On Dynamic Traffic Assignment in Corridor Networks under Heterogeneous Travelers and Modes

By

Zhen Qian
B.S. (Tsinghua University) 2004
M.S. (Tsinghua University) 2006

Dissertation

Submitted in partial satisfaction of the requirements for the degree of

Doctor of Philosophy

in

Civil and Environmental Engineering

in the

Office of Graduate Studies

of the

University of California

Davis

Approved:

H. Michael Zhang, Chair

Roger Wets, Professor

Yueyue Fan, Professor

Committee in Charge

2011

To Hao, and my parents

Contents

List of Figures

List of Tables

Abstract of the Dissertation

On Dynamic Traffic Assignment in Corridor Networks under Heterogeneous Travelers and Modes

This dissertation investigates traveler heterogeneity for dynamic traffic assignment (DTA) in the following four dimensions: travelers' attributes (in the value of time and the value of schedule delay), modal choice, parking choice and route choice. The main focus is on obtaining analytical DTA solutions in simplified networks, particularly in the context of the morning commute problem, with precise sensitivity analysis to derive effective traffic congestion management policies.

First, we solve the morning commute problem with a heterogeneous traveling population whose early/late arrival penalty are continuously distributed. The distribution of the value of schedule delay on each route, freeway or the arterial road, is discussed. It is found that the assumption of homogeneity population overestimates the queuing delay and the total travel time. Every commuter is better off if the freeway capacity or arterial capacity is enlarged, but commuters with high values of early/late arrival penalty generally benefit more than those with low values unless they switch to other routes. We further study the multi-modal morning commute problem with three modes, transit, solo-driving and carpool. Enlarging HOV facilities may reduce transit ridership and increase auto travel, and it does not necessarily reduce the total travel cost when the network is highly congested. The rise of gas price may first entice auto travelers to carpool. However, as the gas price increases further, both carpoolers and solo-drivers will eventually switch to use the transit. In addition, a flat freeway tool can also reduce the total network travel cost.

In addition to the intrinsic distinction among travelers, we also discuss the management measures that can distinguish travelers externally, using parking as an example. The parking fee, parking capacity allocation and accessibility altogether can effectively reduce both the system cost and the queuing delay. If parking lots are owned publicly, then all travelers are better off under the optimal parking setting. This is an advantage that cannot be realized by the system-optimum dynamic toll scheme. If they are owned

privately, then market regulations, such as price-ceiling and quantity tax/subsidy, are suggested to improve the network performance and reduce the congestion.

We finally extend our research to the DTA problem in general networks. We propose a hybrid route choice model for studying non-equilibrium traffic where travelers have different preferences in choosing travel routes. It combines pre-trip route choice and en-route route choice to solve dynamic traffic assignment (DTA) in large-scale networks. We apply the hybrid route choice model in a synthetic medium-scale network and a large-scale real network to assess its effect on the flow patterns and network performances, and compare them with those obtained from Predictive User Equilibrium (PUE) DTA. The proposed route choice model incorporating route choice heterogeneity is capable of solving DTA efficiently in in a realistic size network with satisfactory results. Finally, some suggestions are given on how to calibrate the hybrid route choice model in practice.

Acknowledgments

My deepest gratitude is to my advisor and mentor, Professor H. Michael Zhang, for his guidance and help throughout my PhD study. I have been very fortunate to have an advisor who really supported me in my pursuit of an academic career and encouraged me to explore on my own. Prof. Zhang not only taught me how to conduct research and write papers/grant proposals, but also inspired me greatly to find new ideas and being passionate about the research.

I am grateful to other dissertation committee members, Professor Yueyue Fan and Professor Roger Wets. Both of them also served in my qualify exam committee. They have been always there to listen and give advice. I was fortunate to take Dr. Wets' optimization courses where I take the very first step to my research. His lectures are the most instructive and interesting classes I have ever taken. Dr. Fan has guided me into the world of dynamic programming and the start-of-the-art learning techniques. She also generously shared with me her experience in pursuing her academic career.

I am also grateful to Professor Patricia L. Mokhtarian who provided valuable supports at several stages in my PhD study. I have taken three courses of hers and they are the most elegant courses I had, and she had so many helpful comments on my homework that I benefit greatly from.

I would like to acknowledge my fellow students, Haining Du, Jingtao Ma, Wei Shen, Yi-ru Chen, Martha Shott, Jia Li, Hui Deng, Shikai Tang, Robert Lim and Rachel Carpenter for their valuable help during my PhD study. Thanks also go to my friends who have always been supporting me both professionally and personally. They include but are not limited to: Marco Nie, Pengcheng Fu, Feng Xiao, Song Bai, Jason Cao, Shuang Liu, Peng Wu, Yongxi Huang, Changzheng Liu and Steven Chen.

Last but not least, I would like to thank my beloved wife, Hao Chen and my parents, to whom I dedicate this dissertation. There is nothing more important than a supportive family to my academic career.

This dissertation was supported by a grant from the Sustainable Transportation Center at the University of California Davis, which receives funding from the U.S. Department of Transportation and Caltrans, the California Department of Transportation, through the University Transportation Centers program.

Chapter 1

Introduction

1.1 Background and Motivation

Traffic congestion has been one of the major urban problems. *2010 Urban Mobility Report and Appendices* (2010) reported that traffic congestion, costing Americans 87.2 billion dollars in wasted fuel and lost productivity or $750 per traveler in 2009, is getting worse since 1982. There is apparently no quick and easy solution to congestion. To reverse this trend, Advance Transportation Management System (ATMS), Intelligent Transportation System (ITS) and other dynamic traffic analysis tools are developed intensively in the past decades. Although those systems are developed in a variety of forms, they all require the ability to model the dynamic traffic flow, and they all aim to accurately forecast/estimate traffic conditions on the network, assess the traffic management measures, or evaluate infrastructure investments. The essential idea of those tools is to either make improvement in network infrastructure, the supply side, or manage traffic, the demand side.

Consequently, it is vital to model simultaneously both the demand side and supply side as the foundation of dynamic traffic analysis. Dynamic Traffic Assignment (DTA) serves as one of the fundamental tools. It is an estimation or prediction technique for the time-spatial traffic flow distribution on the network.

From the supply point of view, urban networks, in particular corridor networks, are usually the modeling focus. A corridor network is defined as “a largely linear geographic band defined by existing and forecasted travel patterns involving both people and goods” (FHWA 2006). It is a special network where parallel roadways and transit lines are the

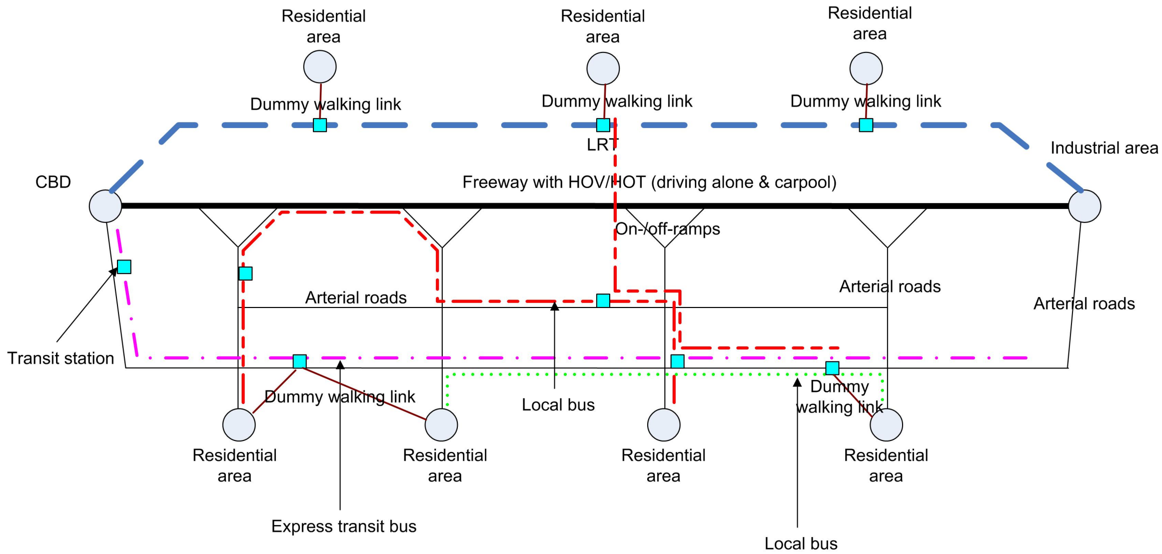


Figure 1.1. A general multi-modal corridor network

main facilities to serve the travel demand from residential areas to employment centers, such as a Central Business District (CBD). Corridors often play an important role in serving commuting travel demand and are often subject to moderate to severe congestion. A corridor network, shown in Figure 1.1, combines various transportation assets, such as freeway networks, arterial roads, local street and transit stations. The Origin-Destination (O-D) nodes are mostly the centroids of CBD areas, industrial areas or residential areas. The freeway connects those O-D nodes to the CBD, and the arterial roads, parallel to the freeway, are local streets connecting those centroids together as well. However, the arterial roads typically have much lower speed limits than the freeway, and thus offers longer travel times for travelers. Besides the roadways for automobile travelers, bus and/or light rail (LRT) services are often provided in a corridor network. Typically the LRT service uses its own guideway but the bus lines share the same roadway network with the private automobile.

On the travel demand side, travelers usually choose from different transportation modes, such as private automobile (driving alone or carpooling) and public transit. A traveler can choose any one or a combination of these modes in his trip, which brings about competition for passengers across modes. Which mode a traveler chooses depends, to a large degree, on the relative costs of using those different modes and the availability/accessibility of transportation vehicles. Generally, the travel demand varies over time during the day, but is believed to be relatively stable after day-to-day adjustment unless incidents occur.

Given the transportation network supply and traveler demand, DTA determines the traffic flow for any physical location and travel times/costs for any traveler on the network by a pre-determined rule of modal choice, departure time choice and route choice. DTA also yields the overall assessment of the network performance, the information at the aggregated level provided for decision makers. Numerous studies has been devoted to DTA over the years, but most of them focus on the automobile travelers while assuming a traveler homogeneity in selecting routes and departure times. The deficiency is obvious. Overlooking traveler heterogeneity with regards to population attributes and travel behavior leads to a rough estimation of the network performance, and the absence of multi-modal transportation modeling results in a heavy bias towards automobile networks. More importantly, the models with homogeneous travelers are only capable of evaluating those traffic management measures applied to all travelers equally and anonymously. In fact, group-specific traffic management measures could be much more efficient in reducing the congestion, which also enables the analysis of social equity.

This dissertation aims to improve dynamic traffic assignment by considering several critical issues of traveler heterogeneity in the scope of corridor networks. The reason of choosing a corridor network, rather than general network or a complete set of urban network, is simple. A corridor network, though concerning special transportation forms, is a multi-modal network with general traffic management problems, and possess many essential features of a general network. More importantly, it has several research advantages: 1) The travel demand in the corridor network consists of mostly commuting traffic during the rush hours, so it is very likely to be stable after day-to-day adjustment. 2) Numerous data have been collected in corridor networks. Compared to general networks, corridor networks are often equipped with sufficient detectors, which provides unique advantages for model calibration and verification. 3) Express transit lines and LRT lines are most likely to be available along a corridor. 4) It makes possible the abstraction of the network to a highly simplified network where we can derive analytical solutions and unique insights.

Traveler heterogeneity in DTA can be categorized by the following criteria,

- Traffic modes. First, each traffic mode is preferred by certain travelers. Some travelers may be captive to a traffic mode, while others may switch between modes freely. Second, travelers using different modes take different routes. On top of the modal choices of the transit and automobile, solo drivers, carpoolers and park-and-ride drivers, either of which is also a separate mode, have different preference on selecting routes and departure times. There might be network access restrictions for travelers using different modes.

- Availability of network information to travelers. Travelers may have different knowledge of the network and real-time traffic information.

- Vehicle attributes. This is also known in vehicle type in the literature. Vehicle attributes is probably the most popular classification in the studies of microscopic traffic models. Vehicles with different attributes can make significant difference in the motion of traffic flow. They can be classified roughly by cars, light-duty trucks and heavy-duty truck, or in more details by maximum speed, size and acceleration/deceleration rates.

- Traveler attributes. Traveler attributes refer to not only the driving characteristics (e.g., in terms of aggressiveness) but also travelers' perception in travel costs. Travelers naturally have their respective preferences on certain factors, such as particular roads, locations, traffic delay, travel time, schedule delay, etc.

Although each type of heterogeneity may have been intensively studied as a separate subject in transportation studies, the heterogeneity is not well considered in the context of DTA. For instance, the modal choice is the focus of discrete choice models and vehicle type heterogeneity is widely studied in microscopic simulation models. Nevertheless, both types of heterogeneity has not received full consideration in DTA.

In addition to the intrinsic distinction among travelers, travelers can also be distinguished externally. Many transportation management measures attempt to separate travelers using a variety of ways so as to increase the efficiency of the measures. One example to distinguish travelers anonymously is to offset the office time for different commuters to relieve the traffic peak. It is also believed that toll rationing, dividing travelers into groups and each group is charged a toll in certain days (Daganzo & Garcia 2000), could reduce the peak congestion. With travelers classified into different groups, it is then even more natural to set different management measures for different groups so as to achieve the least congestion possible.

Although there are many possible perspectives that one can use to look into the traveler heterogeneity problem, this dissertation focuses on just a few crucial factors. The essential idea of this dissertation is to build a more realistic DTA model by considering traveler heterogeneity in the following four categories,

- Traveler modal choice preference
- Traveler attributes in value of time and value of schedule delay
- Traveler's preference on routes
- Heterogeneity resulted from external management measures

The first three are intrinsic heterogeneity, while the last one is one type of management measures that produce external heterogeneity. The objective of this research is to reveal the connections between travelers' choice and those four factors such that targeted management measures can be proposed to optimize the transportation network performance.

1.2 Problem Statement

For narrative convenience, we first introduce the basic notations in DTA. Let a directed graph $G(N, A)$ denote a general network, where N and A are the sets of nodes and directed links respectively. Let R and S denote the set of origins and destinations respectively, K_{rs} the set of paths joining an OD pair rs, and $K = \bigcup K_{rs}, \forall r \in R, s \in S$. Let $[0, T]$ be an *assignment horizon* (i.e., the analysis period). The network is assumed to be empty at $t = 0$, and only travel demands departing within the assignment horizon are considered. Let $q_{rs}(t)$ be the travel demand between O-D pair rs departing at time t, and the total demand for the whole assignment horizon is

$$q_{rs} = \int_0^T q_{rs}(t)dt \tag{1.1}$$

Further, we use $c_a(t)$ to denote the generalized travel cost on link a at the entry time t, and $c_{rs}^k(t)$ to denote the generalized travel cost on path $k \in K_{rs}$ departing at time t. $\mathbf{q}^t = \{q_{rs}^t, r \in R, s \in S\} \forall t \in [0,T]$ is the travel demand pattern at time interval t. Let $\mathbf{f} = \{f_{rs}^{kt}, r \in R, s \in S, t \in [0,T], k \in K_{rs}\}$, $\mathbf{x} = \{x_a^t, a \in A, t \in [0,T]\}$ and $\mathbf{c(f)} = \{c_{rs}^{kt}, r \in R, s \in S\, t \in [0,T], k \in K_{rs}\}$ denote the time-dependent path flow pattern, loading pattern and path cost pattern. The generalized path travel cost is defined as,

$$c_{rs}^k(t) = \alpha \text{fft}(t) + \alpha w(t) + \max\{\beta(t^* - t - w(t)), \gamma(t + w(t) - t^*)\} + \sum_{i=1}^{I} \lambda_i w_{rs}^{i,kt} \quad (1.2)$$

where fft(t) denotes the free-flow travel time of a vehicle on path k departing at time t, while $w(t)$ denotes its travel delay. t^* is the desired arrival time (work start time) for this vehicle. Here α, β and γ measure the generalized cost of one extra minute of travel time, early schedule delay and late schedule delay, respectively. In addition to schedule delay cost and travel time cost, the generalized cost also consists of I number of terms $(w_{1,kt}^{rs}, w_{2,kt}^{rs} \ldots, w_{I,kt}^{rs})$ which represent other types of travel costs that travelers perceive on path p of O-D pari rs departing at time t (such as tolls, gas fees, etc.) and each is weighted by a scaler λ_i.

DTA is usually solved by imposing the user equilibrium (UE) constraints, also known as the extension of the Wardrop's first principle in the dynamic network (Wardrop 1952), which states that every traveler of the network travels on the path and departs at the time with the least cost and cannot unilaterally reduce his path cost by switching to any other path or departure time. The UE condition reads,

$$\text{if } \mathbf{q^t} \text{ is known} \begin{cases} f_{rs}^{kt}(c_{rs}^{kt} - \pi_{rs}^t) = 0, \forall r, s, k, t \in [0,T] \\ c_{rs}^{kt} \geq \pi_{rs}^t, f_{rs}^k \geq 0 \end{cases} \quad (1.3)$$

$$\text{if } \mathbf{q^t} \text{ is unknown} \begin{cases} f_{rs}^{kt}(c_{rs}^{kt} - \pi_{rs}) = 0, \forall r, s, k, t \in [0,T] \\ c_{rs}^{kt} \geq \pi_{rs}, f_{rs}^k \geq 0 \end{cases}$$

where π_{rs}^t is the least path travel cost between O-D pair $r - s$ departing at time t, while π_{rs} is the least path travel cost between O-D pair $r-s$ over the entire assignment horizon.

The above DTA problem using UE as the route choice and/or departure time choice constraint can be casted into a variational inequality (VI) problem (Friesz et al. 1993),

$$\text{Find } \mathbf{f}^* \in \Omega, \text{ such that } \forall \mathbf{f} \in \Omega : \langle \mathbf{c}(\mathbf{f}^*)), \mathbf{f} - \mathbf{f}^* \rangle \geq 0 \quad (1.4)$$

$$\Omega = \{\mathbf{f} \in \mathrm{R}^l_+ | \mathbf{M}\mathbf{f} = \mathbf{q}\} \tag{1.5}$$

where $l = m_a \times \sum_r \sum_s |K_{rs}|$. m_a is the number of assignment time intervals and $\mathbf{M}$ is the path-OD incidence matrix.

The traveler heterogeneity being discussed in this dissertation is essentially embedded in the generalized travel cost function, i.e. Equation 1.2. First, the generalized travel cost function is attached with different terms, $w^{i,kt}_{rs}$, for travelers using different modes. On the other hand, even with the same traffic mode, a traffic management measure may distinguish travelers externally, which attaches different cost terms for different travelers in a similar fashion. We use parking as an example of such traffic management measures. For each traveler, his generalized travel cost determines the probability of choosing each mode, and furthermore the response to the traffic management measures (for example detour or trip cancelation). Second, the value of time (VOT), α, and the value of schedule delay, β and γ, are distinct across the travel population. Third, travelers have different preference on certain factors. In other words, λ_i is also distinct across the travel population. This dissertation will choose one particular problem for each of these categories, discuss the effects of heterogeneity and further propose the policy indication for decision makers.

According to the solution type, conventional DTA equipped with UE conditions, is usually categorized into two classes, *analytical* DTA and *simulation-based* DTA. Analytical DTA looks into highly simplified networks. The solution of the analytical DTA, such as time-varying path flow, can be expressed in explicit formulas, and thus the evaluation of the network performance and sensitivity analysis over certain factors are precise. The classic morning commute problem (Arnott et al. 1988) is a commonly studied analytical DTA problem. The simulation-based DTA, on the other hand, usually do not have analytical solutions and rely on simulations to obtain numerical solutions, but are applicable to general networks. When the network is large, the numerical solution could be very rough since the solution algorithm does not guarantee the achievement of certain routing objectives, such as user equilibrium. Therefore, it is extremely difficult to conduct precise sensitivity analysis for the simulation-based DTA.

Each type of DTA has its own advantages and weaknesses. The merit of the analytical DTA lies not in terms of methodological advancement, but in terms of deriving analytical solutions to study how different factors affect the flow pattern and network performance, so that more insightful discussions can be carried out. Numerical DTA is capable of solving practical problems and can be applied in real large-scale networks, but its ability in capturing the effects of detailed factors is very limited. As a starting stage of studying traveler heterogeneity, the majority of this dissertation follows the analytical DTA approach. It is desirable to learn precisely the effects of heterogeneity before implementing new models for large networks. This dissertation will also study the heterogeneity of driving attributes in selecting routes for the simulation-based DTA because the routing criteria being discussed is directly associated with the generalized travel cost function, Equation 1.2. This will be the beginning of research in implementing traveler heterogeneity in DTA for large-scale networks.

1.3 Contributions

This dissertation advances DTA in several ways.

- Multi-modal morning commute problem is solved analytically. The effects of a variety of dynamic pricing schemes, such as HOV lane setting, transit fare, roadway toll, gass fee and carpool impedance, are discussed intensively. This has many significant policy indications for decision-makers.

- A general distribution of VOT and value of schedule delay are considered in the DTA. This not only helps estimate flow and network performance more accurately, but also enables the analysis of social equity issues, e.g., how different the infrastructure improvement can bring travel time reduction to different travelers.

- Using parking as a traffic management measure, the effects of parking accessibility, availability and fees to both travelers and networks are revealed. We show that parking is capable of reducing the congestion and improve system performance efficiently.

- Last, but not least, the proposed heterogeneous route choice model may be realistic in describing travelers' route choice. The relationship between route choices and queuing patterns was discussed in both theoretical and practical perspectives This allows easy calibration and efficient computation of large-scale networks and possible reproduction of realistic traffic conditions.

1.4 Organization

The structure of this dissertation is as follows. We first present an overview of DTA problems in Chapter 2. The focus is on route choice models and its applications in congestion pricing. We also review the studies on traveler heterogeneity in the literature, including multi-modal multi-class traffic assignment problems. Chapter 3 discusses the distribution of value of time and value of schedule delay in the morning commute. Chapter 4 looks into the DTA problem from the traffic mode point of view. We solve the multi-modal morning commute problem and reveal its policy indications. Then, in Chapter 5, we intensively discuss how the parking accessibility, availability and fees can be used to induce travel heterogeneity and thus reduce the congestion efficiently. This is further followed by Chapter 6 where we propose a new route choice model implemented in large-scale networks. Finally, conclusions are drawn in Chapter 7 with possible future research subjects.

Chapter 2

Literature Review

Originated from the traffic UE definition (Wardrop 1952) and its equivalent mathematical optimization problem (Beckmann et al. 1956), traffic assignment problem has received numerous attentions over the years. The steady-state traffic assignment models are generally used for transportation planning purposes by assuming the traffic conditions on the network are constant. DTA can capture the flow propagation on the network dynamically and thus is more accurate than the steady-state traffic assignment. However, the accuracy does not come without a price. DTA models, if applied in large-scale networks, are usually much more complicated than the steady-state ones and its solutions, even in the context of UE, is not guaranteed. In this chapter, we review the main DTA models, including analytical DTA for the morning commute problem and simulation-based DTA for the general networks. It is common that both types of DTA models use UE as the route choice model. We shall also review a few studies concerning disequilibrium route choice model. Furthermore, focus is given to the three travel heterogeneity categories to be discussed in this dissertation, multi-modal multi-class traffic assignment, VOT heterogeneity, as well as parking regulation as one of the traffic management measures.

2.1 Dynamic Traffic Assignment

2.1.1 Morning commute problem

The problem of eliminating congestion in the morning commute with a time-varying congestion toll dates back to (Vickrey 1969). In his work, Vickrey considered travelers' commuting pattern on a single route with a single bottleneck for a single mode. Suppose the bottleneck has a capacity of s and the total travel demand is N. He showed that there exists an equilibrium departure time pattern, shown in Figure 2.1, if commuters all attempt to minimize their own travel costs, which include the travel time cost and an early/late arrival penalty, and in this equilibrium pattern all commuters incur the same travel cost no matter when he starts his trip. By making commuters choose their departure time based on their marginal costs, which included a congestion externality, he showed that congestion can be completely eliminated in the morning commute.

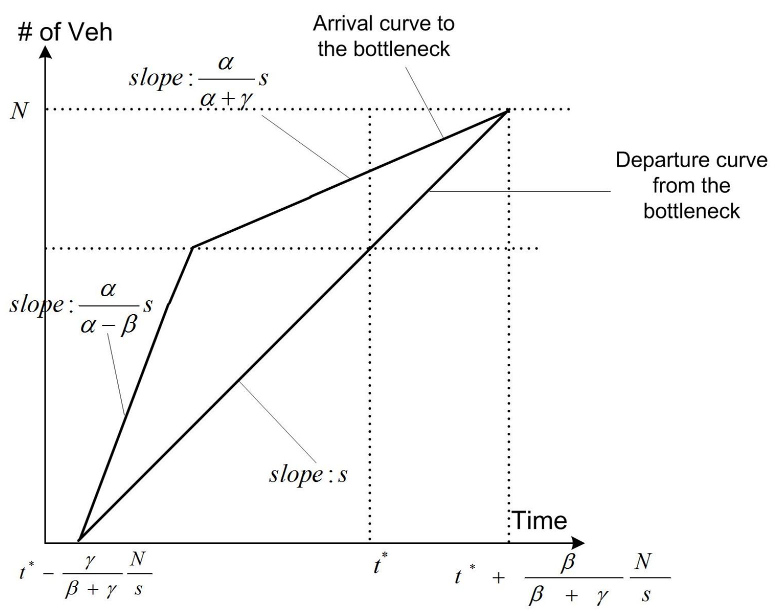


Figure 2.1. Cumulative curves derived from Vickrey's morning commute model

Following this pioneering work, numerous efforts have been devoted to study various extensions of Vickrey's morning commute problem (Arnott et al. 1990, 1988, e.g.), mostly focused on the single mode of auto travel. Henderson (1977) first established a speed-density relationship model for a single-origin-single-destination (SOSD) network and then it was improved by Chu (1994). SOSD with two routes was proposed to study the simultaneous departure time and route choices equilibrium (Mahmassani & Herman 1984), and thereafter, Kuwahara (1990) studied a problem with two origins and a single destination with two continuous bottlenecks. In addition to those studies, another study found a paradox, similar to Braess Paradox, that expanding bottleneck capacity can raise total travel costs, if users only choose when to travel based on fixed routes (Arnott et al. 1993).

2.1.2 General networks

2.1.2.1 A brief review

Route choice, departure time choice and traffic flow evolution are three essential elements of a DTA problem. The traffic flow models will be briefly discussed in Section 2.1.2.4. According to the route/departure time choice criteria, DTA is categorized into system optimal (SO) assignment and user equilibrium (UE) assignment. SO-DTA is to find a temporal and spatial flow pattern that minimizes the total system cost (Merchant & Nemhauser 1978*a*,*b*, Carey 1986, 1987, Friesz et al. 1989, e.g.). The formulation of SO-DTA reads,

$$\min \sum_a \int_0^{T'} x_a(t)dt \qquad (2.1)$$
$$\text{s.t. } \mathbf{f} \in \mathrm{R}^l_+, \mathbf{M}\mathbf{f} = \mathbf{q}, \boldsymbol{\Delta}\mathbf{f} = \mathbf{x}$$

where Δ is the link-path incidence matrix. SO-DTA yields the least total system cost and optimal flow pattern. If travelers are left alone and there is no policy to guide them, SO-DTA is usually not achievable in a network. However, it serves a benchmark of network performance for traffic operational schemes and this problem is at the core of many transportation applications ranging from day-to-day traffic management to emergency evacuations.

In order to estimate or predict the realistic traffic flow pattern, the route choice is usually formulated as a UE flow pattern, i.e. travelers simply choose the cheapest or shortest route that is present to them. There are generally two types of UE in the literature. One is the so-called Boston User Equilibrium (BUE) (Friesz et al. 1993), which is an adaption of the static Wardroppian UE. It assumes a traveler chooses the shortest route only based on the prevailing traffic condition at the time of his choice decision (Ran et al. 1993, Kuwahara & Akamatsu 2001), also known as the minimum instantaneous cost path (Ghali 1995). The other UE type is the so-called Predictive User Equilibrium (PUE). Under this behavioral assumption, travelers choose the shortest route based on "anticipated" travel times, or travel times that they actually experienced from previous days. The result is a UE in which the actual travel times/costs for travelers from any O-D pair are minimal and identical (Friesz et al. 1993), regardless of the routes they take.

2.1.2.2 BUE (en-route) route choice

The DTA in Boston User Equilibrium can be decomposed to all-or-nothing static traffic assignment for each time interval (Kuwahara & Akamatsu 2001). In other words, travelers will first determine their routes based on free-flow travel time of all the links in the network, and then the network is loaded by the O-D demand according to the route choices till the first assignment time interval ends. At the beginning of the next time interval, travelers update their route choices by taking the shortest paths with respect to the prevailing (instantaneous) travel times/costs at that time. Flow on and travel cost of each link will then be updated based on the new route choice decisions. Repeat this process where the update of route choices and network loading alternate till all the travelers end their trips. Thus, only one network loading is needed and each traveler act as if he makes en-route decisions of route to be taken at each assignment time interval according to the instantaneous travel times/costs. Therefore, we also call the route choice model embedded in the BUE en-route route choice.

The en-route route choice assumes that travelers only have real-time information about the current network conditions, and since they do not predict traffic conditions in the future, they always make the route choice decision that is most appealing at the current time interval, although the chosen route can be far from the best route when the actual travel time/cost is considered. This type of en-route route choice may describe the travelers' behavior in response to a sudden change of demand or roadway capacity where travelers are unable to obtain historical day-to-day information and have to deviate from their pre-trip routes. With the adoption of variable message boards, highway advisory radio and mobile internet devices such as smart phones, travelers nowadays can get up-to-the-minute information in some travel market such that they can make en-route route adjustments in response to changes in network traffic conditions rather than sticking to their pre-determined routes.

2.1.2.3 PUE (pre-trip) route choice

PUE assumes travelers choose their routes prior to their departure, and then strictly follow them till they reach their destinations. Their choices are based on the shortest paths with respect to actual travel time/cost they experience. On one hand, a dynamic network loading (DNL) procedure is used to determine the time-varying link flow and the actual travel time/cost, and then to determine the routes traveled; On the other hand, the DNL procedure requires knowing the route choices of all travelers prior to their departure. Therefore, PUE usually requires an iterative process in which DNL and shortest path calculation alternate until both the route choice and flow patterns converge.

According to PUE, all travelers can perfectly predict the traffic conditions of the entire network at any time based on their day to day experiences. They are aware of their actual travel time/cost prior to their departure which is used to determine their shortest routes, regardless of real-time traffic information. This type of route choice is more suitable to model regular commuting traffic without disruptions such as incidents for a prolonged period of time during which the total traffic demand remains relatively constant.

PUE is usually formulated in a VI problem as shown in Equation 1.4. Many algorithms have been developed to solve such a VI problem. One of the commonly used one is projection-type algorithms. A VI problem $VI(\mathbf{c}, \Omega)$ can be transformed to solve the following problem (Nagurney 1999):

$$\max_{\mathbf{f}\in\Omega} \theta(\mathbf{f}) \tag{2.2}$$

where $\theta(\mathbf{f})$ is defined as a nonnegative merit function (gap function):

$$\theta(\mathbf{f}) = \min_{\mathbf{g}\in\Omega} \langle \mathbf{c}(\mathbf{f}), \mathbf{g} - \mathbf{f} \rangle \tag{2.3}$$

Assume $\mathbf{g}^* - \mathbf{f}$ is a descent direction along which $\theta(\mathbf{f})$ can be reduced, where

$$\mathbf{g}^* = argmin_{\mathbf{g}\in\Omega}\langle \mathbf{c}(\mathbf{f}), \mathbf{g} - \mathbf{f}\rangle$$

Note that $\mathbf{g}^* - \mathbf{f}$ is exactly the direction used in F-W algorithm for separable link cost cases. However, in the general case, it does not always hold. Given a possible descent direction, we may apply some heuristic algorithms, such as MSA (Sheffi 1985), to get a convergent solution. The gap function may not be differentiable (Facchinei & Pang 2003), we may define a regularized merit function (Fukushima 1992):

$$\theta_r(\mathbf{f}) = \min_{\mathbf{g}\in\Omega}\{\langle \mathbf{c}(\mathbf{f}), \mathbf{g} - \mathbf{f}\rangle + \frac{\lambda}{2}\langle \mathbf{g} - \mathbf{f}, G(\mathbf{g} - \mathbf{f})\rangle\} \tag{2.4}$$

Where λ is a positive scalar and G is a symmetric P.D. matrix. If $\mathbf{c}$ is Lipschitz continuous and strongly monotone on Ω, then $g_r^*(\mathbf{f}) - \mathbf{f}$ is a strict descent direction for θ_r at $\mathbf{f}$ whenever $\mathbf{f}$ is not a solution to $VI(\mathbf{c}, \Omega)$, and

$$g_r^*(\mathbf{f}) = Proj_\Omega[\mathbf{f} - \frac{1}{r}\mathbf{c}(\mathbf{f})]$$

Similarly to the MSA, some projection-type algorithms may be applied to converge to a solution, such as basic projection algorithm (BPA), extra-gradient method (EGM), hyperplane projection method (HPM), and the like (Nagurney 1999, Nie 2003, 2006).

A PUE problem for general networks is typically solved following the steps below:

Step 0 Initialization. Select an initial flow pattern $\mathbf{f}^\nu$, $\nu = 0$.

Step 1 Conduct dynamic network loading (DNL) based on $\mathbf{f}^\nu$. Obtain the time-varying travel times for each link.

Step 2 Column generation. Solve the time dependent shortest path problem (or the least costly path problem) for each O-D pair rs, $[p_{rs}^*, t_{rs}^*] = \text{argmin}_{p,t} c_{pt}^{rs}$.

Step 3 Update flow pattern $\mathbf{f}^{\nu+1}$ (e.g. the projection-type algorithms) based on the new path set.

Step 4 Convergence criteria (e.g., $||\mathbf{f}^{\nu+1} - \mathbf{f}^\nu|| < \epsilon$ where ϵ is a small positive real number). Terminate if convergent, go to Step 1 otherwise.

2.1.2.4 Traffic flow models

Even with the same route/departure time choice model, various flow models could result in considerably different flow patterns. In DTA, a dynamic network loading (DNL) procedure incorporating a variety of traffic dynamics models is used to determine the time-varying travel cost $\mathbf{c}(\mathbf{f})$. Such traffic dynamics models include microscopic models and macroscopic models. Microscopic models represent each individual vehicles in terms of position speed, acceleration/deceleration rate etc, and thus provide the most details with, however, heavy computation. It is not computationally efficient, sometime even not plausible, in solving DTA for large-scale networks under microscopic traffic flow models. On the other hand, macroscopic models does not represent individual vehicles. Their speed, acceleration/deceleration rate and other attributes are aggregated. This will be the focus of this dissertation.

In the literature, several macroscopic flow models was introduced to formulate the DTA problem, which include exit-flow function based traffic model (Merchant & Nemhauser 1978*a*), the delay-function based model (Friesz et al. 1993), the point queue (PQ) model (Smith 1993), the spatial queue (SQ) model (Kuwahara & Akamatsu 2001) and kinematic wave (LWR) model (Daganzo 1994). Nie & Zhang (2005) compared these various kinds of so-called link models and found that the kinematic wave model, which models queue spillback in the form of shock waves, provides a realistic description of traffic flow propagation. However, queue spillback in large-scale networks, though exists in the real world, may propagate in an unexpected way that could lead to serious network gridlock, which in most cases is caused by unrealistic route choices. Daganzo (1998) explored the case of queue spillover with a PUE route choice under the spatial-queue model, and showed that the results could be naturally chaotic in the sense that flow patterns may be fairly sensitive to small changes in the network settings. This indicates that the stability of the PUE state is closely related to route choice and queue spillback. Therefore, for large-scale networks, it is crucial to analyze how a certain route choice model interacts with the queue spillback mechanism such that it can be appropriately calibrated in real world applications. A more detailed literature review of the traffic flow propagation models can be found in Nie (2003).

2.1.3 Non-equilibrium route choice model

In real life, travelers' route choice behavior is likely to be more complex than what was assumed in both BUE and PUE. For example, travelers may not consider all the possible routes but have several pre-trip routes in mind prior to their departure, which are selected from their day-to-day traveling experiences. Moreover, these pre-selected routes may not be user-optimal ones. Although travel time and schedule delay costs are dominant factors in travelers' route choice decisions, several other factors, such as road accessibility, pavement conditions, and so on, may influence their decisions as well. Besides these factors, a traveler's personality should also play an important role in his or her route choice. For example, a conservative traveler may stick to his chosen route from day to day while an adventurous traveler may be more willing to explore new routes based on his actual travel experiences. Thus real traffic is more likely to be the product of various types of choice decisions rather than cost-minimizing BUE or PUE applied uniformly across the entire traveling population. It is therefore of particular interest to develop a route choice model that combines various types of information and considers various kinds of travelers. However, such route choice models, particularly in the context of traffic disequilibria in large-scale dynamic networks, are rare in the literature.

Peeta & Mahmassani (1995) are among the few who analyzed the combination of BUE and PUE. They distinguished travelers by their route choices following either of system optimal, UE, historical and real-time information, in order to predict the future O-D demands and to optimize the provision of real-time information. Pel et al. (2009) also adopted route choices other than BUE/PUE. They introduced a hybrid route choice model where all travelers have a pre-trip route, but they all consider real-time traffic conditions in seeking the new routes. This model, however, requires path enumeration and its application to large-scale networks is very limited. In addition, Kant (2008) conducted an interesting study that combines BUE and PUE to form a new type of UE. He assumes travelers' travel cost consists of the actual travel costs and the expected travel costs. Then the solution framework of a PUE is directly applied to solve the new "combined" UE. However, this approach requires a pre-determined route set with very limited number of routes for each O-D pair, and is thus also problematic in handling large-scale networks.

2.2 Multi-modal Multi-class Traffic Assignment

Multi-modal or multi-class traffic assignment has been well studied in the steady-state context, but its extension in DTA is rare in the literature.

2.2.1 Morning commute problem

Tabuchi (1993) is among the first to study such extensions in the morning commute problem. His network includes a roadway with a bottleneck and a railway considering economies of scale. He obtained optimal transit fares and road tolls under different system performance goals. Huang & Yang (1999*b*) investigates a parallel auto/transit SOSD network using optimal control theory. Another work in this direction is Huang (2000, 2002), where transit fares and road tolls were obtained to achieve system optimal under a bi-logit modal choice model for two groups of commuters.

Limited research on the morning commute problem with HOV lanes has been done. Based on an single-origin-single-destination network, Yang & Huang (1999) defined a pair of interactive travel time function for two groups of people, and showed the optimal toll considering carpool is significantly different from the original one. This was extended to compare different toll schemes in the presence of HOV lanes (Huang & Yang 1999*a*). However, the network performances under different HOV configurations were not directly compared in this study. In another study, a potential implementation of High-Occupancy-Toll (HOT) lanes is evaluated using the morning commute model (Dahlgren 2002).

2.2.2 General networks

For general networks, "multi-class" (i.e. a type of traveler heterogeneity) is addressed by introducing traveling groups that have different travel time/cost functions.

In the steady-state context, some scholars (Dafermos 1971, Smith 1979, e.g.) gave the general formulations of multi-class user equilibrium problems. Let M denote the set of user groups. Let x_a^m be the flow on link a associated with group m and $c_a^m(\mathbf{x})$ be the group-specific link travel cost function on link a, where $\mathbf{x} = (\mathbf{x}^m, m \in M)^T$ and $\mathbf{x}^m = (x_a^m, a \in A)^T$. Let $B_{rs}^m(\mathbf{q})$ denote the OD benefit (inverse demand) functions where $\mathbf{q} = (\mathbf{q}^m, m \in M)^T$ and $\mathbf{q}^m = (q_{rs}^m)^T$. A typical UE problem with multi-class users can be formulated as the following VI (Yang & Huang 2005, Patriksson 1994, Nagurney 1999): determine $(\mathbf{x}^*, \mathbf{q}^*)$ such that,

$$\sum_{m \in M} \sum_{a \in A} c_a^m(\mathbf{x}^*)(x_a^m - x_a^{m*}) - \sum_{k \in K} \sum_{rs} B_{rs}^m(\mathbf{q})(q_{rs}^m - q_{rs}^{m*}) \geq 0 \quad \forall (\mathbf{x}, \mathbf{d}) \in \Omega \tag{2.5}$$

$$\Omega = \{(\mathbf{x}^m, \mathbf{q}^m), m \in M | \mathbf{x}^m = \mathbf{\Delta f}^m, \mathbf{q}^m = \mathbf{M f}^m, \mathbf{f}^m \geq 0, \mathbf{q}^m \geq 0\} \tag{2.6}$$

This formulation is applicable for vehicle type classifications (e.g. fast vehicles vs. slow vehicles, trucks vs. cars, high-duty vehicles vs. light-duty vehicles) and mode type classifications (e.g. transit vs. private cars). Moreover, a multi-class traffic assignment problem may be reduced to a single-group problem by constructing a new network with k copies of the original network (Dafermos 1971, Nagurney 1999) where k is the number of traveler groups. Following this approach, Palma & Lindsey (2004) formulated a bi-level model with several toll schemes imposed to "reveal the separate influences of heterogeneity, network effects, fiscal effects and welfare-distributional weights", and found that assuming heterogeneity will significantly affect the social welfare in numerical examples.

This idea has been implemented for multi-class DTA by Bliemer (2000). He developed a series of travel cost function for multi-class travelers as the delay-function based traffic dynamics models, and introduced an additional dimension, traffic class indexed by m, in the VI formulation Equation 1.4. However, due to the nature of UE, FIFO must be maintained within each traveler class. The multi-class travel cost functions may violate the FIFO principle.

Although multi-modal traffic assignment could be solved as a special case of multi-class traffic assignment, it does overlook the characteristics of modal choice. One key issue is that modal choice may not strictly follow the UE principle, simply because travelers may be captive to some modes or they may have personal preference on a particular mode. There are basically two ways to resolve this issue.

One way is that we split the transit O-D and auto O-D in passenger units before we assign them separately to their own network. The two separate assignment methods are called auto assignment and transit assignment, respectively. Note that the transit assignment is in the unit of passenger, while the auto assignment is in the unit of standard vehicles. Here we simply apply a loading factor to covert the auto passenger flow into auto standard vehicles. The auto assignment easily follows the steady-state assignment and DTA discussed before, but the transit assignment is a bit different in terms of network modeling and solution algorithms. This is mainly because the public transit has fixed schedules unlike the arbitrary departure time for the automobile. In particular, there are three features that should be taken into account in the transit assignment, 1) waiting time at transit stations is a stochastic variable; 2) bus or trains normally stop at particular locations, i.e. the transit stations; and 3) people may walk or drive to the transit stations, and transfer between different lines. Steady-state transit assignment has been intensively developed (Spiess & Florian 1989, Cea & Fernandez 1993, Wu et al. 1994, Nguyen et al. 1998, Lam et al. 1999, e.g.,). Most of those studies use "hyperpath" to assign passenger flow, rather than a regular path in the auto assignment. A "hyperpath" records not only the travel cost on a path, but also the probability of taking a particular transit route and the "expected" travel cost of a path. Fortunately, dynamic transit assignment can directly follow the methodology used in the automobile DTA without adding the part of "hyperpath", because each traveler, once departed and marked with a particular time stamp, will take a transit vehicle that is also stamped with a departure time. The route choice can be deterministic in a dynamic context rather than being probabilistic as in the static case. Nevertheless, the DNL procedure for transit vehicles should be modified slightly (Tong & Wong 1999).

The other way is to assign the total traffic O-D onto the combined network and the transit O-D, and the auto O-D are determined internally by the assignment model (Lam & Huang 1992, Ferrari 1999, Garcia & Marin 2005, e.g.). This way may be more reasonable than the former way because the modal choice is essentially determined by the perceived travel costs for each mode and those costs are the results of the traffic assignment. Fernandez et al. (1994) gave a classic steady-state assignment model that combines the binary modal choice and passenger flow assignment. They assume that the probability of choose a traffic mode follows a logit model, which states:

$$G^a_{rs}(\mu_{rs}) = \frac{1}{1 + e^{-(\alpha^{ac} + \beta_1(\mu^c_{rs}) - (\mu^a_{rs}))}} \tag{2.7}$$

where the superscript "c" denotes the transit mode, and "a" denotes the car mode. G^a_{rs} is the proportion of auto O-D demands for the $rs-$th O-D pair. Furthermore, they assume that besides the modal choice, travelers can also choose different park-and-ride stations, p, where they parking their car and take the transit line. An additional logit model is introduced, which reads:

$$G^c_{rs,p}(c_{rs}) = \frac{e^{-(\alpha^c_p + \beta_2 \mu^c_{rs,p})}}{\sum_{p'} e^{-(\alpha^c_{p'} + \beta_2 \mu^c_{rs,p'})}} \tag{2.8}$$

Therefore, the number of travelers using the transit between O-D pair is, $q^c_{rs,p} = q_{rs} \cdot G^c_{rs} \cdot G^c_{rs,p}$. This model has been later extended to the case with more general network cases (Ferrari 1998, 1999, Ying & Yang 2005, Hamdouch et al. 2007, e.g.). In a general steady-state combined network, road pricing and transit fare can be added as part of the cost function in each network, which results in a VI problem equivalent to a combined network UE (Ferrari 1999, Hamdouch et al. 2007).

To our best knowledge, multi-modal DTA that uses a stochastic modal choice model is not seen in the literature. Also, carpool, as a single traffic mode along with HOV facilities, has not been modeled in multi-modal traffic assignment problems.

2.3 Traveler Heterogeneity in VOT

In the context of traffic assignment, traveler heterogeneity can be represented by travelers with different value of time (VOT) that is embedded in travel cost/time functions. Travel cost and travel time can be converted with each other by VOT scaling. Some studies used a bi-criteria objective to allow for tradeoff between time and monetary cost of individuals with different VOTs. Travelers with continuously distributed VOTs will only choose several so-called "efficient paths", where no traveler can be better off in both time and cost by unilaterally switching to any other path (Leurent 1993, 1996, Dial 1996, 1997). All the "efficient paths", if ordered by decreasing travel times, have increasing monetary costs. Therefore, each efficient path is used by only those travelers whose VOTs fall within a certain range. Following this approach, Leurent (1993, 1996) considered the heterogenous case with flow-dependent travel times, pre-determined tolls and elastic demand, and showed that the assigned flows are significantly different from those in the homogeneous case, especially when the tolls are set to be high. Dial (1996, 1997) studied the similar problem, but with fixed demand and flow-dependent tolls. He showed that, in a numerical example, the homogeneous model overestimates toll road usage when the toll charge is low and underestimates it when the toll charge is high. This idea was further examined by Mayet & Hansen (2000) and Verhoef & Small (2004) in a one-to-one network with two routes. The former derived analytical results of tolls, user benefits and social benefits. The latter, by deriving second-best tolls, showed that ignoring heterogeneity will underestimate the welfare benefits gaining from second-best pricing, and it further performed a numerical sensitivity analysis to show how the parameters influence the tolls and their efficiency. Furthermore, Dial (1999*a*,*b*) shows that with OD-dependent, continuously distributed VOTs, a tolling problem can be formulated and transformed into an infinite-dimensional VI problem. The solution to this problem reveals that when each traveler uses a path that minimizes his own particular generalized cost, his cost equilibrates to the expected value of the social marginal cost.

Traveler heterogeneity has also been considered in modeling dynamic transportation network problems, mostly concerning morning commute problems. In the case of general networks, the idea of "efficient paths" was extended to time-space networks in dynamic traffic assignment (DTA) (Mahmassani et al. 2005, Lu et al. 2006). However, the "efficient paths" under simultaneous route choice and departure time equilibrium were not obtained precisely as in the static context. In the case of the morning commute problem with a special single-route, single-bottleneck network, Newell (1987) graphically showed, as in the homogeneous case, a stable departure time equilibrium exists if travelers value the early/late arrivals differently, but the resultant queuing pattern is significantly different. Arnott et al. (1988) considered a finite number of traveler groups with non-identical parameters in the morning commute problem, and compared some indicators of network performances (such as total travel cost) under two groups to those in the homogeneous cases. The results show notable differences between homogeneous and heterogeneous cases. Zijpp & Koolstra (2002) provided a numerical algorithm for obtaining the equilibrium solution of the problem discussed by Newell with general cost functions and a finite number of groups of travelers. Recently, Ramadurai et al. (2008) developed a linear complementarity formulation to solve Newell's problem in discrete time with a finite number of groups of travelers. None of these four papers deals with route choices, and nor do they explicitly provide analytical formulae for the travel profiles. In another direction, Lindsey (2004) investigates the existence and uniqueness of departure-time user equilibrium in the morning commute problem with multi-user classes, applying generalized travel cost functions.

2.4 Congestion pricing on parking

There are quite a few descriptive and empirical studies on parking (Thompson et al. 1998, Vianna et al. 2004, e.g.,), but theoretical studies on parking modeling associated with the transportation network are few in the literature. Bifulco (1993) introduced several parking types, fees and average walking times to the static traffic assignment model so as to evaluate the efficacy of several regulatory parking policies in a general urban network. Several others focused on the influences of parking fees to simplified networks. Glazer (1992) analytically derived the social welfare with respect to the parking fees by assuming a constant road-usage fee and constant travel cost for all the travelers. He showed that a lump-sum parking fee may increase the welfare, but a parking fee per unit time may influence travelers who can vary the length of time they park and thus does not help increase the welfare. Verhoef et al. (1995) also assumed a constant travel cost, inclusive of daily parking fee, for all the travelers from each origin node, and conducted diagrammatic analysis on how parking affects the individual travel cost and the modal split. Rather than assuming pre-determined parking demands and constant travel cost for all the travelers, Arnott & Rowse (1999) developed a structural model of parking for a ring-road network. They assumed travelers' choice of parking lot is uniformly distributed on the ring-road, and thereafter derived the expected parking time, driving time and cruising distance for searching available parking spaces. More recently, Anderson & de Palma (2004) studied how drivers cruise and search to find parking spaces, assuming those parking lots are privately owned. Their results are intriguing: "when cruising for parking congests both parkers and through traffic, the benefits from parking pricing are substantially reduced"(p.1). All these studies focus on the daily parking in a steady-state traffic network. Though static models provide a basic idea of traffic congestion and network performance, they do overlook dynamic queuing of traffic flow and time-varying traffic patterns. Also, the availability and accessibility of parking spaces are not explicitly discussed in those models.

Among the few studies looking into the parking in the DTA, Arnott et al. (1991) first studied how a combination of road toll and parking charge affect the morning commute equilibrium pattern. They assumed the parking spaces are continuously distributed along the freeway near the Central Business District (CBD), and the number of parking spaces per unit distance from the CBD is constant. Both the provision and fee structure of parking are centrally planned. Compared to the roadway tolls, parking fees do not eliminate queuing, but can still be fairly efficient. Without parking fees, commuters always occupy parking spots in the increasing order of distance from the destination. An optimal dynamic parking fee scheme can change the order to be decreasing and thus shorten the arrival time window in the CBD area. This thread was followed by Zhang et al. (2008) to derive the dynamic traffic pattern of morning and evening commutes based on road tolls and parking fees.

Chapter 3

Value of Time Heterogeneity in The Morning Commute

This chapter attempts to extend the previous studies in the morning commute problem in two ways: 1) we consider an infinite number of commuter groups by adopting a continuous distribution of parameters in commuters' generalized travel time function; and 2) we take into account both departure time and route choices under heterogeneous cases. We shall first describe our problem set-up and assumptions in Section 2, which also includes a detailed review of Arnott et al. (1988)'s analysis, because their results are central to our study. We first derive the travel profiles and a number of network performance indicators for a route with a single bottleneck. Further, we extend our analysis to a two-route network and discuss its application to infrastructure planning. Finally, we solve a multi-route network with heterogeneous travelers.

3.1 Morning Commute Problem with Continuously Distributed Parameters

In this section we first define the problem set-up and parameters to be used in this chapter, introduce the basic assumptions adopted in our study and review some key results from Arnott et al. (1988), on which our subsequent analysis is based.

As in a typical morning commute problem, a commuter's generalized travel time is composed of his actual travel time (including delay) and weighted early or late arrival time. We first consider a finite number of commuter groups $i = 1, 2, \cdots, n$. For a commuter in group i who departs at time t, his generalized travel time is expressed as,

$$C_i(t) = w(t) + \max\{\beta_i(t^* - t - w(t)), \gamma_i(t + w(t) - t^*)\} \tag{3.1}$$

where $w(t)$ denotes the commuting time, and t^* the desired arrival time (the same work starting time for all travelers). Here β_i and γ_i are group i's early arrival penalty (EAP) and late arrival penalty (LAP), respectively[1]. Similarly, the generalized travel cost of a commuter in group i departing at time t is defined as

$$TC_i(t) = \alpha_i(w(t) + \max\{\beta_i(t^* - t - w(t)), \gamma_i(t + w(t) - t^*)\}) \tag{3.2}$$

where α_i is the VOT of commuters in group i. Here $\alpha_i\beta_i$ and $\alpha_i\gamma_i$ are group i's Values attached to Early Schedule Delay (VESD) and Late Schedule Delay (VLSD). In addition, $w(t)$ reads,

$$w(t) = \int_{t_0}^{t} \frac{r(x)}{s} dx - (t - t_0) \tag{3.3}$$

where s denotes the bottleneck capacity and $r(t)$ denotes the departure rate at time t. t_0 is the beginning time of the departure.

The proportionality assumption 1. *Commuters' preferences for early/late arrival change proportionally, i.e.* $\frac{\gamma_i}{\beta_i} = \eta$ *is constant for any group* i.

This assumption may be reasonable because commuters with large/small VESD ($\alpha\beta$) are more likely to have large/small VLSD ($\alpha\gamma$). Based on the proportionality assumption, and arranging the n groups in an order of increasing β_i ($i = 1, \cdots, n$), Arnott et al. (1988) showed that under UE conditions, a fraction $\eta/(1+\eta)$ of each group arrives early and the remainder $1/(1+\eta)$ of it arrives late. For travelers who arrive early, Group 1 (with lowest β) departs first, then Group 2 and so on, until Group n (with highest β) departs. For travelers who arrive late, the departing order is reversed: Group n departs first, then Group $n-1$ and so on until Group 1 departs. This is shown in Figure 3.1, where travelers in Group i who arrive early depart from $t_{i-1,i}$ to $t_{i,i+1}$ with queuing delay linearly increasing from $w_{i-1,i}$ to $w_{i,i+1}$, and those who arrive late depart from $t_{i+1,i}$ to $t_{i,i-1}$ with queuing delay linearly decreasing from $w_{i+1,i}$ to $w_{i,i-1}$[2]. Travelers

[1] Note that in this chapter β and γ represent the early arrival penalty (EAP) and late arrival penalty (LAP) in the unit of travel time, rather than monetary cost of a unit of schedule delay in the previous literature

[2] $w_{0,1} = w_{1,0} = 0$, $w_{n,n+1}$ and $w_{n+1,n}$ reduce to w_n. $t_{0,1}$ and $t_{n,n+1}$ reduce to t_0 and t_n accordingly, and similarly, $t_{1,0}$ and $t_{n+1,n}$ reduce to t_1 and t_n.

departing from t_0 to t_n will arrive early and those departing from t_n to t_1 will arrive late. Queue persists during the rush period given by $[t_0, t_1]$. Let N represent the total travel demand. We define $\delta = \eta/(1+\eta)^3$, $\phi = 1/(1+\eta)$. It was shown by Arnott et al. (1988) that $t_0 = t^* - \delta N/s$ and $t_1 = t^* + \phi N/s$, and both times remain the same whether the traveling population is homogeneous or not, because the total number of travelers who arrive early and that for those who arrive late do not change.

$w(t)$ is therefore a piece-wise linear function:

$$\frac{dw(t)}{dt} = \frac{\beta_i}{1-\beta_i} \text{ for } t_{i-1,i} \leq t \leq t_{i,i+1} \text{ early arrival, or,} \tag{3.4a}$$

$$\frac{dw(t)}{dt} = -\frac{\gamma_i}{1+\gamma_i} \text{ for } t_{i+1,i} \leq t \leq t_{i,i-1} \text{ late arrival} \tag{3.4b}$$

This result is fairly intuitive because under UE, a lower β (γ) means less weight on the time loss due to early (late) arrival, hence travelers with lower β would prefer arriving early (late) than waiting longer in the queue.

It is worth mentioning that the derivation of travel profiles under UE only relies on generalized travel times which are independent of VOT α. Since our following analysis only concerns UE, we consistently use generalized travel time, as expressed in Equation 3.1, rather than generalized travel cost, throughout the chapter.

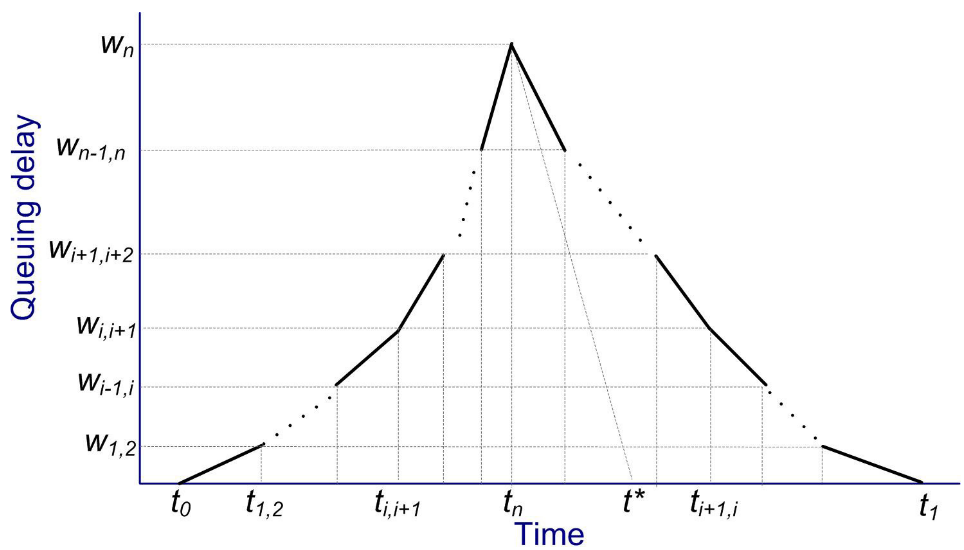


Figure 3.1. Queuing delay with respect to the departure time

[3] In the previous literature, δ is defined to be $\beta\gamma/(\gamma+\beta)$, which is not a constant any more in the context of continuously distributed EAP β. Here, we define $\delta = \gamma/(\gamma+\beta) = \eta/(1+\eta)$

3.2 The Case of A Single Route

3.2.1 Travel profiles

We first extend the aforementioned results from finite number of commuter groups to infinite number of groups by considering a continuously distributed β and γ.

Let f be the probability density function (PDF) of β with a lower limit a and an upper limit b ($B = b - a$). For the case with finite number of groups, groups are ordered increasingly by β_i, and the number of travelers in group i is approximated by $f(\beta_i)\cdot\frac{B}{n}\cdot N, i = 1, \cdots, n$. We divide the commuters into n groups, with n arbitrarily large. Subsequently, Equation 3.4 can be used to derive the waiting times for each group i. By letting the number of groups become infinitely large, we derive the desired waiting time, and consequently the departure profile, for a continuously distributed heterogeneous traveling population.

Differentiation with respect to t for both sides in Equation 3.3 yields:

$$dw(t)/dt = r(t)/s - 1 \tag{3.5}$$

Substitute $dw(t)/dt$ in Equation 3.4, we have:

$$r(t) = \frac{1}{1-\beta_i}s \text{ for } t_{i-1,i} \leq t \leq t_{i,i+1} \text{ for early arrival} \tag{3.6}$$

Because a fraction δ of each group arrives early (Arnott et al. 1988), we have,

$$\delta f(\beta_i)\frac{B}{n}N = (t_{i,i+1} - t_{i-1,i})r(t) = (t_{i,i+1} - t_{i-1,i})\frac{1}{1-\beta_i}s \tag{3.7}$$

Consequently, we get $t_{i,i+1}$ by solving Equation 3.7 recursively

$$t_{i,i+1} = t^* - N\frac{\delta}{s} + N\frac{\delta}{s}\cdot\frac{B}{n}\sum_{k=1}^{i}[(1-\beta_k)f(\beta_k)] \tag{3.8}$$

Similarly, for late arrival, $t_{i+1,i}$ reads,

$$t_{i+1,i} = t^* + N\frac{\phi}{s} - N\frac{\phi}{s}\cdot\frac{B}{n}\sum_{k=1}^{i}[(1+\eta\beta_k)f(\beta_k)] \tag{3.9}$$

The punctual departure time t_n,

$$t_n = t^* - N\frac{\delta}{s}\cdot\frac{B}{n}\sum_{k=1}^{n}[\beta_k f(\beta_k)] \tag{3.10}$$

Because

$$(t_{i,i+1} - t_{i-1,i}) \cdot dw(t)/dt = w(t_{i,i+1}) - w(t_{i-1,i}) \text{ for } t_{i-1,i} \leq t \leq t_{i,i+1} \tag{3.11}$$

we have:

$$w(t_{i,i+1}) = N\frac{\delta}{s} \cdot \frac{B}{n} \sum_{k=1}^{i} \beta_k f(\beta_k) \tag{3.12}$$

$$w(t_n) = N\frac{\delta}{s} \cdot \frac{B}{n} \sum_{k=1}^{n} \beta_k f(\beta_k) \tag{3.13}$$

Therefore, generalizing Equation 3.8, 3.9 and 3.12 if $n \longrightarrow \infty$, i.e. infinite number of traveler groups, yields that a traveler with EAP $\beta (a \leq \beta \leq b)$ will depart at time

$$t(\beta) = t^* - N\frac{\delta}{s} + N\frac{\delta}{s} \int_a^\beta (1-x) f(x) dx \text{ early arrival, or,} \tag{3.14a}$$

$$t(\beta) = t^* + N\frac{\phi}{s} - N\frac{\phi}{s} \int_a^\beta (1+\eta x) f(x) dx \text{ late arrival} \tag{3.14b}$$

and experience a queuing delay

$$w(\beta) = N\frac{\delta}{s} \int_a^\beta x f(x) dx \tag{3.15}$$

Let $G_1(\beta) = \int_a^\beta (1-x) f(x) dx$, $G_2(\beta) = \int_a^\beta (1+\eta x) f(x) dx$ and $w(t)$ denotes the queuing delay with respect to the departure time t similarly as the curve shown in Figure 3.1, however, with β continuously distributed. We know that

$$dw\left(t_0 + N\frac{\delta}{s} G_1(\beta)\right)/dt = \frac{\beta}{1-\beta} \text{ for early arrival, or,} \tag{3.16a}$$

$$dw\left(t_1 - N\frac{\phi}{s} G_2(\beta)\right)/dt = -\frac{\eta\beta}{1+\eta\beta} \text{ for late arrival} \tag{3.16b}$$

Therefore, we can obtain the analytical formula of $w(t)$ by solving the ordinary differential equation:

$$dw(t)/dt = \frac{G_1^{-1}(\frac{(t-t_0)s}{\delta N})}{1 - G_1^{-1}(\frac{(t-t_0)s}{\delta N})} \text{ early arrival, or,} \tag{3.17a}$$

$$dw(t)/dt = -\frac{\eta G_2^{-1}(\frac{(t_1-t)s}{\phi N})}{1 + \eta G_2^{-1}(\frac{(t_1-t)s}{\phi N})} \text{ late arrival} \tag{3.17b}$$

$$\text{where} \quad w(t_0) = 0, w(t_1) = 0$$

We further define $\beta_0 = \int_a^b x f(x) dx$, i.e. the expected value of β, for the following sections.

Proposition 3.1. *If γ/β is assumed to be constant for all the commuters, then under UE condition, only the travelers departing at $t(b)$, t_1 and t_0 (i.e. the traveler with the highest β, b, and the one with the lowest β, a) will experience the same queuing delay as in the homogeneous case where $\beta = \beta_0$ for all the travelers. The queuing delay with respect to any other departure time t is overestimated by the assumption of homogeneity, and thus, the total generalized travel time (TTT) is overestimated as well.*

Proof. Here, we only present the proof for early arrival. A similar proof for late arrival is omitted here to save space.

For the traveler with $\beta = a$, the queuing delay $w(a) = 0$ and the departure times $t(a) = t_0$ and t_1 are the same in both homogeneous and heterogeneous cases. For the traveler with $\beta = b$, $w(b) = N\frac{\delta}{s}\int_a^b xf(x)dx = N\frac{\delta}{s}\beta_0$ and $t(b) = t^* - N\frac{\delta}{s} + N\frac{\delta}{s}\int_a^b(1 - x)f(x)dx = t^* - N\frac{\delta}{s}\int_a^b xf(x)dx = t^* - N\frac{\delta}{s}\beta_0$, the queuing delay and departure time are the same as in the homogeneous case. For any other traveler with $a < \beta < b$ in early arrival, the queuing delay a traveler would experience in the homogeneous case is $(t(\beta) - t_0)\frac{\beta_0}{1-\beta_0}$. Therefore, the second part of the proposition is equivalent to show that $(t(\beta) - t_0)\frac{\beta_0}{1-\beta_0} > w(\beta), \forall a < \beta < b$. Let $g(\beta) = (t(\beta) - t_0)\frac{\beta_0}{1-\beta_0} - w(\beta)$.

$$\begin{aligned} g'(\beta) &= d\{\frac{\beta_0}{1-\beta_0}\int_a^\beta (1-x)f(x)dx - \int_a^\beta xf(x)dx\}/d\beta \\ &= \frac{\beta_0}{1-\beta_0}(1-\beta)f(\beta) - \beta(f(\beta)) \\ &= \frac{(\beta_0 - \beta)f(\beta)}{1-\beta_0} \end{aligned}$$

$g'(\beta) > 0$ when $a < \beta < \beta_0$, $g'(\beta) = 0$ when $\beta = \beta_0$, and $g'(\beta) < 0$ when $b > \beta > \beta_0$. In addition, $g(a) = g(b) = 0$. Therefore, $g(\beta) > 0, \forall a < \beta < b$. Furthermore, when $\beta = \beta_0$, $g(\beta)$ reaches its maximum, i.e., assuming homogeneity of β across the population leads to the maximum overestimation of queuing delay for a traveler departing at $t(\beta_0)$. □

Proposition 3.2. *If γ/β is assumed to be constant for all the commuters, then under UE condition, every commuter is better off if the bottleneck capacity is enlarged. Commuters with high values of β benefit more than those with low values.*

Proof. The generalized travel time of the traveler with $\beta(a \le \beta \le b)$ becomes,

$$C(\beta) = \beta(t^* - t(\beta)) + w(\beta)(1 - \beta) \tag{3.18}$$

Substitute $t(\beta)$ and $w(\beta)$ by Equation 3.14a and 3.15, we obtain,

$$\begin{aligned} C(\beta) &= \beta(N\frac{\delta}{s} - N\frac{\delta}{s}\int_a^\beta (1-x)f(x)dx) + (1-\beta)N\frac{\delta}{s}\int_a^\beta xf(x)dx \\ &= N\frac{\delta}{s}[\beta - \int_a^\beta (\beta - x)f(x)dx] \end{aligned} \tag{3.19}$$

Therefore,

$$\begin{aligned} \frac{\partial C(\beta)}{\partial s} &= -\frac{N\delta}{s^2}[\beta - \int_a^\beta (\beta - x)f(x)dx] < 0 \\ \frac{\partial \frac{\partial C(\beta)}{\partial s}}{\partial \beta} &= -\frac{N\delta}{s^2}(1 - \int_a^\beta f(x)dx) < 0 \end{aligned}$$

which completes the proof. □

3.2.2 System performance

In this section we derive the formulae for computing total generalized travel time. Then we provide numerical examples to illustrate the magnitude of performance difference between that of a homogeneous population and that of a heterogeneous population. Let the index e and o represent the heterogenous case and the homogenous case, respectively.

The total generalized travel time (TTT) of all the heterogeneous travelers reads:

$$\begin{aligned} TTT_e &= N\int_a^b f(\beta)w(\beta)d\beta \\ &= N^2\frac{\delta}{s}\int_a^b f(\beta)[\int_a^\beta xf(x)dx]d\beta \end{aligned} \tag{3.20}$$

while TTT under homogeneous travelers is $TTT_o = \delta\beta_0 N^2/2s$.

Next we provide some numerical examples to illustrate the differences between these network performance indicators under a homogeneous and three types of heterogeneous populations. We consider two possible symmetric PDFs of β with expected value of β, $\beta_0 = 0.5$ (the limits of β are set to be $a = 0.1$ and $b = 0.9$). One is a truncated normal distribution with mean 0.5 and standard deviation 0.12, and the other is a piecewise linear distribution, with the PDF $f(\beta) = (\beta - 0.1)\cdot 25/4$ if $0.1 \leq \beta \leq 0.5$ and $f(\beta) = (0.9 - \beta)\cdot 25/4$ if $0.5 \leq \beta \leq 0.9$. More realistically, the PDF of β may be skewed to the left, as indicated by other empirical studies (Tseng & Verhoef 2008). Hence, we also consider a log-normal distribution of β, and the mean and standard deviation of $\ln \beta$ are -1 and 0.3, respectively ($\beta_0 = 0.3834$). Other parameters are: $\eta = 4$, $N = 10000$ persons and $s = 90$ persons/min(approximately a three-lane freeway).

We plot in Figure 3.2 the changes in queuing delay with respect to the departure time based on aforementioned three PDFs and an identical $\beta (= \beta_0)$. It can be seen that the queuing delay is overestimated by the assumption of homogeneity, and thus TTT is overestimated as well.

We find that compared with a normally distributed β, the homogeneity overestimates TTT by nearly 15%. The maximum queuing delay under log-normal distribution does not coincide with other distributions, because its β_0 is less than others. In this numerical example, "linear distribution" of β may lead to underestimate the total queuing delay, by 5% compared to the normal distribution. This is because βs are more centralized in the normal distribution than "linear distribution", and the travel profiles derived from normal distributions are more likely to be estimated towards the homogeneous case.

3.3 The Case of A Two-route Network

3.3.1 Travel profiles

In this section, we extend the results derived in Section 2 to the case with route choice. We let a corridor freeway (capacity s_f) and arterial roads (AR) (capacity s_a), connecting an origin-destination pair, are open to all the travelers at any time, and travelers make choices between the freeway and the AR to minimize their own travel time/cost. The free flow travel time of the freeway is zero without loss of generality, while that of the AR is set to be τ. N_f and N_a denote the number of travelers on the freeway and the AR, respectively. $N_f + N_a = N$ is assumed to be fixed. f_{f0} and f_{a0} are the normalized PDF of β for freeway and AR commuters, respectively. Therefore, $N_f \cdot f_{f0}(x) + N_a \cdot f_{a0}(x) = N \cdot f(x), \forall x \in [a, b]$. Let f_f and f_a be the original distribution of β for freeway and AR commuters, respectively, such that $f_f(x) + f_a(x) = f(x), \forall x \in [a, b]$. Then we have the following results.

Proposition 3.3. *If there are two routes connecting a one-to-one network and the free*

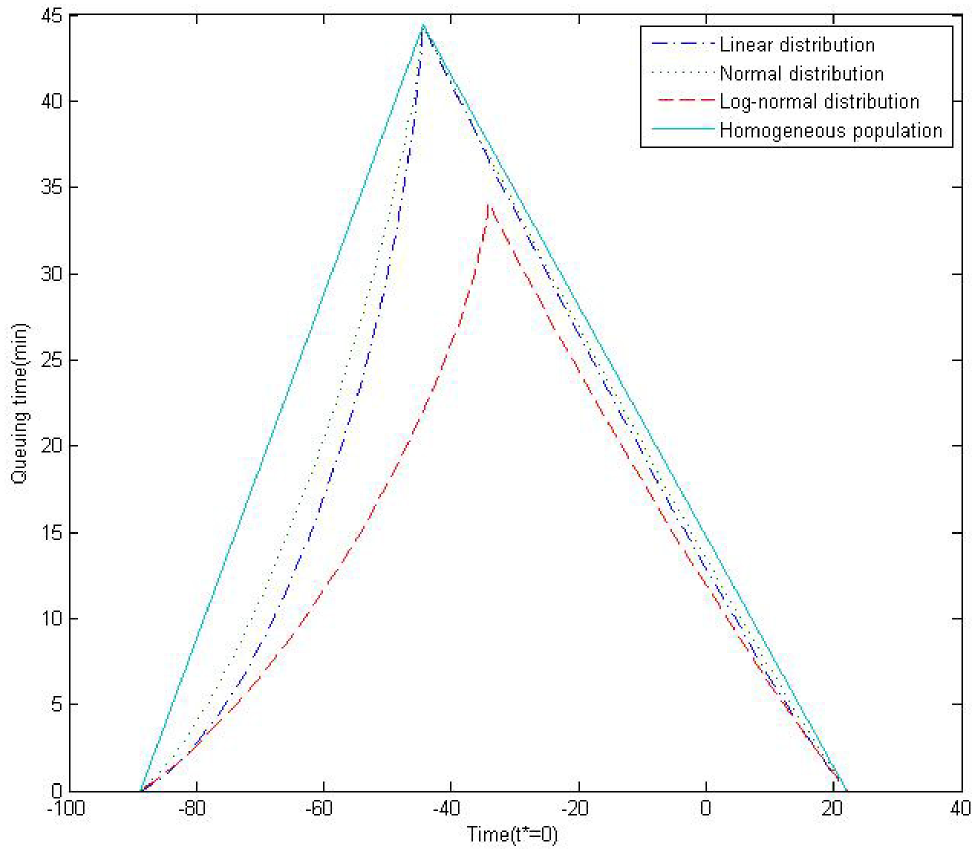


Figure 3.2. Queuing delay with respect to the departure time($N = 10000$)

flow travel time on the AR is larger than that on the freeway, then the travelers with highest β, b, will first shift to the AR, prompted by demand increase. The critical travel demand N^* when travelers start to use the AR remains the same as in the homogeneous case.

Proof. From Equation 3.18, we know that:

$$\frac{dC(\beta)}{d\beta} = N\frac{\delta}{s_f}\left(1 - \int_a^\beta f(x)dx\right) > 0 \tag{3.21}$$

when $a \leq \beta < b$. Therefore, $C(\beta)$ in monotonically increasing with respect to β. Hence, the traveler with the highest β, b, will first shift to the AR when the total demand is progressively increased. When $\beta = b$, $C(b) = N_f\frac{\delta}{s_f}(b - \int_a^b(b-x)f(x)dx) = N_f\frac{\delta}{s_f}\beta_0$. Consequently, $N^* = \tau s_f/(\delta\beta_0)$, which is consistent with N^{***} derived by Arnott et al. (1990). □

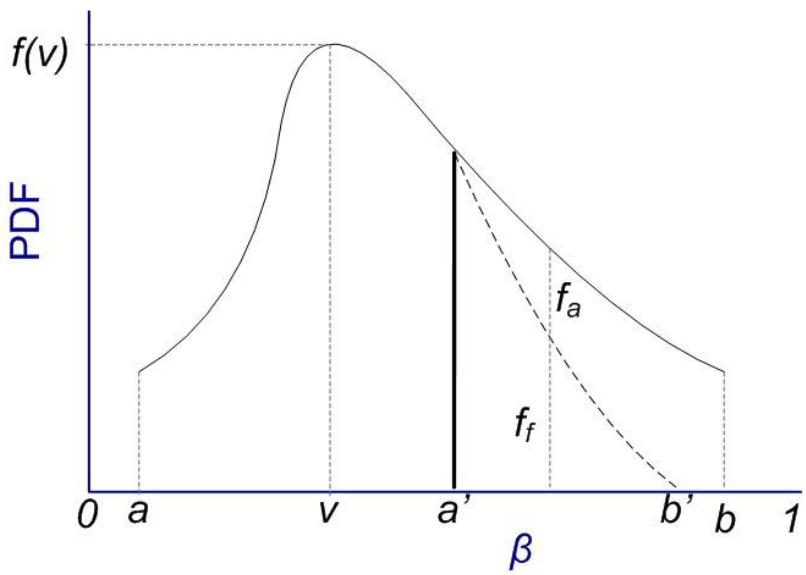


Figure 3.3. The β distribution splits at a'

The first part of the conclusion is somewhat expected because intuitively, travelers who weigh the penalty of schedule delay more importantly are more likely to choose the AR which takes longer time to reach the destination than the freeway but they suffer less schedule delay as the AR is less congested than the freeway. When $N > N^*$ where $N^* = \tau s_f/(\delta\beta_0)$, some travelers with high values of β will start to use the AR.

Proposition 3.4. *There exists a value of β, a', such that a' is the highest value of β of all the travelers on the freeway and is the lowest value of β of all the travelers on the AR. In other words, the β distribution splits, at a', into two parts, as shown in Figure 3.3, such that all the travelers with $\beta \leq a'$ choose the freeway and those with $\beta > a'$ choose the AR.*

Proof. The proof is by contradiction.

Proposition 3.4 indicates that with the increase of total demand, travelers with high values of β shift to the AR while travelers with low values of β stay in the freeway. If Proposition 4 does not hold, the β distribution splits within an interval, not a single point. We show that by contradiction there does not exist such a transition interval of β within which some travelers take the highway and others take the AR.

We hence assume b' is the highest value of β of all the travelers on the freeway and a' is the lowest value of β of all the travelers on the AR, $b' > a'$, as shown in Figure

3.3. The travelers with $a' \leq \beta \leq b'$ can take either the freeway or the AR. Therefore, $N_f \cdot f_{f0}(x) = N \cdot f_f(x), \forall a \leq x \leq b'$. According to Equation 3.18, the travel time of the travelers on the freeway with $a' \leq \beta \leq b'$:

$$
\begin{aligned}
C_f(\beta) &= N_f \frac{\delta}{s_f}[\beta - \int_a^\beta (\beta - x) f_{f0}(x)dx] \\
&= N_f \frac{\delta}{s_f}\beta - N\frac{\delta}{s_f}\int_a^\beta (\beta - x) f_f(x)dx \\
&= N_f \frac{\delta}{s_f}\beta - N\frac{\delta}{s_f}\int_a^{a'} (\beta - x) f(x)dx - N\frac{\delta}{s_f}\int_{a'}^\beta (\beta - x) f_f(x)dx
\end{aligned}
$$

The travel time of the travelers on the AR with $a' \leq \beta \leq b'$:

$$
\begin{aligned}
C_a(\beta) &= \tau + N_a \frac{\delta}{s_a}[\beta - \int_{a'}^\beta (\beta - x) f_{a0}(x)dx] \\
&= \tau + N_a \frac{\delta}{s_a}\beta - N\frac{\delta}{s_a}\int_{a'}^\beta (\beta - x) f_a(x)dx
\end{aligned}
$$

Because under User Equilibrium, $C_f(\beta) = C_a(\beta)$ for $a' \leq \beta \leq b'$, $dC_f(\beta)/d\beta = dC_a(\beta)/d\beta$ for $a' \leq \beta \leq b'$. Now we have:

$$
\begin{aligned}
dC_f(\beta)/d\beta - dC_a(\beta)/d\beta &= N_f \frac{\delta}{s_f} - N_a \frac{\delta}{s_a} - N\frac{\delta}{s_f}\int_a^{a'} f(x)dx \\
&\quad - N\frac{\delta}{s_f}\int_{a'}^\beta f_f(x)dx + N\frac{\delta}{s_a}\int_{a'}^\beta f_a(x)dx = 0
\end{aligned}
$$

When $\beta = a' + d\beta$ where $d\beta$ is an infinitesimal, the last two items becomes infinitesimal as well:

$$
N_f \frac{\delta}{s_f} - N_a \frac{\delta}{s_a} - N\frac{\delta}{s_f}\int_a^{a'} f(x)dx + o(\beta) = 0 \tag{3.22}
$$

On the other hand, $C_f(a') = C_a(a')$ yields

$$
N_f \frac{\delta}{s_f}a' - N\frac{\delta}{s_f}\int_a^{a'} (a' - x) f(x)dx = \tau + N_a \frac{\delta}{s_a}a' \tag{3.23}
$$

Compare Equation 3.23 and 3.22, we obtain that:

$$
\tau = N\frac{\delta}{s_f}\int_a^{a'} x f(x)dx + o(\beta) \tag{3.24}
$$

Because

$$
\begin{aligned}
C_f(a' + d\beta) &= N_f \frac{\delta}{s_f}(a' + d\beta) - N\frac{\delta}{s_f}\int_a^{a'+d\beta} (a' + d\beta - x) f(x)dx \\
&= N_f \frac{\delta}{s_f}a' - N\frac{\delta}{s_f}a'\int_a^{a'} f(x)dx + N\frac{\delta}{s_f}\int_a^{a'} x f(x)dx + o(\beta) + o^2(\beta) \\
&= \tau + o(\beta) + o^2(\beta)
\end{aligned}
$$

and

$$C_a(a' + d\beta) = \tau + N_a \frac{\delta}{s_a} a' + o(\beta)$$

$C_a(a' + d\beta) > C_f(a' + d\beta)$, which contradicts our assumption, $C_f(\beta) = C_a(\beta)$ for $a' \leq \beta \leq b'$. Therefore, the β distribution splits at a single point a' such that all the travelers with $\beta \leq a'$ choose the freeway and those with $\beta > a'$ choose the AR.

□

Based on this proposition, we can obtain a' by solving the following equation which rewrites Equation 3.23 by substituting $N_f = N \int_a^{a'} f(x)dx$ and $N_a = N \int_{a'}^{b} f(x)dx$.

$$N \frac{\delta}{s_f} \int_a^{a'} x f(x)dx = \tau + N \frac{\delta}{s_a} a' \int_{a'}^{b} f(x)dx \tag{3.25}$$

Proposition 3.5. *(Existence and uniqueness of UE) There is at least one solution of a' in Equation 3.25. User Equilibrium (UE) solution is unique, i.e. there is only one solution of a' in Equation 3.25, if, 1) the PDF of β is differentiable, and, 2) there exists v ($a < v < b$) such that $f'(a') \geq 0$ for $a' < v$ and $f'(a') < 0$ for $a' > v$, 3) $\int_v^b f(x)dx \leq v f(v)$*

Proof. Let $\rho(y) = N \frac{\delta}{s_f} \int_a^y x f(x)dx - \tau - N \frac{\delta}{s_a} y \int_y^b f(x)dx$. Then, $\rho(a) = -\tau - N \frac{\delta}{s_a} a < 0$, and $\rho(b) = N \frac{\delta}{s_f} \beta_0 - \tau > 0$. The existence of UE can be obtained immediately due to the continuity of $\rho(a')$. Furthermore,

$$\begin{aligned} \frac{d\rho(y)}{dy} &= N \frac{\delta}{s_f} y f(y) - N \frac{\delta}{s_a} \left(\int_y^b f(x)dx - y f(y) \right) \\ \frac{d^2\rho(y)}{dy^2} &= N\delta \left(\frac{1}{s_f} + \frac{1}{s_a} \right) (f(y) + y f'(y)) + N \frac{\delta}{s_a} f(y) \end{aligned}$$

Because $\frac{d^2\rho(y)}{dy^2} > 0$ where $y < v$ due to condition 2), ρ is convex where $y < v$. $\frac{d\rho(y)}{dy} > 0$ where $y > v$ due to condition 3), so ρ is strictly monotonically increasing where $y > v$. Consequently, there is only one solution of y, which is a', and $a < a' < v$ if $\rho(v) > 0$, $v \leq a' < b$ if $\rho(v) \leq 0$. □

Actually, the three conditions required for the uniqueness in this chapter could be satisfied by many common distributions. We can first easily see that all symmetric distributions satisfy them. For the log-normal distribution, they also hold when $v f(v) =$

$\frac{1}{\sqrt{2\pi}\sigma}c^{-\sigma^2/2} \geq 1 > \int_v^b f(x)dx$. The uniqueness of UE under more general distributions, however, is not guaranteed and there would be worthy of further investigation.

Assuming the solution of a' is unique, we obtain that $d\rho(a')/dy > 0$, because $d\rho(a')/dy \leq 0$ and continuity of β yield another solution of a', and thus contradicts the uniqueness of a'. Therefore,

$$\frac{d\rho(a')}{dy} = (\frac{1}{s_f} + \frac{1}{s_a})a'f(a') - \frac{1}{s_a}\int_{a'}^{b} f(x)dx > 0 \tag{3.26}$$

3.3.2 System performance

Once a' is obtained from Equation 3.25, the β distributions on the freeway and AR are known and we can calculate the TTT by rewriting Equation 3.20 for both the freeway and AR.

The number of travelers on the freeway and on the AR under heterogeneity reads:

$$N_{f,e} = N\int_{a}^{a'} f(x)dx, N_{a,e} = N\int_{a'}^{b} f(x)dx \tag{3.27}$$

and the normalized PDF of β for freeway and AR commuters become,

$$f_{f0}(x) = f(x)\frac{N}{N_{f,e}}, a \leq x \leq a' \tag{3.28a}$$

$$f_{a0}(x) = f(x)\frac{N}{N_{a,e}}, a' \leq x \leq b \tag{3.28b}$$

While the number of travelers on the freeway and on the AR under homogeneity is (Arnott et al. 1990):

$$N_{f,o} = \frac{s_f}{\delta\beta}\tau + (N - \frac{s_f}{\delta\beta}\tau)\frac{s_f}{s_f + s_a} \tag{3.29a}$$

$$N_{a,o} = (N - \frac{s_f}{\delta\beta}\tau)\frac{s_a}{s_f + s_a} \tag{3.29b}$$

The total generalized travel time (TTT) of all the commuters under heterogeneity reads:

$$\begin{aligned} TTT_e &= N^2\frac{\delta}{s_f}\int_{a}^{a'} f(\beta)[\int_{a}^{\beta} xf(x)dx]d\beta + N^2\frac{\delta}{s_a}\int_{a'}^{b} f(\beta)[\int_{a'}^{\beta} xf(x)dx] \\ &+\tau N\int_{a'}^{b} f(x)dx \end{aligned} \tag{3.30}$$

while TTT under homogeneous commuters is :

$$TTT_o = \frac{\delta\beta_0 N^2}{2(s_a + s_f)} - \frac{s_a s_f \tau^2}{2(s_a + s_f)\delta\beta_0} + \frac{s_a}{s_f + s_a} N\tau \tag{3.31}$$

Next we provide some numerical examples to illustrate the differences in TTT between homogeneous and heterogeneous traveling populations. In this section, the AR is not just one route but the aggregate of a collection of parallel arterials, so we assume that s_a = 180veh/min (approximately six lanes in total) $\geq s_f$ = 90 veh/min (approximately a three-lane freeway). The free-flow travel time on the AR is set to be τ = 15min. N changes from 10000 to 18000, where the AR will be used. A log-normal distribution, with the mean -1 and standard deviation 0.3 of $\ln\beta$, is assumed to be the PDF of β.

We first plot in Figure 3.4 the number of travelers on the freeway and AR in both homogeneous and heterogeneous cases, with respect to the total demand N. Generally, considering heterogeneity leads to more travelers on the freeway and less travelers on the AR, compared to not considering heterogeneity. When $N = 17600$, the number of travelers on the freeway is the same as on the AR based on the assumption of homogeneity, since $s_a > s_f$ and the AR serves s_a/s_f times as many travelers as the freeway for the demand beyond N^*. However, for the heterogeneous case, the demand share of the freeway, though slowly decreasing, is still 40% more than the demand share of the AR. This is because the freeway will be used by travelers with low β who weigh the queuing penalty more than early or late arrival penalty when the network is highly congested. Those travelers are more likely to depart early or late to avoid long queuing time. Consequently, the departure times on the freeway will be more spread out, and thus the freeway will serve more travelers under heterogeneity.

Furthermore, we compare the TTT of the total network in both homogeneous and heterogeneous cases with respect to the total demand N, as shown in Figure 3.5. Similarly as in the case of a single-route network, homogeneity overestimates the TTT by approximately 12%.

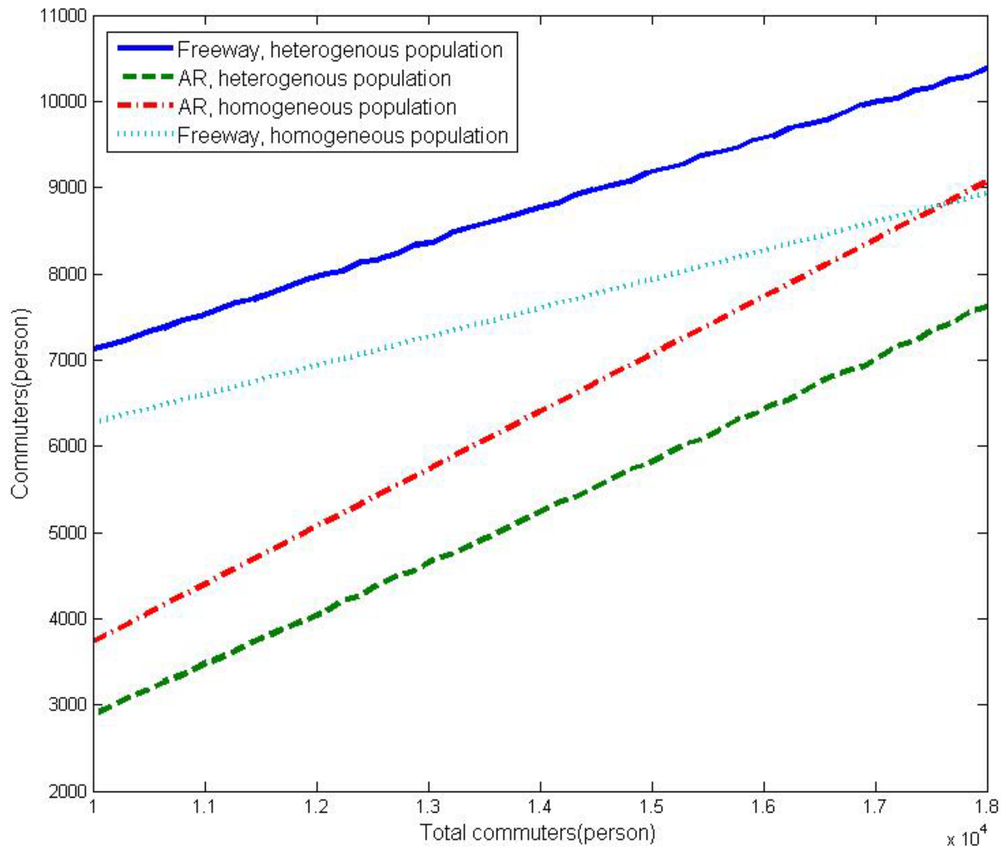


Figure 3.4. Comparisons of the number of travelers on the freeway and AR in both homogeneous and heterogeneous cases

3.4 Application to Infrastructure Planning

It is of particular interest to study, under heterogeneity, how facility improvement will impact the system total travel time (TTT) and demand share of the freeway route, and who would benefit most/least from such improvements. To do this, we examine the marginal system time savings with respect to freeway capacity expansion ($dTTT/ds_f$), arterial road capacity expansion ($dTTT/ds_a$), and arterial free-flow travel time ($dTTT/d\tau$). Furthermore, we define the following quantities for later use: TTT_f and TTT_a are the total travel times on the freeway and arterial road, respectively, and $b_f = \int_a^{a'} f(x)dx$ is the share of the demand from the freeway.

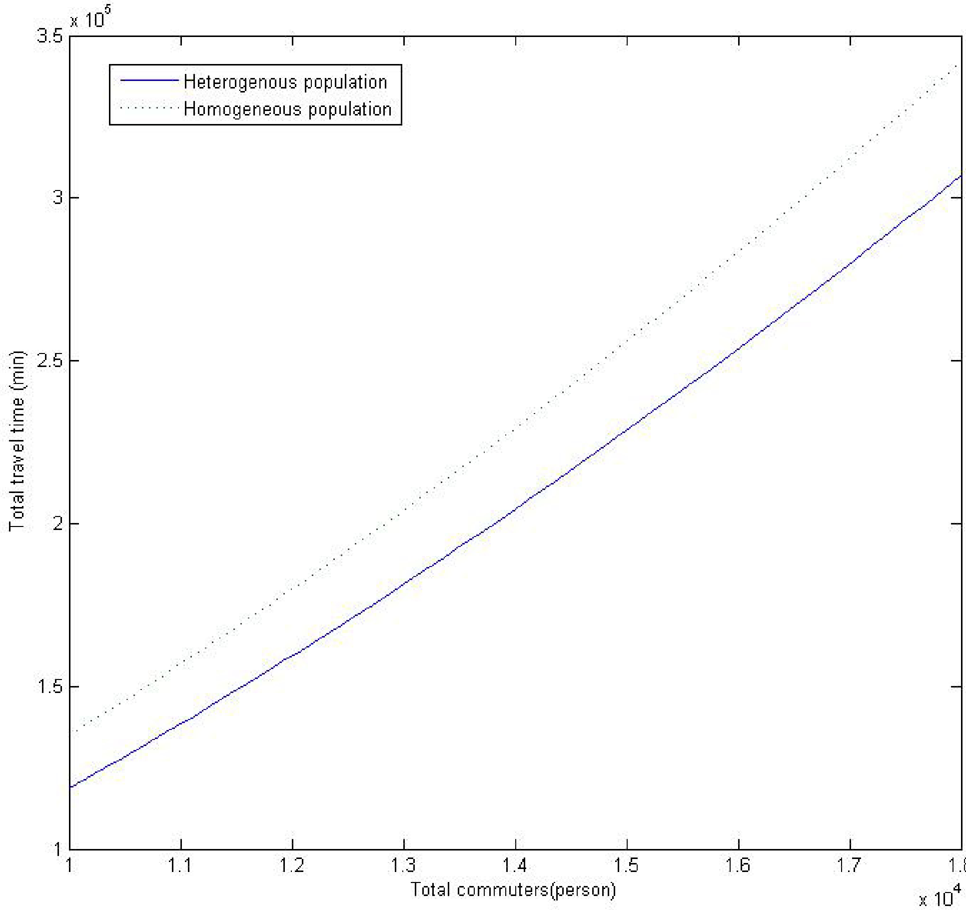


Figure 3.5. Comparisons of TTT in both homogeneous and heterogeneous cases with respect to N

3.4.1 Expanding freeway capacity

We first investigated the impact on the total travel time caused by freeway capacity enlargement. The results are shown in Table 3.1 (please see Appendix B for the derivations).

Table 3.1. Marginal system time savings with respect to freeway capacity enlargement

	$\frac{dTTT_f}{ds_f}$	$\frac{dTTT_a}{ds_f}$	$\frac{dTTT}{ds_f}$	$\frac{db_f}{ds_f}$
Sign	$\pm$	< 0	< 0	> 0

The results indicate that improving freeway capacity will always reduce the TTT of the whole network and TTT_a, and increase the demand share of the freeway. This conclusion under heterogeneity is consistent with that under homogeneity (see Appendix A for those derivatives under homogeneity). Under homogeneity, TTT_f increases with the increase of freeway capacity if and only if $s_a > s_f$. However, under heterogeneity, whether TTT_f will be reduced or not with the increase of freeway capacity is dependent on all the given parameters and the PDF of β, which may lead to quite different conclusions as compared to the homogeneous case.

Now we use an example to illustrate those points. Let $N = 10000$ veh, $\tau = 15$ min,the AR capacity $s_a = 90$ veh/min and the freeway capacity s_f changes from 40 veh/min to 140 veh/min (we still adopt the aforementioned log-normal PDF). The changes of $\frac{dTTT_f}{ds_f}$ and $\frac{dTTT}{ds_f}$, under both homogeneity and heterogeneity, are shown in Figure 3.6. In this numerical example, the assumption of homogeneity will overestimate the impact on TTT caused by the freeway capacity enlargement. Under heterogeneity, TTT_f achieves the minimum value of 101135 veh*min, i.e., $\frac{dTTT_f}{ds_f} = 0$, when $s_f = 72$ veh/min , while the minimum value of TTT_f is 107271 veh*min when $s_f = 90$ veh/min under homogeneity. Therefore, the optimal freeway capacity and the desired total travel time under heterogeneity could be considerably different from the homogeneous case.

3.4.2 Improving arterial road

Now we turn to investigate the impact of AR improvement in the manner of capacity enlargement and free-flow travel time reduction. The results are shown in Table 3.2 (please see Appendix B for the derivations).

Table 3.2. Marginal system time savings with respect to AR improvement

	$\frac{dTTT_f}{ds_a}$	$\frac{dTTT_a}{ds_a}$	$\frac{dTTT}{ds_a}$	$\frac{db_f}{ds_a}$
Sign	< 0	$\pm$	< 0	< 0

	$\frac{dTTT_f}{d\tau}$	$\frac{dTTT_a}{d\tau}$	$\frac{dTTT}{d\tau}$	$\frac{db_f}{d\tau}$
Sign	> 0	$\pm$	> 0	> 0

The results in Table 3.2 indicate that improving AR capacity, or reducing the free-flow travel time on the AR, will always reduce the TTT of the total network and TTT_f, as well as the demand share of the freeway. This is consistent with the conclusion under homogeneity. However, whether s_a and τ have a positive effect on TTT_a depends on all the given parameters and the PDF of β. In fact, according to our numerical examples, we can find some PDFs such that $dTTT_a/ds_a > 0$ and $dTTT_a/d\tau > 0$ for arbitrary values

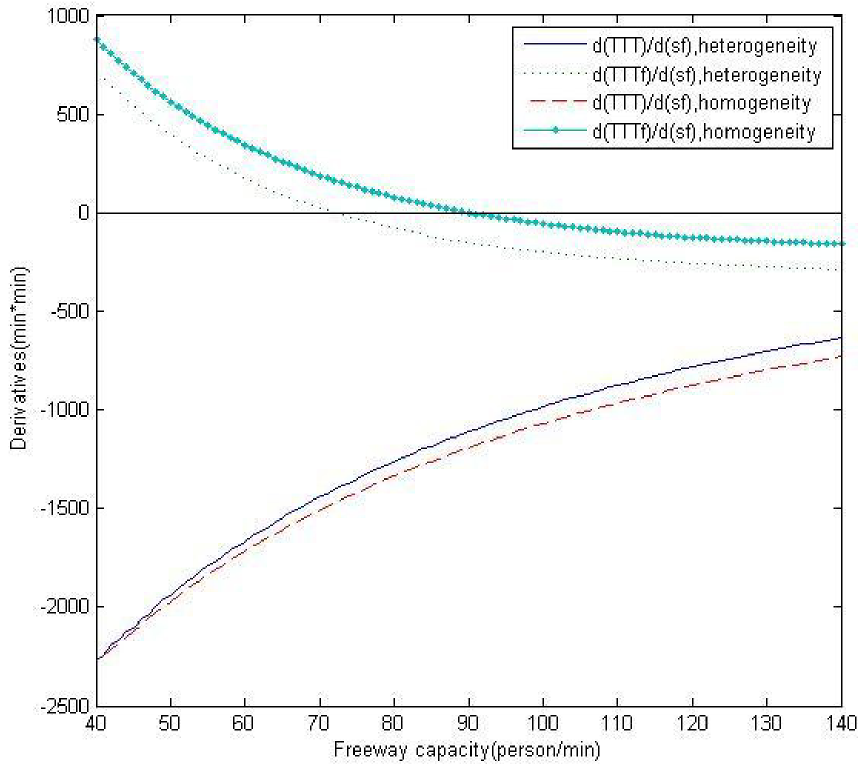


Figure 3.6. Changes of $\frac{dTTT_f}{ds_f}$ and $\frac{dTTT}{ds_f}$ with respect to s_f

of N, s_a, s_f and τ, which implies that sometimes the EAP/LAP (i.e. β or γ) distribution of the population has the dominant effect on changes in total arterial travel time when an improvement is made to the arterial road.

3.4.3 Is the capacity enlargement Pareto-improving?

Since improving roadway capacity reduces total system travel time, we are particularly concerned about whether it is a Pareto improvement, i.e. whether every commuter is better off in regard to capacity enlargement. Let $C_f(\beta)$ and $C_a(\beta)$ denote the generalized travel time of a commuter with EAP β on the freeway and AR, respectively. $\frac{\partial C_f(\beta)}{\partial s_f}$ and $\frac{\partial C_a(\beta)}{\partial s_f}$ represent the changes in generalized travel time with respect to the freeway capacity improvement for a commuter (with EAP β) on the freeway and the AR, respectively. Similarly, $\frac{\partial C_f(\beta)}{\partial s_a}$ and $\frac{\partial C_a(\beta)}{\partial s_a}$ represent the changes in generalized travel time with respect to the AR capacity improvement for a commuter (with EAP β)

on the freeway and the AR, respectively. Furthermore, by analyzing $\partial \frac{\partial C_i(\beta)}{\partial s_j} / \partial \beta$ (where $i = f$ or a, and $j = f$ or a), we are able to show how the reduction of generalized travel time varies for each commuter.

With respect to the freeway capacity improvement, we have (please see Appendix C for the derivations),

$$\frac{\partial C_f(\beta)}{\partial s_f} < 0 \tag{3.32a}$$

$$\frac{\partial \frac{\partial C_f(\beta)}{\partial s_f}}{\partial \beta}|_{\beta=a} < 0, \frac{\partial \frac{\partial C_f(\beta)}{\partial s_f}}{\partial \beta}|_{\beta=a'} > 0, \frac{\partial \frac{\partial C_f(\beta)}{\partial s_f}}{\partial \beta} \text{is monotone w.r.t } \beta \tag{3.32b}$$

$$\frac{\partial C_a(\beta)}{\partial s_f} < 0, \frac{\partial \frac{\partial C_a(\beta)}{\partial s_f}}{\partial \beta} = 0 \tag{3.32c}$$

The derivatives imply that improving freeway capacity makes every commuter better off. In particular, it equally benefits commuters on the AR in travel time reduction. However, for commuters on the freeway, it benefits those with medium values of β more than those with high or low values of β, in terms of travel time reduction. Now we illustrate these with an example. Let $\tau = 15$ min, $s_a = 90$ veh/min, s_f changes from 90 veh/min to 100veh/min. We plot the changes in generalized travel time of all commuters ordered by the value of β in a mildly congested network ($N = 10000$) and a highly congested network ($N = 20000$), as shown in Figure 3.7. The left vertical axis measures the travel time reduction in the unit of minutes, while the right axis measures the reduction ratios(the travel time reduction over the original travel time) in the unit of percentage.

When $N = 10000$, enlarging freeway capacity benefits commuters on the AR equally by 1.4 minutes, whereas it benefits commuters on the freeway differently by ranging from 0.5 minutes to 1.3 minutes. Due to freeway capacity enlargement, those commuters on the AR with β ranging from 0.47 to 0.49 switch to the freeway, and receive less benefits than those remaining on the AR. On the other hand, those commuters with low values of β on the freeway obtain maximal travel time reduction ratios, while the travel time reduction ratio of commuters on the AR keep relatively constant. The changes in both the reduction and reduction ratios over all the commuters when $N = 20000$ seem similar to those when $N = 10000$, only that $N = 20000$ generally yields more travel time reduction with, however, less travel time reduction ratios.

When the capacity of the AR is expanded, we have (please see Appendix C for the derivations),

$$\frac{\partial C_f(\beta)}{\partial s_a} < 0, \frac{\partial \frac{\partial C_f(\beta)}{\partial s_a}}{\partial \beta} < 0 \text{ and is independent of } \beta \tag{3.33a}$$

$$\frac{\partial C_a(\beta)}{\partial s_a} < 0, \frac{\partial \frac{\partial C_a(\beta)}{\partial s_a}}{\partial \beta} < 0 \tag{3.33b}$$

The derivatives imply that improving AR capacity makes every commuter better off as well. In particular, travelers with higher EAP β have a greater reduction in their generalized travel time if they remain on the same route. Now we use a numerical example to illustrate this. Let $\tau = 15$ min, $s_f = 90$ veh/min, s_a changes from 90 veh/min to 100veh/min. We plot the changes in generalized travel time of all commuters ordered by the value of β in a mildly congested network ($N = 10000$) and a highly congested network ($N = 20000$), as shown in Figure 3.8. The left vertical axis measures the travel time reduction in the unit of minutes, while the right axis measures the reduction ratios(the travel time reduction over the original travel time) in the unit of percentage.

When $N = 20000$, enlarging AR capacity benefits commuters on the freeway linearly from 0.3 minutes (where $\beta = 0.1$) to 1.4 minutes (where $\beta = 0.41$), whereas it benefits commuters on the AR by ranging from 1.1 minutes to 1.7 minutes. Commuters with β at approximately 0.41 switch from the freeway to the AR due to the AR capacity enlargement. As expected, the marginal capacity enlargement is more helpful in a highly congested network than the mildly congested one, because both the reduction of each commuter's travel time and the reduction ratio are more significant when $N = 20000$.

3.5 The Case of A Multi-route Network

Now, we consider a one-to-one network connected by n routes, including a freeway route, and $n-1$ arterials. The routes are ordered by their free flow travel time. Suppose the free-flow travel times and the capacity of route i is τ_i and s_i respectively. Hence, $\tau_1 < \tau_2 < \cdots < \tau_n$. τ_1 is set to be 0 without loss of generality. Proposition 3.4 can be immediately extended in the case of multiple routes. Since only travelers with a critical value of β will choose either of two neighbored routes, the domain of β will be split into i intervals (suppose that the freeway route and i arterials are used, $i + 1 < n$),

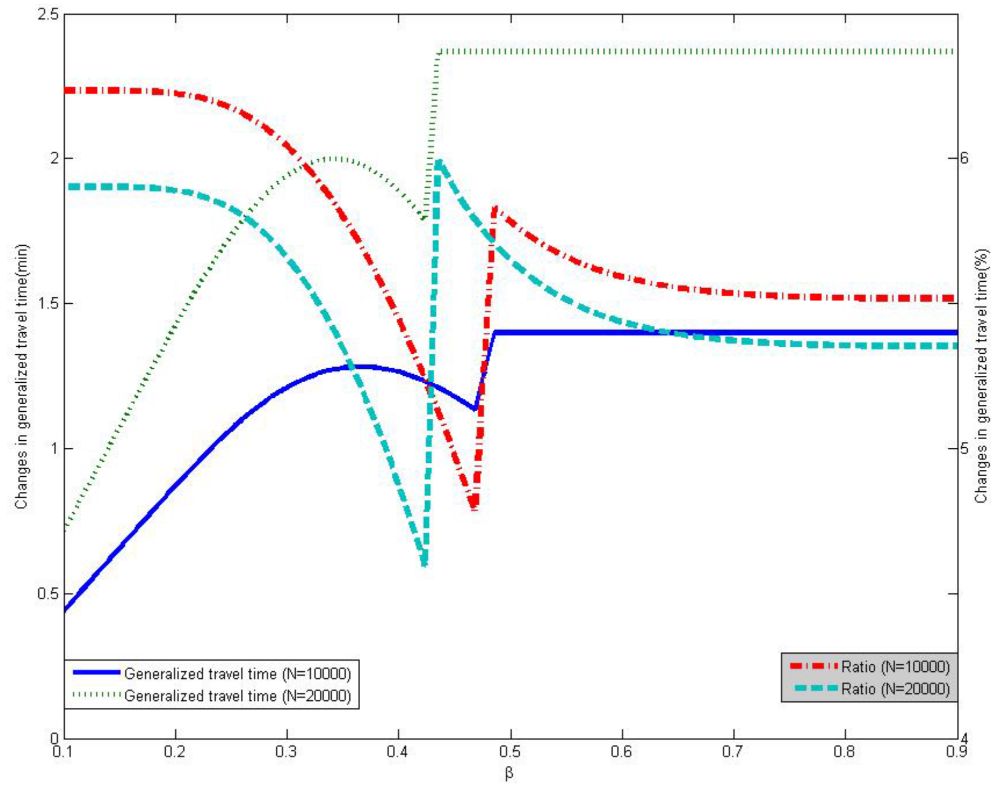


Figure 3.7. Changes in generalized travel time of all commuters in regards to freeway capacity improvement

$[a, a_1], [a_1, a_2] \cdots, [a_i, b]$, such that the freeway route is used by travelers whose values of β are within $[a, a_1]$ and the arterial road k is used by those within $[a_k, a_{k+1}]$ $(a_{i+1} = b)$.

The following i equations can be immediately obtained as an extension of Equation 3.25, in order to solve for $a_1, \cdots, a_i$.

$$N\frac{\delta}{s_k}\int_{a_{k-1}}^{a_k} xf(x)dx = \tau_{k+1} - \tau_k + N\frac{\delta}{s_{k+1}}a_k\int_{a_k}^{a_{k+1}} f(x)dx \quad \forall k = 1, 2, \ldots, i \qquad (3.34)$$

where $a_{i+1} = b$ and $a_0 = a$.

The bi-section method can be used to efficiently solve the multi-route morning commute problem under heterogeneity.

Step 1: Initialization: $a_{1,left} = a$ and $a_{1,right} = b$.

Step 2: Loop and stop criteria: If $a_{1,right} - a_{1,left} >$ a small positive number, let $a_1 = a_{1,left}/2 + a_{1,right}/2$ and go to loop, otherwise stop.

Step 3: Sub-loop:

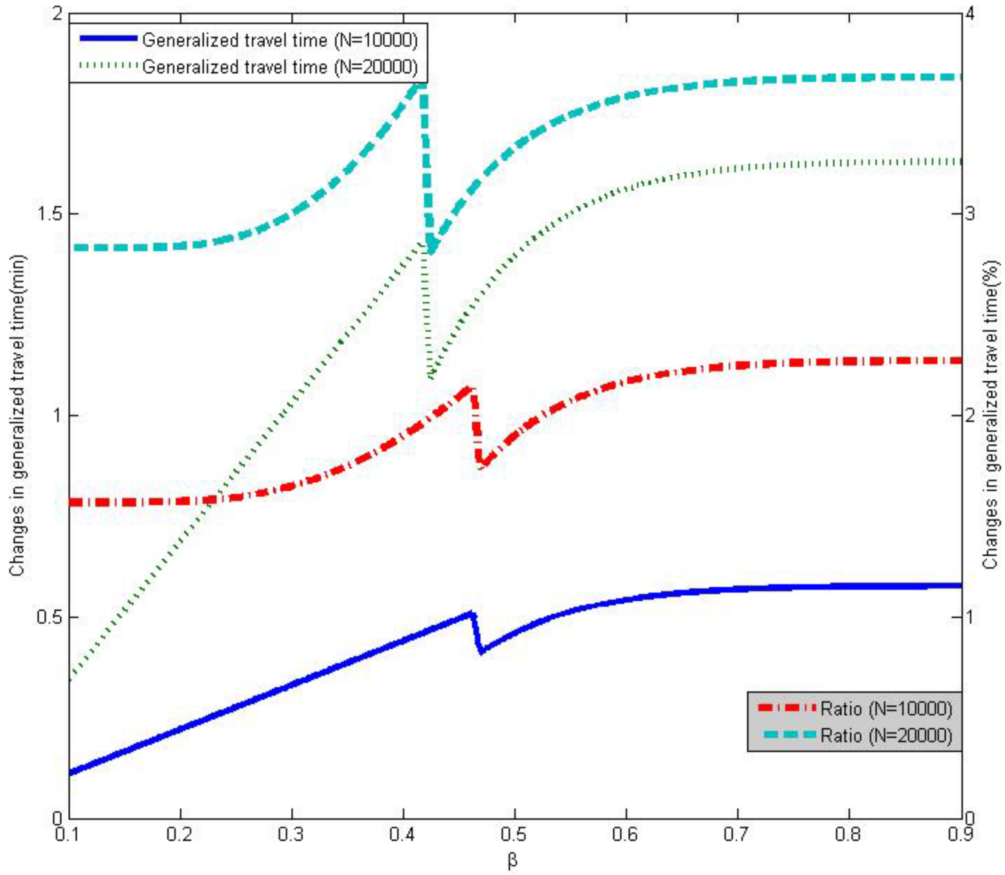


Figure 3.8. Changes in generalized travel time of all commuters in regards to AR capacity improvement

Step 3.1 $k = 1$

Step 3.2 Calculate a_{k+1} by substituting a_k in Equation 3.34.

If $a_{k+1} = b$, stop the loop and solution found. $i = k$.

If $N\frac{\delta}{s_k}\int_{a_{k-1}}^{a_k} xf(x)dx - (\tau_{k+1} - \tau_k) < 0$, then $a_{1,left} = a_1$, stop the sub-loop.

If $N\frac{\delta}{s_k}\int_{a_{k-1}}^{a_k} xf(x)dx > \tau_{k+1} - \tau_k + N\frac{\delta}{s_{k+1}}a_k\int_{a_k}^{b} f(x)dx$, then $a_{1,right} = a_1$, stop the sub-loop.

Step 3.3 $k = k + 1$

3.6 Summary

In this chapter, we study the morning commute problem with a heterogeneous traveling population whose early/late arrival penalty parameters β/γ are continuously distributed. Following Arnott et al. (1988), where the ratio of β over γ is assumed constant across the population and all the travelers have the same work starting time t^*, we derive the user-optimal travel profiles and the corresponding total generalized travel time (TTT) for a one-to-one network with a single route, where β and γ are continuously distributed. We show that assuming homogeneity overestimates the queuing delay and thus the total travel time. In addition, every commuter is better off if the bottleneck's capacity is enlarged, and commuters with high values of β benefit more than those with low values.

We then extend our analysis to a network with two routes, a freeway and an AR connecting a single Origin-Destination (O-D) pair. In this case, we find that the travelers who dislike to arrive early/late more (corresponding to higher β/γ values) will first shift to the AR with the increase of total demand. However, the critical travel demand N^* at which travelers start to use the AR remain the same whether the traveling population is homogeneous or not. Interestingly, there exists a critical a' such that the travelers with $\beta \leq a'$ all choose the freeway and those with $\beta \geq a'$ all choose the AR. Our numerical examples show that considering heterogeneity across the population leads to an estimate of larger route share of the freeway than not considering heterogeneity. We further investigate the impact of the facility improvement on the system and most results are consistent with those under homogeneity with respect to changes in TTT. However, unlike the homogeneous case, whether the marginal freeway(AR) improvement has a positive effect on $TTT_f(TTT_a)$ depends on all the parameters and β distributions, and sometimes β distributions can dominate this effect. Additionally, we show that every commuter is better off if either the freeway capacity or the AR capacity is enlarged.

As a matter of fact, we can easily extend our conclusion to solve a multi-route morning commute problem under heterogeneity where a one-to-one network is connected by n routes. Suppose those routes are ordered by their free flow travel time. Since only travelers with a critical value of β can choose either of two adjacent routes in such an order, the domain of β will be split into i intervals where i is the number of used routes and travelers with the value of β in each interval only use a certain route. This extension allows applications in more general networks.

Chapter 4

Traffic Mode Heterogeneity: Multi-modal morning commute

In this chapter, we address the morning commute problem in a single-origin-single-destination (SOSD) corridor network with three modes, driving-alone, carpool, and light rail transit. Under the equilibrium framework, departure time patterns will be derived for commuters of all three travel modes. The mode shares, on the other hand, are determined by a nested-logit choice model. Analytical and numerical results will then be obtained to study how mode shares and the performance of the corridor network are affected by various "price signals", such as changes in transit fare, road toll and fuel cost.

4.1 Problem set-up

An SOSD corridor, comprised of a light rail line, a freeway route and an arterial route, is shown in Figure 4.1. It is assumed that all the travelers have identical expected work start time, t^* say, 8:00am. The freeway and arterial roads are open to all the travelers at any time, while the light rail line has a fixed schedule without running delay. This is because a light rail route normally has its own physical facilities paralleled with roadways, and the travel time on the railway route is independent of roadway traffic conditions. The transit line is assumed to have several runs centered around the scheduled work starting time t^*. Toll, if exists, is charged to freeway travelers only. The time-varying passenger demand may exceed the capacity of the light rail line, but the light rail passengers are subject to an in-vehicle congestion or reliability penalty σ.

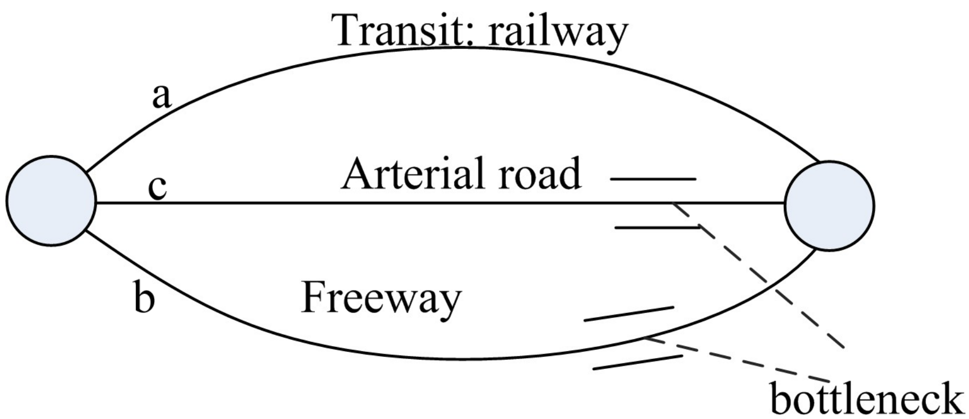


Figure 4.1. An SOSD multi-modal corridor network

Note that the generalized travel cost of a commuter departing at time t is defined to be $C(t) = \alpha w(t) + \max\{\beta(t^* - t - w(t)), \gamma(t + w(t) - t^*)\}$. Since we assume a homogenous population with identical α, γ and β, the UE solution does not change if we define $C'(t) = w(t) + \max\{\beta'(t^* - t - w(t)), \gamma'(t + w(t) - t^*)\}$ where $\beta' = \beta/\alpha$ and $\gamma' = \gamma/\alpha$. Here $C'(t)$ becomes the generalized travel time. In this chapter, we shall use travel time as the measuring unit of travel cost.

Let the index a denote the transit route, b the freeway route, and c the arterial route. The indices $1, 2, 3$ denote the modes of transit, carpool and solo-driving respectively. Let N_1 be the transit passenger flow and $\mathbf{N_{auto}} = (N_{2,b}, N_{2,c}, N_{3,b}, N_{3,c})$ the auto *passenger flow* pattern for the two auto modes on the two routes. Consequently, $N_c = N_{2c}/m + N_{3c}$ and $N_b = N_{2b}/m + N_{3b}$ are the number of vehicles traveling via the arterial road (AR) and the freeway, respectively, where m denotes the average number of passengers in a carpooling vehicle. Let N be the total passenger travel demand in the corridor, which is known and fixed, then $N = N_1 + N_{2b} + N_{3b} + N_{2c} + N_{3c}$.

Unlike in the auto mode where travelers depart continuously over time, a light rail has a fixed schedule and departs in discrete intervals so its passengers can only depart from their boarding stations at these fixed times, although they can depart from their homes at any time. For instance, the scheduled departure time of the transit line is approximately at five-minute intervals from 7:30am to 8:15am. The generalized travel time of a transit user at departure time t, $c_1(t)$, is defined as:

$$c_1(t) = t_a + max[\gamma(t + t_a - t^*), \beta(t^* - t - t_a)] + p + \sigma(t) \tag{4.1}$$

where t_a is the sum of expected waiting time, walking time and in-vehicle time for a transit passenger. t_a is assumed to be fixed. $max[\gamma(t + t_a - t^*), \beta(t^* - t - t_a)]$ is the early arrival or late arrival penalty as mentioned before. There is a uniform transit fare p (denominated in time) for all travelers. $\sigma(t) = \eta \cdot v_k \cdot h$ can be considered as a mode-specific cost related to actual or perceived inconveniences of the transit mode. η is a scalar, h is the fixed transit headway during rush hours, and v_k denotes the transit passenger flow on the k-th run. This formula for $\sigma(t)$ implies that 1) if the transit demand is fixed, the higher the transit frequency, the less likely a passenger would miss a train or waiting longer for a train, hence the less travel cost, and 2) if the transit schedule is fixed, the higher the transit passenger flow on a run, the more crowded the train cars would be, hence the less comfort of a train ride, which in itself is also a cost to the passengers.

Let $c_{2b}(t)$ and $c_{3b}(t)$ be the generalized travel time of a private auto commuter at departure time t via the freeway, for carpooling commuters and driving-alone commuters respectively.

$$c_{2b}(t) = w_b(t) + max[\gamma(t + w_b(t) - t^*), \beta(t^* - t - w_b(t))] + \frac{\xi_b}{m} + \Delta \tag{4.2}$$

$$c_{3b}(t) = w_b(t) + max[\gamma(t + w_b(t) - t^*), \beta(t^* - t - w_b(t))] + \xi_b \tag{4.3}$$

Where $w_b(t)$ is the waiting time at the bottleneck on the freeway departing at time t. $max[\gamma(t + w_b(t) - t^*), \beta(t^* - t - w_b(t))]$ is the early arrival or late arrival penalty, denominated in time. ξ_b consists of fuel cost and a uniform toll charge, all denominated in time, on a vehicle basis. Δ denotes the extra time cost of gathering the people together for carpoolers. The capacity of the GP lanes and the HOV lanes of the bottleneck on the freeway is given by s_f and $\phi \cdot s_f$, respectively. The free flow travel time on the freeway is set to be 0 without loss of generality.

Similarly, let $c_{2c}(t)$ and $c_{3c}(t)$ be the generalized travel time of a private auto trip at departure time t via the AR, for carpooling commuters and driving-alone commuters respectively.

$$c_{2c}(t) = t_c + w_c(t) + max[\gamma(t + w_c(t) - t^*), \beta(t^* - t - w_c(t))] + \frac{\xi_c}{m} + \Delta \tag{4.4}$$

$$c_{3c}(t) = t_c + w_c(t) + max[\gamma(t + w_c(t) - t^*), \beta(t^* - t - w_c(t))] + \xi_c \tag{4.5}$$

Where $w_c(t)$, $max[\gamma(t + w_c(t) - t^*), \beta(t^* - t - w_c(t))]$, and ξ_c are defined similarly as in the generalized travel time for travelers who use the freeway. The capacity of the bottleneck on the AR is denoted by s_{ar}. The free flow travel time on the AR is set to be t_c.

When the freeway is not tolled, we should have $\xi_c > \xi_b$ because the AR is usually longer than the freeway and takes longer to travel as well (hence a vehicle travels on it consumes more fuel). In this chapter, the AR is not just one arterial but the aggregate of a collection of parallel arterials, we assume that $s_{ar} \geq s_f$ holds. The AR, however, have longer free-flow travel time than the freeway since it has a lower speed limit.

Once her generalized travel time is known, a commuter would choose her mode of travel based on the relative travel time differences among the mode. The modal split of the commuter population is given by a nest logit model:

$$\frac{N_{2c} + N_{2b}}{N_{3c} + N_{3b}} = e^{\theta(C_2 - C_3)} \tag{4.6}$$

$$\text{Define } \Gamma = \ln(e^{\theta(C_2 - C_3)} + 1)$$

$$\frac{N_1}{N - N_1} = \frac{e^{\theta' C_1}}{e^{\theta'(C_3 + \Gamma/\theta)}} \tag{4.7}$$

where C_2 and C_3 are the equilibrated generalized travel time of carpooler and solo-drivers, respectively, and θ and θ' are the parameters of the logit model, to be calibrated and are given exogenously in this chapter. The equilibrium travel costs C_2 and C_3 for the auto mode, on the other hand, are obtained from an user-equilibrium analysis based on the dual-mode bottleneck model presented below.

4.2 The Dual-mode Bottleneck Model Considering Carpooling Lanes

In this section we develop a dual-mode morning commute model for an SOSD network with HOV lanes. Note that in this chapter, HOV lanes are assumed to be present along the entire freeway route connecting the origin to the destination. In addition, the capacity of the general purpose lane is assumed not affected by the presence of HOV lanes besides them. As was done in the classical morning commute problem Arnott et al. (1988), the generalized travel cost of a commuter departing at time t is given by

$$C(t) = \alpha w(t) + \max\{\beta(t^* - t - w(t)), \gamma(t + w(t) - t^*)\}$$

Assuming free-flow travel time is zero without loss of generality, where $w(t)$ denotes the commuting time cost, and t^* the desired arrival time (work start time) for all the commuters. Here α, β and γ measure the generalized cost of one extra minute of queuing delay, early schedule delay and late schedule delay, respectively. Solving this problem under user equilibrium (UE) conditions for given fixed travel demand N yields the generalized travel cost for all the commuters, $C = \delta N/s$, where s is the bottleneck capacity and $\delta = \gamma\beta/(\gamma + \beta)$.

We assume that the generalized travel cost applies to all the commuters and there are K driving-alone vehicles that are only allowed to use only the general-purpose (GP) lanes with the capacity s_f, and J carpool vehicles that are free to use any freeway lanes, including the GP lanes and HOV lanes. Furthermore, the capacity of the HOV lanes is assumed to be $\phi \cdot s_f$ where ϕ is the capacity ratio of HOV facilities to the freeway. Let J_1 and J_2 denote the number of carpooling vehicles using GP lanes and HOV lanes, respectively. Under UE conditions, i.e. no commuter can choose any other departure time to reduce his travel cost, we have:

$$C_{GP} = \delta \frac{K + J_1}{s_f} = C_{HOV} = \delta \frac{J_2}{\phi \cdot s_f} \tag{4.8}$$

Therefore,

$$J_1 = \begin{cases} \frac{J - \phi K}{1 + \phi} & \text{if } J > \phi K \\ 0 & \text{if } J \le \phi K \end{cases} \tag{4.9}$$

When the demand share of carpool is larger than its share of capacity, i.e. $\phi < J/K$, the above formulae show that carpool commuters have to use GP lanes as well, hence have no travel advantage over solo-drivers even when HOV lanes are provided, because

$$C = \delta \frac{K + J_1}{s_f} = \delta \frac{J_2}{\phi \cdot s_f} = \delta \frac{K + J}{s_f(1 + \phi)}, \tag{4.10}$$

i.e., the travel cost and equilibrated travel pattern with HOV lanes are exactly what they would be if all the lanes were GP lanes.

If the demand share of carpool does not exceed the share of its capacity in a freeway facility, i.e. $\phi \geq J/K$, an equilibrium departure pattern is achieved if and only if carpooling commuters use only the HOV lanes. The equilibrated traffic patterns are shown in Figure 4.2, and the travel cost of each group becomes:

$$C_1 = \delta \frac{K}{s_f},\ C_2 = \delta \frac{J}{\phi \cdot s_f} \tag{4.11}$$

As shown in Figure 4.2, solo-drivers commute from t_0 to t_1, and those departing at t_2 are subject to the largest queuing delay with punctual arrivals. Carpoolers commute from t'_0 to t'_1, and those departing at t'_2 are subject to the largest queuing delay with punctual arrivals.

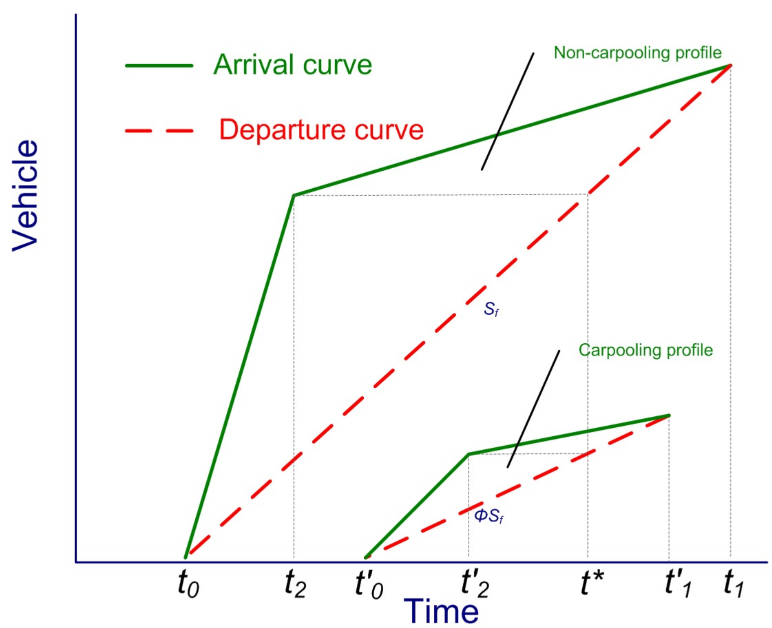


Figure 4.2. Travel profiles of solo-drivers and carpoolers

For the convenience of later exposition, we categorize the dual-mode morning commute into two cases: the case where carpool offers an travel advantage (i.e., less travel cost) and the case where carpool does not offer such an advantage. Case one corresponds to $J < \phi K$. i.e. the share of carpool demand is less than its share of capacity, and case two corresponds to $J \geq \phi K$, i.e. the share of carpool demand is greater than its share of roadway capacity.

4.3 The Morning Commute Model for The Transit Mode

Since we assume that the transit line has several runs, transit passengers between different runs should have the same generalized travel time under UE conditions. This problem can be solved in the same fashion as for its counterpart in the auto mode (the bottleneck in the latter is analogous to the item σ), and we obtain a similarly-shaped passenger flow profile in a staircase form.

Following the Equation 4.1, let k^+ denote the k-th run before the schedule time t^*, and k^- denote the k-th run after the schedule time. $k = 0$ is the run which is scheduled exactly at the scheduled working time. Now the transit travel cost becomes:

$$c_1(k) = \begin{cases} \eta \cdot v_{k^+} \cdot h + \beta(t^* - t) & \text{early arrival} \\ \eta \cdot v_{k^-} \cdot h + \gamma(t - t^*) & \text{late arrival} \\ \eta \cdot v_0 \cdot h & \text{punctual arrival} \end{cases} \tag{4.12}$$

Actually, $t^* - t = k^+h$ and $t - t^* = k^-h$. We substitute t by k and h in the above equation and have:

$$c_1(k) = \begin{cases} \eta \cdot v_{k^+} \cdot h + \beta \cdot k^+h & \text{early arrival} \\ \eta \cdot v_{k^-} \cdot h + \gamma \cdot k^-h & \text{late arrival} \\ \eta \cdot v_0 \cdot h & \text{punctual arrival} \end{cases} \tag{4.13}$$

According the the UE condition, i.e. no matter which run a transit passenger takes, she experiences the same travel cost. Therefore,

$$\begin{aligned} v_{i^+} - v_{j^+} &= \frac{\beta(j^+ - i^+)}{\eta} \\ v_0 - v_{j^+} &= \frac{\beta j^+}{\eta} \forall i, j \end{aligned} \tag{4.14}$$

The passenger flow for the run before the scheduled office time is:

$$(v_0 - \frac{\beta}{\eta}), (v_0 - \frac{2\beta}{\eta}), (v_0 - \frac{3\beta}{\eta}) \ldots\ldots (v_0 - \frac{n_2\beta}{\eta})$$
$$n_2 = int[\frac{v_0\eta}{\beta}]$$

$$v_{i^-} - v_{j^-} = \frac{\gamma(j^- - i^-)}{\eta} \tag{4.15}$$
$$v_0 - v_{j^-} = \frac{\gamma j^-}{\eta} \forall i, j$$

The passenger flow for the run after the scheduled office time is:

$$(v_0 - \frac{\gamma}{\eta}), (v_0 - \frac{2\gamma}{\eta}), (v_0 - \frac{3\gamma}{\eta}) \ldots\ldots (v_0 - \frac{n_1\gamma}{\eta})$$
$$n_1 = int[\frac{v_0\eta}{\gamma}]$$

Given transit demand N_1, $v_0 + \sum_{k=1}^{n_2} v_{k^+} + \sum_{k=1}^{n_1} v_{k^-} = N_1$, we obtain passenger flow v_0. The flow of other runs are obtained through equations 4.14 and 4.15.

$$v_0 = \frac{N_1 + \frac{\beta}{\eta}\frac{n_2^2+n_2}{2} + \frac{\gamma}{\eta}\frac{n_1^2+n_1}{2}}{1 + n_1 + n_2} \tag{4.16}$$

where $n_1 = int[\frac{v_0\eta}{\gamma}]$ and $n_2 = int[\frac{v_0\eta}{\beta}]$. It is somewhat difficult to solve for v_0 analytically, but we can choose initially a large n_1 and n_2. If we obtain a negative flow for any of the runs, n_1 or n_2 are then reduced. Such a process can be applied iteratively to obtain v_0.

Finally, the generalized travel time for a transit passenger at departure-time equilibrium is,

$$C_1 = t_a + p + v_0 \cdot \eta \cdot h \tag{4.17}$$

4.4 The Multi-modal Morning Commute Model

Given the fixed total passenger demand N, we would like to determine the transit passenger flow N_1 and the auto passenger flows $N_{2b}, N_{2c}, N_{3b}, N_{3c}$, as well as the total travel cost (denomitated in travel time) of all the passengers in the multi-modal corridor network:

$$TC = C_1N_1 + C_{2b}N_{2b} + C_{3b}N_{3b} + C_{2c}N_{2c} + C_{3c}N_{3c} \tag{4.18}$$

and that in auto network:

$$TC_{auto} = C_{2b}N_{2b} + C_{3b}N_{3b} + C_{2c}N_{2c} + C_{3c}N_{3c} \tag{4.19}$$

Some observations can be made even before we go into any detailed analysis. 1) Theoretically, there will always be commuters for each mode because the choice of mode is determined by a logit choice model, no matter how small or large N would be; 2) since the AR has higher travel cost even there is no congestion, it will not be used for sufficiently small N, and 3) from the dual-mode bottleneck model we know that carpool may not always has a travel advantage over solo-driving, hence cases where carpool does have an advantage will be analyzed separately. From these observations, we shall analyze the multi-modal morning commute problem under the four following conditions: 1)the arterial road is not used and carpool offers no travel advantage, 2)the arterial road is used and carpool offers no travel advantage, 3)the arterial road is not used and carpool offers travel advantage and 4)the arterial road is used and carpool offers travel advantage.

4.4.1 Arterial road not used, carpool offers no travel advantage

In this case, we have

$N_{2c} = N_{3c} = 0$, $N_{2b}/m\phi \geq N_{3b}$, $N_b = N_{2b}/m + N_{3b}$.

$$N_1 + N_{2b} + N_{3b} = N \tag{4.20}$$

$$C_{2b} = \delta \frac{N_b}{(1+\phi)s_f} + \frac{\xi_b}{m} + \Delta \tag{4.21}$$

$$C_{3b} = \delta \frac{N_b}{(1+\phi)s_f} + \xi_b \tag{4.22}$$

$$\frac{N_{2b}}{N_{3b}} = e^{\theta(C_{2b}-C_{3b})} = e^{\theta(\xi_b/m-\xi_b+\Delta)} \tag{4.23}$$

Define $e^{\theta(\xi_b/m-\xi_b+\Delta)} = \mu_1$ and $\Gamma = \ln(\mu_1+1)$,

$$\frac{N_1}{N_{2b}+N_{3b}} = \frac{e^{\theta' C_1}}{e^{\theta'(C_{3b}+\Gamma/\theta)}} = \frac{e^{\theta'(t_a+p+\eta\cdot h\cdot v_0(N_1))}}{e^{\theta'(\delta\frac{N_b}{(1+\phi)s_f}+\xi_b+\Gamma/\theta)}} \tag{4.24}$$

Solving equations 4.20, 4.23 and 4.24 we would get the three unknown variables, N_1, N_{2b}, N_{3b}. However, this solution makes sense only when, 1) carpool offers no travel advantage, i.e. $\mu_1 \geq m\phi$, and 2) the AR is not used, i.e.

$$C_{2b} = \delta \frac{N_b}{(1+\phi)s_f} + \frac{\xi_b}{m} + \Delta \leq C_{2c} = \frac{\xi_c}{m} + \Delta + t_c \tag{4.25}$$

$$C_{3b} = \delta \frac{N_b}{(1+\phi)s_f} + \xi_b \leq C_{3c} = \xi_c + t_c \tag{4.26}$$

Because $(C_{2c}-C_{2b})-(C_{3c}-C_{3b}) = (1/m-1)(\xi_c-\xi_b) < 0$ given $m > 1$, $\xi_c > \xi_b$, checking Equation 4.25 is sufficient to find if the AR is not used.

If $\mu_1 < m\phi$ or Equation 4.25 does not hold, we must consider other cases to solve for the passenger flows.

4.4.2 Arterial road not used, carpool offers travel advantage

In this case, we have

$N_{2c} = N_{3c} = 0$, $N_{2b}/m\phi < N_{3b}$, $N_b = N_{2b}/m + N_{3b}$.

$$N_1 + N_{2b} + N_{3b} = N \tag{4.27}$$

$$C_{2b} = \delta \frac{N_{2b}}{m\phi s_f} + \frac{\xi_b}{m} + \Delta \tag{4.28}$$

$$C_{3b} = \delta \frac{N_{3b}}{s_f} + \xi_b \tag{4.29}$$

$$\frac{N_{2b}}{N_{3b}} = e^{\theta(C_{2b} - C_{3b})} \tag{4.30}$$

Define $\Gamma = \ln(e^{\theta(C_{2b}-C_{3b})} + 1)$,

$$\frac{N_1}{N - N_1} = \frac{e^{\theta' C_1}}{e^{\theta'(C_{3b} + \Gamma/\theta)}} = \frac{e^{\theta'(t_a + p + \eta \cdot h \cdot v_0(N_1))}}{e^{\theta'(\delta \frac{N_{3b}}{s_f} + \xi_b + \Gamma/\theta)}} \tag{4.31}$$

Solving Equations 4.27, 4.30 and 4.31 we obtain N_1, N_{2b}, N_{3b}.

Note that this solution requires the following two conditions to be true: 1) carpool offers a travel advantage, i.e. $N_{2b}/m\phi < N_{3b}$, and 2) the AR is not used, i.e.

$$C_{2b} = \delta \frac{N_{2b}}{m\phi s_f} + \frac{\xi_b}{m} + \Delta \leq C_{2c} = \frac{\xi_c}{m} + \Delta + t_c \tag{4.32}$$

$$C_{3b} = \delta \frac{N_{3b}}{s_f} + \xi_b \leq C_{3c} = \xi_c + t_c \tag{4.33}$$

Checking both 4.32 and 4.33 is necessary before finding a solution for this case.

4.4.3 Arterial road used, carpool offers no travel advantage

Under these conditions, we have

$N_{2b}/m\phi \geq N_{3b}$, $N_b = N_{2b}/m + N_{3b}$, $N_c = N_{2c}/m + N_{3c}$.

$$C_{2b} = \delta \frac{N_b}{(1+\phi)s_f} + \frac{\xi_b}{m} + \Delta,\ C_{2c} = \delta \frac{N_c}{s_{ar}} + \frac{\xi_c}{m} + \Delta + t_c \tag{4.34}$$

$$C_{3b} = \delta \frac{N_b}{(1+\phi)s_f} + \xi_b,\ C_{3c} = \delta \frac{N_c}{s_{ar}} + \xi_c + t_c \tag{4.35}$$

$(C_{2c} - C_{2b}) - (C_{3c} - C_{3b}) < 0$ given $m > 1$, $\xi_c > \xi_b$. Therefore, $C_{2c} = C_{2b}$ and $C_{3c} = C_{3b}$ cannot hold simultaneously. We shall discuss the following circumstances: 1) If $C_{2c} = C_{2b}$, then $C_{3c} > C_{3b}$. This shows continuity following the case of "AR not used, carpool offers no travel advantage" with the increase of N. 2) If $C_{3c} = C_{3b}$, then $C_{2c} < C_{2b}$, i.e. $N_{2b} = 0$. This conflicts with the condition of carpool offering no travel advantage $N_{2b}/m\phi > N_{3b}$, and thus it will never occur. 3) $C_{2c} > C_{2b}$ and $C_{3c} < C_{3b}$ obviously conflict with $(C_{2c} - C_{2b}) - (C_{3c} - C_{3b}) < 0$. 4) $C_{2c} < C_{2b}$ and $C_{3c} > C_{3b}$ cannot occur because $N_{2b} = 0$.

Consequently, we show that $C_{2c} = C_{2b}$ and $C_{3c} > C_{3b}$ given that $\xi_c > \xi_b$. Moreover, $N_{3c} = 0$, if the AR is used and carpool offers no travel advantage. Now we obtain the following:

$$\delta \frac{N_b}{(1+\phi)s_f} + \frac{\xi_b}{m} = \delta \frac{N_c}{s_{ar}} + \frac{\xi_c}{m} + t_c \tag{4.36}$$

$$N_1 + N_{2b} + N_{3b} + N_{2c} = N \tag{4.37}$$

$$\frac{N_{2b} + N_{2c}}{N_{3b}} = e^{\theta(C_{2b} - C_{3b})} = e^{\theta(\xi_b/m - \xi_b + \Delta)} \tag{4.38}$$

Define $e^{\theta(\xi_b/m - \xi_b + \Delta)} = \mu_1$ and $\Gamma = \ln(\mu_1 + 1)$,

$$\frac{N_1}{N - N_1} = \frac{e^{\theta' C_1}}{e^{\theta'(C_{3b} + \Gamma/\theta)}} = \frac{e^{\theta'(t_a + p + \eta \cdot h \cdot v_0(N_1))}}{e^{\theta'(\delta \frac{N_b}{(1+\phi)s_f} + \xi_b + \Gamma/\theta)}} \tag{4.39}$$

Solving Equations 4.36, 4.37, 4.38 and 4.39, we would get $N_1, N_{2b}, N_{3b}, N_{2c}$.

The result shows that in this case, only carpool vehicles travel on the arterial route when traffic on the freeway is getting so congested that the HOV lanes on the freeway no longer offer a travel advantage over the general purpose lanes. This result may seem counter-intuitive. Yet once considering the factor of fuel cost, which is shared among carpoolers in each vehicle, one can easily understand why this happens: the cost difference between the AR and the freeway for solo-drivers can be considerably larger than that for carpoolers, so there is less incentive for solo-drivers to travel on the arterial route, yet the equilibrium solution under this scenario requires that some travelers would use the arterial route, and they must the carpoolers.

However, as the total passenger demand N increases, even solo-drivers may be forced to take the arterial route to reduce their travel cost. Under certain circumstances, this may create a situation where taking the HOV lanes become advantageous again, as can be seen from the following analysis.

Express $N_c = N_{2c}/m$ with N_b by solving Equation 4.36, then substitute it in Equation 4.38, we get:

$$\frac{N_{2b}}{N_{3b}} = \frac{\frac{\delta\mu_1}{ms_{ar}} - \frac{\delta}{(1+\phi)s_f}}{\frac{\delta}{m(1+\phi)s_f} + \frac{\delta}{ms_{ar}}} + \frac{\frac{\xi_c - \xi_b}{m} + t_c}{(\frac{\delta}{m(1+\phi)s_f} + \frac{\delta}{ms_{ar}})N_{3b}} \tag{4.40}$$

From Equation 4.38 we know that $N_{2b}/N_{3b} < \mu_1$. With the increase of N_{3b}, N_{2b}/N_{3b} may decrease and approach ϕm (we can always find a large N such that N_{3b} is large enough to make N_{2b}/N_{3b} less or equal to $m\phi$ if $\frac{\delta\mu_1}{ms_{ar}} < \frac{\delta}{(1+\phi)s_f}$). Once it's less than ϕm, i.e. with the increase of N, more solo drivers on the margin are assigned to the freeway than carpoolers so that carpool offers travel advantage on the freeway, the increase rate of the travel time on the freeway will be less than the AR for carpooling commuters. Since carpooling commuters can take this advantage, the freeway becomes more attractive to them. As more carpoolers take the freeway, the arterial route now becomes more attractive to solo-drivers and some of them will take it, which is exactly the case in the next subsection.

4.4.4 Arterial road used, carpool offers travel advantage

In this case, we have $N_{2b}/m\phi < N_{3b}$, $N_b = N_{2b}/m + N_{3b}$, $N_c = N_{2c}/m + N_{3c}$.

$$C_{2b} = \delta\frac{N_{2b}}{m\phi s_f} + \frac{\xi_b}{m} + \Delta,\ C_{2c} = \delta\frac{N_c}{s_{ar}} + \frac{\xi_c}{m} + \Delta + t_c \tag{4.41}$$

$$C_{3b} = \delta\frac{N_{3b}}{s_f} + \xi_b,\ C_{3c} = \delta\frac{N_c}{s_{ar}} + \xi_c + t_c \tag{4.42}$$

There are five sub-cases,

1. $C_{2c} = C_{2b}$, $C_{3c} > C_{3b}$. Thus, $N_{3c} = 0$, i.e. no solo drivers will choose the AR.

$$C_{2b} = \delta\frac{N_{2b}}{m\phi s_f} + \frac{\xi_b}{m} + \Delta = C_{2c} = \delta\frac{N_c}{s_{ar}} + \frac{\xi_c}{m} + \Delta + t_c \tag{4.43}$$

$$\frac{N_{2c} + N_{2b}}{N_{3b}} = e^{\theta(C_{2b} - C_{3b})} \tag{4.44}$$

$$N_1 + N_{2b} + N_{3b} + N_{2c} = N \tag{4.45}$$

Solving Equations 4.43, 4.44, 4.45 and 4.31, we obtain N_1, N_{2b}, N_{3b} and N_{2c}.

2. $C_{2c} = C_{2b}$, $C_{3c} < C_{3b}$ or $C_{2c} > C_{2b}$, $C_{3c} < C_{3b}$. $N_{3b} = 0$ conflicts with the condition $N_{2b}/m\phi < N_{3b}$.

3. $C_{2c} > C_{2b}$, $C_{3c} = C_{3b}$. $N_{2c} = 0$, i.e. no carpoolers will choose the AR.

$$C_{3b} = \delta \frac{N_{3b}}{s_f} + \xi_b = C_{3c} = \delta \frac{N_c}{s_{ar}} + \xi_c + t_c \tag{4.46}$$

$$\frac{N_{2b}}{N_{3c} + N_{3b}} = e^{\theta(C_{2b} - C_{3b})} \tag{4.47}$$

$$N_1 + N_{2b} + N_{3b} + N_{3c} = N \tag{4.48}$$

   Solving Equations 4.46, 4.47, 4.48 and 4.31, we obtain N_1, N_{2b}, N_{3b} and N_{3c}.

4. $C_{2c} < C_{2b}$, $C_{3c} = C_{3b}$ or $C_{2c} < C_{2b}$, $C_{3c} > C_{3b}$. $N_{2b} = 0$. This case will never occur because $C_{2b} = \xi_b/m + \Delta < C_{2c} = \delta \frac{N_c}{s_{ar}} + \xi_c/m + \Delta + t_c$

5. $C_{2c} = C_{2b}$, $C_{3c} = C_{3b}$, i.e. carpoolers and solo drivers will choose both the AR and the freeway.

$$\frac{N_{2c} + N_{2b}}{N_{3c} + N_{3b}} = e^{\theta(C_{2b} - C_{3b})} \tag{4.49}$$

$$N_1 + N_{2b} + N_{3b} + N_{2c} + N_{3c} = N \tag{4.50}$$

   Solving Equations 4.43, 4.46, 4.49, 4.50 and 4.31, we obtain $N_1, N_{2b}, N_{3b}, N_{2c}$ and N_{3c}.

Note that $N_{2b}/m\phi < N_{3b}$ always holds for the solution in this condition, given that $e^{\theta(\xi_b/m - \xi_b + \Delta)} = \mu_1 < \phi m$. This is because $N_{2b}/N_{3b} \geq m\phi > \mu_1$ conflicts with the conclusion that $N_{2b}/N_{3b} < \mu_1$ given no solution in the case where "carpool offers no travel advantages", i.e. Sections 4.4.1 and 4.4.3.

4.5 Sensitivity Analysis

In this section, we numerically solve the multimodal morning commute model and study how network performance and the mode split are influenced by several key factors in the model, such as total passenger demand, the existence of HOV lanes, the capacities of the bottlenecks and fuel costs.

The basic model parameters used in the later analysis are: $\eta = 0.02$/veh/person, $h = 5$min, $p = 2$min (of time equivalent), $t_a = 20$min, n_1 and n_2 are set to be no larger than 3 and 8 respectively, $\beta = 0.5$, $\gamma = 2$, $t_c = 15$min, $\phi = 0.15$, $N = 5000$ persons, $s_f = 40$veh/min, $s_{ar} = 60$veh/min, $\xi_b = 10$min (of time equivalent), $\xi_c = 20$min (of time equivalent), $\Delta = 10$min (of time equivalent), $m = 2$person/veh, $\theta = -0.1$, $\theta' = -0.2$. These parameters are fixed in the following numerical calculations unless noted otherwise.

Before presenting our numerical results, we first describe the solution procedure for the multi-modal morning commute model.

4.5.1 Solution procedure

Given a network and the total travel demand N, first we calculate $e^{\theta(\xi_b/m-\xi_b+\Delta)} = \mu_1$ and compare it with ϕm. 1) If $\mu_1 \geq \phi m$, we first assume AR is not used and carpool offers no travel advantage to solve the problem. Then check the solution by Equation 4.25. If Equation 4.25 holds, calculation terminates, otherwise, re-solve the problem under the condition of "AR used, carpool offers no travel advantage". If $N_{2b}/N_{3b} \geq \phi m$ holds, calculation terminates, otherwise, turn to 2). 2) If $\mu_1 < \phi m$ or no solution exists for $\mu_1 \geq \phi m$, we first assume "AR is not used and carpool offers a travel advantage" to solve the problem. If Equation 4.32 and 4.33 hold for the solution, calculation terminates, otherwise, re-solve the problem under the condition of "AR used, carpool offers a travel advantage". The calculation terminates finally.

From equations in all four cases listed in Section 4.4, it's easy to show that, given a total demand N, the solution $N_1, N_{2b}, N_{2c}, N_{3b}, N_{3c}$ exists and is unique.

4.5.2 The influence of total demand on route choice and mode split

Figure 4.3 and 4.4 show the transit passenger flow and auto passenger flow with respect to the total passenger demand N with different HOV capacities. Here the capacity of the general purpose lane is fixed, and as ϕ increases the both the capacity of the HOV lanes and the entire freeway also increases.

When $\phi = 1$, i.e. the share of the HOV capacity is sufficient to support its share of demand, the number of commuters who do carpool is obviously more than that of non-carpool commuters when $N > 2000$, and is nearly the same between $N = 1000$ and 2000. Except for the carpoolers on the AR, the passenger flow of each group on each route increases at a relatively fixed rate after $N > 7000$, meaning that the market share of each route and each mode remains steady after that point. Also, commuters do not use the AR until $N > 6400$. Moreover, carpoolers never choose the AR since the HOV facility can serve them with a travel advantage. On the other hand, solo-drivers start to choose between the freeway and the AR after $N = 6400$. In terms of mode share, carpool has slightly higher share than solo-driving, while transit has the lowest share among all three modes, because of a large t_a. It is noted that about $N = 2000$, the share of carpool roughly equals that of solo-driving.

When $\phi = 0.15$, i.e. the share of the HOV capacity is not sufficient to support its share of demand, solo-driving has the largest mode share. As discussed earlier, it is carpool commuters that would first travel on the arterial route because solo-drivers have less incentive to travel on the arterial route. With the increase of N, the increase in the proportion of carpooling commuters is more pronounced on the AR than on the freeway, and only a small percentage of them chooses to travel via the freeway. Although after $N = 4150$ the HOV lanes offer a travel advantage to carpoolers, it is still not enough to attract them back to the freeway, since their travel costs on the freeway would be higher if they switch to it. It can be seen that the mode shares of carpooling and solo-driving remain nearly constant with the increase of N. But the freeway's route share of solo-driving commuters keeps decreasing. In addition, compared to the case with $\phi = 0.15$, the number of transit commuters in the case with $\phi = 1$ decreases by nearly 25%, this is, adding more HOV capacity (hence the overall freeway capacity)reduced transit ridership and increased auto travel.

4.5.3 The influence of HOV's capacity share on mode shares and total travel cost

Figures 4.5, 4.6 and 4.7 show the relationship between the passenger flow of three travel modes and the total travel cost with respect to ϕ, while keeping the total freeway capacity, $(1+\phi)s_f = 40$veh/min, constant. Recall that ϕ measures the capacity share of

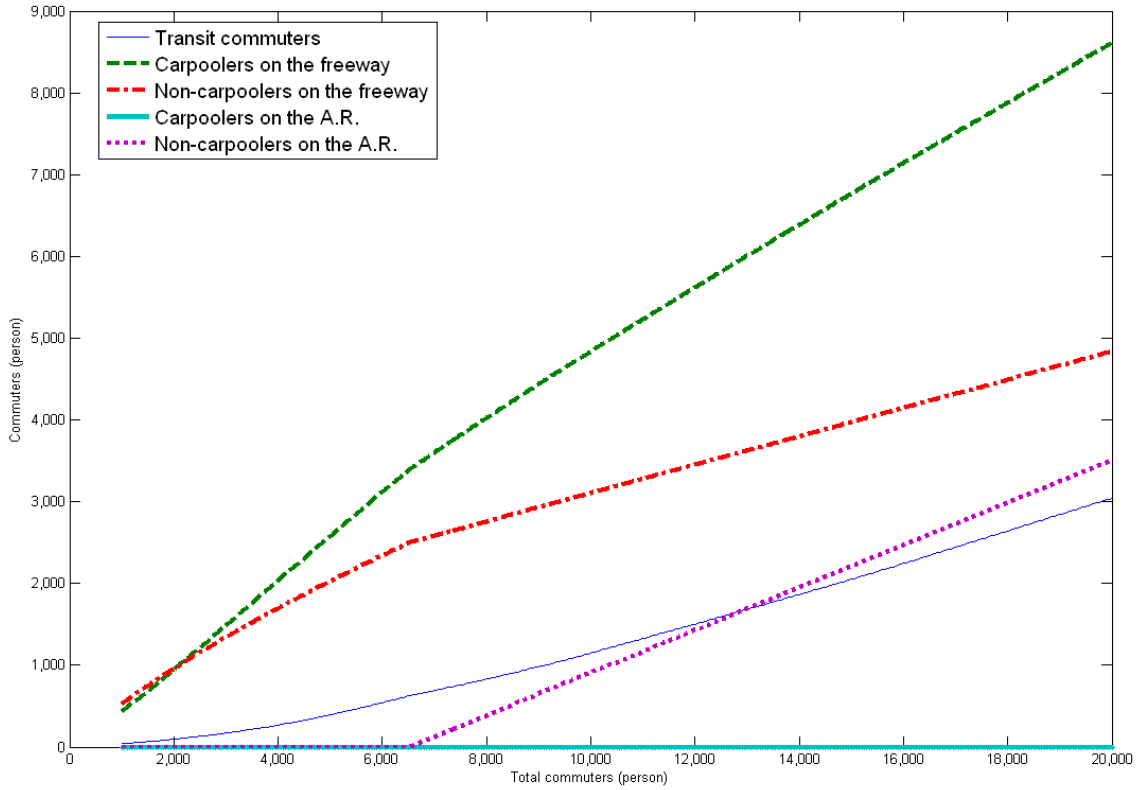


Figure 4.3. Changes in passenger flow with respect to the total demand ($\phi = 1$)

the HOV lanes on the freeway. From Figure 4.5, we can see that the capacity share of HOV lanes has only minor effect on transit ridership, but significant effect on the mode share between solo-driving and carpool, and on the route share between the freeway and the arterial road.

Undoubtedly, the advantages of carpooling, i.e. a large ϕ, will lead carpoolers to shift from the AR to the freeway, and solo drivers to shift from the freeway to the AR. Evidently, based on the case with $N = 10000$, when ϕ changes from 0.05 to 1, the carpoolers on the AR decreases from 3750 to 300, and so does the solo drivers on the freeway from 3750 to 2000. Similarly, the carpoolers on the freeway increases from 300 to 3500, so does the solo drivers on the AR from 300 to 2000. As a matter of fact, the mode share does not change much, regardless of route choices. Figure 4.6 indicates that, for a mildly congested network where no AR unused ($N = 2000$), providing more carpool advantages will significantly reduce the travel cost of the total network and the

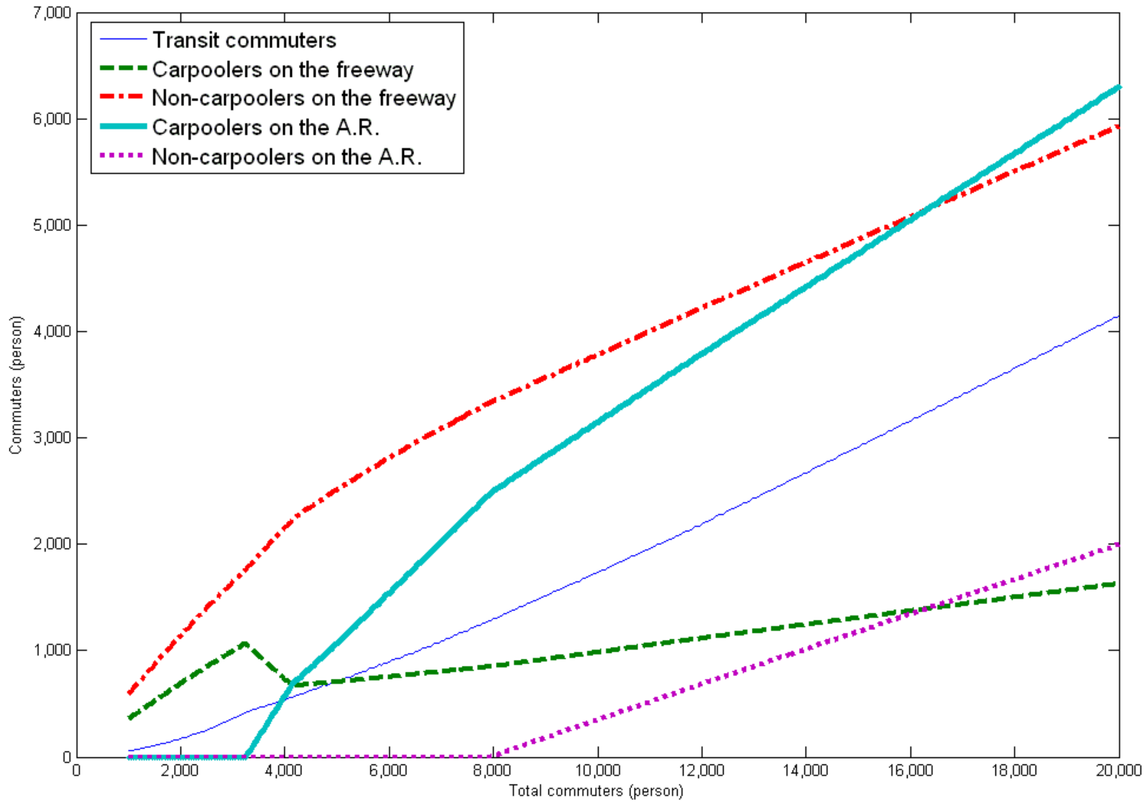


Figure 4.4. Changes in passenger flow with respect to the total demand ($\phi = 0.15$)

auto network. Nevertheless, for a highly congested network with AR highly used (e.g. $N = 8000$), increasing the HOV lanes' capacity share would increase the total travel, as shown in Figure 4.7. This latter paradoxical situation may occur when the total extra cost to the solo drivers who shifted from the freeway to the AR is less than the total cost saved by the carpoolers who shifted in the reverse direction. This implies that, providing more HOV capacity in this case may not necessarily improve the network's overall performance.

We also plotted the passenger flow of three travel modes with respect to Δ, as shown in Figure 4.8. The results indicate that reducing the time of picking up fellow carpoolers from 20min to 5min will increase the mode share of carpooling from 20% to 48% and reduce share of solo-driving from 58% to 37%. The increase of this inconvenience 'cost' also increases transit ridership, from a share of 15% to 23%, because some carpoolers may switch to transit when facing a rising inconvenience 'cost'.

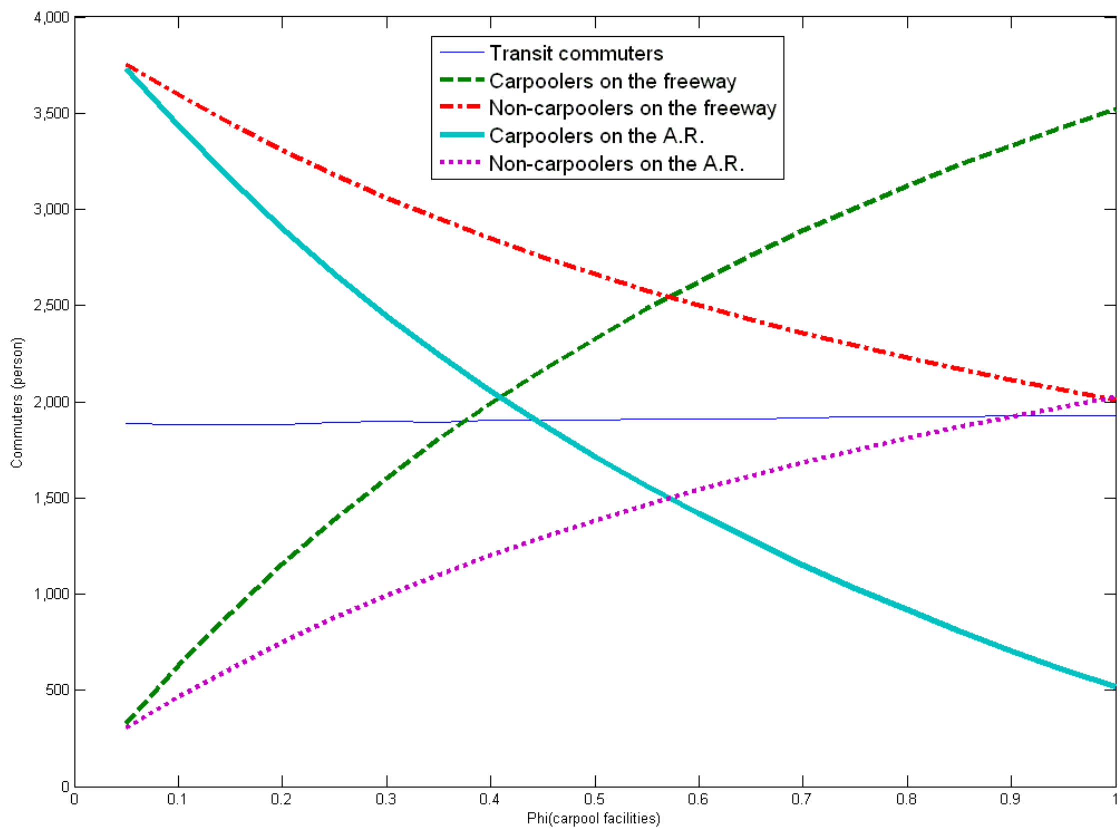


Figure 4.5. Changes in passenger flow of three traffic modes with respect to ϕ ($N =$ 10000)

4.5.4 The influence of transit fare and headway on transit ridership

We first investigate the changes in transit ridership N_1 with respect to the transit fare p (of time equivalent) and the headway h with $N = 10000$. When $N = 10000$, the corridor network is heavily congested. p varies from 1min to 3min (of time equivalent). The results show that increasing an uniform transit fare from 1 min-equivalent to 3 min-equivalent will reduce the transit percentage by 7%. When h changes from 3 min to 15 min, and the transit ridership sharply falls from 2550 persons to 490, by about 80%. This may be a strong indication that the service frequency of a transit line is very crucial to transit ridership. A transit line with more frequent runs in the rush hour may

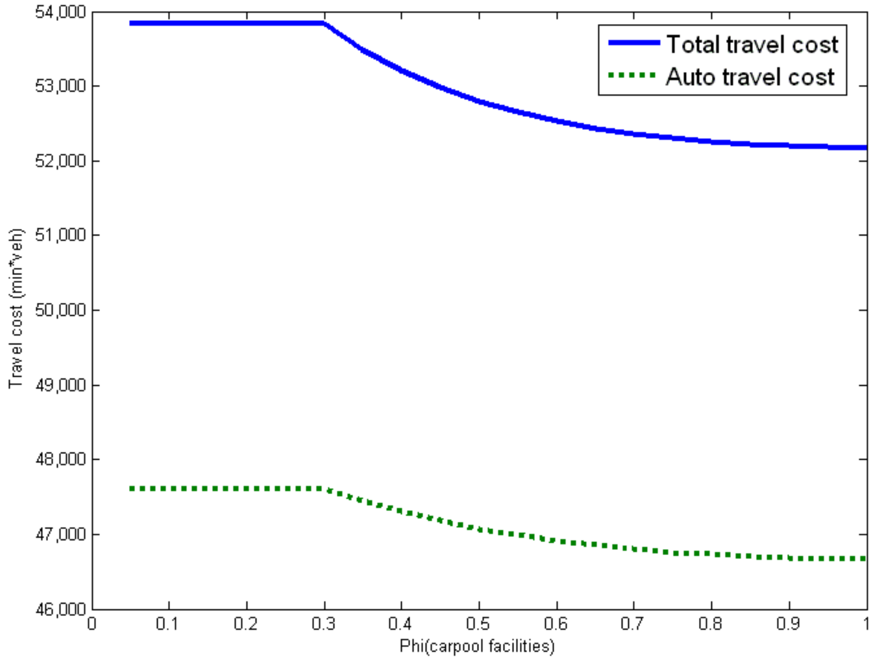


Figure 4.6. Changes in total travel cost with respect to ϕ ($N = 2000$)

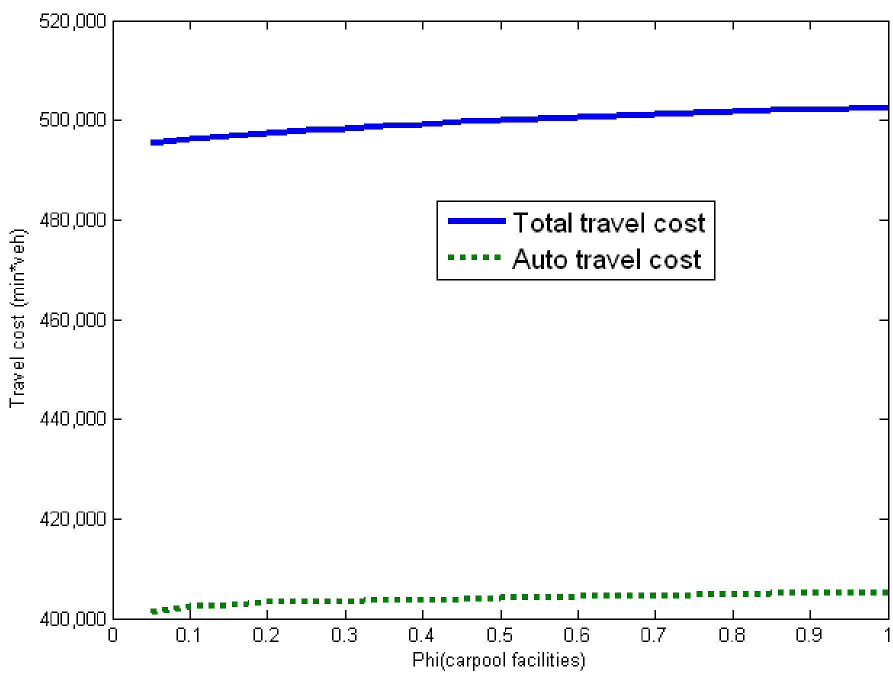


Figure 4.7. Changes in total travel cost with respect to ϕ ($N = 10000$)

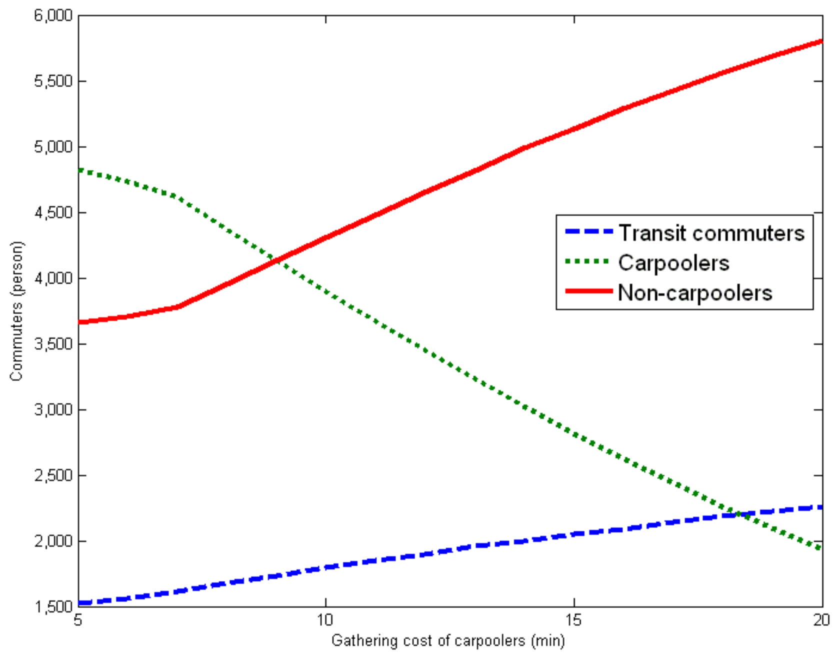


Figure 4.8. Changes in passenger flow of three traffic modes with respect to Δ ($N = 10000$)

lead to much higher mode share, which is expected.

4.5.5 The influence of fuel cost on mode split

In Figure 4.9, we plot the passenger flow of three travel modes with respect to fuel cost ξ_b (of time equivalent) when $N = 10000$. Note that both ξ_b and ξ_c will increase proportionally if gas price rises, assuming that toll is not charged on the freeway. Let $\xi_c = 2\xi_b$ and ξ_b vary from 4 min to 16 min (of time equivalent) [1]. When the fuel cost is below a certain level (12min-equivalent in this case), many solo-drivers switch to carpool and some even shift to transit as fuel cost increases. On average, the number of non-carpooling commuters reduces by about 10% on the current percentage basis, when the gas price rises by 50% on the current price basis. When the gas price rises from 4 min per trip (of time equivalent) to 12 min per trip (of time equivalent), the mode share of non-carpooling commuters reduces from 55% to 37% (a 32% reduction) and that of carpool commuters and transit commuters increase from 30% to 44% (a 47% increase)

[1]Note that under this assumption, $\xi_c \geq \xi_b$, the necessary condition as we discussed before, still hold.

and from 15% to 19% (a 27% increase), respectively. However, when the fuel cost exceeds 12 min (of time equivalent), both non-carpooling and carpooling commuters simply shift to transit. At this point, no solo-driving commuters change to carpool any more. This implies that the rise of gas prices may entice commuters to carpool in the first place, but as the price increases further, auto commuters who drive will eventually shift to the transit, rather than carpooling.

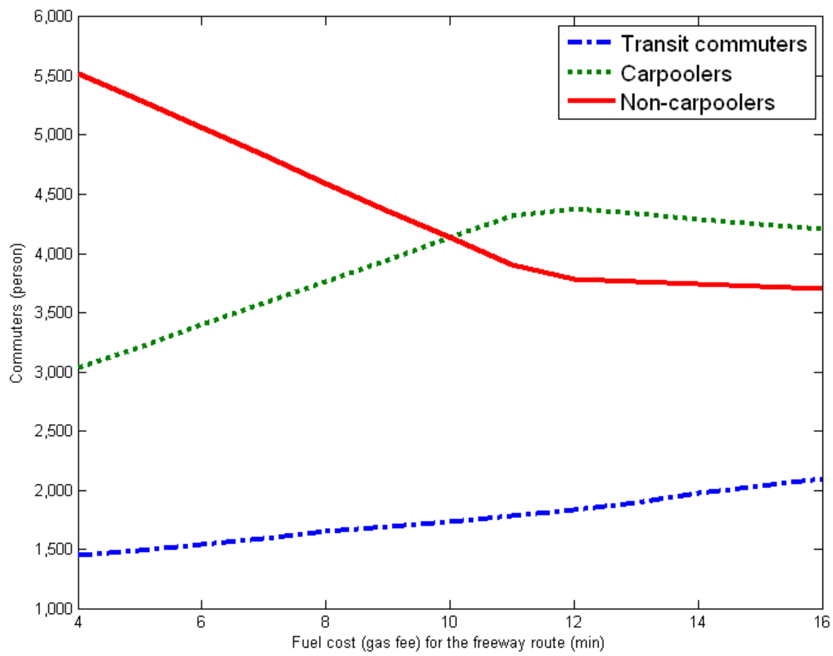


Figure 4.9. Changes in passenger flow of three travel modes with respect to fuel cost ξ_b (of time equivalent)

4.5.6 The influence of bottleneck capacity on network travel cost

To reduce the congestion delay in the network, we may attempt to improve the roadway facilities, such as building extra lanes for the freeway or arterial roads. We therefore investigate the changes in total travel cost with respect to the bottleneck capacity of the freeway and AR. The expansion of the bottleneck capacity of the freeway almost linearly decrease the travel cost of the whole network and the auto network. For a sufficiently high passenger demand ($N = 12000$) that congested the network while maintaining the travel advantage of the HOV lanes, changing the bottleneck capacity of the freeway from

30 veh/min to 50 veh/min, will reduce the network cost TC by 81000 (min·veh) (14%) and the auto network TC by 18000 (min·veh) (7%), while the corresponding numbers are 22000 (min·veh) (3.5%) and 1500 (min·veh) (1%) in a mildly congested network with carpool facilities offering no travel advantage($N = 4000$). That is, the capacity expansion on the freeway is more beneficial under heavy travel demand (heavier congestion). Expanding the bottleneck capacity on the arterial route, however, has much less impact on the reduction of total travel cost, because the number of commuters travel on it is considerably lower than on the freeway. On the other hand, the improved traffic conditions on the roads through expanding their capacities attract more commuters to travel by private auto. The reduction of the travel cost (of time equivalent) in the auto network TC is, therefore, not as pronounced as that in the total network.

4.5.7 The influence of a flat freeway toll on network travel cost

Let ξ_b represent the sum of a fixed fuel cost and a uniform freeway toll, u, which is anonymous to all the commuters, on the vehicle basis. Therefore, from the point view of social welfare, the total travel cost in the total network and auto network, defined previously as Equation 4.18 and 4.19, becomes:

$$TC = C_1 N_1 + C_{2b} N_{2b} + C_{3b} N_{3b} + C_{2c} N_{2c} + C_{3c} N_{3c} - u N_b \tag{4.51}$$

$$TC_{auto} = C_{2b} N_{2b} + C_{3b} N_{3b} + C_{2c} N_{2c} + C_{3c} N_{3c} - u N_b \tag{4.52}$$

because toll revenues will to some extent be returned to transportation systems by various redistribution schemes.

Assuming the fixed fuel cost is 10 min (of time equivalent), and the flat toll changes from 0 min to 10 min(of time equivalent), we plot the network travel cost (of time equivalent) with respect to the toll, as shown in Figure 4.10. The uniform toll applied to all the vehicles is unable to eliminate the queuing delay, but does reduce the total travel cost (exclusive of toll). Since carpooling can reduce the per-person cost by sharing this toll burden, this may attract some solo-drivers to carpool. With such a cost advantage, carpoolers are more likely to use the freeway over the AR, than the solo drivers do. These mode and route shifts under the flat toll lead to less total travel cost.

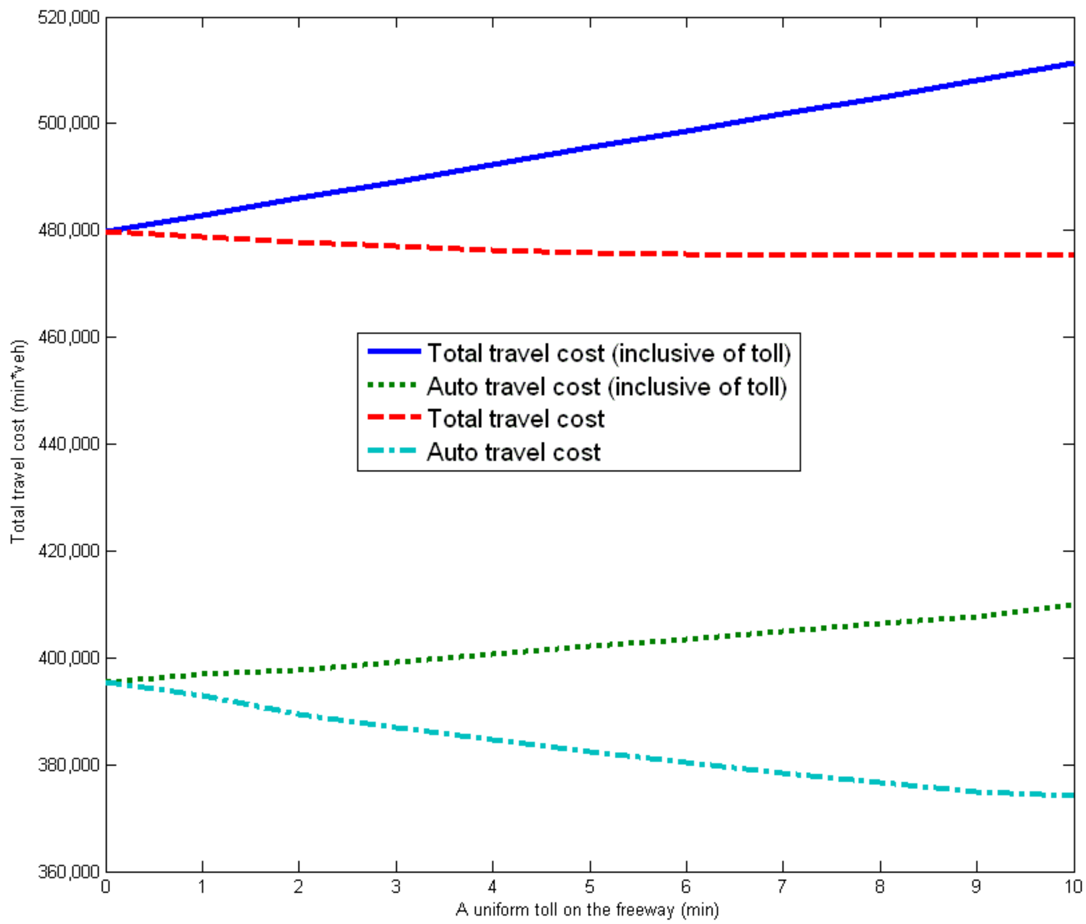


Figure 4.10. Changes in network travel cost (of time equivalent) with respect to the toll ($N = 10000$)

4.6 Eliminating Freeway Queuing With a Time-varying Toll

In this section we derive a time-varying freeway toll to eliminate all the queuing delay on the freeway. To do this, the definitions of $C_{2b}(t)$ and $C_{3b}(t)$ are changed accordingly.

$$C_{2b}(t) = \max[\gamma(t - t^*), \beta(t^* - t)] + \frac{u(t)}{m} + \frac{\xi_b}{m} + \Delta \tag{4.53}$$

$$C_{3b}(t) = \max[\gamma(t - t^*), \beta(t^* - t)] + \zeta(t) + \xi_b \tag{4.54}$$

where $\zeta(t)$ is the toll charged at time t (of time equivalent).

1)If $N_{2b} \geq N_{3b} m\phi$, then carpool offers no travel advantage. The time-varying toll, $\zeta(t)$, should be imposed uniformly for all the automobiles. Based on the UE conditions, carpooling commuters will choose to travel in the middle of the rush hour, while non-carpooling commuters at the beginning and end of the rush hour Arnott et al. (1988)[2]. We readily obtain that:

$$C_{2b}(t) = \delta\frac{N_{3b}}{m(1+\phi)s_f} + \delta\frac{N_{2b}}{m(1+\phi)s_f} + \frac{\xi_b}{m} + \Delta \tag{4.55}$$

$$C_{3b}(t) = \delta\frac{N_{3b}}{(1+\phi)s_f} + \delta\frac{N_{2b}}{m(1+\phi)s_f} + \xi_b \tag{4.56}$$

$$\zeta(t) = \begin{cases} \beta(t-t_0) & t \in [t_0, t_{12}] \\ \delta\frac{N_{3b}}{s_f} + m\beta(t-t_{12}) & t \in [t_{12}, t^*] \\ \delta\frac{N_{3b}}{s_f} + m\gamma(t_{21}-t) & t \in [t^*, t_{21}] \\ \gamma(t_1 - t) & t \in [t_{21}, t_1] \end{cases} \tag{4.57}$$

where $t_0 = t^* - \delta N_b/\beta s_f$, $t_1 = t^* + \delta N_b/\beta s_f$, $t_{12} = t^* - \delta N_{2b}/m\beta s_f$, $t_{21} = t^* + \delta N_{2b}/m\beta s_f$.

2)If $N_{2b} < N_{3b}m\phi$, carpool offers a travel advantage. We are able to impose time-varying tolls, $u_3(t)$ on the GP lanes, and $u_2(t)$ on the HOV lanes respectively.

$$\zeta_i(t) = \begin{cases} \frac{\beta}{1-\beta}(t - t_{i,1}) & t \in [t_{i,1}, t^*] \\ \frac{\gamma}{1+\gamma}(t_{i,2} - t) & t \in [t^*, t_{i,2}] \end{cases} \tag{4.58}$$

where $i = 2$ or 3. $t_{2,1} = t^* - \delta N_{2b}/m\beta\phi s_f$, $t_{2,2} = t^* + \delta N_{2b}/m\gamma\phi s_f$, $t_{3,1} = t^* - \delta N_{3b}/\beta s_f$, $t_{3,2} = t^* + \delta N_{3b}/\gamma s_f$.

Following the same solution procedure proposed in this chapter, we can obtain the market share of all the three modes, as well as network performances after applying the time-varying toll.

4.7 Summary

In this chapter, we studied the morning commute problem with three modes, carpooling, transit and driving-alone, in a network with two auto routes (a freeway and an arterial road) and one dedicated transit route. We simultaneously established equilibrium within each mode and the mode split through applying a nested logit model. Four types of solutions were identified and discussed, and a solution procedure to obtain them

[2]Refer to Arnott et al. (1988) for the case of multiple groups with identical β/γ but different α and β

was also proposed. Various factors, such as transit fares, fuel cost, a flat toll on the freeway and HOV lane capacity, were examined with respect to their effects on network performance, route choices and mode shares.

We found that carpoolers always choose the freeway when carpool offers suffi-cient travel cost savings over solo-driving. When carpoolers no longer save travel cost by using HOV lanes, they are more likely to use the arterial route than solo-drivers. It was also found that when a flat toll is applied, more carpoolers would use the freeway than solo-drivers, because they can share the toll, which reduces their travel cost.

When the capacity of the HOV lane is expanded, hence the overall capacity of the freeway, auto traffic volume would rise significantly while transit ridership would decrease when traffic demand is sufficiently high. When the total capacity of the freeway is fixed but the share of the HOV capacity increases, transit ridership would increase slightly, and carpoolers shift from the arterial to the freeway, and solo-drivers from freeway to the arterial. In addition, expanding HOV capacity may not necessarily reduce the total travel cost on the network.

Our results indicated that the rise of gas price may entice auto commuters to car-pool in the first place. However, the further increase of gas price beyond a certain threshold would force some auto commuters to shift to the transit mode.

Chapter 5

Heterogeneity of Parking Choices: Managing Morning Commute Traffic with Parking

Downtown parking has been a challenging issue in urban transportation system. The price, availability and accessibility of parking spaces may considerably influence commuters' travel behavior. Therefore, parking could be used to manage travel demand and change travel patterns so as to mitigate traffic congestion. Compared to charging a roadway toll, using parking to manage traffic demand can be less controversial. The goal of this chapter is to model parking choices in the morning commute, and to investigate how parking would affect the travel patterns and the network performance in a linear city, and thereafter proposes optimal parking-based pricing schemes or parking regulations that can reduce both travel cost and traffic delay.

Unlike Arnott et al. (1991) where parking is centrally provided and continuously distributed, in this chapter we model morning commute parking from a different perspective. We assume a finite number of parking clusters (areas) in the CBD area, instead of pre-determined continuous distribution of parking spaces towards the CBD area. This setting has two features in addition to Arnott et al. (1991)'s model. First of all, the capacity and the accessibility (as measured by the access distance/time from the parking lots to the final destinations) of the parking lots become variables, rather than predetermined as in Arnott et al. (1991)'s paper. As a matter of fact, both factors can substantially change the commuting pattern and commuting cost. The capacity and access time may be abstracted to more general concepts. For example, outlying parking lots offer a low parking fee and provide shuttle bus service to attract commuters, the "access time" becomes the average travel time of the shuttle bus plus the actual walking time from the bus station to the office, and the capacity becomes the total parking spaces available serving commuters in the outlying parking area. In order to mitigate congestion, the parking lots may be constructed in desired locations with desired capacity such that the network performance of morning commute can be improved or optimized. Second, parking spaces in a CBD area can be clustered, because most commuters have an expected parking cost (inclusive of parking fee and access time) for each cluster and perceive all the parking spaces in that cluster indifferently. Overall, parking spaces within each cluster are assumed to be identical, that is, they cost the same amount to park and have the same access time from the lots to the offices.

We model the parking market in two cases. One is that all the parking lots are city-owned or publicly owned, and the regulatory agency has full control over the parking fee, parking capacity and access time of those parking lots and aims to maximize the social welfare and/or minimize traffic congestion. We show how adjusting the parking fee, capacity and access time can enlarge the social welfare and mitigate congestion. From this perspective, we are able to evaluate the parking as a manner of traffic management compared to other manners (such as a roadway toll) in terms of effectiveness. Also, the best possible settings of parking can serve as the benchmark of cases where a regulatory agency has certain influence on a competitive parking market. The other case is that we assume all the parking lots are privately owned and there exists a parking market where each private parking operator determines the parking fee, parking capacity and accessibility to compete with others. Since such parking provision may not produce the most desirable market outcome in terms of system performance or traffic congestion, we also consider several market regulations and study their effects on the travelers' travel cost and operators' profit/cost in the morning commute. A parking model is first proposed below and then used to discuss both cases.

5.1 The Parking Model

In this section, we first define the problem setup and parameters to be used in this chapter, and introduce the basic assumptions adopted in our study.

As in the classical morning commute problem, in this dissertation we consider a simplified network as depicted in Figure 5.1. It consists of a major highway and multiple clusters of parking lots, connected by local streets. For simplicity we again assume the travel times on the local streets are zero. Moreover, we limit our investigation to the case of two parking clusters, with each cluster represents one or multiple parking lots in either the central area of the city (the central cluster) or a peripheral location of the city (the peripheral cluster). The central cluster usually charges a higher fee but has shorter access time to one's office, compared with the peripheral cluster.

We assume that parking spaces within each cluster are identical, that is, they cost the same amount to park and have the same access time from the lots to the offices in the CBD. This assumption may be reasonable because, 1) from the traveler point of view, most commuters have an expected parking cost (inclusive of parking fee and access time) when they determine a parking area prior to their trips. Although the actual parking fee or access time may vary slightly dependent on the space chosen within an area (cluster), commuters may perceive all the parking spaces in that cluster indifferently. Moreover, the access times among parking spaces within a cluster may vary, but the differences are usually small compared with the overall commuting time; 2) If all the parking lots are owned by private operators, from the parking operator side, even though parking spaces within each cluster may be operated by a number of private companies, the parking prices of those parking lots locating at approximately the same distance from the offices are likely to be the same. This is because those private operators within one cluster offer similar services, so the competition among them would equilibrate their prices. Moreover, they are more likely to cooperate to compete with parking lots located at other clusters. Based on this assumption, if parking market is competitive, then it can be viewed as an "oligopolistic" market where the union of parking operators in the central cluster competes with the union of those in the peripheral cluster.

Here, parking cluster 1 and 2 can be two real parking lots if the targeted offices are within a fairly small area. They could also be two areas of parking spaces if we consider a CBD area where the parking fee and access time within each cluster (zone) is indistinguishable.

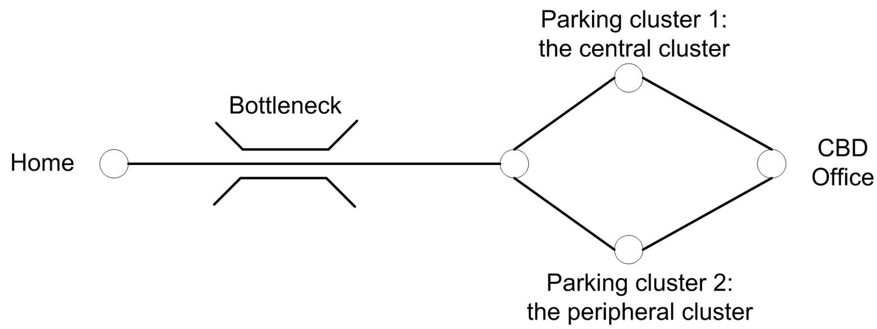


Figure 5.1. A simplified network with a choice of two parking clusters

In the morning rush hour, a total demand of N commuters heading for the CBD area (their offices) first go though a bottleneck, then choose to park his vehicle in either the central parking cluster or the peripheral parking cluster, and finally walk (or use other modes, e.g., take the parking shuttle) to their offices. We use K_1 and K_2 (all in vehicle units) to represent the **effective capacity** of the central and peripheral clusters, respectively, i.e. the number of parking spaces used by the travelers. p_1 and p_2 denote the parking fees, and l_1 and l_2 the access times of the respective parking clusters ($l_1 < l_2$). Here, access time measures the accessibility of parking spaces and may not be walking time alone. For example, the farther parking lots offer a lower charge and provide shuttle bus service to attract commuters, the access time in this case consists of the average travel time of the shuttle bus plus the actual time from the bus station to the office on foot. In addition, the cost of providing a parking space per day in the two clusters is denoted by a_1 and a_2, respectively. It is assumed that $a_1 < p_1, a_2 < p_2$, i.e., both parking clusters are profitable. Since $l_1 < l_2$, it is reasonable to assume that $p_1 \geq p_2$ (that is, the parking fee inside of the CBD is usually no less than in the peripheral area) and $a_1 > a_2$ (that is, it is cheaper to provide a parking space in the peripheral than inside of the CBD).

Adding the costs associated with parking to a commuter's travel cost, a commuter departing at time t and choosing the parking cluster i has the following generalized travel cost,

$$C_i(t) = \alpha w(t) + \max\{\beta(t^* - t - w(t) - l_i), \gamma(t + w(t) + l_i - t^*)\} + p_i + \lambda l_i \tag{5.1}$$

where $i = 1, 2$ and λ is the equivalent monetary cost of one unit of access time. We assume $\beta < \lambda$, which is consistent with Arnott et al. (1991) and supported by empirical evidence. Although in reality commuters may value travel time, schedule delay and access time differently, we do not differentiate the commuters in terms of their values of time. Their work starting time t^* is also assumed to be identical for all commuters.

The dynamic user equilibrium (UE) that Vickrey (1969) and Arnott et al. (1990) defined is actually a day-to-day equilibrium. In other words, in a typical morning commute problem where the bottleneck capacity is given, all commuters are aware of the traffic conditions after sufficient experiences of the commute, and the eventual travel patterns (i.e., commuters' choices of departure times) are such that their generalized travel costs are the same after their day-to-day adjustments. Following this definition of UE, we define a day-to-day user equilibrium incorporating parking choices as follows.

Definition 5.1. Day-to-day User Equilibrium. *Given the bottleneck capacity, parking facilities and parking pricing, the commuters in morning commute achieve a day-to-day user equilibrium if, 1) all commuters are aware of the traffic conditions, parking facilities and parking pricing after a sufficiently long time and they choose their departure times and parking spots such that their generalized travel costs are the same after day-to-day adjustments. 2) No commuter can unilaterally change his parking choice or his departure time to reduce his generalized travel cost.*

Whenever the bottleneck capacity, the parking fee, location or capacity are changed, we assume that a new day-to-day user equilibrium will eventually be reached after a sufficiently long adjustment period. Later on, we derive the travel patterns and network total travel cost under such day-to-day equilibria.

In addition, we use total cost and network queuing delay to measure the performance of the system. If all the parking lots are publicly owned, then a regulator is assumed in charge of managing all parking spaces, and has full control over the parking locations, fees, available parking spaces. We assume that the goal of the regulator is to maximize the social welfare and mitigate traffic congestion (in terms of queuing delay) to the most degree. The social welfare can be represented by the social gains minus travelers' actual travel cost. The social gains are the revenue collected from the parking fee, and can be re-distributed to the public in some way. A traveler's perceived travel cost consists of queuing delay, schedule delay, access time to the office and the parking fee he pays. Therefore, maximizing the social welfare is equivalent to minimizing the total travel cost (exclusive of parking fee) which reads

$$\min TC = NC - K_1 p_1 - K_2 p_2 \tag{5.2}$$

where C is the identical generalized travel cost of all travelers under the day-to-day UE, given by Equation 5.1.

One the other hand, if parking lots are owned by private operators, we define the total travel cost that is slightly different from the case with publicly owned parking. The total commuter travel cost (TCC), represented by the total queuing delay and schedule delay of all travelers, plus the total parking fees paid by those travelers, is given by

$$TCC = N \times C \tag{5.3}$$

The total system cost (TSC) is the total queuing delay and schedule delay of all travelers, plus the total investment cost of those private operators. Therefore,

$$\begin{aligned} TSC &= (TCC - p_1 K_1 - p_2 K_2) + (a_1 K_1 + a_2 K_2) \\ &= TCC - ((p_1 - a_1) K_1 + (p_2 - a_2) K_2) \end{aligned} \tag{5.4}$$

where the second term of the RHS is exactly the profits of private operators.

5.2 Parking Location Preference

For a given set of parking location and charge, commuters would adjust their departure times accordingly based on their preference over parking locations or vice versa. Overall, there are five types of parking location preference. By introducing the following composite "prices",

$$v_1 = p_1 + \lambda l_1 - \beta l_1$$

$$v_2 = p_2 + \lambda l_2 - \beta l_2$$

$$u_1 = p_1 + \lambda l_1 + \gamma l_1$$

$$u_2 = p_2 + \lambda l_2 + \gamma l_2$$

the generalized travel cost of a traveler choosing the central parking cluster becomes,

$$C_1(t) = \begin{cases} \alpha w(t) + \beta\left(t^* - t - w(t)\right) + v_1 & \text{if early arrival} \\ \alpha w(t) + \gamma\left(t + w(t) - t^*\right) + u_1 & \text{if late arrival} \end{cases} \tag{5.5}$$

and the generalized travel cost of a traveler choosing the peripheral parking cluster becomes,

$$C_2(t) = \begin{cases} \alpha w(t) + \beta\left(t^* - t - w(t)\right) + v_2 & \text{if early arrival} \\ \alpha w(t) + \gamma\left(t + w(t) - t^*\right) + u_2 & \text{if late arrival} \end{cases} \tag{5.6}$$

We can characterize five types of parking preference as follows,

1. **Strongly outward**. If $v_2 > v_1$ i.e. $p_1 - p_2 < (\lambda - \beta)(l_2 - l_1)$, which ensures $u_2 > u_1$, then $C_1(t) < C_2(t)$ in both early arrival and late arrival. In this case, commuters will prefer the central parking cluster (i.e. cluster 1) in both early arrival and late arrival. The peripheral one (i.e. cluster 2) will not be used unless the central cluster is used up. Because commuters choose to use the central cluster first and then the peripheral one, parking cluster 1 is **strongly preferred** and we call this type of parking preference **strongly outward** parking. This usually occurs when the peripheral parking cluster is not sufficiently competitive, possibly because either its parking fee is not sufficiently low or it is unacceptably inconvenient to the office. In particular, $p_1 = p_2$, where two parking clusters charge the same parking fee, and thus the central cluster is always strongly preferred.

2. **Weakly outward**. If $v_2 = v_1$ i.e. $p_1 - p_2 = (\lambda - \beta)(l_2 - l_1)$, which ensures $u_2 > u_1$, then $C_1(t) = C_2(t)$ in early arrival but $C_1(t) < C_2(t)$ in late arrival. In this case, the parking fees and access times of the two parking clusters are such that commuters are indifferent to both clusters in early arrival. Since late arrival is weighed more than the early arrival ($\gamma > \beta$), the central cluster offers an advantage in generalized travel cost over the peripheral one in late arrival. Thus the central parking cluster is overall **weakly preferred**. This type of parking preference is defined as **weakly outward** parking. In early arrival where travelers are indifferent to the two parking clusters, we assume using either parking cluster is equally likely for any traveler.

3. **Strongly inward**. If $u_2 < u_1$ i.e. $p_1 - p_2 > (\lambda + \gamma)(l_2 - l_1)$, which ensures $v_2 < v_1$, then $C_1(t) > C_2(t)$ in both early arrival and late arrival. In this case, commuters will prefer the peripheral parking cluster in both early arrival and late arrival. The central cluster will not be used unless the peripheral one is used up. Because commuters choose to use the farther parking cluster first and then the closer one, the peripheral cluster is **strongly preferred** and we call this type of parking preference **strongly inward** parking. This usually occurs when either the central parking charges unacceptably high parking fee or the access time of the peripheral cluster is reasonably close to the central one.

4. **Weakly inward**. If $u_2 = u_1$ i.e. $p_1 - p_2 = (\lambda + \gamma)(l_2 - l_1)$, which ensures $v_2 < v_1$, then $C_1(t) > C_2(t)$ in early arrival and $C_1(t) = C_2(t)$ in late arrival. In this case, the parking fees and access times of the two parking clusters are such that commuters are indifferent to both clusters in late arrival. Since early arrival is weighed less than the late arrival ($\beta < \gamma$), the peripheral parking cluster offers an advantage in generalized travel cost over the central one in early arrival. Thus, the peripheral cluster is overall **weakly preferred**. This type of parking preference is defined as **weakly inward** parking. In late arrival where travelers are indifferent to the two parking clusters, we assume using either parking cluster is equally likely for any traveler.

5. **Hybrid**. If $v_2 < v_1$ and $u_2 > u_1$, i.e. $(\lambda - \beta)(l_2 - l_1) < p_1 - p_2 < (\lambda + \gamma)(l_2 - l_1)$, then $C_1(t) > C_2(t)$ in early arrival but $C_1(t) < C_2(t)$ in late arrival. In this case, commuters will prefer the peripheral cluster in early arrival and the central one in late arrival. We call this type of parking preference **hybrid**, since it is a hybrid of inward parking and outward parking over the entire commuting period. Hybrid parking may occur when the farther parking cluster offers a sufficient advantage for commuters in early arrival, but such an advantage is yet not sufficient in attracting commuters in late arrival (commuters' penalty in late arrival is far larger than in early arrival). In the case of late arrival, travelers would rather pay an additional but reasonable parking fee in the closer cluster than being subject to a high late-arrival penalty by using the farther cluster.

The preference on a parking cluster is essentially determined by parking fees and access times of both parking clusters, regardless of availability of parking spaces. However, parking capacity plays an important role in determining a certain travel pattern and the market shares between the two parking clusters. For example, in strongly inward or strongly outward parking, if sufficient parking spaces ($> N$) in the preferred parking cluster are provided, then the other parking cluster will never be used. If the parking spaces in the preferred parking cluster are limited and cannot accommodate all the travelers, then the travel pattern under user equilibrium is such that travelers who use the preferred parking cluster must depart home earlier than those who use the other. Therefore, three factors of parking, fee, access time and capacity, altogether determine the travel pattern, and each of them influences travelers' departure times and thus system performance.

5.3 Travel Profiles and Their Properties Under User Equilibrium

5.3.1 Overview of parking profiles

By applying the user equilibrium principle, we study all possible travel patterns for each parking location preference. There are in all 20 possible travel patterns, whose conditions of validity and total travel cost (TC) are shown in Table 5.1 [1], and departure and arrival patterns (travel profiles) are drawn in Appendix D. Though those profiles only hold under certain conditions with respect to three parking factors and vary case-by-case, their derivations follow the same logic which we demonstrate using an example in strongly outward parking in Appendix E. Due to space limitation, the derivation of other profiles are omitted. However, there are two possible profiles in hybrid parking worth further discussion. This is because not only do the solutions of those two profiles interestingly differ from those of other profiles, but also the two profiles are central to solve the optimal parking settings in later sections. In the following subsection, we solve

[1] Profile 19 and 20 are the cases where one of the two clusters has zero space while the other accommodates all travelers. These two extreme cases resemble the typical morning commute without parking choices, and can exist for all five parking preferences. Therefore, we do not include them in Table 5.1.

for two possible profiles in hybrid parking.

Each profile is achieved under a certain condition with respect to parking capacity, fee and access time, which is expressed by an inequality in terms of capacity bounds in the second column of Table 5.1. Those conditions can also be rewritten as an inequality in terms of bounds of either fee or access time.

Note that the effective capacity of one parking cluster equals to its actual capacity only if this cluster is preferred. In the table we only use the capacity of the preferred location to express the profile's condition as well as in the TC formula. For example, in strongly outward parking, the closer cluster is always preferred. In other words, the profiles and TC-s under strongly outward parking are dependent on the capacity of the closer cluster (i.e. cluster 1), while the other one does not take effects as long as it can provide sufficient spaces for the total demand[2]. Therefore, in strongly outward parking, we use K_1 to express all the inequality conditions and TC-s, although the effective capacity of the farther cluster, $K_2 = N - K_1$, can do the same job. In each profile of hybrid, weakly outward and weakly inward parking, both K_1 and K_2 may be preferred under certain conditions. In those cases, the conditions and TCs are expressed in terms of either K_1 or K_2, whichever effectively determines the profile.

5.3.2 Two interesting profiles in hybrid parking

Now we solve for two interesting profiles in hybrid parking. In hybrid parking, a traveler who departs at such a time that he arrives earlier than t^* using either of the two clusters prefers the farther parking cluster due to $v_1 > v_2$, while a traveler who departs at such a time that he arrives later than t^* using either either of the two clusters prefers the closer one, due to $u_1 < u_2$. However, some travelers may depart the bottleneck in such a time that they are subject to early arrival if choosing the closer cluster, and late arrival if choosing the farther cluster. We first analyze the parking preferences of those travelers.

Let y denote the duration from the arrival time to the office of the first traveler using

[2]We assume there are always enough parking spaces for all the commuters provided by the public regulator

Table 5.1. An overview: 20 travel profiles and TCs for five types of parking preference

Parking preference	Condition	Total travel cost (TC)	Profile
Outward (strong)	$0 \le K_1 \le \frac{s(v_2-v_1)}{\beta}$	$\frac{N\beta\gamma}{s}\frac{N-K_1}{\beta+\gamma}+\lambda l_2 N - K_1(p_1-p_2)$	1
	$\frac{s(v_2-v_1)}{\beta} \le K_1 \le \frac{s(v_2-v_1)+N\gamma}{\beta+\gamma}$		2
	$\frac{s(v_2-v_1)+N\gamma}{\beta+\gamma} \le K_1 \le \frac{s(u_2-u_1)+N\gamma}{\beta+\gamma}$	$\frac{N\beta}{s}\frac{s(u_2-u_1)+N\gamma}{\beta+\gamma}+(N-K_1)(p_1-p_2)+\lambda N l_1$	3
	$N > K_1 \ge \frac{s(u_2-u_1)+N\gamma}{\beta+\gamma}$		4
Outward (weak)	$0 \le K_2 \le \frac{1}{2}\frac{s(u_1-u_2)+N\gamma}{\beta+\gamma}$	$\frac{N^2\beta\gamma}{s(\beta+\gamma)}+\lambda l_1 N + K_2(p_1-p_2)$	12
	$\frac{1}{2}\frac{s(u_1-u_2)+N\gamma}{\beta+\gamma} \le K_2 \le \frac{1}{2}\frac{N\gamma}{\beta+\gamma}$	$\frac{N\beta}{s}2K_2+\lambda l_2 N-(N-K_2)(p_1-p_2)$	13
	$\frac{u_2-u_1}{\beta+\gamma}s+\frac{1}{2}\frac{N\gamma}{\beta+\gamma} \le K_1 \le N-\frac{1}{2}\frac{N\gamma}{\beta+\gamma}$		14
	$\frac{1}{2}\frac{N\gamma}{\beta+\gamma} \le K_1 \le \frac{u_2-u_1}{\beta+\gamma}s+\frac{1}{2}\frac{N\gamma}{\beta+\gamma}$	$\frac{N^2\beta\gamma}{s(\beta+\gamma)}+\lambda l_2 N - K_1(p_1-p_2)$	15
	$0 \le K_1 \le \frac{1}{2}\frac{N\gamma}{\beta+\gamma}$		16
Inward (strong)	$0 \le K_2 \le \frac{s(v_1-v_2)}{\beta}$	$\frac{N\beta\gamma}{s}\frac{N-K_2}{\beta+\gamma}+\lambda l_1 N + K_2(p_1-p_2)$	5
	$\frac{s(v_1-v_2)}{\beta} \le K_2 \le \frac{s(u_1-u_2)+N\gamma}{\beta+\gamma}$		6
	$\frac{s(u_1-u_2)+N\gamma}{\beta+\gamma} \le K_2 \le \frac{s(v_1-v_2)+N\gamma}{\beta+\gamma}$	$\frac{N\beta}{s}\frac{s(u_1-u_2)+N\gamma}{\beta+\gamma}-(N-K_2)(p_1-p_2)+\lambda N l_2$	7
	$N > K_2 \ge \frac{s(v_1-v_2)+N\gamma}{\beta+\gamma}$		8
Inward (weak)	$0 \le K_2 \le \frac{s(v_1-v_2)}{\beta}$	$\frac{N\beta\gamma}{s}\frac{N-K_2}{\beta+\gamma}+\lambda l_1 N + K_2(p_1-p_2)$	5
	$\frac{s(v_1-v_2)}{\beta} \le K_2 \le \frac{N\gamma}{\beta+\gamma}$		6
	$\frac{N\gamma}{\beta+\gamma} \le K_2 \le \frac{N\gamma+s(v_1-v_2)}{\beta+\gamma}$	$\frac{N^2\beta\gamma}{s(\beta+\gamma)}-(N-K_2)(p_1-p_2)+\lambda N l_2$	7
	$\frac{N\gamma+s(v_1-v_2)}{\beta+\gamma} \le K_2 \le N-\frac{1}{2}\frac{N\beta-s(v_1-v_2)}{\beta+\gamma}$		17
	$0 \le K_1 \le \frac{1}{2}\frac{N\beta-s(v_1-v_2)}{\beta+\gamma}$	$\frac{N^2\beta\gamma}{s(\beta+\gamma)}-K_1(p_1-p_2)+\lambda N l_2$	18
Hybrid	$0 \le K_2 \le \frac{s(v_1-v_2)}{\beta}$	$\frac{N\beta\gamma}{s}\frac{N-K_2}{\beta+\gamma}+\lambda l_1 N + K_2(p_1-p_2)$	5
	$\frac{s(v_1-v_2)}{\beta} \le K_2 \le \frac{s(u_1-u_2)+N\gamma}{\beta+\gamma}$	$\frac{N\beta}{s}\frac{s(u_1-u_2)+N\gamma}{\beta+\gamma}-(N-K_2)(p_1-p_2)+\lambda N l_2$	6
	$\frac{s(u_1-u_2)+N\gamma}{\beta+\gamma} \le K_2 \le \frac{s(u_1-u_2)+N\gamma+s(v_1-v_2)}{\beta+\gamma}$		7
	$\frac{N\beta-s(v_1-v_2)}{\beta+\gamma} \le K_1 \le \frac{N\beta+s(u_2-u_1)-s(v_1-v_2)}{\beta+\gamma}$	$\frac{N\beta}{s}(N-K_1-\frac{s(v_1-v_2)}{\beta+\gamma})-K_1(p_1-p_2)+\lambda N l_2$	9
	$\frac{(u_2-u_1)s}{\beta+\gamma} \le K_1 \le \frac{N\beta-s(v_1-v_2)}{\beta+\gamma}$	$\frac{N^2\beta\gamma}{s(\beta+\gamma)}-K_1(p_1-p_2)+\lambda N l_2$	10
	$0 \le K_1 \le \frac{(u_2-u_1)s}{\beta+\gamma}$		11

Note: dependent on the magnitude of parameters, some profiles may not occur under certain conditions. For example, Profile 2 may not occur if $\frac{s(v_2-v_1)}{\beta} > \frac{s(v_2-v_1)+N\gamma}{\beta+\gamma}$

cluster 1 through t^*. The travel cost of such a traveler is,

$$\alpha w + p_1 + \lambda l_1 + \beta y$$

where w is the travel delay. The duration from t^* through the arrival time to the office of the last traveler using cluster 2 is then $l_2 - l_1 - y$. $0 \leq y \leq l_2 - l_1$. The travel cost of this traveler is,

$$\alpha w + p_2 + \lambda l_2 + \gamma(l_2 - l_1 - y)$$

The critical value of y, $\bar{y}$, such that a traveler is indifferent to the two parking clusters is

$$\alpha w + p_1 + \lambda l_1 + \beta \bar{y} = \alpha w + p_2 + \lambda l_2 + \gamma(l_2 - l_1 - \bar{y})$$

which yields,

$$\bar{y} = \frac{u_2 - u_1}{\beta + \gamma}$$

Therefore, for those whose arrival time to the office is such that $\bar{y} < y \leq l_2 - l_1$, the farther cluster is preferred, and for those whose arrival time to the office is such that $0 \leq y < \bar{y}$, the closer cluster is preferred. This implies that there exists a transition period during which travelers' preference gradually changes from parking cluster 2 to cluster 1, with the elapse of departure time from the bottleneck (thus also the elapse of arrival time to the bottleneck due to first-in-first-out).

Figure 5.2(a) depicts one possible profile in hybrid parking where travelers choose the farther parking cluster until its spaces are used up, and thereafter they have to choose the closer one instead. The capacity of the farther cluster is such that its spaces are used up before travelers change their preference during the transition period. Therefore, $\bar{y} \leq y \leq l_2 - l_1$. During period CA, travelers still prefer the farther cluster and they are subject to late arrival, and thus the slope of the arrival curve to the bottleneck is $\frac{\alpha}{\alpha+\gamma}s$[3]. When travelers start to use the closer cluster, they arrive at their office earlier than t^*. Therefore, the slope of EF equals $\frac{\alpha}{\alpha-\beta}s$. After the traveler whose arrival time to the bottleneck is t_F, travelers arrive later than t^* again, and the slope of the arrival curve to the bottleneck changes to $\frac{\alpha}{\alpha+\gamma}s$ again.

[3] This is a direct application of the typical morning commute model (Vickrey 1969). Travelers who arrive earlier than t^* have the same travel costs (i.e. summation of queuing delay and schedule delay) as long as they arrive the bottleneck in the arrival rate of $\frac{\alpha}{\alpha-\beta}s$. Similarly, travelers who arrive later than t^* have the same travel costs as long as they arrive the bottleneck in the arrival rate of $\frac{\alpha}{\alpha+\gamma}s$.

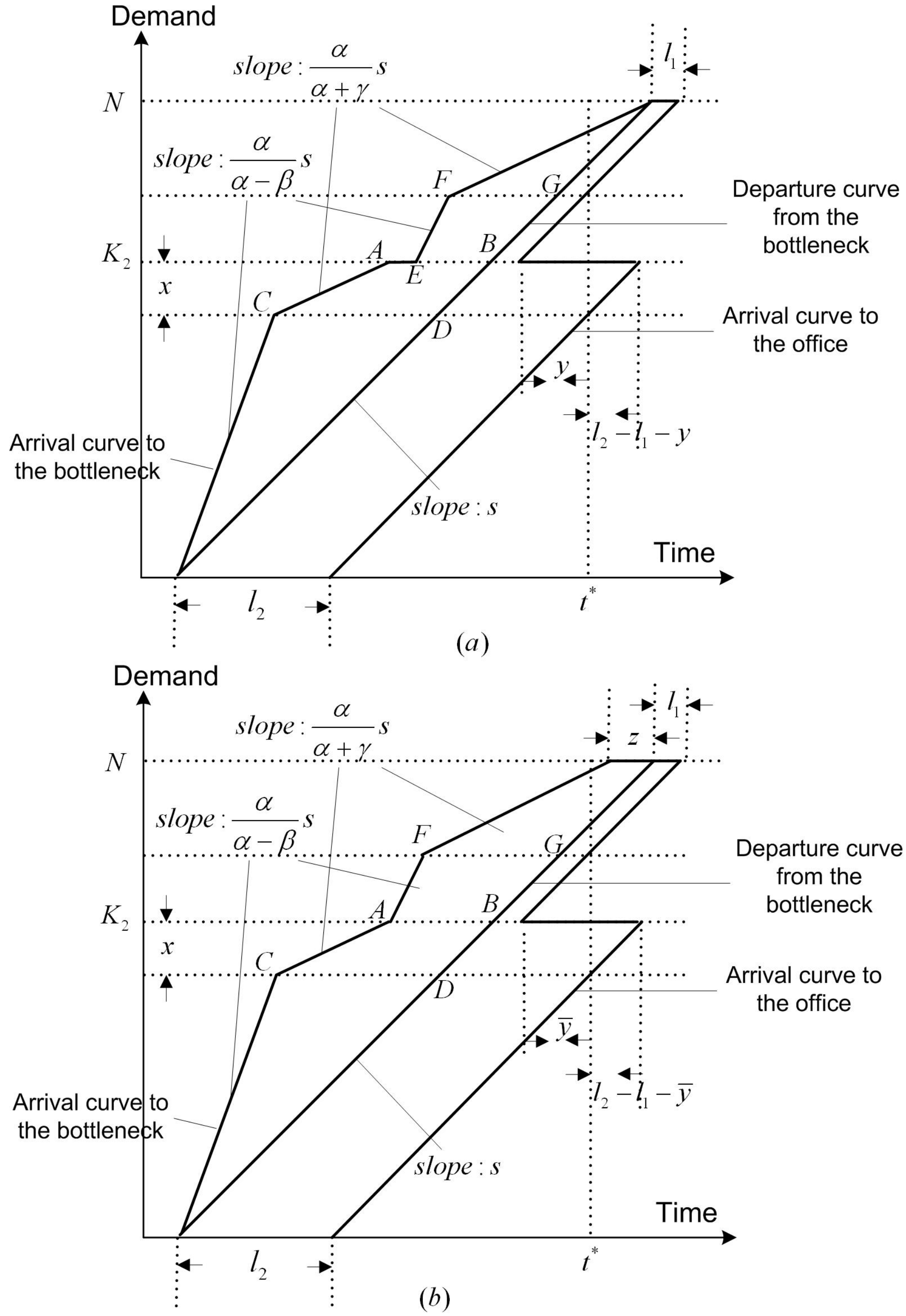


Figure 5.2. Two travel profiles in hybrid parking

Under user equilibrium, the last traveler using parking cluster 2 has the same generalized travel cost as the first traveler using parking cluster 1. Thus,

$$C_1(t_E) - C_2(t_A) = 0 \tag{5.7}$$

where t_E and t_A denote the departure times of the first traveler using parking cluster 1 and the last traveler using parking cluster 2, respectively. Substitute $C_1(t_E)$ and $C_2(t_A)$ by Equation 5.5 and 5.6 into Equation 5.7,

$$\alpha(EB - AB) + [(p_1 + \lambda l_1) - (p_2 + \lambda l_2)] + [\beta y - \gamma(l_2 - l_1 - y)] = 0 \tag{5.8}$$

where $\alpha(EB - AB)$, $(p_1 + \lambda l_1) - (p_2 + \lambda l_2)$ and $\beta y - \gamma(l_2 - l_1 - y)$ represent the difference in queuing delay cost, parking-related cost and schedule delay cost between the two marginal travelers, respectively.

On the other hand, the geometry of the profile yields,

$$EB = FG - y\frac{\beta}{\alpha} \tag{5.9}$$

$$FG = \frac{N - K_2 - ys}{s}\frac{\gamma}{\alpha} \tag{5.10}$$

$$AB = CD - \frac{N}{s}\frac{\gamma}{\alpha} \tag{5.11}$$

$$CD = \frac{K_2 - x}{s}\frac{\beta}{\alpha} \tag{5.12}$$

$$\frac{x}{s} = l_2 - l_1 - y \tag{5.13}$$

Solving for y by Equations 5.8, 5.9, 5.10, 5.11, 5.12 and 5.13, yields,

$$s(l_2 - l_1 - y) = K_2 - \frac{s(u_1 - u_2) + N\gamma}{\beta + \gamma}$$

Because $\bar{y} \leq y \leq l_2 - l_1$, we have

$$\frac{s(u_1 - u_2) + N\gamma}{\beta + \gamma} \leq K_2 \leq \frac{s(u_1 - u_2) + N\gamma + s(v_1 - v_2)}{\beta + \gamma} \tag{5.14}$$

One can also show that

$$EB = \frac{N - K_2}{s}\frac{\gamma}{\alpha} > 0 \tag{5.15}$$

given the range of K_2 expressed by Equation 5.14. This ensures the existence of the profile.

Since all commuters are subject to the same generalized travel cost, the generalized travel cost can be computed in terms of an arbitrary traveler, e.g. the traveler departing at time t_C with punctual arrival,

$$\alpha CD + (p_2 + \lambda l_2)$$

Thus, the total travel cost excluding parking revenue is

$$\begin{aligned} TC &= (\alpha CD + p_2 + \lambda l_2)N - K_2 p_2 - (N - K_2)p_1 \\ &= \frac{N\beta}{s}\frac{s(u_1 - u_2) + N\gamma}{\beta + \gamma} - (N - K_2)(p_1 - p_2) + \lambda N l_2 \end{aligned} \quad (5.16)$$

If K_2 is larger than the right-hand side of Inequality 5.14, the travel profile will then look like Figure 5.2(b). The profile in Figure 5.2(b) considers a similar case as in 5.2(a) where the farther cluster is never used after travelers prefer the closer cluster, but the spaces in the farther cluster are used up exactly by the traveler who treats both clusters indifferently. On the other hand, for this profile to hold, K_2 must also have an upper bound. This is because if K_2 is sufficiently large and the effective K_1 is relatively small, then the closer cluster has insufficient capacity to accommodate all travelers who depart after travelers' parking preference is changed to cluster 1.

In this profile, the spaces in the farther cluster are used up exactly at the departure time from the bottleneck, $t^* - \bar{y} - l_1$, when travelers are indifferent to both clusters but prefer the closer one thereafter. As shown in Figure 5.2(b), the last traveler using the farther cluster arrives at his office at $t^* + (l_2 - l_1 - \bar{y})$ and the first traveler using the closer cluster arrives his office at $t^* - \bar{y}$. Because the two marginal travelers are indifferent to both clusters, they have the same queueing delay under user equilibrium. Thus, in the profile depicted in Figure 5.2(b), there is no such line of AE as in Figure 5.2(a) during which no travelers depart from their homes.

We next show that under the user equilibrium of this profile, there exists a queuing delay (with the duration of z) for the last traveler. Under the day-to-day user equilibrium, the last traveler will not wait to depart until the queue vanishes or dissipates, because if he does, a traveler choosing the farther cluster can reduce his travel cost by departing a bit earlier than the "last traveler" and using the closer one instead. In that case, the "last" traveler has to use the remaining space in the farther cluster that was originally used by the traveler who switched, and thus he will have higher travel cost. Therefore, the last traveler will not risk his chances by departing at the end (or anywhere in the middle) of the queue. Rather, he will depart following the second last traveler to ensure his parking space in cluster 1 will not be occupied by another traveler. Such a queuing delay of the last traveler cannot occur in a typical morning commute problem, because if a traveler is not restricted by the parking capacity of cluster 1, then he does indeed can reduce his travel cost by departing at the end of the queue without any risk of being affected by any other travelers. Note that although the equilibrium with the existence of a queuing delay of the last traveler satisfies our equilibrium definition in Section 2, it is not a Nash Equilibrium because the last traveler does not take the departure times of other travelers as given. Rather, he takes the queuing time as given, and anticipates possible changes in other travelers' departure times.

On the other hand, the queuing delay z is also upper bounded. Under user equilibrium, the travel cost of a traveler using the farther cluster equals to that of the last traveler using the closer cluster, which is $\alpha z + u_1 + \gamma t'$ where t' is the last traveler's departure time from the bottleneck. If $\alpha z + u_1 > u_2$, then for a traveler using the farther cluster, by switching to depart at the end of the queue (i.e. t') and still using the farther one, his travel cost becomes $\gamma t' + u_2$. This traveler can reduce his travel cost by changing this departure time, without any risk of being affected by another traveler. Consequently, $0 \leq z \leq \frac{u_2 - u_1}{\alpha}$.

In addition, K_1 effectively determines this profile. On one hand, keeping K_1 constant, increasing the actual capacity of the farther cluster ($> K_2$) will not change the profile, because all travelers departing after t_A prefer the closer parking cluster and they will not use the farther one unless spaces in cluster 1 are used up. On the other hand, if we enlarge the closer parking cluster by an additional parking space, then a traveler choosing the farther cluster can reduce his travel cost by following the "last" traveler (in the current profile) and using the additional space in the closer cluster, without affecting any other travelers. In this case, when the network achieves UE, eventually one traveler switches from the farther cluster to the closer one. Since K_1 effectively determines the profile, we express TC and capacity bounds in terms of K_1.

In the profile of Figure 5.2(b), we express z in terms of K_1, and letting $0 \leq z \leq \frac{u_2 - u_1}{\alpha}$

yields,

$$\frac{N\beta - s(v_1 - v_2)}{\beta + \gamma} \leq K_1 \leq \frac{N\beta + s(u_2 - u_1) - s(v_1 - v_2)}{\beta + \gamma} \tag{5.17}$$

The derivation of Inequality 5.17 follows a similar procedure as in the derivation of the profile depicted in Figure 5.2(b), and thus is omitted here to save space. The identical generalized travel cost can be obtained by computing the travel cost of an arbitrary traveler, e.g. the traveler departing at time t_C with punctual arrival,

$$\alpha CD + (p_2 + \lambda l_2)$$

Thus, the total travel cost exclusive of parking revenue is

$$\begin{aligned} TC &= (\alpha CD + p_2 + \lambda l_2)N - (N - K_1)p_2 - K_1 p_1 \\ &= \frac{N\beta}{s}(N - K_1 - \frac{s(v_1 - v_2)}{\beta + \gamma}) - K_1(p_1 - p_2) + \lambda N l_2 \end{aligned} \tag{5.18}$$

5.4 The Case of Parking Lots Operated By Public Agencies

5.4.1 The effects of accessibility, capacity and fee settings

Since we obtain all possible travel profiles with corresponding conditions in terms of parking fee, capacity and access time, we are able to examine the effects of all three factors on the system performance. More importantly, we show how each one, when fixing the other two, should be set to achieve the minimum total travel cost.

In the following subsections, we define $\Delta p = p_1 - p_2 \geq 0$, $\Delta l = l_2 - l_1 > 0$, and suppose l_1 and p_2 are fixed without loss of generality. The derivatives of total travel cost with respect to parking fee, capacity and access time are listed in Appendix D.

5.4.1.1 Parking capacity

The parking location preference, outward, inward or hybrid, is determined by the parking fee and access time. Parking capacity will not change the type of parking preference. However, once the parking fee and access time, and thus the type of parking preference, are determined, the changes in the capacity of both parking clusters will change the travel profiles within a certain type of parking preference.

In most profiles, if K_1 determines a profile, then we have $\partial TC/\partial K_1 < 0$, and if K_2 determines a profile, then we have $\partial TC/\partial K_2 > 0$. This indicates that in most profiles, enlarging the effective capacity of the closer parking cluster, or equivalently reducing the effective capacity of the farther parking cluster, generally reduces the total travel cost, regardless of the type of parking preference. However, there are two exceptions. One is when $K_2 \leq \frac{s(v_1-v_2)}{\beta}$ in inward parking and hybrid parking. In that case, enlarging the effective capacity of the farther cluster reduces the total travel cost if $\frac{N\beta\gamma}{s(\beta+\gamma)} > p_1 - p_2$. Intuitively, this is because when the farther parking cluster offers some advantage over the closer one, providing more parking spaces in the farther cluster can induce some commuters to switch from cluster 1 to 2 and reduce the total parking-related travel cost, with merely minor increase in queuing delay and/or schedule delay. This exception implies that enlarging the outer parking spaces may sometimes benefit the social welfare when the parking fee in the outer cluster is sufficiently low. The other exception is when $p_1 = p_2$, a special case of strongly outward parking. When both parking clusters charge the same parking fees, the network travel cost is independent of the capacity of the closer cluster when $K_2 > \frac{s(v_1-v_2)}{\beta}$. In this case, if we enlarge the capacity of the preferred parking cluster (i.e. cluster 1), the reduction in total parking-related travel cost is exactly offset by the increase in queuing delay and schedule delay. This implies that if the parking fee of the closer cluster and outer cluster are the same, building more spaces in closer cluster does not help reduce the total travel cost.

According to the derivatives listed in Appendix D, for any given parking fee and access time of both parking clusters, the desired optimal profile in terms of minimum TC is given as follows,

1. If parking fee and access time are such that parking are outward (strongly or weakly), the optimal profile occurs when the closer cluster can accommodate all travelers[4], i.e.,$K_{1,opt} = N, K_{2,opt} = 0$.

2. If parking fee and access time are such that parking preference are otherwise, the

[4]Though TC is not differentiable when $K_1 = N$ in strongly outward parking, it is easy to show that $TC_{K_1=N} = \frac{N^2\beta\gamma}{(\beta+\gamma)s} + \lambda l_1 N$ is less than TC of all other profiles

optimal profiles occur when[5]

$$K_{1,opt} = N, K_{2,opt} = 0, \text{if } \Delta p > \frac{N\beta\gamma}{s(\beta+\gamma)} \tag{5.19}$$

i.e. the closer cluster accommodates all travelers, or when

$$K_{2,opt} = \frac{s(v_1 - v_2)}{\beta} = \frac{s\left(\Delta p - (\lambda-\beta)\Delta l\right)}{\beta}, \text{if } \Delta p < \frac{N\beta\gamma}{s(\beta+\gamma)} \tag{5.20}$$

as shown in Figure 5.3 which is a special case of Profile 5 where the departure curve from the bottleneck is continuous (also a special case of Profile 6 where the first traveler using the closer parking cluster does not have queuing delay). In addition, the optimal profile is not unique when $\Delta p = \frac{N\beta\gamma}{s(\beta+\gamma)}$, which is depicted in Profile 5

$$0 \le K_{2,opt} \le \frac{s\left(\Delta p - (\lambda-\beta)\Delta l\right)}{\beta}, \text{if } \Delta p = \frac{N\beta\gamma}{s(\beta+\gamma)} \tag{5.21}$$

Equations 5.20 and 5.21 imply that when the central parking area charges an overly high fee or it does not offer too large an accessibility advantage over the outer area, it is clearly that enlarging the closer parking cluster may not be desirable in terms of the total travel cost. Unless the outer parking area is too far inconvenient, we should consider restricting the capacity of central parking, and meanwhile offer some incentives to induce more travelers to use the outer parking.

5.4.1.2 Parking fee

Now we turn to investigate how parking fee affects the travel profile and total travel cost when holding the parking capacity and access time constant. Weakly outward and weakly inward parking hold if and only if Δp takes a single real value, and TC is not differentiable with respect to p_1 (or Δp) for these two types of parking preference. Therefore, we only have $\partial TC/\partial \Delta p$ for strongly outward, strongly inward and hybrid parking choices in Appendix D. We first give a transition order of travel profiles with the increase of Δp (from zero to infinity), and discuss the effects of Δp on the TC. Based on the transition order, we show that the TC reaches minimum at certain profiles during the transition so that the optimal Δp can be solved.

[5]Though TC is not differentiable when $K_2 = N$ in strongly inward parking, it is easy to show that $TC_{K_1=N} = \frac{N^2\beta\gamma}{(\beta+\gamma)s} + \lambda l_1 N$ is less than $TC_{K_2=N} = \frac{N^2\beta\gamma}{(\beta+\gamma)s} + \lambda l_2 N$

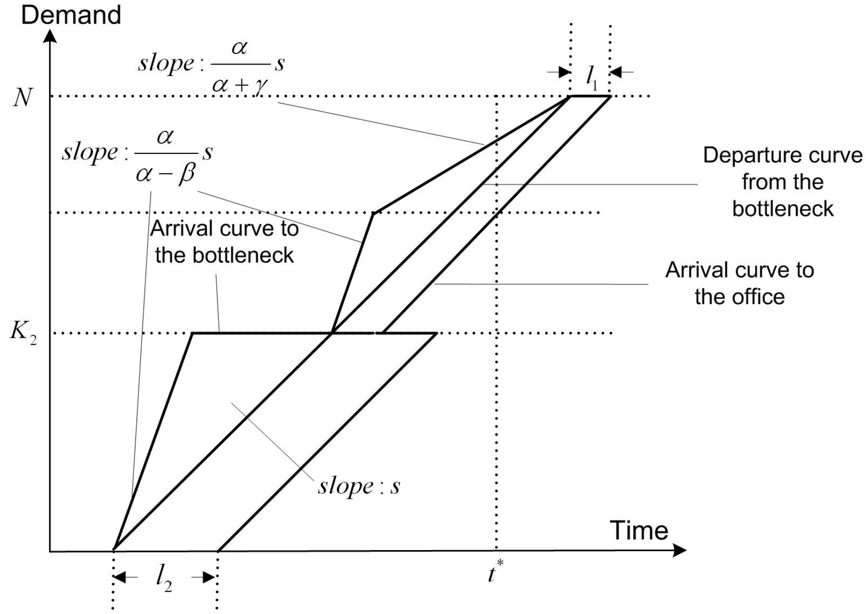


Figure 5.3. Optimal travel profile

A transition order of travel profiles with the increase of Δp

We first rewrite the conditions for all profiles in Table 5.1 in terms of Δp rather than K_1 or K_2. For example, the condition for Profile 1 becomes,

$$p_1 - p_2 \leq (\lambda - \beta)(l_2 - l_1) - \frac{K_1 \beta}{s}$$

Starting from $p_1 = p_2$, with the increase of p_1, outward parking gradually transitions to hybrid parking, and finally inward parking. By checking the bounds of $p_1 - p_2$ for each profile, we show that this transition process consists of several profiles in the following order: $1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow (one\ of\ 12 \sim 16) \rightarrow 11 \rightarrow 10 \rightarrow 9 \rightarrow 8 \rightarrow 7 \rightarrow 6 \rightarrow 5$. Note that Profiles 5, 6 and 7 may occur in both inward parking and hybrid parking. For instance, dependent on the magnitudes of parameters (such as total demand, parking capacities, access times), if we gradually increase Δp so as to induce hybrid parking to switch to inward parking, the critical Δp that changes the parking preference may be such that the condition of one profile of Profiles 7, 6 or 5 is satisfied, i.e. inward parking starts at one of those profiles. If inward parking starts at Profile 6, then Profile 7 in the profile order must occur in hybrid parking and Profile 5 must occur in inward parking.

Dependent on the magnitudes of parameters (such as total demand, parking capacities, access times), the required conditions of Δp may never hold for some profiles. In that case, those profiles do not necessarily occur in this transition order. For example, if the hybrid parking choice is converted to inward parking at Profile 9 with the increase of p_1, then Profile 8 must not occur because the condition of Profile 9 in hybrid parking and Profile 8 in strongly inward parking are mutually exclusive. However, if the increase of Δp is such that the hybrid parking is switched to inward parking at Profile 10, then Profile 9 cannot occur, and Profile 10 in hybrid parking is followed directly by Profile 8 in inward parking. Another example is that if given $K_2 < \frac{N\gamma}{\beta+\gamma}$, then Profiles 10 and 11 in hybrid parking must not occur since $K_2 < \frac{N\gamma}{\beta+\gamma}$ and the conditions of Profile 10 or 11 are mutually exclusive. In this case, the transition order is such that Profile 4 in outward parking is followed by Profile 9 in hybrid parking.

The signs of $\partial TC/\partial\Delta p$ of those profiles are (see $\partial TC/\partial\Delta p$ in Appendix D),

Profile	1	$\to 2$	$\to 3$	$\to 4$	$\to 11$	$\to 10$	$\to 9$	$\to 8$	$\to 7$	$\to 6$	$\to 5$
Derivative sign	< 0	U	U	U	< 0	< 0	< 0	U	U	U	> 0

where U stands for "uncertain" or "to be determined". The signs of Profiles 2, 3, 4 are dependent on $\frac{N\gamma}{\beta+\gamma} - K_1$, while those of profiles 8, 7, 6 dependent on $K_2 - \frac{N\gamma}{\beta+\gamma}$. For inward, outward and hybrid parking, there exists certain conditions where increasing p_1 may have positive and negative effect on the total travel cost, respectively. This indicates that increasing the parking fee in the closer cluster so as to induce travelers to park at the outer cluster may not always be desirable in terms of total travel cost.

Optimal parking prices

We now derive the optimal travel profiles with respect to the parking fee. First, we show that the optimal profile falls in neither Profiles 2, 3 and 4 of strongly outward parking nor Profiles 12∼16 of weakly outward parking (the proof is provided in Appendix F), even though their signs of derivatives are parameter-dependent. Therefore, the optimal profile must be the case of hybrid parking or inward parking.

We now discuss the optimal parking fee in hybrid parking or inward parking with respect to the sign of $K_2 - \frac{N\gamma}{\beta+\gamma}$.

If $0 < K_2 < \frac{N\gamma}{\beta+\gamma}$, then the derivatives of Profile 6, 7 and 8 are negative. In addition, Profiles 5 and 6 must occur. This is because in hybrid parking, when Δp changes from $(\lambda - \beta)(l_2 - l_1)$ to $(\lambda + \gamma)(l_2 - l_1)$, the required condition of Profile 6 changes from $0 \leq K_2 \leq \frac{N\gamma}{\beta+\gamma} - s\Delta l$ to $s\Delta l\frac{\gamma+\beta}{\beta} \leq K_2 \leq \frac{N\gamma}{\beta+\gamma}$ which necessarily includes $\forall K_2 \in [0, \frac{N\gamma}{\beta+\gamma}]$. Meanwhile, the upper bound required for Profile 5 to exist can be arbitrarily large as Δp increases, and thus the condition for Profile 5 can always include such a K_2. Since Profiles 5 and 6 must occur, we conclude that the optimal travel profile is achieved in inward or hybrid parking, at the boundary between Profiles 5 and 6, i.e.,

$$\Delta p_{opt} = \frac{K_2\beta}{s} + (\lambda - \beta)(l_2 - l_1) \tag{5.22}$$

In this case, the optimal profile is shown in Figure 5.3.

If $K_2 = \frac{N\gamma}{\beta+\gamma}$, the derivatives for Profiles 6, 7 and 8 are zero. The Δp achieving the optimal profile is not unique. Because $K_1 = \frac{N\beta}{\beta+\gamma}$ and the upper bound of Profile 10 is always less than $\frac{N\beta}{\beta+\gamma}$ in hybrid parking, Profiles 10 and 11 in hybrid parking cannot occur. In fact, $K_1 = \frac{N\beta}{\beta+\gamma}$ satisfies the required condition of Profile 9 with the increase of Δp. To see this, replace K_1 by $\frac{N\beta}{\beta+\gamma}$ in the required condition of profile 9,

$$\frac{N\beta - s(v_1 - v_2)}{\beta + \gamma} \leq \frac{N\beta}{\beta + \gamma} \leq \frac{N\beta + s(u_2 - u_1) - s(v_1 - v_2)}{\beta + \gamma}$$

The left inequality always holds due to the required condition of hybrid parking, $v_1 > v_2$. Rewriting the right inequality yields,

$$\begin{aligned} & (u_2 - u_1) - (v_1 - v_2) \geq 0 \\ \Rightarrow \quad & \Delta p \leq \frac{2\lambda + \gamma - \beta}{2}(l_2 - l_1) < (\lambda + \gamma)(l_2 - l_1) \end{aligned}$$

Because hybrid parking requires $(\lambda - \beta)(l_2 - l_1) \leq \Delta p \leq (\lambda + \gamma)(l_2 - l_1)$, we always have hybrid parking when Δp changes from $(\lambda - \beta)(l_2 - l_1)$ to $\frac{2\lambda+\gamma-\beta}{2}(l_2 - l_1)$ where the right inequality also holds. Similarly, we also show that $K_1 = \frac{N\beta}{\beta+\gamma}$ also satisfies the required condition of Profile 7 in hybrid parking when Δp changes from $\frac{2\lambda+\gamma-\beta}{2}(l_2 - l_1)$ to $(\lambda + \gamma)(l_2 - l_1)$. Also, it satisfies the required conditions of Profile 5 or 6 in inward parking. Therefore, in the transition order, Profile 9 in hybrid parking is followed by Profile 7 in hybrid parking (Profile 8 cannot occur), and Profiles 6 and 5 must occur in inward parking. The critical Δ_p that converts the hybrid parking to inward parking is such that the inward parking starts at Profile 6 or 5. The upper bound of Δp_{opt} is the bound between Profiles 5 and 6 in inward parking, and its lower bound is the bound between Profile 9 and 7 in hybrid parking. Expressing Δp in terms of parking fee and access time, we have,

$$\frac{2\lambda+\gamma-\beta}{2}(l_2-l_1) \leq \Delta p_{opt} \leq \frac{N\beta\gamma}{s(\beta+\gamma)} + (\lambda-\beta)(l_2-l_1) \tag{5.23}$$

If $N > K_2 > \frac{N\gamma}{\beta+\gamma}$, i.e. $0 < K_1 < \frac{N\beta}{\beta+\gamma}$, the derivatives of Profiles 6, 7 and 8 are positive. In hybrid parking, when Δp changes from $(\lambda-\beta)(l_2-l_1)$ to $(\lambda+\gamma)(l_2-l_1)$, the required condition of Profile 9 changes from $\frac{N\beta}{\beta+\gamma} \leq K_1 \leq \frac{N\beta}{\beta+\gamma} + s\Delta l$ to $\frac{N\beta}{\beta+\gamma} - s\Delta l \leq K_1 \leq \frac{N\beta}{\beta+\gamma} - s\Delta l$. Therefore, Profile 9 must occur for $\forall K_1 \in [\frac{N\beta}{\beta+\gamma} - s\Delta l, \frac{N\beta}{\beta+\gamma})$. In that case, the optimal travel profile achieves when Δp is such that the upper bound of the required condition of Profile 9 equals K_1 or equivalently $N - K_2$, i.e.

$$\Delta p_{opt} = \frac{2\lambda+\gamma-\beta}{2}(l_2-l_1) + \frac{K_2(\beta+\gamma)-N\gamma}{2s}, \forall K_2 \in (\frac{N\gamma}{\beta+\gamma}, \frac{N\gamma}{\beta+\gamma} + s\Delta l]$$

However, Profile 9 may not occur for $\forall K_1 \in (0, \frac{N\beta}{\beta+\gamma} - s\Delta l)$, i.e. $\forall K_2 \in (\frac{N\gamma}{\beta+\gamma} + s\Delta l, N)$. In that case, only Profiles 10 and 11 may occur in hybrid parking. When Δp gradually increases in the way that hybrid parking transitions to inward parking, Profile 10 or 11 in hybrid parking then transitions to one of Profile 5, 6, 7 and 8. Therefore, the optimal profile achieves weakly inward parking where

$$\Delta p_{opt} = (\lambda+\gamma)(l_2-l_1), \forall K_2 \in (\frac{N\gamma}{\beta+\gamma} + s\Delta l, N)$$

Combing the two cases of $N > K_2 > \frac{N\gamma}{\beta+\gamma}$ yields,

$$\Delta p_{opt} = \min\left(\frac{2\lambda+\gamma-\beta}{2}(l_2-l_1) + \frac{K_2(\beta+\gamma)-N\gamma}{2s}, (\lambda+\gamma)(l_2-l_1)\right) \tag{5.24}$$

The optimal profile achieves in either weakly inward parking or the special case of Profile 9 (i.e., Figure 5.2(b) where $z = 0$) in hybrid parking.

In addition, if $K_2 = 0, K_1 = N$ or $K_1 = 0, K_2 = N$, then Δp_{opt} can be arbitrary since $\partial TC/\partial \Delta p = 0$.

In a nutshell, if the parking facilities (locations and capacities of both parking clusters) are pre-determined, the minimum total travel cost is achieved in hybrid or inward parking unless $K_2 = 0, K_1 = N$ or $K_1 = 0, K_2 = N$. If both parking clusters have parking spaces available, increasing the parking fee of the closer parking cluster in such a way that outward parking switches to hybrid or inward parking, can always reduce the total travel cost. Compared to outward parking, inward parking can shorten the arrival time window. As a result, this concentrates travelers' arrivals to the office closer to the work starting time. Therefore, inward parking reduces the total schedule delay cost without changing much the queuing delay and therefore yields a significantly less total travel cost than outward parking. This conclusion is consistent with that under the assumption of continuous parking spaces (Arnott et al. 1991). However, increasing the parking fee of the closer parking cluster in a small vicinity, as reflected by the sign of $\partial TC/\partial \Delta p$, may not necessarily reduce the total travel cost.

5.4.1.3 Access time

Access time from a parking spot to one's office also plays an important role in determining parking preference and travel departure times. To examine the effect of access time on the system performance, we suppose that parking fees and capacities of both parking clusters are fixed. Weakly outward and weakly inward parking hold if and only if l_2 equals a single real value, and TC is not differentiable with respect to l_2 (or Δl). It is easy to verify that given K_1, K_2 and Δp, the total travel cost of weakly outward (inward) parking is always larger than that of strongly outward (inward) parking. Therefore, we only need to consider the derivatives, $\partial TC/\partial \Delta l$, for strongly outward, strongly inward and hybrid parking as shown in Appendix D.

As indicated by the signs of those derivatives with respect to access time, reducing the access time of the farther parking cluster can always reduce the total travel cost in inward, outward and hybrid parking with only one exception. The exception is associated with $K_2 \leq \frac{s(v_1 - v_2)}{\beta}$ in hybrid and inward parking, where TC is independent of l_2. Intuitively, this is because the farther cluster is preferred at the beginning of the commuting period, and the reduction in parking-related travel costs by only shortening its access time is offset by the increase in queuing delay or schedule delay caused by travelers taking advantage of this accessibility improvement. Similarly, reducing the access time of the closer parking cluster can reduce the total travel cost in all types of parking preference with the exception that $K_2 \leq \frac{s(v_1-v_2)}{\beta}$ where TC is independent of l_1.

The access time in this dissertation refers to the actual time spent on the way from the parking lot to the office by any traffic mode, such as shuttle buses or walking. Our analysis indicates that, in order to obtain the optimal travel pattern in the sense of minimum total travel cost, the public parking operators should always improve the accessibility of parking spaces, i.e., reduce the access times of both parking clusters, by providing more frequent shuttle bus services or locate the parking lots closer to offices, for instance.

5.4.2 Optimal provision of parking

We already discussed how an individual factor, when holding the other two factors constant, should be set to induce the travel profile that minimizes total travel cost. In this section, we show how all three factors, parking capacity, fee and access time, should be set jointly to obtain the optimal travel pattern by a parking operator who has full control over parking supply.

5.4.2.1 Optimal parking fees, capacities and access times

First of all, the access time of both parking clusters should always be shorten to achieve the optimum as discussed in Section 5.4.1.3. In reality, the access time is usually restricted by the locations of the parking clusters and thus can only be reduced up to a certain level. In order to obtain the optimal travel profile, l_2 and l_1 will always be set to the lowest possible values, $\hat{l}_2$ and $\hat{l}_1$, which we consider as given.

If the parking fee and access time are given, Equations 5.19, 5.20 and 5.21 indicates that the optimal K_2 is always smaller than $\frac{N\gamma}{\beta+\gamma}$. Therefore, Equations 5.23 and 5.24 never hold in the optimal profile where all three factors can be jointly adjusted. As a result, the optimal parking fee, capacity and access time must be such that Equation 5.22 holds, i.e.

$$\Delta p_{opt} = \frac{K_{2,opt}\beta}{s} + (\lambda - \beta)(\hat{l}_2 - \hat{l}_1) \tag{5.25}$$

and the minimum total travel cost is always achieved in inward or hybrid parking unless $K_2 = 0, K_1 = N$ or $K_1 = 0, K_2 = N$. Let $TC(\Delta p, K_2)$ represent the total travel cost in terms of parking fee ($\Delta p = p_1 - p_1$) and the effective capacity of the farther cluster.

If $(\lambda-\beta)(\hat{l}_2-\hat{l}_1) \geq \frac{N\beta\gamma}{s(\beta+\gamma)}$, then $K_{2,opt} = 0$. To see this, for any $K_2' \neq 0$ and any Δp, since $\frac{K_2'\beta}{s} + (\lambda-\beta)(\hat{l}_2-\hat{l}_1)$ is the optimal parking fee for any given K_2', thus,

$$TC(\Delta p, K_2') \geq TC(\frac{K_2'\beta}{s} + (\lambda-\beta)(\hat{l}_2-\hat{l}_1), K_2')$$

Because $\frac{K_2'\beta}{s} + (\lambda-\beta)(\hat{l}_2-\hat{l}_1) > \frac{N\beta\gamma}{s(\beta+\gamma)}$ and Equation 5.19, we have,

$$TC(\frac{K_2'\beta}{s} + (\lambda-\beta)(\hat{l}_2-\hat{l}_1), K_2') > TC(\frac{K_2'\beta}{s} + (\lambda-\beta)(\hat{l}_2-\hat{l}_1), 0)$$

Therefore, in this case,

$$K_{2,opt} = 0, K_{1,opt} = N, \Delta p_{opt} > \frac{N\beta\gamma}{s(\beta+\gamma)} \tag{5.26}$$

and the optimal profile is shown as Profile 20 in Appendix D. When the closer parking cluster offers an overwhelming advantage in the access time, the optimal network performance achieves when all travelers park in the closer parking cluster and the closer parking cluster charges a relatively high parking fee.

Now we turn to the case where $(\lambda-\beta)(\hat{l}_2-\hat{l}_1) < \frac{N\beta\gamma}{s(\beta+\gamma)}$. If $\Delta p_{opt} > \frac{N\beta\gamma}{s(\beta+\gamma)}$, then $K_{2,opt} = 0, K_{1,opt} = N$ and Equation 5.25 cannot hold, the corresponding total travel cost is thus,

$$TC(\Delta p_{opt}, 0) = \frac{N^2\beta\gamma}{s(\beta+\gamma)} + \lambda\hat{l}_1 N \tag{5.27}$$

If $\Delta p_{opt} \leq \frac{N\beta\gamma}{s(\beta+\gamma)}$, then Equation 5.25 holds and the corresponding total travel cost becomes,

$$TC(\Delta p_{opt}, K_{2,opt}) = \frac{N^2\beta\gamma}{s(\beta+\gamma)} + \lambda\hat{l}_1 N + K_{2,opt}(\Delta p_{opt} - \frac{N\beta\gamma}{s(\beta+\gamma)})$$

which is always no larger than the TC in Equation 5.27 under $\Delta p_{opt} \leq \frac{N\beta\gamma}{s(\beta+\gamma)}$. Therefore, the case with $\Delta p_{opt} > \frac{N\beta\gamma}{s(\beta+\gamma)}$ never occurs in the optimal travel profile. Express Δp_{opt} in terms of $K_{2,opt}$ by Equation 5.25, and then substitute it in the total travel cost,

$$TC(\Delta p_{opt}, K_{2,opt}) = \frac{s}{\beta}\Delta p_{opt}^2 - \frac{s}{\beta}\Delta p_{opt}\left(\frac{N\beta\gamma}{s(\beta+\gamma)} + (\lambda-\beta)(\hat{l}_2-\hat{l}_1)\right) + C$$

where C is a term independent of Δp_{opt}. Solving for Δp_{opt} by $\partial TC/\partial\Delta p_{opt} = 0$ and $\partial TC^2/\partial^2\Delta p_{opt} > 0$ yields,

$$\Delta p_{opt} = \frac{N\beta\gamma}{2s(\beta+\gamma)} + \frac{1}{2}(\lambda-\beta)(\hat{l}_2-\hat{l}_1) \tag{5.28}$$

$$K_{2,opt} = \frac{N\gamma}{2(\beta+\gamma)} - \frac{s}{2\beta}(\lambda-\beta)(\hat{l}_2-\hat{l}_1) \tag{5.29}$$

The resulting optimal profile is shown in Figure 5.3. It is easy to check that the obtained Δp_{opt} and $K_{2,opt}$ satisfy that $\Delta p_{opt} \leq \frac{N\beta\gamma}{s(\beta+\gamma)}$ and $N > K_{2,opt} > 0$ given that $(\lambda-\beta)(\hat{l}_2 - \hat{l}_1) < \frac{N\beta\gamma}{s(\beta+\gamma)}$.

When the farther parking cluster offers a reasonable access time compared to the closer one, the optimal travel profile is achieved in inward or hybrid parking where the optimal capacities and parking fees are determined by Equations 5.28 and 5.29 such that both parking clusters will be used.

As can be seen from the optimal profile, the traffic at the bottleneck now has two peaks, instead of one overly concentrated arrival peak in the case without parking choices. Each of the two peaks is caused by traffic demand targeting a particular parking area. Therefore, the traffic congestion at the bottleneck has been mitigated by shifting traffic demand and directing them to different parking choices. In particular, in order to achieve the best system performance, the traffic targeting the closer parking cluster does not depart until the queue of the traffic targeting the farther one vanishes in the middle of the rush hour. Meanwhile, no bottleneck capacity waste should occur (i.e. the bottleneck discharging rate should always be the maximum flow rate so as to fully utilize the roadway facility).

To sum up, when the closer parking cluster offers overwhelming advantages in accessibility, the optimal travel pattern is achieved by having all travelers park in the closer parking cluster. When the farther parking cluster offers competitive accessibility compared to the closer one (or the closer parking cluster does not offer a far advantageous accessibility than the farther one), the optimal travel profile is such that both parking clusters should be used and the resulting total travel cost is less than the case without parking choices.

In addition, we also show that,

$$\frac{\partial TC(\Delta p_{opt}, K_{2,opt})}{\partial s} = -\frac{1}{s^2}\frac{N^2\beta\gamma(4\beta+3\gamma)}{4(\beta+\gamma)^2} - \frac{\lambda-\beta}{2\beta}(\hat{l}_2 - \hat{l}_1) < 0$$

It is interesting to see from Equations 5.28 and 5.29 that if such an optimal travel profile is obtained given a travel demand and a bottleneck capacity, and thereafter the bottleneck capacity is able to be enlarged, then the parking operators should accordingly reduce the actual capacity of the farther cluster and meanwhile increase the parking fee of the farther parking cluster, in order to maintain optimal network performance. Such a setting of parking capacities and fees ensure that the enlargement of the bottleneck capacity will always reduce the total travel cost.

5.4.2.2 Properties of the optimal parking setting

In the typical morning commute problem, the system-optimal dynamic toll scheme can completely eliminate all the queuing delay, but such a optimal toll does not reduce travelers' individual travel cost. We now examine whether the optimal settings of parking location, fee and capacity can lead to less queuing delay than the case without parking choices (i.e. the typical morning commute problem), and how it affects individual travel cost which includes the parking fee.

The optimal parking fees and capacities by Equations 5.28 and 5.29 can effectively reduce the travel delay as compared to the case where all travelers park in the same cluster. To see this, we derive the travel delay cost under such an optimal setting (the optimal profile is shown in Figure 5.3),

$$TD = \frac{\beta}{2s\alpha}\left(K_{2,opt}^2 + (N - K_{2,opt})^2 \frac{\gamma}{\beta+\gamma}\right)$$

Arnott et al. (1990) gives that the travel delay of a typical morning commute problem is $\frac{N^2\beta\gamma}{2(\beta+\gamma)s\alpha}$. Therefore,

$$\frac{TD}{\frac{N^2\beta\gamma}{2(\beta+\gamma)s\alpha}} = 1 - \frac{K_{2,opt}}{N}\left(2 - \frac{K_{2,opt}}{N}\left(2 + \frac{\beta}{\gamma}\right)\right)$$

On the other hand, substitute $K_{2,opt}$ by Equation 5.29, we have

$$\frac{K_{2,opt}}{N}\left(2 + \frac{\beta}{\gamma}\right) = \left(\frac{\gamma}{2(\beta+\gamma)} - \frac{s(\hat{l}_2 - \hat{l}_1)(\lambda - \beta)}{2\beta N}\right)\left(2 + \frac{\beta}{\gamma}\right) < \frac{1}{2} \times 3 < 2$$

Therefore,

$$\frac{TD}{\frac{N^2\beta\gamma}{2(\beta+\gamma)s\alpha}} < 0$$

and the reduction in queuing delay achieved by the optimal setting of parking fee and capacity, as compared to the case where parking is not used to manage the travel demand, is up to

$$\frac{1}{(2 + \frac{\beta}{\gamma})} \times 100\%, \text{ which is achieved when } \frac{K_{2,opt}}{N} = \frac{1}{(2 + \frac{\beta}{\gamma})}$$

If $\gamma = 5\beta$ (this relationship of β and γ is usually adopted in the literature), then the reduction in queuing delay is fairly significant. For example, when $K_{2,opt}$ equals half of the demand N, the reduction of total delay is 40%. Although the optimal setting of parking is unable to completely eliminate all the queuing delay, it can indeed reduce the congestion effectively.

Under such an optimal travel profile, the total travel cost inclusive of parking fee is,

$$\begin{aligned} TC(\Delta p_{opt}, K_{2,opt}) &= \frac{N^2\beta\gamma}{s(\beta+\gamma)} + \lambda\hat{l}_1 N + Np_{1,opt} - \frac{NK_{2,opt}\beta\gamma}{s(\beta+\gamma)} \\ &< \frac{N^2\beta\gamma}{s(\beta+\gamma)} + \lambda\hat{l}_1 N + Np_{1,opt} \end{aligned} \tag{5.30}$$

The right side of Inequality 5.30 is the total travel cost inclusive of parking fee, $p_{1,opt}$, if all travelers use the closer parking cluster. This implies that instead of having all travelers park at the closer parking cluster, encouraging some travelers to use the farther one can, as a matter of fact, reduce the total system travel cost (inclusive of parking fee) even if the revenue collected from the parking fee is not re-distributed to the public. Thus, compared to the case where parking is not used to manage the travel demand, every traveler is better off under such an optimal parking policy. Unlike the system-optimal dynamic toll under which the individual travel cost remains the same as a typical morning commute problem, the optimal parking is able to reduce travelers' individual travel cost. This implies that using parking to manage traffic may be welcomed by travelers and less controversial than toll schemes.

5.4.3 Numerical examples

Next we use a numerical example to illustrate how the optimal capacity and parking fee should be set under a realistic setting in the morning commute. The basic model parameters are as follows: A total demand of $N = 10,000$ vehicles commute in morning peak hour and go through a freeway bottleneck with capacity $s = 80$ veh/min (approximately a three-lane freeway), so that the morning peak commuting time lasts around 2 hours. $\alpha = 10\$/hour, \beta = 4\$/hour, \gamma = 20\$/hour, \lambda = 10\$/hour$ are chosen according to the literature. The closer parking cluster is on average 2 minutes away from the office, while the outer parking cluster is about 20 minutes away from the office (probably by shuttle bus, or walking) for free.

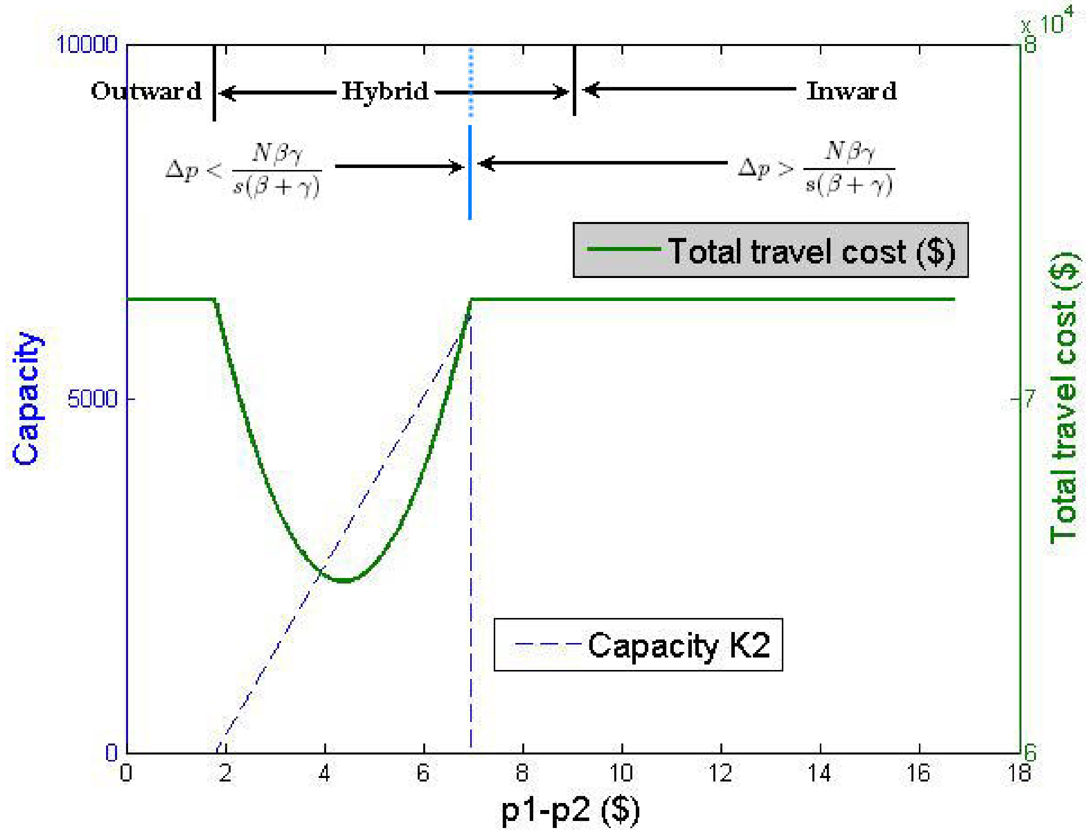


Figure 5.4. The changes in the optimal K_2 and optimal TC with respect to pre-determined Δp

Figure 5.4 depicts the changes in the optimal capacity of the outer parking cluster and optimal TC with respect to pre-determined parking fees. When the central parking cluster charges a small amount of fee that is less than $1.8 (in this case, the closer parking is preferred, and parking is outward) or a high parking fee larger than $6.9, one should build a sufficiently large lot in the central cluster to accommodate all travelers to obtain an optimal performance. In both cases, the TC exclusive of parking fees is $436,667, since the resulting travel profile is independent of parking fees. When the parking fee changes from $1.8 to $6.9, the optimal parking capacity of the outer cluster is non-zero and ranges from 0 to 6,174 out of the required 10,000 parking spaces. Moreover, in some cases, more parking spaces in the peripheral cluster should be provided than in the central cluster to achieve optimal outcome. Meanwhile, if parking fee is medium such that both parking clusters are used, then the resultant optimal TC is less than that when it is set too high or too low, and the reduction in optimal TC can be as much as

11%. In this numerical example, because $\frac{N\beta\gamma}{s(\beta+\gamma)} = 6.9 < 9 = (\lambda+\gamma)\Delta l$, all the optimal profiles with non-zero outer parking capacity are in the form of hybrid parking.

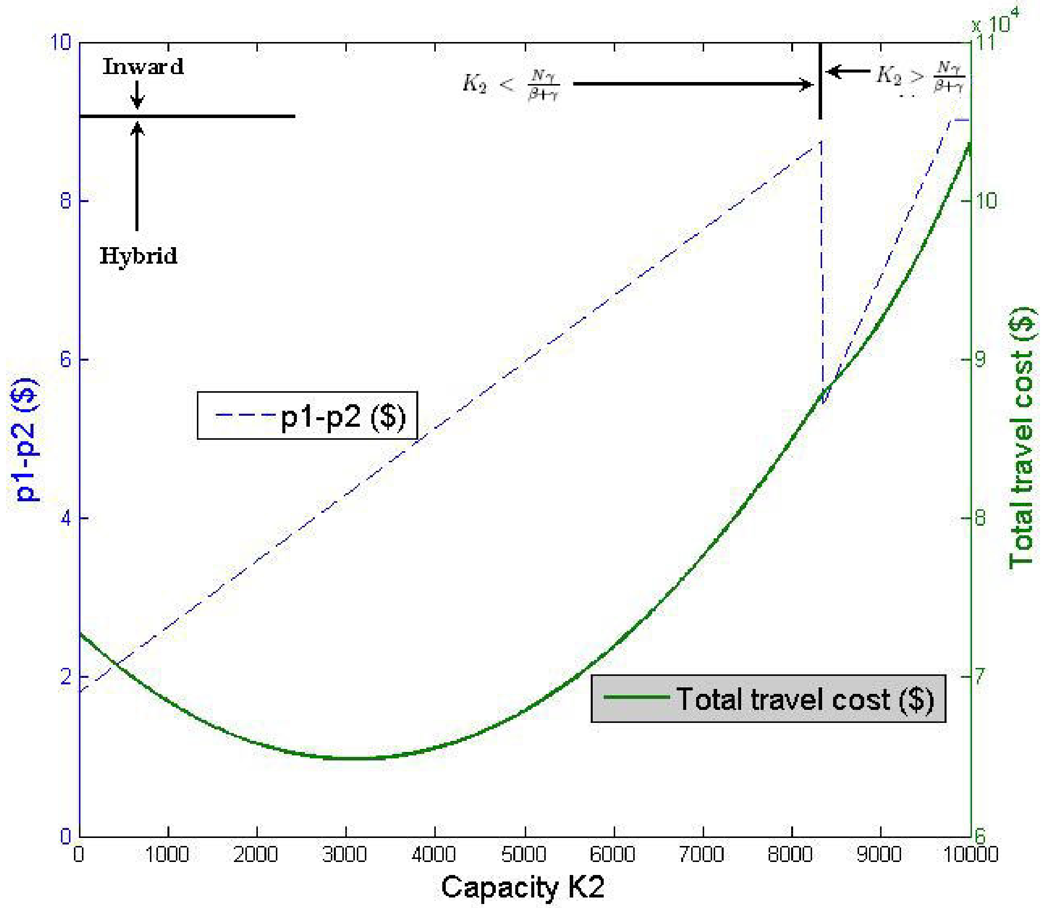


Figure 5.5. The changes in the optimal Δp and optimal TC with respect to pre-determined parking capacity for the peripheral (farther) parking cluster

We also plot Figure 5.5 to show how the optimal parking fee of the central cluster and the optimal TC change with respect to pre-determined parking capacities. When the outer parking cluster has up to $K_2 = \frac{N\beta\gamma}{s(\beta+\gamma)} = 8,333$ parking spaces, the optimal parking fee for the central cluster ranges from \$1.8 to \$8.7, and the resultant optimal travel profiles are in the form of hybrid parking. If the outer parking cluster provides exactly 8,333 parking spaces, then the optimal parking fee of the central cluster could vary from \$5.4 to \$8.7. The optimal profiles occur in weakly inward parking only when $K_2 > 9,773$, i.e. the outer parking cluster overwhelmingly dominates the parking market. With the increase of (pre-determined) K_2, the TC decreases in the first place, reaches the minimum when $K_2 = 3,087$ and then increases up to \$616,667. Adjusting parking fee to obtain the optimal profile when $K_2 = 3,087$ can get up to 37% reduction of TC compared to the optimal TC when no parking spaces are provided in the central parking cluster.

Overall, if the parking operators can adjust both parking capacity and parking fee, then optimal network performance achieves when $K_2 = 3,087$ and $\Delta p = \$4.4$. The minimum TC is \$389,027, a 11% reduction of travel cost compared to the case where all travelers use the central parking cluster. Meanwhile, such an optimal parking capacity and fee can yield 40.8% reduction of queuing delay and every traveler is better off in the sense of a 5.2% reduction in individual travel cost.

5.5 The Case of Parking Lots Owned Privately

5.5.1 Travel profiles and total commuter cost

As discussed before, there are in all 20 possible types of travel profiles (i.e. cumulative departure curve and arrival curve) under those five types of parking preference. The travel profiles and their corresponding TCCs are listed in Table 5.2 (TSC will simply be the sum of TCC and operators' investment cost). Note that both TCC and TSC are used particularly for the case with privately owned parking, which is different from the TC in the case with publicly owned parking. In all five types of parking preference, if $K_1 = 0, K_2 = N$ ($K_2 = 0, K_1 = N$), then the profile and its TCC are shown in the row of "other" in the table.

5.5.2 The combined parking/departure-time equilibrium model

Now we are ready to give the definition of the equilibrium, the **competitive parking equilibrium** associated with the parking market defined above, based on the **day-to-day dynamic user equilibrium** defined in Section 5.1.

Definition 5.2. *In a parking market, the allocation of parking capacities, $\bar{K_1}, \bar{K_2}$, parking fees, $\bar{p_1}, \bar{p_2}$, and access times, $\bar{l_1}, \bar{l_2}$, constitute a* ***competitive parking equilibrium***

Table 5.2. An overview: 20 travel profiles and TCCs for five types of parking preference

Parking preference	Condition	Total commuter cost	Profile type
Outward (strong)	$0 \le K_1 \le \frac{s(v_2-v_1)}{\beta}$	$\frac{N\beta\gamma}{s}\frac{N-K_1}{\beta+\gamma}+\lambda l_2 N+Np_2$	1
	$\frac{s(v_2-v_1)}{\beta} \le K_1 \le \frac{s(v_2-v_1)+N\gamma}{\beta+\gamma}$		2
	$\frac{s(v_2-v_1)+N\gamma}{\beta+\gamma} \le K_1 \le \frac{s(u_2-u_1)+N\gamma}{\beta+\gamma}$	$\frac{N\beta}{s}\frac{s(u_2-u_1)+N\gamma}{\beta+\gamma}+\lambda N l_1+Np_1$	3
	$N > K_1 \ge \frac{s(u_2-u_1)+N\gamma}{\beta+\gamma}$		4
Outward (weak)	$0 \le K_2 \le \frac{1}{2}\frac{s(u_1-u_2)+N\gamma}{\beta+\gamma}$	$\frac{N^2\beta\gamma}{s(\beta+\gamma)}+\lambda l_1 N+Np_1$	12
	$\frac{1}{2}\frac{s(u_1-u_2)+N\gamma}{\beta+\gamma} \le K_2 \le \frac{1}{2}\frac{N\gamma}{\beta+\gamma}$	$\frac{N\beta}{s}2K_2+\lambda l_2 N+Np_2$	13
	$\frac{u_2-u_1}{\beta+\gamma}s+\frac{1}{2}\frac{N\gamma}{\beta+\gamma} \le K_1 \le N-\frac{1}{2}\frac{N\gamma}{\beta+\gamma}$		14
	$\frac{1}{2}\frac{N\gamma}{\beta+\gamma} \le K_1 \le \frac{u_2-u_1}{\beta+\gamma}s+\frac{1}{2}\frac{N\gamma}{\beta+\gamma}$	$\frac{N^2\beta\gamma}{s(\beta+\gamma)}+\lambda l_2 N+Np_2$	15
	$0 \le K_1 \le \frac{1}{2}\frac{N\gamma}{\beta+\gamma}$		16
Inward (strong)	$0 \le K_2 \le \frac{s(v_1-v_2)}{\beta}$	$\frac{N\beta\gamma}{s}\frac{N-K_2}{\beta+\gamma}+\lambda l_1 N+Np_1$	5
	$\frac{s(v_1-v_2)}{\beta} \le K_2 \le \frac{s(u_1-u_2)+N\gamma}{\beta+\gamma}$		6
	$\frac{s(u_1-u_2)+N\gamma}{\beta+\gamma} \le K_2 \le \frac{s(v_1-v_2)+N\gamma}{\beta+\gamma}$	$\frac{N\beta}{s}\frac{s(u_1-u_2)+N\gamma}{\beta+\gamma}+\lambda N l_2+Np_2$	7
	$N > K_2 \ge \frac{s(v_1-v_2)+N\gamma}{\beta+\gamma}$		8
Inward (weak)	$0 \le K_2 \le \frac{s(v_1-v_2)}{\beta}$	$\frac{N\beta\gamma}{s}\frac{N-K_2}{\beta+\gamma}+\lambda l_1 N+Np_1$	5
	$\frac{s(v_1-v_2)}{\beta} \le K_2 \le \frac{N\gamma}{\beta+\gamma}$		6
	$\frac{N\gamma}{\beta+\gamma} \le K_2 \le \frac{N\gamma+s(v_1-v_2)}{\beta+\gamma}$	$\frac{N^2\beta\gamma}{s(\beta+\gamma)}+\lambda N l_2+Np_2$	7
	$\frac{N\gamma+s(v_1-v_2)}{\beta+\gamma} \le K_2 \le N-\frac{1}{2}\frac{N\beta-s(v_1-v_2)}{\beta+\gamma}$		17
	$0 \le K_1 \le \frac{1}{2}\frac{N\beta-s(v_1-v_2)}{\beta+\gamma}$	$\frac{N^2\beta\gamma}{s(\beta+\gamma)}+\lambda N l_2+Np_2$	18
Hybrid	$0 \le K_2 \le \frac{s(v_1-v_2)}{\beta}$	$\frac{N\beta\gamma}{s}\frac{N-K_2}{\beta+\gamma}+\lambda l_1 N+Np_1$	5
	$\frac{s(v_1-v_2)}{\beta} \le K_2 \le \frac{s(u_1-u_2)+N\gamma}{\beta+\gamma}$	$\frac{N\beta}{s}\frac{s(u_1-u_2)+N\gamma}{\beta+\gamma}+\lambda N l_2+Np_2$	6
	$\frac{s(u_1-u_2)+N\gamma}{\beta+\gamma} \le K_2 \le \frac{s(u_1-u_2)+N\gamma+s(v_1-v_2)}{\beta+\gamma}$		7
	$\frac{N\beta-s(v_1-v_2)}{\beta+\gamma} \le K_1 \le \frac{N\beta+s(u_2-u_1)-s(v_1-v_2)}{\beta+\gamma}$	$\frac{N\beta}{s}(N-K_1-\frac{s(v_1-v_2)}{\beta+\gamma})+\lambda N l_2+Np_2$	9
	$\frac{(u_2-u_1)s}{\beta+\gamma} \le K_1 \le \frac{N\beta-s(v_1-v_2)}{\beta+\gamma}$	$\frac{N^2\beta\gamma}{s(\beta+\gamma)}+\lambda N l_2+Np_2$	10
	$0 \le K_1 \le \frac{(u_2-u_1)s}{\beta+\gamma}$		11
Other	$K_1=0, K_2=N$	$\frac{N^2\beta\gamma}{s(\beta+\gamma)}+\lambda N l_2+Np_2$	19
	$K_1=N, K_2=0$	$\frac{N^2\beta\gamma}{s(\beta+\gamma)}+\lambda N l_1+Np_1$	20

if,

1. *Private parking operators in each cluster locally maximize their profits by setting their own parking fee, capacity and access time, given the price of the other cluster, i.e.* $\max K_i p_i - K_i a_i$ *(*$i = 1, 2$*)*
2. *Travelers maximize their utilities (i.e. minimize the generalized travel cost in this case) by choosing a departure time and parking location, given the parking fees, capacity allocations and access times of both clusters, which is represented by a day-to-day dynamic user equilibrium defined in Section 5.1.*
3. *Market clearing, i.e.* $K_1 + K_2 = N$*. Each of the commuters will choose one of the parking cluster to park his car.*

5.5.3 Parking provision without regulations

In this section, we study the case of a parking market without regulatory intervention. We first solve the competitive parking equilibrium based on the UE traffic profiles presented in Section 5.5.1, then discuss its properties and finally examine how changes in parking fee, access time and parking capacity affects market performance through a sensitivity analysis. Because access time is a less flexible control factor than parking fee and capacity, we first assume it is fixed in the short term in deriving the competitive parking equilibrium (that is, the parking operators adjust the parking fees and effective capacities to achieve the competitive equilibrium, taking the access time as given in the short term), then we can show how it affects the competitive equilibrium when it changes in the long term.

5.5.3.1 The competitive parking equilibrium

Let $\Delta l = l_2 - l_1 > 0$ and $\Delta p = p_1 - p_2 \geq 0$. $\bar{p}_1, \bar{p}_2$ and $\bar{K}_1, \bar{K}_2$ denote the parking fee and capacity allocation of the central and peripheral clusters under the competitive equilibrium, respectively. We will show that there are in all four equilibria: Type I in strongly outward parking, Type II in weakly outward parking, Type III in weakly inward parking, and Type IV in hybrid parking. We now discuss the competitive equilibria under each of the five types of parking location preference.

As can be seen from Table 5.1, if the competition leads to parking fees and access times such that the strongly outward or strongly inward parking occurs, then there exists a preferred parking cluster where the private operators who manage it will always build N parking spaces so that all the travelers will use their parking spaces and their profits are maximized. Since the private operators in the other cluster will never have any consumers in that case, under such a parking market, they can always attract consumers by reducing their parking charge to induce a transition from the strongly outward or inward parking to other types of parking under which their lots are preferred by some travelers. However, there is one (and the only) case that the less favored cluster is unable to secure a market share no matter how its spaces are priced. This occurs when $p_1 \leq (\lambda - \beta)\Delta l$. If the parking charge of the closer cluster is set to be sufficiently low, say $p_1 < (\lambda - \beta)\Delta l$, then the farther parking cluster will not attract any commuters even if its spaces are free. When $p_1 = (\lambda - \beta)\Delta l$, the private operators will not build any lot in the farther cluster because they will not make any profit, although travelers are indifferent to parking locations in early arrival when $p_2 = 0$. Therefore, we show that a competitive equilibrium of Type I occurs when,

$$\bar{K_1} = N \tag{5.31a}$$

$$\bar{K_2} = 0 \tag{5.31b}$$

$$\bar{p_1} = (\lambda - \beta)(l_2 - l_1) \tag{5.31c}$$

where the operators in the closer cluster set a low parking price so that it can attract all the travelers and the farther parking cluster is never used, and meanwhile the profits of the operators in the closer parking cluster are locally maximized.

When the parking fee in the closer cluster is priced higher than $(\lambda - \beta)\Delta l$, the operators in the farther cluster can always set a price to get a market share with some profits. Therefore, both strongly inward and strongly outward parking cannot occur in the competitive equilibrium other than Type I.

Weakly outward or weakly inward parking may occur in the competitive equilibrium. For weakly outward parking, in Profiles No. 12 and 13, K_2 effectively determines the parking usage and thus, the operators in the farther cluster are always willing to build as many parking spaces as possible to attract commuters so as to increase their profits;

while in Profiles No. 14, 15 and 16, K_1 similarly determines the parking usage and the operators in the closer cluster will also build as many parking spaces as possible. There exists a capacity allocation at the boundary between Profiles No. 13 and 14 such that the operators in both clusters cannot further increase their market share by building more spaces and keeping a constant relative parking price (i.e. $\Delta p = (\lambda - \beta)(l_2 - l_1)$ as the required condition of weakly outward parking), which satisfies the definition of the competitive parking equilibrium. Therefore, a competitive equilibrium of Type II occurs when,

$$\bar{K_1} = N - \frac{1}{2}\frac{N\gamma}{\beta + \gamma} \tag{5.32a}$$

$$\bar{K_2} = \frac{1}{2}\frac{N\gamma}{\beta + \gamma} \tag{5.32b}$$

$$\bar{p_1} - \bar{p_2} = (\lambda - \beta)(l_2 - l_1) \tag{5.32c}$$

Similarly, a competitive equilibrium of Type III occurs in weakly inward parking at the boundary between Profiles No. 17 and 18,

$$\bar{K_1} = \frac{1}{2}\frac{N\beta - (\beta + \gamma)(l_2 - l_1)s}{\beta + \gamma} \tag{5.33a}$$

$$\bar{K_2} = N - \frac{1}{2}\frac{N\beta - (\beta + \gamma)(l_2 - l_1)s}{\beta + \gamma} \tag{5.33b}$$

$$\bar{p_1} - \bar{p_2} = (\lambda + \gamma)(l_2 - l_1) \tag{5.33c}$$

Note that the type III competitive equilibrium may not exist if $N\beta < (\beta + \gamma)(l_2 - l_1)s$.

More importantly, a competitive parking equilibrium of Type IV occurs in hybrid parking at the boundary between Profiles 7 and 9 where,

$$K_1 = \frac{N\beta + s(u_2 - u_1) - s(v_1 - v_2)}{\beta + \gamma} \tag{5.34}$$

$$K_2 = \frac{N\gamma - s(u_2 - u_1) + s(v_1 - v_2)}{\beta + \gamma} \tag{5.35}$$

In this case, the competition leads to a situation where the parking fees are priced to be such that each parking cluster is preferred in a certain time period, i.e., the farther parking cluster is preferred in early arrival, and the closer parking cluster is preferred in late arrival. Meanwhile, the number of parking spaces built in both clusters are such that a further increase in the capacity in either cluster will not increase its the market share and profits.

We now turn to solve the Type IV competitive parking equilibrium. Given the equilibrated price of the closer parking cluster, $\bar{p_1}$, the profit maximization problem of the farther cluster reads,

$$\begin{aligned}\max_{p_2} K_2(p_2-a_2) &= \max_{p_2}\left(\frac{N\beta+s(u_2-u_1)-s(v_1-v_2)}{\beta+\gamma}(p_2-a_2)\right)\\ &= \max_{p_2}\left(-\frac{2s}{\beta+\gamma}p_2^2+\frac{N\gamma+2s\bar{p_1}+2sa_2-(2\lambda+\gamma-\beta)s\Delta l}{\beta+\gamma}p_2+C\right)\end{aligned}$$

where C is a term independent of p_2. Therefore,

$$\bar{p_2} = argmax_{p_2}K_2(p_2-a_2) = \frac{N\gamma}{4s} - \frac{2\lambda+\gamma-\beta}{4}\Delta l + \frac{\bar{p_1}+a_2}{2} \tag{5.36}$$

and substitute p_2 in Equation 5.35 by $\bar{p_2}$ in Equation 5.36, we have

$$\bar{K_2} = \frac{\frac{N\gamma}{2} + s(\bar{p_1} - a_2 - \frac{2\lambda+\gamma-\beta}{2}\Delta l)}{\beta+\gamma} \tag{5.37}$$

Similarly, maximizing the profits of the operators who own the closer cluster, given the equilibrated price of the farther cluster, $\bar{p_2}$, yields,

$$\bar{p_1} = \frac{N\beta}{4s} + \frac{2\lambda+\gamma-\beta}{4}\Delta l + \frac{\bar{p_2}+a_1}{2} \tag{5.38}$$

$$\bar{K_1} = \frac{\frac{N\beta}{2} + s(\bar{p_2} - a_1 + \frac{2\lambda+\gamma-\beta}{2}\Delta l)}{\beta+\gamma} \tag{5.39}$$

Adding up Equations 5.37 and 5.39 should satisfy the market clearing condition,

$$\bar{p_1} + \bar{p_2} = \frac{N}{2s}(\beta+\gamma) + a_1 + a_2 \tag{5.40}$$

Combining Equations 5.40 and 5.38 (or 5.36) solves the competitive equilibrium,

$$\bar{p_1} = \frac{N}{6s}(2\beta+\gamma) + \frac{2\lambda+\gamma-\beta}{6}\Delta l + \frac{2a_1+a_2}{3} \tag{5.41a}$$

$$\bar{p_2} = \frac{N}{6s}(\beta+2\gamma) - \frac{2\lambda+\gamma-\beta}{6}\Delta l + \frac{a_1+2a_2}{3} \tag{5.41b}$$

$$\bar{K_1} = \frac{N(\gamma+2\beta)+s(2\lambda+\gamma-\beta)\Delta l - 2s(a_1-a_2)}{3(\beta+\gamma)} \tag{5.41c}$$

$$\bar{K_2} = \frac{N(2\gamma+\beta)-s(2\lambda+\gamma-\beta)\Delta l + 2s(a_1-a_2)}{3(\beta+\gamma)} \tag{5.41d}$$

Because the Type IV competitive equilibrium is the result of hybrid parking, $(\lambda-\beta)\Delta l < (\bar{p_1}-\bar{p_2}) < (\lambda+\gamma)\Delta l$. It exists only if

$$\frac{N}{2s}(\gamma-\beta) - (\gamma+2\beta-\lambda)\Delta l < a_1 - a_2 < \frac{N}{2s}(\gamma-\beta) + (2\gamma+\beta+\lambda)\Delta l \tag{5.42}$$

In other words, if the difference in the investment cost between the two clusters is too small or too large, Type IV equilibrium will not exist. This is somewhat expected. An extremely high investment cost in the farther cluster can lead to a high parking price set by those operators (as a result of Equation 5.36). When the farther cluster is overpriced, travelers are not willing to use the farther parking spaces and consequently the parking preference becomes outward, which destroys the equilibrium. On the other hand, if those operators who own the farther cluster set a lower price so that they do make some profits (still under hybrid parking) but the profits are not maximized, they can increase the parking fee to make more profits. This, however, place them into the risk of losing their market share because a fee increase beyond a certain threshold can eventually change travelers' parking preference. In this sense, such a Type IV competitive equilibrium may not be reachable and the parking market may be unstable in terms of parking prices and capacity allocations. The same logic applies to the case where the operators who own the closer cluster are subject to an extremely high investment cost. In both cases where a Type IV equilibrium is not stable, regulations can be used to create a stable market in the manner of, for instance, restricting the parking fee or capacity of both clusters. This can ensure the private operators in both clusters earn reasonable profits in a stable parking market. More importantly, market regulation may help further increase the system performance and reduce congestion than what would be achieved if the market is left alone, as will be discussed later.

5.5.3.2 Stability of the equilibria

There are in all four types of competitive parking equilibrium. Not all of them, however, are stable. We examine them here one by one.

The Type II and Type III equilibrium occurs when the parking preference is weakly outward or weakly inward, and the two types of preference only hold when Δp equals a single real value, i.e. $\Delta p = (\lambda - \beta)\Delta l$ or $\Delta p = (\lambda + \gamma)\Delta l$. Such equilibria are obviously not stable because a slight price change in either parking cluster can result in a change in parking preference and thus completely changes the travel profile and market shares. Both types of competitive equilibrium are unlikely to occur in reality since their traffic patterns are unstable.

Type I occurs where $\bar{p}_1 = (\lambda - \beta)\Delta l$. This type of equilibrium is theoretically stable in the sense that the operators in the closer cluster can set a reasonable price and build N spaces such that they attract all the commuters and locally maximizes their profits, no matter how the farther cluster lowers its price or changes its capacity. However, Type I may not exist (or sometimes may not be stable/desired) in a practical sense for the following reasons: 1) The fixed investment cost per parking space is normally high in the closer cluster. It is very likely that $\bar{p}_1 = (\lambda - \beta)\Delta l < a_1$. Under this condition the operators in the closer cluster make no profit. To make a profit they have to increase the parking fee, but this destroys the Type I equilibrium; 2) Even if $\bar{p}_1 = (\lambda - \beta)\Delta l > a_1$ and the operators in the closer cluster can make profits by setting a low price and achieving Type I equilibrium, the profits they make under Type I (since locally maximized) may be less than what they can make under a Type IV equilibrium in which they can charge a substantially higher price, and a smaller market share. Therefore, a Type IV equilibrium is more desirable; and 3) now we only consider the case where Δl is fixed in the short run. In reality, if the farther parking cluster manages to improve its accessibility in the long run, it can easily destroy the Type I equilibrium to get a market share, which leads to fundamentally different flow patterns. Type I equilibrium, therefore, is unlikely to occur in practice.

The Type IV competitive parking equilibrium is stable both in theory and in practice, because its travel preference and profile exists under a broad range of prices and capacities, and it offers the opportunity for operators in both clusters to make a profit. Since this type of equilibria is most likely to occur in practice, we focus on the analysis of the Type IV equilibrium in the rest of the chapter.

5.5.4 The Type IV competitive parking equilibrium

In this section, we examine in detail the properties of the Type IV equilibrium from several perspectives: market share, profits, and system performance.

5.5.4.1 Market share, profits and travel profile

As can be seen from Equation 5.41, given the attributes of the commuter population (i.e. β, λ, γ) and travel demand N, improving the accessibility of parking lots (i.e., reducing the "access time") is desired for both parking clusters since a shorter access time in one cluster can, holding the access time of the other cluster constant, increase both its market share and equilibrated parking fee, hence its profit. Therefore, the private operators in both clusters have the incentive to reduce their access times through providing frequent shuttle bus services or other means.

As for the investment cost, a higher investment cost of one cluster can lead to a smaller market share and profit for this cluster, but higher profit for the other cluster. Therefore, private operators in both clusters have the incentive to reduce their investment costs. In addition, an increase in the investment cost of one cluster indeed can lead to higher parking fees for both clusters, which does not favor travelers. This is because the operators in this cluster have to raise the parking price to pay off a higher investment cost, and therefore the competitors will also raise their parking fee, however not as much as the former (by a half of the increased fee of the former to be exact), which is proven to enlarge their market shares and profits.

We plot the travel profile of the Type IV equilibrium in Figure 5.6 before we show several of its features.

This profile is actually a special case of both Profiles No. 7 and 9. As discussed before, in hybrid parking, a traveler who departs at such a time that he arrives earlier than t^* using either cluster prefers the farther parking cluster due to $v_1 > v_2$, while a traveler who departs at such a time that he arrives later than t^* using either cluster prefers the closer one, due to $u_1 < u_2$. However, some travelers may depart the bottleneck in such a time that they are subject to early arrival if choosing the closer cluster, and late arrival if choosing the farther cluster. For those travelers, there exists a transition period (i.e. the time period from D to B on the departure curve from the bottleneck) during which travelers' parking preference switches gradually from the farther cluster to the closer one. The traveler who departs home at time t_A (the departure time of the traveler marked by A in the profile) is indifferent to both clusters and $\bar{y} = \frac{u_2 - u_1}{\beta + \gamma}$, and the travel profile under competitive market equilibrium is such that this exact traveler fills up the

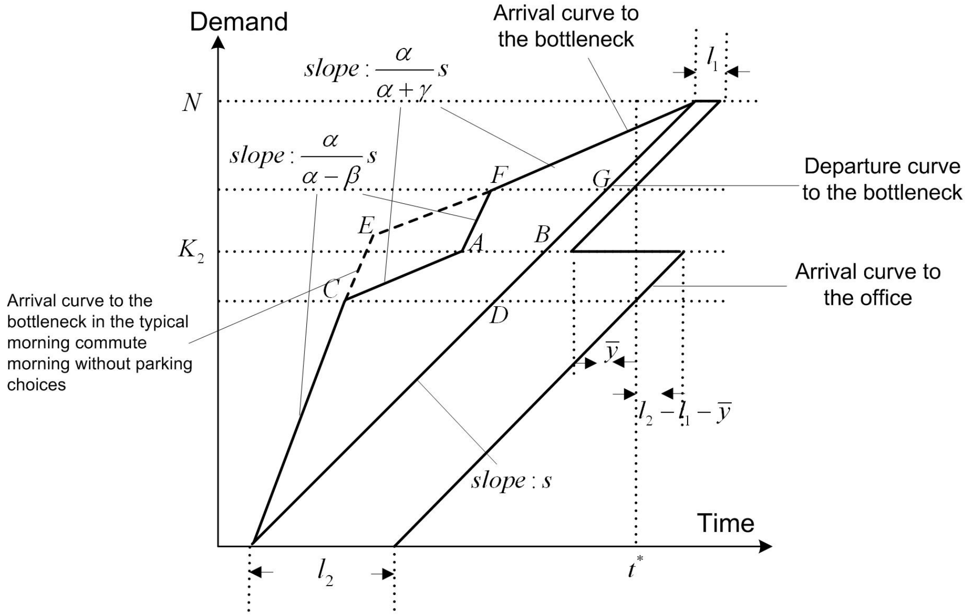


Figure 5.6. The travel profile under Type IV competitive equilibrium

farther parking cluster. As can be seen from the profile, during the morning commute, the queue at the bottleneck has two peaks. The first peak is due to the demand for parking at the farther cluster when it is preferred, while the second one is caused by those travelers parking at the closer cluster when the preference on the farther cluster is later transferred to the closer one.

The travel pattern achieved under the competitive equilibrium indicates that compared to the morning commute without parking choices, a parking market has the ability to mitigate the congestion in the middle of the rush hour, i.e. ACEF in Figure 5.6 is the queuing delay saved by such a competition between parking clusters. While it may not necessarily reduce the schedule delay cost (as the departure curve from the bottleneck can move along the time axle within a certain range, dependent on the value of $\bar{y}$), it can always reduce the queuing delay. Incorporating parking choices in the morning commute essentially distinguishes travelers in groups such that each group of commuters will travel within a certain time period different from others, and hence the queue can be shortened without necessarily reducing the travel demand nor spreading the peak time. This has the same effect as tolling travelers selectively. In the latter, each of the travelers is charged a toll in a fraction of working days so as to change travelers' departure times (Daganzo & Garcia 2000). However, using parking choices to distinguish travelers during the morning commute may be less controversial since travelers can change their own departure times voluntarily in a parking market, rather than being selected to be tolled in certain days.

5.5.4.2 Total travel cost and total queuing delay

Now we compute the total travel cost and total queuing delay of a Type IV equilibrium, and compare them to the case where all travelers use either the closer or farther cluster. Let $\overline{TCC}$ ($\overline{TSC}$), $TCC_{o,1}$ ($TSC_{o,1}$) and $TCC_{o,2}$ ($TSC_{o,2}$) denote the total commuter cost (total system cost) under the Type IV equilibrium and the cases without parking choices where all travelers use parking cluster 1 or 2, respectively.

Under the travel profile as shown in Figure 5.6, if the parking fees of the two clusters are set to be p_1 and p_2 and Equation 5.34 and 5.35 hold, then the total commuter travel cost is[6],

$$\begin{aligned} TCC &= \frac{N^2\beta\gamma}{s(\beta+\gamma)} + \lambda N l_2 + \frac{N}{\beta+\gamma}(\beta p_1 + \gamma p_2 - \beta(\lambda+\gamma)\Delta l) \\ &= \frac{N^2\beta\gamma}{s(\beta+\gamma)} + \lambda N l_1 + \frac{N}{\beta+\gamma}(\beta p_1 + \gamma p_2 + \gamma(\lambda-\beta)\Delta l) \end{aligned} \tag{5.43}$$

and the total system cost is,

$$\begin{aligned} TSC &= \frac{N^2\beta\gamma}{s(\beta+\gamma)} + \lambda N l_2 - \frac{N}{\beta+\gamma}\beta(\lambda+\gamma)\Delta l + a_1K_1 + a_2K_2 - \Delta ps(2\bar{y} - \Delta l) \\ &= \frac{N^2\beta\gamma}{s(\beta+\gamma)} + \lambda N l_1 + \frac{N}{\beta+\gamma}\gamma(\lambda-\beta)\Delta l + a_1K_1 + a_2K_2 - \Delta ps(2\bar{y} - \Delta l) \end{aligned} \tag{5.44}$$

where K_1 and K_2 are determined by Equation 5.34 and 5.35, and

$$\bar{y} = \frac{(\lambda+\gamma)\Delta l - (p_1 - p_2)}{\beta+\gamma}$$

Since the parking market reaches a competitive equilibrium such that $\bar{p_1}$, $\bar{K_1}$ and $\bar{p_2}$, $\bar{K_2}$ can be represented by Equation 5.41, then,

$$\begin{aligned} \overline{TCC} &= \frac{N^2\beta\gamma}{s(\beta+\gamma)} + \lambda N l_2 + \frac{N}{\beta+\gamma}(\beta\bar{p_1} + \gamma\bar{p_2} - \beta(\lambda+\gamma)\Delta l) \\ &= \frac{N^2\beta\gamma}{s(\beta+\gamma)} + \lambda N l_1 + \frac{N}{\beta+\gamma}(\beta\bar{p_1} + \gamma\bar{p_2} + \gamma(\lambda-\beta)\Delta l) \end{aligned} \tag{5.45}$$

[6] We first solve the geometry of the profile and then compute the TCC and TSC. The derivation is fairly lengthy and is omitted here.

$$\overline{TSC} = \frac{N^2\beta\gamma}{s(\beta+\gamma)} + \lambda N l_2 - \frac{N}{\beta+\gamma}\beta(\lambda+\gamma)\Delta l + a_1\bar{K}_1 + a_2\bar{K}_2 - \Delta\bar{p}s(2\bar{y}-\Delta l) \tag{5.46}$$
$$= \frac{N^2\beta\gamma}{s(\beta+\gamma)} + \lambda N l_1 + \frac{N}{\beta+\gamma}\gamma(\lambda-\beta)\Delta l + a_1\bar{K}_1 + a_2\bar{K}_2 - \Delta\bar{p}s(2\bar{y}-\Delta l)$$

If all the travelers use either the closer or the farther cluster, then,

$$TCC_{o,1} = \frac{N^2\beta\gamma}{s(\beta+\gamma)} + \lambda N l_1 + N p_{o,1}$$
$$TCC_{o,2} = \frac{N^2\beta\gamma}{s(\beta+\gamma)} + \lambda N l_2 + N p_{o,2}$$
$$TSC_{o,1} = \frac{N^2\beta\gamma}{s(\beta+\gamma)} + \lambda N l_1 + N a_1$$
$$TSC_{o,2} = \frac{N^2\beta\gamma}{s(\beta+\gamma)} + \lambda N l_2 + N a_2$$

Due to monopoly, the parking fee charged by the monopolistic cluster is usually higher than what is achieved under the competitive market equilibrium, i.e. $p_{o,1} \geq \bar{p}_1$ and $p_{o,2} \geq \bar{p}_2$. It is easy to verify that $\overline{TCC} < TCC_{o,1}$ and $\overline{TCC} < TCC_{o,2}$ given the required condition of hybrid parking, i.e. $(\lambda-\beta)\Delta l < \bar{p_1} - \bar{p_2} < (\lambda+\gamma)\Delta l$. However, the sign of $\overline{TSC} - TSC_{o,1}$ (or $\overline{TSC} - TSC_{o,2}$) is actually undetermined, and depends on the values of the parameters, such as the access times and investment costs. This implies that such a competitive equilibrium, naturally, can reduce the commuter travel cost compared to the case where all travelers use the same parking cluster, due to the competition among private parking operators. But it does not necessarily reduce the total social cost. Therefore, a competitive parking market, under certain conditions can lead to an undesirable market outcome.

We already show graphically that the competitive equilibrium can reduce the queuing delay as compared to the typical morning commute problem without parking choices. Let $\overline{TD}$ and TD_o denote the total queuing delay under the Type IV competitive equilibrium and the typical morning commute problem, respectively. Under the travel profile shown in Figure 5.6, if the parking fees of the two clusters are set to be p_1 and p_2, then the queuing delay is,

$$TD = \frac{N^2\gamma\beta}{2\alpha s(\beta+\gamma)} - \bar{y}(\Delta l - \bar{y})\frac{s}{\alpha}(\beta+\gamma) \tag{5.47}$$
$$\text{where } \bar{y} = \frac{(\lambda+\gamma)\Delta l - (p_1 - p_2)}{\beta+\gamma}$$
$$TD_o = \frac{N^2\gamma\beta}{2\alpha s(\beta+\gamma)}$$

Therefore, when a Type IV equilibrium is achieved,

$$\overline{TD} - TD_o = -\bar{y}(\Delta l - \bar{y})\frac{s}{\alpha}(\beta + \gamma) < 0 \tag{5.48}$$

Given the access times of both clusters, Δl, we show that the parking fees that minimize the queuing delay, denoted by Δp^{**}, are such that $\bar{y} = \frac{1}{2}\Delta l$, i.e.

$$\Delta p^{**} = \frac{2\lambda + \gamma - \beta}{2}\Delta l \tag{5.49}$$

This can serve as one of the targeted parking prices for a regulatory agency in order to achieve minimum queuing delay.

5.5.4.3 The effects of investment cost, parking fee and access time

Now we perform a sensitivity analysis to study how changes in parking fee, access time and investment cost affect TSC, TCC and TD. The derivatives of TSC, TCC and TD with respect to all three factors are shown in Appendix G. First we examine the effect of investment cost. The investment cost, including the real estate value/tax and maintenance cost, may be adjusted by both the private operators and the public regulator. Because

$$\frac{d\overline{TCC}}{da_2} > \frac{d\overline{TCC}}{da_1} > 0$$

lowering the investment cost per parking space in either parking cluster can always reduce the commuter travel cost as it reduces the parking fee, and it seems such a reduction in investment cost in the farther cluster can be more efficient than it is in the closer cluster. However, reductions in the investment cost may not necessarily reduce both the total system cost and queuing delay. This is because the savings in operators' cost may not pay off the increase in the queuing delay or schedule delay cost that travelers are subject to.

When the difference of parking fee between the two clusters is not large (i.e. $2\Delta\bar{p} < (2\lambda + \gamma - \beta)\Delta l$), the regulator can set a higher investment cost in the central cluster or lower it in the outer cluster in order to reduce the queuing delay. On the other hand, a higher investment cost in the central area or a lower cost in the outer area can lead to less total social cost only if the difference of investment cost is sufficiently large.

In other words, to achieve a better system performance, the regulator should consider enlarging the difference of investment cost only if it indeed already makes sufficiently large difference.

Now we examine the effect of access time. Whether or not reducing access time of the farther parking cluster can benefit the commuter cost is dependent on the population attributes, i.e. β, γ, λ. In addition, under a certain condition, shortening the access time may not necessarily reduce the queuing delay nor the system cost: the outcome depends on not only the population attributes but also some other parameters, such as total demand$N$ and the bottleneck capacity s. Therefore, under the Type IV equilibrium, improving the accessibility of the parking lots may not benefit the commuters nor the entire system, but it would benefit some parking operators because lowering access times can attract more customers and bring in higher profits.

Finally, by differentiating Equation 5.43 with respect to p_1 and p_2, we have,

$$\frac{\partial TCC}{\partial p_1} > 0, \frac{\partial TCC}{\partial p_2} > 0 \tag{5.50}$$

implying that reducing the parking fee in both clusters under the Type IV competitive equilibrium can always benefit commuters. While if the difference of parking fee between the two clusters is sufficiently high, then reducing the parking fee of the farther cluster (and hence also the difference of parking fee) can benefit the overall system. If the market is regulated, a regulator can require one cluster to reduce its parking fee, then the other cluster will also lower its price in order to remain competitive. This results in a lower cost for commuters but probably less profits for parking operators (as we can see later), and can benefit the entire system under certain conditions.

To summarize, a parking market can result in less queuing delay than the morning commute without parking choices. However, a parking market without regulation can lead to a market equilibrium where the total system cost and commuter cost are higher than the case without parking choices and can be further reduced. This can be remedied through market regulation, such as a price ceiling or capacity floor, as we shall discuss in the next section. It is worth noting that reducing the access times of both clusters may not benefit the travelers in terms their travel costs, but can bring more profits to the operators.

5.5.5 The regulated parking market

Since the competitive parking equilibrium may not produce the best market outcome, we explore in this section how market regulation can be used to improve the market outcome in terms of total cost and congestion. Several types of regulations will be discussed, they include price-ceiling, capacity-floor or ceiling, quantity tax or subsidy. Each of these regulations is introduced to achieve the following objectives: 1) reduce the total system cost; 2) reduce the queuing delay (i.e. network congestion); 3) reduce the commuters' total cost (or equivalently, commuters' individual travel cost), particularly the parking fee; and 4) maintain a certain level of profitability for the private operators. The resultant total cost and queuing delay are expected to be reduced compared to the following two cases, 1) the morning commute without parking choices (i.e. the typical morning commute problem) and 2) competitive equilibrium without regulation.

5.5.5.1 The price-ceiling regulation

As shown in the previous section, a price-ceiling applied to the farther parking cluster can reduce the parking fees of both clusters, and consequently the total commuter travel cost. It can also ensure a stable market when a competitive equilibrium does not exist. Suppose that a regulator sets a price ceiling, p_2', for the farther cluster, we have

$$p_2 \leq p_2' \tag{5.51}$$

Let $\bar{p}_{1,c}, \bar{p}_{2,c}$ and $\bar{K}_{1,c}, \bar{K}_{2,c}$ denote the parking fee and capacity allocation of the closer and farther cluster under the Type IV competitive market equilibrium with the price-ceiling regulation, respectively. Given the equilibrated price of the closer cluster, $\bar{p}_{1,c}$, the profit maximization problem for the farther cluster reads,

$$\max_{p_2 \in [0, p_2']} K_2(p_2 - a_2) = \max_{p_2 \in [0, p_2']} \left(-\frac{2s}{\beta+\gamma} p_2^2 + \frac{N\gamma + 2s\bar{p}_{1,c} + 2sa_2 - (2\lambda+\gamma-\beta)s\Delta l}{\beta+\gamma} p_2 + C \right)$$

where C is a term independent of p_2. In order to let such a p_2' take effect in reducing the equilibrated price on the farther cluster, we have

$$p_2' < \bar{p}_2 = \frac{N}{6s}(\beta + 2\gamma) - \frac{2\lambda + \gamma - \beta}{6}\Delta l + \frac{a_1 + 2a_2}{3} \tag{5.52}$$

Therefore,

$$\bar{p}_{2,c} = argmax_{p_2 \in [0, p_2']} K_2(p_2 - a_2) = p_2' \tag{5.53}$$

$$\bar{K}_{2,c} = \frac{N\gamma + \frac{N\beta}{2} - s(p_2' + \frac{2\lambda + \gamma - \beta}{2}\Delta l - a_1)}{\beta + \gamma} \tag{5.54}$$

Similarly, maximizing the profits of private operators in the closer cluster, given the equilibrated price of the farther cluster, p_2', yields,

$$\bar{p}_{1,c} = \frac{N\beta}{4s} + \frac{2\lambda + \gamma - \beta}{4}\Delta l + \frac{p_2' + a_1}{2} \tag{5.55}$$

$$\bar{K}_{1,c} = \frac{\frac{N\beta}{2} + s(p_2' + \frac{2\lambda + \gamma - \beta}{2}\Delta l - a_1)}{\beta + \gamma} \tag{5.56}$$

and the condition of market clearing is satisfied. Because such an equilibrium only exists in hybrid parking, i.e. $(\lambda - \beta)\Delta l < \bar{p}_{1,c} - \bar{p}_{2,c} < (\lambda + \gamma)\Delta l$, we solve for p_2' by combing Equation 5.55 and 5.53, which yields

$$\frac{N\beta}{2s} - \frac{\Delta l}{2}(2\lambda + 3\gamma + \beta) + a_1 < p_2' < \frac{N\beta}{2s} + \frac{\Delta l}{2}(\gamma - 2\lambda + 3\beta) + a_1 \tag{5.57}$$

Inequality 5.52 and 5.57 altogether determine an appropriate range for p_2' that the regulator can choose from. If p_2' is out of this range, the market does not attain a stable equilibrium, or the regulation may not take effect.

Compared to the unregulated parking market, in the market with the price-ceiling regulation, private operators in the closer cluster also lower their parking fee in order to compete with the farther cluster. Although the closer parking cluster now has a higher market share, its operators make less profits than in the unregulated market. To see this, we note that

$$\begin{aligned}
\frac{\partial \max K_1(p_1 - a_1)}{\partial p_2} &= -\frac{s}{\beta + \gamma}(p_2 - \frac{N\beta}{2s} - a_1 - \frac{2\lambda + \gamma - \beta}{4}\Delta l) \\
\frac{\partial \max K_1(p_1 - a_1)}{\partial p_2}\Big|_{p_2 = p_2'} &> 0 \text{ due to Inequality 5.57.} \\
\frac{\partial \max K_1(p_1 - a_1)}{\partial p_2}\Big|_{p_2 = \bar{p}_2} &> 0 \text{ due to Equality 5.36.}
\end{aligned}$$

As a consequence, the farther parking cluster has a smaller market share with a lower parking fee, and therefore also earns less profits.

Because the competitive parking equilibrium under the price-ceiling regulation is achieved under the same profile type (i.e. Figure 5.6) as in the unregulated market, we can compute the total commuter cost using Equation 5.43 and total system cost using Equation 5.44,

$$\begin{aligned} TCC_c &= \frac{N^2\beta\gamma}{s(\beta+\gamma)} + \lambda N l_2 + \frac{N}{\beta+\gamma}(\beta\bar{p}_{1,c} + \gamma\bar{p}_{2,c} - \beta(\lambda+\gamma)\Delta l) \qquad (5.58) \\ TSC_c &= \frac{N^2\beta\gamma}{s(\beta+\gamma)} + \lambda N l_2 - \frac{N}{\beta+\gamma}\beta(\lambda+\gamma)\Delta l + a_1\bar{K}_{1,c} + a_2\bar{K}_{2,c} - \Delta p_c s(2\bar{y}_c - \Delta l) \end{aligned}$$

From this we can obtain the derivatives of the total cost with respect to the price bound and access time as follows:

$$\begin{aligned} \frac{dTCC_c}{dp_2'} &= \frac{N}{\beta+\gamma}\left(\frac{1}{2}\beta+\gamma\right) > 0 \\ \frac{dTSC_c}{dp_2'} &= \frac{s}{\beta+\gamma}(p_2' - a_2) - \frac{N\beta}{2(\beta+\gamma)} \end{aligned}$$

As long as the price ceiling p_2' is set to be such that the competitive parking equilibrium still exists, the regulator may want to lower the ceiling as this can always reduce the total commuter cost. If the net profit per parking space for the farther cluster is beyond a threshold (i.e. $\frac{N\beta}{2s}$), then a lower p_2' is also desirable from the system point of view. A lower pricing ceiling p_2' can also reduce both the commuters' individual travel costs and parking charge (in both clusters). However, the ceiling may not be set too low because, 1) a low ceiling p_2' does not necessarily produce less queuing delay than a high ceiling; 2) a low ceiling p_2' can squeeze the profits of operators in both clusters and in fact result in a higher total system cost; and 3) the ceiling has a lower bound given by Equation 5.57. Although p_2' can be set to be asymptotically approaching the lower bound or $\frac{N\beta}{2s} + a_2$ to further reduce the total travel cost, whichever comes smaller. Such a minimum cannot be achieved due to the discontinuity of the travel profile at the lower bound. The price ceiling regulation transfers the benefits from operators to commuters through a lower parking fees rather than less queuing delay. Hence a too low price bound may not be desirable if the regulator's objective is to use it to manage traffic congestion.

We show that $\frac{dTD_c}{dp_2'}$ can take any sign. Therefore, the price-ceiling regulation may not necessarily reduce the total queuing delay as compared to the case without regulation. However, we can set a $p_2'^{**}$ to minimize the queuing delay. Such an optimal price bound can be obtained by combining Equation 5.53, 5.55 and 5.49,

$$p_2'^{**} = a_1 + \frac{N\beta}{2s} - \frac{2\lambda+\gamma-\beta}{2}\Delta l \qquad (5.59)$$

$p_2'^{**}$ satisfies Inequality 5.57 but may or may not satisfy Inequality 5.52. Therefore, it is possible that such a minimum-queuing-delay inducing bound may not exist in some cases.

5.5.5.2 The capacity-floor or capacity-ceiling regulation

Since the price-ceiling regulation can efficiently reduce the social cost and commuter cost, and ensure a stable parking market, a question naturally arises, can a capacity-floor or capacity-ceiling regulation achieve the same goal?

First, the capacity-ceiling regulation will not work. This is because if a cluster is regulated to build a certain number of spaces that must not exceed such a "ceiling", then the operators in the other cluster can increase the parking price as much as possible to maximize their profits. This may lead to an unreasonably high parking fee in the other cluster and market failure.

On the other hand, a capacity-floor regulation sets a minimum number of parking spaces for a cluster, e.g. $K_2 \geq K_2'$. We would expect the competitive equilibrium (hopefully with less total travel cost) achieves when the regulation takes effect, i.e., the equilibrated capacity $\bar{K}_{2,cf} = K_2'$. However, if $K_2' \leq \bar{K}_2$, then the equilibrium with the capacity-floor regulation still achieves at $\bar{K}_2$ so it does not take effect. If otherwise $K_2' > \bar{K}_2$, then the equilibrium is such that only $\bar{K}_2$ spaces will be used by commuters and the effective equilibrated capacity is still $\bar{K}_2$. Therefore, the capacity-floor regulation does not work either.

Unfortunately, restriction on only the capacity of parking clusters seems not effective or sometimes unnecessary. This is essentially because the regulator can only influence the **actual** parking capacities, rather than the **effective** parking capacities, while the competitive equilibrium determines the **effective** parking capacities, not the **actual** one.

5.5.5.3 The quantity tax/subsidy regulation

In addition to the price-ceiling regulation, taxation may also be an option for the public regulator to adopt in pursuit of a desired market outcome. A tax/subsidy can sometimes be more efficient than the price-ceiling regulation as the tax collected can be re-distributed to the public, or the subsidy paid to the private operators can reduce the deadweight loss. We note that a lump-sum tax or subsidy imposed on operators in either of the two parking clusters does not change the parking fee and capacity allocation at the competitive equilibrium.

Suppose the regulator imposes a quantity tax/subsidy π_1 and π_2 per space (per vehicle) on the closer and farther parking clusters, respectively. If the tax/subsidy is set in favor of the farther cluster, then $\pi_1 > \pi_2 > 0$ in the case of a tax, and $\pi_2 < \pi_1 < 0$ in the case of a subsidy. When the tax/subsidy is set in favor of the closer cluster, then π_1 and π_2 exchange their positions in the above inequalities. A quantity tax/subsidy scheme is equivalent to changing investment costs under the competitive equilibrium. Therefore, by replacing a_1 with $a_1 + \pi_1$ and a_2 with $a_2 + \pi_2$ in Equation 5.41, we have (let the subscript "t" to represent the case of the quantity tax regulation),

$$\bar{p}_{1,t} = \bar{p}_1 + \frac{2\pi_1 + \pi_2}{3} \tag{5.60a}$$

$$\bar{p}_{2,t} = \bar{p}_2 + \frac{\pi_1 + 2\pi_2}{3} \tag{5.60b}$$

$$\bar{K}_{1,t} = \bar{K}_1 - \frac{2s}{3(\beta + \gamma)}(\pi_1 - \pi_2) \tag{5.60c}$$

$$\bar{K}_{2,t} = \bar{K}_2 + \frac{2s}{3(\beta + \gamma)}(\pi_1 - \pi_2) \tag{5.60d}$$

Also, such an equilibrium must satisfy the condition for hybrid parking, i.e., $(\lambda - \beta)\Delta l < \bar{p_1} - \bar{p_2} < (\lambda + \gamma)\Delta l$. We have,

$$\begin{aligned} \frac{N}{2s}(\gamma - \beta) - (\gamma + 2\beta - \lambda)\Delta l - (a_1 - a_2) &< \pi_1 - \pi_2 \\ &< \frac{N}{2s}(\gamma - \beta) + (2\gamma + \beta + \lambda)\Delta l - (a_1 - a_2) \end{aligned} \tag{5.61}$$

Now we can compare this new equilibrium under the tax/subsidy regulation with the competitive market equilibrium without regulation. Let us first consider a quantity tax/subsidy regulation in favor of the farther cluster, $\pi_1 > \pi_2$. By introducing a quantity tax (subsidy), the equilibrated parking prices in the closer cluster and the farther cluster increase (decrease) by $\frac{2\pi_1+\pi_2}{3}$ and $\frac{\pi_1+2\pi_2}{3}$, respectively. The closer parking cluster will have a smaller market share and some commuters may switch to use the farther one. Both operators increase their prices to transfer the regulatory tax to their customers, or both operators reduce their prices to compete for more customers and some of the benefits of the subsidy will also be transferred to the commuters. The profitable parking fee exclusive of tax/subsidy, $p_2 - \pi_2$, in the farther cluster actually increases by $\frac{\pi_1 - \pi_2}{3}$. Since its market share also increase, its operators earn more profits than the case without regulation. However, such a taxation/subsidy reduces both the market share and the parking fee of the closer cluster, and the private operators in the closer cluster are made worse off by this regulation. Therefore, a tax/subsidy in favor of the farther (closer) cluster can efficiently reduce (increase) traffic demand to the central city (CBD). Additionally, as indicated by Equation 5.61, the tax/subsidy must be appropriately set in order to ensure the existence of the competitive equilibrium (also a stable parking market).

According to the derivatives of TSC, TCC and TD with respective to a_1 and a_2, we show that a tax in either of the clusters can increase the parking fee, which essentially increase the commuter travel cost, while a subsidy in contrast can benefits commuters. If the difference of the investment cost between the two clusters is significant, then the regulator should set a quantity tax/subsidy that favors the farther cluster, i.e. $\pi_1 > \pi_2$, to produce a lower total system cost than in the unregulated parking market. Not only does such a regulation tend to balance the profits of both clusters in the market, it also can reduce the total system travel cost. In contrast, if the difference of the investment cost between the two clusters is insignificant, and hence the farther cluster has a more advantageous investment cost than a certain threshold, then the regulator should tax (or subsidize) the closer cluster less (or more) heavily, i.e. $\pi_1 < \pi_2$.

We also obtain a $\Delta\pi^{**}$ that minimizes the queuing delay and in the same time guarantees less congestion than the case without regulation (given Δl) by combining Equations 5.60a, 5.60b and 5.49,

$$\Delta\pi^{**} = \frac{N}{2s}(\gamma - \beta) + \frac{2\lambda + \gamma - \beta}{2}\Delta l - (a_1 - a_2) \tag{5.62}$$

Any setting of tax/subsidy towards $\Delta\pi^{**}$ is desirable in terms of reduced congestion. Because $\Delta\pi^{**}$ satisfies Inequality 5.61, such a tax/subsidy can always be achieved, but it normally does not entail a minimum total social/commuter cost.

Finally, from the travelers' perspective, a subsidy to the operators can reduce the parking fees of both clusters in addition to its capability of reducing total commuter (individual) travel cost, whereas a tax in contrast increase the parking fees. Thus, though both a subsidy and a tax may be able to achieve a better market outcome in terms of total commuter/social cost, a subsidy may be more preferred by the commuters more than a tax.

5.5.6 Numerical examples

In this section we present numerical examples of competitive market equilibrium under realistic network parameters, and show how a price-ceiling regulation and quantity tax regulation can influence the parking market, social welfare and the network congestion.

The basic model parameters are as follows: a total demand of $N = 10,000$ vehicles commute in the morning rush hour and go through a freeway bottleneck with capacity $s = 120$ veh/min (approximately a six-lane freeway with three lanes per direction). $\alpha = \$10/hour, \beta = \$4/hour, \gamma = \$20/hour$ are set to be consistent with the literature. We assume $\lambda = \$15/hour$. The central parking cluster is fairly convenient and it only takes 2 minutes on average to reach the office from the parking space, while the farther parking cluster is about 20 minutes away from the office. Suppose the investment cost of the closer cluster and the farther one is $a_1 = \$10/commuting\ peak/space$ and $a_2 = \$1/commuting\ peak/space$ respectively. In order to have a stable competitive equilibrium in cases with or without regulations, $(\lambda - \beta)\Delta l < \Delta p < (\lambda + \gamma)\Delta l$ (i.e., $3.3\$ < \Delta p < 10.5\$$).

First we solve the competitive market equilibrium without regulation. The equilibrated parking price of the closer cluster and the farther cluster is \$15.8 and \$11.9, respectively. Under such an equilibrium, the closer cluster is used by 3,469 travelers, and the rest 6,531 travelers use the farther one. The total travel cost is \$204,142, and the total queuing delay cost of the network is \$2,196. The resultant queuing delay cost is less than the queuing delay cost of the case without parking choices, \$2,315, by around 5%. As a matter of fact, if no regulation is imposed to the parking market, the parking price can be surprisingly higher than what travelers can accept in practice. Overall, such a market may under-provide parking services. In order to further reduce the total travel cost and queuing delay, and preferably the parking fee as well, a regulator may introduce into the market a price-ceiling or a quantity tax/subsidy regulation.

Figure 5.7 gives the changes in parking prices, capacity, profits, total travel cost

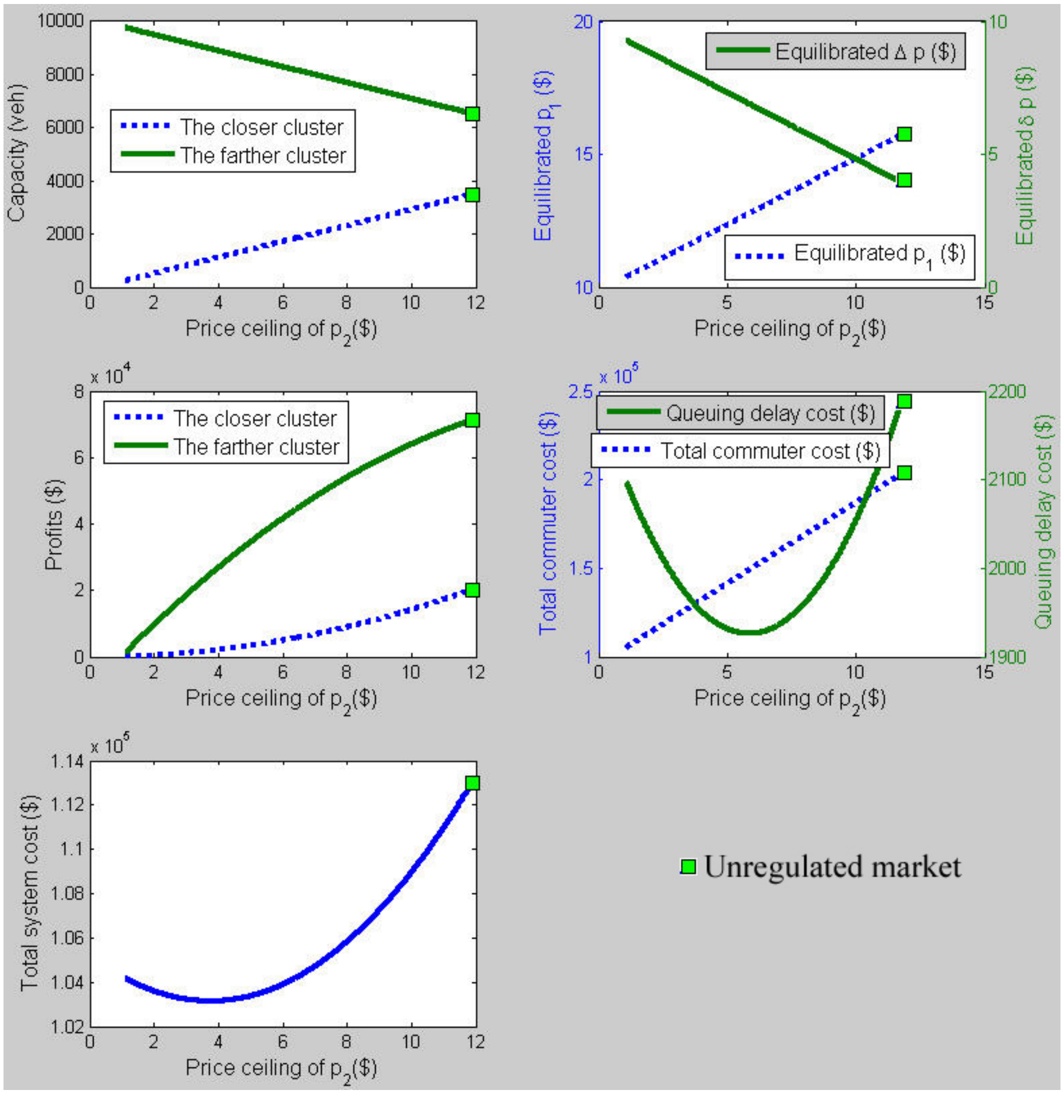


Figure 5.7. The changes in parking prices, capacity allocations, profits, TSC, TCC and TD in a regulated market with respect to a price ceiling

and queuing delay with respect to the price ceiling p_2'. Inequality 5.52 and 5.57 gives that $\$1 < p_2' < \$11.9$ to ensure the existence of an equilibrium. When the price ceiling decreases so that the parking fee of the farther cluster changes from \$11.8 to \$1.1, the number of travelers using the closer cluster reduces from 3,469 to nearly only 200, its parking fee from \$15.8 to \$10.2, and its profits from \$19,761 to \$91. Meanwhile, although the farther cluster tends to dominate the parking market with the decrease of the price ceiling, its profits also reduce from \$70,812 to nearly \$976 due to the decreasing parking fee. Therefore, as proved in previous sections, the profits of private operators in both clusters are squeezed due to the price ceiling. However, from the system point of view, the price-ceiling regulation is very promising. It can reduce the total travel cost, compared to the unregulated market, by up to 10% at a \$3.78 price ceiling. It reduces total commuter cost by nearly half when the price ceiling is \$1.1 and by a quarter when it is \$6.2. In addition, the congestion is also mitigated. When the ceiling is \$5.88, the congestion is minimized with a reduction in delay by 58%. Therefore, the regulatory agency may want to first set a price ceiling ranging from \$1.1 to \$5.88 for the farther cluster, then choose a ceiling based on the tradeoffs between social/commuter cost and congestion.

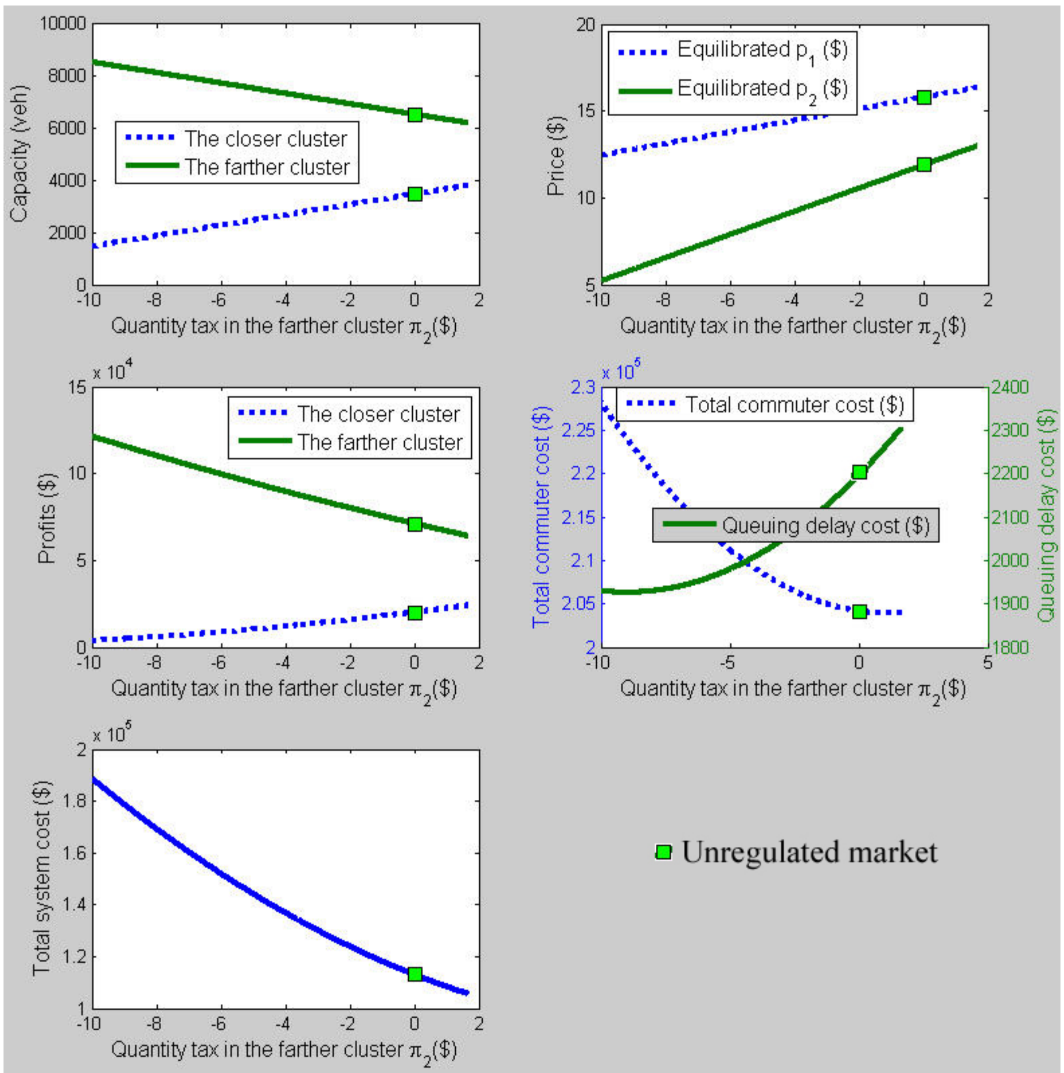


Figure 5.8. The changes in parking prices, capacity allocations, profits, TSC, TCC and TD in the regulated market with respect to a quantity tax/subsidy

We also plot in Figure 5.8 the changes in parking prices, capacity, profits of both clusters, total cost and queuing delay with respect to a quantity tax/subsidy. Without loss of generality, we assume the regulator only charges a tax to or subsidizes travelers of the farther cluster. Inequality 5.61 gives that $-\$19.2 < \pi_2 < \$1.87$ to ensure the existence of a competitive market equilibrium. A tax charged on the farther cluster favors the closer cluster, while a subsidy favors the farther cluster. Overall, by reducing the tax on the farther cluster from \$1.8 to zero, and further subsidizing travelers from zero up to \$10, the farther cluster tends to dominate the market. However, it makes the operators in the farther cluster gain profits (unlike the price-ceiling regulation) while the closer cluster loses profits. Unfortunately, it is unable to reduce the total travel cost effectively. When a \$1.2 tax is charged on the farther cluster, the reduction in the total commuter cost achieves the maximum, but by merely \$221 (around 1.1%). Meanwhile, tax on the farther cluster can yield up to 8% reduction in the system cost. When a subsidy is offered, it always increases both the total system cost (by up to 80%) and commuter cost (by up to 12%), which is not desired from the public's perspective. Although a quantity tax/subsidy is unable to effectively reduce the social cost, the subsidy can mitigate congestion compared to the case of no regulation. The queuing delay can be reduced up to approximately the same amount as in the price-ceiling regulation. Consequently, under a subsidy (to the farther cluster), travelers can suffer less congestion and pay lower parking fee with, however, a considerably higher schedule delay cost.

To sum up, in our numerical example, the price-ceiling regulation overall outperforms, in terms of market efficiency, the quantity tax/subsidy regulation in this experiment. The price-ceiling regulation favors travelers over private operators, while the quantity tax/subsity regulation favors some operators over the travelers who will pay a higher travel cost under this regulation. Even when it offers reduction in the social cost, the latter regulation may not be welcome by the traveling public because a tax is charged on the farther cluster will raise the parking fee and queuing delay cost substantially while it reduces the travel cost only slightly[7].

[7]Note that the quantity tax/subsidy may effectively reduce the total travel cost in other cases, dependent on the network parameters, such as total demand, population attributes and so forth

5.6 Summary

This chapter discusses how parking affects the morning commute patterns and how it can be used to manage traffic in the morning commute. The parking lots in the city are first abstracted to a finite number of parking clusters with each parking cluster capacitated. Within each parking cluster, the same fee is charged for each space and the access time (used to measure the accessibility) from each parking space to the office is assumed to be identical. Dependent on parking fee and access time of the closer and farther parking cluster, there are five types of parking location preference, strongly inward, weakly inward, strongly outward, weakly outward and hybrid. Given parking allocations, fees and access times, we derive all 20 possible travel profiles and corresponding total travel cost under UE, and we study how each factor influences the network performance and the commuting patterns. The cases of parking owned by the public or owned by private operators are discussed separately.

5.6.1 Publicly owned parking

The optimal value of each of parking capacity, fee and access time, holding the other two constant, is extensively discussed. If parking fee and access time are given and only the parking capacity is adjustable, the optimal capacity of the father parking cluster is either 0 or $\frac{s(v_1-v_2)}{\beta}$. Under certain conditions, enlarging the closer parking cluster is not desirable. We may consider restricting its capacity and offering some advantages for the farther parking cluster so as to reduce the total travel cost. If parking capacity and access time are given and only the parking fee is adjustable, the minimum total travel cost is achieved in inward or hybrid parking (unless $K_2 = 0$ or N), and both parking clusters should be used. Changing the types of parking location preference, from the outward parking to hybrid or inward parking, in the manner of increasing the parking fee of the closer one, can always reduce the total travel cost. However, increasing the parking fee of the closer parking cluster in a small range may not necessarily reduce the total travel cost. In addition, a shorter access time of both parking lots is always desirable in all travel profiles.

More importantly, from the parking regulator point of view, we show how all three factors should be set altogether in order to obtain the optimal network performance. When the closer parking cluster offers overwhelming advantages in the access time, the optimal travel pattern is achieved by having all travelers park in the closer parking cluster. When the farther parking cluster offers competitive access time compared to the closer one, the optimal travel profile is such that both parking clusters should be used. In the latter case, the traffic congestion at the bottleneck can be mitigated by shifting traffic demand and directing them to different parking choices. Compared to the case where parking is not used to manage traffic, such an optimal setting can effectively reduce both total travel cost and queuing delay. Although it can neither reduce the total travel cost by more than a half nor eliminate all the queueing delay as can be realized by a time-varying toll, the optimal setting of parking can reduce the total cost (inclusive of parking fee) and thus individual travel cost, i.e. every traveler can be better off under parking regulations. This feature is more advantageous to the travelers than an optimal time-varying toll.

5.6.2 Privately owned parking

When the parking lots are owned privately, we first derived the competitive equilibrium for a parking market. There are in all four types of competitive market equilibrium and only one is proven to be stable, but its existence is not always guaranteed. The stable competitive market equilibrium can yield less total commuter cost and less queuing delay than the case without parking choices, but not necessarily less total social cost. Our sensitivity analysis indicates that the reduction of access times of the parking lots benefits private operators, but does not necessarily reduce the total system cost and congestion. In addition, lowering the investment cost per parking space in either parking cluster can always reduce the commuter travel cost as it reduces the parking fees. To achieve a better system performance, the regulator should consider raising the investment cost of the closer cluster only if there is sufficiently large difference in the investment costs between the two clusters.

Overall, we showed that both total commuter/social cost and queuing delay can be further reduced by introducing regulations on the market, and the regulations can also be used to ensure a stable parking market, to influence the market in favor of the commuters and to balance the profits of private operators.

We considered three types of market regulation, price-ceiling, capacity-floor or capacity-ceiling and quantity tax/subsidy to ensure market stability, improve market outcome, and balance the interests of operators and the traveling public. The price-ceiling regulation looks promising as it can effectively reduce the total commuter cost as compared to no regulation, and the lower the pricing ceiling is, the lower the total commuter cost it achieves. Although the price ceiling regulation does not necessarily reduce traffic congestion when it is compared to the case of no regulation, we showed that there always exists a range of the ceiling where queuing delay is reduced. Under such a price-ceiling regulation, improvement of accessibility of the farther cluster is always desirable from both the regulator's and private operators' perspective. In addition, since it can reduce travelers' individual travel cost and the parking fees charged by both parking clusters, it is also desirable from the travelers's perspective. However, the price-ceiling regulation always squeezes the private operators' profits. Unlike the price-ceiling regulation, a quantity tax/subsidy always favors one parking cluster over the other. It may be able to reduce the total commuter/social cost, and the queuing delay when the tax/subsidy is set in a certain range. Compared to taxation, travelers may favor a subsidy scheme since it can reduce the parking fees charged by both clusters.

Chapter 6

Route Choice Heterogeneity: A New Hybrid Route Choice Model

Now we turn to study DTA for large-scale networks. As discussed in the literature review, most studies assume a route choice model satisfying UE conditions, either BUE or PUE. In real life, travelers' route choice behavior is likely to be more complex than what was assumed in both BUE and PUE. For example, travelers may not consider all the possible routes but have several pre-trip routes in mind prior to their departure, which are selected from their day-to-day traveling experiences. Moreover, these pre-selected routes may not be user-optimal ones. Although travel time and schedule delay costs are dominant factors in travelers' route choice decisions, several other factors, such as road accessibility, pavement conditions, and so on, may influence their decisions as well. Besides these factors, a traveler's personality should also play an important role in his or her route choice. For example, a conservative traveler may stick to his chosen route from day to day while an adventurous traveler may be more willing to explore new routes based on his actual travel experiences. Thus real traffic is more likely to be the product of various types of choice decisions rather than cost-minimizing BUE or PUE applied uniformly across the entire traveling population. It is therefore of particular interest to develop a route choice model that combines various types of information and considers various kinds of travelers.

Rather than treating all travelers identically, this chapter speculates that some travelers are likely to follow their pre-determined routes while others update their routes en-route in response to real-time information. Following this, we propose a hybrid route choice model coupled with the CTM implementation of the kinematic wave traffic flow model (Lighthill & Whitham 1955, Richards 1956), and show how route choices and queue spillback affect the resulting flow patterns. Computational procedures to implement the models are explored with several numerical examples, and issues related to model calibration are also discussed.

6.1 A Hybrid Route Choice Model

6.1.1 The general model

Let $[0, T]$ be an assignment horizon (i.e., the analysis period). The network is assumed to be empty at $t = 0$. Corresponding to the assignment period, we define a loading horizon $[0, T']$, where T' marks the time when the network is empty. Furthermore, let ϕ_a denote an assignment interval, a discrete duration during which the departure flow rate for any O-D pair is assumed to be constant (m_a is the number of assignment intervals, i.e., $T = m_a\phi_a$). ϕ_l is the loading interval, a discrete duration during which network conditions are assumed to be stationary (a loading horizon consists of m_l loading intervals of uniform length, i.e., $T' = m_l\phi_l$). ϕ_a must be a multiple of ϕ_l.

We introduce two groups of travelers: travelers who are willing to deviate from their pre-determined routes and those who are not. The reason is simple. Some conservative travelers, once they determine which routes to take and get familiar with those particular routes, would rather stick to them than risking on finding new (or unknown) routes that may actually turn out to be worse than their previous routes, unless the congestion they experience in their current routes becomes unacceptable to them. Those travelers are normally reluctant to deviate from their prescribed routes. We call this group of travelers **habitual travelers** (the proportion of this type of travelers is $1 - \theta$). On the other hand, some adventurous travelers are more willing to explore new routes in response to their travel experience and/or up-to-date traffic information. They may be equipped with devices that offer real-time navigation, or they may be familiar with the entire network and are able to change their routes to avoid the congestion. We call this group of travelers **adaptive travelers** (the proportion of this type of travelers is θ). In a real network, the proportion θ, also referred to as the **Diversion Ratio**, may not be a constant and may change with respect to network conditions. For example, the diversion ratio can increase in the event of a major accident or highway reconstruction project. Nevertheless, we expect the diversion ratio to be relatively stable for a network at least in the short run barring the occurrences of various major incidents.

For the habitual travelers, their routes are determined based on a number of factors, such as travel distance, historical travel times, and personal preference for major streets and freeways. These routes, however, may not be the same as the DUE routes when everyone is a habitual traveler, since now some of the travelers are adaptive travelers, PUE is no longer achievable. Let P_t^{rs} denote the set of those routes that habitual travelers departing at time t between O-D pair rs strictly follow (The generation of P_t^{rs} will be discussed in Section 6.1.2). The proportion of travelers who use a path $p \in P_t^{rs}$ in the group of habitual travelers, also known as the prescribed route rate (Pel et al. 2009), is assumed to follow the Logit model with respect to the generalized travel cost,

$$\xi_p(t) = \frac{\exp(-c_p(t))}{\sum_{p' \in P_t^{rs}} \exp(-c_{p'}(t))} \tag{6.1}$$

Therefore, the number of habitual travelers who depart at time t between O-D pair rs and use path $p \in P_t^{rs}$ is,

$$q_p(t) = (1-\theta) q_s^{rs} \xi_p(t) \tag{6.2}$$

For adaptive travelers, we assume that they always take their respective shortest path with respect to the instantaneous travel cost at each time interval. Adaptive travelers behave in a similar way as in the en-route route choice embedded in BUE, but the time period at which travelers update their shortest paths using the instantaneous travel cost can be relaxed from the assignment interval, ϕ_a in the BUE, to an arbitrary time interval in multiples of the loading time interval, i.e. $\gamma\phi_l$ (where γ is an integer). $\gamma\phi_l$ indicates how frequent the adaptive travelers are able to obtain up-to-date traffic information and choose an alternative route if necessary. It is easy to see that if all the travelers are adaptive travelers (i.e. $\theta = 1$) and let $\lambda = \phi_a/\phi_l$, then the hybrid route choice model essentially solves BUE.

It is crucial to define the instantaneous travel time of link a at entry time t, $\tau_a(t)$, which equals $l_a/s_a(t)$ where $s_a(t)$ is the instantaneous travel speed of link a at entry time t and l_a the length of link a. Given the density of link a at time t, $k_a(t)$, $s_a(t)$ is estimated by,

$$s_a(t) = \begin{cases} u_a & \text{if } k_a(t) \leq k_{a,m} \\ \dfrac{k_{a,j} - k_a(t)}{k_{a,j} - k_{a,m}} \dfrac{q_m}{k_a(t)} & \text{if } k_a(t) > k_{a,m} \end{cases} \tag{6.3}$$

where $k_{a,j}$ is the jam density of link a, $k_{a,m}$ the critical density and q_m the maximum flux of link a (also known as the capacity). If the density of link a at time t is smaller than $k_{a,m}$, then the instantaneous travel speed is the free-flow speed of link a, i.e. u_a; otherwise, it equals the division of the flux of link a at time t by its density at time t where the flux can be solved using the triangular fundamental diagram of link a given $k_a(t)$.

A DTA with this hybrid route choice model no longer requires an iterative solution procedure and the resultant flow pattern does not satisfy user equilibrium conditions in any sense. Instead, a one-shot DNL is applied to perform the assignment. During the DNL process, a shortest path calculation is needed in every assignment interval to obtain the new routes for those adaptive travelers, and the DNL is finished when every traveler reaches her destination.

6.1.2 Generating pre-trip route sets

We discuss in this subsection how those habitual travelers choose their pre-trip routes prior to the execution of the DTA.

An easy and reasonable way of generating pre-trip routes is to include the K shortest paths (with respect to free-flow travel time) in the set of pre-trip routes for each O-D pair rs, and then apply Equation 6.2 to stochastically assign some habitual travelers with one of those routes. This method is referred as “ordinary K shortest path (KSP) generation” in the rest of the chapter.

Though the ordinary KSP generation provides several alternative routes for habitual travelers, it usually does not result in a realistic route set because 1) those travelers may have a special affinity to the freeway and major arterial streets, and 2) they may use their travel experiences in addition to the free-flow travel times to determine their routes. To address these problems, we proposed two modifications to the the ordinary KSP generation method.

6.1.3 Road-hierarchy-based route set generation

Habitual travelers, being conservative, may prefer the freeway and major arterials over minor streets even if the latter may have shorter travel times. Some empirical observations seem to support this statement. For example, when a freeway is highly congested but both the on-ramp and off-ramp adjacent to the congested section are not used, one can bypass the congestion by first taking the off-ramp then use the on-ramp to get back to the freeway to save travel time. However, few travelers actually do this in reality, because most travelers would rather not bother to make extra efforts (such as changing lanes, switching to unfamiliar links, etc.) to reduce their delay unless this reduction is significant. Furthermore, when travelers select their own set of possible routes, they may mainly look at those major roadways and those minor streets are negligible. For instance, the least costly (first best) route choice of a traveler is a freeway connecting his house and office. Although a route consisting of the same freeway and with, however, other minor links from house/office to the freeway is theoretically his second least costly choice in terms of travel time/cost, he may take the route with the major arterial connecting his house and office as the second best choice. When the first best route is highly congested, one may consider the major arterial rather than the theoretical second best route. Now, various navigation tools also provide alternative routes that differ in terms of freeways and major arterials, rather than those that only differ in minor streets.

The above empirical evidence suggests that habitual travelers' route choice may be hierarchical, that is, they divide the network into multi-level networks, with freeways and major arterials constitute the high-level network and minor streets the lower-level network. They'll favor links in the high-level network over links in the lower-level network when they choose their routes.

To implement this route choice hierarchy, one can use free-flow speed and/or lane capacity to separate links into different levels (for example, 45 miles per hour and 3 lanes each way). An easy way of implementing such a two-level hierarchical route choice model is to weight significantly less on those major links than those minor links, so that the major links can attract more flow as compared to the case where links in both levels are weighted equally. Therefore, in a network with hierarchical roadway facilities (such as freeway, major arterials, minor arterials, collectors and so forth), we construct the following static travel cost for habitual travelers (on path p) when using KSP to generate their pre-trip route sets,

$$c_p^{rs} = \sum_{a \in A} \delta_{ap} \lambda_a c_a \tag{6.4}$$

where δ_{ap} is the link-route incidence indicator, equal to 1 if link a is on path p (equal to 0 otherwise). c_a is the generalized travel cost on link a. In particular, if only travel time is concerned, then c_a is the free-flow travel time of link a at time t. Let ϵ represent a small positive number. $\lambda_a = \epsilon < 1$ if link a is a major link and $\lambda_a = 1$ if link a is a minor link. This method is referred as "hierarchical KSP generation". The choice of the parameter ϵ is by trial-and-error so as to produce the DNL results that best match the reality. The effect of ϵ deserves further investigation, but it is beyond the scope of this dissertation.

6.1.4 PUE route set generation

The KSP generation methods discussed above use free-flow travel times to generate the route set for habitual travelers. This may be reasonable if the habitual travelers have no knowledge of actual travel times on those routes. In reality, however, those travelers often experience longer travel times on free-flow travel times due to traffic congestion, and will not determine their routes solely on free-flow travel times. In order to generate such a route set that incorporates historical traffic information, we can first perform a PUE-DTA and use the resultant routes as the pre-trip route set for habitual travelers. We assume the proportions of flow that are assigned to each route under the PUE will be applied to the group of habitual travelers. This method is referred as "PUE route set generation".

6.2 Queue Spillback in The Hybrid Route Choice

In this section, we use an example to show that under the CTM implementation of the LWR kinematic wave model, queue spillback is closely related to the route choice and the diversion ratio is crucial in determining queuing patterns. For simplicity, we assume the travel time is the sole factor in constructing the generalized travel cost.

The LWR model considers the physical space a vehicle takes, so queues in this model takes up space and can block traffic from entering a downstream link if space runs out on that link. Moreover, the LWR model describes queue growth in a more realistic way through its shock wave mechanism. Here we adopt the CTM implementation of the LWR model for traffic movement on links Daganzo (1994). For traffic flow through junctions, we use a general node traffic model given in Nie & Zhang (2007). It should be pointed out that node models play an important role in determining queuing spillback.

Suppose there are M approaches at an intersection and those approaches are marked as $1, 2, 3, ...M$. Flows leaving any exit will diverge to any other approaches and meanwhile, any approach also serves as a merge point of flow from any other approaches. Therefore, merges and diverges of all the approaches occur simultaneously. Let v_{ij} be the number of vehicles moving from approach i to j during a loading time interval. In order to describe traffic movements through intersections, we first define the demand (supply) of a link at time t, $D(t)$ ($S(t)$), as the maximum number of vehicles that are allowed to leave (enter) the link at t. Assume the vehicle proportion at upstream link i heading for downstream link j, a_{ij}, is known according to the vehicles in the last cell of link i under the LWR model, and demand and supply for each cell is computed based on the fundamental diagram. v_{ij} reads (Nie & Zhang 2007),

$$\text{virtual demand } \overline{vd}_i(t) = \min\left(D_i(t), \min_j\{\frac{S_j(t)}{a_{ij}(t)}\}\right) \tag{6.5}$$

$$\text{virtual supply } \overline{vs}_i(t) = \min\left(S_i(t), \sum_j a_{ij}(t) D_j t\right)$$

$$v_{ij}(t) = \min\left(\overline{vd}_i(t) a_{ij}(t), \overline{vs}_j(t) \frac{\overline{vd}_i(t) a_{ij}(t)}{\sum_k \overline{vd}_k(t) a_{kj}(t)}\right)$$

which is a generalization or streamlined version of several node models (Daganzo 1994, 1995, Lebacque 1996, Jin & Zhang 2003, 2004). As a matter of fact, Equation 6.5 does not strictly enforce FIFO at diverge nodes (Nie 2006), but since we do not pursue PUE in this chapter, the FIFO condition can be relaxed.

Now consider the network of Figure 6.1 (b), where travelers from link 0 heading for the destination can choose either link 1 or link 3 to go through an intermediate node. Link 1 has a lower free-flow travel time than link 3 and it is preferred by travelers under slight congestion. Link 2 serves as a bottleneck due to a lane drop and a queue may grow and back up to both link 1 and link 3. Figure 6.1 (a) gives the fundamental diagram of link 1 with the capacity C_m, free-flow speed u_f and jam density k_j.

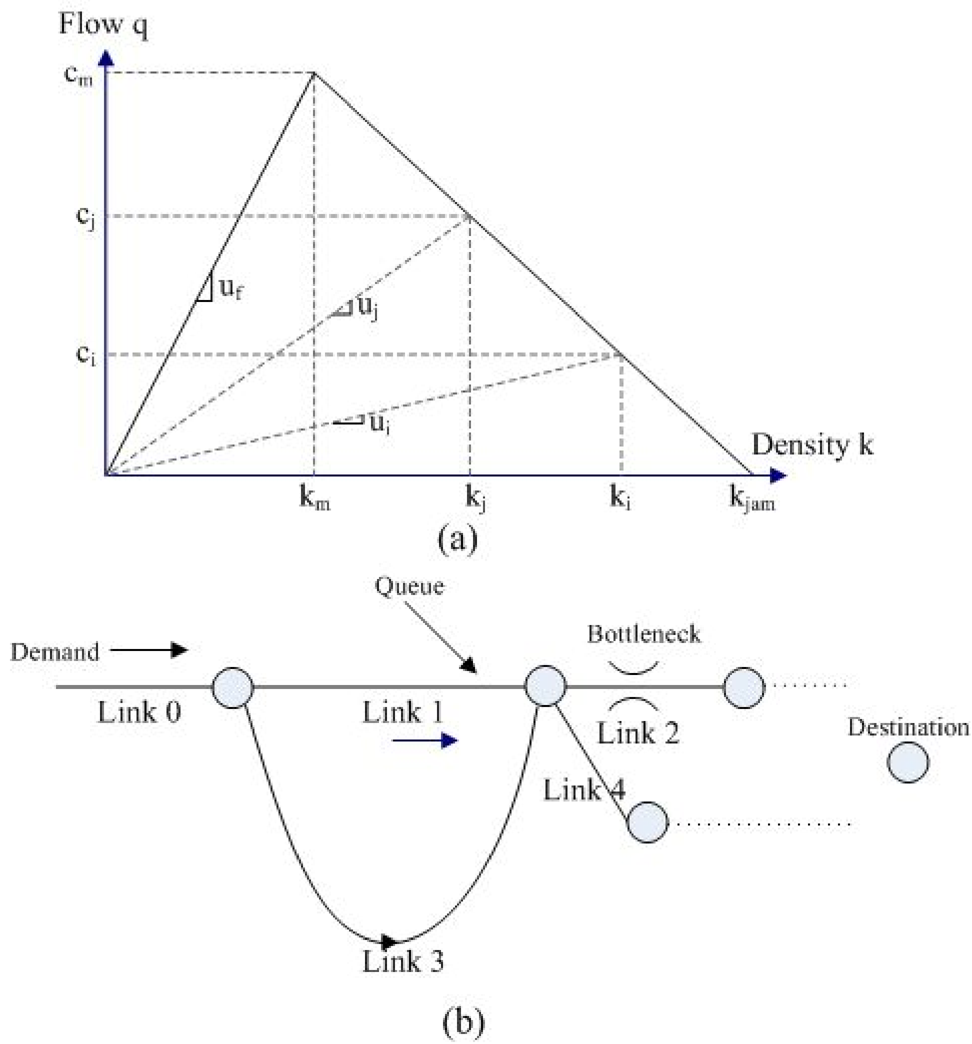


Figure 6.1. (a) The fundamental diagram of link 1; (b) A sample network

Due to the shorter free-flow travel times on link 1 than link 3, habitual travelers will take link 1 rather than link 3. There could be a case that all adaptive travelers also use link 1 regardless of the diversion ratio. To see this, suppose a worst case on link 1 where the outflow rates of link 1 achieves minimum when it equals to the capacity of link 2 (i.e. $C_i = C_{m,2}$) and link 4 is not used, so that the travel speed on link 1 achieves the minimum. Therefore, flow on link 1 is at density k_i and the traversal speed is u_i. Link 3 will never be used if its free-flow travel time is larger than the travel time on link 1, i.e.

$$\frac{l_1}{u_i} < \frac{l_3}{u_{f,3}} \tag{6.6}$$

Therefore, link 1 is over-saturated (i.e. the queue regulates its inflow) and a queue could spillover to upstream links. In this case, the queue may spill back upstream, and this can be unrealistic when link 3 is not that much longer than link 1 and in practice it may be an acceptable choice to some travelers. No diversion ratio can reduce such oversaturation under this setting.

Even though Equation 6.6 does not hold (i.e. link 3 offers relatively competitive free-flow travel time), there is still a case that the queue on link 1 spills over and link 3 is not used, if the diversion ratio is not properly set. If the diversion ratio is low, then most of demands will be loaded on the pre-trip route, i.e. link 0 to link 1 to link 2. It is easy to see that the queue spills back to link 0 due to the low diversion ratio. On the other hand, if the diversion ratio is high, most travelers are willing to switch to a route with shortest instantaneous travel time. Suppose both links 2 and 4 are used and link 2 is still a bottleneck link. In this case, the outflow rates of link 1 becomes C_j and $C_j > C_i$. Therefore, the instantaneous travel speed on link 1 is now u_j. Link 3 will never be used if its free-flow travel time is larger than the travel time on link 1, i.e.

$$\frac{l_1}{u_j} < \frac{l_3}{u_{f,3}} \tag{6.7}$$

Only if the diversion ratio is in a reasonable range, not too low and not too high, then both links 1 and 3 will be used and the queue on link 1 will not spill back. One can also find other examples to illustrate a similar phenomenon. Therefore, an appropriate diversion ratio is crucial in determining queuing patterns, and should be carefully calibrated in practice.

6.3 Numerical Experiments

In this section, we perform several numerical experiments to evaluate the proposed hybrid route choice model with different diversion ratios and with three methods of generating pre-trip route sets for the habitual travelers: ordinary KSP, hierarchical KSP and PUE route set generation. As a benchmark case, the PUE route choice will also be implemented and its results are used to assess the performance of the hybrid route choice model. The numerical experiments are carried out in two networks, one medium-size synthetic network and another large-size real network. We also assume the travel time is the sole factor in constructing the generalized travel cost.

6.3.1 A synthetic network

We synthesize a corridor network which consists of three residential areas and a Central Business District (CBD) connected by a three-lane freeway and two-lane major arterial road, as shown in Figure 6.2.

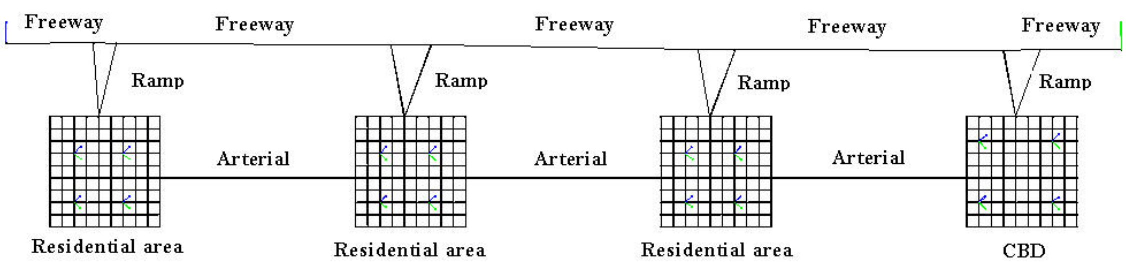


Figure 6.2. A synthetic corridor network

We assume in the morning peak hour (1 hour assignment horizon), travelers commute, either from the residential area or the external area (represented by the origin at the beginning of the freeway), to the CBD area. The external origin has a total demand of 400 vehicles to each of the residential areas, 1200 vehicles to the CBD and 100 vehicles to the external destination (at the end of the freeway). Each residential area also has a total demand of 400 vehicles to other residential areas, 1280 vehicles to the CBD and 200 vehicles to the external destination. Further, we allocate the total demands to 12 five minute assignment intervals following a trapezoidal flow pattern to mimic the flow profile in the peak hour.

Each residential area and the CBD area are represented by 9×9 grid subnetworks and four origins and four destinations. Each link of the grid subnetworks stands for a minor local street and is set to have an identical capacity of 1500 vehicles per hour and a speed limit of 25 mile per hour, but its length is randomly chosen from a uniform distribution within $[0.08, 0.12]$ (miles). The grid subnetwork connects the origins and destinations to the freeway ramps and the major arterial road, so that travelers have choices of choosing either the freeway or the arterial to go to the CBD in the morning commute. There are in all 478 nodes, 1528 links and 129 O-D pairs. Each segment of the freeway and the arterial has the same length, but the freeway has a total capacity of 6000 vph and 65 mph speed limit, while the arterial road has a total capacity of 3600 vph and a speed limit of 40 mph. For each segment, the free-flow travel time via the ramps and the freeway and via the arterial road is set to be 3 min and 5 min, respectively, so that the freeway is a preferred route and the arterials may also be used if the freeway is congested.

For all the scenarios studied, the loading time interval is 5 sec, and adaptive travelers update their travel information/routes every 90 seconds. The diversion ratio is set to be 0.2, 0.5 or 0.8 to represent a low, medium or high proportion of adaptive travelers. For those habitual travelers, we let $K = 2$ in the KSP computation. The results in total travel cost and total delay[1] for all scenarios are shown in Table 6.1.

	TTT(hr)	TD(hr)
PUE	812	166

	Diversion ratio=0.2		Diversion ratio=0.5		Diversion ratio=0.8	
	TTT(hr)	TD(hr)	TTT(hr)	TD(hr)	TTT(hr)	TD(hr)
PUE generation	749	104	842	191	935	261
Ordinary KSP generation	1719	1070	961	302	992	312
Hierarchical KSP generation	1514	843	980	302	1007	321

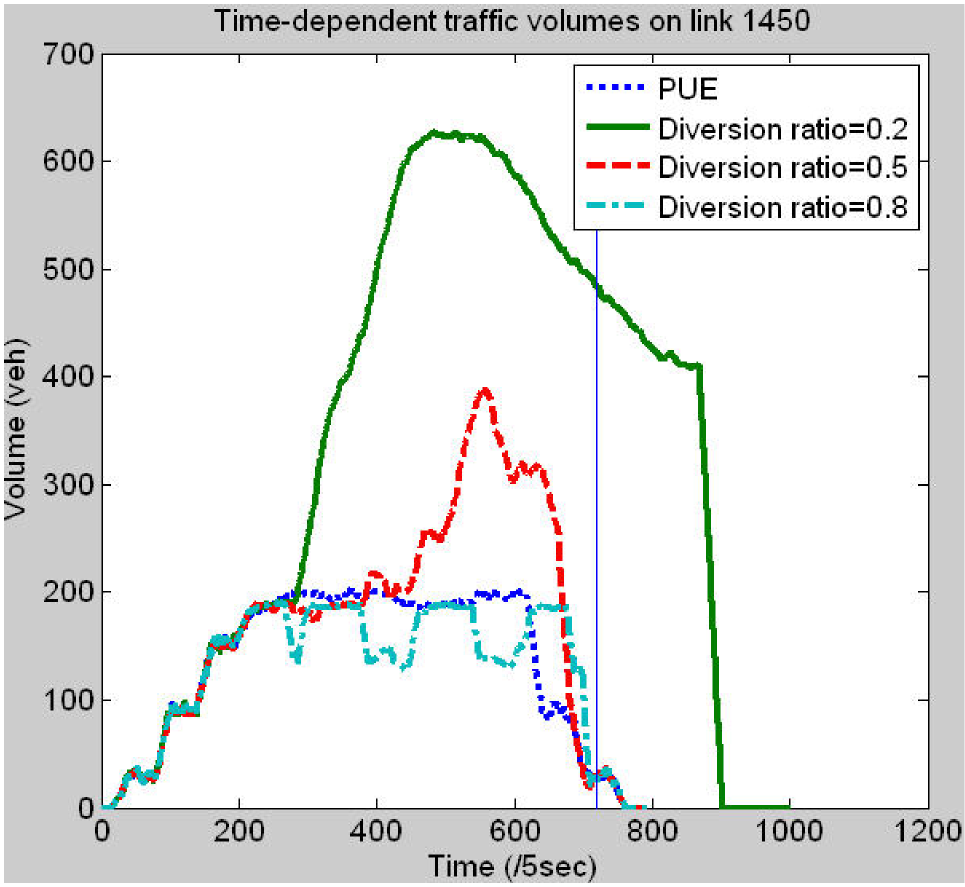
Table 6.1. The total travel time and total delay of selected scenarios (TTT: Total travel time; TD: Total delay)

[1]Total delay equals the total vehicle-hour-traveled (VHT), subtracted by the total free-flow travel time of all vehicle trips taking the same routes as where VHT is computed

6.3.1.1 The effects of diversion ratio

As seen from Table 6.1, under the KSP route set generation, a 0.2 diversion ratio leads to significantly larger total travel time and total delay than a 0.5 or 0.8 diversion ratio. On the other hand, a high diversion ratio yields slightly higher total free-flow travel time (which equals $TTT - TD$) than a low diversion ratio, because more travelers may deviate to a longer route with less congestion under a higher diversion ratio. Interestingly, a 0.5 diversion ratio yields less queuing delay then a 0.2 or 0.8 diversion ratio, which may be explained by the following: 1) as explained in Section 6.2, a medium diversion ratio can prevent queuing spillback under certain conditions; and 2) intuitively, a diversion ratio, if too small, can lead to severe queuing on the freeway due to the high demand of habitual travelers. If the diversion ratio is set to be too high, then compared to a medium diversion ratio, some bottleneck links are used more intensively by adaptive travelers after they update their routes. This explains why in practice, a moderate diversion ratio may produce a flow pattern with less queuing delay than a too high or a too low one.

We plot in Figure 6.3 and 6.4 the time-varying volumes on a freeway link and an arterial link in the middle section against three diversion ratios under the ordinary KSP route set generation. All four scenarios yield approximately the same time-varying volumes in the first 200 loading intervals (around 20 min). This is because when the network is not congested or mildly congested, travelers within the same O-D pair, regardless of their willingness to switch routes, will use the same routes that are preferred in terms of free-flow travel time. However, the volumes of four scenarios start to differ significantly after the network is loaded with high demands. Under the low diversion ratio, a large percentage of travelers will still follow the freeway, which leads to much more severe congestion on the freeway than the case of a high diversion ratio where most travelers will switch routes and the case of PUE. The freeway link cumulates up to 600 vehicles when the diversion ratio is 0.2, as compared to 400 vehicles when the diversion ratio 0.5, and 200 vehicle when it is 0.8 or in PUE. The low diversion ratio also results in a long loading tail after the assignment horizon ends, i.e. the network does not clear up until 25min after the assignment horizon ends, while the network clears up within 6 minutes in all other scenarios.

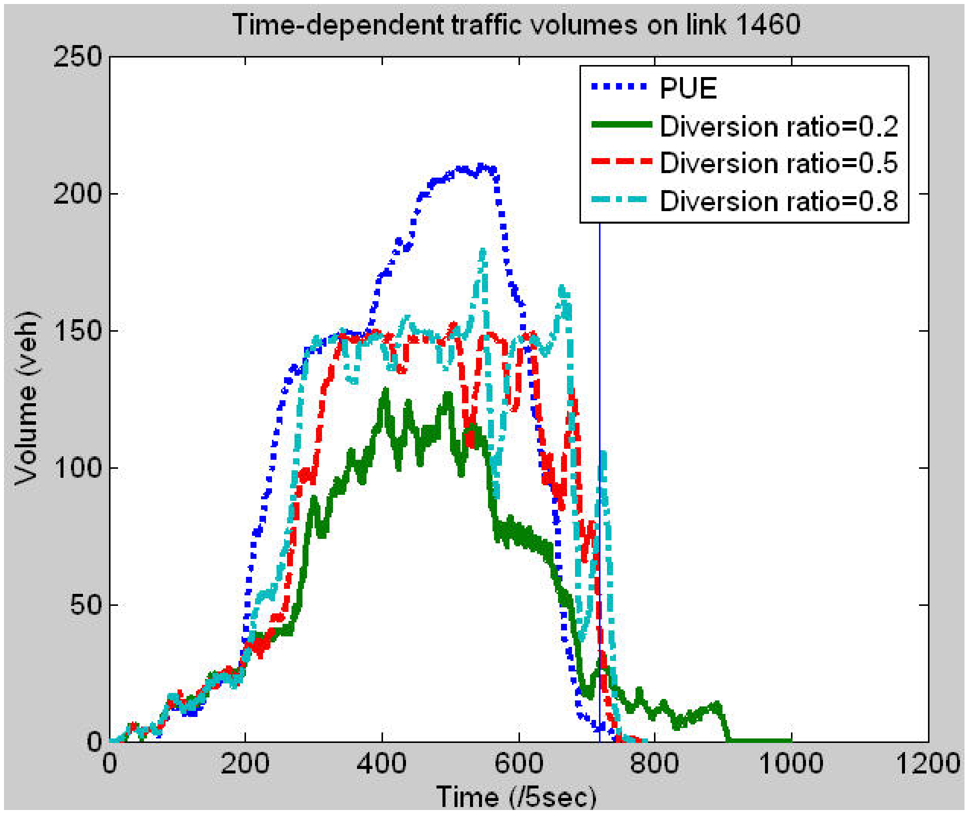


Note: The vertical line indicates the time when the assignment horizon ends

Figure 6.3. Time-varying volumes on a freeway link w.r.t. different diversion ratio

Generally, of all four scenarios, PUE yields the highest share of arterial road usage, while the low diversion ratio has the lowest share. By checking the cumulative number of vehicles on the arterial link, PUE almost triples the number of travelers who use the arterial as compared to the case of 0.2 diversion ratio. Because PUE assumes all travelers can predict the network condition after day-to-day experience, such a user optimum leads to less congestion and higher usage of the arterial than the hybrid route choice model. For the hybrid route choice model, a medium diversion ratio has substantially higher arterial share than a low one, slightly higher arterial share than a high diversion ratio. This is consistent with the result that a medium diversion ratio produce less overall congestion than a high or low one.

As can be seen from Figure 6.4, compared to the PUE, the hybrid route choice model yields less volumes on the arterial road during the time (from 200th interval to 240th



Note: The vertical line indicates the time when the assignment horizon ends

Figure 6.4. Time-varying volumes on a arterial link w.r.t. different diversion ratio

interval) when the congestion on the freeway starts to propagate quickly, and it also yields more volumes during the time (the 610th interval to the end) when the queue on the freeway starts to dissipate. This is because in PUE, travelers can predict the actual traffic congestion on all links and can therefore better respond to the congestion propagation. As for the hybrid route choice model, because travelers make en-route route choices and their prediction is based on instantaneous travel time, they may underestimate the congestion on the freeway when the queue builds up and overestimate the congestion when the queue starts to dissipate. Thus, compared to PUE, the route diversion in the hybrid route choice model always lags behind the time they would divert to optimize their travel cost in response to the actual congestion propagation, and obviously such a route choice may be far from user optimal.

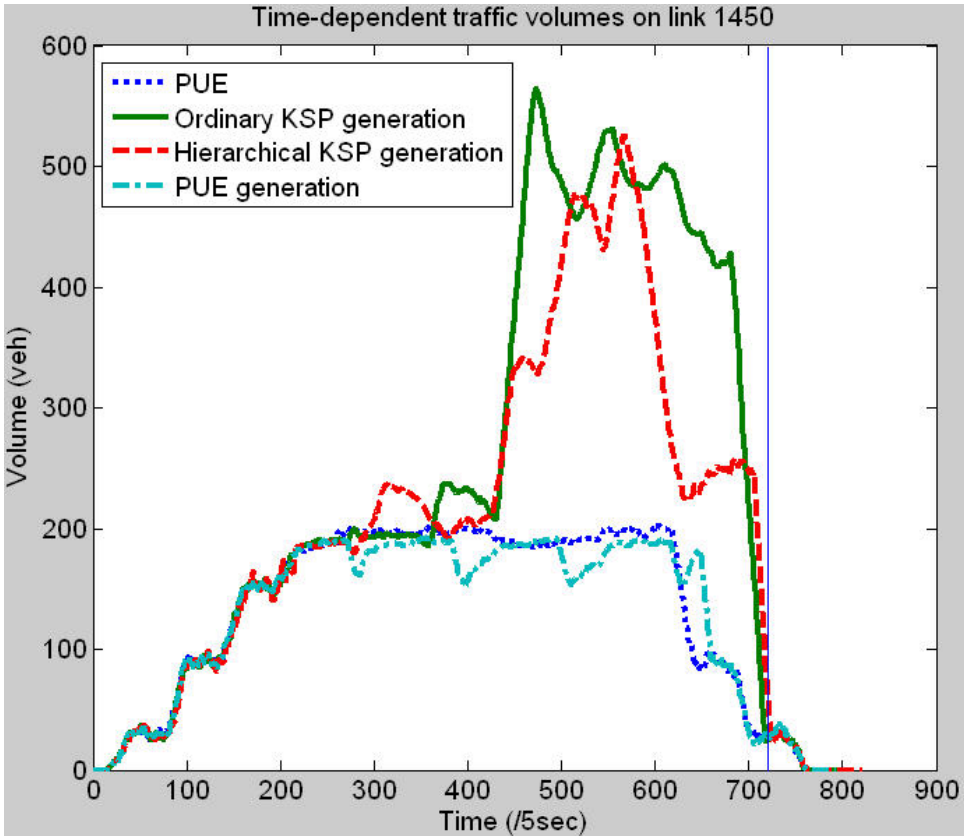
We also observe a large fluctuation of volumes on both the freeway and the arterial road when the diversion ratio is set to be high. Because travelers will update their current traffic information every 18 loading time intervals, if the diversion ratio is high, then a large number of adaptive travelers will deviate to a certain route (e.g. the arterial road) at the beginning of certain time intervals and result in unexpected congestion of the new route. This congestion may spread quickly and makes the freeway route advantageous again, which leads a large number of adaptive travelers to again switch back to the freeway route. Thus, under a high diversion ratio, a large number of travelers intermittently choose either the freeway or the arterial, and that is why we see a large fluctuation of volumes on both links.

6.3.1.2 The effects of pre-trip routes

We plot in Figure 6.5 and 6.6 the time-varying volumes and travel times on a freeway link in the middle section against the three methods of generating pre-trip route sets for habitual travelers where the diversion ratio is set to be 0.5. The weighting factor on the major links (freeway links and arterial links) are set to be $\lambda_a = 0.01$ in the hierarchical KSP calculation.

Generally, if half of the travelers are aware of the historical information and follow their pre-trip routes, then time-varying flows on this freeway link is close to the results obtained from PUE. However, if those travelers only use the free-flow travel time to determine their pre-trip routes so that the freeway is overwhelmingly used, then the freeway could be far more congested than the case where the historical information is used to determine the pre-trip routes, particularly after the 400th time interval where the congestion becomes severe on the freeway. Compared to the maximum 60 seconds in the case of PUE or the case where the historical information is used by habitual travelers, using KSP to generate pre-trip route sets yields up to 135 seconds travel time on the freeway link, and thus the TTT and TD of the network are significantly larger as well.

An interesting result is that using the hierarchical KSP generation indeed significantly reduce the queuing as compared to the case with only the ordinary KSP generation. When such a hierarchical way of generating pre-trip routes is considered, habitual travelers commuting from the residential area to the CBD have choices of either the freeway or the arterial road. However, if no hierarchical road preference is incorporated, then the first two shortest paths for most O-D pairs will always be such that the freeway

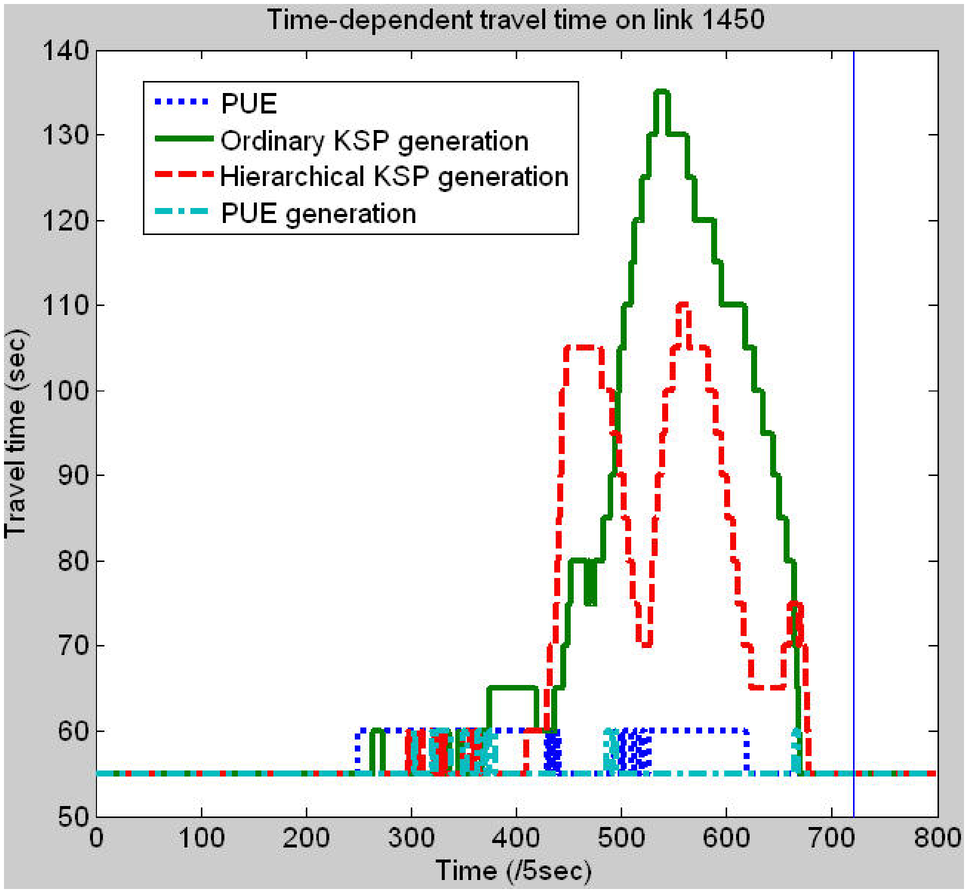


Note: The vertical line indicates the time when the assignment horizon ends

Figure 6.5. Time-varying volumes on a freeway link w.r.t. different methods of generation pre-trip route sets

is chosen in both routes with distinctions on some local streets. Therefore, by using the hierarchical KSP to generate the route sets, more habitual travelers will use the arterial road instead of the freeway than the ordinary KSP generation.

As seen from Table 6.1, the PUE route generation with a 0.2 diversion ratio has less TTT and TD than PUE. This indicates that if we only provide real-time information to a small proportion of travelers and meanwhile assume that most of travelers use the historical information to travel, then this may considerably benefit the network. However, if the real-time information is offered to a large number of the travelers, then the network performance may be worse off. This is because the optimal route at the current time interval, if used by an excessive number of travelers, may build up an unexpected queue and sometimes this is not desirable.



Note: The vertical line indicates the time when the assignment horizon ends

Figure 6.6. Time-varying travel times on a freeway link w.r.t. different methods of generation pre-trip route sets

6.3.2 The Sacramento Metropolitan Area Network

We also applied the hybrid route choice model to a large network which covers the Sacramento metropolitan area. The network consists of 2556 nodes, 7221 links and 729 O-D pairs. We collected 5 min traffic counts for a period of 24 hours on 25 segments of major freeways and 32 major arterials on the periphery of downtown Sacramento. The time-varying counts are then used to estimate a morning peak 6-hour (6:00am-12:00pm) time-dependent O-D demands by a logit path flow estimator algorithm Bell et al. (1997). Other parameters are: loading time interval 10 seconds; $K = 3$ in the KSP calculation; and the traffic information will be updated every 5 minutes. The weighting scaler on the major links is set to be 0.2 in the hierarchical KSP route generation. Table 6.2 shows the average travel time, average travel delay and average travel distance using different diversion ratios and pre-trip route generation methods[2].

	DR	0.3	0.4	0.45	0.5	0.55	0.6	0.7
Ordinary KSP generation	ATT(min)	G	41.53	38.87	37.74	35.61	G	G
	AD(min)	G	15.65	12.87	11.58	9.36	G	G
	ADS (mile)	G	25.49	25.49	25.46	25.51	G	G
Hierarchical KSP generation	ATT(min)	63.93	47.13	42.77	38.18	37.16	37.58	G
	AD(min)	37.67	20.92	16.49	12.01	10.84	10.96	G
	ADS (mile)	25.72	25.67	25.70	25.60	25.61	25.67	G

Table 6.2. The total travel time and total delay of the Sacramento network w.r.t DR (DR: Diversion ratio; ATT: Average travel time; AD: Average travel delay; ADS: Average travel distance; G: Gridlock)

As discussed before, the diversion ratio, if set too high or too low, may cause serious queuing. In this case, the DTA procedure terminates without gridlock only when the diversion ratio is within a certain range, i.e., from $0.35 \sim 0.55$ if the ordinary KSP generation is used, or from $0.25 \sim 0.65$ if the hierarchical KSP generation is used. The wider range of the acceptable diversion ration provided by the hierarchical KSP generation indicates that it can better prevent excessive concentration of queuing than the ordinary KSP generation. Overall a 0.55 diversion ratio looks reasonable in the sense that the resultant ATT and AD and time-varying link flow on designated links (25 segments of freeways and 32 major arterials) approximately match the actual observation. When the diversion ratio is less than 0.55, less travelers are willing to switch routes and travelers are subject to more queuing delay on average. In contrast, when the diversion ratio is larger than 0.55, more travelers respond to real-time information and switch routes, but this may cause excessive queuing on certain routes or links due to herding, which may even produce a gridlock.

If hierarchical KSP is used to generate the route sets, then for most O-D pairs, the second or the third shortest path is more likely to include the arterial links as an alternative to the preferred freeway route, as compared to the ordinary KSP generation. It assigns more travelers on the major arterials, which tends to distribute the flow more evenly for those major links (i.e. those weighed less in computing shortest paths, freeway or major arterials) and prevent excessive queuing that leads to gridlock. Given the same

[2]We do not include the results from PUE because a gridlock occurs during the iterative procedure of DNL, and solving the issue of gridlock in the DTA algorithm is beyond the scope of the dissertation.

diversion ratio, using hierarchical KSP always yields longer average distance and larger average delay and cost than the ordinary KSP. This is because hierarchical KSP induces travelers to travel more on major links rather than the minor links (e.g. local streets), which generally take slightly longer distances. In addition, under hierarchical KSP, it is more likely that flows are concentrated on those major links with fewer minor links assigned with flows, which results in a larger queuing delay, as compared to the ordinary KSP.

6.4 Summary

Travelers' route choice behavior in real life is neither Boston UE or Predictive UE. Rather, the route choice may be such that they used both real-time information and historical experience. This chapter proposes a hybrid route choice model where travelers are divided into two groups, habitual travelers and adaptive travelers. Habitual travelers strictly follow their pre-trip routes. We speculate that route choice of habitual travelers may be such that the major links, such as the freeway or major arterials, are favored over minor road links and they also consider historical traffic information based on their day-to-day driving experiences. Therefore, we propose two new methods of generating their pre-trip route sets, hierarchical K shortest path and Predictive-UE generation. On the other hand, adaptive travelers are responsive to real-time information and willing to change routes. The hybrid model may be more realistic in describing travelers' route choices in regards to travel costs. Unlike the PUE, the new hybrid model requires only one shot dynamic network loading.

We study how the choice of diversion ratio and generation methods for habitual travelers' route set affect the resulting flow and queuing patterns and found that,

- The hybrid model is easy to calibrate and can work efficiently for large-scale networks. It is likely to produce realistic results as indicated by the large-scale numerical example.

- A medium diversion ratio can reduce excessive concentration of queuing compared to a high or low diversion ratio. If it is set too high or too low, then queues tend to concentrate on certain links, which can eventually lead to gridlock.

- In most cases, PUE yields significantly less congestion and queue spillovers than the hybrid route choice model, because travelers in the former can anticipate what would happen for their entire trip and hence can make better choices, while in the latter the adaptive travelers make myopic choice decisions based on the prevailing traffic conditions that may prove to be a bad choice as traffic conditions change.

- Pre-trip routes generated by PUE (i.e. using historical information) for habitual travelers tend to yield considerably less delay and queues in the network than by ordinary KSP which only uses free-flow travel time information. Pre-trip routes generated by hierarchical KSP may reduce queuing compared to those generated by ordinary KSP, thanks to the former's ability to spread demand among freeways and major arterials.

- When habitual travelers use historical information to determine their pre-trip routes, offering real-time information to a small portion of travelers can actually achieve better network performance than PUE where all travelers are assumed to have perfect information all the time. However, if the majority of travelers are offered with real-time information, network performance may actually become much worse than offering no one real-time information.

It is hoped that these findings can help practitioners choose a route choice model and calibrate its parameters against real data. While more experiments need to be carried out to confirm these initial findings, it does seem that the use of historical information, hierarchial network, and moderate level of diversion can help spread congestion and avoid excessive concentration of queuing and gridlock.

Chapter 7

Conclusions

This dissertation investigates traveler heterogeneity for DTA from four perspectives, travelers' attributes (in value of time and value of schedule delay), modal choice preference, parking choice preference and route choice preference. We derived analytical DTA solutions in simplified networks, particularly in the context of the morning commute problem, and analyzed the effects of various factors to both the flow solutions and network performance. Some intriguing findings are obtained.

We first study the morning commute problem with a heterogeneous traveling population whose early/late arrival penalty parameters are continuously distributed. The ratio of Value of Early Schedule Delay (VESD) over Value of Late Schedule Delay (VLSD) is assumed to be constant across the population. In the context of the morning commute problem, assuming homogeneity overestimates the queuing delay and thus the total travel time. The travelers who weigh schedule delay high will first shift to the arterial road, prompted by an increase in total demand. However, the critical travel demand at which travelers start to use the arterial road remains the same if the common EAP for homogeneous users is equal to the expected value of EAP distribution for heterogeneous users. Interestingly, there exists a critical value of EAP parameter, such that all the travelers whose EAP is less than it choose the freeway, and those whose EAP is larger than it choose the arterial road. Sensitivity analysis on total travel time (TTT) indicate that enlarging freeway (AR) capacity will always reduce the TTT of the whole network and the TTT of the AR (freeway), and increase the demand share of the freeway (AR). We also show that every commuter is better off if either the freeway capacity or the AR capacity is enlarged.

Second, we study the morning commute problem with three modes: transit, driving alone and carpool. The transit mode uses its own separate guideway, but the auto modes can access two parallel routes to reach the destination, a freeway and an arterial road (AR). Carpoolers are assumed to share their fuel costs and road tolls, in addition to their advantage of using specially provided facilities (HOV lanes). However, there is an added cost of carpooling: the cost associated with pick-up and drop-off. We analyze the interactions among the three modes and how different factors affect their mode shares and network performance. This is achieved by first deriving the departure time equilibrium for the transit mode (in the same fashion as is done for the auto modes), then establishing equilibrium within each mode. The shares among the modes are determined by a nested logit model. Finally, a time-varying toll on the freeway is proposed to completely eliminate the congestion on the freeway. Some intriguing findings of this study include, 1) enlarging HOV facilities, which offers a travel advantage for carpoolers over solo-drivers, may reduce transit ridership and increase auto travel, and it does not necessarily reduce the total travel cost on the network when the network is highly congested; 2) the rise of gas price may first entice auto commuters to carpool. But as the gas price increases further, both carpoolers and solo-drivers will eventually switch to use the transit mode; and 3) a flat freeway toll is capable of reducing the total network travel cost.

Furthermore, we investigate how parking regulations can be implemented to mitigate traffic congestion, as well as to improve the system performance. The parking lots in the downtown area of a city are first abstracted to two parking clusters (areas) according to their distances to travelers' destination. Dependent on parking fees and access times (or accessibility) of parking clusters, there are five types of parking location preference in the morning commute. Travel patterns under different parking capacities, parking fees and accessibility to the destination are derived. We then discuss the parking management in two cases.

One case is where all the parking lots are publicly owned. Our analysis indicates that under certain conditions, enlarging the central parking lots is not desirable. In terms of total travel cost, an inward or hybrid parking preference and a shorter access time are always preferred. Finally, we derive the optimal parking fee, capacity and access time which altogether yield the minimum total travel cost. When the closer parking cluster does not have too large an accessibility advantage over the farther one, the optimal travel profile is such that both parking clusters should be used. As a result, the optimal parking setting can effectively reduce both the system cost and the queuing delay. Even more intriguing is that, compared to the case without parking choices, all travelers are better off under the optimal parking setting, which cannot be realized by the system-optimal dynamic toll scheme. The other case is that private operators own the parking lots and they compete each other to attract travelers and maximize their profits. We found only one stable solution for the competitive parking market. Under it, each parking area is preferred by the commuters during certain time periods. Compared to the case without parking choice, provision of parking through a competitive market is able to reduce commuters' travel cost and queuing delay, but it does not necessarily lead to the most desirable market outcome that minimizes social cost or commuter cost. This issue can be addressed through market regulations, such as price-ceiling, capacity-floor or capacity-ceiling, and a quantity tax/subsidy regulation. It is found that both price-ceiling and quantity tax/subsidy regulations can efficiently reduce both the system cost and commuter cost under certain conditions, and help ensure the stability of the parking market. Overall, parking is capable of managing traffic efficiently to improve the system performance and to reduce the congestion.

We finally extend our research to the DTA in general networks. We focused our scope to the traveler heterogeneity in their preference on route choices. Previous studies usually suggest network user equilibrium or user optimum as the route choice model, which is an ideal state that can hardly be achieved in real traffic. More often than not, every day traffic tends to be in disequilibrium rather than equilibrium, thanks to uncertainties in demand and supply of the network. We propose a hybrid route choice model for studying non-equilibrium traffic. It combines pre-trip route choice and en-route route choice to solve dynamic traffic assignment (DTA) in large-scale networks. Travelers are divided into two groups, habitual travelers and adaptive travelers. Habitual travelers strictly follow their pre-trip routes which can be generated in the way that major links, such as freeways or major arterial streets, are favored over minor links, while taking into account historical traffic information. Adaptive travelers are responsive to real-time information and willing to explore new routes from time to time. We apply the hybrid route choice model in a synthetic medium-scale network and a large-scale real network to assess its effect on the flow patterns and network performances, and compare them with those obtained from Predictive User Equilibrium (PUE) DTA. The results show that PUE-DTA usually produces considerably less congestion and less frequent queue spillback than the hybrid route choice model. The ratio between habitual and adaptive travelers is crucial in determining realistic flow and queuing patterns. Consistent with previous studies, we found that supplying a small number of travelers real-time information is more beneficial to network performance than supplying the majority of travelers with real-time information. Finally, some suggestions are given on how to calibrate the hybrid route choice model in practice.

There are several other issues concerning heterogeneity in morning commute problem that are worthy of further investigation, for instance, when both traffic modal choice and VOT/VOS heterogeneity are concerned, it is of interest to know how a heterogeneous population will be split in terms of modal choices and route shares. Would a pricing scheme, such as transit fare, gas fee, carpool impedance etc., improve the system performance and meanwhile achieves Pareto-improvement? When toll schemes other than the system optimal toll is implemented, it is very likely that the orders of departure time will change as well. It would be interesting to find out what kind of performance improvement such schemes can achieve and who will benefit most/least from them.

While the proposed parking model yields interesting results, it made many simplifying assumptions that we hope to address in our future research. First, we did not consider commuters' time spent on searching for parking spaces. The search for available parking spaces constitutes a wasteful commuting component that contributes to congestion. Future work should take into account the search time into travelers's commute cost. Second, in the real world, parking is supplied by both private firms and public entities. It would be interesting to study this mixed market and compare it with the two extreme cases of either private-only or public-only parking provision. Third, it would be of particular interest to extend this model to consider other traffic modes, for instance transit. Travelers may be able to avoid the congestion at the bottleneck by taking transit at park-and-ride stations. A combination of transit fares and parking fees can be used to achieve a desired market outcome.

Our future work will turn attention to modeling heterogeneity in general networks. There seems a great need to build a realistic DTA framework for general networks incorporating traveler heterogeneity in vehicle attributes, traveler attributes, modal choices and parking choices. With such an analysis model, we are able to evaluate in real networks a variety of traffic operations (such as transit subsity/fare, capacity enlargement, congestion tolls, parking management, work zones and other operational schemes), to estimate or predict network traffic delay, Vehicle-Miles-Traveled/Vehicle-Hours-Traveled (VMT/VHT) and emissions, and to assess social equity issues.

Appendix A

Sensitivity Analysis of Capacity Improvement for Homogeneous Travelers

Based on Arnott et al. (1990)'s derivations, we compute the impact on the total travel time (TTT) caused by the roadway capacity improvement in a two-route network under a homogeneous population. $dTTT/ds_f$, $dTTT/ds_a$ and $dTTT/d\tau$ represent the system time savings of marginal improvement of freeway capacity, arterial road capacity, arterial road free-flow travel time, respectively. Let TTT_f and TTT_a denote the total travel time on the freeway and the arterial road, respectively. Assuming that both the freeway and the AR are used by the commuters, we have $N - s_f \frac{\tau}{\delta\beta_0} > 0$.

$$
\begin{aligned}
TTT_f &= \frac{\delta\beta_0}{2}\frac{s_f}{(s_f+s_a)^2}(N + s_a\frac{\tau}{\delta\beta_0})^2 \\
TTT_a &= \frac{\delta\beta_0}{2}\frac{s_a}{(s_f+s_a)^2}(N - s_f\frac{\tau}{\delta\beta_0})^2 + \frac{s_a}{s_f+s_a}(N - s_f\frac{\tau}{\delta\beta_0})\tau \\
\frac{dTTT}{ds_f} &= -\frac{\delta\beta_0}{2(s_a+s_f)^2}(N + s_a\frac{\tau}{\delta\beta_0})^2 < 0 \\
\frac{dTTT_f}{ds_f} &= \frac{\delta\beta_0}{2}\frac{s_a - s_f}{(s_f+s_a)^3}(N + s_a\frac{\tau}{\delta\beta_0})^2 \\
\frac{dTTT_a}{ds_f} &= -\frac{s_a\delta\beta_0}{(s_f+s_a)^3}(N + s_a\frac{\tau}{\delta\beta_0})^2 < 0
\end{aligned}
$$

Under homogeneous commuters, improving freeway capacity will always reduce the TTT of the whole network and the TTT_a. However, whether or not TTT_f increases is only dependent on s_a and s_f.

$$\begin{aligned}
\frac{dTTT}{ds_a} &= -\frac{\delta\beta_0}{2(s_a+s_f)^2}(N - s_f\frac{\tau}{\delta\beta_0})^2 < 0 \\
\frac{dTTT_f}{ds_a} &= -\delta\beta_0\frac{s_f}{(s_f+s_a)^3}(N + s_a\frac{\tau}{\delta\beta_0})(N - s_f\frac{\tau}{\delta\beta_0}) < 0 \\
\frac{dTTT_a}{ds_a} &= \frac{\delta\beta_0}{2(s_f+s_a)^3}(N - s_f\frac{\tau}{\delta\beta_0})(Ns_f + \frac{\tau}{\delta\beta_0}s_f^2 + \frac{3\tau}{\delta\beta_0}s_f s_a - Ns_a)
\end{aligned}$$

Under homogeneous commuters, improving AR capacity will always reduce the TTT of the whole network and the TTT_f. However, whether or not TTT_a increases is dependent on all the parameters.

$$\begin{aligned}
\frac{dTTT_a}{d\tau} &= \frac{s_a}{(s_f+s_a)^2}(Ns_a - s_f^2\frac{\tau}{\delta\beta_0} - 2s_f s_a\frac{\tau}{\delta\beta_0}) \\
\frac{dTTT_f}{d\tau} &= \frac{s_f s_a}{(s_f+s_a)^2}(N + s_a\frac{\tau}{\delta\beta_0}) > 0 \\
\frac{dTTT}{d\tau} &= \frac{s_a}{s_f+s_a}(N - s_f\frac{\tau}{\delta\beta_0}) > 0
\end{aligned}$$

Under homogeneous commuters, reducing the free-flow travel time on the AR will always reduce the TTT of the whole network and the TTT_f. However, whether or not TTT_a increases is dependent on all the parameters.

Appendix B

Sensitivity Analysis of Capacity Improvement for Heterogeneous Travelers

We derive the impact on the total travel time caused by roadway capacity improvement.

In order to solve for $dTTT/ds_f$, $dTTT/ds_a$ and $dTTT/d\tau$, we first define a function $g(a', s_f, s_a, \tau)$,

$$g(a', s_f, s_a, \tau) = N\frac{\delta}{s_f}\int_a^{a'} xf(x)dx - \tau - N\frac{\delta}{s_a}a'\int_{a'}^{b} f(x)dx$$

According to Equation 3.25, $g(a', s_f, s_a, \tau) = 0$. We also define $\lambda = (\frac{1}{s_f} + \frac{1}{s_a})a'f(a') - \frac{1}{s_a}\int_{a'}^{b} f(x)dx > 0$ by Equation 3.26. Hence,

$$\frac{\partial a'}{\partial s_f} = -\frac{\frac{\partial g(\cdot)}{\partial s_f}}{\frac{\partial g(\cdot)}{\partial a'}} = \frac{\frac{1}{s_f^2}\int_a^{a'} xf(x)dx}{\lambda} > 0$$

Similarly,

$$\frac{\partial a'}{\partial s_a} = -\frac{\frac{\partial g(\cdot)}{\partial s_a}}{\frac{\partial g(\cdot)}{\partial a'}} = -\frac{\frac{1}{s_a^2}a'\int_{a'}^{b} f(x)dx}{\lambda} < 0$$

$$\frac{\partial a'}{\partial \tau} = \frac{1}{\lambda} > 0$$

Therefore, the derivatives with respect to freeway capacity are:

$$\begin{aligned} \frac{dTTT_f}{ds_f} &= \frac{\partial TTT_f}{\partial a'}\frac{\partial a'}{\partial s_f} + \frac{\partial TTT_f}{\partial s_f} \\ &= \frac{N^2\frac{\delta}{s_f^3}f(a')(\int_a^{a'} xf(x)dx)^2}{\lambda} - \frac{N^2\delta}{s_f^2}\int_a^{a'} f(\beta)(\int_a^{\beta} xf(x)dx)d\beta \end{aligned} \tag{B.1}$$

$$\begin{aligned} \frac{dTTT_a}{ds_f} &= \frac{\partial TTT_a}{\partial a'}\frac{\partial a'}{\partial s_f} + \frac{\partial TTT_a}{\partial s_f} \\ &= -\frac{N^2\frac{\delta}{s_f^3}f(a')(\int_a^{a'} xf(x)dx)^2}{\lambda} < 0 \end{aligned} \tag{B.2}$$

$$\begin{aligned} \frac{dTTT}{ds_f} &= \frac{dTTT_f}{ds_f} + \frac{dTTT_a}{ds_f} \\ &= -\frac{N^2\delta}{s_f^2}\int_a^{a'} f(\beta)(\int_a^{\beta} xf(x)dx)d\beta < 0 \end{aligned} \tag{B.3}$$

$$\frac{db_f}{ds_f} = \frac{\partial b_f}{\partial a'}\frac{\partial a'}{\partial s_f} = f(a')\frac{\partial a'}{\partial s_f} > 0 \tag{B.4}$$

The derivatives with respect to AR capacity are:

$$\begin{aligned} \frac{dTTT_f}{ds_a} &= \frac{\partial TTT_f}{\partial a'}\frac{\partial a'}{\partial s_a} + \frac{\partial TTT_f}{\partial s_a} \\ &= -\frac{N^2\frac{\delta}{s_a^2 s_f}a'f(a')(\int_a^{a'} xf(x)dx)(\int_{a'}^{b} f(x)dx)}{\lambda} < 0 \end{aligned} \tag{B.5}$$

$$\begin{aligned} \frac{dTTT_a}{ds_a} &= \frac{\partial TTT_a}{\partial a'}\frac{\partial a'}{\partial s_a} + \frac{\partial TTT_a}{\partial s_a} \\ &= \frac{N^2\frac{\delta}{s_a^2 s_f}a'f(a')(\int_a^{a'} xf(x)dx)(\int_{a'}^{b} f(x)dx)}{\lambda} - \frac{N^2\delta}{s_a^2}\int_{a'}^{b} f(\beta)(\int_{a'}^{\beta} xf(x)dx)d\beta \end{aligned} \tag{B.6}$$

$$\begin{aligned} \frac{dTTT}{ds_a} &= \frac{dTTT_f}{ds_a} + \frac{dTTT_a}{ds_a} \\ &= -\frac{N^2\delta}{s_a^2}\int_{a'}^{b} f(\beta)(\int_{a'}^{\beta} xf(x)dx)d\beta < 0 \end{aligned} \tag{B.7}$$

$$\frac{db_f}{ds_a} = \frac{\partial b_f}{\partial a'}\frac{\partial a'}{\partial s_a} = f(a')\frac{\partial a'}{\partial s_a} < 0 \tag{B.8}$$

The derivatives with respect to AR free-flow travel time are:

$$\frac{dTTT_f}{d\tau} = \frac{\partial TTT_f}{\partial a'}\frac{\partial a'}{\partial \tau} + \frac{\partial TTT_f}{\partial \tau} \tag{B.9}$$

$$= \frac{N\frac{1}{s_f}f(a')\int_a^{a'} xf(x)dx}{\lambda} > 0$$

$$\frac{dTTT_a}{d\tau} = \frac{\partial TTT_a}{\partial a'}\frac{\partial a'}{\partial \tau} + \frac{\partial TTT_a}{\partial \tau} \tag{B.10}$$

$$= N\int_{a'}^{b} f(x)dx - \frac{N\frac{1}{s_f}f(a')\int_a^{a'} xf(x)dx}{(\frac{1}{s_f}+\frac{1}{s_a})a'f(a') - \frac{1}{s_a}\int_{a'}^{b} f(x)dx}$$

$$= \frac{[\frac{1}{s_f}a'f(a') - \frac{1}{s_a}\int_{a'}^{b} f(x)dx]\int_{a'}^{b} f(x)dx - f(a')\frac{\tau}{N\delta}}{\lambda} \tag{B.11}$$

$$\frac{dTTT}{d\tau} = \frac{dTTT_f}{d\tau} + \frac{dTTT_a}{d\tau} \tag{B.12}$$

$$= N\int_{a'}^{b} f(x)dx > 0$$

$$\frac{db_f}{d\tau} = \frac{\partial b_f}{\partial a'}\frac{\partial a'}{\partial \tau} = f(a')\frac{\partial a'}{\partial \tau} > 0 \tag{B.13}$$

Appendix C

Derivation of Pareto-improving

Here, we show that capacity enlargement on either the freeway or AR is Pareto-improving and how commuters benefit differently from the capacity enlargement.

Let $C_f(\beta)$ and $C_a(\beta)$ denote the generalized travel time of a commuter with EAP β on the freeway and AR, respectively. By Equation 3.19,

$$\begin{aligned} C_f(\beta) &= N_f \frac{\delta\beta}{s_f} - N\frac{\delta\beta}{s_f}\int_a^\beta f(x)dx + N\frac{\delta}{s_f}\int_a^\beta xf(x)dx \quad a \le \beta \le a' \\ C_a(\beta) &= \tau + N_a\frac{\delta\beta}{s_a} - N\frac{\delta\beta}{s_a}\int_{a'}^\beta f(x)dx + N_a\frac{\delta}{s_a}\int_{a'}^\beta xf(x)dx \quad a' \le \beta \le b \end{aligned}$$

We first show that by Equation 3.25,

$$\frac{s_a}{s_f} = \frac{\frac{s_a\tau}{N\delta} + a'\int_{a'}^b f(x)dx}{\int_a^{a'} xf(x)dx} > \frac{a'\int_{a'}^b f(x)dx}{\int_a^{a'} xf(x)dx} > \frac{\int_{a'}^b f(x)dx}{a'f(a')} \tag{C.1}$$

Now we investigate how the changes in freeway capacity s_f affect the individual generalized travel time.

$$\frac{\partial C_f(\beta)}{\partial s_f}\frac{s_f^2}{N\delta} = \beta f(a')\frac{\frac{1}{s_f}\int_a^{a'} xf(x)dx}{(\frac{1}{s_f}+\frac{1}{s_a})a'f(a') - \frac{1}{s_a}\int_{a'}^b f(x)dx} - \int_a^\beta xf(x)dx - \beta\int_\beta^{a'} f(x)dx$$

$$\frac{\partial \frac{\partial C_f(\beta)}{\partial s_f}}{\partial\beta}\frac{s_f^2}{N\delta} = f(a')\frac{\partial a'}{\partial s_f}s_f - \int_\beta^{a'} f(x)dx$$

Therefore,

$$\frac{\partial \frac{\partial C_f(\beta)}{\partial s_f}}{\partial\beta} \text{ is monotone with respect to } \beta$$

$$\frac{\partial \frac{\partial C_f(\beta)}{\partial s_f}}{\partial\beta}|_{\beta=a} < 0, \frac{\partial \frac{\partial C_f(\beta)}{\partial s_f}}{\partial\beta}|_{\beta=a'} > 0$$

$$\begin{aligned}
\frac{\partial C_f(\beta)}{\partial s_f}\frac{s_f^2}{N\delta}|_{\beta=a} &= \frac{\frac{a}{a'}\int_a^{a'} xf(x)dx}{\frac{s_a}{s_f}+1-\frac{\int_{a'}^b f(x)dx}{a'f(a')}} - a\int_a^{a'} f(x)dx \\
&< \frac{\frac{a}{a'}\int_a^{a'} xf(x)dx}{\frac{s_a}{s_f}+1-\frac{\int_{a'}^b f(x)dx}{a'f(a')}} - \frac{a}{a'}\int_a^{a'} xf(x)dx \\
&= \frac{a}{a'}\int_a^{a'} xf(x)dx\left(\frac{1}{\frac{s_a}{s_f}+1-\frac{\int_{a'}^b f(x)dx}{a'f(a')}} - 1\right) \\
&< 0 \text{ by Equation C.1}
\end{aligned}$$

Hence,

$$\frac{\partial C_f(\beta)}{\partial s_f}\frac{s_f^2}{N\delta} < 0$$

$$\begin{aligned}
\frac{\partial C_f(\beta)}{\partial s_f}\frac{s_f^2}{N\delta}|_{\beta=a'} &= \int_a^{a'} xf(x)dx\left(\frac{\frac{s_a}{s_f}}{\frac{s_a}{s_f}+1-\frac{\int_{a'}^b f(x)dx}{a'f(a')}} - 1\right) \\
&< 0 \text{ by the assumption in Proposition 3.5}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial C_a(\beta)}{\partial s_f} &= -\frac{\partial a'}{\partial s_f}\frac{N\delta}{s_a}a'f(a') < 0 \\
\frac{\partial \frac{\partial C_a(\beta)}{\partial s_f}}{\partial \beta} &= 0
\end{aligned}$$

Now we investigate how the changes in AR capacity s_a affect the individual generalized travel time.

$$\begin{aligned}
\frac{\partial C_f(\beta)}{\partial s_a} &= \frac{\partial a'}{\partial s_a}\frac{N\delta}{s_f}\beta f(a') < 0 \\
\frac{\partial \frac{\partial C_f(\beta)}{\partial s_a}}{\partial \beta} &= \frac{\partial a'}{\partial s_a}\frac{N\delta}{s_f}f(a') < 0 \text{ is independent of } \beta
\end{aligned}$$

$$\frac{\partial C_a(\beta)}{\partial s_a}\frac{s_a^2}{N\delta} = a'f(a')\frac{\frac{1}{s_a}a'\int_{a'}^b f(x)dx}{(\frac{1}{s_f}+\frac{1}{s_a})a'f(a') - \frac{1}{s_a}\int_{a'}^b f(x)dx} - \int_{a'}^{\beta} xf(x)dx - \beta\int_{\beta}^b f(x)dx$$

$$\frac{\partial \frac{\partial C_a(\beta)}{\partial s_a}}{\partial \beta}\frac{s_a^2}{N\delta} = -\int_{\beta}^b f(x)dx < 0$$

Because,

$$\begin{aligned}
\frac{\partial C_a(\beta)}{\partial s_a}\frac{s_a^2}{N\delta}|_{\beta=a'} &= a'f(a')\frac{\frac{1}{s_a}a'\int_{a'}^{b}f(x)dx}{(\frac{1}{s_f}+\frac{1}{s_a})a'f(a')-\frac{1}{s_a}\int_{a'}^{b}f(x)dx} - a'\int_{a'}^{b}f(x)dx \\
&= a'\int_{a'}^{b}f(x)dx(\frac{1}{\frac{s_a}{s_f}+1-\frac{\int_{a'}^{b}f(x)dx}{a'f(a')}}-1) < 0 \text{ by Equation C.1}
\end{aligned}$$

therefore,

$$\frac{\partial C_a(\beta)}{\partial s_a} < 0$$

Appendix D

Travel profiles and the total travel cost with parking choices

All possible 20 travel profiles are shown in Figure D.1, D.2 and D.3. In all five types of parking preference, if $K_1 = 0, K_2 = N$ ($K_2 = 0, K_1 = N$), then the travel profile is shown in profile 19 (20).

The derivatives of total travel cost with respect to the parking capacity, parking fee and access time are as follows.

For strongly outward parking (where $p_1 = p_2$ is a special case),

$$\frac{\partial TC}{\partial K_1} = \begin{cases} -(p_1 - p_2) \leq 0 & \text{if } K_1 \geq \dfrac{s(v_2 - v_1)}{\beta}, K_1 \neq N \\ -\dfrac{N\beta\gamma}{s(\beta+\gamma)} - (p_1 - p_2) < 0 & \text{if } K_1 \leq \dfrac{s(v_2 - v_1)}{\beta} \end{cases} \tag{D.1}$$

$$\frac{\partial TC}{\partial \Delta p} = \begin{cases} N\dfrac{\gamma}{\beta+\gamma} - K_1 & \text{if } K_1 \geq \dfrac{s(v_2 - v_1)}{\beta} \\ -K_1 < 0 & \text{if } K_1 \leq \dfrac{s(v_2 - v_1)}{\beta} \end{cases} \tag{D.2}$$

$$\frac{\partial TC}{\partial \Delta l} = \begin{cases} N\dfrac{\beta}{\beta+\gamma}(\lambda+\gamma) > 0 & \text{if } K_1 \geq \dfrac{s(v_2 - v_1)}{\beta} \\ \lambda N > 0 & \text{if } K_1 \leq \dfrac{s(v_2 - v_1)}{\beta} \end{cases} \tag{D.3}$$

For strongly inward parking where $p_1 > p_2$,

$$\frac{\partial TC}{\partial K_2} = \begin{cases} p_1 - p_2 > 0 & \text{if } K_2 \geq \dfrac{s(v_1 - v_2)}{\beta}, K_2 \neq N \\ -\dfrac{N\beta\gamma}{s(\beta+\gamma)} + (p_1 - p_2) & \text{if } K_2 \leq \dfrac{s(v_1 - v_2)}{\beta} \end{cases} \tag{D.4}$$

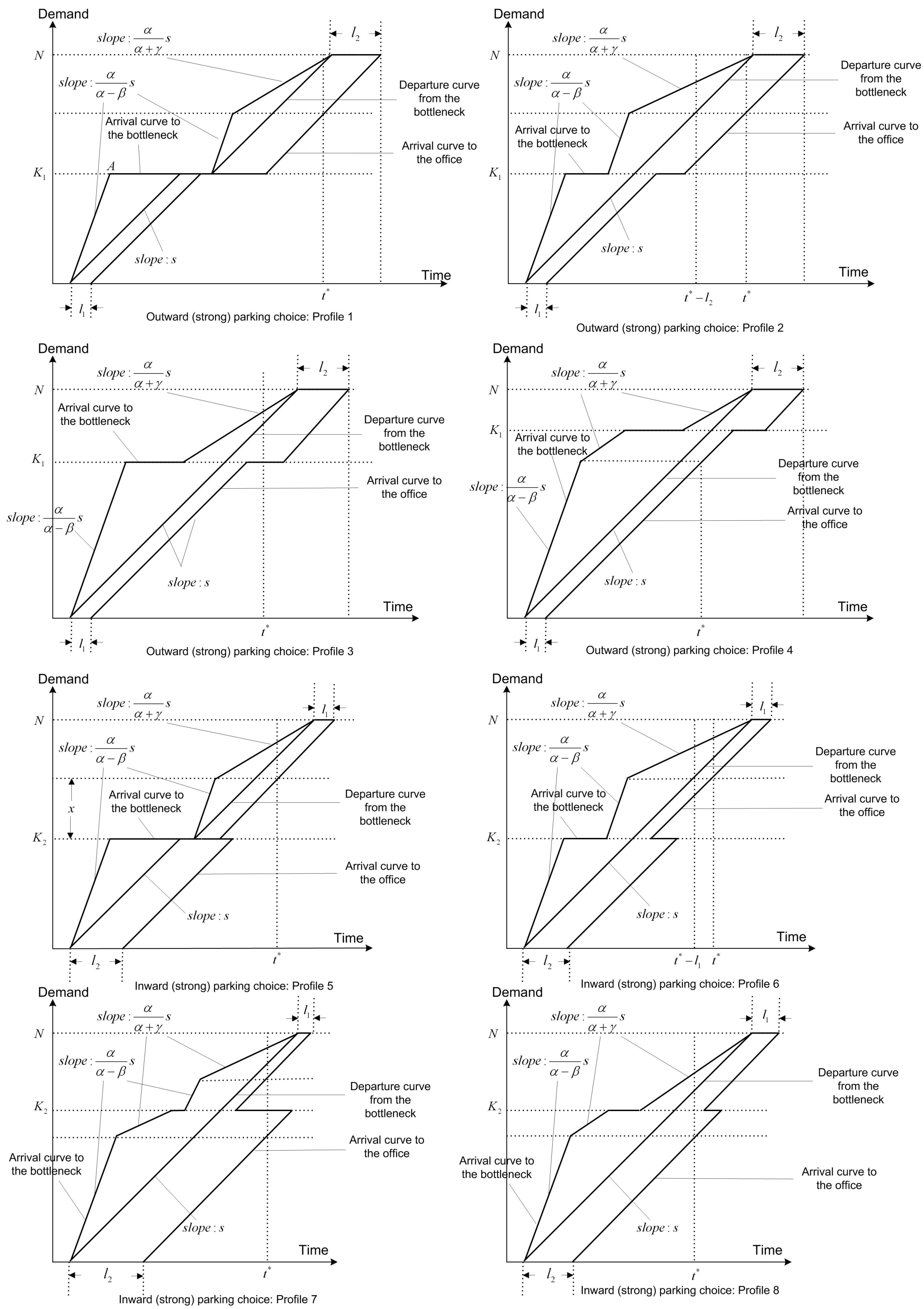


Figure D.1. Travel profiles of strongly outward parking and strongly inward parking

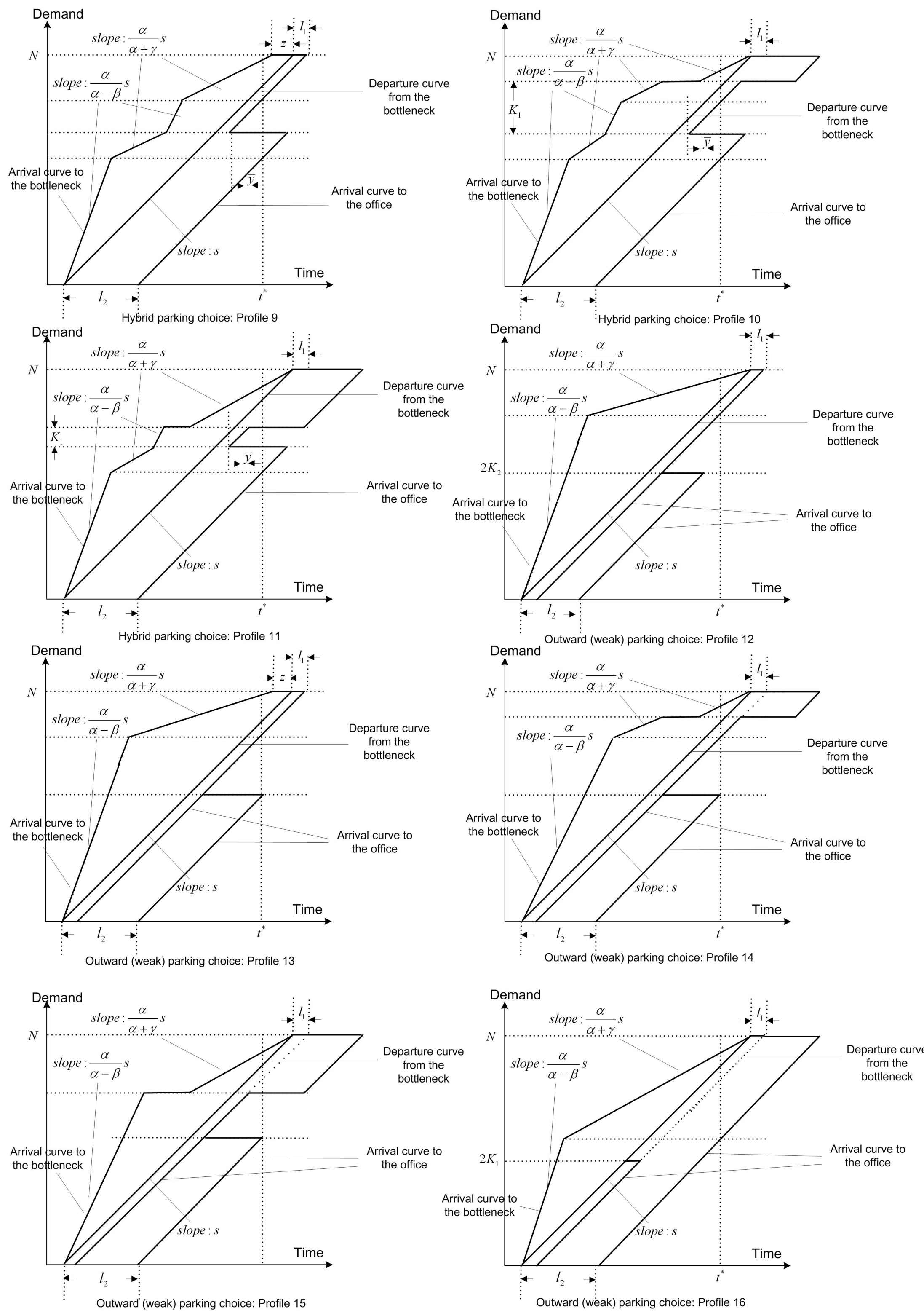


Figure D.2. Travel profiles of hybrid parking and weakly outward parking

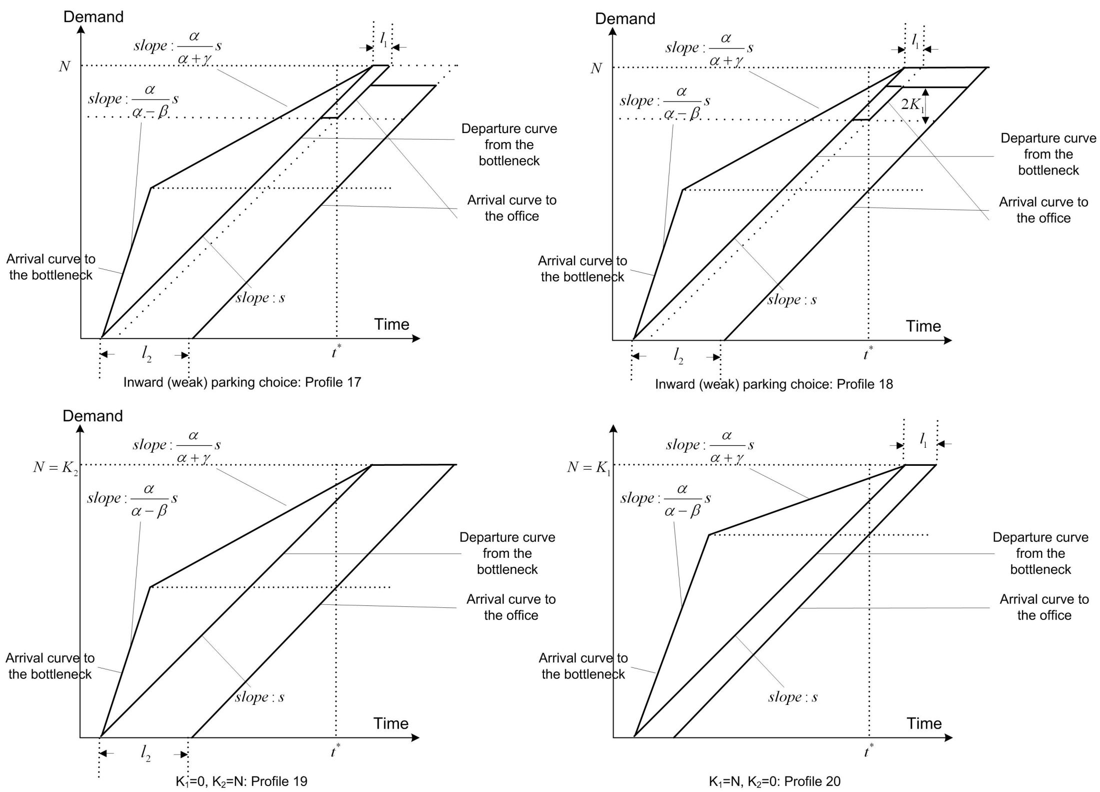


Figure D.3. Travel profiles of weakly inward parking

$$\frac{\partial TC}{\partial \Delta p} = \begin{cases} -N\dfrac{\gamma}{\beta+\gamma} + K_2 & \text{if } K_2 \geq \dfrac{s(v_1-v_2)}{\beta} \\ K_2 > 0 & \text{if } K_2 \leq \dfrac{s(v_1-v_2)}{\beta} \end{cases} \tag{D.5}$$

$$\frac{\partial TC}{\partial \Delta l} = \begin{cases} -N\dfrac{\beta}{\beta+\gamma}(\lambda+\gamma) + \lambda N = \dfrac{N\gamma}{\beta+\gamma}(\lambda-\beta) > 0 & \text{if } K_2 \geq \dfrac{s(v_1-v_2)}{\beta} \\ 0 & \text{if } K_2 \leq \dfrac{s(v_1-v_2)}{\beta} \end{cases} \tag{D.6}$$

For hybrid parking where $p_1 > p_2$,

$$\frac{\partial TC}{\partial K_1} = \begin{cases} -\dfrac{N\beta}{s} - (p_1-p_2) < 0 & \text{if } \dfrac{N\beta - s(v_1-v_2)}{\beta+\gamma} \leq K_1 \leq \dfrac{N\beta - s(v_1-v_2) + s(u_2-u_1)}{\beta+\gamma} \\ -(p_1-p_2) < 0 & \text{if } K_1 \leq \dfrac{N\beta - s(v_1-v_2)}{\beta+\gamma} \\ 0 & \text{otherwise} \end{cases} \tag{D.7}$$

$$\frac{\partial TC}{\partial K_2} = \begin{cases} -\dfrac{N\beta\gamma}{s(\beta+\gamma)} + (p_1-p_2) & \text{if } K_2 \leq \dfrac{s(v_1-v_2)}{\beta} \\ p_1 - p_2 > 0 & \text{if } \dfrac{s(v_1-v_2)}{\beta} \leq K_2 \leq \dfrac{N\gamma + s(v_1-v_2) + s(u_1-u_2)}{\beta+\gamma} \\ 0 & \text{otherwise} \end{cases} \tag{D.8}$$

$$\frac{\partial TC}{\partial \Delta p} = \begin{cases} K_2 > 0 & \text{if } K_2 \le \frac{s(v_1 - v_2)}{\beta} \\ -\frac{N\gamma}{\beta+\gamma} + K_2 & \text{if } \frac{s(v_1 - v_2)}{\beta} \le K_2 \le \frac{N\gamma + s(v_1 - v_2) + s(u_1 - u_2)}{\beta+\gamma} \\ -\frac{N\beta}{\beta+\gamma} - K_1 < 0 & \text{if } \frac{N\beta - s(v_1 - v_2)}{\beta+\gamma} \le K_1 \le \frac{N\beta - s(v_1 - v_2) + s(u_2 - u_1)}{\beta+\gamma} \\ -K_1 < 0 & \text{if } K_1 \le \frac{N\beta - s(v_1 - v_2)}{\beta+\gamma} \end{cases} \tag{D.9}$$

$$\frac{\partial TC}{\partial \Delta l} = \begin{cases} 0 & \text{if } K_2 \le \frac{s(v_1 - v_2)}{\beta} \\ \frac{N\gamma}{\beta+\gamma}(\lambda - \beta) > 0 & \text{if } \frac{s(v_1 - v_2)}{\beta} \le K_2 \le \frac{N\gamma + s(v_1 - v_2) + s(u_1 - u_2)}{\beta+\gamma} \\ \lambda N + \frac{N\beta}{\beta+\gamma} > 0 & \text{if } \frac{N\beta - s(v_1 - v_2)}{\beta+\gamma} \le K_1 \le \frac{N\beta - s(v_1 - v_2) + s(u_2 - u_1)}{\beta+\gamma} \\ \lambda N > 0 & \text{if } K_1 \le \frac{N\beta - s(v_1 - v_2)}{\beta+\gamma} \end{cases} \tag{D.10}$$

For weakly outward parking where $p_1 > p_2$,

$$\frac{\partial TC}{\partial K_1} = \begin{cases} -(p_1 - p_2) < 0 & \text{if } K_1 \le N - \frac{1}{2}\frac{N\gamma}{\beta+\gamma} \\ 0 & \text{otherwise} \end{cases} \tag{D.11}$$

$$\frac{\partial TC}{\partial K_2} = \begin{cases} p_1 - p_2 > 0 & \text{if } K_2 \le \frac{1}{2}\frac{s(u_1 - u_2) + N\gamma}{\beta+\gamma} \\ (p_1 - p_2) + \frac{2N\beta}{s} > 0 & \text{if } \frac{1}{2}\frac{s(u_1 - u_2) + N\gamma}{\beta+\gamma} \le K_2 \le \frac{1}{2}\frac{N\gamma}{\beta+\gamma} \\ 0 & \text{otherwise} \end{cases} \tag{D.12}$$

For weakly inward parking where $p_1 > p_2$,

$$\frac{\partial TC}{\partial K_1} = \begin{cases} -(p_1 - p_2) < 0 & \text{if } K_1 \le \frac{1}{2}\frac{N\beta - s(v_1 - v_2)}{\beta+\gamma} \\ 0 & \text{otherwise} \end{cases} \tag{D.13}$$

$$\frac{\partial TC}{\partial K_2} = \begin{cases} (p_1 - p_2) - \frac{N\beta\gamma}{s(\beta+\gamma)} & \text{if } K_2 \le \frac{s(v_1 - v_2)}{\beta} \\ (p_1 - p_2) > 0 & \text{if } \frac{s(v_1 - v_2)}{\beta} \le K_2 \le N - \frac{1}{2}\frac{N\beta - s(v_1 - v_2)}{\beta+\gamma} \\ 0 & \text{otherwise} \end{cases} \tag{D.14}$$

In all parking patterns, $\frac{\partial TC}{\partial \Delta p} = 0$ if $K_1 = 0$ or $K_2 = 0$.

Appendix E

An example of deriving travel profiles with parking choice

Recall that in strongly outward parking, the closer parking cluster, i.e. parking cluster 1, is always preferred in both early arrival and late arrival. Given parking fees and access times of two parking clusters that leads to strongly outward parking, the travel profile is determined by the actual (also effective) capacity of parking cluster 1, K_1. This is because the farther cluster will not be used unless the closer one is used up. Overall, there exist four patterns in strongly outward parking:

1. After the closer cluster is used up by travelers with early arrival, the first traveler using the farther cluster also arrives earlier than t^*. The capacity of cluster 1 is so low that the queuing delay of the last traveler using cluster 1 is fairly small, and the first traveler using cluster 2 would rather depart after the queue vanishes (i.e. this traveler has no queuing delay). This profile occurs when K_1 is low and is shown in Profile 1.

2. After the closer cluster is used up by travelers with early arrival, the first traveler using the farther cluster also arrives earlier than t^*. However, the queuing delay of the last traveler using cluster 1 is sufficiently long so that the first traveler parking in cluster 2 is willing to wait in the queue to arrive. Therefore, the first traveler parking in cluster 2 has a queuing delay. This profile occurs when K_1 is low but larger than it is in Profile 1, and is shown in Profile 2.

3. After the closer cluster is used up by early-arrival travelers, the first traveler using

the farther cluster arrives later than t^*. This profile occurs when K_1 is medium and is shown in Profile 3.

4. The closer cluster is used by travelers with both early arrival and late arrival, and all travelers using the farther cluster thereafter arrive later than t^*. This profile occurs when K_1 is high, and is shown in Profile 4.

In this example, we derive the departure/arrival curves and TC of Profile 3 (shown in Figure E.1), where all the travelers choosing the closer cluster experience early arrival (or punctual arrival) and all the travelers choosing the farther cluster experience late arrival (or punctual arrival).

We first consider only the early arrival. In a typical morning commute problem, travelers have the same travel cost (i.e. summation of queuing delay and schedule delay) as long as they reach the bottleneck at the arrival rate of $\frac{\alpha}{\alpha-\beta}s$. In this example, because travelers using the closer cluster in early arrival have identical additional cost (parking fee and access time) as compared to the travel cost in the typical morning commute problem, the arrival curve to the bottleneck for those travelers follows the same slope, i.e., $\frac{\alpha}{\alpha-\beta}s$. The similar logic applies to all travelers arriving later than t^*, their arrival curve to the bottleneck also follows the same slope $\frac{\alpha}{\alpha+\gamma}s$ as in the late arrival of a typical morning commute problem. Because the bottleneck is always used throughout the morning peak, the departure curve to the bottleneck is always continuous with the slope of s (bottleneck capacity). It takes l_1 and l_2 amount of time for travelers choosing parking clusters 1 and 2 to walk to their offices, respectively, which is depicted by the arrival curve to the office in Figure E.1. Let y represent the duration from the arrival time to the office of the last traveler using the closer cluster to t^*. Thus, the duration from t^* to the time the first traveler using the farther cluster is $l_2 - l_1 - y$.

Under user equilibrium, since all commuters, regardless of parking choices and departure times, have the same generalized travel cost, the travel cost of the last traveler using parking cluster 1, $C_1(t_C)$, must equal that of the first traveler using the farther cluster, $C_2(t_A)$, i.e.,

$$C_1(t_C) - C_2(t_A) = 0 \tag{E.1}$$

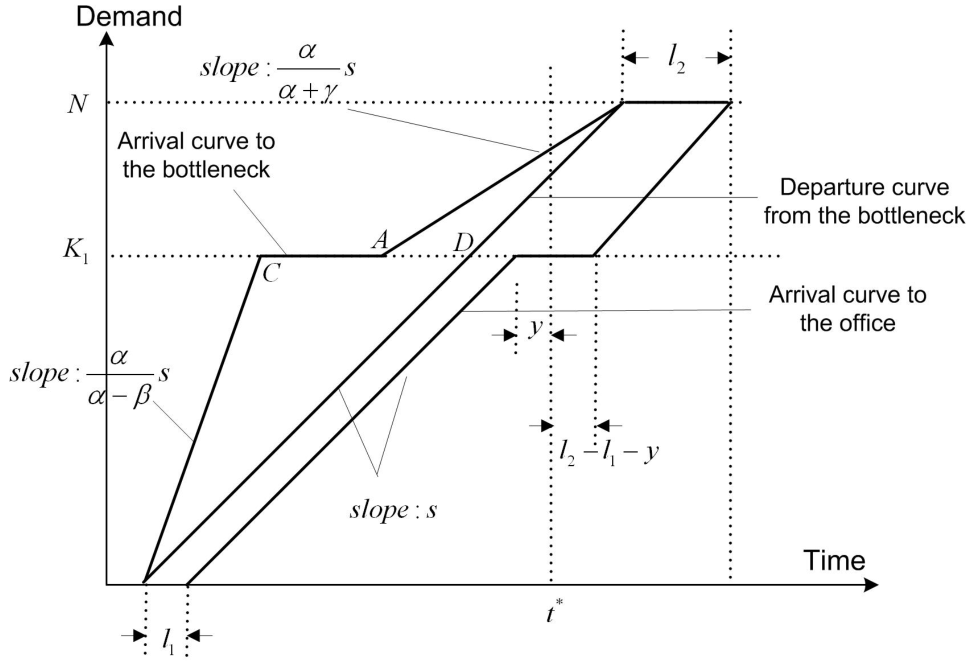


Figure E.1. A travel profile in outward parking: Profile 3

Substitute $C_1(t_C)$ and $C_2(t_A)$ by Equation 5.5 and 5.6 into Equation E.1,

$$\alpha(CD - AD) + [(p_1 + \lambda l_1) - (p_2 + \lambda l_2)] + [\beta y - \gamma(l_2 - l_1 - y)] = 0 \tag{E.2}$$

where $\alpha(CD - AD)$, $(p_1 + \lambda l_1) - (p_2 + \lambda l_2)$ and $\beta y - \gamma(l_2 - l_1 - y)$ represent the difference in queuing delay cost, parking-related cost and schedule delay cost of the two marginal travelers, respectively. Because

$$\begin{aligned} CD &= \frac{K_1}{s}\frac{\beta}{\alpha} \\ AD &= \frac{N - K_1}{s}\frac{\gamma}{\alpha} \end{aligned}$$

solving y in Equation E.2 yields,

$$y = -\frac{K_1}{s} + \frac{N\beta}{s}\frac{s(u_2 - u_1) + N\gamma}{\beta + \gamma}$$

Since $0 \leq y \leq l_2 - l_1$ according to the condition of this profile, we have

$$\frac{s(v_2 - v_1) + N\gamma}{\beta + \gamma} \leq K_1 \leq \frac{s(u_2 - u_1) + N\gamma}{\beta + \gamma} \tag{E.3}$$

Because all commuters are subject to the same generalized travel cost, the identical generalized travel cost can be computed in terms of an arbitrary traveler, e.g. the traveler departing at time t_C,

$$\alpha CD + (p_1 + \lambda l_1) + \beta y$$

Therefore, the total generalized travel cost of all commuters, exclusive of revenues from parking fees, becomes

$$\begin{aligned} TC &= (\alpha CD + p_1 + \lambda l_1 + \beta y)N - K_1 p_1 - (N - K_1)p_2 \\ &= \frac{N\beta}{s}\frac{s(u_2 - u_1) + N\gamma}{\beta + \gamma} + (N - K_1)(p_1 - p_2) + \lambda N l_1 \end{aligned} \tag{E.4}$$

Though the effective capacity of the farther cluster, $K_2 = N - K_1$, can also be used to express Equation E.3 and E.4, we rather use K_1 since the closer cluster essentially determines the profile.

Appendix F

Proof: The optimal profile will not be achieved in outward parking

This section proves that the optimal profile will not be achieved in outward parking.

Without loss of generality, let $l_1 = 0, p_2 = 0$. l_2 is fixed. Let p_1 and p_1' denote the parking fee of the closer parking cluster in outward and inward (or hybrid) parking, respectively. Due to the conditions of outward parking, $(\lambda - \beta)l_2 > p_1$. Let TC denote the total travel cost of Profiles 2, 3 or 4 in outward parking, and TC' denote the total travel cost of Profiles 6, 7 or 8 in inward (or hybrid) parking. By eliminating l_1 and p_2, TC and TC' become,

$$TC = \frac{N\beta}{s(\beta+\gamma)}[-sp_1 + s(\lambda+\gamma)l_2 + N\gamma] + (N - K_1)p_1$$
$$TC' = \frac{N\beta}{s(\beta+\gamma)}[sp_1' - s(\lambda+\gamma)l_2 + N\gamma] - (N - K_2)p_1' + \lambda l_2 N$$

$$TC - TC' = \frac{N\beta}{\beta+\gamma}[2(\lambda+\gamma)l_2 - p_1 - p_1'] + N(p_1 + p_1') + \lambda l_2 N + K_2 p_1' - K_1 p_1$$

Because $p_1 < (\lambda - \beta)l_2 < \lambda l_2$

$$\begin{aligned} TC - TC' &> \frac{N\beta}{\beta+\gamma}[2(\lambda+\gamma)l_2 - p_1 - p_1'] + N(p_1 + p_1') + p_1 N + K_2 p_1' - K_1 p_1 \\ &> \frac{N\beta}{\beta+\gamma}2(\lambda+\gamma)l_2 + (p_1 + p_1')\frac{N\gamma}{\beta+\gamma} + K_2 p_1' + p_1(N - K_1) \\ &> 0 \end{aligned}$$

Since given K_1 and K_2, the total travel cost of Profiles 2, 3 or 4 is always larger than that of Profiles 6, 7 or 8, the optimal profile never falls in strongly outward parking.

Similarly, it is easy to verify that given K_1 and K_2, the total travel cost of Profiles $12 \sim 16$ is also larger than that of Profiles 6, 7 or 8. Therefore, the optimal profile will not be achieved in weakly outward parking. Overall, the optimal profile must be the case of inward parking or hybrid parking.

Appendix G

Sensitivity analysis of investment cost, parking fee and access time

Here, we derive the derivatives of TSC,TCC and TD under Type IV competitive equilibrium with respect to the investment cost, parking fee and access time, respectively.

By Equation 5.41 and 5.48, we have (let $\Delta\bar{p} = \bar{p_1} - \bar{p_2}$),

$$\frac{d\overline{TCC}}{da_1} = \frac{\partial\overline{TCC}}{\partial\bar{p_1}}\frac{\partial\bar{p_1}}{\partial a_1} + \frac{\partial\overline{TCC}}{\partial\bar{p_2}}\frac{\partial\bar{p_2}}{\partial a_1} = \frac{N(2\beta+\gamma)}{3(\beta+\gamma)} > 0$$

$$\frac{d\overline{TCC}}{da_2} = \frac{\partial\overline{TCC}}{\partial\bar{p_1}}\frac{\partial\bar{p_1}}{\partial a_2} + \frac{\partial\overline{TCC}}{\partial\bar{p_2}}\frac{\partial\bar{p_2}}{\partial a_2} = \frac{N(\beta+2\gamma)}{3(\beta+\gamma)} > 0$$

Similarly,

$$\frac{d\overline{TSC}}{da_1} = \frac{-8s(a_1-a_2) + N(8\beta+\gamma) + 4s\Delta l(2\lambda+\gamma-\beta)}{9(\beta+\gamma)}$$

$$\frac{d\overline{TSC}}{da_2} = \frac{16s(a_1-a_2) + N(5\beta+4\gamma) - 2s\Delta l(2\lambda+\gamma-\beta)}{9(\beta+\gamma)}$$

The sign of $\frac{d\overline{TSC}}{da_1}$ and $\frac{d\overline{TSC}}{da_2}$ is dependent on the values of all parameters.

We also have,

$$\frac{d\overline{TD}}{da_1} = \frac{d\overline{TD}}{d\Delta\bar{p}}\frac{d\Delta\bar{p}}{da_1} = \frac{2}{3}\frac{s}{\alpha(\beta+\gamma)}(2\Delta\bar{p} - (2\lambda+\gamma-\beta)\Delta l)$$

$$\frac{d\overline{TD}}{da_2} = \frac{d\overline{TD}}{d\Delta\bar{p}}\frac{d\Delta\bar{p}}{da_2} = -\frac{2}{3}\frac{s}{\alpha(\beta+\gamma)}(2\Delta\bar{p} - (2\lambda+\gamma-\beta)\Delta l)$$

The sign of $\frac{d\overline{TD}}{da_1}$ and $\frac{d\overline{TD}}{da_2}$ is dependent on the values of all parameters.

Suppose l_1 is fixed in the closer cluster, and the private operators in the farther cluster may improve the accessibility so as to reduce l_2. Our derivation yields,

$$\begin{aligned}
\frac{d\overline{TCC}}{dl_2} &= \frac{N}{6(\beta+\gamma)}\left((\beta+\gamma)(2\lambda-\gamma-\beta)+2\gamma(\lambda-\beta)\right) \\
\frac{d\overline{TSC}}{dl_2} &= \frac{N\gamma(\beta+\lambda)}{\beta+\gamma}+\frac{2\lambda+\gamma-\beta}{9(\beta+\gamma)}\left(2s\Delta a-2s\Delta l(2\lambda+\gamma-\beta)-\frac{N(\gamma-\beta)}{2}\right) \\
\frac{d\overline{TD}}{dl_2} &= \frac{2(\gamma+2\beta-\lambda)(\lambda+2\gamma+\beta)}{9}+(\frac{N(\beta-\gamma)}{2s}+\Delta a)\frac{\gamma+2\lambda-\beta}{9}
\end{aligned}$$

The sign of $\frac{d\overline{TCC}}{dl_2}$, $\frac{d\overline{TSC}}{dl_2}$ and $\frac{d\overline{TD}}{dl_2}$ is dependent on the values of all parameters.

Finally, by differentiating Equation 5.43 and 5.44 with respect to p_1 and p_2, we get,

$$\begin{aligned}
\frac{d\overline{TCC}}{dp_1} &= \frac{N\beta}{\beta+\gamma} > 0 \\
\frac{d\overline{TCC}}{dp_2} &= \frac{N\gamma}{\beta+\gamma} > 0 \\
\frac{d\overline{TSC}}{dp_1} &= \frac{s}{\beta+\gamma}\left(4\Delta p-(2\lambda+\gamma-\beta)\Delta l-2\Delta a\right) \\
\frac{d\overline{TSC}}{dp_2} &= \frac{s}{\beta+\gamma}\left(-4\Delta p+(2\lambda+\gamma-\beta)\Delta l+2\Delta a\right)
\end{aligned}$$

References

2010 Urban Mobility Report and Appendices (2010), Technical report, Texas Transportation Institute.

Anderson, S. P. & de Palma, A. (2004), 'The economics of pricing parking', *Journal of Urban Economics* **55**, 1–20.

Arnott, R., de Palma, A. & Lindsey, R. (1988), 'Schedule delay and departure time decisions with heterogeneous commuters', *Transportation Research Record* **1197**, 56–67.

Arnott, R., de Palma, A. & Lindsey, R. (1991), 'A temporal and spatial equilibrium analysis of commuter parking', *Journal of Public Economics* **45**, 301–335.

Arnott, R., Palma, A. D. & Lindsey, R. (1990), 'Departure time and route choice for the morning commute', *Transportation Research Part B* **24**, 209–228.

Arnott, R., Palma, A. D. & Lindsey, R. (1993), 'Properties of dynamic traffic equilibrium involving bottlenecks, including a paradox and metering', *Transportation Science* **27**, 148–160.

Arnott, R. & Rowse, J. (1999), 'Modeling parking', *Journal of Urban Economics* **45**, 97–124.

Beckmann, M., McGuire, C. & Winsten, C. (1956), *Studies in the Economics of Transportation,*, Yale University Press, New Haven, Connecticut.

Bell, M. G. H., Shield, C. M., Busch, F. & Kruse, G. (1997), 'A stochastic user equilibrium path flow estimator', *Transportation Research Part C: Emerging Technology* **5**(3-4), 197 – 210.

Bifulco, G. N. (1993), 'A stochastic user equilibrium assignment model for the evaluation of parking policies', *European Journal of Operational Research* **71**, 269–287.

Bliemer, M. (2000), Analytical dynamic traffic assignment with interacting user-classes, PhD thesis, Delft University of Technology.

Carey, M. (1986), 'A constraint qualification for a dynamic traffic assignment problem', *Transportation Science* **20**, 55–58.

Carey, M. (1987), 'Optimal time-varying flows on congested networks', *Operation Research* **35**, 58–69.

Cea, J. D. & Fernandez, E. (1993), 'Transit assignment for congested public transport systems: an equilibrium model', *Transportation Science* **27**(2), 133–147.

Chu, X. (1994), 'Endogenous trip scheduling: the henderson approach reformulated and compared with the vickery approach', *Journal of urban economics* **37**, 324–343.

Dafermos, S. C. (1971), *The traffic assignment problem for multiclass-user transportation networks*, Cornell University, Dept. of Operations Research.

Daganzo, C. (1998), 'Queue spillovers in transportation networks with a route choice', *Transportation Science* **32**(1), 3–11.

Daganzo, C. F. (1994), 'The cell transmission model: a dynamic representation of highway traffic consistent with the hydrodynamic theory', *Transportation Research Part B* **28**, 269–287.

Daganzo, C. F. (1995), 'The cell transmission model, part II: Network traffic', *Transportation Research Part B* **29**, 79–93.

Daganzo, C. F. & Garcia, R. C. (2000), 'A pareto improving strategy for the time-dependent morning commute problem', *Transportation Science* **34**(3), 303–311.

Dahlgren, J. (2002), 'High-occupancy/toll lanes: Where should they be implemented?', *Transportation Research Part A* **36**, 239–255.

Dial, R. B. (1996), 'Bicriterion traffic assignment: Basic theory and elementary algorithms', *Transportation Science* **30**, 93–111.

Dial, R. B. (1997), 'Bicriterion traffic assignment: Efficient algorithms plus examples', *Transportation Research Part B* **31**, 357–379.

Dial, R. B. (1999*a*), 'Minimal-revenue congestion pricing part i: a fast algorithm for the single-origin case', *Transportation Research Part B* **33**, 189–202.

Dial, R. B. (1999*b*), 'Network-optimized road pricing: Part i a parable and a model', *Operations research* **47**(1), 54–64.

Facchinei, F. & Pang, J.-S. (2003), *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Vol. I, II, Springer.

Fernandez, E., Cea, J. D., Florian, M. & Cabrera, E. (1994), 'Network equilibrium models with combined modes', *Transportation Science* **28**(3), 182–192.

Ferrari, P. (1998), Congestion pricing for urban binodal transportation network, *in* 'Proceedings of the 14th ISTTT meeting', pp. 409–430.

Ferrari, P. (1999), 'A model of urban transport management', *Transportation Research Part B* **33**, 43–61.

FHWA (2006), *Integrated corridor management concepts development and fundational research*.

Friesz, T. L., Bernstein, D., Mehta, N. J., Tobin, R. L. & Ganjalizadeh, S. (1989), 'Dynamic network traffic assignment considered as a conintuous time optimal control problem', *Operations Research* **37**, 893–901.

Friesz, T. L., Bernstein, D., Smith, T. E., Tobin, R. L. & Wei, B. W. (1993), 'A variational inequality formulation of the dynamic network equilibrium problem', *Operation Research* **41**, 179–191.

Fukushima, M. (1992), 'Equivalent differentiable optimization problems and descent methods for asymmetric variational inequality problems', *Mathematical Programming* **53**, 99–110.

Garcia, R. & Marin, A. (2005), 'Network equilibirum with combined modes: models and solution algorithms', *Transportation Research Part B* **39**, 223–254.

Ghali, M. (1995), 'A note on the minimum instantaneous cost path of the dynamic traffic assignment problem', *Annals of Operations Research* **60**, 115–120.

Glazer, A. (1992), 'Parking fees and congestion', *Regional Science and Urban Economics* **22**, 123–132.

Hamdouch, Y., Florian, M., Hearn, D. W. & Lawphongpanich, S. (2007), 'Congestion pricing for multi-modal transportation systems', *Transportation Research Part B* **41**, 275–291.

Henderson, J. (1977), *Economic theory and cities*, Academic Press, New York, chapter 8.

Huang, H.-J. (2000), 'Fares and tolls in a competitive system with transit and highway: the case with two groups of commuters', *Transportation Research Part E* **36**, 267–284.

Huang, H.-J. (2002), 'Pricing and logit-based mode choice models of a transit and highway system with elastic demand', *European journal of operational research* **140**, 562–570.

Huang, H.-J. & Yang, H. (1999*a*), Carpooling and pricing in a multilane highway with high-occupancy-vehicle lanes and bottleneck congestion, *in* 'Proceedings of the 14th ISTTT meeting', pp. 489–513.

Huang, H.-J. & Yang, H. (1999*b*), 'Optimal utilization of a transport system with auto/transit parallel modes', *Optimal Control Applications and Methods* **20**, 297–313.

Jin, W. & Zhang, H. (2003), 'On the distribution schemes for determining flows through a merge', *Transportation Research Part B* **37**, 521–540.

Jin, W. & Zhang, H. (2004), 'Multicommodity kinematic wave simulation model for network traffic flow', *Transportation Research Record* **1883**, 59 – 67.

Kant, P. (2008), 'Route choice modelling in dynamic traffic assignment'.
URL: *http://essay.utwente.nl/58303/*

Kuwahara, M. (1990), 'Equilibrium queuing patterns at a two-tandem bottleneck during the morning peak', *Transportation Science* **24**(3), 217–229.

Kuwahara, M. & Akamatsu, T. (2001), 'Dynamic user optimal assignment with physical queues for many-to-many od pattern', *Transportation Research Part B* **35**, 461–479.

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

Lam, W. H. & Huang, H.-J. (1992), 'A combined trip distribution and assignment model for multiple user classes', *Transportation Research Part B* **26**(4), 273–287.

Lebacque, J. (1996), 'The godunov scheme and what it means for first order traffic flow models', *Procedings of International Symposium of Transport and Traffic Theory* pp. 79 – 102.

Leurent, F. (1993), 'Cost versus time equilibrium over a network', *European journal of operational research* **71**, 205–221.

Leurent, F. (1996), 'Congestion pricing with continuously distributed values of time', *Journal of transport economics and policy* **34**, 359–370.

Lighthill, M. J. & Whitham, G. B. (1955), 'On kinematic waves. ii. a theory of traffic flow on long crowded roads', *Proceedings of the Royal Society* **229**, 317–345.

Lindsey, R. (2004), 'Existence, uniqueness, and trip cost function properties of user equilibrium in the bottleneck model with multiple user classes', *Transportation Science* **38**, 293–314.

Lu, C.-C., Zhou, X. & Mahmassani, H. (2006), 'Variable toll pricing and heterogeneous users: Model and solution algorithm for bicriterion dynamic traffic assignment problem', *Transportation Research Record* **1964**, 19–26.

Mahmassani, H. & Herman, R. (1984), 'Dynamic user equilibrium departure time and route choice on idealized traffic arterials', *Transportation Science* **18**, 362–384.

Mahmassani, H., Zhou, X. & Lu, C.-C. (2005), 'Toll pricing and heterogeneous users: Approximation algorithms for finding bicriterion time-dependent efficient paths in large-scale traffic networks', *Transportation Research Record* **1923**, 28–36.

Mayet, J. & Hansen, M. (2000), 'Congestion pricing with continuously distributed values of time', *Journal of transport economics and policy* **34**, 359–370.

Merchant, D. & Nemhauser, G. (1978*a*), 'A model and an algorithm for the dynamic traffic assignment problem', *Transportation Science* **12**, 183–199.

Merchant, D. & Nemhauser, G. (1978*b*), 'Optimality conditions for a dynamic traffic assignment model', *Transportation Science* **12**, 200–207.

Nagurney, A. (1999), *Network Economics: A variational inequality approach*, Kluwer Academic publishers.

Newell, G. F. (1987), 'Morning commute for nonidentical travelers', *Transportation Science* **21**, 74–88.

Nguyen, S., Pallottino, S. & Gendreau, M. (1998), 'Implicit enumeration of hyperpaths in a logit model for transit networks', *Transportation Science* **32**(1), 54–64.

Nie, X. (2003), The Study of Dynamic User-equilibrium Traffic Assignment, PhD thesis, University of California at Davis.

Nie, X. & Zhang, H. M. (2005), 'A comparative study of some macroscopic link models used in dynamic traffic assignment', *Networks and Spatial Economics* **5**, 89–115.

Nie, Y. (2006), A Variational Inequality Approach For Inferring Dynamic Origin-Destination Travel Demands, PhD thesis, University of California at Davis.

Nie, Y. & Zhang, H. (2007), 'Solving the dynamic user optimal assignment problem considering queue spillback', *Networks and Spatial Economics* **DOI:10.1007/s11067-007-9022-y**.

Palma, A. D. & Lindsey, R. (2004), 'Congestion pricing with heterogeneous travelers: A general-equilibrium welfare analysis', *Networks and spatial economics* **4**, 135–160.

Patriksson, M. (1994), *The traffic assignment problem: models and methods*, VSP, Utrecht, The Netherlands.

Peeta, S. & 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.

Pel, A. J., Bliemer, M. C. & Hoogendoorn, S. P. (2009), 'Hybrid routing choice modeling in dynamic traffic assignment', *Transportation Research Record* **2091**, 100–107.

Ramadurai, G., Ukkusuri, S., Zhao, J. & Pang, J.-S. (2008), 'Dynamic equilibrium in multi-user class single bottleneck models: A complementarity formulation', *Proceedings of The 87th Annual Meeting of Transportation Research Board* .

Ran, B., Boyce, D. E. & Leblanc, L. J. (1993), 'A new class of instantaneous dynamic user-optimal traffic assignment models', *Operations Research* **41**(1), 192–202.

Richards, P. I. (1956), 'Shock waves on the highway', *Proceedings of the Royal Society* **4**, 42–51.

Sheffi, Y. (1985), *Urban transportation networks: Equilibrium analysis with mathematical programming methods*, Prentice-Hall Inc.

Smith, M. (1979), 'The marginal cost taxation of a transportation network', *Transportation Research Part B* **13**, 237–242.

Smith, M. (1993), 'A new dynamic traffic model and the existence and calculation of dynamic user equilibria on congested capacity-constrained road networks', *Transportation Research Part B* **26**, 49–36.

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

Tabuchi, T. (1993), 'Bottleneck congestion and modal split', *Journal of urban economics* **34**, 414–431.

Thompson, R. G., Takada, K. & Kobayakawa, S. (1998), 'Understanding the demand for access information', *Transportation Research Part C* **6**, 231–245.

Tong, C. O. & Wong, S. C. (1999), 'A stochastic transit assignment model using a dynamic schedule-based network', *Transportation Research Part B* **33**, 107–121.

Tseng, Y.-Y. & Verhoef, E. T. (2008), 'Value of time by time of day: A stated-preference study', *Transportation Research Part B* **42**, 607–618.

Verhoef, E., Nijkamp, P. & Rietveld, P. (1995), 'The economics of regulatory parking policies: the (im)possibilities of parking policies in traffic regulation', *Transportation Research Part A* **29**, 141–156.

Verhoef, E. T. & Small, K. A. (2004), 'Product differentiation on road constrained congestion pricing with heterogeneous users', *Journal of transport economics and policy* **38**, 127–156.

Vianna, M. M. B., da Silva Portugal, L. & Balassiano, R. (2004), 'Intelligent transportation systems and parking management: implementation potential in a brazilian city', *Cities* **doi:10.1016/j.cities.2004.01.001**.

Vickrey, W. (1969), 'Congestion theory and transport investment', *American Economic Review* **59**, 251–261.

Wardrop, J. (1952), 'Some theoretical aspects of road trac research', *Proceedings of the Institute of Civil Engineers, Part II* **1**, 325–378.

Wu, J. H., Florian, M. & Marcotte, P. (1994), 'Transit equilibrium assignment: a model and solution algorithms', *Transportation Science* **28**(3), 193–203.

Yang, H. & Huang, H.-J. (1999), 'Carpooling and congestion pricing in a multilane highway with high-occupance-vehicle lanes', *Transportation Research Part A* **33**, 139–155.

Yang, H. & Huang, H.-J. (2005), *Mathematical and economic theory of road pricing*, Elsevier Press.

Ying, J. Q. & Yang, H. (2005), 'Sensitivity analysis of stochastic user equilibrium flows in a bi-modal network with application to optimal pricing', *Transportation Research Part B* **39**, 769–795.

Zhang, X., Huang, H.-J. & Zhang, H. (2008), 'Integrated daily commuting patterns and optimal road rolls and parking fees in a linear city', *Transportation Research Part B* **42**, 38–56.

Zijpp, N. V. D. & Koolstra, K. (2002), 'Multiclass continuous-time equilibrium model for departure time choice on single-bottleneck network', *Transportation Research Record* **1783**, 134–141.

Zhen Qian
August 2011
Civil and Environmental Engineering

On Dynamic Traffic Assignment in Corridor Networks under Heterogeneous Travelers and Modes

Abstract

This dissertation investigates traveler heterogeneity for dynamic traffic assignment (DTA) in the following four dimensions: travelers' attributes (in the value of time and the value of schedule delay), modal choice, parking choice and route choice. The main focus is on obtaining analytical DTA solutions in simplified networks, particularly in the context of the morning commute problem, with precise sensitivity analysis to derive effective traffic congestion management policies.

First, we solve the morning commute problem with a heterogeneous traveling population whose early/late arrival penalty are continuously distributed. The distribution of the value of schedule delay on each route, freeway or the arterial road, is discussed. It is found that the assumption of homogeneity population overestimates the queuing delay and the total travel time. Every commuter is better off if the freeway capacity or arterial capacity is enlarged, but commuters with high values of early/late arrival penalty generally benefit more than those with low values unless they switch to other routes. We further study the multi-modal morning commute problem with three modes, transit, solo-driving and carpool. Enlarging HOV facilities may reduce transit ridership and increase auto travel, and it does not necessarily reduce the total travel cost when the network is highly congested. The rise of gas price may first entice auto travelers to carpool. However, as the gas price increases further, both carpoolers and solo-drivers will eventually switch to use the transit. In addition, a flat freeway tool can also reduce the total network travel cost.

In addition to the intrinsic distinction among travelers, we also discuss the management measures that can distinguish travelers externally, using parking as an example. The parking fee, parking capacity allocation and accessibility altogether can effectively reduce both the system cost and the queuing delay. If parking lots are owned publicly,

then all travelers are better off under the optimal parking setting. This is an advantage that cannot be realized by the system-optimum dynamic toll scheme. If they are owned privately, then market regulations, such as price-ceiling and quantity tax/subsidy, are suggested to improve the network performance and reduce the congestion.

We finally extend our research to the DTA problem in general networks. We propose a hybrid route choice model for studying non-equilibrium traffic where travelers have different preferences in choosing travel routes. It combines pre-trip route choice and en-route route choice to solve dynamic traffic assignment (DTA) in large-scale networks. We apply the hybrid route choice model in a synthetic medium-scale network and a large-scale real network to assess its effect on the flow patterns and network performances, and compare them with those obtained from Predictive User Equilibrium (PUE) DTA. The proposed route choice model incorporating route choice heterogeneity is capable of solving DTA efficiently in in a realistic size network with satisfactory results. Finally, some suggestions are given on how to calibrate the hybrid route choice model in practice.