with a possible transition to $w' = [1,2]$. As the ending state for a_2 must be 2 so the passenger 2 will be picked up automatically among any feasible solutions and there is no benefit at pick-up nodes to avoid double counting of service utilities.

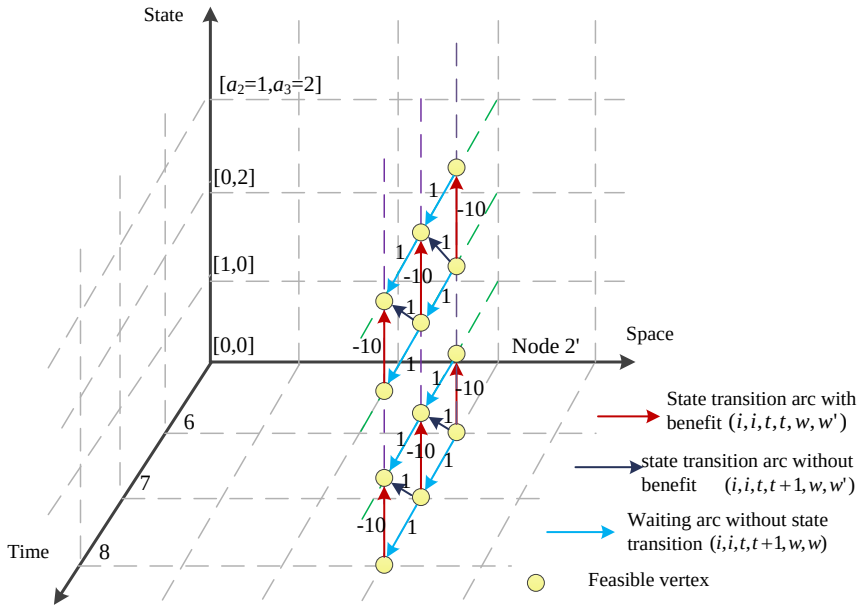


Figure 4-6 Feasible Arcs at Node 2' in a Space-time-state Network

4.2.3 Conceptual Illustration of Case C

As an extension of Cases A and B, Case C strictly honors the travelling arc capacity in the space-time network and space-time-state network, similar to the consideration in the recent papers along this line (Lu et al., 2016; Liu and Zhou, 2016). Compared with the constant link free-flow travel time, the underlying time-varying congestion could dramatically

affect the passenger/vehicle's departure time, route choice, mode choice, destination choice, and even activity generation.

Without loss of generality, we adopt a time-invariant network (Liu and Zhou, 2016) shown in Figure 4-7 to illustrate the congestion effect for two households with two different activities, where household 1 (household 2) has one member who departs from home node H_1 (H_2) to perform activity A_1 (A_2) then go back home, respectively.

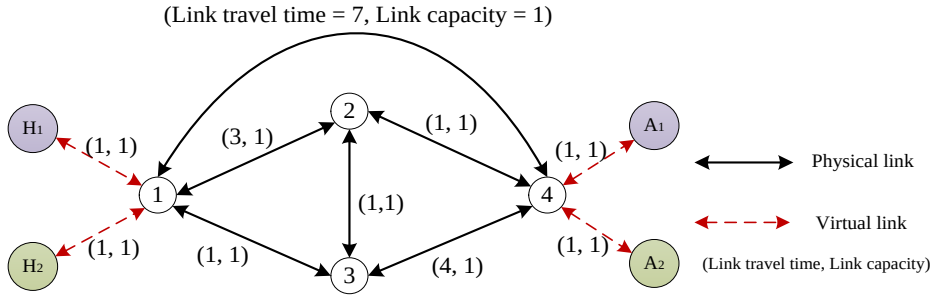


Figure 4-7 A Simple Illustrative Network for Case C

What can be observed in Table 4-4 is summarized as follows and those observation can also be applicable to space-time and space-time-state networks.

(1) When the link capacity is not taken into account, the vehicles from both households choose their own shortest path. The physical path node sequences of households 1 and 2 is $1 \rightarrow 3 \rightarrow 2 \rightarrow 4 \rightarrow 2 \rightarrow 3 \rightarrow 1$ with path travel time of 6.

(2) When the link capacity is considered, the system optimal objective of Case C could make household 1 change its path as $1 \rightarrow 2 \rightarrow 4 \rightarrow 2 \rightarrow 1$ with a larger path cost of 8. Meanwhile, household 2 would switch a new path as $1 \rightarrow 3 \rightarrow 4 \rightarrow 3 \rightarrow 1$ with an increased cost of 10.

Table 4-4 Result Analysis of Different Cases

Different cases		Physical path selection: Path 1	Path 2	Path 3	Path 4	Remarks
Case A/B: without link capacity constraint		×	×	×	√; √	Benchmark
Case C: system optimal with link capacity constraint		√	√	×	×	Compared with Case A/B, household has possible departure time change, route choice change, and possible activity cancel or mode choice change due to link capacity constraints.
Case D: household equilibrium with link capacity constraint	With links 3 ↔ 2	×	×	√	√	Compared with Case C, the total travel time is increased.
	Without links 3 ↔ 2	√	√	×	×	Braess paradox occurs, as the system-wide cost reduces without the link.

√: One household (vehicle) chooses the corresponding path; ×: No household (vehicle) chooses the corresponding path

(3) In observation (1), the travel time of household 2 from 1 to 4 is 3, but now it will increase to 5 due to link capacity constraint in observation (2). If the passenger of household 2 has a strict time window for activity 2, the increased path travel time from 1 to 4 could make passenger depart earlier to satisfy the time window.

(4) If the time budget of household 2 from 1 to 4 is less than 5, the passenger would cancel activity 2 or may change to an alternative by switching to other possible travel modes.

(5) If Case D is considered for possible equilibrium conditions, one household could choose the previous shortest path and the other has to accept the longer path, $1 \rightarrow 4 \rightarrow 1$, with total travel time of 14. It also could lead to changes in departure time, activity cancel or mode choice. In addition, Braess paradox exists in the network above, so blocking links

$3 \rightarrow 2$ and $2 \rightarrow 3$ definitely could improve the transportation efficiency and further influence household activity patterns from the perspective of traffic managers.

4.3 Mathematical Programming Models

4.3.1 Space-time Network-based Optimization Model for Case A

Based on the two-dimension space-time network constructed in section 4.2.2, we formulate our mathematical programing model that satisfies all requirements in *Case A*, which aims to optimize vehicle selection, activity-performing selection and route guidance for each household member so as to minimize the total household travel cost.

Model 1:

Objective function

$$\min \sum_{p} \sum_{(i,j,t,s)\in E} (c_{i,j,t,s}^{p} \times x_{i,j,t,s}^{p}) \tag{4.1}$$

Subject to

(1) Flow balance constraint for each person:

$$\sum_{i,t:(i,j,t,s)\in E} x_{i,j,t,s}^{p} - \sum_{i,t:(j,i,s,t)\in E} x_{j,i,s,t}^{p} = \begin{cases} -1 & j = O(p), s = DT(p) \\ 1 & j = D(p), s = T \\ 0 & otherwise \end{cases}, \forall p \tag{4.2}$$

(2) Vehicle selection constraint at vehicle selection node:

$$\sum_{p} \sum_{i,t:(i,j,t,s)\in E} x_{i,j,t,s}^{p} \le 1, \forall j \in N_v \tag{4.3}$$

(3) Mandatory activity participation for one specific household member:

$$\sum_{i,t:(i,j,t,s)\in E(p,a_m)} x^p_{i,j,t,s} = 1, \forall p \in P_m \tag{4.4}$$

(4) Mandatory activity with multiple candidates for one household member:

$$\sum_{a_n \in A_n(p) i,t:(i,j,t,s)\in E(p,a_n)} \sum x^p_{i,j,t,s} = 1, \forall\ p \in P_n \tag{4.5}$$

(5) Discretionary activity for each household member

$$\sum_{i,t:(i,j,t,s)\in E(p,a_q)} x^p_{i,j,t,s} \le 1, \forall\ p \in P_q \tag{4.6}$$

(6) Binary variable: $x^p_{(i,j,t,s)} = \{0,1\}$

The objective function is to minimize the total system travel cost of all household members, where the travel cost $c_{i,j,t,s}$ on each arc has been predefined in the space-time network construction stage. Eq. (4.2) is the standard person-based flow balance constraint. Eq. (4.3) means that each household member can only choose one vehicle or don't choose any vehicles (a.k.a. staying at home in our example). Eq. (4.4) represents that the activity duration arc of each mandatory activity of a specific household member should be visit exactly once by that household member. For example, if one household member must go to a company for work, one of working arcs must be visited exactly once by the household member. Eq. (4.5) ensures that if one household member needs to perform one type of activity with multiple candidate locations/time durations, he/she must choose one candidate to complete one activity instance among all options. For example, if one household member needs to go shopping and there are two candidate shopping malls, finally only one shopping mall should be visited exactly once to mark the completion state

of the shopping activity. Inequality (4.6) represents the flexibility associated with those optional activities, as they could be performed or not, depending on the availability of those eligible household members and the required travel cost to reach those locations. In short, the proposed model in this section is a 0-1 integer linear programing model, or more precisely, a multi-commodity flow optimization problem with a limited set of side constraints. This compact formulation enables the use of standard optimization solvers for a real-world transportation network.

Table 4-5 offers a systematic comparison for detailed modelling techniques between our proposed model and classical model proposed by Recker (1995), specifically between our Case A and Case IV of Recker.

4.3.2 Space-time-state Network-based Optimization Model for Case B

Before presenting the model for case B, it should be emphasized that the space-time-state network needs be pre-built and satisfies the given time windows of each activity and the predefined arc attributes, such as, the location of each node, the travel time or travel cost of each arc, and the logically feasible state transition in the three-dimension network.

More importantly, the slate of passengers' activity-performing states in the final solution for each vehicle exactly depends on the type of different activities, mandatory activity vs. discretionary activity. As shown in Figure 4-5(a) in section 4.2.3, when the two activities are mandatory for passenger 2, the super starting state at the origin and super ending state at the destination are [0,0] and [2,2], respectively.

Table 4-5 Comparison between Case IV (Recker, 1995) and Our Case A

Modelling constraints	Model R4: Case IV (Recker, 1995)	Model 1 for our Case A	Remarks
(1) Time representation	Continuous	Discretized Time-discretized space-time	
(2) Network representation	Abstract physical traffic network	physical traffic network	
(3) Objective function	Eqns (1a)-(1f) with multiple goals	Eqn (1) with travel cost only	
(4) Coupling constraints for vehicle selection of household member	Constraints (40a)-(40b)	Embedded in the modified physical network	
(5) Vehicle spatial connectivity constraints	Constraints (2), (3), (4'), (5') and (6)		
(6) Vehicle temporal constraints	Constraints (7)-(10)	Constraints (2)-(6) in the space-time network for modelling constraints (5)-(9)	Model 1 needs to build one specific activity duration link for each activity to represent the activity process
(7) Household spatial constraints	Constraints (26)-(30)		
(8) Household temporal constraints	Constraints (31)-(33)		
(9) Illogical activity constraints	Constraints (21)-(24) and (36)-(39)		
(10) Vehicle capacity constraints	Constraints (14)-(17)	Always satisfied (solo driving pattern)	
(11) Activity time window constraints	Constraints (11)-(13) and (34)-(35)	Embedded in the space-time network	Model R4 provides a starting time window and return-home window for each activity, but in Model 1 each activity only has a starting time window and does not have the return-home window. Instead, each household member has a return-home window for his/her arrival at home
(12) Travel cost/time budget constraint	Constraints (18)-(19)	Not considered but can be easily added	
(13) Variable definitional constraints	Binary and continuous variables	Binary variables only	Model R4 is a mixed integer linear programming model. Model 1 is a 0-1 integer linear programming model.

On the other hand, when only one of two activities needs to be executed in a daily schedule in Figure 4-5(b), the final arrival state could be [2,0] or [0,2], with a virtual ending state shown in Figure 4-8a). Similarly, if the two activities are optional, the final state could be

one of four possible alternatives [0,0], [2,0], [0,2] or [2,2], while the final selection of the optimal activity states is highly depending on the vehicle and time resources it consumed along the daily activity chain as well as the corresponding objective function in terms of benefit and travel costs.

To satisfy the flow balance constraint for a network flow programming model, we need to build a virtual super ending state, as shown in Figure 4-8(b), with connections from those possible ending states at the physical destination, e.g., four states in the above example, [0,0], [2,0], [0,2] and [2,2]. As a remark, there is no benefit/utility during the state transition to the virtual super ending state. This state-transition based modeling paradigm could systematically capture the complicated possible interactions between multiple household members in a daily scheduling process.

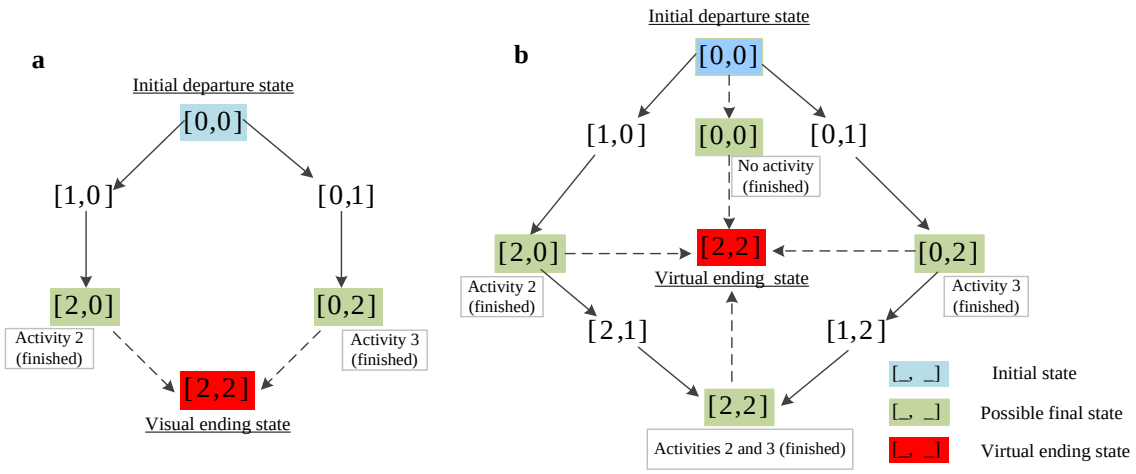


Figure 4-8 (a) One of Two Activities Is Performed; (b) Two Activities Are Optional

Based on the prebuilt 3D space-time-state network and given ride-sharing patterns, we now present our optimization model that satisfies all requirements in Case B that provides the

optimal vehicle route guidance to the driver(s) to enable the scheduling of everyone's activities.

Model 2:

Objective function

$$\min \sum_{v} \sum_{(i,j,t,s,w,w') \in E} (c^{v}_{i,j,t,s,w,w'} \times x^{v}_{i,j,t,s,w,w'}) \tag{4.7}$$

Subject to,

(1) Flow balance constraint for each vehicle:

$$\sum_{i,t,w:(i,j,t,s,w,w') \in E} x^{v}_{i,j,t,s,w,w'} - \sum_{i,t,w:(j,i,s,t,w',w) \in E} x^{v}_{i,j,t,s,w,w'} = \begin{cases} -1 & j = O(v), s = DT(v), w = [0,0,\ldots,0] \\ 1 & j = D(v), s = T, w = [2,\ldots,2] \\ 0 & otherwise \end{cases}, \forall v \tag{4.8}$$

(2) Mandatory activity performing constraint for the driver on the activity arcs (including ride-sharing):

$$\sum_{i,t,w:(i,j,t,s,w,w') \in E(v,a_m)} x^{v}_{i,j,t,s,w,w'} = 1, \ \forall v, \forall a \in A(v) \tag{4.9}$$

(3) Binary variable: $x^{v}_{i,j,t,s,w,w'} = \{0,1\}$

The objective function in Eq. (4.7) aims to minimize the total travel cost of the household, including the travel cost of vehicles and the benefit from everyone's performed activities. Eq. (4.8) is the standard vehicle-based flow balance constraint. With the given initial departure state $[0,0,\ldots,0]$ and virtual ending states for each vehicle, the given activity requirements of each passenger have been embedded in the space-time-state network.

Similar to Eq. (4.4), Eq. (4.9) ensures that the household driver can finish his/her mandatory activity with given time windows and time duration, which means that the activity duration arc of each mandatory activity should be visit exactly once by the driver/vehicle. The decision variable $x^{v}_{i,j,t,s,w,w'}$ is a binary variable that indicates whether or not the arc (i, j, t, s, w, w') will be chosen in the space-time-activity path of vehicle v. Finally, the model we proposed is also a 0-1 integer linear programing model, which has one more dimension compared to Case A but still can be directly solved in GAMS in a reasonable-size network.

In our Case B, the ridesharing pattern is prescribed, so the case can be viewed as a sub-problem of Case V in the paper (Recker, 1995). In Case V (Recker, 1995), it requires to build drop-off and pick-up nodes at each activity location and the set of available vehicles is expanded by designating driver seat and passenger seat(s) for each vehicle. The corresponding model has six categories of constraints, including vehicle temporal constraints, household member temporal constraints, vehicle spatial constraints, household member spatial constraints, vehicle capacity and budget constraints, and vehicle and household member coupling constraints. In our Case B, we also build drop-off and pick-up nodes for each activity with specific time windows. Since the ridesharing pattern is given a priori and modelled as a pair of drop-off-first then pick-up actions, we do not need identify the specific driver seat and passenger seat(s), and the coupling constraints for vehicle and household member is automatically coded through the state transition graph or explicitly taken as activity-performing constraints in Eq. (4.9). The temporal and spatial constraints of vehicle and household member are all embedded in the well-structured

space-time-state network where the state transition graph defines the possible activity visit sequences of passengers/vehicles.

It should be reminded that if we treat the start node and the end node of the activity duration link for the driver as a drop-off node and pick-up node, respectively, the driver's activities can also be added into the cumulative activity-performing state. As a result, side constraints (4.9) can also be embedded in the space-time-state network, and the mathematical model above is reduced to be a time-dependent state-dependent least cost path-finding problem, which could be efficiently solved by dynamic programming with parallel computing technology on large-scale networks. One multi-loop label correcting algorithm is designed in Appendix B.

As a remark, it is also possible to define another state instead of cumulative activity-performing state to model Case B. Based on the specific requirements in one problem, different state definitions could lead to different model formulations (less or more side constraints), different network structure and computation complexity. One specific example can be found in recent papers by Mahmoudi and Zhou (2016) and Mahmoudi et al., (2016) where they applied vehicle carrying state $\{0,1\}$ and vehicle cumulative service state $\{0,1,2\}$ to solve the VRPPDTWs, respectively, with different model formulation, networks, and algorithms. Therefore, our proposed formulation for Cases A and B is not the only possible modelling choice, and one should examine the size of state variables and nature of complex constraints to reformulate the problem based on the preferred network structure and available space and time complexity requirements.

4.3.3 Link Capacity Constraints of Case C

Since Case A considers a solo-driving pattern, one vehicle can only carry one person. In the mathematical model of Section 4.3.1, the person-based formulation is equivalent to the vehicle-based model. After converting the hourly road capacity into "specific time interval-based travelling arc capacity in the space-time network, the tight arc capacity constraint can be formulated as,

$$\sum_{p} x^{p}_{i,j,t,s} \le cap_{i,j,t,s}, \forall (i,j,t,s) \in E \tag{4.10}$$

To consider the "queue spillback" phenomenon, additional inequality needs be added to represent the link storage capacity constraint by using cumulative arrival counts and cumulative departure counts on that link. The detailed formulation can be found in the paper by Li et al. (2015).

Similarly, since the mathematical model of Section 4.3.2 is vehicle-based formulation, the tight capacity constraint can be formulated as,

$$\sum_{v} x^{v}_{i,j,t,s,w,w'} \le cap_{i,j,t,s,w,w'}, \forall (i,j,t,s,w,w') \in E \tag{4.11}$$

Regarding the queue spillback and congestion propagation property from Newell's simplified Kinematic wave model, the specific formulation is similar to the constraints in the paper by Li et al. (2015) which doesn't consider the merge and diverge issues, but with one more dimension *w*.

As stated at the end of Section 4.3.2, the driver's activity participation constraint can also be embedded in the space-time-state networks so that case B becomes a time-dependent

state-dependent least cost path-finding problem. When the road resource capacity constraint (4.11) is recognized, there are two research directions to solve our proposed system optimal problem for large-scale real-world applications:

(1) Lagrangian relaxation: the link/arc capacity constraints can be dualized to objective function (4.7), so a new time-dependent state-dependent least cost path problem is transformed in the Lagrangian relaxation framework to obtain a lower bound. Since the optimal sub-gradient in binary integer programming model is hard to be obtained, the gap between the lower bound and the optimal solution cannot be well analytically proved. Meanwhile, when more side constraints from queue spillback consideration are taken into account, dualizing those constraints might not be a suitable approach.

(2) Queue-based simulation: Since our proposed model is a system optimal problem considering complex traffic dynamics, we can apply event-based simulation to solve the large-scale problem where (i) the event-based simulation process is consistent with the time-discretized space-time-state network, (ii) different travel flow models can be handled, and (iii) the marginal cost analysis (Ghali and Smith, 1995) can be used to find the least marginal cost path for system optimal solutions. The specific algorithm design can refer to the paper by Lu et al (2016), which proposed a simulation framework to solve agent-based eco-system optimal traffic assignment in congested networks.

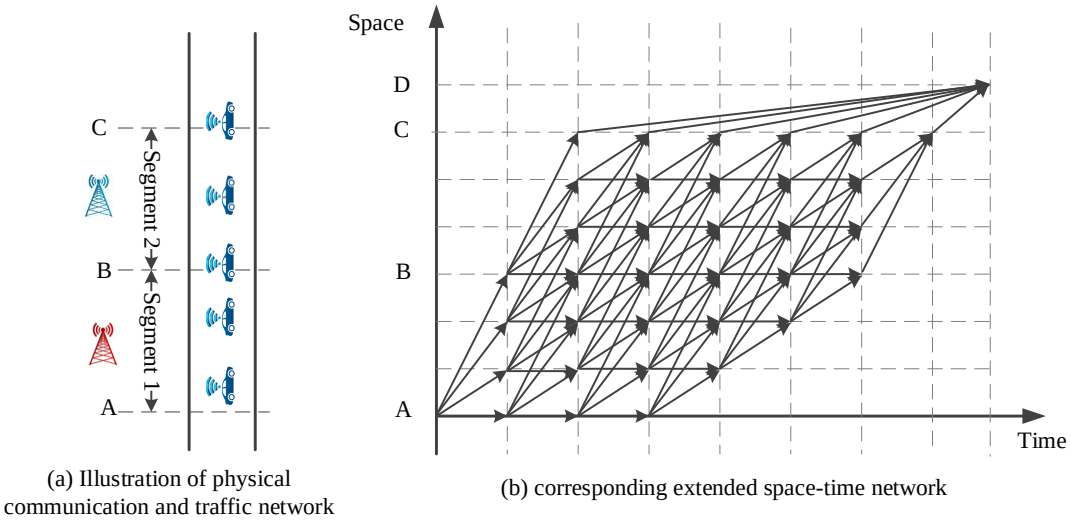
4.4 Longitude Trajectory Optimization for Autonomous Vehicles

4.4.1 Problem Statement

In this research, we only consider autonomous vehicles along a single-lane facility with a given communication system. As shown in

Figure 4-9(a), there are two segments AB and BC, and each segment has a communication system, which provides the reaction time of those autonomous vehicles or backward wave speed in those segments. The corresponding time-discretized space-time network is built in

Figure 4-9(b).



(a) Illustration of physical communication and traffic network

(b) corresponding extended space-time network

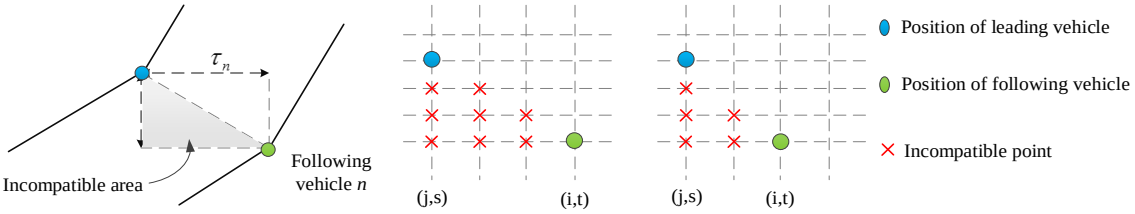
Figure 4-9 Illustration of Traffic Network and Its Space-time Network

At each time stamp along time horizon from 0 to *T*, one vehicle can travel in one of *K* (*K*=4) alternative speed values from 0 to its speed limit, which can be represented by travelling arcs and waiting arcs. The travel cost of each arc in the space-time network is defined as its travel time in advance. At each physical destination node, a corresponding virtual super-destination node is built at the big time T. The travel cost of arcs from the destination node to the virtual super node is 0. To avoid the obstacles such as the time period of traffic red signals, we can delete those travelling arcs which go through or depart at the signal red time at the signal location in the space-time network.

Figure 4-10(a) shows Newell's simplified car-following model, where the displacement of time and space establishes an incompatible area so that both vehicles cannot exist there simultaneously. The detail is further drawn in Figure 4-10 (b) when the reaction time and minimum spacing are 3-time interval and 3 space interval for the following vehicle, respectively.

The relation between the lead vehicle and the following vehicle is displayed in

Figure 4-10(c) shows the incompatible area with reduced displacement time from 3-time interval to 2-time interval. The set of incompatible points at each vertex can be enumerated based on the given reaction time or backward wave speed w in each segment with a built communication system. In short, our problem aims to optimize all autonomous vehicles' trajectories to minimize the total system travel cost while satisfying Newell's simplified car-following constraints.



(a) space and time displacement in Newell's car-following model

(b) restricted area between lead vehicle and following vehicle

(c) restricted area between lead vehicle and following vehicle with reduced displacement time

Figure 4-10 Illustration of Restricted Points Based on Newell's Car-following Model

4.4.2 Mathematical Programming Formulation

Table 4-6 lists the notation used in this model.

Table 4-6 Notation Used for This Section

Indices	Definition
i, j	Index of nodes, $i, j \in N$
(i, j)	Index of physical link between two adjacent nodes, $(i, j) \in L$
a	Index of agents/vehicles, $a \in A$
t, s	Index of time intervals in the space-time network
w	Index of communication time of autonomous vehicles
k	Index of segments with different communication time
$o(a)$	Index of origin node of agent a
$d(a)$	Index of destination node of agent a
Sets	
N	Set of nodes in the physical transportation network
L	Set of links in the physical transportation network
A	Set of agents
W	Set of communication time of self-driving cars
S	Set of segments
V	Set of vertices in the space-time network
E	Set of edges/arcs in the space-time network
Parameters	
DT^a	The departure time of agent a
AT^a	The assumed arrival time of agent a
$c_{i,j,t,s}$	Travel cost of traveling arc (i, j, t, s) in the space-time network
T	The time horizon in the space-time network
$\varphi_{(w,i,j,t,s)}$	$\varphi_{(w,i,j,t,s)} = 1$, if the vertex (j, s) visited by the following vehicle with reaction time w, the vertex (i, t) cannot be visited by its leading vehicle.
Variables	
$\theta_{i,t}^a$	Binary variable, indicator of vertex (i, t) visited by agent a
$x_{i,j,t,s}^a$	$= 1$, if Agent a is assigned on traveling/waiting arc (i, j, t, s) in the space-time network; $= 0$ otherwise

Integer Programming Model:

Objective function:

$$\min \sum_a \sum_{(i,j,t,s)} x_{i,j,t,s}^a \times c_{i,j,t,s} \tag{4.12}$$

Subject to,

Vehicle-based flow balance constraint:

$$\sum_{i,t:(i,j,t,s)\in E} x_{i,j,t,s}^a - \sum_{i,t:(j,i,s,t)\in E} x_{j,i,s,t}^a = \begin{cases} -1 & j = o(a), s = DT^a \\ 1 & j = d(a), s = T \\ 0 & otherwise \end{cases}, \forall a \tag{4.13}$$

Indicator of vertex visited by vehicles:

$$\theta_{i,t}^{a} = \sum_{(j,s)} x_{i,j,t,s}^{a}, \forall (i,t), \forall a \quad (4.14)$$

Simplified car-following safety constraints:

$$\sum_{(i,t)} (\varphi_{(w,i,j,t,s)} \times \theta_{i,t}^{a}) + \theta_{j,s}^{a+1} \leq 1, \forall (j,s) \in \varphi_{(w,i,j,t,s)}, \forall a \quad (4.15)$$

Binary variables: $x_{i,j,t,s}^{a} = \{0,1\}$;

The objective function in Eq. (4.12) is to minimize the total generalized travel cost of all autonomous vehicles under centralized control. Eq. (4.13) is a standard vehicle-based flow balance constraint, similar to recent studies by Liu and Zhou (2016) and Lu et al. (2016). Eq. (4.14) defines whether or not vehicle a visits vertex (i,t) by $\theta_{i,t}^{a}$. Specifically, if $\theta_{i,t}^{a} = 1$, vehicle a visits vertex (i,t) and $\sum_{(j,s)} x_{i,j,t,s}^{a} = 1$, which indicates that only one arc from vertex (i,t) is chosen. Otherwise, $\theta_{i,t}^{a} = 0$, and no arcs from vertex (i,t) will be chosen by vehicle a. Inequality (4.15) represents the safe driving constraints of a pair of lead and following vehicles based on Newell's simplified car-following model. $\varphi_{(w,i,j,t,s)}$ is a parameter with value of 1 that defines the incompatible relation among vertexes (i,t) visited by the lead vehicle and the vertex (j,s) visited by the following vehicle under given reaction time/backward wave speed w at different road segments. As a result, our proposed model is a 0-1 integer linear programming model, which could be directly solved by standard optimization solvers, such as CPLEX.

As a remark, traffic boundary condition (closed or semi-open) is an important input for autonomous vehicle trajectory control/optimization. Our proposed model can handle not only the closed boundary condition in the space-time network but also the semi-open

boundary condition through building a virtual super-destination node with the objective of travel time, fuel consumption, emissions, etc.

4.5 Numerical Experiments

4.5.1 Small-scale Experiment for Case A

The proposed model for Case A in Section 4.3.1 will be tested in the following network shown in Figure 4-11(a), where there are two household members p_1 and p_2, two available vehicles v_1 and v_2, and four candidate activities a_1, a_2, a_3 and a_4. Household member p_1 can choose any one of the two vehicles, and has one mandatory activity a_1 to meet with others and one optional activity to swim. Household member p_2 can only choose v_2 and will go to one of the two shopping malls. The corresponding modified network is constructed in Figure 4-11(b) where nodes 1 and 2 are origin nodes, nodes 3 and 4 are vehicle nodes, and node 5 is the final destination node. It is observed from the activity links that the time durations and costs for performing activities 1 to 4 are (60, -20), (30, -10), (30, -15), and (20, -20), respectively. The specific time windows are listed in Table 4-7. The waiting cost at each time interval is 0 at origin and destination nodes and 1 at activity nodes.

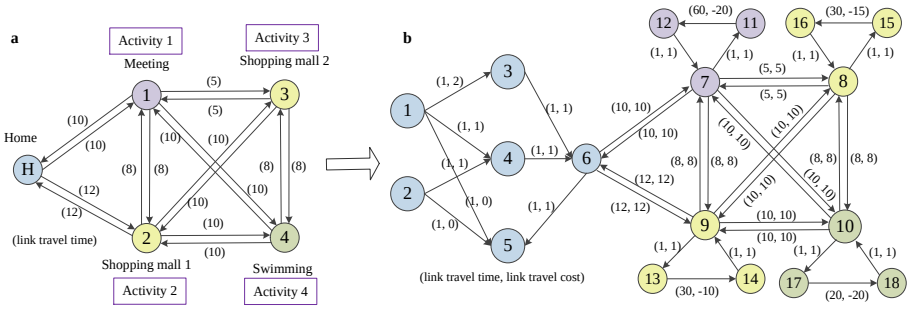


Figure 4-11 (a) Physical Network; (b) Corresponding Modified Network

Table 4-7 The Specific Time Window for Each Event

Location (node)	1	2	5	11	13	15	17
Time window	[1, 3]	[1, 3]	[1, 130]	[15, 18]	[15, 18]	[18, 20]	[86, 90]

Our proposed 0-1 integer linear programming model for this example is solved in GAMS. The related source code can be downloaded at the website: https://www.researchgate.net/publication/306459026_Experiment_1_1. Finally, the total travel cost of this household is 24. The specific optimal solution is listed in Table 4-8, and can be also illustrated in Figure 4-12.

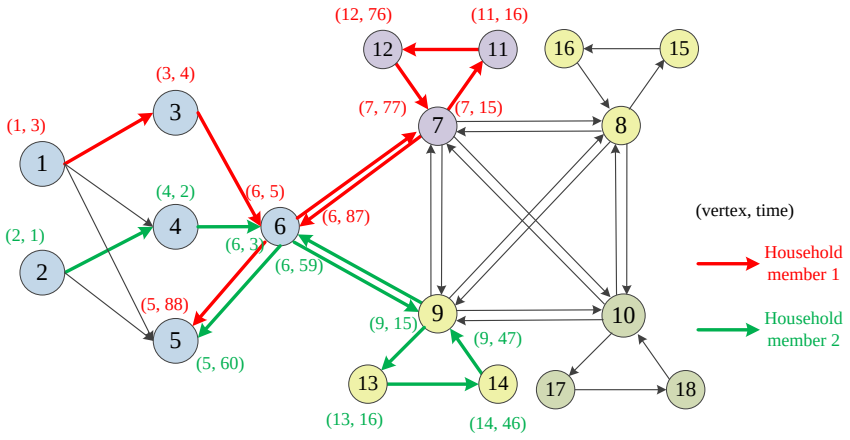


Figure 4-12 The Trajectories of Two Household Members

It is observed that p_1 should not go to activity 4 (swimming) and p_2 does not need to go to activity 3 (shopping mall 2) due to the trade-off between the required travel costs and corresponding activity benefits. Therefore, if we increase the benefits of activities 3 and 4 to 17 and 23, respectively, the optimal solution will be that (i) the total cost is 22, (ii) household member p_1 will visit activities 1 and 4 sequentially and then go back home, and (iii) household member p_2 will visit activity 3 (shopping mall 2) rather than activity 2.

Meanwhile, if we assume that the link travel time increases due to tight link capacity constraints when more other household activity trips are considered, the activity pattern of this household is expected to change again. In short, the final activity selection and route guidance are comprehensively evaluated and selected based on the possible time-varying travel cost in the physical network, available time windows, and the benefits of performing individual available activities.

Table 4-8 The Optimal Solution for Each Household Member

Household member p_1: $x^1_{i,j,t,s} = 1$				Household member p_2: $x^2_{i,j,t,s} = 1$			
i	j	t	s	i	j	t	s
1	3	3	4	2	4	1	2
3	6	4	5	4	6	2	3
6	7	5	15	6	9	3	15
7	11	15	16	9	13	15	16
11	12	16	76	13	14	16	46
12	7	76	77	14	9	46	47
7	6	77	87	9	6	47	59
6	5	87	88	6	5	59	60

4.5.2 Small-scale Experiment for Case B

This section will test our proposed model for Case B in Section 4.3.2 based on the network shown in Figure 4-13(a), where there are three household members with one driver and two passengers. They will share one vehicle to perform their daily activities. The driver p_1 has one mandatory activity a_1 and needs to drop off and pick up two passengers to conduct their activities. Passenger p_2 has one mandatory activity a_2 and one optional activity a_3, and passenger p_3 has one mandatory activity a_4. The corresponding modified network is plotted in Figure 4-13(b) where the activity of the driver is represented by one specific activity link and each activity node of passengers is added with two additional nodes as drop-off node and pick-up node.

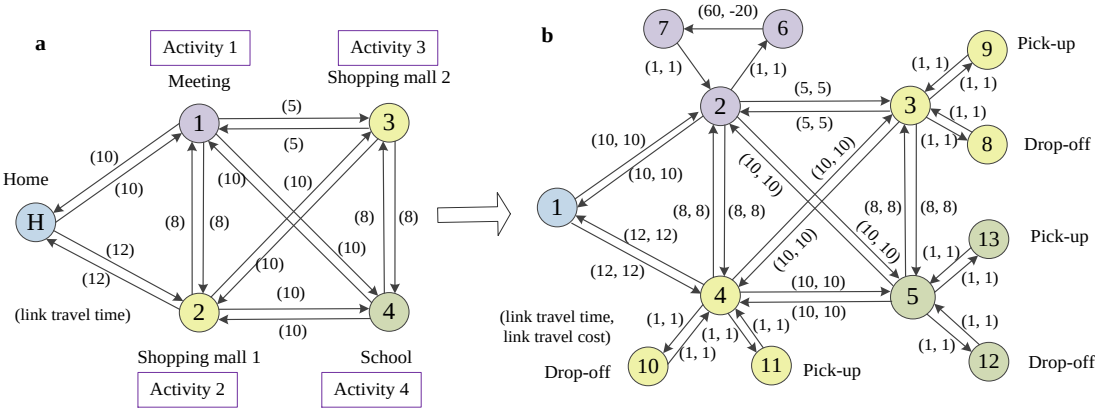


Figure 4-13 (a) Physical Network; (b) Corresponding Modified Network

Table 4-9 Enumeration of All Possible States

State ID	State representation	State ID	State representation	State ID	State representation
1	[0, 0, 0]	9	[0, 0, 1]	17	[0, 0, 2]
2	[1, 0, 0]	10	[1, 0, 1]	18	[1, 0, 2]
3	[2, 0, 0]	11	[2, 0, 1]	19	[2, 0, 2]
4	[0, 1, 0]	12	[0, 1, 1]	20	[0, 1, 2]
5	[0, 2, 0]	13	[0, 2, 1]	21	[0, 2, 2]
6	[2, 1, 0]	14	[2, 1, 1]	22	[2, 1, 2]
7	[1, 2, 0]	15	[1, 2, 1]	23	[1, 2, 2]
8	[2, 2, 0]	16	[2, 2, 1]	24	[2, 2, 2]

Based on the procedure explained in Section 4.3.2, the state has three slots as [_, _, _], of which the first two slots are for activities 2 and 3 of passenger p_2 and the last slot is for activity 4 of passenger p_3. We still use cumulative activity-performing state of {0,1,2} as before. All possible states that can be generated by algorithms are listed in Table 4-9.

In addition, as noted in Section 4.2.2, specific time windows can also eliminate those impossible transitions. In this example, the time window for each event is listed in Table 4-10.

Table 4-10 The Specific Time Window for Each Event

Location (node number)	Node 1 (departure)	Node 1 (arrival)	Node 10	Node 11	Node 12	Node 13	Node 8	Node 9	Node 6
Time window	[1, 3]	[1, 170]	[15, 16]	[114, 115]	[28, 30]	[127, 139]	[127, 129]	[137, 139]	[41, 43]

Based on the time window information, it is impossible that (i) activity 3 happens before activity 2, (ii) the drop-off event and pick-up event of activity 4 happens before those of activity 2, respectively, and (iii) the drop-off event and pick-up event of activity 3 happens before those of activity 4. Therefore, the remaining possible states will be [0, 0, 0], [1, 0, 0], [1, 0, 1], [2, 0, 1], [2, 0, 2], [2, 1, 1], [2, 1, 2], and [2, 2, 2]. For the convenience of implementation in algorithms, we can label each state with one corresponding ID, such as, using 1 to 8 to represent the eight states above sequentially. The final possible state transition is demonstrated in

Figure 4-14, where virtual arcs with virtual ending state are also built for developing a single-origin-to-single-destination problem.

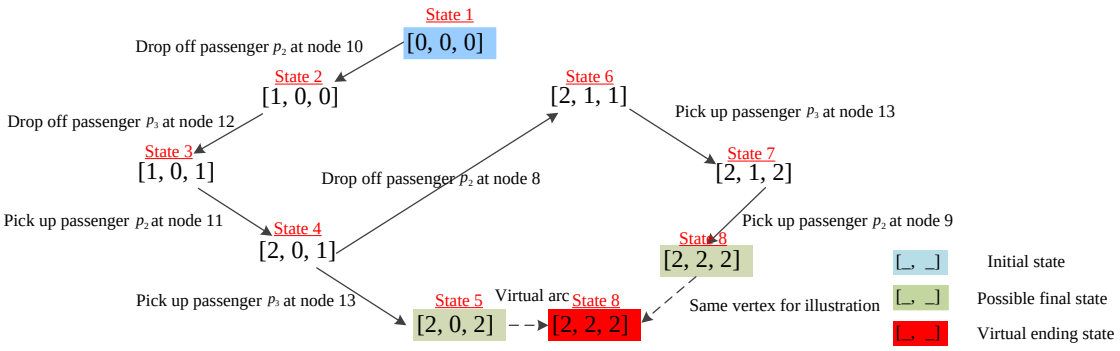


Figure 4-14 The Possible State Transition Graph

In short, the possible state transition can be reduced with consideration of time windows before constructing a space-time-state network. Meanwhile, it is also feasible and straightforward to list all possible state transition without involving time windows and then consider the relation among activity locations, time windows, and state transition to construct the space-time-state network directly. As a result, those impossible state transition can also be eliminated when solving our mathematical model, but the price of this method is to require a bigger computer memory to store all possible state transition and a more complex space-time-state network.

In addition, the benefit or negative cost of performing activity for all passengers is assumed to occur during the state transition at drop-off nodes, as illustrated in Section 4.2.3. The negative travel costs for activities 1,2, 3, and 4 are given as -20, -10, -15, and -15. Based on the constructed space-time-state network, our proposed 0-1 integer linear programming model for this example is solved in GAMS. The related source code can be downloaded at the website: https://www.researchgate.net/publication/306458887_Experiment_2_1. Finally, the total travel cost of this household is 40. The specific optimal solution is listed in Table 4-11.

It is observed that passenger p_2 will not perform activity 3 due to the trade-off between the required travel costs and corresponding activity benefits. If we increase the benefit of activity 3 from 15 to 20, the optimal solution will change to be that (i) the total cost is 37, and (ii) activity 3 will be performed by passenger p_2. Moreover, when the link travel time is modelled as a time-dependent attribute due to road congestion effect, the final household activity pattern is expected to change accordingly.

Table 4-11 The Optimal Solution for the Household

The only vehicle: $x^1_{i,j,t,s,w,w'} = 1$													
i	j	t	s	w	w'	Remarks	i	j	t	s	w	w'	Remarks
1	4	3	15	1	1	Depart at home at time 3	7	2	103	104	3	3	
4	10	15	16	1	1		2	4	104	112	3	3	
10	10	16	16	1	2	State transition (passenger p_2 is dropped off at node 10 for activity 2)	4	4	112	113	3	3	
10	4	16	17	2	2		4	11	113	114	3	3	
4	5	17	27	2	2		11	11	114	114	3	4	State transition (passenger p_2 is picked up at node 11 for activity 2)
5	12	27	28	2	2		11	4	114	115	4	4	
12	12	28	28	2	3	State transition (passenger p_3 is dropped off at node 12 for activity 4)	4	5	115	125	4	4	
12	5	28	29	3	3		5	5	125	126	4	4	
5	2	29	39	3	3		5	13	126	127	4	4	
2	2	39	40	3	3		13	13	127	127	4	5	State transition (passenger p_3 is picked up at node 13 for activity 4)
2	2	40	41	3	3		13	5	127	128	5	5	
2	6	41	42	3	3		5	2	128	138	5	5	
6	6	42	43	3	3		2	1	138	148	5	5	Arrive at home at time 148
6	7	43	103	3	3	The driver p_1 performs activity 1	1	1	148	149	5	8	State transition (from final state to assumed final state, the virtual arc cost is 0)

4.5.3 Medium-scale Experiment Within a Lagrangian Relaxation Framework Using Cumulative Activity-performing State

This section aims to examine the computation efficiency of using cumulative activity-performing state in a medium-scale transportation network. We choose a subarea of

Phoenix regional network as our study case with 1186 nodes, 3164 links and 387 activity locations, shown in Figure 4-15. The given input data for this experiment are listed in Table 4-12.

Table 4-12 The Input Data of This Experiment

agent_id	agent_type	from_node_id	to_node_id	departure_time_start	departure_time_window	arrival_time_start	arrival_time_window	base_profit	optional
1	0	23	23	30	5	40	5	150	0
2	0	24	24	10	20	70	10	133.33	0
3	0	26	26	40	10	60	5	83.33	0
4	0	25	25	20	20	80	5	150	0
5	0	39	39	70	5	90	5	133.33	0
6	0	35	35	20	5	110	5	183.33	0
7	0	38	38	35	10	120	5	133.33	1
Veh 1	1	13	13	1	1	120	1		
Veh 2	1	13	13	1	1	120	1		

“agent_id” could be activity id or vehicle id. “agent_type” = 1 for vehicles, and 0 means activities. Field “from_node_id” and “to_node_id” are the same and define (i) the activity performing location or (ii) vehicle’s origin/destination (home). “departure_time_start” defines the start time of activity or vehicle departure, and “depature_time_window” is the feasible time window duration. “arrival_time_start”, and “arrival_time_window” defines the activity/vehicle end time window. “base_profit” is the benefit/utilities of performing the corresponding activity. The “optional” flag indicates that if an activity is optional, its value is 1, otherwise it is mandatory as 0. As a result, the problem becomes that two vehicles at home (node 13) plans to perform 6 mandatory activities and 1optional activity.

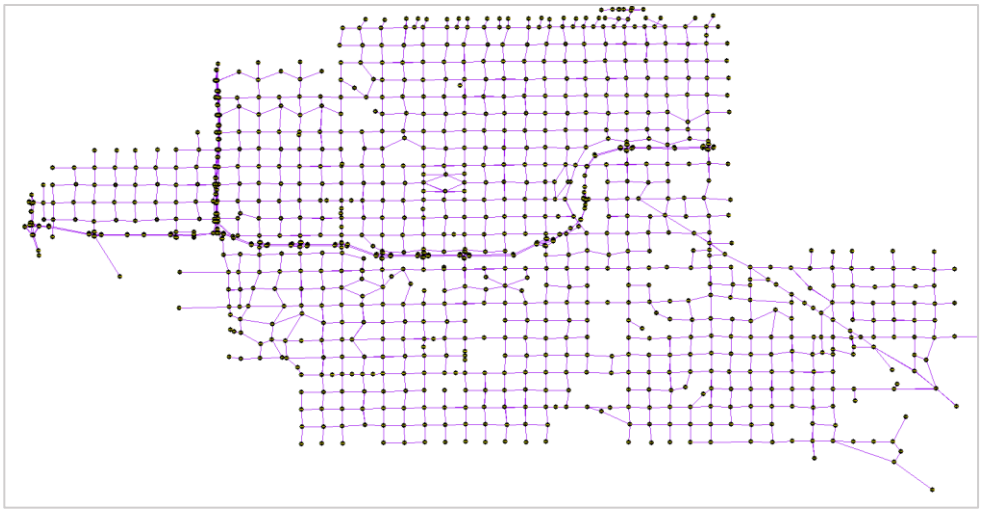


Figure 4-15 One Subarea of Phoenix Regional Transportation Network

To solve this problem, we use cumulative activity-performing state {0,1,2} to record the activity completion process. It is reminded that the maximum number of possible states could be 3^7 for 7 activities. In order to model the competition for one activity by two vehicles simultaneously, we dualize that constraint to our objective function and adopt the forward dynamic programming algorithm within a Lagrangin relaxation framework, which can refer to the process of solving the VRPPDTW for multiple vehicles by Mahmoudi and Zhou (2016). The related C++ source code and data set can be downloaded at the website: https://github.com/xzhou99/Agent-Plus/tree/master/HAPP. Table 4-13 lists the impact of different number of activities on the CPU computation time of 5 Lagrangian iterations (for distributing different tasks to two vehices) and computer memory usage. In the above case, the vehicle/activity preference for household members is not considered. If a pre-specified vehicle-to-activity mapping is given, the search space in the space-time-state network could be further reduced.

Table 4-13 CPU Computation Time and Memory Use under Different Number of Activities

# of activities	Maximum numbers of activity performing states	CPU time (seconds)	RAM (GB)
4	81	15.5	0.3
5	243	38.2	1.3
6	729	112.3	3.6
7	2187	337.4	11.3

4.5.4 Large-scale Experiment Within a Simulation-based Framework with Simplified Activity Representation and Road Capacity Constraints

This section aims to present the initial test result of the simulation-based approach for system optimal dynamic vehicle routing under road capacity constraints. The Salt Lake City regional traffic network is selected shown in Figure 4-16 where there are 13,923 nodes, 26,768 links and 2,302 zones. The total number of simulated vehicles is about 1.35 million from 15:00 to18:00. The traffic flow model chooses point queue model, which just considers the tight road capacity constraints. The details of implementing spatial queue model and Newell's simplified kinematic wave model by simulation can be found in the paper (Zhou and Taylor, 2014).

This experiment can be treated as a special version of Case A. Each origin zone is analogous to one household and those destination zones can be viewed as those mandatory activity locations. The process that vehicles depart from origin to destination is like that household members complete their mandatory activities with flexible time windows. The simulated average trip time index (mean trip simulated travel time/trip free-flow travel time) of 100 iterations is depicted in Figure 4-17 and finally shows a convergence pattern.

Figure 4-16 Salt Lake City Regional Traffic Network (Lu et al., 2016)

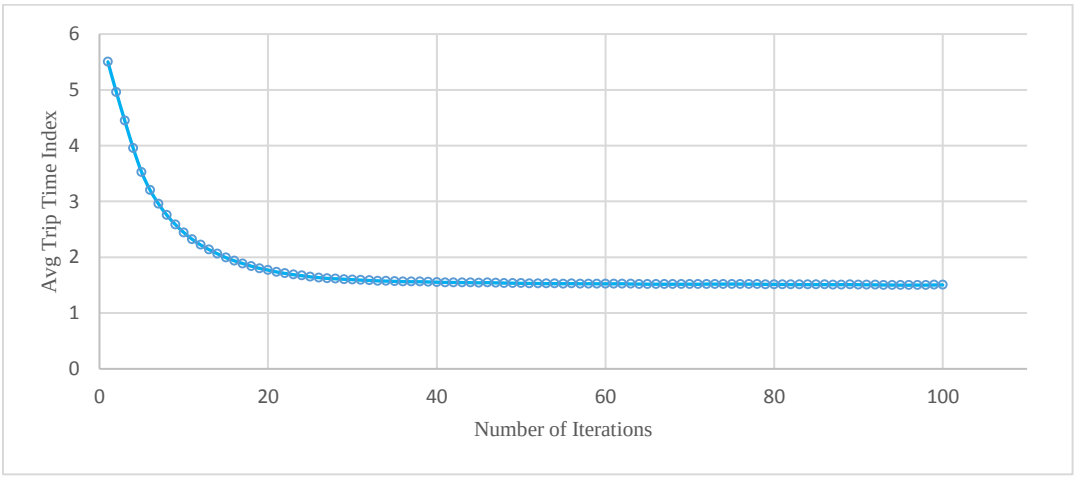


Figure 4-17 Average Trip Time Index of Each Iteration

A parallel computing technique (Qu and Zhou, 2017) is embedded in the simulation process, and the search process for single activity is extremely simple compared to the full scale space-time-state search presented in the medium-scale example, so the computational time for one iteration is just 1 min 25sec in our workstation with 40 available CPU threads and 192G memory. As stated in the paper (Lu et al., 2016), this simulation algorithm still

needs further improvements on path marginal travel time calculation and step size optimization of each iteration.

4.5.5 Tests on the Trajectory Control of Multiple Autonomous Vehicles

The proposed mathematical model in section 4.5.4 is applied in the following tests, where autonomous vehicles move along a one-lane roadway with three segments (AB, BC, and CD) and total 30 space units (e.g., 2 or 3 meters). Each segment has one specific communication facility, which indicates that each vehicle can only have one particular reaction time to its leading vehicle. In these tests, the total time horizon is 40 time intervals (e.g., 20s or 40 s). Each vehicle is assumed to have four speed values to be selected at each time stamp, including 0, 1 space unit, 2 space units, and 3 space units. The minimum spacing d_{jam} between two autonomous vehicles is assumed as 2 space units. The reaction and operation times w of segments AB, BC, and CD are 1 time unit, 2 time units and 3 time units, respectively. Tests 1 and 2 focus on two autonomous vehicles in order to perform the sensitivity of departure time with the impact on system-wide travel times. In all tests, all vehicles depart at node/space 1 and should arrival at node 30.

When departure times of vehicles 1 and 2 are 1 and 4, respectively, the problem is directly solved in GAMS based on our proposed mathematical model. Figure 4-18(a) shows the optimized vehicles' trajectories. It is expected that vehicle 1 can always move in speed limit as 3 space units at each time interval in this discretized space-time network. However, the vehicle reduces its speed at time 8, because if it moves in speed limit, it will reach at node/space 31 after 10 time units rather than node 30 as the destination due to the discretization property of space-time networks. The red rectangle defines the incompatible

zone for the following vehicle (vehicle 2) when its reaction time is 3 time units. As shown in Figure 4-18(a), at the beginning, vehicle 1 stays at the incompatible zone of vehicle 2 if vehicle 2 has communication time of 3 time units, which indicates that vehicle 2 should reduce its speed to ensure there is no conflict with vehicle 1 after node 20 (at segment CD). As a result, the total system travel time is 21 where vehicles 1 and 2 have travel time of 10 and 11, respectively. The shock wave is not obviously found in the result. One possible reason is that the solution above is just one of multiple optimal solutions for our system optimal problem, so vehicles could reduce or increase its speed at different locations with a same minimum system travel time. In other words, once all vehicles can be controlled as autonomous vehicles, vehicles can reduce its speed at any locations before the bottleneck and then drive with a higher speed to pass the following roadway, and it is possible that traditional shock wave happened at bottlenecks will not be obviously observed in future.

In Test 2, the departure time of vehicle 2 is changed to be 3. Then the final optimization result from GAMS is shown in Figure 4-18(b). It can be also observed that vehicle 2 needs to reduce its speed to accommodate the high communication time at segment CD. The total travel time will increase to 22 instead of 21 in Test 1.

Further, we consider Test 3 with 5 autonomous vehicles, whose departure times are at 1st, 3rd, 6th, 9th, and 11en time interval. The optimization result from GAMS (with CPLEX solver) shows that vehicle 5 waits and moves with a high speed at segment AB. As remarked in Test 2, the optimization result could be one of multiple optimal solutions, so vehicle 5 can also drive with medium speeds at segment AB without violating car-following constraints. The model statistics of three tests, including number of equations,

number of variables, optimal objective value, and computation time, are listed in Table 4-14.

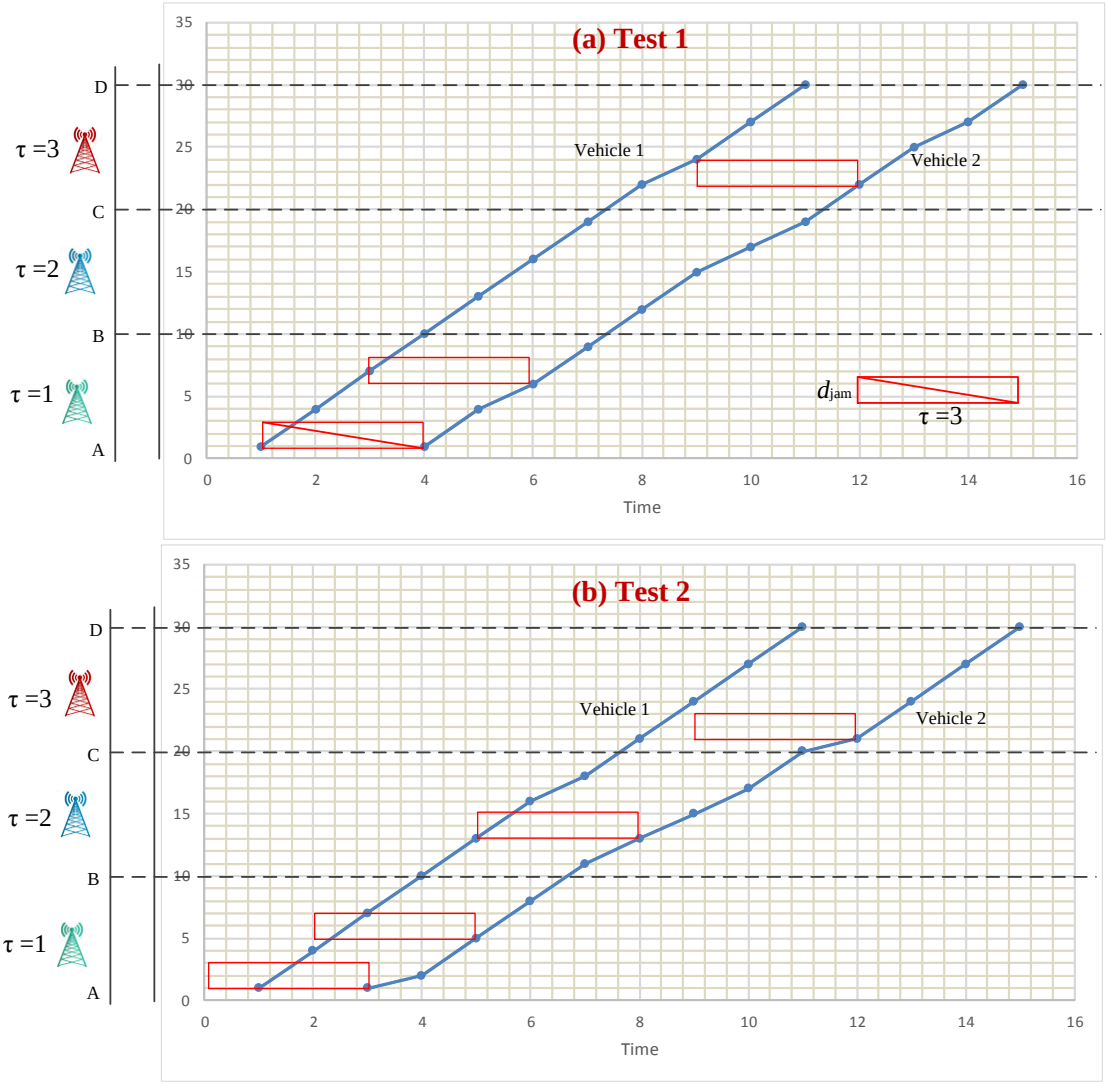


Figure 4-18 Optimized Vehicle Trajectories of Tests 1 and 2 in Discretized Space-time Grid from Integer Programming Model

Table 4-14 Model Statistics of Three Tests with 30 by 40 Space time grid Based on GAMS with CPLEX Solver

	number of equations	number of binary variables	objective value (time unit)	computation time (second)
Test 1 (2 vehicles)	6,083	8,972	21	1.3
Test 2 (2 vehicles)	6,083	8,972	22	1.4
Test 3 (5 vehicles)	17,006	22,430	65	2.8

CHAPTER 5

OBSERVABILITY QUANTIFICATION OF DYNAMIC URBAN TRANSIT SYSTEMS

5.1 Introduction

The currently rapid innovations and developments of transportation system intelligence in multi-source sensing and information sharing continuously generates huge volumes of various data and information for planners and managers to better observe time-varying traffic conditions and accordingly propose adaptive travel demand management and supply (capacity) control strategies. On the other hand, in face of the overwhelming data, the leverage of data on supporting system operational decisions should be carefully examined. Three fundamentally key questions are following: (1) Is the big data useful enough? (2) Under what goals, one kind of data is more useful that others? (3) How to fuse multi-source data to model your interested systems? In this chapter, those questions will be systematically addressed from the perspective of state estimation in urban rail transit systems. The usefulness of data will be reflected by quantifying the uncertainty of correspondingly estimated system states.

Specifically, the urban transit has been generally acknowledged as a major mode to relieve traffic congestion and air pollutions, so it is crucial to improve its service level by reliably estimating and predicting network-wide system conditions. In addition, with increasingly available data and information from heterogeneous sources in systems, especially in OD trip information and vehicle scheduling records, it provides us an opportunity to deeply

dissect the inner mechanism of system state uncertainty for better evaluating the current operations and optimizing the future strategies.

5.2 Conceptual Illustration

Through applying some concepts from game theory and control theory into our problem (LaValle, 2012), a **state space** is defined as the all possible internal system state based on the external physical transportation world, and an **information space** is a place where the internal states live when available information is involved. A **state** is specifically defined and can be associated with the available information. As shown in Figure 5-1, the information space is formed by the available information, and the states (1-d state and H-d states) are well connected by different projection functions, which mathematically define the states according to the managers' needs. The bound among all possible states represents the state uncertainty under current available information.

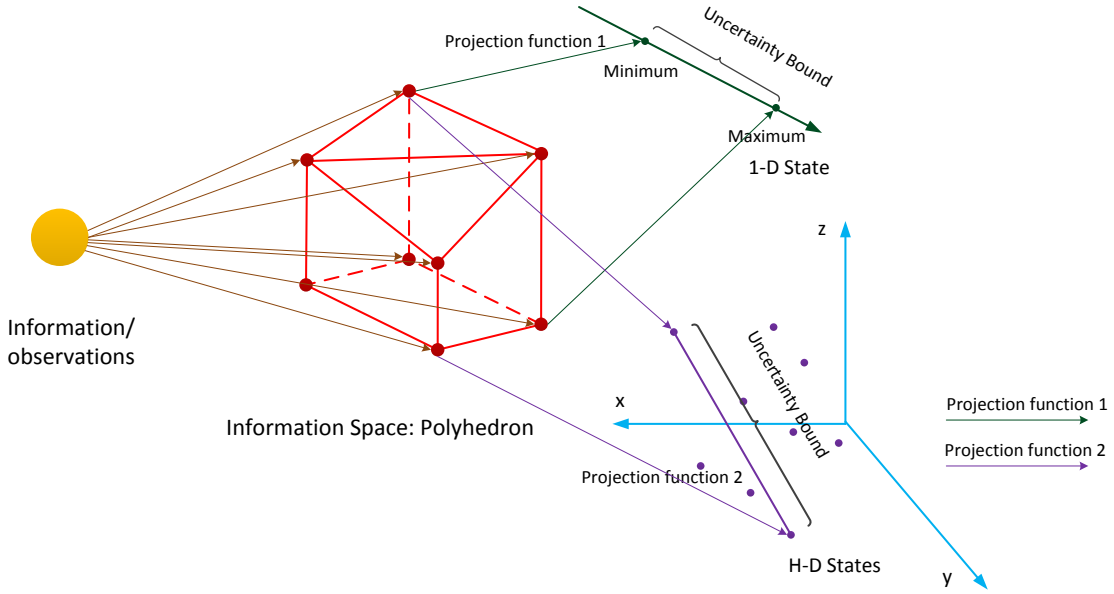


Figure 5-1 Relation among Information, Information Space, and Flexible States

When the information space is generated to be one single point, it is said that the system is observable; otherwise, it is unobservable or partially observable. In this chapter, we aim to build the connection between the internal states and the external states (observation or information) by information space as a bridge or communication channel, and further quantify the corresponding uncertainty of states defined by users. How to design sensor network design to change information space and further increase the state estimation accuracy is out of this paper's scope.

For illustrative purpose, Figure 5-2 (a) depicts a simple transportation network with four nodes and five links. The link travel time and capacity are also given as physical network attributes. Let x_1, x_2 and x_3 represent the path flow on paths 1, 2 and 3. Based on the tight capacity constraints, the following relation can be obtained: $0 \leq x_1 \leq 2$, $0 \leq x_2 \leq 3$, and $0 \leq x_3 \leq 1$, which defines the system state space shown as a blue cuboid in Figure 5-2(b).

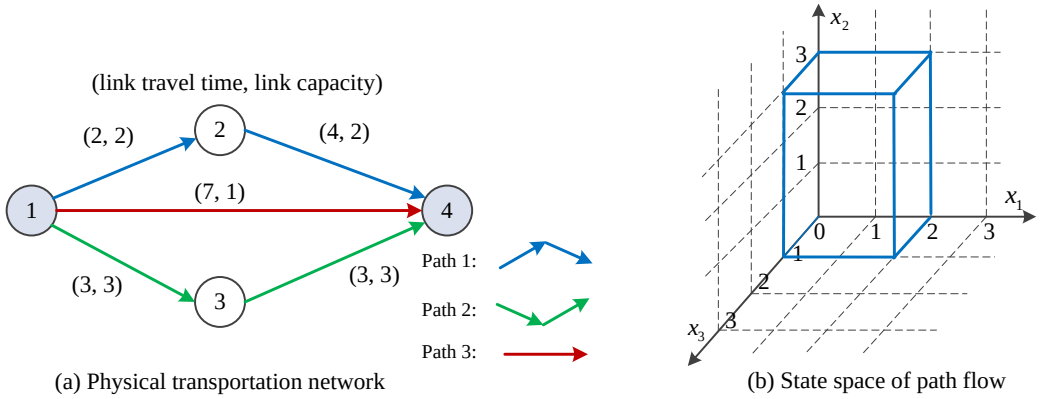


Figure 5-2 An Illustrative Transportation Network and Its State Space

Suppose the OD information is available through survey that there are four vehicles departing from node 1 to node 4. Then, one corresponding constraint will be $x_1 + x_2 + x_3 = 4$. Figure 5-3(a) displays the information space as the intersection of the red triangle

and the blue cuboid based on the available OD information. Two scenarios are designed as follows to analyze the relation between system state and information.

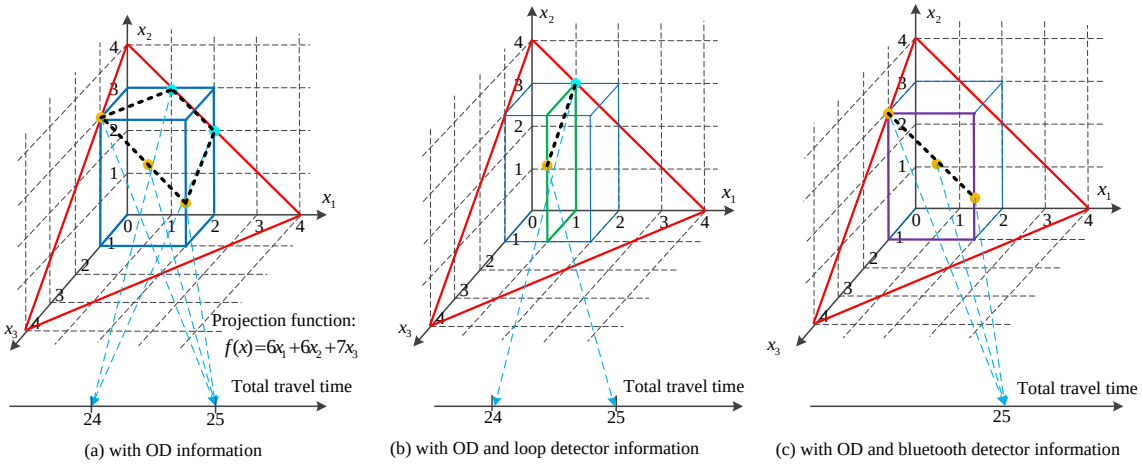


Figure 5-3 Information Spaces and Its Projected Bound under Different Available Information

Scenario 1: Assume that there is one flow count detector on link $(2, 4)$ and its link count is 1. The relation gets updated as follows: $1 + x_2 + x_3 = 4$, $0 \leq x_2 \leq 3$, and $0 \leq x_3 \leq 1$, so the corresponding information space is reduced to be the intersection of the red triangle and the green rectangle shown in Figure 5-3(b), which implies that different system states can reach one same observation. Actually, this many-to-one mapping is the main source of state uncertainty. Of course, the observation/measurement noise is the other source that cannot be ignored to address the state uncertainty, which will be studied in Section 5.4. From Figure 5-3(a) and Figure 5-3(b), it can observe that the information space is reduced when the link count information is provided, which means that the link count information is useful and the value of information is the difference of the two information spaces. However, if one more loop detector is added on link $(1, 2)$ and its link count is 1 as well,

the information space doesn't change, which indicates that the power of big data is on its information's usefulness rather than its volume size.

Scenario 2: Suppose that the automatic vehicle identification (AVI) detectors are available at nodes 1 and 4. One vehicle's travel time is observed as 7min. Since only path 3's travel time is 7 and its capacity is just 1, it implies that path flow $x_3 = 1$. As a result, the relation changes as follows: $x_1 + x_2 + 1 = 4$, $0 \leq x_1 \leq 2$, and $0 \leq x_2 \leq 3$. The corresponding information space becomes the intersection of the purple rectangle and the red triangle displayed in Figure 5-3(c).

As shown in Figure 5-3, the information spaces are generated as polyhedrons based on different available information. A projection function is defined to map the information space into one dimension state (total travel time). Actually, the projection functions could be different goals interested by the managers based on the real-world needs in transportation planning, operation, and controls, such as, the total system travel time, the number of vehicles in one area, etc. Finally, we can possibly find one bound as the measure of the information space mapped in that dimension. In Fig. 5-3, one projection function is defined as $f(\boldsymbol{x}) = 6x_1 + 6x_2 + 7x_3$, which means that the total system travel time is the analysis goal. Then, different optimization models are solved by maximizing and minimizing $f(\boldsymbol{x})$ subject to different information spaces.

As demonstrated in Figure 5-3, there are five feasible integer solution in the information space in Fig. 5-3(a), and the bound of total travel time formed by projection is $[24,25]$. When the link count data is added in Fig. 5-3(b), the information space by integer solutions is reduced, but the projected bound is still same. It indicates that the new information from

link count doesn't contribute to reduce the uncertainty of this state estimation, even a smaller information space is generated. In addition, the point-to-point Bluetooth data (end-to-end passenger id detector data) in Fig. 5-3(c) makes the projected bound converged to be one unique value, which implies that the point-to-point data is more powerful than the loop detector for quantifying the uncertainty of system travel time in this case. Therefore, evaluating the values of different information should be based on which states you really care. One information that seems worthless for your current goal may be much useful for other state estimations, and the volume of information space is not the criteria to judge the bound of state estimate uncertainty.

Except for those information from physical sensors above, the previous travel experiences or currently published traffic information from transportation agencies could also take important roles in quantifying the state uncertainty for managers' further actions. For example, if everyone has a perfect information over the network attributes based on their experiences and each one aims to find the best route for his/her trip, which is usually entitled Wardrop's first principle, the information space will be redefined as, $x_1 + x_2 = 4, 0 \leq x_1 \leq 2, 0 \leq x_2 \leq 3$. Compared with the two scenarios above, the information space gets further reduced by this assumed travel behavior. Therefore, one accurate travel behavior can also provide much rich information to determine system states, and that is why a number of studies focus on travel behavior estimation to better understand the system.

As a remark, the information space concept in game theory and control theory for storing the amounts of ambiguity in state also may have a connection with Shannon's information

theory using entropy-based constructs. For example, the entropy-maximization approach is usually used to represent the most likely traffic state, a kind of stable equilibrium state, for origin-destination travel demand estimation problems. However, in information space theory, the system state is unknown and gradually constructed based on available information. It is possible that the system state derived from the information space is finally same as the equilibrium state based on maximum entropy.

5.3 Problem Statement

Table 5-1 lists the general indices, sets, parameters and variables in our proposed models appeared in Sections 5.3 and 5.4.

5.3.1 Space-time Network Construction in Public Transit Systems

Consider a physical transit network with a set of nodes (stops/stations) N and a set of links L. Each link can be denoted as a directed link (i, j) from upstream node i to downstream node j. A deterministic transit schedule is supposed to be obtained from Automatic Vehicle Location (AVL) data from train tracking systems. We then construct a space-time network, where V is the set of vertices and E is the set of arcs. Node i is extended to a set of vertices (i, t) at each time interval t in the study horizon, $t = 1,2, \dots, T$, where T is the length of the optimization horizon. The transit schedule from node i to node j from time t to time s can be represented by a travelling arc (i, j, t, s) where $(s - t)$ is the exact scheduled/running link travel time and should be integer multipliers of one time interval. The capacity of travelling arcs can be viewed as the transit vehicle's carrying capacity. In addition, waiting arc is built from (i, t) to $(i, t + 1)$ at node i with waiting time of 1 time unit and its capacity is defined as the station/platform storage capacity.

Table 5-1 Indices, Sets, Parameters and Variables

Indices	Definition
i, j	Index of nodes, $i, j \in N$
(i, j)	Index of physical link between two adjacent nodes, $(i, j) \in L$
a	Index of passenger group, $a \in A$
$o(a)$	Index of origin node of group a
$d(a)$	Index of destination node of group a
t, s	Index of time intervals in the space-time network
τ	Index of time period for the observed passenger flow
p	Index of paths, $p \in P$
r	Index of transit companies
Sets	
N	Set of nodes in the physical transit network
L	Set of links in the physical transit network
A	Set of passenger groups
V	Set of vertices in the space-time network
E	Set of edges/arcs in the space-time network
G	Set of time period for the observed passenger flows
$S_{p,a}$	Set of paths p of group a
$G(i, j, \tau)$	Set of arcs on observed link (i, j) at time period τ
Parameters	
β_1, β_2	The weights on target passengers' travel time and observed link/arc flows, respectively
μ_a	The observed aggregated travel time of group a from smart card data
$\mu_{i,j,\tau}$	The observed aggregated passenger count on link (i, j) during time period τ
w_p	The travel time of path p
c_p^r	The earning collected on path p of transit company r
$Cap_{i,j,t,s}$	Capacity of traveling arc (i, j, t, s) in the space-time network
DT^a	The departure time of group a
AT^a	The assumed arrival time of group a
D_a	The number of passengers in group a
$c_{i,j,t,s}$	Travel cost of traveling arc (i, j, t, s) in the space-time network
T	The time horizon in the space-time network
$\delta_{(i,j,t,s)}^{p,a}$	Path-link incidence index of route p of group a on arc (i, j, t, s)
w^p	The path travel time of path p
Variables	
$x_{i,j,t,s}^a$	The number of passengers in group a is assigned on traveling/waiting arc (i, j, t, s) in the space-time network
$\theta_a, \theta_{i,j,\tau}$	Continuous positive deviation variables for group a's travel time and link (i, j) during time period τ, respectively
x_a^p	The number of passengers of group a choosing their feasible path p
μ_a^*	The corrected aggregated travel time of group a from smart card data
$\mu_{i,j,\tau}^*$	The corrected aggregated passenger count on link (i, j) during time period τ

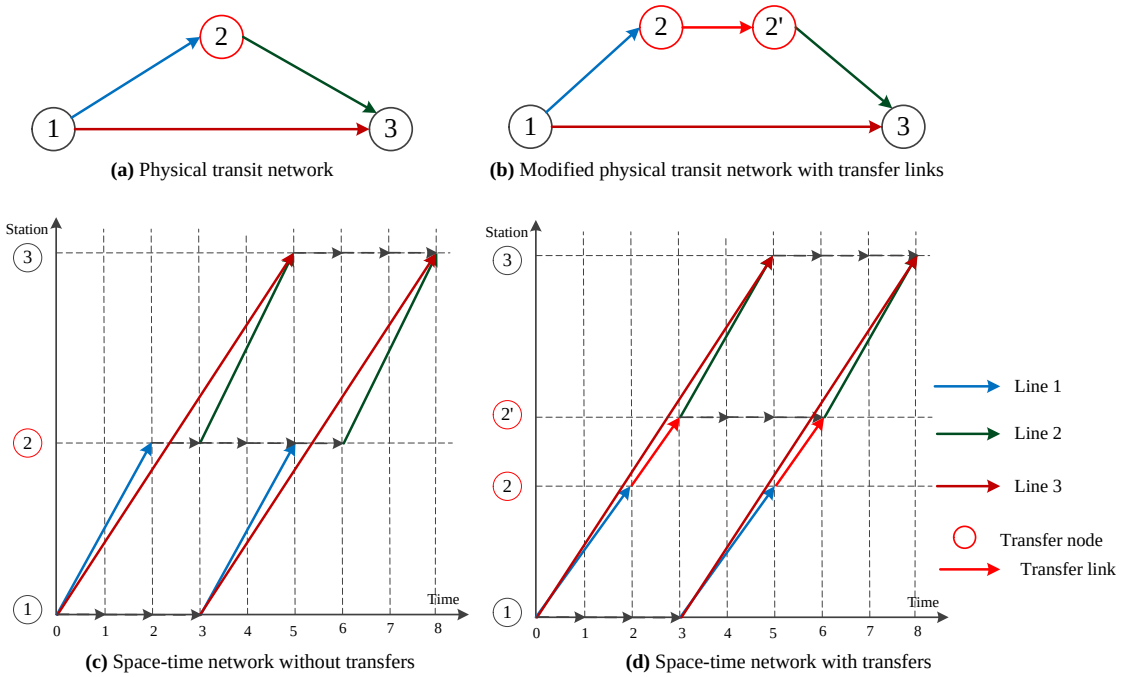
In urban rail transit systems, individual passenger should have a trip record with origin, departure time, destination and arrival time from the smart card. However, transit agencies may just provide aggregated trip data for groups of passengers. Each group a with D_a passengers has a departure time DT^a at origin node $o(a)$ to its destination node $d(a)$. At

each destination node, there is one assumed large arrival time T for all groups so that the following proposed model will be one-origin-one destination problem in the space-time network. It should be noted that the travel cost of waiting arcs on the destination node is 0, which means that once the passengers in a group arrive at the destination, the waiting cost to the super-destination (at larger arrival time T) should be 0.

About how to address the common line issue in transit systems, it can be referred to the papers by Poon et al. (2004), Hamdouch and Lawphongpanich (2008), and Liu and Zhou (2016). When addressing the many-origin-to-many-destination networks, dummy node is required as the centroid for those destination nodes. More specific details about the extended space-time network with dummy nodes can be found in a major early paper by Drissi-Kaïtouni and Hameda-Benchekroun (1992).

One transfer node can be divided as multiple nodes, depending on how many transit lines intersect at this node. One illustrative example is shown in Figure 5-4 (a) where two lines intersect at node 2 and make it as a transfer station. Then node 2 is split as node $2'$ and node 2" and the modified physical network is drawn in Figure 5-4(b). The travel time of transfer links could be the actual walking time, and its capacity is the maximum passenger throughput at transfer corridors. As a remark, based on the maximum transfer distance accepted by passengers, it is possible to connect different stops by transfer link or extended to multimodal networks. As a result, the modified physical network can be built in advance for the time-extended space-time network construction. Based on the given transit schedule and physical transit network, the space-time network is constructed in Figure 5-4(c). Further, a transfer process is considered in Figure 5-4(d) where the transfer time is assumed

to be 1 time unit. In addition, it is also feasible to consider the uncertainty of walking time on transfer links or from station entry to the platform in our proposed space-time network framework through constructing more service/travelling arcs with different arc travel times.



(a) Physical transit network

(b) Modified physical transit network with transfer links

(c) Space-time network without transfers

(d) Space-time network with transfers

Figure 5-4 Physical Transit Networks and Correspond Space-time Networks

Note that the estimated trip time that will be obtained in the later models based on the space-time network should be equal to the observed trip time of each group from smart card data. However, due to the observation noise/error, the observed trip time needs to be estimated to guarantee a feasible solution.

5.3.2 Information Space Generation Based on Multi-source Sensor Data

As mentioned before, information spaces arise as a communication channel to connect the external physical world and the internal system states. The external physical world is sensed by heterogeneous sensors in terms of different observations or information, which

finally forms a corresponding information space. Meanwhile, the internal system states are reflected in the information space based on the specific state definitions.

In addition to the physical transit lines and schedules, the possible observations in the urban transit systems are summarized in Table 5-2.

Table 5-2 Available Trip Information in Urban Transit Systems

Bus transit system	Rail transit system
(1) origin, boarding time, destination, alighting time for each person in individual bus	(1) origin, entry time, destination, exit time for each person or aggregated for each passenger group in transit network
(2) origin, and boarding time for each person in individual bus	(2) origin and boarding time for each person or aggregated for each passenger group in transit network
(3) historical time-dependent OD information for transit networks	

In bus transit systems, (1) when the origin, boarding time, destination, alighting time for each person in individual bus are available from smart card data, the state in vehicle can always be observed and the transit assignment problem disappears, but those accurate trip information provides great values for the operational transit planning. (2) If only the origin, and boarding time can be recorded, algorithms are needed to estimate the individual destination. Since the transit schedule or transit vehicle trajectory can be obtained, once the destination is estimated as the nearest stop to the boarding stop of traveler's next route based on the continuous riding records, the corresponding alighting time is also available. However, if travelers just have a single transit trip, it is still difficult to estimate the destination, which causes a large uncertainty for state estimation. (3) If the market penetration rate of smart card is very low or the goal is for operational transit planning, the historical time-dependent OD information has to be used to perform transit network

assignment with assumed travel behaviors (Szeto and Jiang, 2014a; Jiang and Szeto, 2016; Cats et al., 2016; Liu and Zhou, 2016; Codina et al., 2017), which could create a larger uncertainty in the system and needs to be carefully calibrated and validated by real-world survey and observations.

In rail transit systems, (1) the origin, entry time, destination, exit time usually are available for each passenger or group, but the path/vehicle/transfer selection in the network level still has a large uncertainty. (2) When the destination and exit time to stations are not recorded, the uncertainty will be increased more in the network level.

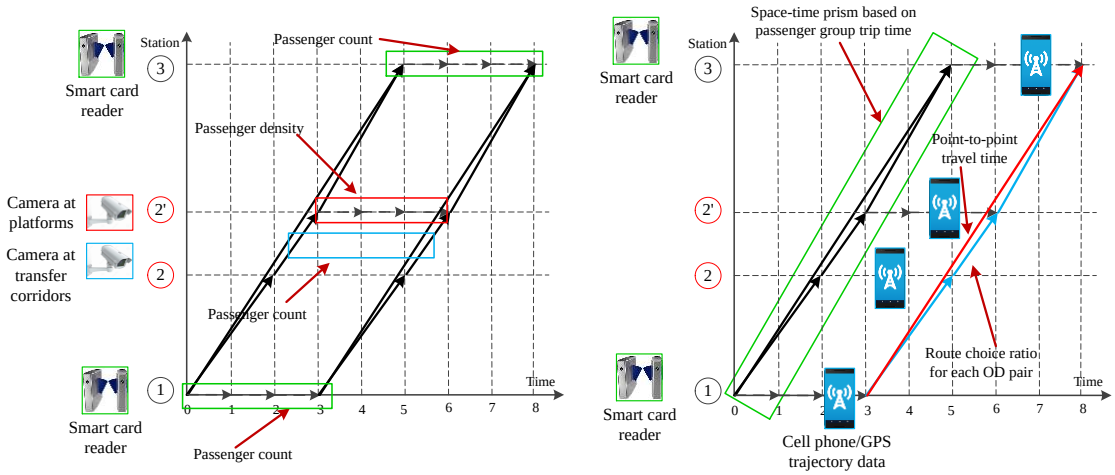
In addition, with the development of sense technologies, more available sensor information can be used in the transit systems.

(1) Video data processed to gain the aggregated passenger flow at key points during different time periods, such as, transfer corridors, the entry and exit of stations, or the stop/platforms.

(2) Cellphone/GPS based point-to-point trajectory data. A general path choice ratio bound when the penetration/sample rate of cell phone used as sensors is big enough. The granularity of the trajectory points is highly depended on the cell tower locations. Also, Bluetooth data can provide a point-to-point travel time and general path choice ratio.

(3) General travel behavior data (e.g. preference) through survey. It can provide the path choice of some specific passengers, so the path choice uncertainty of all passengers can be reduced, to some extent.

Taking the network in subsection 5.3.1 as an example, the possible observations are illustrated in Figure 5-5 (a) and (b) for fixed sensors and mobile sensors, respectively.



(a) Observations from fixed sensors

(b) Observations from mobile sensors

Figure 5-5 Observations from Multi-source Sensors

The specific modeling on generating information space based on those available sensor data is developed as follows.

Assumption: the first-in-first-out (FIFO) rule is not incorporated in our proposed space-time network, so it may happen that some passengers wait for a longer time than the passengers ahead, or some passengers alight and then aboard for the next available vehicle in a same route. However, the available trip time data could greatly reduce the possibility of those unrealistic results. Also, in our model, we can define the waiting cost for each passenger or group at those stops/stations which are not their transfer nodes as an infinity in the space-time network to avoid those unrealistic behaviors. Actually, passengers still have the non-FIFO behavior under certain conditions. For example, passengers are waiting at a platform for their next desirable transit vehicle, it may not actually be the next available

vehicle. Therefore, it needs more exogenous priority rules to accurate model the passenger loading process. However, as mentioned before, once more available information can be obtained, it is possible to directly and completely observe the system, so those assumed priority rules and traveler behaviors may become unnecessary.

Taking the rail transit system as our modeling object, the formulation is proposed in the following.

(i) According to the physical network, transit schedule, and dynamic OD information from smart card data, the standard flow balance constraint can be given as

$$\sum_{i,t:(i,j,t,s)\in E} x^a_{i,j,t,s} - \sum_{i,t:(j,i,s,t)\in E} x^a_{j,i,s,t} = \begin{cases} -D_a & \forall a, j = o(a), s = DT^a \\ D_a & \forall a, j = d(a), s = T \\ 0 & otherwise \end{cases} \quad (5.1)$$

(ii) Strict vehicle and station platform capacity constraint

$$\sum_a x^a_{i,j,t,s} \le Cap_{i,j,t,s}, \ \forall (i,j,t,s) \in A \quad (5.2)$$

(iii) As stated in subsection 5.3.1, the estimated trip time of each group in the model should be consistent with the observation (average trip time of each group) from smart card.

$$\sum_{(i,j,t,s)} (x^a_{i,j,t,s} \times c_{i,j,t,s}) = D_a \times \mu_a, \forall\, a \quad (5.3)$$

(iv) Estimated aggregated passenger flow count on link (i,j) during time period τ is expected to be the observation from video data or counting by people.

$$\sum_a \sum_{t\in\tau} x^a_{i,j,t,s} = \mu_{i,j,\tau}, \forall (i,j,\tau) \quad (5.4)$$

(v) Non-negative arc flow variables

$$x_{i,j,t,s}^{a} \geq 0 \tag{5.5}$$

Note that if a passenger is viewed as a group of passengers and $x_{i,j,t,s}^{a}$ is a binary variable, the above modelling is still available, and agent-based trajectory data can provide more point-to-point travel time information rather than just path choice. The above presents an arc-based formulation for constructing the information space. Also, a path-based formulation can be offered in the following based on the feasible path enumeration.

(i) Flow balance constraint:

$$\sum_{p} x_{a}^{p} = D_{a}, \forall a \tag{5.6}$$

(ii) Capacity constraint:

$$\sum_{(p,a)\in S_{(p,a)}} (\delta_{(i,j,t,s)}^{p,a} \times x_{a}^{p}) \leq Cap_{i,j,t,s}, \forall (i,j,t,s) \in A \tag{5.7}$$

(iii) Trip time constraint:

$$\sum_{p} x_{a}^{p} * w^{p} = D_{a} \times \mu_{a}, \forall a \tag{5.8}$$

(iv) Aggregated passenger flow count constraint:

$$\sum_{a} \sum_{t\in\tau} (\delta_{(i,j,t,s)}^{p,a} \times x_{a}^{p}) = \mu_{i,j,\tau}, \ \forall (i,j,\tau) \tag{5.9}$$

(v) Non-negative path flow:

$$x_{a}^{p} \geq 0 \tag{5.10}$$

Given the time-expanded space-time network constructed in subsection 5.3.1, the general feasible path set for a passenger group with specific OD pair and a departure time can be

generated by a forward label correcting algorithm from the vertex (origin and departure time) to its destination node based on the observed trip time of that group as a prism.

As a discussion, when considering the bus transit systems, the smart card data usually only have the origin and departure time information without passengers' destination and arrival time information. To model this condition that dynamic OD trips are unknown, D_a will be a variable in equation (1) and the summation of D_a with same origin and departure time should be equal to the recorded total trip production at this origin and departure time from smart card data. If the structure of each OD pair with departure time is given based on the historical OD information, the number of unknown OD variables will be greatly reduced.

5.4 Uncertainty Quantification of State Estimates under Heterogeneous Data Sources

The system state uncertainty mainly arises from two sources: one is the measurement error due to the noise and disturbance in sensing systems, and the other is lack of useful information, which results in the many-to-one mapping between the many possible system states and one partial observation. This section will address the measurement error issues and further quantify the uncertainty of state estimates.

5.4.1 The Sensor Measurement Estimation Problem

The analyses on small card data (Trépanier et al., 2007; Barry et al., 2009) show that the data must be thoroughly validated and corrected prior to the practical use. Therefore, it might happen that no feasible solution exists when the observed data are directly used in built models. Actually, the infeasibility of solutions may arise from the obvious error of observations or the missed details of the built models. Assuming that the proposed models

are accurate, a good observation should be feasible in the models, but it doesn't mean that the observations don't have any errors anymore, which can be viewed as hidden errors. In addition, even each observation is tested in the model and can provide feasible solutions, but it is still possible to have infeasible solutions when different observations are considered simultaneously, because the inconsistency among different kinds of sensors may still exist. Hence, this situation leads to the measurement estimation problem, which aims to obtain estimates as close as possible to the corresponding measurements under real-world physical constraints. There are different approaches to clean and verify those measurements in advance. The approach adopted is the generalized least squares. Based on the proposed constraints in subsection 5.3.2, a nonlinear estimation model is presented as follows.

$$\text{Min } \beta_1 \sum_a \left(\sum_{(i,j,t,s)} (x^a_{i,j,t,s} \times c_{i,j,t,s}) - D_a \times \mu_a\right)^2 + \beta_2 \sum_{(i,j,\tau)} \left(\sum_a \sum_{t\in\tau} x^a_{i,j,t,s} - \mu_{i,j,\tau}\right)^2 \quad (5.11)$$

Subject to constraints (5.1), (5.2) and (5.5).

The objective function is to minimize the weighted total deviations between the estimated and observed data, where β_1 and β_2 are the weights reflecting different degrees of confidence on observations. Those weights can be viewed as the inverses of the variances of the heterogeneous sources of measurements adopted by Lu et al. (2013).

Another technique used to measure the deviation is to quantify the absolute difference as least absolute deviations (LAD). The corresponding objective function will change to be

$$\text{Min } \beta_1 \sum_a \left|\sum_{(i,j,t,s)} (x^a_{i,j,t,s} \times c_{i,j,t,s}) - D_a \times \mu_a\right| + \beta_2 \sum_{(i,j,\tau)} \left|\sum_a \sum_{t\in\tau} x^a_{i,j,t,s} - \mu_{i,j,\tau}\right| \quad (5.12)$$

The analysis on the attributes of the methods of least squares and least absolute deviation can be found in the paper by Gorard (2005) and Koenker and Hallock (2001). Specifically, least absolute deviation treats all observation equally, but least squares gives more emphasis to large residuals by squaring the residuals, which could be a better choice when dealing with outliers in which estimated values are far from real-world sensor observations. Note that the least absolute deviations can be solved by linear programming through transforming the model. For example, the new model based on formula (5.12) will be minimizing $\beta_1 \sum_a \theta_a + \beta_2 \sum_{(i,j,\tau)} \theta_{i,j,\tau}$ while adding new constraints $-\theta_a \leq \sum_{(i,j,t,s)}(x^a_{i,j,t,s} \times c_{i,j,t,s}) - D_a \times \mu_a \leq \theta_a$, $-\theta_{i,j,\tau} \leq \sum_a \sum_{t\in\tau} x^a_{i,j,t,s} - \mu_{i,j,\tau} \leq \theta_{i,j,\tau}$, and $\theta_a \geq 0, \theta_{i,j,\tau} \geq 0$.

5.4.2 Projection Function-based State Estimate Uncertainty Quantification

The estimated measurements ensure to form an information space as a bounded or unbounded polyhedron. In this section, different projection functions will be chosen as the mapping between the feasible information space and specific system states. The states we will model are introduced in Table 5-3 and illustrated in Figure 5-6.

The state uncertainty quantification is modeled in detail as follows.

Projection function 1 for state 1: the number of passengers on one specific arc (i, j, t, s) (station platform, vehicle, transfer corridor) in the space-time network is represented as $\sum_a x^a_{i,j,t,s}$, so the arc flow/density uncertainty can be quantified by maximizing and minimizing $\sum_a x^a_{i,j,t,s}$, subject to constraints (5.1) to (5.5) where μ_a and $\mu_{i,j,\tau}$ are replaced by the estimated measurements μ^*_a and $\mu^*_{i,j,\tau}$.

Table 5-3 The Focused States and Its Motivations

Focused states	Motivations
(1) passenger density on station platform, vehicle, transfer corridors	(i) identify possible dangerous spots for safety; (ii) make decisions on vehicle updates and timetable changes
(2) the number of passengers taking one specific line segment	(i) assign the ticket fare to each company based on the service they provide; (ii) evaluate the current liquidation policy and quantify the unreasonable income bound for each company
(3) the path flow range of each OD pair	(i) compare or verify the traditional logit route choice model for better understanding travel behavior
(4) network-level time-dependent passenger density states on serval key stations/vehicles	(i) distribute the network-level transit condition; (ii) evaluate network-level control

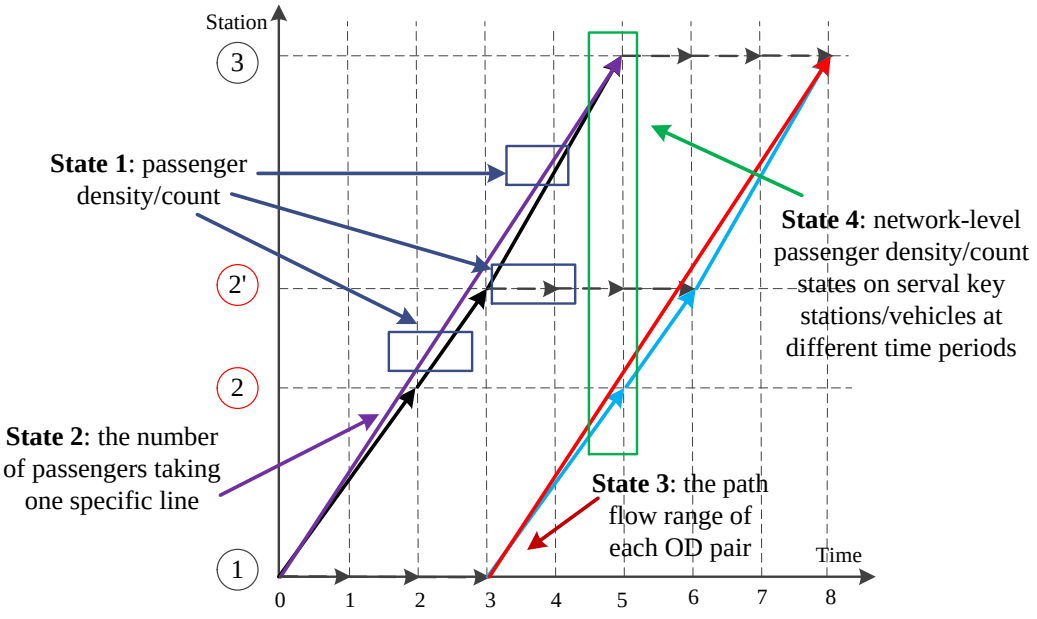


Figure 5-6 States illustration in a space-time network

Projection function 2 for state 2: the earnings that one transit company r can obtain is represented as $\sum_a \sum_p (x_a^p \times c_p^r)$, where c_p^r is the income of using the segment in company r's operation area of path p. It can be calculated as a parameter in advance based on the

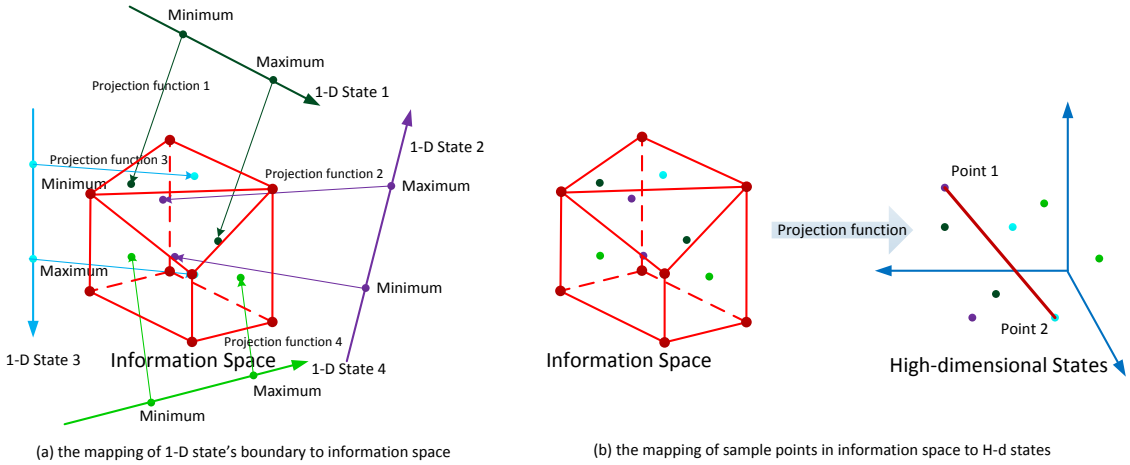
ticket price and segment and path distance. Therefore, the earning bound is estimated by maximizing and minimizing$\sum_a \sum_p (x_a^p \times c_p^r)$ subject to constraints (5.6) to (5.10) where μ_a and $\mu_{i,j,\tau}$ are replaced by the estimated measurements μ_a^* and $\mu_{i,j,\tau}^*$.

Projection function 3 for state 3: the flow rate on path p is $\sum_a x_a^p$, so the uncertainty bound of path flow is measured by maximizing and minimizing $\sum_a x_a^p$, subject to constraints (5.6) to (5.10) where μ_a and $\mu_{i,j,\tau}$ are replaced by the estimated measurements μ_a^* and $\mu_{i,j,\tau}^*$.

Projection function 4 for state 4: the passenger flow (density) states on key station platforms at one time index (e.g., at 7:30am) will be a high-dimensional vector $\{\boldsymbol{q}(\boldsymbol{i},\boldsymbol{t})\}$ where i is one of key stations. For one specific station i, $q(i,t) = \sum_a x_{i,j,t,s}^a$ is the number of passengers at station i at time t. Since the state is not one dimension anymore, the concept of the Maximal Possible Relative Error (MPRE) first introduced by Yang et al. (1991) is adopted to quantify the state uncertainty of high-dimensional variables. As shown in Figure 5-7(a), the state solution (vector $\{\boldsymbol{x}_{\boldsymbol{i,j,t,s}}^{\boldsymbol{a}}\}$) based on different projection functions for one-dimensional state above is one feasible solution in the information space, so each solution (vector $\{\boldsymbol{x}_{\boldsymbol{i,j,t,s}}^{\boldsymbol{a}}\}$) can be mapped to high-dimensional states to generate new state points (vector $\boldsymbol{q}(\boldsymbol{i},\boldsymbol{t})$)illustrated in Figure 5-7(b), which are used as sample points to approximately obtain the MPRE. Specifically, we need to calculate the average relative error between any two points, and find the maximal one as the MPRE.

For example, the average relative error between point 1 and point 2 is calculated as follows (Yang et al., 1991), where $\boldsymbol{q_1}(\boldsymbol{i},\boldsymbol{t})$ and $\boldsymbol{q_2}(\boldsymbol{i},\boldsymbol{t})$ are a m-dimensional vector recording m

stations' passenger flow at time t. The relative deviation is $\lambda_{(1,2,i,t)} = \frac{q_1(i,t)-q_2(i,t)}{q_1(i,t)}$ and the average relative deviation $AV(\lambda_{(1,2,t)}) = \sqrt{\frac{\phi(\lambda_{(1,2,t)})}{m}}$, where$\phi(\lambda_{(1,2,t)}) = \sum_{i=1}^{m} \lambda_{(1,2,i,t)}{}^2$and $\lambda_{(1,2,t)} = \{\lambda_{(1,2,1,t)}, \lambda_{(1,2,2,t)}, \dots, \lambda_{(1,2,m,t)}\}$. In addition, Yang et al. (1991) defined the concept of Estimation Reliability as a measure about the state uncertainty; that is, $Re = \frac{1}{1+AV(\lambda)}$, which shows that when the $AV(\lambda)$ is 0, the reliability of the estimated state is 1. In contrast, when $AV(\lambda)$ tends to infinity, there is almost no reliability guarantee. The result is just based on some sample points, so it is an approximation approach.



(a) the mapping of 1-D state's boundary to information space

(b) the mapping of sample points in information space to H-d states

Figure 5-7 Relation of Information Space and Different Types of States

In addition, we can also assume that any one of the points in the high-dimensional states is the estimated one, and then adopt the method of finding one feasible state which could have a MPRE as the upper bound of our state estimate uncertainty. The problem can be formulated as a quadratic programming model (Yang et al., 1991). Note that the above is also an approximation since the estimated state is selected from sample points.

5.4.3 Discussion on the Real-time State Uncertainty Quantification

The uncertainty of real-time system state increases the difficulty of real-time state prediction and optimal control. Compared with the offline state estimation in this chapter, the challenges in the real-time condition include that (1) the real-time rail transit OD travel information is not available and (2) the state transition along the time is highly required.

(1) Real-time OD demand estimation: Based on day-to-day historical and accurate dynamic OD demands in urban rail transit systems, we can classify $k$representatives $OD_{o,d,\tau}^{k}$for each OD pair at different time periods, so the estimated real-time OD demand is $OD_{o,d,\tau} = \sum_k (w_k \times OD_{o,d,\tau}^{k})$ where w_k is a binary variable, which indicates that only one OD candidate k will be chosen. As a result, the dynamic OD travel demand's spatial structure can be well captured, compared with those OD estimation models which mainly optimize one departure time profile for all or one-class static total OD trips. In addition, the real-time trip generation at each station/origin with departure time is available from the smart card data, so $\sum_d OD_{o,d,\tau} = OD_{o,\tau}^{obs}$ provides more information to generate the real-time information space.

(2) Real-time state transition: the rolling horizon approach has been widely chosen for real-time transportation operations and control (Peeta and Mahmassani, 1995; Zhou and Mahmassani, 2007; Meng and Zhou, 2011). Under this mechanism, when focusing on one time period, it needs a look-back period and a look-ahead period, because the generated passengers from the look-back period could still in the transit network during our focused time period, and in the look-ahead period we can assume that all passengers can arrive at their destination for our network modeling. Along the planning time horizon, once some

trips are finished at our focused time period, their true OD information can be obtained in real time, so the corresponding estimated OD trips can be replaced by the real ones, which can also reduce the information space for our state uncertainty estimation.

Similar to the offline modelling, the states can be flexibly defined based on the managers' analysis goals. The min/max models on one-dimensional state and the MPRE for multi-dimensional states are also available for quantifying the real-time state uncertainty, which provides a fundamental input for the measure of future real-time prediction and optimal control.

5.5 Solution Algorithms

The models proposed in Section 5.4 generally include:

M1: a linearly constrained quadratic measurement estimation model considering the squared difference between estimate and measurement. Objective function (5.11), subject to constraints (5.1), (5.2) and (5.5).

M2: a linear measurement estimation model considering the absolute difference between estimate and measurement. Objective function (5.12), subject to constraints (5.1), (5.2) and (5.5).

As discussed in subsection 5.4.1, we prefer choosing M1 for the measurement estimation problem due to its senstivity to the large difference between estimates and observations.

M3: linear programming models based on projection functions 1-3. Objective functions: passenger flow, company's earning collected from ticket fare, path flow of one OD pair;

Constraints: arc-based formulation (5.1)-(5.5) or path-based formulation (5.6)-(5.10) with estimated measurements rather than the original observed.

M4: a quadratic programming model for finding the MPRE based on selected one of sample points. Objective function is to minimize $\phi(\lambda_t)$ of one given sample point and the unknown point in the high-dimensional states, subject to constraints (5.1)-(5.5) with estimated measurements.

Therefore, this paper will address two kinds of models. One's objective function is nonlinear (quadratic), and the other's is linear. The both models has similar side constraints; that is, one flow balance constraint and other linear side constraints, such as, capacity constraints, observation constraints.

5.5.1 Frank-wolfe Algorithm for Nonlinear Programming Models

The Frank-Wolfe algorithm is used to solve the optimization problem where the objective function is convex differentiable real-valued function and the feasible region of side constraints is compact convex (Frank and Wofle, 1956). Therefore, M1 and M4 can be well solved under the framework of the Frank-Wolfe algorithm. For simplicity, M1 is represented as follows,

Objective function:

$$\text{Min} \quad f(x) = \beta_1 \sum_a \Big(\sum_{(i,j,t,s)} (x^a_{i,j,t,s} \times c_{i,j,t,s}) - D_a \times \mu_a\Big)^2 \; + \; \beta_2 \sum_{(i,j,\tau)} \Big(\sum_a \sum_{t\in\tau} x^a_{i,j,t,s} - \mu_{i,j,\tau}\Big)^2$$

Subject to, flow-balance constrains: $AX = B$, capacity constraints: $CX \le D$, and $x \ge 0$.

The algorithm procedure is described as follows.

Step 1: initialization: $k = 0$, and find one feasible solution as $\boldsymbol{x}_0$;

Step 2: Based on the first-order Taylor approximation of $f(x)$ aroundx_k, minimizing the linear approximation: min $\boldsymbol{s}_k^T \nabla f(x_k)$, and $\boldsymbol{s}_k$ is subject to all constraints.

Step 3: Find γ that minimizes $f(x_k + \gamma(s_k - x_k))$, subject to $0 \leq k \leq 1$.

Step 4: Update: $x_{k+1} = x_k + \gamma(s_k - x_k)$. If $|x_{k+1} - x_k| \leq \Delta$ or $k = K$, stop. Otherwise, $k = k + 1$ and go to step 2.

Specifically, at step 2, $\nabla f(x_k) = \beta_1 \sum_a \sum_{(i,j,t,s)} [2 \times c_{i,j,t,s} \times c_{i,j,t,s} \times x^a_{i,j,t,s}(k)] + \beta_2 \sum_{(i,j,\tau)} [\sum_a \sum_{t \in \tau} (2 \times x^a_{i,j,t,s}(k))]$ is constant, so $\boldsymbol{s}_k^T \nabla f(x_k) = \beta_1 \sum_a \sum_{(i,j,t,s)} [2 \times c_{i,j,t,s} \times c_{i,j,t,s} \times x^a_{i,j,t,s}(k) \times s^a_{i,j,t,s}(k)] + \beta_2 \sum_{(i,j,\tau)} [\sum_a \sum_{t \in \tau} (2 \times x^a_{i,j,t,s}(k) \times s^a_{i,j,t,s}(k))]$. Finally, the model proposed in step 2 is a linear programming model with the flow-balance constraint.

In addition, at step 1, for finding one feasible solution, we can define a simple linear objective function, so the model will be a linear programming model with the flow balance constraint. Also, M3 is this kind of linear programs, which will be solved by the Dantzig-Wolfe algorithm due to the special block structure of the flow balance constraint in next subsection.

5.5.2 Dantzig-Wolfe Decomposition for Linear Programming Models

The Dantzig–Wolfe decomposition is originally proposed by Dantzig and Wolfe (1960) for solving linear programming problems with special structure. A general primal linear

program can be represented as: min $c^T \boldsymbol{x}$, subject to, $A\boldsymbol{x} \leq \boldsymbol{b}$, $D\boldsymbol{x} \leq \boldsymbol{d}$, and $\boldsymbol{x} \geq 0$. According to Minkowski-Weyl's Theorem, given the convex set $X = \{x \in \mathbb{R}^n | Ax \leq b\}$ where $Ax \leq b$ is a special block, X can be represented by the extreme points and extreme rays of X: $X = \{x = \sum_i \lambda_i x^i + \sum_j \mu_i y^j | \sum_i \lambda_i = 1, \lambda_i \geq 0, \mu_j \geq 0\}$. When X is a bounded polyhedron, X can be represented by the extreme points, $X = \{x = \sum_i \lambda_i x^i | \sum_i \lambda_i = 1, \lambda_i \geq 0\}$.

Substituting the expression above to the original model leads to the following Master Problem: min $\sum_i c^T \lambda_i \boldsymbol{x}^i$, subject to, $\sum_i D\lambda_i \boldsymbol{x}^i \leq \boldsymbol{d}$, $\sum_i \lambda_i = 1$ and $\lambda_i \geq 0$. Suppose that a subset of extreme points P is available. The Restricted Master Problem (RMP) can be obtained: min $\sum_{i \in P} c^T \lambda_i \boldsymbol{x}^i$, subject to, $\sum_{i \in P} D\lambda_i \boldsymbol{x}^i \leq \boldsymbol{d}$, $\sum_{i \in P} \lambda_i = 1$ and $\lambda_{i \in P} \geq 0$. Assume that λ^* and (π, μ) is the optimal and dual solutions to the RMP, respectively. The reduced cost is defined as $\gamma(\boldsymbol{x}) = c^T \boldsymbol{x} - \pi^T A\boldsymbol{x} - \mu$. Solve the subproblem: min $c^T \boldsymbol{x} - \pi^T A\boldsymbol{x} - \mu$, subject to $A\boldsymbol{x} \leq b$ and $\boldsymbol{x} \geq 0$. If the reduced cost is non-negative, the solution is optimal; otherwise, the solution can be viewed as a new extreme point and added to the RMP until the reduced cost is non-negative.

The flow-balance constraint in network flow models (Larsson and Patriksson, 1992; Larrson et al., 2004; Desrosiers and Lubbecke, 2005;) can be viewed as a special block solved by classical shortest path algorithms, so the Dantzig–Wolfe decomposition will be adopted for our proposed linear programs. Specifically, the flow on a particular path (or path flow for a passenger group a) can represent one extreme point. A path flow uniquely corresponds its path, so a particular path implicitly indicates a specific extreme point. This enables us to express the arc flow of group a on arc (i, j, t, s) as $x^a_{i,j,t,s} = \sum_h (\delta^{p(h),a}_{(i,j,t,s)} \times$

$x_a^{p(h)} \times \lambda_{(a,h)})$, where $x_a^{p(h)} = D_a$ for each generated extreme point h , and $\sum_{h\epsilon H(a)} \lambda_{(a,h)} = 1$. Since variable $x_{i,j,t,s}^a$ is continuous rather discrete, it should be a continuous combination of extreme points and $\lambda_{(h,a)} \geq 0$ according to Minkowski-Weyl's Theorem. On the other hand, the link flow vector of each group $\boldsymbol{x}_{i,j,t,s}^a$ can also be seen as one extreme point, as it is the result of one specific path flow vector. In addition, for the maximum problem, we can transform it as a minimum problem by changing the positive arc costs to be negative.

The detailed formulation for our models by the Dantzig-Wolfe decomposition will not presented due the limited space in this chapter. However, the general procedure of the algorithm is described as follows:

Step 1: initialization. Find one feasible passenger flow on the shortest paths as extreme points.

Step 2: Solve the restricted master problem to obtain the duals of side constraints.

Step 3: Solve each sub-problem to calculate its reduced cost as a time-dependent shortest path problem. If its reduced cost is negative, add the solution of the sub-problem to the restricted master problem at step 2. When the reduced costs of all sub-problems are non-negative, the optimal solution is achieved.

In the initialization step, in order to find one feasible solution from the shortest path problem as initial extreme points, we can introduce artificial variables for those coupling constraints and solve the problem by the Dantzig-Wolfe decomposition again (Kalvelagen, 2003). For example, the coupling side constraint is $\sum_j D_{i,j} x_i \leq d_i$, so we can add artificial

variable $y_i \geq 0$ to have $\sum_j D_{i,j} x_i - y_i \leq d_i$, and minimize $\sum_i y_i$ as a master problem. Based on the Dantzig-Wolfe decomposition algorithm, when the $\sum_i y_i$ is equal to 0 or less than 0.0001, we can conclude that one feasible solution for our primal problem is obtained and can be used for step 2.

5.6 Experiments

This section will demonstrate the proposed models and algorithms in Sections 5.4 and 5.5 which are implemented in a general purpose optimization package GAMS. The experiments are performed in the following transit network shown in Fig. 5-8(a), where 7 urban rail lines exist in the transit systems. In order to model the passenger count observation at transfer corridors, specific transfer links are built as shown in Fig. 5-8(b).

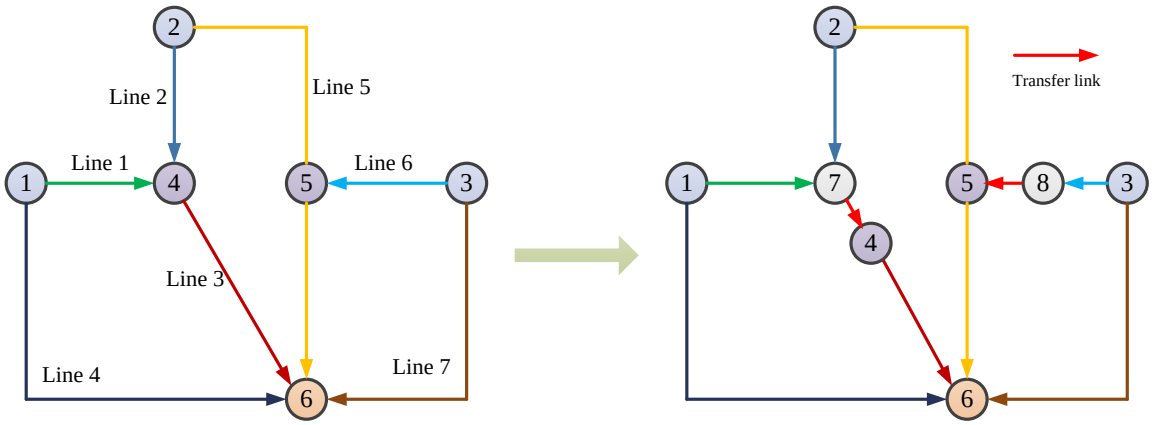


Figure 5-8 Hypothetic Urban Rail Transit Network

5.6.1 Given Multi-source Sensor Data

(1) Table 5-4 lists the existing transit service arcs based on given the timetable of the seven transit lines, and the corresponding space-time network is constructed in Figure 5-9.

Table 5-4 Hypothetic Transit Service Arcs List

Service Arc	Start Time	End Time	Service Arc	Start Time	End Time
(1,7)	0	3	(1,7)	3	6
(7,4)	3	4	(7,4)	6	7
(4,6)	4	6	(4,6)	7	9
(1,6)	0	8	(1,6)	3	11
(2,7)	0	3	(2,7)	3	6
(2,5)	0	4	(2,5)	3	7
(5,6)	4	7	(5,6)	7	10
(3,8)	0	3	(3,8)	3	6
(8,5)	3	4	(8,5)	6	7
(3,6)	0	8	(3,6)	3	11

(2) The origin, destination, departure time and aggregated trip time of each passenger group are listed in Table 5-5, and each group represents 100 passenger in this test.

(3) The vehicle capacity of each line is assumed in

Group No	OD Pair	Departure Time	Average Trip Time	Group No	OD Pair	Departure Time	Average Trip Time
1	1 → 6	0	6	15	1 → 6	3	7.5
2	1 → 6	0	7	16	1 → 6	3	7
3	1 → 6	0	8	17	1 → 6	3	8
4	1 → 6	0	6.5	18	2 → 6	3	6
5	2 → 6	0	7	19	2 → 6	3	7
6	2 → 6	0	7.5	20	2 → 6	3	6.5
7	2 → 6	0	6.5	21	2 → 6	3	7.5
8	2 → 6	0	6	22	2 → 6	3	8
9	3 → 6	0	7	23	2 → 6	3	6.8
10	3 → 6	0	7.5	24	3 → 6	3	7
11	3 → 6	0	8	25	3 → 6	3	7.5
12	1 → 6	3	6	26	3 → 6	3	7.4
13	1 → 6	3	7	27	3 → 6	3	7.8
14	1 → 6	3	6.5	28	3 → 6	3	8

Table 5-6, where it can be observed that the capacity of rail transit vehicles could have its adjustment at different time periods by increasing or decreasing the number of train units.

(4) The passenger count data from video processed data at transfer corridor (7,4) is available; that is, 450 and 810 passengers are observed at time points 3 and 6.

5.6.2 Focused States for Estimation Uncertainty Quantification

The states we focused in this experiment are listed as follows.

State 1: passenger count (congestion) in transfer corridor (8,5) at time points 3 and 6.

State 2: the passenger flow departing at node 2 and time 0 to use line 1.

State 3: the earning collected in the ticket for company line 1 on its first vehicle.

State 4: the system-wide passenger count (congestion) on the running vehicles at time point 5.

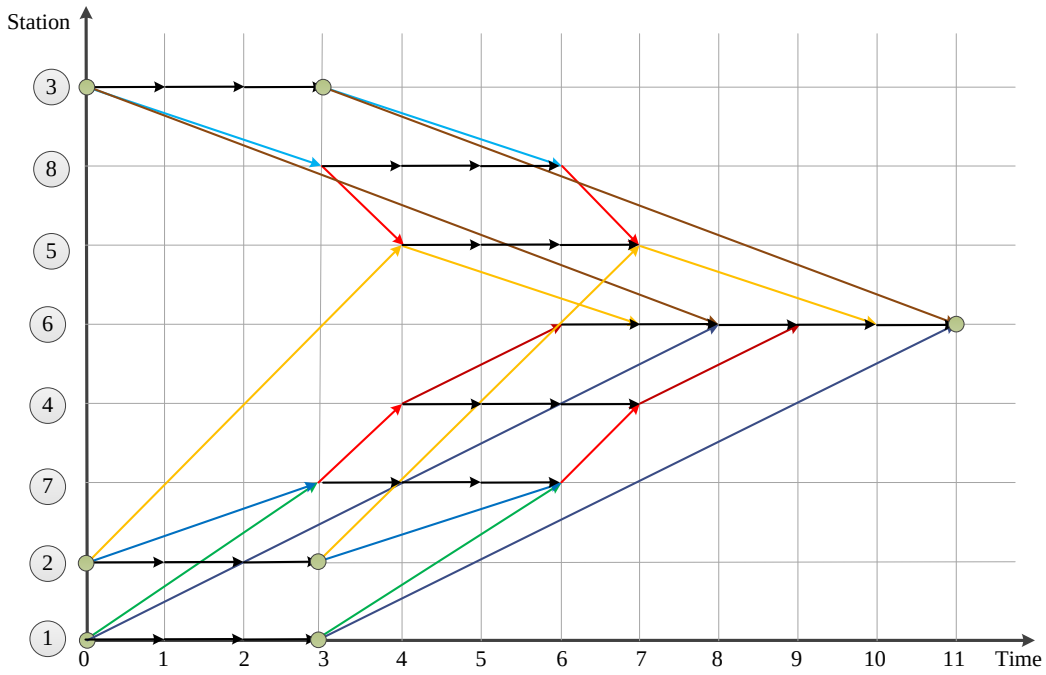


Figure 5-9 The Corresponding Space-time Transit Service Network

Table 5-5 Trip Attributes of Each Passenger Group

Group No	OD Pair	Departure Time	Average Trip Time	Group No	OD Pair	Departure Time	Average Trip Time
1	1 → 6	0	6	15	1 → 6	3	7.5
2	1 → 6	0	7	16	1 → 6	3	7
3	1 → 6	0	8	17	1 → 6	3	8
4	1 → 6	0	6.5	18	2 → 6	3	6
5	2 → 6	0	7	19	2 → 6	3	7
6	2 → 6	0	7.5	20	2 → 6	3	6.5
7	2 → 6	0	6.5	21	2 → 6	3	7.5
8	2 → 6	0	6	22	2 → 6	3	8
9	3 → 6	0	7	23	2 → 6	3	6.8
10	3 → 6	0	7.5	24	3 → 6	3	7
11	3 → 6	0	8	25	3 → 6	3	7.5
12	1 → 6	3	6	26	3 → 6	3	7.4
13	1 → 6	3	7	27	3 → 6	3	7.8
14	1 → 6	3	6.5	28	3 → 6	3	8

Table 5-6 Vehicle Capacity of Transit Lines

Line No	L1	L2	L3	L4	L5	L6	L7
Capacity of vehicles departing at time 0	300	300	600	200	400	300	200
Capacity of vehicles departing at time 3	400	400	800	300	600	400	300

5.6.3 Scenario Design

As a short summary, based on the available supply and demand data, we aim to (i) estimate the measurements in case there is no feasible solution due to the possible existence of measurement errors in step 1, and (ii) quantify the uncertainty of our focused states in step 2. Five scenarios are designed to demonstrate the value of information based on our proposed models.

Scenario 1 (S1: base case): it is assumed that the origin, destination, and departure time of each passenger group is given, and no other information is available.

Scenario 2 (S2: base case + count): based on scenario 1, the passenger count data from video processed data at transfer corridor (7, 4) is available.

Scenario 3 (S3: base case + end-to-end travel time): based on scenario 2, the averaged group trip time from smart card is available.

Scenario 4 (S4: base case + end-to-end travel time + count): based on scenario 1, both the passenger count data and average group trip time data are available.

Scenario 5 (S5: ground truth): since the observed data could have its measurement errors, we assume that a ground truth can be obtained and will be compared with other scenarios. The ground truth is assumed as the system conditions based on maximizing state 1 at time point 3 in scenario 3.

5.6.4 Result Analysis

It needs two steps to obtain the final results. In step 1, the measurement estimation is performed to best estimate the observation and avoid to have infeasible solutions in step 2. Then, in step 2, we compute the uncertainty range of states 1-3 by maximizing and minimizing the state goals, and state 4 is addressed based on the solutions from the previous three states as a sample-based approximation. Before analyzing different state results in different scenarios, it is important to clearly illustrate the conditions under which those results are obtained from our proposed models.

(1) In scenario 1, there is no available sensor data, so the measurement estimation is not necessary.

(2) In scenario 2, the measurement estimation is performed for the passenger count data at transfer corridor (7, 4). The total squared errors in objective function (5.11) in step 1 is not equal to 0, which indicates that there will be no feasible solution if the observed

measurement is directly used to step 2. The estimated passenger counts at transfer corridor (7, 4) at time points 3 and 6 is 450 and 800, respectively, compared with the observed values of 450 and 810. The total absolute error for the observed passenger count is 10.

(3) In scenario 3, when step 1 is conducted using the observed average trip time, the total error is also not equal to 0. The estimated average group trip time for each group is shown in Table 5-7. The total absolute error for the average group trip time is 2.58.

(4) In scenario 4, in step 1, there are two different sensor data, so it will require weights on different measurements. As discussed by Lu et al. (2013), the weights should reflect the degrees of confidence on different observed data and can be represented by the inverses of the variances of the distinct sources of measurements. Therefore, the weights on aggregated average trip time and passenger count are calculated as 2.36 and 0.31, respectively. Finally, the total absolute errors for observed average group trip time and passenger count are 3.83 and 273, which are greater than the absolute errors in scenario 1 and scenario 2, respectively. It shows that the inconsistency among multi-source data makes the model to find a balance among those observation.

(5) In scenario 5, the estimated group trip time in step 1 is used as the input to maximize the passenger count in transfer corridor (88,55) at time points 3, and the corresponding system condition is assumed as the ground truth in this dynamic transit system.

Figure 5-10 shows that the estimated maximum and minimal flow rates on each focused arc under different scenarios. As the increase of available information, the uncertainty range of passenger flows on transfer corridor (8, 5) is reduced. Meanwhile, both scenarios

3 and 4 can assert that their estimated state uncertainty is 0 and the state is completely observable. However, the different estimated unique states on arc (8,5,6,7) seem conflicted.

Table 5-7 The Observed and Estimated Average Group Trip Time for Each Passenger Group

Passenger group No	Observed values	Estimated values in scenario 3	Passenger group No	Observed values	Estimated values in scenario 3
1	6	6	15	7.5	7.5
2	7	7	16	7	7
3	8	8	17	8	8
4	6.5	6.5	18	6	6
5	7	7	19	7	6.9
6	7.5	7	20	6.5	6.4
7	6.5	6.5	21	7.5	7
8	6	6	22	8	7
9	7	7	23	6.8	6.7
10	7.5	7.5	24	7	7.08
11	8	8	25	7.5	7.57
12	6	6	26	7.4	7.47
13	7	7	27	7.8	7.87
14	6.5	6.5	28	8	8

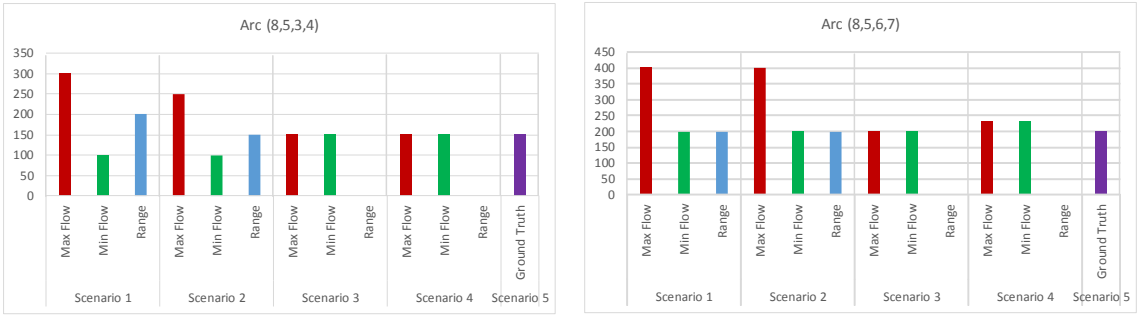


Figure 5-10 The Estimated Flow Uncertainty Range on Each Focused Arc

Specifically, in scenario 3, the observed trip time is corrected due to its measurement error and finally the estimated states on transfer corridor (8, 5) is consistent with the states in the ground truth. Note that the estimated states may not be totally consistent with the ground truth, even though the observed data is same as the corresponding data in ground truth, because the observation is only a part reflection of the whole system condition. It is also

possible that the corrected measurement is not consistent with that in this ground truth if other measurement correction approaches are used in reality in step 1.

In addition, in scenario 4, the inconsistency of observed link count data and observed trip time data makes the corrected measurement different with the corresponding data in the ground truth, so the final estimated unique state in step 2 is not the real-world condition anymore. Therefore, in reality, when the transportation system state is estimated by different sensor data, the data quality and assigned weight on each data source in step 1 is important and should be clearly stated. In other words, one bad observation could decrease the accuracy of those final estimated states when it is incorporated in the estimation models. Also, different weights on each data source could make the estimated state different and unrealistic.

Focusing on the passenger flow departing at node 2 and time 0 to use line 1, it is actually the path flow of path (2,0) →(5,4)→ (6,7). The path flow uncertainty is shown in Figure 5-11. The uncertainty range is similar to the arc flow above. The estimated unique state in scenario 4 is not consistent with the state value in ground truth. In addition, if line 1 is managed by one company and the other lines are managed by other different company, it needs to assign the fare to each company based on their service. However, the number of passengers using one specific line is uncertain in the transit system, so based on our proposed method, we can quantify the uncertainty and estimate the general fare earning for each company rather than just using just some simple rules for fare clearing (Gao et al., 2011; Zhou, 2014). One simple rule is to calculate the shortest path and then assume that passengers will choose the shortest path as their selected lines.

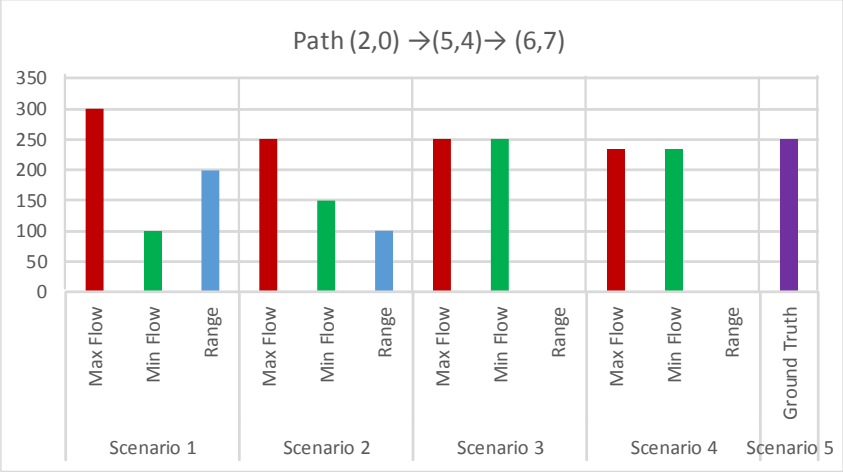


Figure 5-11 Estimated Flow Uncertainty Range on the Focused Path

In each scenario, we maximize and minimize the passenger flow on arc (8,5,3,4), arc (8,5,6,7), and path of line 1 as 6 cases, respectively, so six feasible solutions of „ can be obtained as sample points to estimate our defined system-level state uncertainty. For state 4, the system-wide passenger count (congestion) on the running vehicles at time point 5 could be represented by the passenger flow on arcs (1,6,3,11), (1,6,0,8), (1,7,3,6), (2,5,3,7), (2,7,3,6), (3,6,3,11), (3,6,0,8), (3,8,3,6), (4,6,4,6), and (5,6,4,7), which are listed in Table 5-8 for scenario 1.

Table 5-8 Estimated Passenger Flows on Arcs under Six Objectives in Scenario 1

Arc(i,j,t,s)	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
(1,6,3,11)	300	300	300	300	200	200
(1,6,0,8)	200	200	200	200	100	100
(1,7,3,6)	300	300	300	300	400	400
(2,5,3,7)	400	400	200	400	200	200
(2,7,3,6)	200	200	400	200	400	400
(3,6,3,11)	300	300	100	300	100	100
(3,6,0,8)	0	200	200	200	0	200
(3,8,3,6)	200	200	400	200	400	400
(4,6,4,6)	500	300	300	300	600	400
(5,6,4,7)	400	400	400	400	400	400

Based on the definitions of Maximal Possible Relative Error (MPRE) and Estimation Reliability (Re), the values of MPRE and Re are 16.38 and 5.753%, respectively. As shown in Table 5-8, the possible flow on arc (3,6,0,8) is from 0 to 200, which creates a huge uncertainty and makes the estimation reliability extremely low. One possible reason is that our proposed models don't assume any travel behavior, so all solutions are based on the physical constraints and available sensor observations.

In scenario 2, with passenger count information, the estimated results of 6 cases for the system-level state are shown in Table 5-9. The corresponding values of MPRE and Re are 0.267 and 78.93%, respectively. It shows that the estimation reliability gets significantly improved when passenger counts from one key location (transfer corridor) are available, which could avoid a large uncertainty range occurred in scenario 1. Also, scenarios 3 and 4 are performed and their values of MPRE and Re are 0 and 100%, respectively, but it is still emphasized that the MRPR and Re should be clearly explained with its correspondingly different measurement estimation errors (assigned weights) and adopted approach.

Table 5-9 The Estimated Passenger Flows on Arcs under Six Objectives in Scenario 2

Arc(i,j,t,s)	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
(1,6,3,11)	200	200	300	300	200	200
(1,6,0,8)	200	100	200	200	200	100
(1,7,3,6)	400	400	300	300	400	400
(2,5,3,7)	200	200	200	400	200	200
(2,7,3,6)	400	400	400	200	400	400
(3,6,3,11)	300	300	100	300	100	100
(3,6,0,8)	50	200	200	200	50	150
(3,8,3,6)	200	200	400	200	400	400
(4,6,4,6)	450	450	300	300	450	450
(5,6,4,7)	400	350	400	400	400	400

5.6.5 Results from Frank-Wolfe Algorithm and Dantzig-Wolfe Decomposition

In this section, we implement Frank-Wolfe algorithm in step 1 and Dantzig-Wolfe decomposition algorithm in step 2 in GAMS. The case of minimizing the passenger flow on arc(8,5,3,4) in scenario 4 is treated as an example to analyze the performance of those algorithms.

In section 5.6.4, the case is solved by the solver MINOS in GAMS directly. In step 1, the solved model is a non-linear programming model, and the minimal total generalized least square error in objective function is 5.968. When the model is solved by Frank-Wolfe algorithm as a linear programming model, the result shown in Figure 5-12 finally converge to 7.069 after 20 iterations. The gap is probably caused by the optimal step size, which is found as a constant value at each iteration rather than a constant value vector for each variable. Hence, it could make the final solution converge to a local optimal solution.

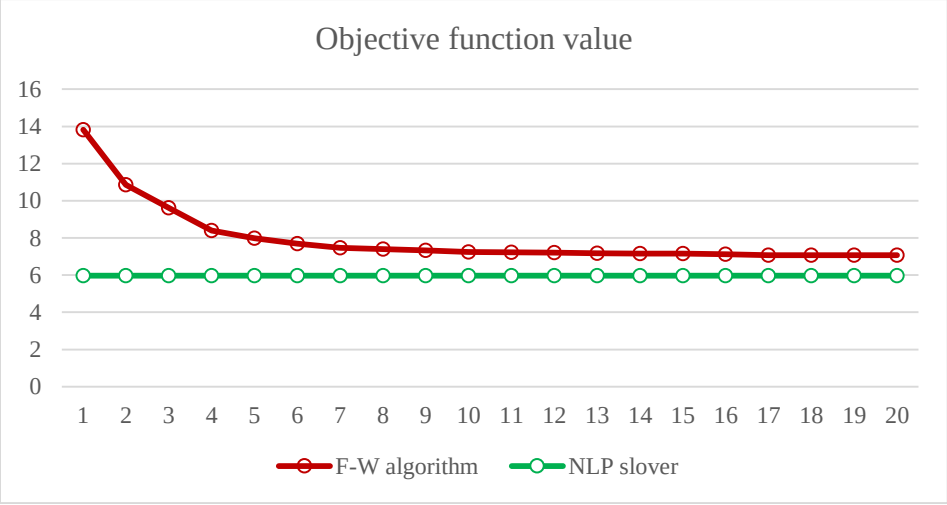


Figure 5-12 Objective Function Values under Different Solving Approaches

Table 5-10 Generated Extreme Points and Optimal Weights in Dantzig-wolfe Decomposition

Passenger Group No	Extreme points (path node sequence (ii, tt))	Optimal weights on extreme points	Passenger Group No	Extreme points (path node sequence (ii, tt))	Optimal weights on extreme points
1	(1,0)→(6,8);	0.06	15	(1,3)→(7,6) →(4,7) → (6,9);	0.30
	(1,0)→(7,3) →(4,4) → (6,6);	0.94		(1,3)→(6,11);	0.70
2	(1,0)→(6,8);	0.54	16	(1,3)→(7,6) →(4,7) → (6,9);	0.51
	(1,0)→(7,3) →(4,4) → (6,6);	0.46		(1,3)→(6,11);	0.49
3	(1,0)→(6,8);	0.93	17	(1,3)→(7,6) →(4,7) → (6,9);	0.07
	(1,0)→(7,3) →(4,4) → (6,6);	0.07		(1,3)→(6,11);	0.93
4	(1,0)→(6,8);	0.29	18	(2,3)→(7,6) →(4,7) → (6,9);	1
	(1,0)→(7,3) →(4,4) → (6,6);	0.71		(2,3)→(5,7) →(6,10);	0
5	(2,0)→(5,4) →(6,7);	0.93	19	(2,3)→(7,6) →(4,7) → (6,9);	0.15
	(2,0)→(7,3) →(4,4) → (6,6);	0.07		(2,3)→(5,7) →(6,10);	0.85
6	(2,0)→(5,4) →(6,7);	0.93	20	(2,3)→(7,6) →(4,7) → (6,9);	0.62
	(2,0)→(7,3) →(4,4) → (6,6);	0.07		(2,3)→(5,7) →(6,10);	0.38
7	(2,0)→(5,4) →(6,7);	0.58	21	(2,3)→(7,6) →(4,7) → (6,9);	0
	(2,0)→(7,3) →(4,4) → (6,6);	0.42		(2,3)→(5,7) →(6,10);	1
8	(2,0)→(5,4) →(6,7);	0.13	22	(2,3)→(7,6) →(4,7) → (6,9);	0
	(2,0)→(7,3) →(4,4) → (6,6);	0.87		(2,3)→(5,7) →(6,10);	1
9	(3,0)→(6,8);	0.08	23	(2,3)→(7,6) →(4,7) → (6,9);	0.36
	(3,0)→(8,3) →(5,4) → (6,7);	0.92		(2,3)→(5,7) →(6,10);	0.64
10	(3,0)→(6,8);	0.56	24	(3,3)→(6,11);	0.09
	(3,0)→(8,3) →(5,4) → (6,7);	0.44		(3,3)→(8,6) →(5,7) →(6,10);	0.91
11	(3,0)→(6,8);	0.93	25	(3,3)→(6,11);	0.55
	(3,0)→(8,3) →(5,4) → (6,7);	0.07		(3,3)→(8,6) →(5,7) →(6,10);	0.45
12	(1,3)→(7,6) →(4,7) → (6,9);	0.94	26	(3,3)→(6,11);	0.45
	(1,3)→(6,11);	0.06		(3,3)→(8,6) →(5,7) →(6,10);	0.55
13	(1,3)→(7,6) →(4,7) → (6,9);	0.50	27	(3,3)→(6,11);	0.85
	(1,3)→(6,11);	0.50		(3,3)→(8,6) →(5,7) →(6,10);	0.15
14	(1,3)→(7,6) →(4,7) → (6,9);	0.76	28	(3,3)→(6,11);	0.93
	(1,3)→(6,11);	0.24		(3,3)→(8,6) →(5,7) →(6,10);	0.07

In step 2, when the linear programming is directly solved by CPLEX, the objective function (minimal passenger flow on arc (8,5,3,4)) is 150. When Dantzig-Wolfe decomposition is applied to generate extreme points for time-dependent OD pairs, the minimal passenger flow is 142.5 based on the estimated measurements by Frank-Wolfe algorithm rather than by the NLP solver. The generated extreme points (feasible paths) and the correspondingly optimal weights are listed in Table 5-10. However, if the estimated measurements in step 1 are directly obtained from the NLP solver, the final minimal passenger count on arc (8,5,3,4) from Dantzig-Wolfe algorithm is 150 as well.

5.6.6 Experiment on a Large-scale Transit Network

In order to address the computational challenges in large-scale networks, we will propose an approximation-based approach, which provides a k-shortest path set as extreme points for each passenger group (in each OD pair with time-dependent departure time) in advance rather than using Dantzig-Wolfe decomposition to generate extreme point iteration by iteration. In this section, the public Google Transit Feed Specification (GTFS) data from Alexandria Transit Company in 2015 is used as our tested large-scale transit network (https://transitfeeds.com/p/alexandria-transit-company). As shown in Figure 5-13, it has 12 routes, 1638 trips (866 trips on weekdays, 423 trips on Saturdays, 261 trips on Sundays, and 88 trips on the Christmas day), and 629 stops.

In this experiment, the trips on weekdays are only considered as the provided schedule. Then, 32,029 vertexes and 713,650 arcs are generated in the corresponding space-time network for one whole weekday. The arcs include vehicle running arcs, passengers' walking arcs from origin to transit stops and from transit stops to destination, transfer arcs,

and waiting arcs. The space-time arc generation rules contain that (i) the trip (path) travel time is less than 120min; (ii) the maximum number of transfer times is 3; (iii) the maximum transfer/walking time is 30min; (iv) the maximum transfer/walking distance is 0.5mile.

Figure 5-13 Alexandria Transit Network Read from GTFS, in Virginia, USA

In order to obtain the time-dependent transit demand, we map the traffic analysis zones in the city of Alexandria to the transit network as the activity locations. As a result, 42 OD pairs are matched. Plus, the time period of 7:00am to 9:00am is divided by 15 time intervals, so the time-dependent OD demand is defined by each 5 mins. Finally, 1484 time-dependent OD pairs are obtained based on the arc generation rules above.

In addition, as an approximation for those extreme points in Dantzig-Wolfe decomposition for each time-dependent OD pair, we generate 3-shortest paths using our developed k-shortest path algorithm. Finally, 4452 paths will be generated with 7,868 arcs in the space-time network. The general process for each time-dependent OD pair is shown as follows.

(i) Based on the origin vertex (origin node and departure time) in the space-time network, the label correcting algorithm can be used to generate a shortest path tree from origin vertex to all possible vertexes selected on the basis of the space-time arc generation rules.

(ii) According to the destination physical location, we can find a number of candidate vertexes (stop id and stop time in schedule) connecting the destination node by walking arcs. Then we can add the label costs of those candidate vertexes and its corresponding walking arc costs to the destination, so the destination will have a number of vertexes (destination node and arrival time) with different label cost.

(iii) Sort those label costs of the destination node and select k least-cost destination vertexes and back trace to the origin vertex. Finally, the k-shortest path set will be generated for one time-dependent OD pair.

For simplicity, we assume that all transit vehicle capacity is 35 and the walking, waiting and transfer arc capacity is 9999. Also, the time-dependent demand of each OD pair is assumed to be 1, which means that one passenger will arrive every 5 mins for each OD pair. The observed passenger trip time is assumed and generated as a random value between the minimal and the maximal path costs of 3-shorest paths. Focusing on the uncertainties of passenger flow state on transfer links from stop 370 to stop 553 and from stop 447 to stop 290 based on the 3-hour transit demand, our models are solved by CPLEX in GAMS

as a linear programming problem on the workstation with Intel(R) Xeon(R) CPU E5-2680 v2 @ 2.8GHz processors. For each model, there are 10,837 equations and 4,452 variables, and the computation time is around 19 seconds. The results are shown in Figure 5-14.

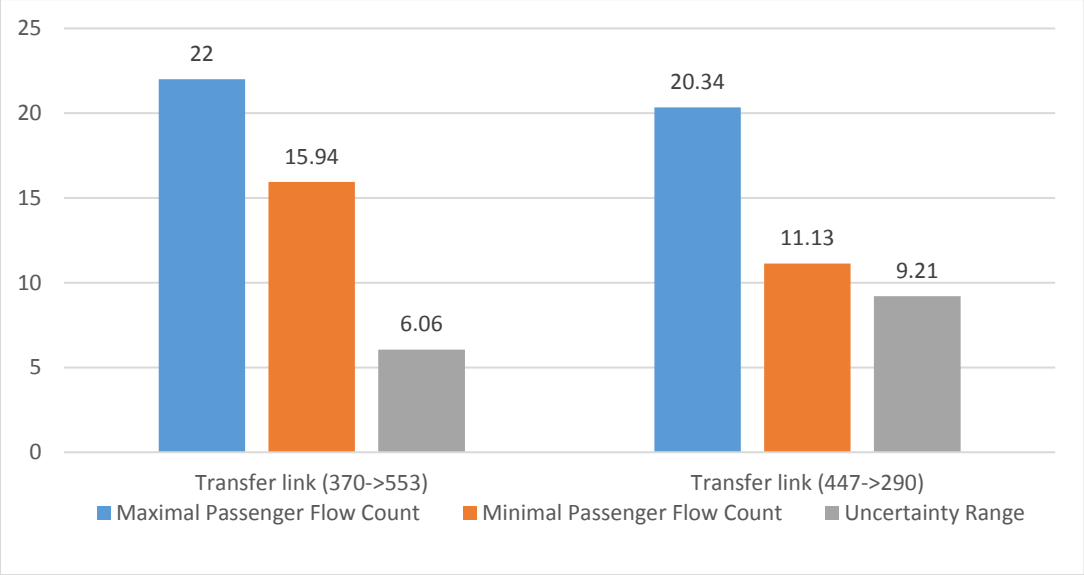


Figure 5-14 Uncertainties of Passenger Flow Count on Two Transfer Links

CHAPTER 6 CONCLUSION AND FUTURE RESEARCH

6.1 Conclusions

This dissertation aims to study the passenger-focused scheduled transportation systems, including public transit systems and upcoming optimal planning systems of household daily scheduled activities, from the increased observability due more available multi-source sensor data, to the operational planning-level service network design, finally to the more efficiently shared mobility.

In the operational planning level, Chapter 3 reveals how the tight capacity constraints can invoke bounded rationality travel behavior and how to consider them in dynamic capacitated transit serviced network design. Finally, a single-level integer linear programming model is proposed and can be further decomposed as two sub-problems by Lagrangian decomposition, namely, a time-dependent least cost path problem and a 0-1 knapsack problem, for improve the computation efficiency. The numerical experiments demonstrate the impacts of different transit demand levels, agent's tolerance value/indifference band, and transit vehicle capacity on the final service network decisions. In addition, the transportation network design problem is usually modeled as a bi-level programming problem. The proposed framework from modelling to algorithm design is also helpful to problems requiring the similar bi-level programming structure, such as, dynamic tolling design problem, signal optimization problem, and dynamic OD demand matrix estimation problem.

Motivated by the recently emerging trend of self-driving vehicles and information sharing technologies, the household activity pattern problems are examined in Chapter 4. By

embedding a set of hard constraints into a well-structured (space-time and space-time-state) network structure, we reformulate two difficult cases using a person-based or vehicle-based network flow programming model with very few side constraints, which could be directly solved by some standard optimization solvers, or directly further solved by time-dependent state-dependent shortest path algorithms. Meanwhile, the tight road capacity is highly considered for capturing the underlying congestion effect for the cases above. The numerical experiments demonstrate the proposed methodology and analyze the impacts of different activity benefits on the final vehicle routing and household member activity selection.

Going back to transportation system observability quantification, the developments presented in Chapter 5 provide insights on the relationship among multi-source information, information space, state estimation, and estimation uncertainty quantification by taking the urban rail transit systems as the analysis object. The information space and information errors are highly respected for state estimation, and project functions-based approaches are presented to quantify the uncertainty of different states under same information space. It should be emphasized that the final estimated states and the uncertainty of state estimates need to be provided with the correspondingly specific total measurement estimation errors due to the different hard observed errors from heterogeneous data sources. Further, the proposed models can explain that the value of information highly relies on its aimed specific estimated states and sensor location rather than just its high volume. It provides the analysis base for how to better use available information for different state estimates and how to design the sensor network for future estimate improvement.

6.2 Future Research

Passenger-focused scheduled transportation system aims to serve passengers by shared rides, accurate schedules, and optimal activity completion planning, which are still challenges in theory and need to be further studied.

Specifically, the uncertainty quantification of state estimation is just the first step for better observing and controlling the system. The following questions are currently under my considerations for future research: (1) what is the balance among the uncertainty of different system states, the minimally needed information, and the required accuracy of future controls? (2) What is the balance of the sensor data cost, value of information and its computational efficiency in proposed models and algorithms? (3) How to integrate the heterogeneous sensor network design with the real-time system control? (4) How to visualize the real-time uncertainty of different system states in a straightforward way for the public.

For better designing transit service networks, it needs to (1) calibrate the boundedly rational behavior by taking additional data collection efforts, (2) consider the limited number of transfers accepted by each passenger, (3) address the responsive travel demand issue under different transit system performances, and (4) consider the future most likely seamless multimodal scheduled transportation system with autonomous vehicle applications. In order to implement more deployable service network design systems, it also needs to incorporate realistic transit route/mode/departure choice models, e.g., by specifically taking into account of parking availability (Ruan et al., 2016), stochastic transfer activities (Yang et al., 2016), and uncertain time-dependent passenger demand (Yin et al., 2016).

To optimally complete those household scheduled activities, our research needs to further consider vehicle selection (mode choice) and ride-sharing simultaneously so that an optimal travel pattern, such as driving alone or ride-sharing, can be found for each household. More importantly, when available vehicles are automated and may or may not belong to each household, it needs research to find out how to find the optimally solution for household level scheduled activities rather than single travel requests. Finally, it is also extremely crucial to accurately design and efficiently operate the future scheduled and shared multimodal transportation system with public transit and flexible autonomous vehicles.

REFERENCES

Abdul Aziz, H.M., Ukkusuri, S., 2013.An Approach to Assess the Impact of Dynamic Congestion in Vehicle Routing Problems. In: Advances in Dynamic Network Modeling in Complex Transportation Systems. Springer US, 265-285.

Alsger, A., Assemi, B., Mesbah, M., Ferreira, L., 2016. Validating and improving public transport origin–destination estimation algorithm using smart card fare data. Transportation Research Part C, 68, 490-506.

Arentze, T. A., Timmermans, H. J., 2004. A learning-based transportation oriented simulation system. Transportation Research Part B 38(7), 613–633

Ban, X., Herring, R., Margulici, J.D., Bayen, A., 2009. Optimal sensor placement for freeway travel time estimation. Transportation and Traffic Theory, Chapter 34 (W.H.K. Lam, S.C. Wong, H.K. Lo eds.), Springer, 697-721.

Ban, X.J., Hao, P. and Sun, Z., 2011. Real time queue length estimation for signalized intersections using travel times from mobile sensors. Transportation Research Part C: Emerging Technologies, 19(6), 1133-1156.

Bard, J., Moore, J., 1990. A branch and bound algorithm for the bilevel programming problem. SIAM Journal on Scientific and Statistical Computing 11 (2), 281–292.

Barry, J., Freimer, R. and Slavin, H., 2009. Use of entry-only automatic fare collection data to estimate linked transit trips in New York City. Transportation Research Record: Journal of the Transportation Research Board, (2112), 53-61.

Beckmann, M. J., McGuire, C. B., Winsten, C. B., 1956. Studies in the Economics of Transportation. Yale University Press, New Haven, CT.

Bhat, C. R., Guo, J. Y., Srinivasan, S., Sivakumar, A., 2004. Comprehensive econometric micro-simulator for daily activity-travel patterns. Transportation Research Record: Journal of the Transportation Research Board 1894(1), 57–66.

Bianco, L., Confessore, G. and Reverberi, P., 2001. A network based model for traffic sensor location with implications on O/D matrix estimates. Transportation Science, 35(1), 50-60.

Bierlaire, M., 2002. The total demand scale: a new measure of quality for static and dynamic origin–destination trip tables. Transportation Research Part B: Methodological, 36(9), 837-850.

Boyles, S., Waller S. T., 2011. Optimal information location for adaptive routing. Networks and Spatial Economics 11, 233-254.

Castillo, E., Conejo, A.J., Menéndez, J.M. and Jiménez, P., 2007. The observability problem in traffic network models. Computer- Aided Civil and Infrastructure Engineering, 23(3), 208-222.

Castillo, E., Conejo, A.J., Pruneda R., Solares C., 2008. Observability in linear systems of

equations and inequalities: Applications. Computer & Operations Research, 34, 1708-1720.

Castillo, E., Nogal, M., Rivas, A. and Sánchez-Cambronero, S., 2013. Observability of traffic networks. Optimal location of counting and scanning devices. Transportmetrica B: Transport Dynamics, 1(1), 68-102.

Castillo, E., Grande, Z., Calviño, A., Szeto, W.Y. and Lo, H.K., 2015. A state-of-the-art review of the sensor location, flow observability, estimation, and prediction problems in traffic networks. Journal of Sensors, 2015.

Cats, O., West, J. and Eliasson, J., 2016. A dynamic stochastic model for evaluating congestion and crowding effects in transit systems. Transportation Research Part B: Methodological, 89, 43-57.

Cepeda, M., Cominetti, R., Florian, M., 2006. A frequency-based assignment model for congested transit networks with strict capacity constraints: characterization and computation of equilibrium. Transportation Research Part B 40 (6), 437–459.

Chow, J. Y., Recker, W. W., 2012. Inverse optimization with endogenous arrival time constraints to calibrate the household activity pattern problem. Transportation Research Part B 46(3), 463–479.

Chow, J.Y., Djavadian, S., 2015. Activity-based market equilibrium for capacitated multimodal transport systems. Transportation Research Part C 59, 2-18.

Codina, E. and Rosell, F., 2017. A heuristic method for a congested capacitated transit assignment model with strategies. Transportation Research Part B: Methodological, 106, pp.293-320.

Correa, J. R., Schulz, A. S., Stier Moses, N. E., 2004. Selfish routing in capacitated networks. Mathematics of Operations Research 29(4), 961–976.

Danczyk, A., Liu, H., 2011. A mixed-integer linear program for optimizing sensor locations along freeway corridors. Transportation Research Part B, 45 (1) 208–217.

Dantzig, G.B. and Wolfe, P., 1960. Decomposition principle for linear programs. Operations research, 8(1), 101-111.

De Cea, J. and Fernández, E., 1993. Transit assignment for congested public transport systems: an equilibrium model. Transportation science, 27(2), 133-147.

Desrosiers, J. and Lübbecke, M.E., 2005. A primer in column generation. In Column generation (1-32). Springer US.

Di, X., He, X., Guo, X., Liu, H.X., 2014. Braess paradox under the boundedly rational user equilibria. Transportation Research Part B 67, 86–108.

Di, X., Liu, H.X., Ban, X., 2016. Second best toll pricing within the framework of bounded rationality. Transportation Research Part B 83, 74–90.

Di, X., Liu, H.X., Pang, J.S., Ban, X., 2013. Boundedly rational user equilibria (BRUE): mathematical formulation and solution sets. Transportation Research Part B 57, 300–313.

Drezner, Z., Wesolowsky, G.O., 2003. Network design: selection and design of links and facility location. Transportation Research Part A 37 (3), 241–256.

Drissi-Kaïtouni, O., Hameda-Benchekroun, A., 1992. A dynamic traffic assignment model and a solution algorithm. Transportation Science 26, 119–128.

Eisenman, S., M., Fei, X., Zhou, X., Mahmassani, H. S., 2006. Number and Location of Sensors for Realtime Network Estimation and Prediction: A Sensitivity Analysis. Transportation Research Record 1964, 253-259.

Facchinei, F., Kanzow, C., 2010. Generalized Nash equilibrium problems. Annual of Operations Research. 175, 177–211.

Farahani, R.Z., Miandoabchi, E., Szeto, W., Rashidi, H., 2013. A review of urban transportation network design problems. European Journal of Operational Research 229 (2), 281–302.

Farvaresh, H., Sepehri, M.M., 2011. A single-level mixed integer linear formulation for a bi-level discrete network design problem. Transportation Research Part E 47 (5), 623–640.

Florian, M., Hearn, D., 1999. Network equilibrium and pricing. In Handbook of Transportation Science. Springer US, 361-393.

Frank, M. and Wolfe, P., 1956. An algorithm for quadratic programming. Naval Research Logistics (NRL), 3(1- 2), 95-110.

Fu, X., Lam, W.H., Xiong, Y., 2016. Modelling intra-household interactions in household's activity-travel scheduling behaviour. Transportmetrica A: Transport Science 12(7), 1-21.

Gan, L.P., Recker, W., 2008. A mathematical programming formulation of the household activity rescheduling problem. Transportation Research Part B 42 (6), 571–606.

Gao, L., Wang, B., and Zhang, C., 2011. Comparison of urban rail transit fare clearing model based on travel survey. Modern urban transit, 11, 97-99.

Gao Z, Wu J, Sun H, 2005. Solution algorithm for the bilevel discrete network design problem. Transportation Research Part B 39, 479–95.

Gao, Z., Sun, H., Shan, L., 2003. A continuous equilibrium network design model and algorithm for transit systems. Transportation Research Part B 38, 235– 250.

Gentili, M. and Mirchandani, P.B., 2012. Locating sensors on traffic networks: Models, challenges and research opportunities. Transportation research part C: emerging technologies, 24, 227-255.

Gentili, M., Mirchandani, P. B., 2005. Locating Active Sensors on Traffic Networks. Annals of Operations Research 136 (1): 229-257.

Ghali, M.O., Smith, M.J., 1995. A model for the dynamic system optimum traffic assignment problem. Transp. Res. Part B 29 (3), 155–170.

Gorard, S., 2005. Revisiting a 90-year-old debate: the advantages of the mean deviation. British Journal of Educational Studies, 53(4), 417-430.

Guihaire, V., Hao, J.K., 2008. Transit network design and scheduling: A global review. Transportation Research Part A 42 (10), 1251–1273.

Guo, X., Liu, H., 2011. Bounded rationality and irreversible network change. Transportation Research Part B 45 (10), 1606–1618.

Hamdouch, Y., Lawphongpanich, S., 2008. Schedule-based transit assignment model with travel strategies and capacity constraints. Transportation Research Part B 42, 663–684.

Hamdouch, Y., Marcotte, P., Nguyen, S., 2004. Capacitated transit assignment with loading priorities. Mathematical Programming 101B, 205–230.

Hamdouch, Y., Szeto, W.Y., and Jiang, Y., 2014. A New Schedule-Based Transit Assignment Model with Travel Strategies and Supply Uncertainties. Transportation Research Part B 67, 35-67.

Han, K., Piccoli, B., Friesz, T.L., 2015c. Continuity of the path delay operator for dynamic network loading with spillback. Transportation Research Part B, DOI: 10.1016/j.trb.2015.09.009.

Han, K., Sun, Y., Liu, H., Friesz, T.L., Yao, T., 2015b. A bi-level model of dynamic traffic signal control with continuum approximation. Transportation Research Part C 55, 409–431.

Han, K., Szeto, W.Y., Friesz, T.L., 2015a. Formulation, existence, and computation of boundedly rational dynamic user equilibrium with fixed or endogenous user tolerance. Transportation Research Part B 79, 16–49.

Hearn, D. W., 1982. The gap function of a convex program. Operations Research Letters 1, 67-71.

Hearn, D.W., Ribera, J., 1980. Bounded flow equilibrium problems by penalty methods. In Proceedings of the 1980 IEEE International Conference on Circuits and Computers, 162–166.

Hu, S.R., Peeta, S. and Chu, C.H., 2009. Identification of vehicle sensor locations for link-based network traffic applications. Transportation Research Part B: Methodological, 43(8-9), 873-894.

Ibarra-Rojas, O.J., Delgado, F., Giesen, R., Muñoz, J.C., 2015. Planning, operation, and control of bus transport systems: a literature review. Transportation Research Part B 77, 38–75.

Jiang, Y. and Szeto, W.Y., 2016. Reliability-based stochastic transit assignment: Formulations and capacity paradox. Transportation Research Part B: Methodological, 93, 181-206.

Kalman, R., 1959. On the general theory of control systems. IRE Transactions on Automatic Control 4(3), 110-110

Kalvelagen, E., 2003. Dantzig-Wolfe Decomposition with GAMS.

Kang, J.E., Chow, J.Y., Recker, W.W., 2013. On activity-based network design problems. Transportation Research Part B 57:398-418

Kang, J.E., Recker, W.W., 2013. The location selection problem for the household activity pattern problem. Transportation Research Part B 55, 75-97.

Kepaptsoglou, K., Karlaftis, M., 2009. Transit route network design problem: review. Journal of Transportation Engineering 135 (8), 491–505.

Kusakabe, T. and Asakura, Y., 2014. Behavioural data mining of transit smart card data: A data fusion approach. Transportation Research Part C: Emerging Technologies, 46, 179-191.

Kusakabe, T., Iryo, T. and Asakura, Y., 2010. Estimation method for railway passengers' train choice behavior with smart card transaction data. Transportation, 37(5), 731-749.

Koenker, R. and Hallock, K.F., 2001. Quantile regression. Journal of economic perspectives, 15(4), 143-156

Lam, W.H.K., Gao, Z.Y., Chan, K.S., Yang, H., 1999. A stochastic user equilibrium assignment model for congested transit networks. Transportation Research Part B 33, 351–368.

Larsson, T. and Patriksson, M., 1992. Simplicial decomposition with disaggregated representation for the traffic assignment problem. Transportation Science, 26(1), 4-17.

Larsson, T., Patriksson, M. and Rydergren, C., 2004. A column generation procedure for the side constrained traffic equilibrium problem. Transportation Research Part B: Methodological, 38(1), 17-38.

Larsson, T., Patriksson, M., 1995. An augmented Lagrangean dual algorithm for link capacity side constrained traffic assignment problems. Transportation Research Part B 29, 433-455.

Larsson, T., Patriksson, M., 1998. Side constrained traffic equilibrium models—traffic management through link tolls. Equilibrium and advanced transportation modelling. Springer US, 125-151.

LaValle, S.M., 2012. Sensing and filtering: A fresh perspective based on preimages and information spaces. Foundations and Trends® in Robotics, 1(4), 253-372.

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

Li, P., Mirchandani, P., Zhou, X., 2015. Solving simultaneous route guidance and traffic signal optimization problem using space-phase-time hypernetwork. Transportation Research Part B 81(1), 103–130.

Li, X., Ouyang, Y. 2011. Reliable sensor deployment for network traffic surveillance. Transportation Research Part B 45 (1), 218–231.

Liao, F., Arentze, T., Timmermans, H., 2013a. Incorporating space–time constraints and activity-travel time profiles in a multi-state supernetwork approach to individual activity-travel scheduling. Transportation Research Part B 55, 41–58.

Liao, F., Arentze, T., Timmermans, H., 2013b. Multi-state supernetwork framework for the two-person joint travel problem. Transportation, 40(4), 813-826.

Lin, D.Y., Eluru, N., Waller, S., Bhat, C., 2008. Integration of activity-based modeling and dynamic traffic assignment. Transportation Research Record: Journal of the Transportation Research Board 2076, 52-61.

Liu, J., Kang, J.E., Zhou, X. and Pendyala, R., 2017. Network-oriented household activity pattern problem for system optimization. Transportation Research Part C: Emerging Technologies.

Liu, J., Zhou, X., 2016. Capacitated transit service network design with boundedly rational agents. Transportation Research Part B 93, 225-250.

Lo, Andrew W. 2013. The origin of bounded rationality and intelligence. Proceedings of the American Philosophical Society 157, 269–280.

Lou, Y., Yin, Y., Lawphongpanich, S., 2009. Robust approach to discrete network designs with demand uncertainty. Transportation Research Record 2090, 86–94.

Lou, Y., Yin, Y., Lawphongpanich, S., 2010. Robust congestion pricing under boundedly rational user equilibrium. Transportation Research Part B 44 (1), 15–28.

Lu, C.C., Liu, J., Qu, Y., Peeta, S., Rouphail, N.M., Zhou, X., 2016. Eco-system optimal time-dependent flow assignment in a congested network. Transportation Research Part B 94, 217–239.

Lu, C.C., Zhou, X., Zhang, K., 2013. Dynamic origin–destination demand flow estimation under congested traffic conditions. Transportation Research Part C, 34, 16-37.

Luathep, P., Sumalee, A., Lam, W.H., Li, Z.C., Lo, H.K., 2011. Global optimization method for mixed transportation network design problem: a mixed-integer linear programming approach. Transportation Research Part B 45 (5), 808–827.

Ma, X.L., Liu, C., Liu, J.F., Chen, F. Yu, H., 2015. Boarding stop inference based on transit IC card data. Journal of Transportation Systems Engineering and Information Technology, 15(4), 78-84.

Ma, X., Wu, Y.J., Wang, Y., Chen, F. and Liu, J., 2013. Mining smart card data for transit riders' travel patterns. Transportation Research Part C: Emerging Technologies, 36, 1-12

Ma, X.L., Wang, Y.H., Chen, F., Liu, J.F., 2012. Transit smart card data mining for passenger origin information extraction. Journal of Zhejiang University Science C, 13(10), 750-760.

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

Mahmassani, H., Chang, G., 1987. On boundedly rational user equilibrium in transportation systems. Transportation Science 21 (2), 89–99.

Mahmoudi, M., Chen, J., Zhou, X., 2016. Embedding Assignment Routing Constraints through Multi-Dimensional Network Construction for Solving the Multi Vehicle Routing Problem with Pickup and Delivery with Time Windows. arXiv:1607.01728.

Mahmoudi, M., Zhou. X, 2016. Finding Optimal Solutions for Vehicle Routing Problem with Pickup and Delivery Services with Time Windows: A Dynamic Programming Approach Based on State-space-time Network Representations. Transportation Research Part B 89, 19-42.

Marcotte, P., Nguyen, S., Schoeb, A., 2004. A strategic flow model of traffic assignment in static capacitated networks. Operations Research 52 (2), 191–212.

Martello, S., Toth, P., 1990. Knapsack Problems: Algorithms and Computer Implementation, John Wiley and Sons.

Meng, L. and Zhou, X., 2011. Robust single-track train dispatching model under a dynamic and stochastic environment: a scenario-based rolling horizon solution approach. Transportation Research Part B: Methodological, 45(7), 1080-1102.

Miller, E. J., Roorda, M. J., 2003. Prototype model of household activity-travel scheduling. Transportation Research Record: Journal of the Transportation Research Board 1831 (1), 114–121.

Mokhtarian, P. L., Chen, C., 2004. TTB or not TTB, that is the question: a review and analysis of the empirical literature on travel time (and money) budgets. Transportation Research Part A 38(9), 643-675.

Morikawa, T., Miwa, T., Kurauchi, S., Yamamoto, T., Kobayashi, K., 2005. Driver's route choice behavior and its implications on network simulation and traffic assignment. Simulation Approaches in Transportation Analysis, 341–369.

Munizaga, M.A. and Palma, C., 2012. Estimation of a disaggregate multimodal public transport Origin–Destination matrix from passive smartcard data from Santiago, Chile. Transportation Research Part C: Emerging Technologies, 24, 9-18.

Nassir, N., Hickman, M. and Ma, Z.L., 2015. Activity detection and transfer identification for public transit fare card data. Transportation, 42(4), 683-705.

Ng, M., 2012. Synergistic sensor location for link flow inference without path enumeration: A node-based approach. Transportation Research Part B: Methodological, 46(6), 781-788.

Nguyen S., Pallottino, S., 1988. Equilibrium Traffic Assignment for Large Scale Transit Networks. European Journal of Operations Research 37(2), 176–186.

Nguyen, S., Pallottino, S., Malucelli, F., 2001. A modeling framework for passenger assignment on a transport network with timetables. Transportation Science 35, 238–249.

Nie, Y., Wu, X., Zissman, J., Lee, C., Haynes, M., 2010. Providing reliable route guidance:

Phase ii. Technical report, Center for the Commercialization of Innovative Transportation Technology, Northwestern University, Evanston, IL.

Nie, Y., Zhang, M., Lee, D.H., 2004. Models and algorithm for the traffic assignment problem with link capacity constraints. Transportation Research Part B 38 (4), 285–312.

Niu, H., Zhou, X., 2013. Optimizing urban rail timetable under time-dependent demand and oversaturated conditions. Transportation Research Part C 36, 212–230.

Niu, H., Zhou, X., Gao, R., 2015. Train scheduling for minimizing passenger waiting time with time-dependent demand and skip-stop patterns: nonlinear integer programming models with linear constraints. Transportation Research Part B 76, 117–135.

Nunes, A.A., Dias, T.G. and e Cunha, J.F., 2016. Passenger journey destination estimation from automated fare collection system data using spatial validation. IEEE transactions on intelligent transportation systems, 17(1), 133-142.

Nuzzolo, A., Crisalli, U., 2009. The schedule-based modeling of transportation systems: recent developments. In Schedule-Based Modeling of Transportation Networks (1-26). Springer, Boston, MA.

Nuzzolo, A., Crisalli, U., Rosati, L., 2012. A schedule-based assignment model with explicit capacity constraints for congested transit networks. Transportation Research Part C 20(1), 16-33.

Nuzzolo, A., Russo, F., Crisalli, U., 2001. A doubly dynamic schedule-based assignment model for transit networks. Transportation Science 35, 268–285.

Pallottino, S., Scutellà, M. G., 1998. Shortest path algorithms in transportation models: classical and innovative aspects. Equilibrium and Advanced Transportation Modelling, (Eds.) Marcotte, P., Nguyen, S., Kluwer Academic Publishers, MA, 245-281.

Pelletier, M.P., Trépanier, M. and Morency, C., 2011. Smart card data use in public transit: A literature review. Transportation Research Part C: Emerging Technologies, 19(4), 557-568.

Pendyala, R., Kitamura, R., Kikuchi, A., Yamamoto, T., Fujii, S., 2005. Florida activity mobility simulator: overview and preliminary validation results. Transportation Research Record: Journal of the Transportation Research Board 1921(1), 123–130.

Pendyala, R., Konduri, K., Chiu, Y.C., Hickman, M., Noh, H., Waddell, P., Wang, L., You, D., Gardner, B., 2012. Integrated land use-transport model system with dynamic time-dependent activity-travel microsimulation. Transportation Research Record: Journal of the Transportation Research Board 2303, 19-27.

Pendyala, R., You, D., Garikapati, V., Konduri, K., Zhou, X., 2017. Paradigms for integrated modeling of activity-travel demand and network dynamics in an era of dynamic mobility management. The 96th Annual Meeting of the Transportation Research Board, accepted for presentation only.

Peeta, S. and Mahmassani, H.S., 1995. Multiple user classes real-time traffic assignment

for online operations: a rolling horizon solution framework. Transportation Research Part C: Emerging Technologies, 3(2), 83-98.

Poon, M.H., Wong, S.C., Tong, C.O., 2004. A dynamic schedule-based model for congested transit networks. Transportation Research Part B 38, 343–368.

Poorzahedy, H., Rouhani, O.M., 2007. Hybrid meta-heuristic algorithms for solving network design problem. European Journal of Operational Research 182 (2), 578–596.

Pribyl, O., Goulias, K.G., 2005. Simulation of daily activity patterns incorporating interactions within households: Algorithm overview and performance. Transportation Research Record: Journal of the Transportation Research Board 1926, 135-141.

Psaraftis, H.N., 1983. An exact algorithm for the single-vehicle many-to-many dial-a-ride problem with time windows. Transportation Science 17 (3), 351–357.

Qu, Y., Zhou, X., 2017. Large-scale dynamic transportation network simulation: a space-time-event parallel computing approach. Transportation Research Part C 75, 1-16.

Ramming, M.S., 2001. Network knowledge and route choice. Ph.D. Thesis. Massachusetts Institute of Technology.

Recker, W.W., 1995. The household activity pattern problem: general formulation and solution. Transportation Research B 29 (1), 61-77.

Recker, W.W., 2001. A bridge between travel demand modelling and activity-based travel analysis. Transportation Research Part B 35(5), 481–506.

Recker, W.W., Chen, C., McNally, M.G., 2001. Measuring the impact of efficient household travel decisions on potential travel time savings and accessibility gains. Transportation Research Part A 35 (4), 339-369.

Rosenthal, R., 2015. GAMS: A User's Guide. GAMS Development Corporation.

Ruan, J.M., Liu, B., Wei, H., Qu, Y., Zhu, N. and Zhou, X., 2016. How many and where to locate parking lots? A space–time accessibility-maximization modeling framework for special event traffic management. Urban Rail Transit, 2(2), 59-70.

Seaborn, C., John, A., and Nigel H. M. Wilson.2009. Analyzing Multimodal Public Transport Journeys in London with Smart Card Fare Payment Data. Transportation Research Record: Journal of the Transportation Research Board 2121, 55–62.

Spiess, H., Florian, M., 1989. Optimal strategies: a new assignment model for transit networks. Transportation Research Part B 23 (2), 83–102.

Sun, Y., Xu, R., 2012. Rail transit travel time reliability and estimation of passenger route choice behavior: Analysis using automatic fare collection data. Transportation Research Record (2275), 58-67.

Szeto, W.Y, 2003. Dynamic traffic assignment: formulations, properties, and extensions. PhD Thesis, The Hong Kong University of Science and Technology, H.K.

Szeto, W.Y., Jiang, Y., 2014a. Transit assignment: Approach-based formulation, extragradient method and paradox. Transportation Research Part B 62, 51–76.

Szeto, W.Y., Jiang, Y., 2014b. Transit route and frequency design: Bi-level modeling and hybrid artificial bee colony algorithm approach. Transportation Research Part B 67, 235-263.

Tong, C. O., Richardson, A. J., 1984. A computer model for finding the time- dependent minimum path in a transit system with fixed schedules. Journal of Advanced Transportation, 18(2), 145-161.

Tong, L., Zhou, L., Liu, J., Zhou, X., 2017. Customized bus service design for jointly optimizing passenger-to-vehicle assignment and vehicle routing. Transportation research part C, 85, 451-475.

Tong, L., Zhou, X., Miller, H.J., 2015. Transportation network design for maximizing space–time accessibility. Transportation Research Part B 81(2), 555–576.

Trépanier, M., Tranchant, N. and Chapleau, R., 2007. Individual trip destination estimation in a transit smart card automated fare collection system. Journal of Intelligent Transportation Systems, 11(1), 1-14.

Verbas, I., Mahmassani, H., 2015. Integrated frequency allocation and user assignment in multi-modal transit networks: methodology and application to large-scale urban systems. Transportation Research Record 2498, 37-45.

Wang D.Z.W., Liu H, Szeto W. Y., 2015. A novel discrete network design problem formulation and its global optimization solution algorithm. Transportation Research Part E 79, 213-230.

Wang, D.Z.W., Lo, H.K., 2010. Global optimum of the linearized network design problem with equilibrium flows. Transportation Research Part B 44 (4), 482–492.

Wang, S., Meng, Q., Yang, H., 2013. Global optimization methods for the discrete network design problem. Transportation Research Part B 50, 42–60.

Wei, Y., Avcı, C., Liu, J., Belezamo, B., Aydın, N., Li, P.T. and Zhou, X., 2017. Dynamic programming-based multi-vehicle longitudinal trajectory optimization with simplified car following models. Transportation Research Part B: Methodological, 106, pp.102-129.

Xing, T., Zhou, X., Taylor, J., 2013. Designing Heterogeneous Sensor Networks for Estimating and Predicting Path Travel Time Dynamics: An Information-Theoretic Modeling Approach. Transportation Research Part B, 57, 66-90.

Xiong, Y., Schneider, J.B., 1992. Transportation network design using a cumulative algorithm and neural network. Transportation Research Record 1364, 37–44.

Xu, W., He, S., Song, R., and Sohail, S., 2012. Finding the K shortest paths in a schedule-based transit network. Computers & Operations Research 39(8), 1812–1826.

Xu, X., Liu, J., Li, H., and Hu, J. 2014. Analysis of subway station capacity with the use

of queueing theory, Transportation Research Part C 38, 28–43

Xu, X., Lo, H.K., Chen, A. and Castillo, E., 2016. Robust network sensor location for complete link flow observability under uncertainty. Transportation Research Part B: Methodological, 88, 1-20.

Yang, H., Bell, M.G.H., 1998. Models and algorithms for road network design: a review and some new developments. Transport Reviews 18 (3), 257–278.

Yang, H., Zhou, J., 1998. Optimal Traffic Counting Locations for Origin-Destination Matrix Estimation. Transportation Research Part B 32, 109-126

Yang, H., Iida, Y. and Sasaki, T., 1991. An analysis of the reliability of an origin-destination trip matrix estimated from traffic counts. Transportation Research Part B: Methodological, 25(5), 351-363.

Yang, L., Zhang, Y., Li, S., Gao, Y., 2016. A two-stage stochastic optimization model for the transfer activity choice in metro networks. Transportation Research Part B 83, 271-297.

Yen, J., 1971. Finding the k shortest loopless paths in a network. Management Science 17, 712–716.

Yin, J., Tang, T., Yang, L., Gao, Z., Ran, B., 2016. Energy-efficient metro train rescheduling with uncertain time-variant passenger demands: An approximate dynamic programming approach. Transportation Research Part B 91, 178-210.

Yuan, N.J., Wang, Y., Zhang, F., Xie, X. and Sun, G., 2013, December. Reconstructing individual mobility from smart card transactions: A space alignment approach. In Data Mining (ICDM), 2013 IEEE 13th International Conference, 877-886.

Zhou, F., Xu, R.H., 2012. Model of passenger flow assignment for urban rail transit based on entry and exit time constraints. Transportation Research Record (2284), 57-61.

Zhou, L., 2014. Verification and optimization of network clearing model. Urban mass transit, 11, 59-66.

Zhou, X., List, G.F., 2010. An information-theoretic sensor location model for traffic origin-destination demand estimation applications. Transportation Science, 44(2), 254-273.

Zhou, X. and Mahmassani, H.S., 2007. A structural state space model for real-time traffic origin–destination demand estimation and prediction in a day-to-day learning framework. Transportation Research Part B: Methodological, 41(8), 823-840.

Zhou, X., Mahmassani, H. S., Zhang, K., 2008. Dynamic micro-assignment modeling approach for integrated multimodal urban corridor management. Transportation Research Part C 16(2), 167-186.

Zhou, X., Taylor, J., 2014. DTALite: A queue-based mesoscopic traffic simulator for fast model evaluation and calibration. Cogent Engineering 1 (1), 961345.

Zhu, S., 2011. The Roads Taken: Theory and Evidence on Route Choice in the Wake of the I-35w Mississippi River Bridge Collapse and Reconstruction. Ph.D. thesis, University

of Minnesota.

Zhu, Y., Koutsopoulos, H.N. and Wilson, N.H., 2017a. A probabilistic Passenger-to-Train Assignment Model based on automated data. Transportation Research Part B: Methodological, 104, 522-542.

Zhu, Y., Koutsopoulos, H.N. and Wilson, N.H., 2017b. Inferring left behind passengers in congested metro systems from automated data. Transportation Research Part C: Emerging Technologies.

Ziliaskopoulos, A., Mahmassani, H. S., 1993. A time-dependent shortest path algorithm for real-time intelligent vehicle/highway systems applications. Transportation Research Record 1408, 94-100.

APPENDIX A

KKT CONDITION OF THE MODIFIED BMW MODEL

$$\sum_{(i,j)\in L}(c_{i,j} + \delta_{(\mathrm{i,j})}^{w,p} \times \mu_{i,j}) \geq \pi_w \,, \forall\, p \in P_w, \forall (i,j) \in L \qquad \text{(A.1)}$$

$$\mu_{i,j} \geq 0, \forall (i,j) \in L \qquad \text{(A.2)}$$

$$\sum_{p\in P_w} h_{w,p} = d_w, \forall w \in W \qquad \text{(A.3)}$$

$$h_{w,p} \geq 0, \forall\, p \in P_w, \forall w \in W \qquad \text{(A.4)}$$

$$\sum_{w\in W} \sum_{p\in P_w} (\delta_{(\mathrm{i,j})}^{w,p} \times h_{w,p}) \leq Cap_{i,j}, \forall (i,j) \in L \qquad \text{(A.5)}$$

$$h_{w,p}\left(c_{i,j} + \delta_{(\mathrm{i,j})}^{w,p} \times \mu_{i,j} - \pi_w\right) = 0, \forall\, p \in P_w, \forall w \in W, \forall (i,j) \in L \qquad \text{(A.6)}$$

$$\mu_{i,j}\left(Cap_{i,j} - \sum_{w\in W} \sum_{p\in P_w} \left(\delta_{(\mathrm{i,j})}^{w,p} \times h_{w,p}\right)\right) = 0, \forall (i,j) \in L \qquad \text{(A.7)}$$

APPENDIX B

MULTI-LOOP LABEL-CORRECTING ALGORITHM

$L_{i,t,w}(v) := 0$ // label cost at vertex (i, t, w) for vehicle v

$L_{j,s,w'}(v) := +\infty$; for each vertex $(j, s, w') \in R - \{(i, t, w)\}$ // label cost at vertex (j, s, w') for vehicle v

node pred of vertex $(.,.,.,.) := -1$;

time pred of vertex $(.,.,.,.) := -1$;

state pred of vertex $(.,.,.,.) := -1$;

$LIST := \{(i, t, w)\}$;

While $LIST \neq \emptyset$ do

for each time $t \in [0, T]$ do // adding time window of each activity can further reduce the searching region

begin

for each state w do // the number of states can be reduced by the activity sequence for each passenger

begin

for each link (i, j) do //

begin

derive downstream state w' based on the feasible state transition

derive arrival time $s = t + TT_{i,j,t,}$;

if $(L_{i,t,w}(v) + c_{i,j,t,s,w,w'}(v) < L_{j,s,w'}(v))$

begin

$L_{j,s,w'} := L_{i,t,s}(v_u) + c_{i,j,t,s,w,w'}(v)$; // label update

node pred of vertex $(v, j, s, w') := i$;

time pred of vertex $(v, j, s, w') := t$;

state pred of vertex $(v, j, s, w') := w$;

if vertex $(j, s, w') \notin LIST$ then add vertex (j, s, w') to $LIST$

end;

end; // for each link

end; // for each state

end; // for each time