The Hong Kong Polytechnic University

Department of Civil and Structural Engineering

ADVANCED MODELS FOR TRANSIT NETWORK DESIGN AND OPERATION UNDER UNCERTAINTIES

ZHANG Yuqing

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Philosophy

March 2011

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written nor material which has been accepted for the award of any other degree or diploma, expect where due acknowledgement has been made in the text.

Name of the Student: ZHANG Yuqing

Signature: ________________

To my parents.

ABSTRACT

Transit network uncertainties commonly exist in transit systems due to day-to-day and within-day demand variation, congestion, adverse weather, and road incidents. Typical phenomena include frequency instability at downstream stops such as vehicle bunching, unexpectedly lengthy passenger waiting times, and over-crowded services followed by empty runs. Under such circumstances, the interests of both transit passengers and service suppliers are affected. Transit passengers may be subjected to travel time unreliability, and service suppliers may be subjected to profit fluctuation and poorly specified levels of service. Hence, uncertainty is an inevitable aspect of transit planning, especially affecting passenger flow prediction and transit network design.

In this study, transit network uncertainties have been examined from the perspectives of both transit passengers and suppliers. Two new dynamic transit network assignment models have been developed to reflect passenger reaction to network uncertainties. The first is a single-class reliability-based transit assignment model, developed to reflect risk-averse passenger travel decisions as to departure time and route choices. In order to account for travel time reliability, passenger effective travel time, which includes average travel time plus a safety margin to cope with uncertainty, is adopted as dis-utility function. This model is formulated as a fixed point problem which can be solved by a heuristic solution algorithm. A numerical example shows the existence of service deviations under transit network equilibrium conditions, such as vehicle bunching and overtaking.

The second is a new multi-class reliability-based transit assignment model. A safety margin is differentiated for different passenger classes, as passengers have different risk-taking attitudes towards random generalized travel costs (including both travel times and monetary costs). Network congestion is also reflected in this model by introducing an overload parameter on vehicle design capacity constraint on random passenger boarding demand. A network example connecting the Kowloon area to Hong Kong International Airport illustrates the ability of this model to demonstrate that different passenger risk-taking attitudes greatly impact passenger route and departure time choices, and subsequently both monetary and time costs.

With taking account of the effects of network uncertainties, another two transit line scheduling models have been developed to serve a transit supplier's different planning purposes. Of concern to the transit authority, social welfare aspects, such as passenger travel time efficiency and reliability, are improved by a proposed line scheduling model. The interaction between service supply and passenger behavior response is reflected in the bi-level formulation of the scheduling problem. The upper level of the model optimizes the integrated transit service attributes, while the lower level predicts passenger travel decisions under transit network uncertainties. The bi-level problem is solved by applying the genetic algorithm (GA). The numerical results show that transit service reliability under network uncertainties can be improved by the adjustment of line schedules, without the need for extra vehicle resources.

Of relevant to the private transit operator, a new transit line scheduling model has been proposed to reflect the competition between operators in the deregulated transit market under conditions of network uncertainty. The operator's profit is considerably affected by service irregularity, as well as the passenger's response to the irregular service. The operator's risk preference determines how the variability of random profit is measured. Thus, the objective of the operator in the transit line scheduling model is to maximize the α-confident profit, defined as the stochastic profit within a confidence threshold. The passenger's response to the change of line schedules is formulated as a reliability-based user stochastic equilibrium (RSUE) constraint. The α-confident profit maximization model is formulated as a variational inequality (VI) problem and solved by an adapted diagonalization algorithm. This model shows that the ignorance of network uncertainties and operator risk preferences can result in over-optimism on profit when developing the transit line schedules.

PUBLICATIONS ARISING FROM THE THESIS

JOURNALS

Zhang, Y.Q., Lam, W.H.K., and Sumalee, A. (2010) Transit schedule design in dynamic transit network with demand and supply uncertainties. *Journal of the Eastern Asia Society for Transportation Studies*, Vol.8, pp. 1425-1435.

Zhang, Y.Q., Lam, W.H.K., Sumalee, A., Lo, H.K., and Tong, C.O. (2010) The multi-class schedule-based transit assignment model in network with uncertainties. *Public Transport*, Vol. 2, pp. 69-86.

CONFERENCES

Zhang, Y.Q., Lam, W.H.K., and Sumalee, A. (2009) Dynamic transit assignment model for congested transit networks with uncertainties. *The 88rd Transportation Research Board Meeting*, CD-ROM, no. 09-2486.

Zhang, Y.Q., Lam, W.H.K., and Sumalee, A. (2008) Schedule-based transit assignment model in transit network with recurrent uncertainties. *Proceedings of the 13th Conference of the Hong Kong Society for Transportation Studies (HKSTS),* pp. 41-50.

Zhang, Y.Q., Lam, W.H.K., and Sumalee, A. (2009) Transit scheduling design in dynamic transit network with demand and supply uncertainties. Proceedings of *the 8th International Conference of Eastern Asia Society for Transportation Studies (EASTS).* no. 100383

Zhang, Y.Q., Lam, W.H.K., Sumalee, A., Lo, H.K., and Tong, C.O. (2009) The multi-class schedule-based transit assignment model in network with uncertainties.

Proceedings of *the 11th International Conference on Advanced Systems for Public Transport (CASPT).* MS-93.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my supervisor, Professor William H.K. Lam, for his valuable advices and constant encouragements. He is always sincerely involved in the wellbeing of the students and brings the greatest potential one student has. Throughout my studies in Hong Kong, Professor Lam’s attitude towards doing research in a scientific and critical way has deeply influenced me. I would like also to express my great gratitude to my co-supervisor, Professor Agachai Sumalee, one of the most brilliant young scholars. They will always be the role models in my scientific research.

My appreciation is also extended to the Research Grants Council of the Hong Kong Special Administrative Region for the support to this research project (RGC Ref. PolyU 5202/06E and Project No. B-Q01R).

Finally, special thanks must be given to Mrs. Elaine Anson for her sincere help in proofreading this thesis. My colleagues and friends at The Hong Kong Polytechnic University deserve my appreciation as well. They are Dr. Z.C. Li, Dr. H. Shao, Dr. Z.Y. Wu, L.Q. Ouyang, B.Y. Chen, Lucia Pan, R.X. Zhong, J.S. Zhu, D.J. Ye, Y.L. Xiong, D. Liu, K.S. Chan and S.M.L. Tam. They provided me warm friendship, precious caring and inspiring discussions.

This thesis is dedicated to my parents, and also to Prof. Zhang Guowu, one of my

referees for this research study, for their endless support and encouragement during the past years.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

NOTATION

The following notations are used throughout the thesis unless otherwise specified.

Acronyms

AEL	Airport Express Line
BO	Bus Operator
CBD	Central Business District
CDF	Cumulative Density Function
CV	Correlated Variation
DSUE	Dynamic Stochastic User Equilibrium
GA	Genetic Algorithm
GNE	General Nash Equilibrium
HKIA	Hong Kong International Airport
IID	Independently and Identically Distributed
KMB	Kowloon Motor Bus Company Ltd.
MSA	Method of Successive Average
MTR	Mass Transit Railway
OD	Origin and Destination
PAB	Passenger Arriving and Boarding
QVI	Quasi-Variational Inequality
RSUE	Reliability-based Stochastic User Equilibrium
RA	Risk-Averse

RN	Risk-Neutral
RO	Rail Operator
SD	Standard Deviation
SUE	Stochastic User Equilibrium
VI	Variational Inequality

Network sets and variables

$\mathbf{I}$	Set of transit runs i
$\mathbf{L}$	Set of transit lines l
$\mathbf{J}$	Set of nodes (transit stops) j
$\mathbf{\Omega}$	Set of graph to represent a transit network, $\mathbf{\Omega(I,J,L)}$
$\mathbf{R}$	Set of origins r
$\mathbf{S}$	Set of destinations s
$\mathbf{U}$	Set of transit route u
$\boldsymbol{\tau}$	The time interval of the passenger arriving process, $(...\tau_{i-1}, \tau_i, \tau_{i+1},...) \subseteq \boldsymbol{\tau}$
$\lambda(\boldsymbol{\tau})$	Passenger arrival rate at the time interval $\boldsymbol{\tau}$
$\mathbf{M}$	Set of the passenger class m
$\mathbf{K}$	Set of the private transit operator k

Stochastic variables and Functions

Q^{rs}	The total random passengers demand between an origin-destination (OD) pair r-s over the study period

Q_l	The total random number of passengers for a transit line l
$Q_l(\tau)$	Inhomogeneous Poisson process of passengers arriving within time interval τ of line l
Q_l^j	The random passengers flow in j th run of line l
Al	Passengers alighted for transfer
$N(\tau_i)$	The stochastic number of passengers boarded within time interval τ_i
$B(\tau_i)$	Duration for the passenger boarding during time interval τ_i
B_{per}	The stochastic boarding time for each passenger
$B_{per}(x)$	The cumulative density function of the boarding time for each passenger
$B_{per}^{N(\tau)}(x)$	The $N(\tau)^{\text{th}}$ convolution of $B_{per}(x)$
$B_{per}^{n_2}(x)$	The n_2^{th} convolution of $B_{per}(x)$ at Phase 2 of PAB process
$B_s^l(\tau)$	The total passenger boarding time at stop s of line l
$T_{i,j,l}^a$	The actual arrival time and departure time of the i^{th} transit vehicle of line l at stop j
$T_{i,j,l}^d$	The actual arrival time and departure time of the i^{th} transit vehicle of line l at stop j
$Av_j^l(t)$	The earliest vehicle arrival time with respect to the passenger arrival time t of line l at stop j
$Tv(t)$	The transit stochastic vehicle on-road time between two adjacent stops
$Tw(t)$	The stochastic passenger waiting time for vehicle

$\tilde{T}w$	The delayed waiting time if there is overload delay from the previous vehicle
$Tr(t)$	The stochastic passenger transfer time is the transfer of line is needed
C	The stochastic travel time including passengers waiting time, in-vehicle travel time, and transfer time
ε	The passenger perception error
TC	The total stochastic passenger travel time, including the effective travel time, early or late arrival penalty, and the perception error
GC	Passenger generalized travel cost composed of the total stochastic passenger travel time and the monetary cost
$\varphi(x)$	The Normal cumulative probability function
$P(\cdot)$	Passenger flow proportion function
$F_{l,i}$	The stochastic passenger flow of i^{th} transit vehicle of line l
$F_{t,u}$	The passenger stochastic flow on route u departing at time t
$\mathbf{F}$	The set of passenger flow variable, $F_{t,u} \in \mathbf{F}$
ΔCAP_l^j	The difference between on-board passengers and vehicle capacity
Cv_l	The stochastic vehicle on-road time of operation cost
Cw_l^i	The stochastic vehicle dwell time of operation cost
D_p	The stochastic variable of dispatching an unplanned vehicle
Cd_l	The cost of dispatching an extra vehicle
CO_l	The stochastic operation cost for a transit line

Cv_l	The stochastic running time cost
Cw_l^i	The stochastic dwelling time cost
Φ^k	Operator k's stochastic net profit
R^k	The stochastic revenue of operator k
C^k	The stochastic cost of operator k
S^k	The operator's risk preference margin

Deterministic Variables and Functions

n_1	The mean number of passengers boarding at Phase 1 of the PAB process
n_2	The mean number of passengers boarding at Phase 2 of the PAB process
q_j'	The number of passengers waiting at the beginning of Phase 2 of the PAB process in interval τ_j
t	The passenger arrival time at a stop, also denoted as the passenger departure time as access time is not allowed for in this study
t^s	The passenger's arrival time at destination s
Δ_1^s	The passenger's early arrival window
Δ_2^s	The passenger's late arrival window
$t^s - \Delta_1^s$	The lower bound of the desired arrival time window for the passengers arriving at destination s
$t^s + \Delta_2^s$	The upper bound of the desired arrival time window

$tp(t)$	The early or late penalty with respect to passenger arrival time t
ett	The effective passenger travel time
gc	The summation of the deterministic variables in passenger generalized travel cost GC
$\mathbf{f}$	The vector of mean of passenger flow f
$\mathbf{q}$	The vector of expected passenger O-D demand q^{rs}
cap_l	The actual vehicle capacity
ga	The gap function of passenger flow between successive iteration in the algorithm
$sf(\cdot)$	The step function specifying multi-class passengers' safety margin a
h	The dispatching headway of lines
$\mathbf{h}$	The vector of transit line headways
Z	The objective function of designer's integrated optimization of transit service efficiency and reliability
$a_{l,u}$	The line-route incidence variable
$\mathbf{A_{l,u}}$	The line-route incidence matrix $\mathbf{A_{l,u}} = (..., a_{l,u}, ...)$
$b_{l,k}$	The line-agency incidence variable
$\mathbf{B_{l,k}}$	The line-agency incidence matrix $\mathbf{B_{l,k}} = (..., b_{l,k}, ...)$
$\mathbf{\Pi_{l,i}}$	The line-run incidence matrix
p^d	The probability the extra vehicle is used according to the current line schedule

ϕ	The operator's desired profit
cf_l	The transit fare of line l
co_l	The operation cost of line l
cc_l	The capital cost of line l
ψ	The utility function of one operator
$g_{l,i}$	The dispatching time of i th vehicle of line l
G^k	The strategy set of line schedules of operator k, $G^k \in R^I$
G	The full Cartesian product of operators' line schedule strategy sets
$\mathbf{y}^k$	The vector of variables for α confident profit maximization

Parameters

μ_b	The mean of an individual passenger's boarding time
σ_b	The standard deviation of an individual passenger's boarding time
ω	The probability of over-crowding in vehicles
q'_j	The number of passengers waiting at the beginning of interval τ_j.
$\boldsymbol{\alpha}$	The vector of the confidence level of the passenger class i, $\alpha_i \in \boldsymbol{\alpha}$; and operator k, $\alpha^k \in \boldsymbol{\alpha}$
β'	The unit value of time of arriving early at destination
β''	The unit value of time of arriving late at destination
β_1	The weighting coefficient of passenger waiting time

β_2 The weighting coefficient of passenger on-road travel time

β_3 The weighting coefficient of passenger transfer time

β_4 The weighting coefficient of early or late penalty

β_5 The weighting coefficient of transit fare

θ The OD demand multiplier

γ_1 The convention factor for unit equivalence of passenger network travel time and the variance

μ_l^{cd} The mean of dispatching cost of an extra transit vehicle

σ_l^{cd} The standard deviation of dispatching cost of an extra transit vehicle

β_v The coefficient variation of vehicle stochastic on-road travel time

β_b The coefficient variation of passenger stochastic individual boarding time

μ_{GC} The mean of passenger generalized travel cost

σ_{GC} The standard deviation of passenger generalized travel cost

CHAPTER 1

INTRODUCTION

1.1 STATEMENT OF THE PROBLEM

Hong Kong is a densely populated city, with a total population of 7.026 millions, and with over 11 million personal trips being made daily. The public transport system carries over 90% of these trips for various activities. However, the current increasing travel demand provides great challenges as regards the provision of efficient and reliable transit services, particularly, as far as Hong Kong is concerned, for bus modes in congested road network. Insufficient capacity and unreliable service at peak periods are frequent bus passenger complaints. Dissatisfaction with the service offered is likely to cause a loss of bus patronage and a change for other transport modes.

The causes of insufficient bus capacity and unreliable service during peak and off-peak periods often lie in transit network uncertainties. The transit network uncertainty phenomena, such as long passenger queues, passenger overload delays, vehicle bunching and vehicle overtaking at bus stops, commonly exist in most metropolitan cities. The sources of transit network uncertainties, analogous with road network uncertainties, are derived from both demand and supply sides. The differences, however, lie in many aspects. On the demand side, uncertainties stem not only from within-day or day-to-day demand variation, but also from passenger random arrival patterns at bus stops. On the supply side, congestion, adverse weather conditions, and road incidents result not only in stochastic on-road travel times, but also in vehicle unbalanced dwell times at bus stops and unbalanced passenger loads. All of the above lead to further poor transit system performance.

To tackle these uncertainty problems, in the short term, a dynamic transit assignment model is needed to predict passenger flow and the system operation status at peak periods. Passengers' temporal travel characteristics, such as departure times at the origins of journeys and estimated arrival times at destinations, are important factors which affect their choices of travel routes and modes. Passengers' individual reliability evaluations, such as travel time reliability and arrival time punctuality, are also determining factors in their travel choices. Further, the consistency of line schedules in the provision of transit service especially affects passenger travel decisions.

Transit line schedule design is an important factor promoting system improvement at the operational planning stage. The design of vehicle dispatching times at the origin terminal can balance passenger boarding demand and maintain the reliability of passenger waiting time. Consideration of network uncertainties, such as the with-in day or day-to-day passenger flow fluctuations and service irregularities, can be built into line schedule design models with the aim to improve the level of transit service. Transit line schedule design allows for different market regimes, i.e. regulated or deregulated transit markets. The regulated market transit system is often operated by government authorities and aims to improve total social welfare, whereas the deregulated market system is often operated by private operators and aims to maximize individual company profits. For both types of service suppliers, transit line schedule design, allowing for network uncertainties, can improve service efficiency and effectiveness. The optimal transit line schedules can save passenger waiting time, improve service travel time reliability, attract passenger patronage and balance fleet size between lines.

1.2 LITERATURE REVIEW

Interests in dynamic transit operation and design models stem from the need to predict daily passenger travel times and peak period passenger flows under specific service configurations. Transit assignment and line schedule problems are basic and typical problems affecting transit short-term planning. Models proposed for such problems, however are mostly built, assuming either that network configurations are precisely known without variations, or that the transit system runs in an ideal totally predictable way without uncertainties. Such assumptions overlook transit network uncertainties and the resulting influences brought by these uncertainties to a transit system. The following brief and general literature review aims at identifying a potential research area. The advantages and limitations of existing models given and explored in the literature are discussed. Additional examples, specifically relevant to the contents of each of the following chapters are presented for ease of access, at the head of each chapter.

Transit planning models presented in the literature, mostly differ as regards the following: (a) the planning horizon, and (b) the network environment. Regarding the planning horizon viewpoint, public transportation planning has been categorized broadly into long-term and short-term problems (Black, 1981). The essential elements considered in long-term transit planning are mainly resource consumption (budget, subsidy, and pricing) and environmental changes (land use policy and population density). The researchers however have focused on short-term transit planning as affected by transit network characteristics and passenger behavioral responses. As regards the network environment viewpoint, problems that received most attention are network congestion (Ghoseiri et al., 2004), connectivity (Beimborn et al., 2003), and uncertainty (Teodorovic et al., 1994; Yan and Tang, 2008). The effects of network uncertainties, in particular, have received much recent attention (Hickman, 2001; Nuzzolo et al., 2001; Yang and Lam, 2006; Yan and Tang, 2008). It appears, however, that the influences of these effects in a dynamic environment have not been sufficiently studied.

The classification of existing transit network assignment and design models as regards both the planning horizons and network environments is shown in Table 1.1.

It is assumed in the static transit assignment (STA) and strategic transit network design (STND) models that both the transit network attributes and passenger flow patterns are deterministic in relation to long-term planning. The reliability-based static transit assignment (RSTA) and strategic stochastic transit network design (SSTND) models relaxed the assumption of an ideal network environment with its deterministic passenger demand and actual service configuration. Instead, passengers are allowed to select transit routes on the basis of travel time budget (the average travel time plus a safety margin for on-time arrival reliability). Operators are allowed to select transit design schemes which improve transit service reliability or accommodate network uncertainties by their risk preferences, respectively from the design objectives by public or private operators. However, these models ignore day-to-day and within-day demand and service changes. The transit network was only studied taking a long-term static perspective when models were formulated.

To model temporal passenger demand and transit service changes, it is necessary to extend the transit network from a static one to a dynamic one. The demand variations, service attribute evolutions and passenger behavioral responses are typical aspects which need to be considered at the short-term planning stage. When the dynamic transit assignment (DTA) and transit network operation (TNO) models were formulated, the network environment was assumed reliable. Neither of these models was influenced by adverse weather conditions and road incidents. Such models may be applicable to cities with low public transport demand, where transit services arrive on schedule and service frequency is low. However, for more frequent transit services in cities where passenger demand is high and road networks are also congested, transit network uncertainties certainly exist and affect transit system operation. To the best of the author's knowledge, no assignment and network design models have been developed under network uncertainties and dynamic modeling horizon. To fill this research gap, the present study proposes two reliability-based dynamic transit assignment (RDTA) transit assignment models for the single and multiple passenger classes and two stochastic transit network operation (STNO) transit network design models, seen from government authority and private company's perspectives.

Table 1.1 Classifications of Transit Network Assignment and Design Models

			Network uncertainties?	
			No	Yes
Modeling Horizon?	Long-term (Static)	Assignment	STA	RSTA
		Operation and Design	STND	SSTND
	Short-term (Dynamic)	Assignment	DTA	RDTA
		Operation and Design	TNO	STNO

where

STA	=	Static Transit Assignment Model (Last and Leak, 1976; Cominetti and Correa, 2001)
STND	=	Strategic Transit Network Design Model (Murray et al., 1998)
DTA	=	Dynamic Transit Assignment Model (Hamdouch and Lauphongpanich, 2008)
TNO	=	Transit Network Operation Model (Fu et al., 2003; Lee and Vuchic, 2005)
RSTA	=	Reliability-based Static Transit Assignment Model (Yang and Lam, 2006)
SSTND	=	Strategic Stochastic Transit Network Design Model (Li et al., 2009, Sumalee et al., 2006)
RDTA	=	Reliability-based Dynamic Transit Assignment Model
STNO	=	Stochastic Transit Network Operation Model

1.3 RESEARCH QUESTIONS AND OBJECTIVES

In this study, three important aspects of the transit design and operation processes have been investigated: 1) passenger travel behaviors under network uncertainties, 2) transit line schedule design by service suppliers allowing for their own risk preferences, and 3) passenger responses and feedbacks on the reliability of transit service configurations.

To achieve the objectives (set out immediately below), the following research questions have to be answered: “How do network uncertainties arise and how do they affect transit operators and passengers?” Hence the resulting research problems have three dimensions: the sources of uncertainties, passenger behavior responses, and operator planning strategies.

In line with the above dimensions, the objectives of this research are as follows:

(1) To specify sources of network uncertainties from both the demand and supply sides and to discuss the impacts on transit passengers and operators

(2) To develop a dynamic transit assignment model for short-term planning with explicit consideration of passenger responses to network uncertainties in terms of travel time reliability

(3) To extend the single-class stochastic transit assignment model to multi-class for estimating passenger flows in congested transit networks

(4) To investigate the transit line scheduling problem in an unreliable network environment and hence to optimize transit system performance as measured by service efficiency and reliability

(5) To optimize and predict private operators' transit schedule design schemes taking account of their risk preferences in a competitive transit market.

1.4 STRUCTURE OF THE THESIS

A novel dynamic transit assignment model under uncertainties is described in Chapter 2. Sources of uncertainties derived from demand and supply sides are discussed, and the resulting stochastic service configurations are given. The schedule-based dynamic transit assignment approach is used to model the dynamic transit network uncertainties and passenger behavioral responses. A dynamic network loading procedure embedding the stochastic passenger arriving and boarding (PAB) process is also presented. The proposed reliability-based stochastic user equilibrium (RSUE) transit assignment model is formulated as a fixed-point problem and solved using a heuristic algorithm. Some key findings are illustrated using a test network.

The transit assignment model given in Chapter 2 is further developed in Chapter 3. A new multi-class RSUE dynamic transit assignment model is proposed. Heterogeneous passenger risk-taking attitudes towards random travel time are considered. Passenger route and departure time choices differ, not only from their travel time perception, but also their individual reliability requirements. The vehicle design capacity constraint on the stochastic passenger demand is also discussed, aiming to reflect the variation of passenger in-vehicle loading. The merits of the proposed model are illustrated by a numerical example, based on a simplified transit network connecting Tsing Yi new town to Hong Kong International Airport.

A transit line scheduling problem with network uncertainties is explored and described in Chapter 4. The objective of the line scheduling problem is to integrate service reliability improvements into an overall average target of saving total passenger network travel time. A bi-level problem is formulated to find the optimal schedule scheme. An array of uneven headways for each line is optimized in the upper level model, while passenger responses to line schedule schemes are considered in the lower level model. The bi-level problem is solved by a Genetic Algorithm (GA). The numerical result shows that the proposed model can utilize existing fleet resources to reach an optimization objective in terms of operational efficiency and service reliability.

The line scheduling problem from a transit operator perspective is explored and the results are given in Chapter 5. Each transit operator’s line schedules under a competitive market and uncertain network conditions are optimized. Both operator and passenger risk preferences are taken into account. This line schedule problem is modeled as a Variational Inequality (VI) problem with equilibrium constraints. The objective is to maximize the individual operator's α-confidence profit, defined as the stochastic profit within a confidence threshold. The equilibrium constraint is a RSUE problem. The diagonalization algorithm is adapted to solve the VI problem. A simple network is used to illustrate the performance of the model and solution algorithm. The structure of the thesis is shown in Figure 1.1.

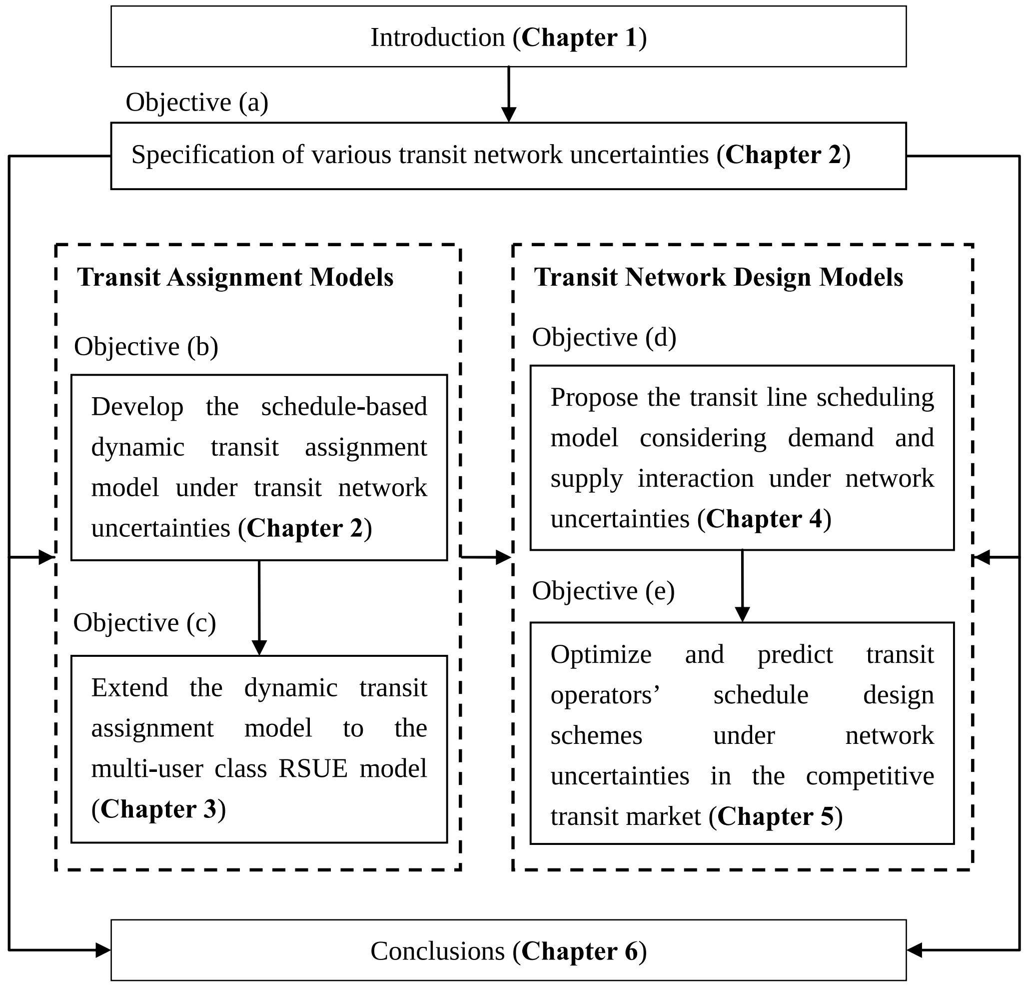


Figure 1.1 Structure of the Thesis

CHAPTER 2

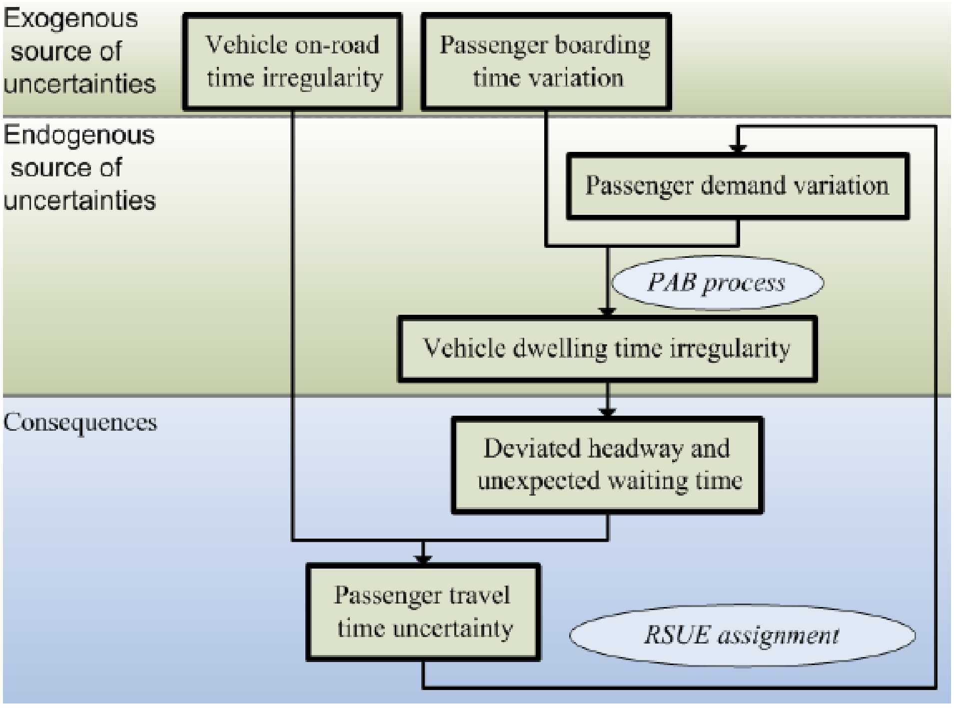
DYNAMIC TRANSIT ASSIGNMENT MODEL FOR CONGESTED TRANSIT NETWORKS WITH UNCERTAINTIES

A novel dynamic transit assignment model with demand and supply uncertainties is proposed and described in this chapter. The demand uncertainty is due to passenger random arrival at transit stops and day-to-day travel demand variation. The supply uncertainty is due to vehicle on-road travel time and dwelling time variability. The concept of passenger effective travel time (travel time budget) was adopted in the passenger travel choice model to account for passenger risk-taking attitudes towards the unreliable travel time. The interaction of demand and supply uncertainties, represented as passenger arriving, queuing and boarding stochasticity, is also modeled explicitly in the passenger travel choice model. In particular, the analytical expression of passenger waiting time and vehicle dwelling time are derived. A new network loading procedure is presented to capture the evolution of transit service configurations. The proposed model is formulated as a fixed-point problem and solved by heuristic solution algorithm. The numerical results show that this model can generate passenger travel route and departure time choices under uncertainties and generate the resulting transit service spatial and temporal attributes. This chapter is an edited version of: Zhang, Y.Q., Lam, W.H.K. and Sumalee, A. (2009) Dynamic transit assignment model for congested transit networks with uncertainties. *The 88rd Transportation Research Board Meeting, CD-ROM,* 09-2486.

2.1 INTRODUCTION

A reliable transit system can provide efficiency and productivity for both passengers and transit agencies. Deterioration of transit networks, however, commonly exists and severely affects level of transit service. For example, vehicle double-heading, knocking-on (bunching) and overtaking phenomena usually occur, causing unexpectedly long waiting times for passengers and low productivity for transit runs. Various topics, such as route performance enhancement (Powell and Sheffi, 1983; Sumalee et al., 2006), optimal vehicle holding time (Hickman, 2001), and optimal slack time (Carey, 1998; Zhao et al., 2006) have been investigated, as described in the literature, to alleviate the impact of uncertainty on the performance of transit systems.

Recently, attention has been given to the effects of uncertainties related to what has been termed, the transit network assignment problem. Nuzzolo et al. (2001) developed a doubly dynamic assignment model to investigate the demand and supply interaction with regular and irregular transit services. The source of uncertainties considered in their paper, however, was mainly the interaction between transit vehicles and private cars on the roads, while the irregular service configuration was given exogenously. Latterly, Yang and Lam (2006) have proposed a probit-type reliability-based transit assignment model for a congested transit network with unreliable services. The transit vehicle running time was assumed to follow a normal distribution, and passenger travel choice was based on the quadratic travel disutility function proposed by Yin and Ieda (2001). However, the dynamic effects in transit networks have not yet been considered, as their model was mainly designed for strategic planning rather than operational improvements.



Note: PAB = Passenger Arriving and Boarding

RSUE = Reliability Stochastic User Equilibrium

Figure 2.1 Sources, Evolutions and Interactions of Transit Network Uncertainties

Systematically, sources of uncertainties can be divided into the supply side (road conditions, weather, and incidents) and demand side (within-day or day-to-day variation of passenger demand). As shown in Figure 2.1, passenger boarding time variation and vehicle on-road running time irregularity are regarded as exogenous uncertainties. These factors are independent of passenger route and departure time choices, but do affect these choices in turn. Hence the stochastic passenger arrivals over time are endogenously decided by passenger behavior responses. Vehicle dwell time uncertainty is derived from the stochastic passenger boarding process, which is determined by the random passenger boarding time (exogenously) and the number of boarding passengers (endogenously).

By the explicit consideration of passenger waiting time and the transit deviations from timetables, a dynamic transit assignment model is proposed for transit operation improvement. Specifically, a dynamic network loading model embedding a stochastic process is proposed to allow for the uncertainty in passenger arriving and boarding (PAB) process. This process is explicitly described in Section 2.2.1 and 2.2.2. By embedding the PAB process, the proposed model can assign time-dependent passenger flow and illustrate transit system evolution in an uncertain and congested transit network.

In Section 2.2 of this chapter, a dynamic transit assignment model allowing for the exogenous and endogenous uncertainties is formulated. The analytical expressions for stochastic transit service variables are derived in Subsection 2.2.1. Passenger attitude towards risk and the related effective travel time is presented in Subsection 2.2.4. The solution algorithm embedding a novel dynamic loading model is presented in Section 2.3. Finally, a numerical example is presented, followed by the study conclusions and discussion of further research.

2.2 MODEL FORMULATION

The passenger demand day-to-day variation is assumed to follow the Normal distribution. Q^{rs} denotes the passenger demand between an origin-destination (OD) pair r-s on a specific day. For each time interval $\tau_i \in (...\tau_{i-1}, \tau_i, \tau_{i+1},...)$ of the period under investigation, it is further assumed that passenger arrivals during each time interval τ_i for a line l is an inhomogeneous Poisson process $\{Q_l(\tau), \tau \geq 0\}$ with arrival rate $\lambda(\tau)$. According to the Central Limit Theorem, the total number of passengers for each line l over the whole period Q_l can be approximated by $Q_l(\tau)$. Passenger demand for all OD and for all lines can then be defined as:

$$\sum_l (Q_l - Al_l) = \sum_{r,s} Q^{rs}, \tag{2.1}$$

where Al_l is the number of transfer passengers joining line l.

When the transit vehicle arrives, if the vehicle has spare capacity, passengers waiting in the queue will start boarding. However, the number of passengers boarding and the total boarding time can not be easily ascertained because the time interval between successive passenger arrivals and the boarding time for each passenger are stochastic. Even though the boarding time mean value and passenger arrival rate are easier to obtain, these mean values are not adequate for the representation of transit network configurations and passenger travel behavior.

To model the passenger boarding process, Renewal theory and the M/G/1 queue technique are applied to obtain the analytical expression for boarding passenger loads and total passenger boarding time, and thereby the stochastic vehicle dwelling time and run headway. The Markovian property of the PAB process is introduced below. The formulation of the other stochastic variables mentioned above is then presented.

2.2.1 Passenger Arriving and Boarding (PAB) Process

In the following discussion, the stochastic process of passenger arriving, queuing and boarding is described by a set of variables. The notation used particularly for this section is given below:

$\mathbf{E}$	The state space of Markov chain, $\mathbf{E} = \{0,1,2,...\}$
π	The time period of the PAB process
t	The time interval for passenger boarding, $t \in \pi$
N_t^+	The stochastic number of passengers boarding the vehicle at time interval t
v_t	Passengers get aboard at time interval t, $v = 0,1,...,i,k,...$
$\lambda(t)$	The passenger arrival rate at time interval t
$G(t)$	The cumulative density function of individual passenger boarding time
p_k	The probability that the number of boarding passengers equals k

p_{ik} The one-step transition probability of the number of boarding passengers from i to k

B The stochastic busy period of the PAB process:

D The stochastic number of boarding passengers during the PAB process

μ_b The mean of passenger boarding time (passenger boarding rate)

σ_b The standard deviation of passenger boarding time

ρ The traffic intensity of the PAB process

Proposition The passenger arriving and boarding process $\{N_t^+, t \geq 1\}$ generated at each time interval t is an irreducible, periodic, and time-homogenous Markov chain and the one-step probability is:

$$P\{v = k - i + 1\} = \begin{cases} \int_0^\infty \frac{(\lambda(t)t)^k}{(k)!} e^{-\lambda(t)t} dG(t), & i = 0, \\ \int_0^\infty \frac{(\lambda(t)t)^{k-i+1}}{(k-i+1)!} e^{-\lambda(t)t} dG(t), & k \geq i-1,\ i \geq 1, \\ 0, & k < i-1,\ i \geq 1. \end{cases} \tag{2.2}$$

Proof. During the time period π, assuming v_t represents the number of passengers boarding at time interval t, it is obvious that $\{v_t, t \geq 1\}$ is IID (independently and identically distributed):

$$p_k = P\{v_t = k\} = \int_0^\infty \frac{(\lambda(t)t)^k}{k!} e^{-\lambda(t)t} dG(t),\ k \geq 0,\ t \in \pi, \tag{2.3}$$

which means the number of boarding passengers at time interval t is independently and identically following the right-hand-side distribution.

Let N_t^+ represent the number of passengers boarding at time interval t, the number of passengers boarding at time interval $t+1$ is then:

$$N_{t+1}^+ = \begin{cases} N_t^+ - 1 + v_{t+1}, & if \quad N_t^+ > 0, \\ v_{t+1,} & if \quad N_t^+ > 0, \end{cases} \quad t \geq 1. \tag{2.4}$$

Since $\{v_t, t \geq 1\}$ is IID, define $v_t = v$, $t \geq 1$, there is:

$$N_{t+1}^+ = \begin{cases} N_t^+ - 1 + v \ , & if \quad N_t^+ > 0, \\ v_, & if \quad N_t^+ > 0, \end{cases} t \geq 1. \tag{2.5}$$

It can be observed from the above equation, that when N_t^+ is known, N_{t+1}^+ is simply related to the arriving process, but not related to the previous $N_1^+, N_2^+, ..., N_t^+$. Thus $\{N_t^+, t \geq 1\}$ is a Markov chain with state space $\mathbf{E} = \{0,1,2,...\}$.

The one-step transition probability is:

$$p_{ik} = P\{N_{t+1}^+ = k \mid N_t^+ = i\} = \begin{cases} P\{v = k - i + 1\}, & i \geq 1 \\ P\{v = k\}, & i = 0 \end{cases}. \tag{2.6}$$

When $i \geq 1$:

$$P\{v = k - i + 1\} = \begin{cases} \int_0^\infty \frac{(\lambda(t)t)^{k-i+1}}{(k-i+1)!} e^{-\lambda(t)t} dG(t), & k \geq i - 1 \\ 0, & k < i - 1 \end{cases}. \tag{2.7}$$

When $i = 1$:

$$P\{v = k\} = \int_0^\infty \frac{(\lambda(t)t)^k}{k!} e^{-\lambda(t)t} dG(t). \tag{2.8}$$

Hence Equation (2.2) is obtained.

It is obvious from the one-step transition expression that p_{ik} $(i, k = 0,1,2,...)$ is not related to the time origin. $p_{ik} > 0$ means that there is a non-zero probability from one state to the other. Thus, the two auxiliary states are accessible. Hence,

$\{N_t^+, t \geq 1\}$ is an irreducible, periodic, and time-homogenous Markov chain.

To maintain integrity, the following definitions of the M/G/1 queue are presented below (Medhi, 2003):

Definition 1 In the M/G/1 queuing system, the mean and variance of busy period B is:

$$E[B] = \begin{cases} 1/(\mu_b - \lambda), & \rho < 1, \\ \infty, & \rho \geq 1, \end{cases} \tag{2.9}$$

$$\mathrm{var}[B] = \begin{cases} ({\sigma_b}^2 + \rho{\eta_b}^2)/(1-\rho)^3, & \rho < 1, \\ \infty, & \rho \geq 1, \end{cases} \tag{2.10}$$

where λ and μ_b are the rates of arrival and service respectively, and $\rho = \mu_b / \lambda$ is the traffic intensity. σ_b is the standard deviation of the service time.

Definition 2 In the M/G/1 queue, during the busy period, the mean and variance of the number of customers served D is:

$$E[D] = \begin{cases} 1/(1-\rho), & \rho < 1, \\ \infty, & \rho \geq 1, \end{cases} \tag{2.11}$$

$$\mathrm{var}[D] = \begin{cases} \rho(1+\rho)/(1-\rho)^3, & \rho < 1, \\ \infty, & \rho \geq 1. \end{cases}. \tag{2.12}$$

Recent transit assignment models (Larrain and Muñoz, 2008) allow for passenger service time (boarding and alighting time) to be taken into consideration with vehicle dwelling time. The connection between vehicle dwelling time and the number of boarding passengers was regarded as deterministic and linear. However, the transit vehicle dwelling is a stochastic process involving passengers arriving, queuing, boarding and alighting. Lam et al. (1998) investigated the train dwelling time at several main rail stations in Hong Kong, and found that the train dwelling time followed a normal distribution.

The bus and train dwelling time are quite different for inward and outward trips. On the outward trip at peak travel time, bus boarding time has a greater weighting than alighting time, and vice versa for the inward trip. It is also common for a bus to wait for a rushing passenger, belatedly attempting to join the bus. The PAB process discussed below is particularly modeled for outward bus trips; otherwise both the service time and boarding passengers are likely to be inaccurately estimated in the stochastic environment.

The behavior of passengers arriving, queuing and boarding and relating these activities to vehicle dwelling time, were analyzed using the PAB process. The process is divided into two consecutive phases. Phase 1 relates to a continuous boarding process and Phase 2 relates to those occasions when vehicles wait for passengers, hurrying to catch the vehicle. Both phases are constrained by vehicle capacity. Within the small time interval τ_i (say 1 minute), passenger boarding is a Renewal process, and the mean and variance of passenger boarding time $B(\tau_i)$ can be derived as:

$$\mathrm{E}[B(\tau_i)] = \int_0^\infty x dB_{per}^{N(\tau_i)}(x), \tag{2.13}$$

and

$$\mathrm{var}[B(\tau_i)] = \int_0^{\infty} x^2 dB_{per}^{N(\tau_i)}(x^2) + [\int_0^{\infty} x dB_{per}^{N(\tau_i)}(x)]^2 \ , \tag{2.14}$$

where $B_{per}^{N(\tau_i)}(t)$ is the $N(\tau_i)$ th convolution of $B_{per}(t)$. $N(\tau_i)$ is the stochastic number of passengers boarding within time interval τ_i. $B_{per}(t)$ is the cumulative density function of the boarding time for each passenger B_{per}, which follows the Normal distribution $B_{per} \sim N(\mu_b, \sigma_b^2)$. Note that the coincidence of the mean of total passenger boarding time and the time interval τ_i (the number of passengers boarding equals the service rate):

$$\mathrm{E}[B(\tau_i)] = \int_0^{\infty} x dB_{per}^{N(\tau_i)}(x) = \mathrm{E}[B_{per}^{N(\tau_i)}] = \mathrm{E}[N(\tau_i)] \cdot \mathrm{E}[B_{per}] = \tau_i \ . \tag{2.15}$$

Hence, the mean number of boarding passengers in Phase 1 is:

$$n_1 = \sum_i \mathrm{E}[N(\tau_i)] = \sum_i \frac{\tau_i}{\mathrm{E}[B_{per}]} \ . \tag{2.16}$$

The PAB process is completed in Phase 2, where the number of boarding passengers in this phase is:

$$n_2 = \lambda(\tau_j) + q'_j \ , \tag{2.17}$$

where q'_j is the number of passengers waiting at the beginning of interval τ_j. Applying the same logic as that applied above, the mean and variance of passenger boarding time for this interval is:

$$\mathrm{E}[B(\tau_j)] = \int_0^{\infty} x dB_{per}^{n_2}(x) \tag{2.18}$$

$$\mathrm{var}[B(\tau_j)] = \int_0^{\infty} x^2 dB_{per}^{n_2}(x^2) + [\int_0^{\infty} x dB_{per}^{n_2}(x)]^2 \ . \tag{2.19}$$

The $(n_1 + n_2)$ th convolution of B_{per} still follows a Normal distribution, so the total passenger boarding time at stop s of line l follows the Normal distribution:

$$B_s^l(\tau) \sim N[(n_1 + n_2)\mu_b, (n_1 + n_2)\sigma_b^2]. \tag{2.20}$$

During the morning peak hour, when outward trips comprise the heaviest travel demand, the above PAB model can capture the main service process. During the afternoon peak, when alighting occupies the main service time, the alighting time model developed by Adamski (1992) and Lam et al. (1998) can be added to the passenger choice models.

2.2.2 Transit Route and Departure Time Choice Model

The transit network is redefined as a hyper-graph in which the passenger transit travel decisions are similar to those in the road network, i.e. choice of travel route and departure time. The principle of DSUE (Dynamic Stochastic User Equilibrium) was adopted for the representation of transit network temporal characteristics. The total passenger travel time consists of (i) passenger waiting time, (ii) in-vehicle travel time, (iii) in-vehicle waiting time, (iv) transfer time (if transfer is needed), (v) the early and late arrival penalty at destination, and (vi) the passenger travel perception error. Passenger travel strategy is assumed pre-determined, which is appropriate when passengers make travel decisions taking into account travel time reliabilities.

The schedule-based transit network is illustrated in the diachronic graph (Nuzzolo et al., 2003). The transit line operation together with incidents during operation (such as schedule coordination, vehicle encountering and overtaking phenomena) can be represented on such a graph by the time and space illustrations of vehicle trajectories. Given a transit network $\Omega(I,J,L)$ and a defined transit vehicle i of line l at stop j, $T^a_{i,j,l}$ and $T^d_{i,j,l}$ can uniquely define the random vehicle arrival and departure times.

Passenger waiting time is derived from the difference between passenger arrival time t and vehicle arrival time $T^a_{i,j,l}$. $\mathbf{T^a_{i,j,l}}$ is denoted as the vector of vehicle arrival time at stop j for all vehicles of line l. Denote $\tilde{T}w$ is denoted as the delayed waiting time if there is an overload delay from a previous vehicle. $\tilde{T}w$ equals the headway between two sequential vehicles of the same line.

$$Av^l_j(t) = \min[\mathbf{T^a_{i,j,l}} - t - \tilde{T}w]^+ \quad (2.21)$$

represents the earliest vehicle arrival time after passenger arrived at time t of line l at stop j. The passenger waiting time with respect to vehicle arrival time is:

$$Tw(t) = Av^l_j - t. \quad (2.22)$$

The passenger waiting time also follows the Normal distribution, drawn from the distribution of vehicle arrival times. The mean and variance of the passenger waiting time can be defined as:

$$E[Tw(t)] = \mathrm{E}(Av^l_j) - t, \quad (2.23)$$

$$\mathrm{var}[Tw(t)] = \mathrm{var}(Av^l_j). \quad (2.24)$$

The in-vehicle travel time and the in-vehicle waiting time can be derived from the vehicle running time model as:

$$T^{a}_{i+1,j,l} = T^{d}_{i,j,l} + Tv\,, \tag{2.25}$$

$$T^{d}_{i,j,l} = T^{a}_{i,j,l} + Av^{l}_{j}\,, \tag{2.26}$$

where Tv is the transit stochastic vehicle on-road time between two adjacent stops. Similar to the assumption made in Chen et al. (1999) and Lam et al. (2008), Tv is also assumed to follow the Normal distribution. For generality, the on-road time for each transit line segment is independently and identically distributed (IID).

The early or late penalty is deduced from passenger departure time, perceived travel time and desired arrival time at destination:

$$tp(t) = \begin{cases} \beta'(t^{s} - \Delta^{s}_{1} - ett - t) & \text{if } t^{s} - \Delta^{s}_{1} \geq ett + t, \\ \beta''(t + ett - t^{s} - \Delta^{s}_{2}) & \text{if } t^{s} + \Delta^{s}_{2} < ett + t, \\ 0 & \text{otherwise.} \end{cases} \tag{2.27}$$

where $[t^{s} - \Delta^{s}_{1}, t^{s} + \Delta^{s}_{2}]$ is the desired arrival time window at destination s carrying no schedule delay penalty. β' (β'') is the unit value of time of arriving early (late) (i.e. schedule delay) at the destination.

ett is the effective travel time which consists of the mean travel time plus a safety margin which ensures travel time reliability. It is plausible to assume that most passengers in the modeling period are risk-averse since that period consists of the morning or evening peak when most travelers are commuters. Thus, $\alpha = 0.95$ is set for the passenger confidence interval (95% confident of a punctual arrival at destination). The effective travel time is defined as:

$$ett = E[C] + \varphi^{-1}(\alpha) \cdot Std[C], \tag{2.28}$$

where the total passenger travel time C is the sum of passenger waiting time, in-vehicle travel time, in-vehicle waiting time, and transfer time. Each element is multiplied by weighting coefficients to convert each component to an equivalent unit of time:

$$C = \beta_1 Tw + \beta_2 Tv + \beta_3 Tr \, . \tag{2.29}$$

Denoting the perception error as ε, which is a stochastic variable following the Normal distribution, the total perceived travel time of a passenger departing at t, choosing transit route u is:

$$TC(t,u) = ett(t,u) + \beta_4 tp(t,u) + \varepsilon \, . \tag{2.30}$$

The parameter β_4 is the weighting coefficients for early or late penalty.

2.2.3 Fixed-point Problem for RSUE

$\mathbf{F}$ is denoted as the set of stochastic passenger flow variables on route u between OD pair departing at time t, which satisfies:

$$\sum_t \sum_u E[F_{t,u}^{rs}] - \sum_t \sum_u E[Al_{t,u}^{rs}] = E[Q^{rs}] \, , \tag{2.31}$$

where $F_{t,u}^{rs} \in \mathbf{F}$, $Al_{t,u}^{rs} \in \mathbf{Al}$, and Al is the number of alighting passengers at each transit route. The following fixed-point problem can then be derived as a dynamic transit assignment model under uncertainty. In the proposed model, both the transit route and departure time choices are considered simultaneously:

$$\mathbf{f}_{\mathbf{t,u}} - \mathbf{q} \cdot \mathbf{P}_{\mathbf{t,u}}(\mathbf{f}_{\mathbf{t,u}}) = \mathbf{0} \tag{2.32}$$

where **q** is the vector of expected passenger OD demand; **f** is the vector of the mean passenger flow; and **P** is the vector of passenger departure time and route choice probabilities:

$$f_{t,u} = q \cdot P_{t,u}(f_{t,u})\,. \tag{2.33}$$

Theorem. At least one solution of the fix-point problem exists.

Proof. **F** is a convex and compact set, and **P(f)** is continuous on **F**, then following the Fixed-Point Theorem (Gasinski and Papageorgiou, 2005), at least one solution exists for the above fixed-point problem.

2.3 DYNAMIC NETWORK LOADING AND ALGORITHM

The loading of passengers is obviously triggered when transit vehicles arrive at stops. Passengers are loaded according to the two boarding phases described in Section 2.2. Passengers loaded at previous stops affect the configuration of transit vehicles arriving thereafter. This is because the available vehicle capacity and the deviation from schedules at downstream stops are usually determined by service configurations at upstream stops. The evolution of transit service configurations are explicitly taken into account throughout the simulation process.

Lam et al. (2008) proposed an algorithm to solve the multi-class reliability-based stochastic user equilibrium (RSUE) problem for a road network. This algorithm has been adapted to solve the fixed-point model proposed and described in this chapter. Uncertainty effects are recorded in the time-incremental micro-simulation procedure to enable the modeling of boarding delays and schedule deviations. The framework of the solution algorithm is shown in Figure 2.2.

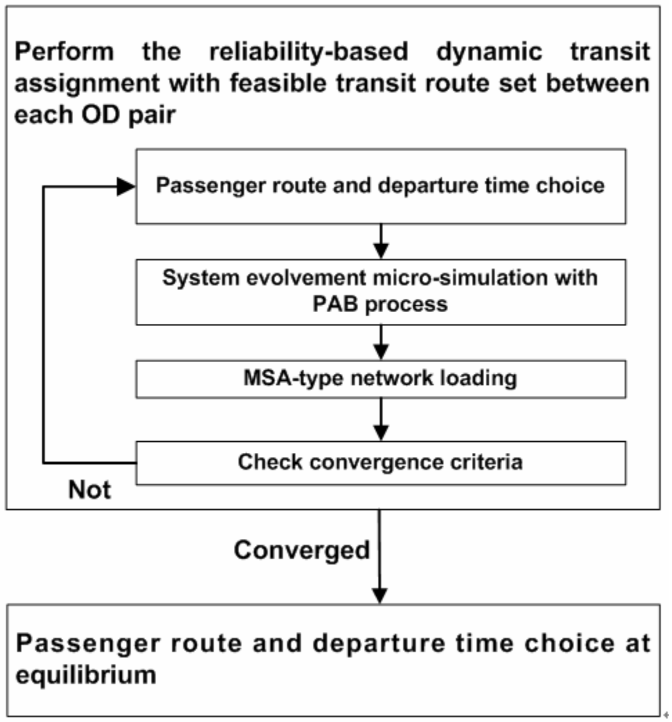


Figure 2.2 Flow Chart of the Solution Algorithm

2.4 NUMERICAL EXAMPLE

The small transit network shown in Figure 2.3 was used for the numerical test. This network is similar to the test network adopted by Lam et al. (1999) and De Cea and Fernandez (1993). However, the network has been slightly altered, in that transit services in Line 4 have been deleted, and Line 1 links every node between origin node N_1 and destination node N_4 instead of linking them directly. This alternative better represents service lines in a CBD area at a morning peak period.

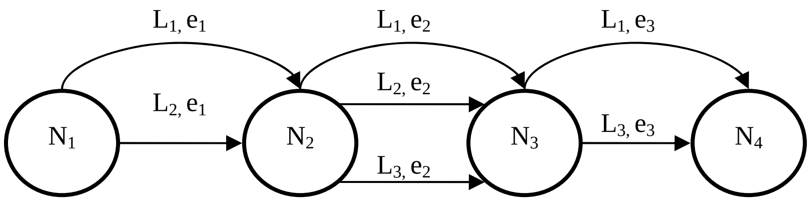


Figure 2.3 Example Transit Network

Table 2.1 Transit Routes List by Transit Lines and Links

Route	Order of transit links	Transfer node	OD pair
1	$L_{1,} e_1 – L_{1,} e_2 – L_{1,} e_3$	--	N_1--N_4
2	$L_{2,} e_1 – L_{2,} e_2 – L_{1,} e_3$	N_3	
3	$L_{2,} e_1 – L_{1,} e_2 – L_{1,} e_3$	N_2	
4	$L_{2,} e_1 – L_{3,} e_2 – L_{3,} e_3$	N_3	
5	$L_{2,} e_1 – L_{2,} e_2 – L_{3,} e_3$	N_2	
6	$L_{1,} e_2 – L_{1,} e_3$	--	N_2--N_4
7	$L_{2,} e_2 – L_{1,} e_3$	N_3	
8	L2, $e_2 – L_{3,} e_3$	N_3	
9	$L_{3,} e_2 – L_{3,} e_3$	--	
10	$L_{1,} e_3$	--	N_2--N_4
11	$L_{3,} e_3$	--	

The morning peak period between 8:00-9:00 was simulated in this test. Three OD pairs were included: OD 1 from node N_1 to N_4, OD 2 from node N_2 to N_4, and OD 3 from node N_3 to N_4. It was assumed that the maximum number of transfers is 1. The routes associated with OD pairs represented by lines, segments, and transfer nodes are shown in Table 2.1. Some input data include:

the confidence level of passengers: $\alpha = 0.95$;

the waiting time weighting coefficients: $\beta_1 = 2$;

the in-vehicle travel time weighting coefficients: $\beta_2 = 1$;

the transfer weighting coefficients: $\beta_3 = 2$;

the early or late penalty: $\beta_4 = 1$, $\beta' = 0.5$, $\beta'' = 2$;

the vehicle capacities for Line1, Line2 and Line3 are the same: $cap_{l1} = cap_{l2} = cap_{l3} = 120$;

the original headways for each line: $h_{l1} = 7$ min, $h_{l2} = 8$ min, $h_{l3} = 10$ min; and the mean of OD demand: $q^{14} = 400$, $q^{24} = 600$, $q^{34} = 300$.

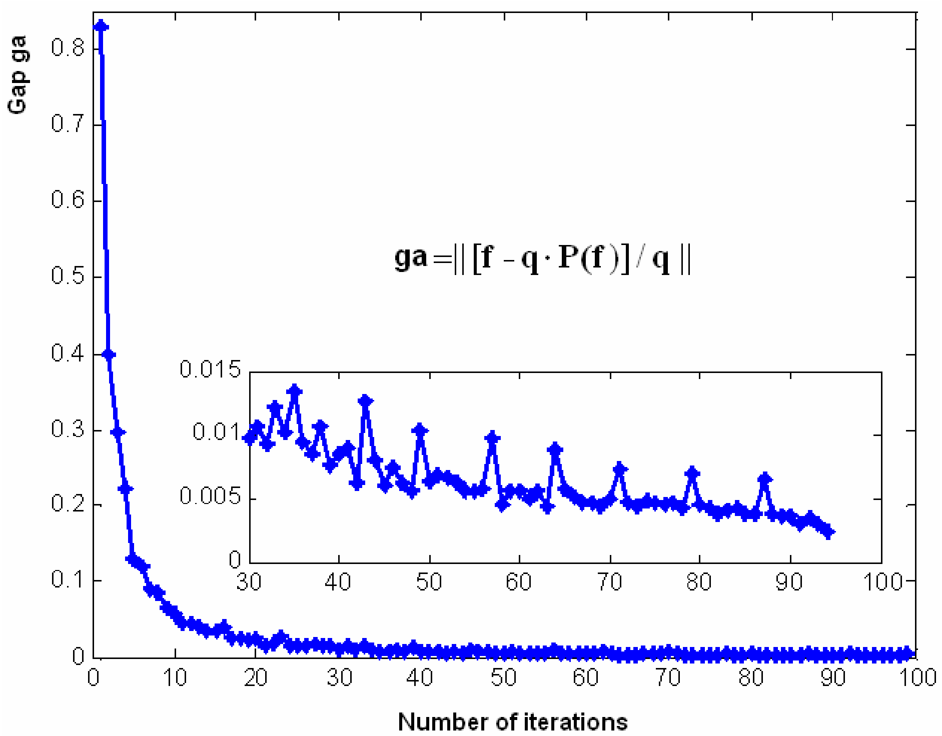


Figure 2.4 Convergence Results of the Solution Algorithm

Some "jumps" of occasional increases was observed in Figure 2.4 of the thesis. The frequency and range of "jumps" are affected by the number of Monte Carlo simulations, the expected value of passenger perception error, and the property of passenger generalized cost function in the proposed problem. These "jumps" shown in Figure 2.4 have two reasons. The first reason is that the descent direction in the method of successive algorithm (MSA) is generated by the inner Monte Carlo simulation. The number of the Monte Carlo simulation determines the efficiency and effectiveness of the inner Monte Carlo simulation. The trade off between the number of simulation and the algorithm efficiency causes the "jumps" in the convergence process. The second reason is that the moving step is fixed in MSA algorithm, which can overshoot the optimal moving step size and thus cause "jumps". Figure 2.5 shows the test results of algorithm convergence at different number of inner Monte Carlo simulations. It shows that as the number of Monte Carlo simulations increases, the proposed MSA-type algorithm converges to the optimal solution more quickly and smoothly.

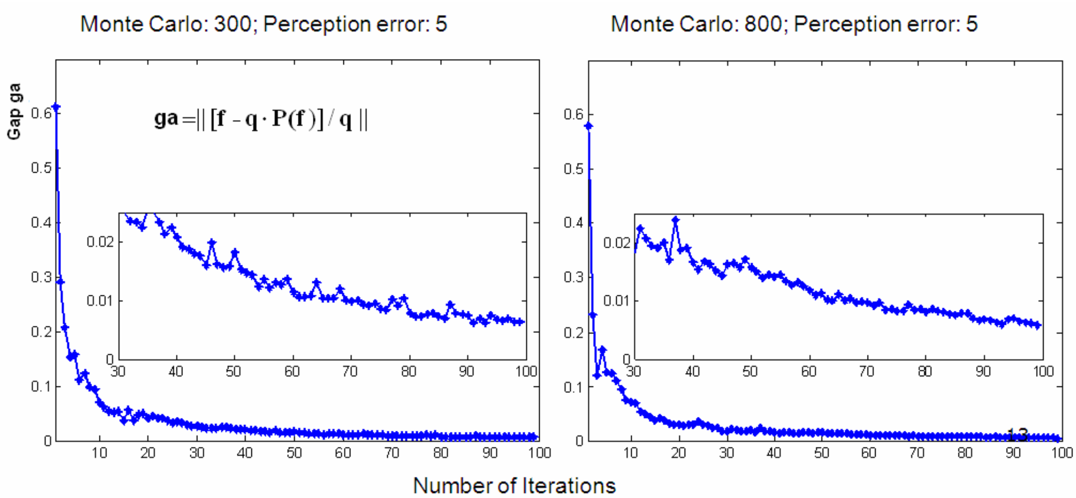


Figure 2.5 Convergence with Different Number of Monte Carlo Simulations

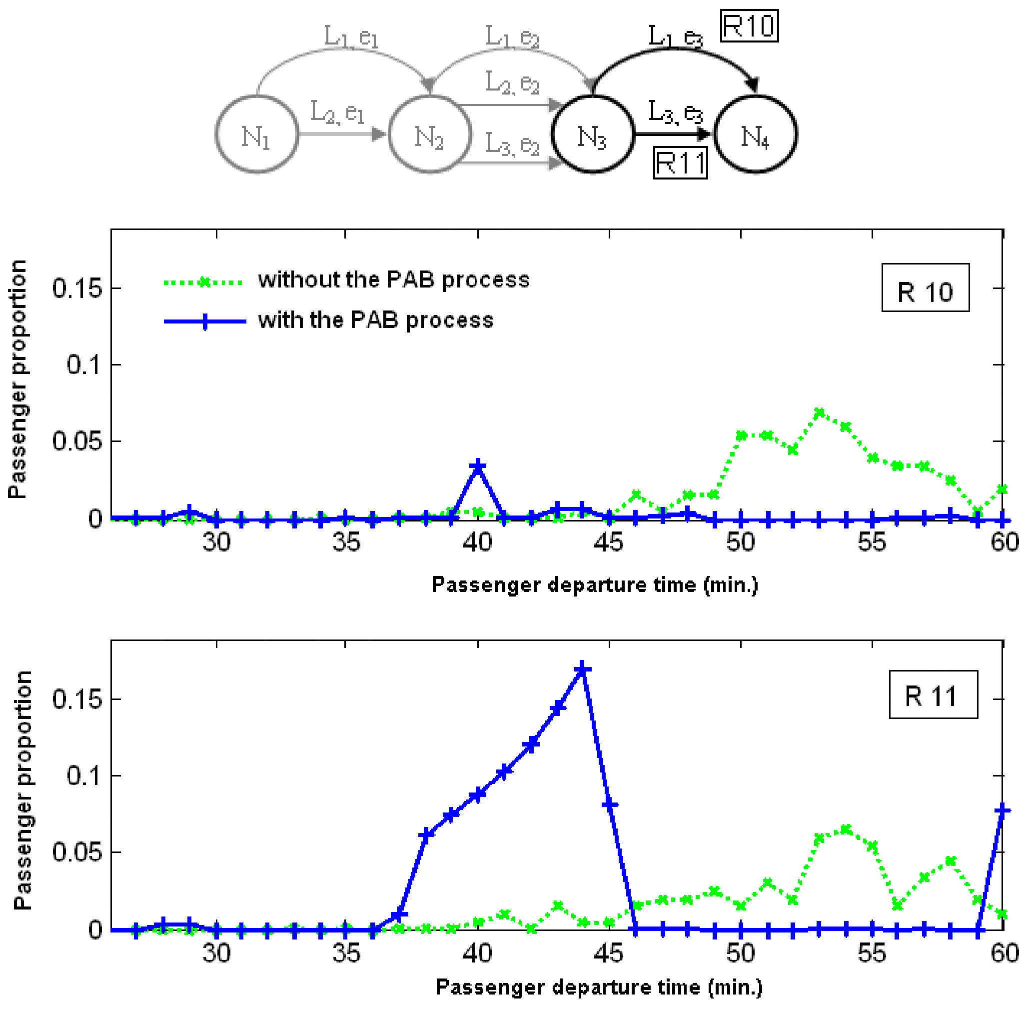


Figure 2.6 Passenger Departure Time Choices with and without the PAB Process

The temporal passenger flow on two routes for travel between OD pair N_3 -- N_4, with and without the simulation of the PAB process, is shown in Figure 2.6. With the PAB process simulation, some passengers switch from Route 10 to Route 11. The shift of passenger flows between the two routes is because of the aggregated randomness of Line 1 at downstream node N_3. The peak period also shifts from around 55min (in the upper figure) to around 45min (in lower figure) after taking into account the PAB process. This is because the boarding delay and the overload delay within the PAB process is taken into account in the travel decisions. The result indicates that neglecting transit network uncertainties is likely to overestimate passenger flows on the long lines and inaccurately predict a posterior peak passenger departure time. In fact, passengers tend to choose the more reliable short lines and depart early to ensure the on-time arrivals.

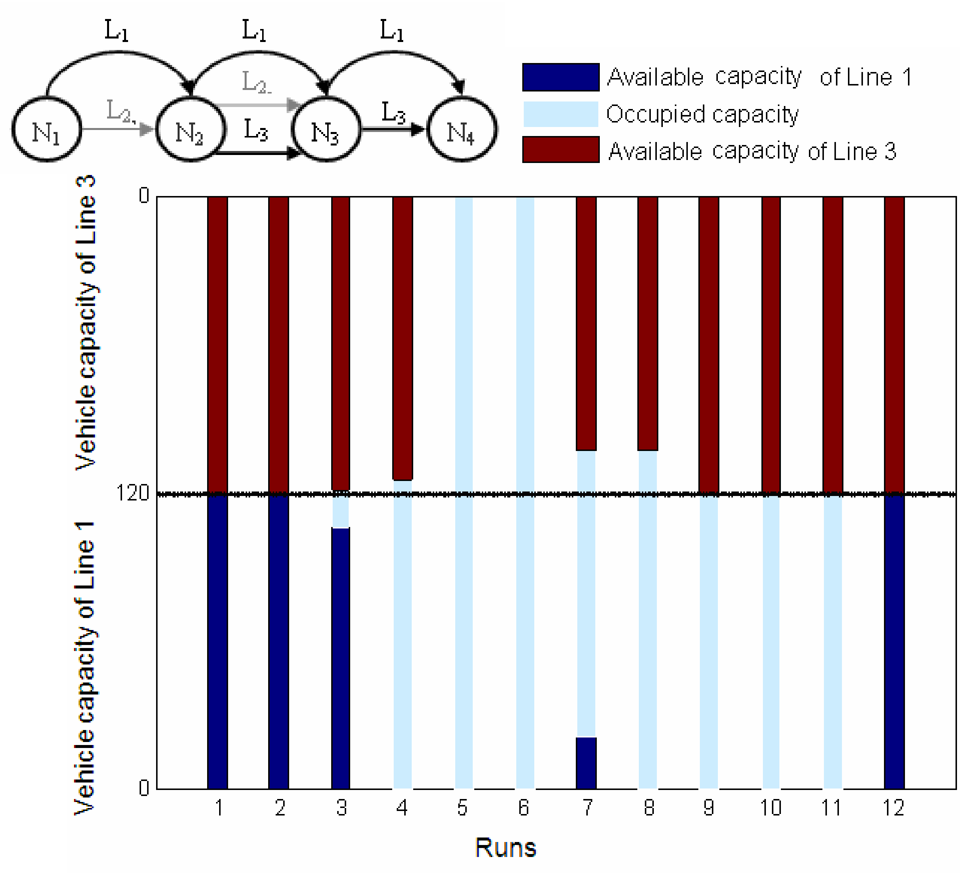


Figure 2.7 Vehicle Capacity Utilization of Line 1 and Line 3 at Terminal N_4

The other reason for the shift of passengers from Route 10 to Route 11 illustrated in

Figure 2.6 might be the capacity constraints. As shown in Figure 2.7, the 3rd to 11th vehicles of Line 1 are mostly occupied, mainly because of the travel demand of OD pair N_1--N_4 and OD pair N_2--N_4. The above situation means the passengers who intended to join Line 3 have to choose other transit lines because of the capacity shortage. Only the 5th and 6th vehicles of Line 3 are full loaded, which means Line 3 has sufficient capacity for passenger travel demand. Thus, passengers considering travel time reliability and overload delay shift their travel demand from the long line, Line 1 to the short line, Line 3.

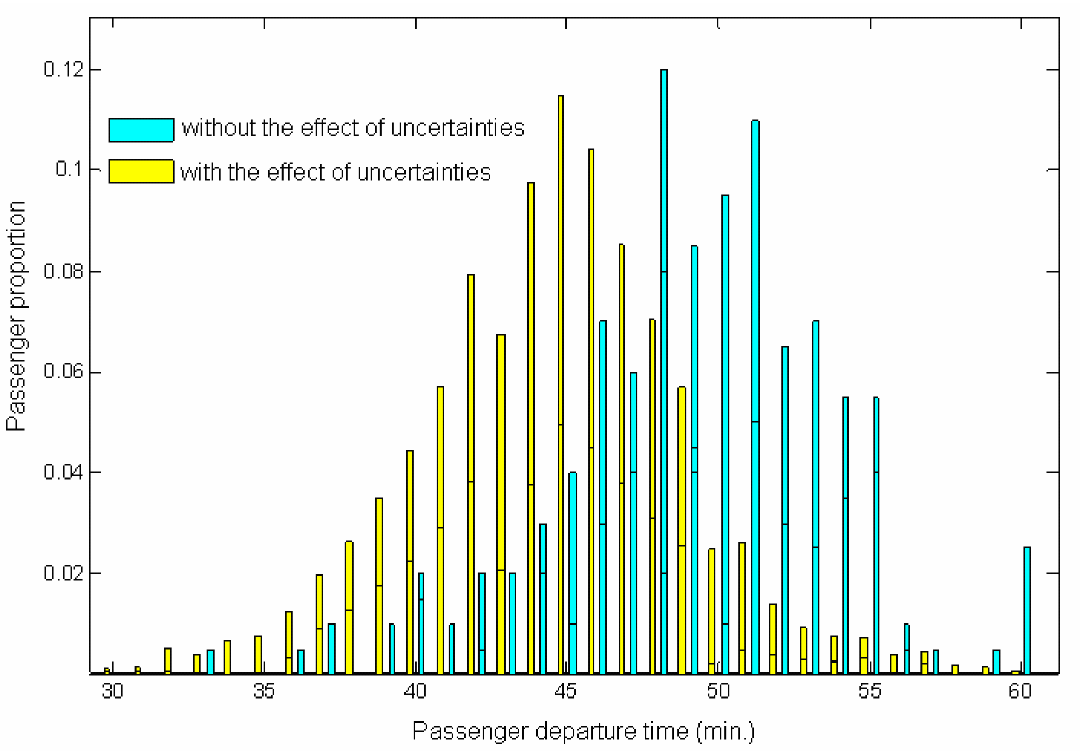


Figure 2.8 Passenger Departure time Choices with and without the Effect of Uncertainties

Figure 2.8 shows the proportion of passengers departing at the study time period and between OD pair $N_2 -- N_4$, with and without taking random network attributes into account. Passengers depart earlier when they take random network attributes into account than when they don’t. This is because the risk-averse passengers include additional safety margins in to their travel time budget to ensure the punctuality. The result indicates that the neglect of passenger responses to network uncertainties will cause the later estimation of peak period, which further affect the transit operational efficiency.

Table 2.2 Mean and Variance of Vehicle Dwell Time of Line 1

Stops \ Runs	1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}	6^{th}	7^{th}	8^{th}
Stop 1	0.25/0.22	0.35/0.26	0.55/0.33	1.0/0.45	3.10/0.79	***4.60/0.96***	0/0	0/0
	0/0	0.25/0.22	0.35/0.26	2.70/0.73	***4.20/0.92***	0/0	0/0	0/0
	--	--	--	--	--	--	--	--
Stop 2	0.95/0.44	1.00/0.45	1.55/0.56	***4.50/0.95***	2.90/0.76	1.40/0.53	0/0	0/0
	0/0	0/0	0.55/0.33	0.70/0.37	0/0	0/04	0/0	0/0
	0/0	0.45/0.30	0.85/0.41	**6.00/1.10**	**6.00/1.10**	0/0	0/0	0/0
Stop 3	1.70/0.58	1.65/0.57	3.90/0.88	0.45/0.30	0/0	0/0	0/0	**6.00/1.10**
	--	--	--	--	--	--	--	--
	0/0	0.35/0.26	2.75/0.74	0/0	0/0	**6.00/1.10**	**6.00/1.10**	**6.00/1.10**

Note: The Stops 1, 2, and 3 are the nodes N_1, N_2, and N_3 representatively

The bunching problem, caused by uncertainties in vehicle dwelling time and journey time, also decreases the operational efficiency of transit services. The mean and variance of vehicle dwelling times for each line and run at each stop is given in Table

2.2. The vehicle dwelling time is up to 6 min for several runs, due to the highly congested situation in the test case. Such vehicle dwelling delay can cause severe vehicle stop congestion and affect road traffic externally. In practice, the service capacity and travel demand model should be calibrated against actual data.

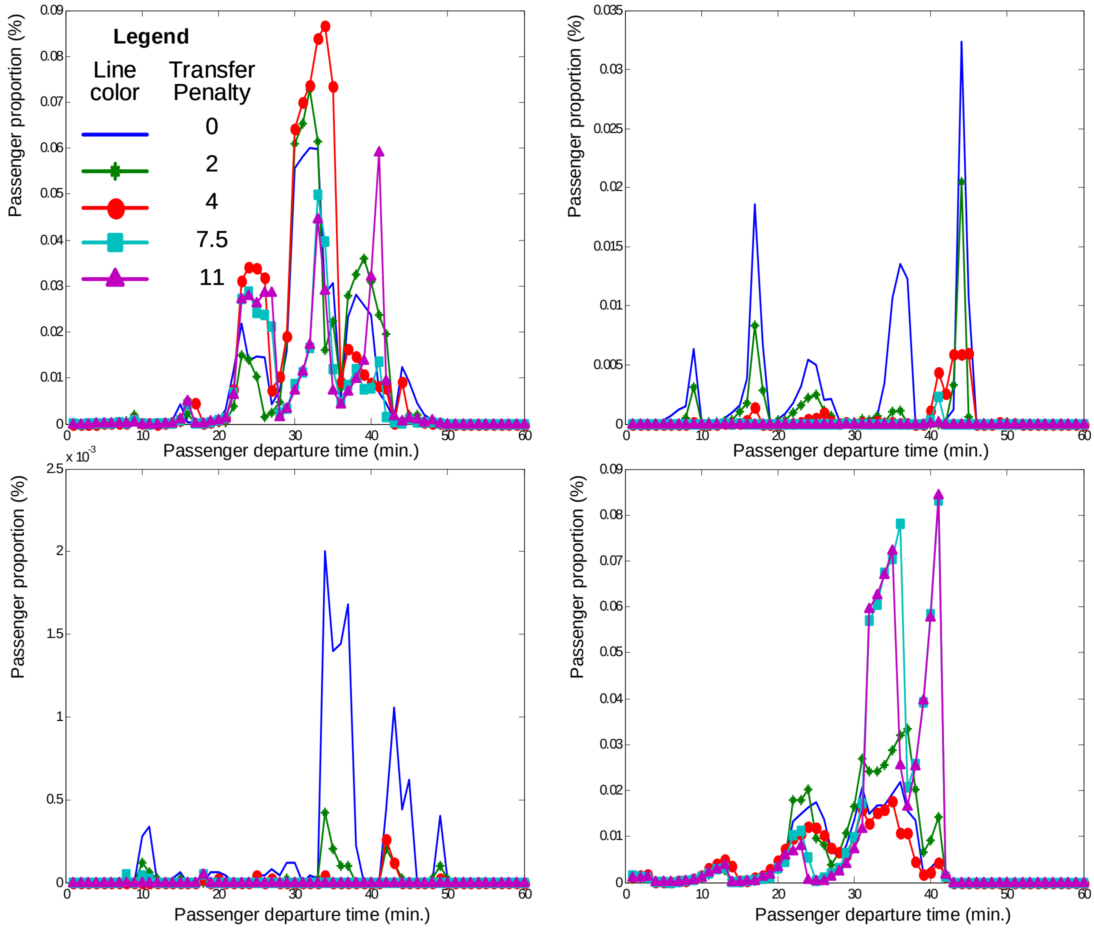


Figure 2.9 Sensitivity Analysis for Passenger Transfer Penalty

Passenger transfer penalty has been investigated in the previous related studies on the basis of different definition of transfer penalty and transfer context. Alger et al. (1975) examined the penalty for transfers between subway, rail, and bus in Stockholm, Sweden. Variables related to passenger behavior comfort and convenience such as waiting time, number of transfers, and seat availability are considered during the transfer. Han (1987) tested the influence of transfer on bus path choice in Taipei, Taiwan. A binary choice model is used to estimate the transfer penalty based on the data collected from bus passenger interviews. Liu et al. (1997) examined modal choice and transfer between auto and transit using survey data in New Jersey. Wardman et al. (2001) collected the data from stated preference survey among bus, rail, and auto in Edinburgh and Glasgow, United Kingdom. Transfer penalty between each two modes are specified. Because of the different definition and research context of transfer penalty, the survey results were found quite different, varying from less than 2 minutes to up to 50 minutes. They may even be ignored in the previous transit assignment and network design models. In this study, the sensitivity analysis is carried out to reveal the shift of passenger flows between long and short routes affected by different values of passenger transfer penalty for assessing their effects on route choices of transit passengers.

Figure 2.9 shows the sensitivity analysis results for passenger transfer penalty and the impact of transfer penalty on passenger departure time and route choices of OD 2. When transfer penalty is equal to 0, passengers prefer to choose transfer routes (particularly transferring to short bus route) rather than non-transfer short or long bus routes; As transfer penalty arises moderately (from 0 to 4), passengers who chosen non-transfer routes increase; When transfer penalty is very high (7.5 and 11 equivalent minutes), more passengers choose non-transfer routes rather than transfer routes. This implies that the route choices of passengers at downstream are affected by the route choices of passengers at upstream.

2.5 SUMMARY

The model proposed in this chapter assigned passenger demand to a transit network, and generated the transit service configurations in a congested network under uncertainties. The properties of the PAB process at transit stops were first analyzed and then used to estimate the mean and variance of passenger waiting time and vehicle dwelling time at stops. The analytical expressions for the mean and variance of these delays were combined into the passenger travel dis-utility function as the determinants for the transit travel choice. The dynamic stochastic user equilibrium model for representing passenger route and departure time choices was then formulated.

The perception error term was assumed to follow the Normal distribution, which provided the Probit-type network equilibrium model. Passengers in this model are assumed to consider both the mean travel time and a safety margin (derived from the mean and variance of travel time) in their travel decisions. The proposed model was tested on a simple network to highlight the effect resulting from passenger arriving, queuing, and boarding processes, being included in the transit assignment model.

The results showed great differences between the passenger flow profiles (combined route and departure time choices) obtained from the models when network uncertainty was allowed for. The results also showed a shift in the travel demand profile to earlier time periods to ensure on-time arrivals. The other key feature of the model was its ability to replicate the bus bunching phenomena which underpins transit service quality in practice.

Though the numerical example was unrealistic so far as a general transit network is concerned, it still represents the typical transit network states when congestion becomes extreme. From this perspective, the numerical example is sensible in showing that extreme congestion needs special attention as it leads to severe impacts on a transit network. A practical transit network, connecting Kowloon urban area to the Hong Kong International Airport (HKIA), will be given in Chapter 3 to show passenger travel route and departure time choices under transit network uncertainties and congestion.

Further research based on this chapter includes:

i) Extending to a multi-class model with a random distribution of the parameters governing passenger risk-taking behaviors.

ii) Incorporating the dynamic transit assignment model into the transit short-term network design model to study the interactions between short-term operational schemes and passenger behavior responses.

Extension to multi-user class problem will be presented in following Chapter 3, and further extensions will be described in Chapters 4 and 5, with respect to the regulated and deregulated market regimes in transit networks under uncertainties.

CHAPTER 3

THE MULTI-CLASS SCHEDULE-BASED TRANSIT ASSIGNMENT MODEL UNDER NETWORK UNCERTAINTIES

Demand and supply uncertainties at schedule-based transit network levels strongly impact passenger travel behaviors. In this chapter, a new multi-class reliability-based dynamic transit assignment model is proposed. Passenger travel behaviors vary because of the heterogeneous risk-taking attitudes towards random generalized travel cost in congested network with uncertainties. Passenger transit route and departure time choices are affected by respective passenger reliability requirements. Vehicle design capacity constraint on stochastic passenger demand is captured by an overload congestion parameter. The proposed model is formulated as a fixed-point problem, and solved by a heuristic algorithm. The numerical results show that passenger risk-taking attitudes will greatly impact passenger travel route and departure time choices, as well as monetary and time costs. This chapter is an edited version of: Zhang, Y.Q., Lam, W.H.K., Sumalee, A., Lo, H.K., and Tong, C.O. (2010) The multi-class schedule-based transit assignment model in network with uncertainties. *Public Transport*, Vol. 2, pp. 69-86.

3.1 INTRODUCTION

In the previous chapter, passenger travel route and departure time choices with the singular risk-averse attitude toward network uncertainties have been studied. However, passengers may have different risk perceptions when facing network uncertainties in their travels, as it is likely that they will value travel time reliability differently, depending both on their income levels and trip purposes (Noland and Polak, 2002; Lam et al., 2008). In this chapter of the research study, multi-class passengers' different attitudes toward stochastic generalized travel cost are considered as 1) risk-prone, 2) risk-neutral and 3) risk-averse, and embedded in the RSUE model by a step function with anticipated possibility of on-time arrival.

In literature, Modeling techniques for transit assignment problems are largely categorized as frequency-based (De Cea and Fernandez, 1993; Cominetti and Correa, 2001; Schmöcker et al., 2008) and schedule-based (Wilson and Nuzzolo, 2004; Poon et al., 2004; Hamdouch and Lawphongpanich, 2008) methods. These two modeling methods serve different planning purposes. The former aims at long-term planning such as land use and transport development projects, while the latter is better suited to short-term transit operations and service planning such as transit timetabling and vehicle scheduling.

Transit assignment models have recently emphasized the influence of uncertainties in frequency-based frameworks. The vertex failure in transit networks has been studied by Bell et al. (2002). A particular type of vertex failure (failure to board a full service) has been investigated by the absorbing Markov chain model. The notion of failure-to-board has been further applied in frequency-based transit assignment models with common line problems (Kurauchi et al., 2004). Yang and Lam (2006) proposed a probit-type reliability-based transit assignment model for congested networks with unreliable transit services. Stochastic passenger in-vehicle travel time, impacted by vehicle on-road running time uncertainties, was considered by Szeto et al. (2009). The stochastic passenger waiting time and the stochastic capacity, from the perspective of a line rather than a run, were studied. Similar to the research of Spiess and Florian (1989), stochastic passenger waiting time was found to be due to random passenger arrival and stochastic distributed line headways. Stochastic vehicle capacity stems from headway variation of the line due to congestion on the roads and delays at bus stops.

The above frequency-based models can be used to study aggregated stochastic effects of a specific transit line from the static perspectives. However, uncertainties exist in both vehicle running and dwelling process in line operation. The influences of uncertainties are also different for each run. The schedule-based model provides a means to investigate uncertainties within the vehicle operation process. Nuzzolo et al. (2001) have investigated the dynamic transit systems with regular and irregular services. The road congestion uncertainty resulting from irregular service is defined exogenously. Teklu et al. (2007) studied the day-to-day passengers learning processes regarding stochasticity in transit networks by a micro-simulation-based approach. The variance of passenger perceived cost is given by an equation of line frequency and passenger in-vehicle travel time, but without justification. These models represent uncertain passenger perceived travel cost and transit network operation status, do not, however, cover the evolution and interaction of uncertainties and impacts on passenger travel behavior.

In this chapter, a new multi-class reliability-based dynamic transit assignment model is developed. Passenger travel choice, such as mode, line, route, and departure time, varies in accordance with heterogeneous risk-taking attitudes. The stochastic passenger generalized travel cost consists of stochastic passenger in-vehicle travel time (composed of vehicle dwelling and running time), waiting time, transfer time, early or late penalty, passenger cost perception error, and out-of-pocket fares. Passenger stochastic waiting time can be the waiting time for the first arriving vehicle, or vehicles arriving thereafter, due to the passenger's inability to board the first vehicle. Congestion and vehicle capacity constraints are the main reasons for this situation (overload delay). Under demand uncertainties, the deterministic physical vehicle constraint is adjusted by imposing a performance parameter to constrain stochastic in-vehicle passengers.

This chapter is organized as follows. Section 3.2 introduces the formulation of the proposed multi-class reliability-based stochastic user equilibrium (RSUE) model on the schedule-based transit modeling framework. The capacity constraint problem for stochastic passenger demand is also discussed in this section. Section 3.3 introduces the heuristic solution algorithm. In Section 3.4, a numerical example based on the transit network from the Kowloon area to Hong Kong International Airport is carried out to illustrate the application of the model, solution algorithm, and some important insights.

3.2 MODEL FORMULATION

3.2.1 Capacity Constraint Problem for Stochastic Passenger Demand

Wirasinghe (2003) revealed that bus load status can vary from being underutilized to being overloaded with respect to the different dispatching time and elapsed travel time. In this stochastic network, as passenger boarding demand and the resulting overload situation are stochastic, a pre-assumed overload parameter is used to represent the possibility of vehicle full-loaded situation. The overload parameter produces a new practical capacity constraint for the stochastic boarding demand instead of the design capacity constraint. The practical vehicle capacity constraint with overload parameter ω is able to represent the overall transit vehicle overload situation with ω possibility. The equation then implies that the probability of passengers loads (Q_l^j) in jth run of line l exceeds the vehicle actual capacity

cap_l by ω:

$$P\{Q_l^j \geq cap_l\} = \omega. \quad (3.1)$$

The difference between on-board passengers and vehicle capacity is $\Delta CAP_l^j = Q_l^j - cap_l$. cap_l is a constant and Q_l^j follows Normal distribution, so the mean and standard deviation thereby can be written as: $E(\Delta CAP_l^j) = E(Q_l^j) - cap_l$ and $Std(\Delta CAP_l^j) = Std(Q_l^j)$. Standardizing this variable with the given confidence interval, the probability of overloading is:

$$P\{\Delta CAP_l^j \geq 0\} = 1 - \varphi(\frac{0 - E(Q_l^j) + cap_l}{Std(Q_l^j)}) = \omega \quad (3.2)$$

Be reminded that the mean and variance both equal the square of standard deviation of the number of passengers loaded under the assumption of Poisson distribution. Hence the standard deviation is substituted by the square root of the mean:

$$E(Q_l^j) + \sqrt{E(Q_l^j)} \cdot \varphi^{-1}(1 - \omega) - cap_l = 0. \quad (3.3)$$

The unique value of $E(Q_l^j)$ can be found by solving the square root equation, revealing the practical capacity constraint in the stochastic network loading.

3.2.2 Modeling Transit Demand and Service

The schedule-based transit network is illustrated in the diachronic graph (Nuzzolo et al., 2003). Passenger movement and transit line running attributes (such as the schedule coordination and the vehicle encountering or overtaking phenomena) can be represented in the graph by a time and space illustration of vehicle trajectories. Given a transit network $\Omega(I, J, L)$, the stochastic vehicle arrival time and departure time of vehicle $V_{i,j,l}$, the i^{th} transit vehicle of line l at stop j, are described as $T^{a}_{i,j,l}$ and $T^{d}_{i,j,l}$. The i^{th} vehicle and $(i+1)^{\text{th}}$ vehicle of the same line may meet on the road or stop due to the stochastic vehicle dwell time and on-road running time.

As described in Section 2.3, passenger arriving process at each time interval τ_i for each line l is assumed to follow the inhomogeneous Poisson process $\{Q_l(\tau), \tau \geq 0\}$. On the arrival of the transit vehicle, if that vehicle has spare capacity, the passenger waiting in the queue will start boarding. However, it is difficult to ascertain the number of passengers getting aboard and the total boarding time for the process, when the time interval between successive passengers and time for each passenger boarding are stochastic. Although the mean values of boarding time and arrival rate are more easily obtained, they may not be sufficiently adequate for representing transit network attributes and passengers' reaction under various uncertain issues.

To model the boarding process, the M/G/1 queue in queuing theory and renewal theory is applied to get the analytical expression of the stochastic passenger boarding time, thereby the stochastic dwelling time and run headway of transit vehicles. The Markovian property of the passenger arriving and boarding (PAB) process and the derivation of the total passenger boarding time, vehicle dwell time, as well as the number of passenger boarding the bus have been described in Chapter 2, Equations (2.1-2.20).

3.2.3 Modeling Passengers' Risk-taking Behaviors and Travel Choices

Consider the following general class of passengers: 1) They have a desired arrival time, and know the travel time is not certain; 2) They choose the best departure time and transit route as long as the α (percent) confidence of on-time arrival is met. To represent such travel choices considering the reliability requirements, the Chance-constrained model is applied to convert the following stochastic programming problem into a deterministic presentation:

$$\text{Min } c \tag{3.4 a}$$

$$\text{s.t. } P\{C \leq c\} \geq \alpha, \tag{3.4 b}$$

where C is the stochastic travel time.

Classify passengers into m classes. Such passengers are taken as having different confidence levels and are able to introduce different safety margins by a step function:

$$a = sf(\alpha_m),\ (\ldots, \alpha_m, \ldots) \in \alpha,\ \alpha \in [0,1]. \tag{3.5}$$

ett is the effective travel time (ETT) which consists of the mean passenger travel time and safety margin:

$$ett_m(\cdot) = \mathrm{E}(C(t,u)) + \varphi^{-1}(\alpha_m) \cdot \mathrm{Std}(C(t,u)), \tag{3.6}$$

where α_m represents the confidence level that m^{th} class passengers hold for their on-time arrival requirement. $C(t,u)$ represents passengers' stochastic travel time on route u and departure time t. Waiting time consists of (i) passenger waiting time at stops, (ii) the in-vehicle travel time, including passenger in-vehicle waiting time after boarding, (iii) passenger transfer time (if transfer is needed). Each element is multiplied by a weighting coefficient to convert each component to the equivalent unit of time, as shown in Equation (2.29).

Passenger generalized travel cost of class m is the summation of ETT, the early or late arrival penalty at destination, and the fares on the transit route. Mathematically, passenger generalized travel cost is defined by passenger classes:

$$GC_m = ett_m + \beta_4 tp(t) + \beta_5 cf + \varepsilon \,, \qquad (3.7)$$

The parameters β_4 and β_5 are the weighting coefficients of early or late penalty and transit fares.

Passenger waiting time is derived from the difference between passenger arrival time t and vehicle arrival time. Denote $\mathbf{T^a_{i,j,l}}$ as the vector of vehicle arrival time at stop j for all vehicles of line l and $\tilde{T}w$ as delayed waiting time if there is overload delay from the previous vehicle, and $\tilde{T}w$ equals the headways between two sequential vehicles of the same line. The passenger waiting time, with respect to the vehicle arrival time, has been defined in Equation (2.22), and also follows the Normal distribution.

The in-vehicle travel time and the in-vehicle waiting time can be derived from the vehicle running time models, which have been presented in Equations (2.25-2.26).

Passenger transfer time consists of the waiting time at the transfer stop multiplied by the transfer penalty coefficient β_3. The early or late penalty is deduced from passenger departure time, perceived travel time and desired destination arrival time, given in Equation (2.27).

ε represents passengers perception error when making travel decisions. It is a stochastic variable following the Normal distribution. Probabilities of passengers of class m choosing route *r* for travel at time *t* can be expressed as follows:

$$P_{t,u}^{m} = \Pr\{GC^{m}(t,u) \le GC^{m}(t',u'),\ \forall t \ne t',\ u \ne u'\}. \tag{3.8}$$

The stochastic equilibrium condition has been characterized by the following equation (Sheffi, 1985):

$$f_u = q \cdot P_u\ , \tag{3.9}$$

where q is the average passenger demand for a single OD pair, f_u and P_u are the respective passenger flow and passenger flow probabilities of a route connecting the OD. The single-class dynamic RSUE condition developed in Chapter 2 is extended to the multi-class situation in this chapter, written as a fixed-point problem:

$$\mathbf{f} - \mathbf{q} \cdot \mathbf{P}(\mathbf{f}) = \mathbf{0}\,, \tag{3.10}$$

where $\mathbf{f}$ is the vector of $f_{t,u}^{m}$, representing passenger class m choosing route u with the departure time t; $\mathrm{P}(\mathbf{f})$ is the vector of $P_{t,u}^{m}(f)$, representing the probability of passenger class m choosing route u with the departure time t; and $\mathbf{q}$ is the vector of expected passenger OD demand. In addition, the regular network flow conservation holds:

$$q^m = \sum_{t,u} q \cdot P_{t,u}^m(f), \tag{3.11}$$

$$q = \sum_m \sum_{t,u} q \cdot P_{t,u}^m(f). \tag{3.12}$$

3.3 SOLUTION ALGORITHM

Recently, Lam et al. (2008) proposed an algorithm for their reliability-based stochastic user equilibrium (RSUE) model on the road network, the framework of which can be adapted for solving the fixed-point problem given above in this chapter. Under congested conditions in the transit network, when passenger boarding process is delayed at stops, the uncertainty effects should be considered in the estimation of passenger loading distribution.

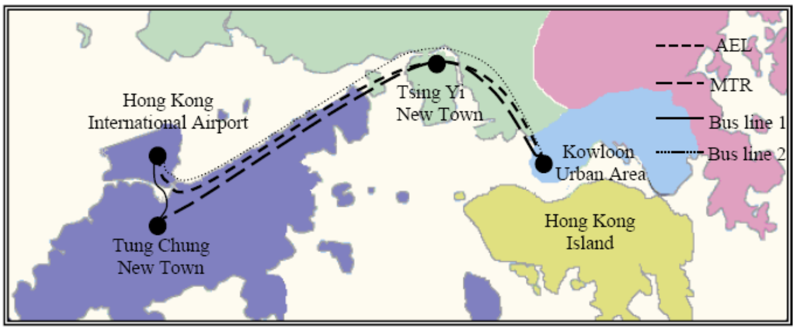
The loading of passengers, is triggered when the transit vehicle arrives at a stop, forming the first time interval. Passengers are loaded according to the two boarding phases previously described in Section 2.2.1. The practical vehicle capacity, set according to the probability of vehicle full-loading, constrains the random passenger boarding demand. Passenger loading process at upstream stops, affect transit service configuration thereafter, i.e. the occupied passenger loads and deviated schedules, affect the generalized travel cost at the downstream stops. The service-load dependency is taken into account through the explicit loading process.

3.4 NUMERICAL EXAMPLE

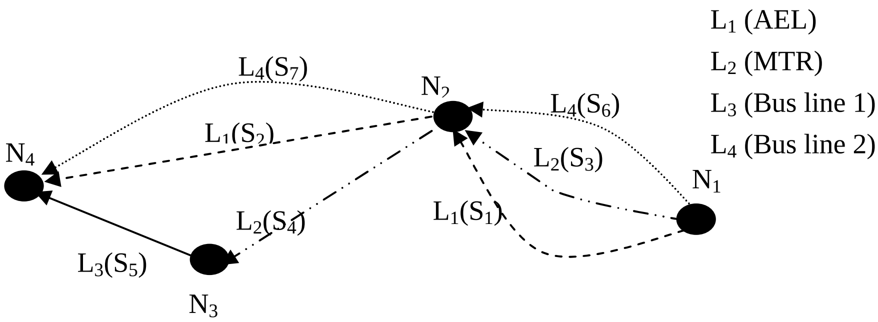
A simple transit network is used to present the impact of network uncertainties on different passenger travel behaviors. The transit network connects Kowloon urban area to the Hong Kong International Airport (HKIA) as shown in Figure 3.1(a). Four transit lines were considered, the Airport Express Line (AEL), Mass Transit Railway (MTR), Bus line 1 (Bus-1) and Bus line 2 (Bus-2). There are two OD demand pairs connecting the two origins (Kowloon and Tsing Yi) to the destination HKIA.

Passengers go to the airport for multiple purposes, primarily when making a plane trip but also to, pick up passengers or visit the nearby museum. Their awareness of trip time and the dependence on trip time reliability are distinguishable. The numerical example is designed to: (1) analyze the effects of demand variation on departure time and route choice in the multi-class network; (2) show how the transit service reliability, by different modes, affects the passenger departure time, route choices and waiting time (3) compare, in terms of average actual travel time and effective travel time, the assignment result under several modeling scenarios.

The Hong Kong air flight departure peak period is from 11:00 am to 1:00 pm, and the transit network rush hour to HKIA is around 2 hours prior to flight departures. The transit network study period of the example transit network is thus chosen to be the morning rush period, from 8:00 am to 12:00 noon. Passenger check-in at the airport is usually, 1 hour before flight departure time. Hence, for this study, passenger desired arrival time is set at 12:00 am.



(a) The simplified transit network between Kowloon and HKIA



(b) The alternative representation of the transit network by transit links

Figure 3.1 Transit Network of the Numerical Example

Total passenger demands during the above rush hour have one destination but two origins. They are: 1) from Kowloon (node N_1) to HKIA (node N_4) q^{14}=20000 (pass), and 2) from Tsing Yi (node N_2) to HKIA (node N_4) q^{24}=10000 (pass). Figure 3.1 (b) shows the alternative representation of the example transit network in terms of transit lines and links.

Table 3.1 Basic Transit Line Data for the Example Transit Network

<table>
<tr><th>Transit line</th><th colspan="2">AEL (L_1)</th><th colspan="2">MTR (L_2)</th><th>Bus-1 (L_3)</th><th colspan="2">Bus-2 (L_4)</th></tr>
<tr><td>K_l (pass/veh)</td><td colspan="2">500</td><td colspan="2">1500</td><td>120</td><td colspan="2">120</td></tr>
<tr><td>Transit link</td><td>S_1</td><td>S_2</td><td>S_3</td><td>S_4</td><td>S_5</td><td>S_6</td><td>S_7</td></tr>
<tr><td>In-vehicle time (min)
Mean /
Standard Deviation</td><td>8/
1</td><td>12/
1</td><td>9/
1.732</td><td>11/
1.732</td><td>24/
2.828</td><td>28/
3.162</td><td>35/
4</td></tr>
<tr><td>Dwell time (sec)
Mean /
Standard Deviation</td><td colspan="4">23.377/2.241</td><td>-/-
(shuttle service)</td><td colspan="2">$2.410 \cdot Boardings$ /
$1.828 \cdot Boardings$</td></tr>
<tr><td rowspan="2">Transit fare (HK$)</td><td>-</td><td>60</td><td>9</td><td>14</td><td rowspan="2">3.5</td><td>26</td><td>17</td></tr>
<tr><td colspan="2">90</td><td colspan="2">17</td><td colspan="2">33</td></tr>
</table>

Table 3.2 Transit Routes List by Transit Links

Route	Order of transit links	OD pair
1	S_1– S_2	
2	S_3– S_2	
3	S_3– S_7	
4	S_3– S_4– S_5	N_1---N_4
5	S_6– S_2	
6	S_6– S_7	
7	S_6–S_4– S_5	
8	S_2	
9	S_7	N_2---N_4
10	S_4– S_5	

Table 3.1 gives the basic transit line data for the transit network example. All available transit routes and line attributes are listed in Table 3.2. AEL is operated strictly according to the given timetable. Owing to their exclusive right-of-way operation, AEL and MTR are more reliable than bus lines as the running time

variance is small. The data provided in Table 3.1 are either real data from transit agencies or that gained from practical experience and information systems such as EasyGo (a research and development product from the Land Surveying and Geo-Informatics Department of Hong Kong Polytechnic University).

The OD demand multiplier is denoted as θ to represent various passenger demand levels. Other input data include:

the confidence level of risk-neutral, moderate risk-averse and high risk-averse passengers: $\alpha_1 = 0.5, \alpha_2 = 0.7, \alpha_3 = 0.95$;

the waiting time parameter: $\beta_1 = 2$;

the in-vehicle travel time parameter: $\beta_2 = 1$;

the transfer parameter: $\beta_3 = 2$;

the early or late penalty parameter: $\beta_4 = 1$, $\beta' = 0.2$, $\beta'' = 2$;

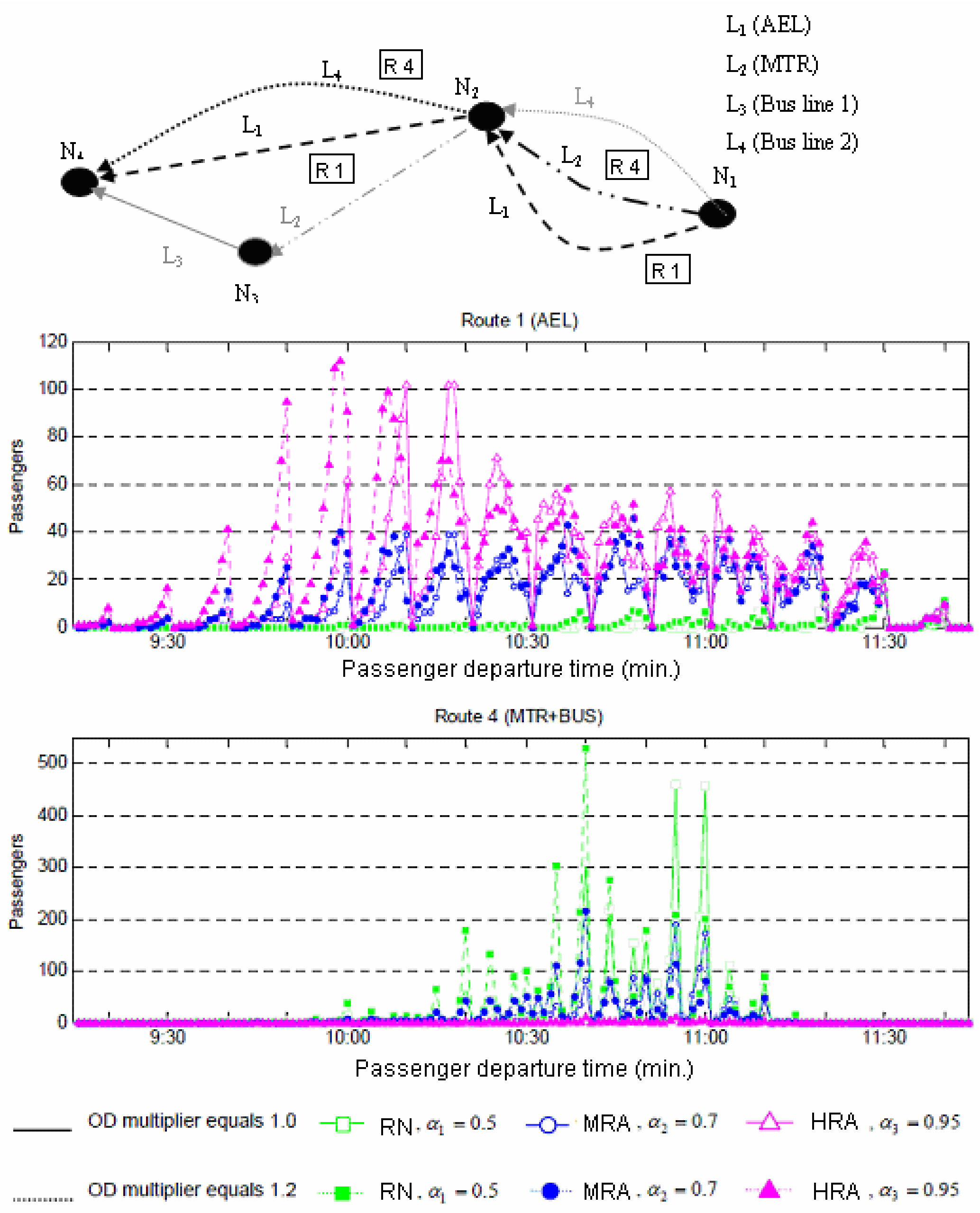
the fare parameter: $\beta_5 = 0.5$; and

the dispatching headways of AEL, MTR, Bus Line 1, and Bus Line 2 respectively: $h_{AEL} = 10\min$, $h_{MTR} = 5\min$, $h_{bus-1} = 4\min$, $h_{bus-2} = 12\min$.

The passenger departure time choices are illustrated in Figure 3.2. Different classes of passengers are included. It is shown that high risk-averse passengers did not choose Route 4 (taking the MTR first and then transferring to the bus). Their choices are likely to have been influenced by the substantial unreliable waiting time and transfer delays in Route 4. Passenger demand for Route 1 varied significantly in accordance with different passenger classes. Most risk-averse passengers but only a few risk-neutral passengers chose the more expensive yet more reliable route. The indication is that the risk-averse passengers choose the more expensive routes for the sake of reliability. They try to avoid risks by including safety margin in travel decisions. The extra monetary cost to ensure travel time reliability during peak period could be considered as an unfair penalty imposed on risk-averse passengers.

It is also observed from Figure 3.2 that risk-neutral and moderate risk-averse passengers increased due to the congestion caused by increased demand. The departure time range did not change for the risk-neutral passengers, but was expanded for the moderate risk-averse passengers. The expansion of the departure time range indicates that risk-averse passengers try to avoid the demand driven uncertainties: vehicle dwelling times at transit stops.

Table 3.3 shows the proportion of passengers who have been forced to wait at the Kowloon station (i.e. N_1) owing to the insufficient capacity in the first arriving vehicle. The assignment results of single-class RSUE and multi-class RSUE models are compared. From the table, it is seen that most congested periods were the same for both models (from 10:00 to 11:00). However the multi-class RSUE model showed an alleviation of overload congestion as the percentage of passengers on the same journey decreased. This decrease is due to the travel decisions made by risk-averse passengers, who possibly chose to change to earlier departure times or more reliable transit modes.



Notes: RN=Risk Neutral MRA=Moderate Risk Averse HRA=High Risk Averse

Figure 3.2 Departure Time Choices of Multi-Class Passengers

Table 3.3 Estimated Proportions of Passengers Waiting at Node N_1

Time / Model	8:00 -- 8:30	8:30 -- 9:00	9:00 -- 9:30	9:30 -- 10:00	10:00 – 10:30	10:30 -- 11:00	11:00 -- 11:30	11:30 -- 12:00
Single-class RSUE	0	0	2%	9%	**33%**	**16%**	10%	2%
Multi-class RSUE	0	1%	8%	9%	**16%**	**24%**	11%	1%

Table 3.4 shows passenger generalized travel cost, effective travel time, expected travel time, and each cost component of the generalized travel cost, when the demand multiplier $\theta = 1.0$ and $\theta = 1.5$ respectively. It should be noted that the effective travel time and expected travel time were the same for risk-neutral passengers, which means that network uncertainties have no impact on passenger average in-vehicle travel time and average waiting time. This demonstrates that the SUE model presents a special case of the multi-class RSUE model when network uncertainties are not considered.

However the risk-neutral passengers had the highest generalized travel cost for both the above scenarios ($\theta = 1.0$ and $\theta = 1.5$). The highest component of their generalized travel cost were early or late penalties, suggesting that these passengers did not recognize possible network uncertainties and unreliable travel time and therefore failed to add a safety margin to their expected travel time. As a result, such passengers departed during the most congested time period (35% waiting time when $\theta = 1.5$) or chose the time-consuming route (29% in-vehicle travel time when $\theta = 1$) for travel.

Of the three passenger classes, the high risk-averse passengers were subjected to the lowest generalized travel cost but the highest monetary cost, compared with that of other passengers. The high risk-averse passengers appeared willing to pay extra money to ensure travel time reliability. Thus, their cost of early or late penalty is the lowest among three passengers classes (takes 27% of generalized travel cost), hence maintaining the reliability of on-time arrival. The highest cost component for moderate risk-averse passengers lay in early or late arrival penalties and waiting time, indicating the possibility of overload delays as the result of later departures.

Table 3.4 Generalized Travel Cost by Passenger Classes

Parameters \ Passenger Classes ($\boldsymbol{\alpha}$)		OD demand multiplier $\theta = 1$			OD demand multiplier $\theta = 1.5$		
		RN 0.5	MRA 0.7	HRA 0.95	RN 0.5	MRA 0.7	HRA 0.95
Generalized travel cost (min)		119.36	118.50	107.79	338.95	256.55	162.49
Effective travel time (min)		61.51	60.54	47.042	122.89	115.36	73.53
Expected travel time (min)		61.51	54.85	43.06	122.89	109.88	69.21
Monetary cost (%)		9%	20%	**35%**	3%	11%	**27%**
Early or late penalty (%)		**41%**	**37%**	27%	**57%**	**43%**	**27%**
In-vehicle time (%)		**29%**	15%	10%	13%	11%	12%
Waiting time (%)		21%	**28%**	**28%**	**27%**	**35%**	**34%**
Standard deviation (min)	In-vehicle	4.09	2.58	1.52	4.08	2.51	1.55
	Waiting	1.56	0.73	0.69	1.58	0.82	0.77

Notes: RN=Risk Neutral MRA=Moderate Risk Averse HRA=High Risk Averse

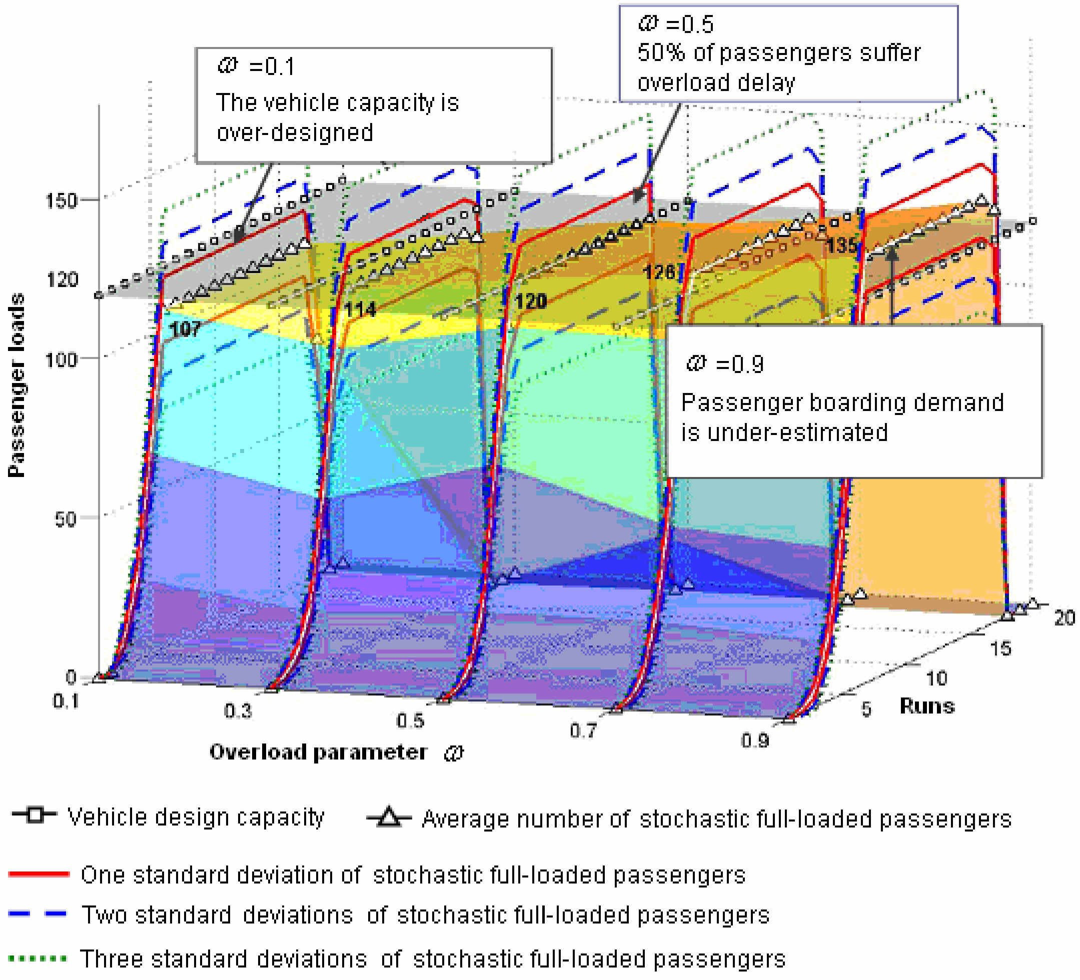


Figure 3.3 Stochastic Passenger Loads under Vehicle Design Capacity Constraint

Passenger loads in vehicles with deterministic capacity under stochastic passenger demand are illustrated in Figure 3.3. The standard deviations of stochastic full-loaded passengers represent different degree of variations on full-loaded passengers. The average number of stochastic full-loaded passengers under vehicle design capacity varies considerably in accordance with different overload parameter ω. For example, when $\omega = 0.5$, the average number of stochastic full-loaded passengers under vehicle design capacity constraint equaled the design capacity 120. It means that the probability of vehicle overload at peak period was 50%. When $\omega = 0.1$, the average number of stochastic full-loaded passengers under vehicle design capacity constraint was 107, which was less than the design capacity (i.e. equivalent to the average number of stochastic full-loaded passenger plus one standard deviation as shown in Figure 3.3). It indicates that the probability of vehicle overload at peak period was comparatively small. Thus, the ignorance of stochastic passenger boarding demand resulted in low profitability of transit runs. When $\omega = 0.9$, the average number of stochastic full-loaded passengers under vehicle design capacity constraint was 135, exceeded the design capacity. It implies that nearly 67% of waiting passengers could not be able to get on the first-coming vehicle. Severe underestimation of passenger boarding demand could lead to a degraded level of transit service. Thus, the overload parameter ω in the congested transit network under uncertainties should be carefully calibrated to meet the acceptable level of transit service.

3.5 SUMMARY

A new reliability-based dynamic transit assignment model has been presented in this chapter to investigate multi-class passengers travel decisions including route and departure time choices in congested and stochastic transit networks with uncertainties. The proposed model has been shown to be capable of accounting for the impact of uncertainties with respect to passenger travel time, waiting time, vehicle dwelling time, as well as transit service reliability. It is noted that the SUE model presents a special case (passengers are risk-neutral) to those presented by the multi-class RSUE model. The results of the multi-class RSUE transit assignment model show that when traveling on the same transit network, risk-neutral passengers suffer the highest generalized travel cost, in comparison with that suffered by the risk-averse passengers. Risk-averse passengers are likely to choose the less risky options, regardless of the higher monetary cost, and appear willing to spend more money in the hope of benefits of reliability.

This chapter also demonstrated the importance of demand and demand-driven uncertainties in transport networks. The effects of such uncertainties influence passenger departure time choices and determine passengers' travel reliability of reaching destination on time. The ignorance of such uncertainties, when transit planners making the short-term passenger flow prediction or transit operational design, could lead to a rapid degradation of transit services in the ever changing situations of any city.

The above study has provided a new fundamental tool for transit service evaluation and transit network design under demand and supply uncertainties in the schedule-based network framework. Further research investigates transit planners operational network design with the input of passenger flow produced in the RSUE models developed in Chapters 2 and 3. The typical operational transit network design, transit schedule design, with the advantages of operational flexibility and economical feasibility are proposed to meet the planners' interests. The regulated and deregulated transit systems, as the progress of liberalization of transit market, are investigated respectively in Chapters 4 and 5.

CHAPTER 4

TRANSIT LINE SCHEDULE DESIGN IN DYNAMIC TRANSIT NETWORK WITH DEMAND AND SUPPLY UNCERTAITIES

A novel transit schedule design model for dynamic transit networks with uncertainties is proposed in this chapter. An array of uneven headways for each transit line is designed to optimize the integrated transit service efficiency and reliability. As passenger route and departure time choices are impacted by transit line schedules, the line schedule design problem is formulated as a bi-level problem to enable demand and supply interaction. The objective of the upper-level model is the integrated optimization of the transit network. The reliability-based dynamic transit assignment model is formulated in the lower-level model to present passenger behavior responses to transit network uncertainties, arising from both demand and supply sides. The above bi-level problem is solved by the Genetic Algorithm (GA) approach. A numerical example is used to illustrate the performance of the proposed model and solution algorithm. The numerical results demonstrate that the optimal line schedule can save the total network travel time, improve travel time reliability, and balance fleet size between lines. This chapter is an edited version of: Zhang, Y.Q., Lam, W.H.K., and Sumalee, A. (2010) Transit schedule design in dynamic transit network with demand and supply uncertainties. *Journal of the Eastern Asia Society for Transportation Studies*, Vol.8, pp. 1425-1435.

4.1 INTRODUCTION

In Chapters 2 and 3, the schedule-based transit assignment models considering passenger risk-taking attitudes have been developed. The models can predict passenger temporal and spatial flow in an unreliable transit network during a study period. Such prediction produces a solid basis for the short-tem transit network design and operation as passenger flow influences both the source of profit and the major cause of network congestion and uncertainties. In this chapter, passenger flow by transit route and passenger departure time is utilized to evaluate the potential of transit line schedule schemes proposed by a transit authority. To this end the transit line scheduling model can produce optimal transit line schedules accounting for passenger behavior responses under uncertainties.

The growth of city population density and the subsequent high mobility requirements call for efficient and reliable public transit services. Setting transit routes and frequencies are the main components of transit network design. However the two planning tasks lie on different planning levels. The former aims at long-term planning, such as the change of transit lines at network levels, and is implemented over long term periods. The latter aims at short-term planning and usually includes the setting of transit frequencies or headways. The planning purpose is to ensure adequate adaptability for both the day-to-day and within-day passenger demand variation.

Time-varying passenger demand, adverse weather, and traffic incidents often cause uncertainties in transit networks. These uncertainties from both demand and supply sides severely affect transit services and cause vehicle double-heading, bunching and overtaking phenomena. The unexpected prolonged waiting time and overload delay for passengers severely impact transit service reliability. Efficiency is another significant concern of transit agencies. Empty seats in the vehicle are an obvious cause of low productivity, while the overloaded vehicles during peak make it impossible for all waiting passengers to board. Thus, when considering transit scheduling problems, the saving of network travel time should be considered, not only in average situations, but also in uncertain network environments with embedding reliability components.

Studies have found the setting of a reasonable timetable could save on the number of operating vehicles (Ceder, 2003; Gao et al., 2004; Uchida et al., 2007) and improve the level of transit service, such as maximizing social benefit (Furth and Wilson, 1981), maintaining headway regulation (Ding and Chien, 2001), improving line connection and timetable synchronization (Ceder et al., 2001; Fleurrent et al., 2005). To combine the passenger travel decisions, in response to different service configurations, into the transit network design process, more researchers have proposed the multi-level programming (Fernandez et al. 2008; Zhou and Lam, 2001; Gao et al., 2004) or applied the iterative approach (Lee and Vuchic, 2005; Yan and Tang, 2008).

The bi-level model developed by Gao et al. (2004) and Uchida et al. (2007) considered the interaction between service and demand, and carried out sensitivity analysis to determine the transit frequencies. Normally, the frequency for each line is a fixed number and the transit assignment model at the lower level is static. The implication is that passenger demand is constant over time, and the running of the transit vehicles matches that set out in the timetable. However, the bi-level transit scheduling model could be more practical, if demand and service uncertainties are considered on the basis of the schedule-based transit modeling framework. It would also better allow for transit scheduling flexibility and rapid response to demand variation.

The bi-level model presented in this chapter is a reliability-based dynamic transit schedule design model. The objectives of the model are to 1) minimize the integrated value of network travel time and uncertainties and 2) balance the number of vehicles dispatched on each line. The lower-level model is a schedule-based transit assignment model under network uncertainties, which generates the passenger time-dependent demand and stochastic passenger travel time. The uneven headway given in the upper-level can also be reflected in the lower-level model by the time-space network representation.

Setting uneven dispatching headways, however, is a multivariable problem, too complex for mathematical programming, because of the large number of variables and constraints. A mathematical programming or optimization method is feasible when dealing with frequency setting problems with even headways (Gao et al., 2004; Uchida et al., 2007). In view of the above, heuristic methods (Ceder, 2003; Yan and Chen, 2002) are used instead of analytical methods to find a feasible and reasonable solution for uneven headway setting problems.

The genetic algorithm (GA), in particular, provides a robust search and a near optimal solution in a reasonable time. Hence it is widely applied in transit route design and scheduling problems (Pattnaik et al., 1998; Kidawi et al., 2005; Shrivastava and O'Mahon, 2006). GA is an adaptive heuristic search algorithm which produces solutions by natural evolution and selection. It is also capable of solving bi-level problems with large variable dimensions. In this chapter of the study, GA is applied to find the best headways for each transit line.

This chapter is organized as follows. In section 4.2, the reliability-based stochastic user equilibrium (RSUE) model in the lower-level and the optimization model in the upper-level are described in the formulation of the bi-level scheduling problem. The GA with the intelligence of assigning transit demand and dealing with scheduling constraints is discussed in section 4.3. The numerical example is carried out to show the performance of the model in section 4.4. The summary of this chapter is given in Section 4.5.

4.2 THE BI-LEVEL MODEL: INTEGRATED OPTIMIZATION OF TRANSIT EFFICIENCY AND RELIABILITY

4.2.1 Network Presentation

The transit service network is normally represented by either a time-space trajectory graph (Powell and Sheffi, 1983; Ceder, 2007), time-space diachronic model (Nuzzolo et al., 2001), or time-extended model (Hamdouch and Lawphongpanich, 2008). The time-space trajectory model, particularly, can illustrate the departure and arrival of transit vehicles spatially and temporally. Hence the operation incidents, such as schedule deviation, vehicles encountering and overtaking, can also be presented. As shown in Figure 4.1, even headways may not be maintained throughout the line run. However, the uneven dispatching headways, which take into account passenger demand and vehicle run time variations over time periods, are shown to better maintain transit service regularity.

Given a transit network $\Omega(I,J,L)$, the i^{th} transit vehicle of line l at j^{th} stop, $V_{i,j,l}$, can uniquely define the related arrival time and departure times: $T^a_{i,j,l}$, $T^d_{i,j,l}$. The i^{th} vehicle and $(i+1)^{\text{th}}$ vehicle of the same line may meet at the j^{th} stop owing to the variations in vehicle dwelling time and on-road travel time. The stochastic vehicle on-road time, due to traffic accident and incidents on the road (Chen et al., 1999) or the adverse weather (Lam et al., 2008) is evident in the real world and has been studied extensively. Modeling the stochastic vehicle dwelling time, however, is more complicated and has yet to be given attention.

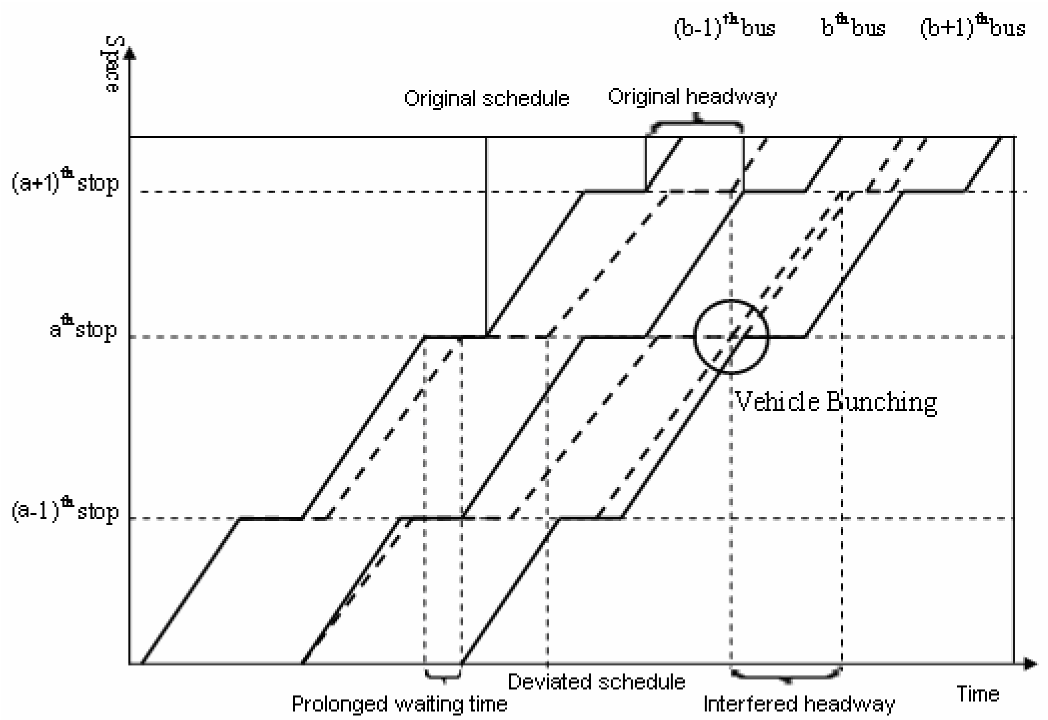


Figure 4.1 Transit Vehicle Time-Space Trajectories with Vehicle Bunching

4.2.2 The Upper-Level Model

In general, network design problems are concerned with two groups: network planners (government authorities) and network users (passengers). Passengers' travel behaviors follow the dynamic RSUE principle proposed in Chapter 2. The government authorities try to save the total passenger network travel cost in order to enhance the level of transit service. The demand and service interaction in a congested network with demand and supply uncertainties is considered by formulating a bi-level model as shown in Equation (4.1).

The upper-level model Equation (4.1 a) is an optimization model with explicit and implicit decision variables. $\mathbf{f}$ is the implicit decision variable deduced from the lower-level RSUE model. $\mathbf{h}$ is the vector of headways, constituting the connection of the upper and lower level models. $\mathbf{h}$ is assumed to be integers and the feasible lower and upper constraints for the headways of each line held. γ_1 is the convention factor for unit equivalence of network travel time and the variance. The network travel time efficiency and reliability is represented by the mean and weighted variance of total network travel cost, $\mathrm{E}[\mathbf{TC}\cdot\mathbf{P(f,h)}]$ and $\mathrm{var}[\mathbf{TC}\cdot\mathbf{P(f,h)}]$ respectively. The convention factor (γ_1) can be adjusted to meet the evaluation demand of different transit networks.

(Upper-Level) min $Z=\mathrm{E}[\mathbf{TC}\cdot\mathbf{P(f,h)}]+\gamma_1\ \mathrm{var}[\mathbf{TC}\cdot\mathbf{P(f,h)}]$ (4.1 a)

where $\mathbf{P}(\mathbf{f},\mathbf{h})$ solves:

(Lower-Level) $\mathbf{f}-\mathbf{q}\cdot\mathbf{P}(\mathbf{f},\mathbf{h})=\mathbf{0}$. (4.1 b)

4.2.3 The Lower-Level Model

As shown in Section 2.2.4, $\mathbf{F}$ is the set of passenger flow variables on route u between the OD pair departing at time t. The following fixed-point problem has been derived for the dynamic transit assignment model under uncertainty. Both the route and departure time choices are simultaneously considered:

$$\mathbf{f} - \mathbf{q} \cdot \mathbf{P}(\mathbf{f}) = \mathbf{0} \tag{4.2}$$

where **q** is the vector of expected passenger OD demand, **f** is the vector of the mean passenger flow; and **P** is the vector of passenger departure time and route choice probabilities:

$$f = qP(f)\,. \tag{4.3}$$

q is the vector of expected passenger OD demand, and **P(f)** is the vector of $P_{t,u}$, representing the probability of passengers choosing route u with the departure time t. The probabilities are obtained by finding optimal passenger travel routes and departure times, which also generate the minimum passenger generalized travel cost:

$$P_{t,u} = \Pr\{TC(t,u) \le TC(t',u'),\ \forall t \ne t',\ u \ne u'\}\,. \tag{4.4}$$

where $TC(t,u)$ is the stochastic passenger generalized travel cost:

$$TC(t,u) = ett(t,u) + \beta_4 tp(t,u) + \varepsilon\,. \tag{4.5}$$

ett is the summation of effective travel time, tp is the early or late arrival penalty, and ε is passenger perception error. Parameter β_4 is the weighting coefficient of early or late penalty. The definition and formulation of transit service, passenger demand, and their interaction, have been elaborated in Section 2.2, From Equation (2.1) to Equation (2.33).

4.3 GENETIC ALGORITHM

Although the transit network design problems are well modeled as the bi-level problems, the non-linearity and large-size nature limits the development of a solution algorithm. The adapted heuristic approaches (Gao et al. 2004; Uchida et al., 2005; Fan and Machemehl, 2006; Shrivastava and O'Mahon, 2006) or practical approaches (Ceder, 2003) are widely applied in finding solutions for transit network design problems.

Ceder (2003) proposed three graphical procedures to determine uneven headways to balance passenger count. The passenger travel decision and transit service attributes interaction, however, is not considered in these procedures. Gao et al. (2004) designed a heuristic solution algorithm based on sensitivity analysis to solve the bi-level model. The nonlinear and implicit function of passenger flow in the lower-level model is approximated by a linear formulation in order to refine the sensitivity of the upper-level model. Yan et al. (2006) developed the stochastic demand scheduling model. The scenario decomposition method was applied and stochastic events were decomposed into predetermined stochastic passenger demand scenarios.

In the non-convex, non-linear and stochastic optimization problem, it is robust to use genetic algorithm (GA) to search for the global optimal solution within a reasonable computational time. The advantages of GA over conventional optimization algorithms, in solving transportation network design problem, have been reported by Fan and Machemehl (2006) and Pattnaik et al. (1998). Table 4.1 summarizes the pros and cons of using GA for solving transit line schedule design problems.

Table 4.1 Pros and Cons of Genetic Algorithm

Algorithm	Literature	Pros	Cons
Genetic Algorithm (GA)	Shrivastava and O'Mahon (2006)	◆ Good global search capability ◆ Fast random search in solution space	◆ Limitation in local search ◆ Feedback information is not used in following search
Simulated Annealing (SA)	Friesz et al. (1992)	◆ Good at local search capability	◆ Susceptible to parameters ◆ Limitation in global search
Graphical procedures	Ceder (2003)	◆ Simple mathematical manipulation	◆ Ignorance of modeling information like attributes interaction and alternation
Heuristic algorithm based on sensitivity analysis	Gao et al. (2004)	◆ The uniqueness solution of global optimization ◆ Good time efficiency	◆ Restrict requirements on objective function convexity

In the application of GA, the decision variables are usually represented by chromosomes, constituted by genes. These chromosomes are generated randomly and evaluated to find their fitness values. The translation of chromosomes from numerical strings into technical operational forms is the key issue in GA. The most common coding method is to transform the variables to a binary string (Goldberg, 1989). The population generation is operated by three main operators: reproduction, crossover, and mutation. Reproduction selects best strings in the population, crossover exchanges the information among mating pool strings and mutation, making the local search near the current solution. At the end of each generation process, if the termination criterion (length of generation, computation time, use of memory) is met, the genetic process terminates. Otherwise, the population is iteratively generated and evaluated by the above three operators.

The formulated bi-level problem shown in Section 4.2 is solved by using an altered Genetic Algorithm (GA) approach. The lower level Probit-based dynamic assignment model is solved by the MSA-type of algorithm, for a new generation of feasible headways. The genetic algorithm for solving the bi-level problem is outlined as follows:

Step 0 At the initial generation of the GA, a number of populations are produced.

Step 1 Perform the reliability-based dynamic transit assignment with given feasible transit route set between each OD pair:

Step 1.1 Initialize the transit passenger flow on transit routes and departure time;

Step 1.2 Simulate the PAB process and obtain the passenger flows based on the current generalized travel cost and transit system attributes, using the Monte Carlo simulation;

Step 1.3 Update the transit passenger flows of the two classes using the method of successive averages (MSA);

Step 1.4 Check the convergence of the inner iteration, and calculate the mean and variance of the network travel time.

Step 2 Choose the best two designs according to the fitness function from the last generation (the parents), simply the objective function, are kept in set.

Step 3 Crossover and mutation points are randomly chosen for the evolution of the next generation.

Step 4 Check the stopping criteria of the outer iteration. Go to Step 2 if the stopping criterion is not met; Stop, otherwise.

4.4 NUMERICAL EXAMPLE

The small transit network used for the numerical test is the same as that shown in Figure 2.3. The morning peak period between 8:00-9:00 is considered in this example. The routes associated with lines, links, transfer nodes, and OD pairs are as that shown in Table 2.1.

The GA parameters were tuned for the proposed objective function. The best combination of crossover and mutation probabilities was selected, based on the value of the objective function (lowest). The population size was decided on the combined basis of the generation number, required for convergence, and the fitness value calculation time, for each individual in the population pool. The tuning of parameters was executed by several trials with different sets of parameter values. The following values were adopted:

Size of each chromosome strings: 18 (the total number of line headways of three lines);

Seed: $h_{l1} = (8,8,8,8,8,8,8)$, $h_{l2} = (8,8,8,8,8,8,8)$, $h_{l3} = (15,15,15,15)$ (min, the original headways of each line);

Population size: 30;

Elite count: 2;

Crossover probability: 0.8;

Mutation probability: 0.2; and

Number of generations: 50 generations or till convergence.

Other input data include: the parameters for early or late penalties $\beta' = 0.5$, $\beta'' = 2$; vehicle capacity for Line1, Line2 and Line3 is given by: $cap_{l1} = cap_{l2} = cap_{l3} = 120$ (passengers/vehicle); and a mean OD demand is $q^{14} = 400$, $q^{24} = 600$, $q^{34} = 200$ (passengers/hour).

Table 4.2 shows the service comparison attributes of the three lines before and after the transit schedule design, with the convention factor of variance of passenger network travel time $\gamma_1 = 0.5$ and different OD multipliers. It can be seen that the optimal transit schedules, not only saved passenger total travel time and enhanced service reliability, but also enabled the reschedule of vehicles to other lines to alleviate congestion. For the normal (OD multiplier $\theta = 1.0$) and congested ($\theta = 1.2$) network conditions, both total passengers' generalized travel cost and the cost of uncertainties (the weighted variance of total passengers' generalized travel cost) decreased following the schedule design. It is also shown that one vehicle in Line 2 was scheduled to Line 1 when demand increased.

Table 4.2 Service Attributes Comparison before and after Line Schedule Design (γ_1 = 0.5)

OD multiplier	1.0		1.2	
Dispatching headways (min)	Even	Balanced	Even	Balanced
Line 1	8	11,8,9,10,6,5,9	8	4,4,3,11,7,10,10,7
Line 2	8	10,9,5,5,9,4,10	8	13,7,5,9,6,10
Line 3	15	15, 10, 11, 20	15	12,9,11,11
Total passengers' generalized travel cost	9614.5	9043	12265	11382
Cost of uncertainties	1266.8	1103.5	1505.3	1450.5

Note: Cost of uncertainties: the weighted variance of total passengers' generalized travel cost.

To better understand when to reschedule vehicles from Line 2 to Line 1, occupied and available vehicle capacity before and after the schedule design are illustrated in Figure 4.2. When the network was not congested, the utilization of vehicles was improved, but not to a large extent. Two vehicles in Line 2 were not utilized at all before and after the schedule design. This means that the number of vehicles is more than enough to accommodate passenger demand. However, when demand increased by 20%, there was vehicle shortage on Line 1. Therefore one vehicle was scheduled from Line 2 to Line 1 to alleviate this shortage. In addition, one vehicle in Line 2 still remained unutilized, but was not rescheduled to any other lines. This is a consequence of maximum headway and fleet size constraints.

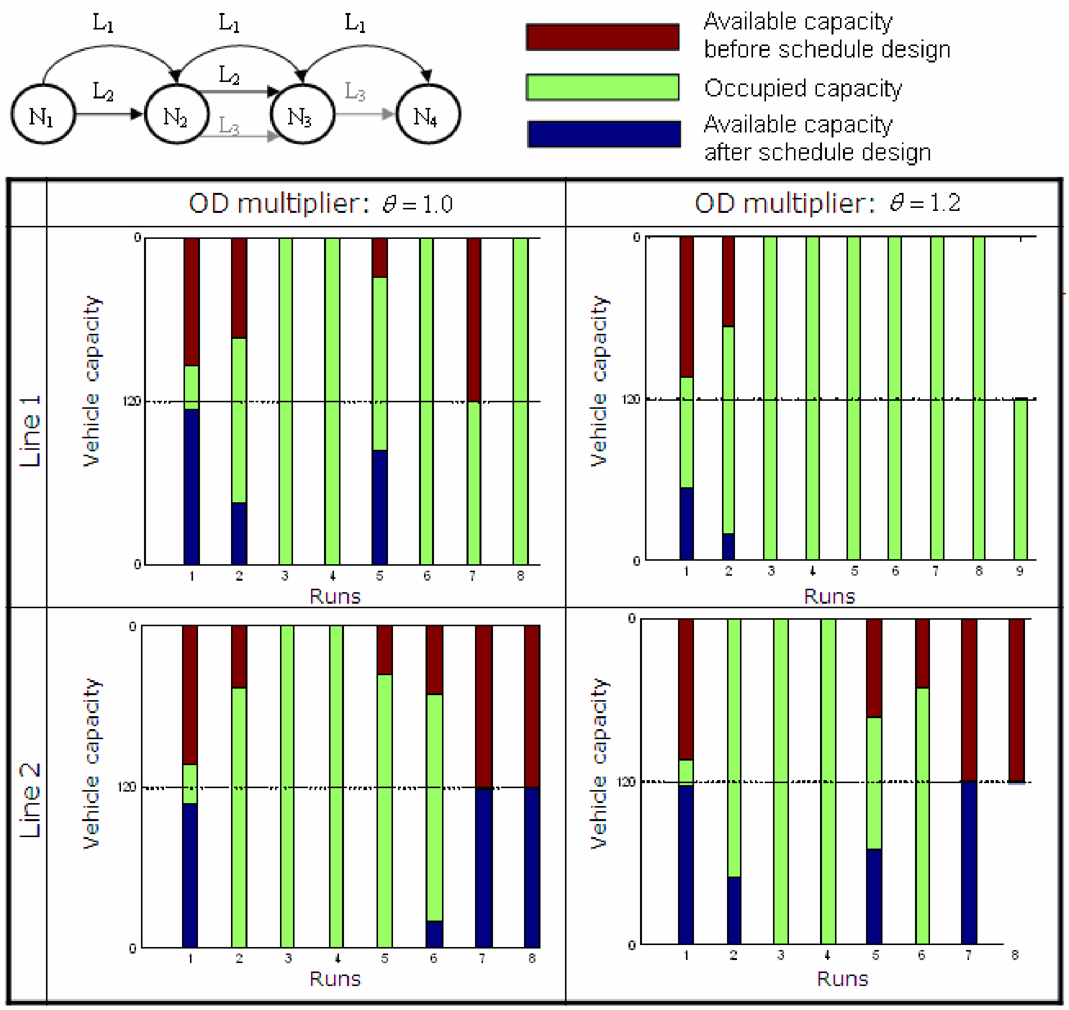


Figure 4.2 Vehicle Capacity Illustration for Different OD Multipliers

The change of vehicle schedules impacted passenger travel decisions. It can be seen from Figure 4.3, passenger traveling between node N_1 and N_4 on Route 4 was smoother after the transit network schedule design. The more discretized passenger flow distribution is intuitively beneficial in saving passenger waiting time and travel time, as well as in maintaining transit service reliability. Passengers depart later after schedule design in Page 75 because Route 4 is a transfer route from Line 2 to Line 3. The line schedule of Line 3 after design is postponed comparing to the schedule before design.

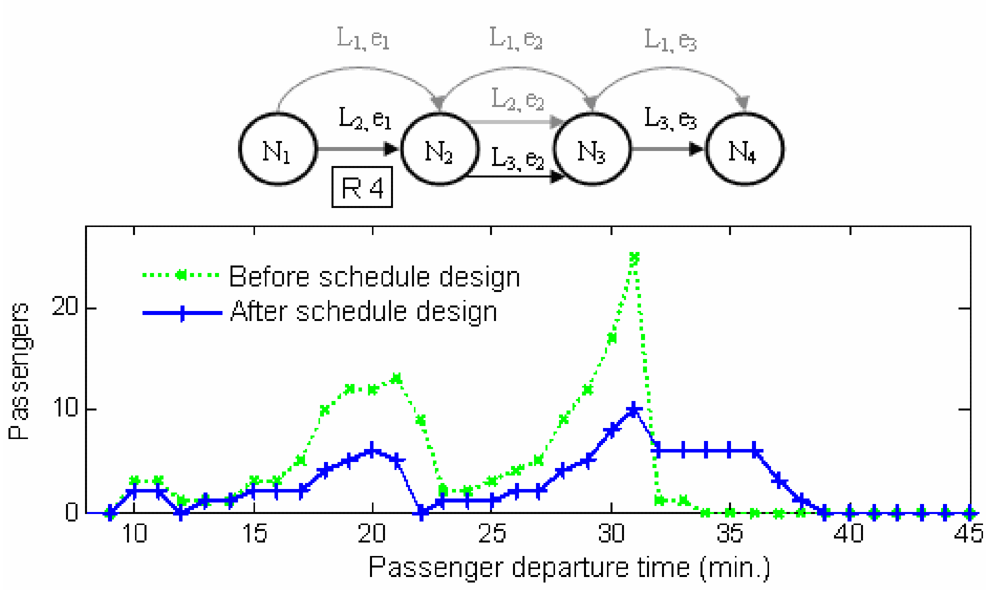


Figure 4.3 Passenger Departure Time Choices on Route 4 before and after Line Schedule Design

Figure 4.4 shows the sensitivity analysis results for different passenger lateness penalty under respective optimal transit line schedules. Passenger departure time is influenced by both passenger lateness penalty and the designed transit vehicle schedule. The results of the sensitivity analysis indicate that passengers taking Line 1 (Routes 1, 2 and 3) of OD 1 depart early because the line schedule after optimization is ahead of the schedule before optimization. When the lateness penalty is equal to 4, passengers depart earlier after optimization because the line schedule accommodated passengers' needs of early departure when the lateness penalty is high.

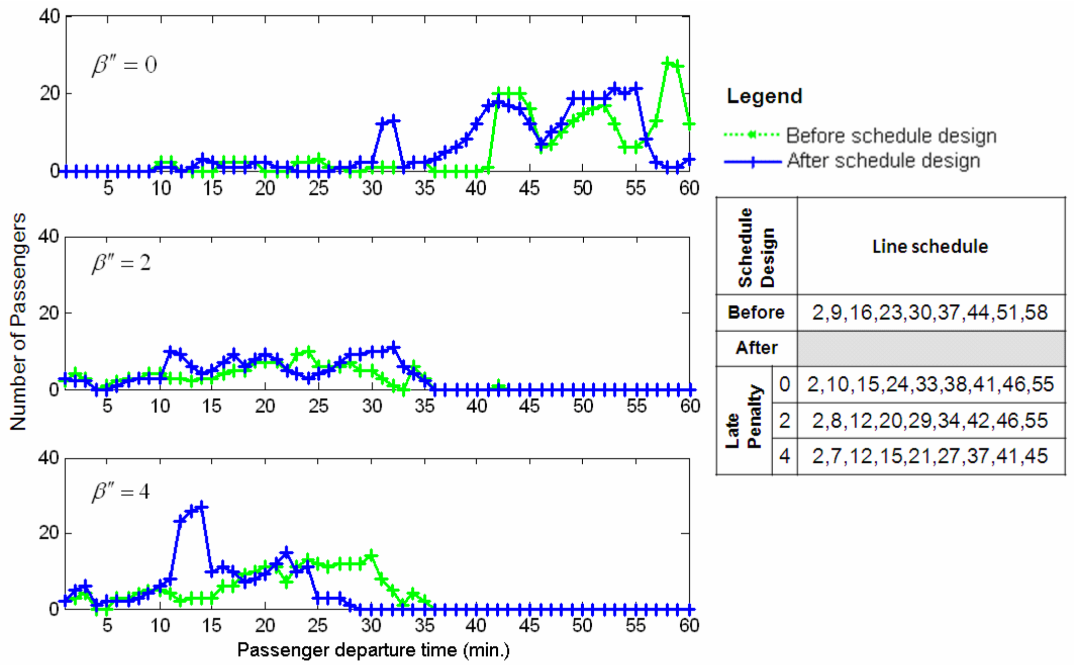


Figure 4.4 Passenger Departure Time of Line 1 (Route 1, 2 and 3) with Different Passenger Lateness Penalty

4.5 SUMMARY

A new bi-level model for solving transit scheduling problems in dynamic and stochastic transit networks has been proposed in this chapter. The upper-level model involves the change of transit schedule to simultaneously optimize the transit network efficiency and reliability. The lower-level model explicitly considers the demand and supply uncertainties, as well as passenger behavioral responses. The bi-level model was solved by Genetic algorithm, which produced a stabled approximation of the optimal transit line schedules.

The numerical example showed the effectiveness of the transit scheduling model and the solution algorithm. After the transit line schedule design, the data showed a saving in total passengers' generalized travel cost, while the number of vehicles on different lines was also better balanced. The passenger load profile, after the vehicle schedule alternation, was smoother, which implied a better level of transit service had been developed.

CHAPTER 5

TRANSIT LINE SCHEDULE DESIGN UNDER NETWORK UNCERTAINTIES IN OLIGOPOLY TRANSIT MARKET

In this chapter, a new model is proposed for solving the line schedule design problems under transit network uncertainties in a deregulated competitive transit market. In a regulated transit market, where transit service is run by government authorities, transit line schedule design aims at saving passenger travel time and improving passenger travel time reliability. However, this may not be true in a deregulated transit market, in which private operators aim at profit maximization. Scheduling problems have been investigated in the literature mostly from deterministic perspectives. In the proposed model, both operator and passenger risk preferences are considered under network uncertainties with respect to the demand and supply sides. The objective is to maximize individual operators' α-confident profit, defined as the stochastic profit within a confidence threshold α. The proposed model is expressed as an equivalent variational inequality (VI) problem with equilibrium constraint. The equilibrium constraint is the RSUE problem proposed in Chapter 3. A diagonalization algorithm is adapted to solve the VI problem. A simple network is used to illustrate the performance of the model and solution algorithm together with discussion on some insightful findings. This chapter is an edited version of: a working paper prepared by Zhang, Y.Q., Lam, W.H.K., and Sumalee, A. (2011) for submission to SCI journal.

5.1 INTRODUCTION

In transit markets which are not fully deregulated, transit companies operate under franchise grants or concessions by authorities (for example, the subway in London, the bus and railway in Hong Kong, and the railway and tram in Melbourne). Franchises can alleviate some negative effects such as safety hazards and informal ownership structures (Gomez-Lobo, 2007). However, the scope of operation is limited when the route, fare and frequency are ruled by authorities by granting franchises. Thus, the method for transit private operators to improve service quality and attract passenger patronage lies in short-term transit operational planning.

Transit operational planning (including frequency and timetable setting, vehicle and crew scheduling) provides a mean by which existing resources can be used in response to variations in transit networks and markets to enhance service efficiency. The flexibility of transit operational planning, however, has not yet been well investigated (Guihaire and Hao, 2008). Only a small number of airline competition models (Powell, 1982; Powell and Winston, 1983) have investigated transit fare, frequency and capacity optimization problems, accounting for passenger demand variations. Recently, many frequency optimization and timetable synchronization models have been studied by transit researchers (Ceder et al., 2001; Gao et al., 2004; Li et al., 2009). However, these models deal mostly with static transit problems for long-time planning purposes. The dynamics and randomness of passenger demand

and transit service still needs research attention.

In franchised transit markets, each operator aims to maximize profit, but also ensures that profit variation over the course of time is acceptable. The operator’s revenue and cost, constituting net profit, are adversely impacted by network uncertainties from demand and supply sides. Revenue is random, mainly due to the variability of passenger patronage. The sources of such variability are identified as (a) day-to-day passenger total demand variation and (b) passenger choice variation as regards departure time and travel route. The cost variability lies in the operation processes, such as dispatching and traveling. In addition, the shortage of vehicles, resulting from insufficient slackness time, prolonged on-road time, terminal congestion, and emergency events, can further lead to vehicle availability problems (Higgins, et al., 1996). In such instances, extra vehicles are needed to meet the frequency constraints of franchise requirements. The extra dispatching cost varies significantly, possibly in accordance with overtime work, vehicle renting or vehicle sharing agreements (Zuckerman and Tapiero, 1980).

The transit line schedule can impact both a transit operator’s revenue and cost. A transit schedule which reflects passenger need, such as acceptable waiting time and fair in-vehicle congestion, can attract more passengers and generate more revenue. However, such schedules may conflict with dispatching reliability or encounter severe road congestion. These risks will be in accordance with the different risk-taking attitudes and may, therefore be valued differently by transit operators Behavioral characteristics of competitive transit operators, the resulting transit line schedules, and the reaction of passenger travel decisions are the practical problems which may be better addressed by scientific appraisal methods.

However, few studies, reported in the literature, are seen to investigate the risk-taking behavior of transit companies. The behavioral characteristics of operators, however, have been studied as regards road traffic network design (Chen et al. 2007) or airline competition models (Powell and Winston, 1983). Different criteria have been used in the examination of stochasticity in network design problems. Included are the mean-variance optimizing models (Chen et al., 2003) and probability maximizing models (Sumalee et al., 2006; Chootinan et al., 2005). Chen et al. (2007) adopted the Value-at Risk measure to model planners risk preferences. The confidence level α as regards total travel time is used to identify planner's level of risk. A variant of chance-constrained model is used to minimize the total travel time.

In this chapter, the α-confident profit is considered rather than the generalized utility or average profit in previous models. The α-confident profit is defined by operator risk preferences. Chance-constrained programming is applied to reflect the α probability marginal effect. The advantages of the α-confident profit model, compared to the previous transit operation models, are 1) the ability to reflect operators' varying risk preferences and (2) the absence of the need for exogenous parameters for the consideration of profit reliability when stochastic profit characteristics are examined.

A bi-level model is formulated in the form of the Stackelberg game to represent the interaction between operator line schedules and passenger travel decisions. Uncertainties in dynamic transit networks are investigated by considering time-dependent demand and supply interaction in the schedule-based modeling framework. In the lower-level model, the RSUE transit assignment model proposed in Chapter 3 is applied to represent passengers' route and departure time choices. In the upper-level model, operators decide the best line schedule schemes to maximize their α-confident profit, given the pre-defined level of confidence α.

The transit line scheduling problem, with the objective to maximize the α-confident profit for each competitive operator, is a multi-variable bi-level transit network optimization problem. Such problems are difficult to solve using mathematical optimization methods because of their discrete, non-linearity, and combinatorial natures (Baaj and Mahmassani, 1991; Zhao and Zeng, 2008). One of the most common approaches used to solve the nonlinear and asymmetry equilibrium problems is the diagonalization method (Dafermos, 1982; Friesz et al., 1984; Harker, 1984). The diagonalization method is adapted in this chapter to find the equilibrium solution for operators with risk preferences, while convexity of the objective function is demonstrated.

This chapter is organized as follows: the basic stochastic assumptions are given and discussed in Section 5.2. The bi-level modeling framework is presented in Section 5.3 and included are a lower-level RSUE model, upper-level α-confident profit maximization model, and bi-level formulation. The competition model is described in Section 5.4, together with the proof of solution existence and uniqueness. The diagonalization algorithm is presented in Section 5.5. Finally a simple transit network in Hong Kong, from TsingYi to Hong Kong International Airport, is used to show the performance of the proposed model and algorithm.

5.2. BASIC ASSUMPTIONS AND NOTATION

The schedule-based transit network is used to present the transit service temporal and spatial evolution. A vehicle's arrival and departure can be explicitly represented in the schedule-based transit network by the illustration of vehicle trajectory by time and space dimensions. The vehicle dwell time, as the response of passenger arriving, boarding, and alighting processes, particularly, is also accessible. The illustration of the vehicle time and space trajectory in the schedule-based transit network is shown in Figure 4.1 of Chapter 4.

Define $A_{l,u} = (..., a_{l,u}, ...)$ as the line-route incidence matrix, and $B_{l,k} = (..., b_{l,k}, ...)$ as the line-agency incidence matrix:

$$a_{l,u}\begin{cases}1 & \text{if line } l \text{ forms part of route } u \\ 0 & \text{Otherwise}\end{cases}, \quad (5.1)$$

$$b_{l,k}\begin{cases}1 & \text{if line } l \text{ is run by operator } k \\ 0 & \text{Otherwise}\end{cases}. \quad (5.2)$$

Using these vector operations, the incidence relationships can now be written in matrix notation as:

$$F_l = \sum_u \sum_i a_{l,u} F_u^i , \quad (5.3)$$

$$\psi^k = \sum_l b_{l,k} \psi_l . \quad (5.4)$$

Several assumptions for studying the stochastic effects are made and given in this chapter:

A1. Operators dispatch vehicles exactly in accordance with the schedule. However, vehicle arrival, departure, and availability at terminals are affected by many factors, such as stochastic run time, level of slackness, terminal congestion, and emergency events (Higgins, et al., 1996). As vehicle dispatching availability is random, extra vehicles may be needed from stock, or by renting or sharing agreements with other transit companies (Zuckerman and Tapiero, 1980). It is assumed that the stochastic variable of dispatching an unplanned vehicle D_p follows the Bernoulli distribution:

$$D_p \sim (p^d , p^d (1 - p^d)) . \quad (5.5)$$

p^d is the probability the extra vehicle is used according to the current line schedule when the stochastic time of vehicle availability for dispatching is later than the scheduled dispatching time:

$$p_{l,i}^{d} = P(g_{l,i} < T_{l,i,1}^{a}) . \tag{5.6}$$

The cost of running the extra vehicle is also stochastic, as the cost may stem from such as hiring vehicles from other transport agencies or from drivers' overtime payment. The cost of dispatching each vehicle is also assumed to follow the Normal distribution:

$$Cd_{l} \sim N(\mu_{l}^{cd}, \sigma_{l}^{cd}) . \tag{5.7}$$

Assuming the independence of the probability of dispatching an extra vehicle and the cost for dispatching, the penalty for dispatching the extra vehicle (dispatching penalty for short) is presented as:

$$Cd_{l,i} \sim N(p_{l,i}^{d}\mu_{l}^{cd}, (1-p_{l,i}^{d})p_{l,i}^{d}\sigma_{l,i}^{cd^2}) . \tag{5.8}$$

A2. The vehicle on-road running and individual passenger boarding times are both stochastic and assumed to follow the Normal distribution. Both standard deviations are assumed to be the positive linear function of the mean value. This can be interpreted as the longer the trip/boarding time, the greater the uncertainties. Their corresponding connections, evaluated by the coefficient variation (CV, the quotient of mean and standard deviation), are constant. Following this assumption, the CV of these two stochastic variables are given as β_v and β_b. The distributions of these two stochastic variables are:

$$Tv_{l,i} \sim N(E(T_{l,i,n}^{a} - T_{l,i,1}^{a}),\ \beta_v^2 E^2 (T_{l,i,n}^{a} - T_{l,i,1}^{a})) , \tag{5.9}$$

and

$$B(\tau) \sim N(\sum_{\tau} \mu_b, \sum_{\tau} \beta_b^2 \mu_b^2) \,. \tag{5.10}$$

5.3 THE BI-LEVEL FRAMEWORK FOR A SINGLE TRANSIT OPERATOR

In order to make clear the relationships between transit operators and passengers, an interaction between a single operator and passengers is firstly investigated. In such a case, the transit operator optimizes his interest by structuring the line schedule. At the same time, both stochastic passenger demand and travel choice, in response to line schedule changes, are taken into account. The line scheduling problem is represented as a leader-follower Stackelberg game, in which the transit operator is the leader and passengers are the followers. It is assumed that the operator can influence but cannot control passenger departure time and route choices.

The operator's objective function adopted for this task is to maximize the net profit (revenue minus cost) by structuring efficient line schedules. The operator's revenue R and cost C, as well as the net profit Φ:

$$\Phi = R - C \tag{5.11}$$

are all stochastic. In a stochastic demand and supply transit network environment, the operator aims not only to gain the expected profit, but also to meet the reliability requirement measured by a confidence level. Given such confidence level α, the operator's profit maximization can be formulated as a chance-constrained programming as follows:

$$\text{Max } \phi \tag{5.12 a}$$

$$\text{s.t. } P\{\Phi \geq \phi\} \geq \alpha\,, \tag{5.12 b}$$

where Φ is the stochastic profit and φ is the threshold of the stochastic profit. Equation (5.12) can then be converted into a deterministic presentation as:

$$\text{Max } \psi = E(\Phi) + \varphi^{-1}(\alpha) \cdot Std(\Phi)\,, \tag{5.13}$$

where ψ is the α - confident profit defined before.

Given the probability distribution of the stochastic revenue and cost, the above α - confident (Chen et al., 2007) profit is expanded as:

$$\psi = E(R - C) + \varphi^{-1}(\alpha) \cdot \{Std(R) + Std(C)\} \tag{5.14}$$

where the mean stochastic revenue is the sum of the transit fares of the full passenger flow on run i line l. The standard deviation is the square root of the mean, according to the Poisson process assumption:

$$E(R) = \sum_{l \in A_{l,u}} \sum_{i \in \Pi_{l,i}} cf_l \cdot E(F_{l,i}(\cdot))\,, \tag{5.15}$$

$$Std\,(R) = \sqrt{\sum_{l \in A_{l,u}} \sum_{i \in \Pi_{l,i}} cf_l^2 \cdot E(F_{l,i}(\cdot))}\,. \tag{5.16}$$

In the above two equations, $F_{l,i}$ is the stochastic passenger flow of i^{th} transit vehicle of line l. The stochastic passenger flow is determined by the equilibrium constraint. cf_l is the transit fare of line l.

The cost is the sum of the capital and operation costs for the study period. The operation cost consists of vehicle on-road operation cost (including running and dwelling time) and the cost of extra vehicles introduced to cover the temporary vehicle unavailability. The increased run cost may result from renting or sharing agreements with other transit companies. Additional cost may also be generated from the need for additional crew, vehicles, or fuel. For representation simplicity, the additional cost, possibly generated at dispatching, is termed the dispatching penalty. The hourly capital and operation costs for each line are denoted as cc_l and co_l, respectively. The cost is the sum of capital cost cc_l, dispatching penalty Cd_l, running time cost Cv_l, and dwelling time cost $Cw_{l,i}(\cdot)$. The mean and variance of total cost are:

$$E(C) = cc_l + E(Cd_l) + E(Cv_l) + \sum_{i=\Pi_{l,i}} E(Cw_{l,i}(F_{l,i}(\cdot))), \tag{5.17}$$

$$Std(C) = \sqrt{\operatorname{var}(Cd_l) + \beta_v^2 \cdot (E(Cv_l) + \sum_{i=\Pi_{l,i}} E(Cw_{l,i}(F_{l,i}(\cdot))))^2}\,. \tag{5.18}$$

The stochastic vehicle dwell time is shown as a function of passenger flow and individual passenger boarding time in Equation (5.10), Section 5.2:

$$Cw_{l,i}(F_{l,i}(\cdot)) \sim N(E(F_{l,i}(\cdot)) \cdot \mu_b, E(F_{l,i}(\cdot)) \cdot \beta_b^2 \cdot \mu_b^2). \tag{5.19}$$

β_v and β_b represents the coefficient variations of the vehicle on-road running time and individual passenger boarding time.

According to the above derivations in Equations (5.15-5.19) and vehicle stochastic dispatching penalty function in Equation (5.8), the α - confident profit function of one operator ψ^k, consisting of the revenue R^k, cost C^k, and the operator's risk preference margin S^k, is:

$$\psi^k(\alpha)=\mathrm{E}(R^k)-\mathrm{E}(C^k)-S^k(\alpha). \tag{5.20}$$

s.t.

$$R^k=\sum_{l\in B_{l,k}}(cf_l-\mu_b co_l)\cdot\sum_{i\in\Pi_{l,i}}F_{l,i}(g_{l,i}), \tag{5.21}$$

$$C^k=cc_l+\sum_{i\in\Pi_{l,i}}\mu_l^{cd}p_{l,i}^d(g_{l,i})+co_l\cdot Tv_l\ , \tag{5.22}$$

and

$$S^k(\alpha)=-\varphi^{-1}(\alpha)\sqrt{\begin{array}{l}\sum_{l\in B_{l,k}}\{(cf_l^2+\beta_b^2co_l^2\mu_b^2)\cdot\sum_{i\in\Pi_{l,i}}E(F_{l,i}(g_{l,i}))\\+\sum_{i\in\Pi_{l,i}}\sigma_l^{cd2}p_{l,i}^d(g_{l,i})^2+\beta_v co_l^2\cdot E(Tv_l\ )\}\end{array}}. \tag{5.23}$$

The α-confident profit maximization problem for a single operator k given confidence level α^k is:

$$\max_{g^k}\quad \psi^k(\mathrm{g}^k,\alpha^k). \tag{5.24}$$

5.4 THE BI-LEVEL FRAMEWORK FOR COMPETITIVE TRANSIT OPERATORS

The introduction of private operators for transit services has been adopted in many cities to alleviate the government's fiscal burden. The objective of private operators, however, is profit maximization rather than welfare gain or efficient utilization of vehicles. Consequently, a Nash game is initiated between different transit operators. As the profit fluctuates following the network demand and service variation, operators are likely to have different profit confidence levels, when considering and deciding upon transit network design schemes. In this section, a bi-level problem is formulated. The upper-level problem represents the Nash game between transit operators and the lower-level problem represents passenger responses to operators' line schedule schemes. The Stackelberg game between operators and passengers is described in Section 5.3.

Operator line schedule strategies, including the transit line schedule and risk preference, have significant effects on passenger route choices and transit line flow patterns. The interaction between operators and passengers lead to a competition in which each operator seeks to maximize his own revenue. The revenue directly relates to the volume of boarding passengers on the line. In turn, passenger flow on the line is dependent on vehicle arrival time. Each operator in this competitive market seeks to attract as many passengers as possible onto his own service lines, while the total passenger demand is stochastic, but follows the fixed mean and other recognized stochastic properties.

5.4.1 Convexity of the α-Confident Profit Function

Many researchers have studied the existence and uniqueness of a VI problem by demonstrating the objective function to be strictly increasing, continuously differentiable, and convex. In the following part of this section, the convexity of each component of the operator α -confident profit function is shown, so that the convexity of profit function can remain by the summation. The second derivative of the mean of the stochastic revenue function is:

$$\frac{\partial^2 E(R^k)}{\partial^2 g_{l,\bar{i}}} = \sum_{l \in B_{l,k}} (cf_l - \mu_b co_l) \cdot \frac{\partial^2 E(F_{l,\bar{i}}(g_{l,\bar{i}}))}{\partial^2 g_{l,\bar{i}}} \tag{5.25}$$

As discussed in Section 3.1, the passenger flow is assigned by a RSUE model. Passenger perception error is assumed to follow the Normal distribution and the probability of passenger flow distribution $P(gc(t,u))$ is a function of passenger generalized travel cost $gc(t,u)$. Thus the second derivative of the expected passenger flow is:

$$\begin{aligned}
\frac{\partial^2 E(F_{l,\bar{i}})}{\partial^2 g_{l,\bar{i}}} &= \frac{E(Q^{rs}) \cdot a_{l,u} \cdot \partial^2 P(gc(t,u))}{\partial^2 g_{l,\bar{i}}} \\
&= E(Q^{rs}) \cdot \frac{\partial^2 \int_{-\infty}^{gc} \phi_{\mu_{GC},\sigma_{GC}^2}(x)dx}{\partial^2 gc(t,l)} \cdot \frac{\partial^2 gc(t,l)}{\partial^2 g_{l,\bar{i}}} \\
&= E(Q^{rs}) \cdot \frac{\partial \phi_{\mu_{GC},\sigma_{GC}^2}(gc(t,l))}{\partial gc(t,l)} \cdot \frac{\partial^2 ett_l}{\partial^2 g_{l,\bar{i}}} \\
&= E(Q^{rs}) \cdot \frac{-(gc(t,l) - \mu_{GC})}{\sigma_{GC}^3 \sqrt{2\pi}} \cdot \exp\left(\frac{-(gc(t,l) - \mu_{GC})^2}{2} \right) \\
&\quad \cdot \frac{\partial^2 (E(C(t,l)) + \varphi^{-1}(\alpha) \cdot Std(C(t,l))}{\partial^2 g_{l,\bar{i}}}
\end{aligned} \tag{5.26}$$

When the mean of the stochastic term (passenger perception error in generalized travel cost) ε is larger than zero, the second term in Equation (5.26) $\frac{-(gc(t,l) - \mu_{GC})}{\sigma_{GC}^3 \sqrt{2\pi}}$ is negative. The convexity of passenger flow with respect to the change of line schedule is thus determined by the last term. The passenger stochastic travel time $C(t,l)$, defined in Equation (2.29) consists of passengers waiting time, in-vehicle travel time, and transfer time. These terms and the vehicle arrival time have linear relationships, which can be observed from Equations (2.23-2.27). Thus the second derivatives of the mean and standard deviation of the $C(t,l)$ equals zero:

$$\frac{\partial^2 (E(C(t,l)) + \varphi^{-1}(\alpha) \cdot Std(C(t,l))}{\partial^2 g_{l,\bar{i}}} = 0 \tag{5.27}$$

and Equation (5.26) equals zero:

$$\frac{\partial^2 E(F_{l,\bar{i}})}{\partial^2 g_{l,\bar{i}}} = 0 . \tag{5.28}$$

The second derivative of the mean of the stochastic operator cost function is

$$\frac{\partial^2 E(C^k)}{\partial^2 g_{l,\bar{i}}} = \mu_l^{cd} \frac{\partial^2 p_{l,\bar{i}}^d (g_{l,\bar{i}})}{\partial^2 g_{l,\bar{i}}} cc_l \quad . \tag{5.29}$$

As stated in Section 5.2, the probability of punctually dispatching the vehicle $V_{l,i}$ is the function of the cumulative density function (CDF) of the Normal distribution:

$$p_{l,\bar{i}}^d (g_{l,\bar{i}}) = \int_{-\infty}^{g_{l,\bar{i}}} \phi_{\mu,\sigma^2}(u) du . \tag{5.30}$$

The second derivative of the probability of punctually dispatching is:

$$\begin{aligned} \frac{\partial^2 p_{l,\bar{i}}^d (g_{l,\bar{i}})}{\partial^2 g_{l,\bar{i}}} &= \frac{\partial \varphi_{\mu_l^{cd}, \sigma_l^{cd^2}}(g_{l,\bar{i}})}{\partial g_{l,\bar{i}}} \\ &= \frac{1}{\sigma_l^{cd}\sqrt{2\pi}} \cdot \frac{\partial \exp\left(\frac{-(g_{l,\bar{i}} - \mu_l^{cd})^2}{2\sigma_l^{cd^2}} \right)}{\partial \left(\frac{-(g_{l,\bar{i}} - \mu_l^{cd})^2}{2\sigma_l^{cd^2}} \right)} \cdot \frac{\partial \left(\frac{-(g_{l,\bar{i}} - \mu_l^{cd})^2}{2\sigma_l^{cd^2}} \right)}{\partial g_{l,\bar{i}}} . \\ &= \frac{-(g_{l,\bar{i}} - \mu_l^{cd})}{\sigma_l^{cd^3}\sqrt{2\pi}} \cdot \exp\left(\frac{-(g_{l,\bar{i}} - \mu_l^{cd})^2}{2} \right) \end{aligned} \tag{5.31}$$

It can be observed from the above equation, if $g_{l,\bar{i}} \geq \mu_l^{cd}$, $\dfrac{\partial^2 p_{l,\bar{i}}^d(g_{l,\bar{i}})}{\partial^2 g_{l,\bar{i}}} \leq 0$; otherwise, $\dfrac{\partial^2 p_{l,\bar{i}}^d(g_{l,\bar{i}})}{\partial^2 g_{l,\bar{i}}} > 0$. Thus, the negative of the mean of the stochastic cost function is convex, as the cost is a negative component in the α - confident profit function.

The second derivative of operator α -confident risk preference measure:

$$\frac{\partial^2 S^k(\alpha)}{\partial^2 g_{l,\bar{i}}} = \frac{\partial^2 \varphi^{-1}(\alpha) \cdot \sqrt{\begin{array}{l}\sum\limits_{l \in B_{l,k}} \{ (cf_l^2 + \beta_b^2 co_l^2 \mu_b^2) \cdot \sum\limits_{\bar{i} \in \Pi_{l,\bar{i}}} E(F_{l,\bar{i}}(g_{l,\bar{i}})) \\ + \sum\limits_{\bar{i} \in \Pi_{l,\bar{i}}} \sigma_l^{cd\,2} p_{l,\bar{i}}^d (g_{l,\bar{i}})^2 + \beta_v co_l^2 \cdot E(Tv_l\ ) \}\end{array}}}{\partial^2 g_{l,\bar{i}}} . \tag{5.32}$$

The second derivative of the square-root function is definitely positive. The formula inside the square root, which determines the signs of the second derivative, is also positive as there is no negative component. Thus, the confident value $\varphi^{-1}(\alpha)$ is the only factor determining whether the second derivative is positive or negative.

In this chapter, heterogeneous risk-aversion operators are considered in the oligopoly transit market. These risk-averse operators can also represent the majority of transit agencies, as reliability of profit is an important factor in the transit system design and operation. The risk preference α is from the highest confidence (lowest risk) to the average situation (risk is not considered), which means the sign of the second derivative of the operator’s α -confident risk preference measure is positive. Similar to the cost component, the negative of the risk preference measure is convex, as it is a negative component in the operator’s α -confident profit function.

The convexity of the mean revenue, cost, and the operator's risk preference has been demonstrated in the section above. Thus the operator k's α-confident profit, the summation of the three components, is also convex.

5.4.2 The Formulation of Generalized Nash Equilibrium

The strategies of each operator are affected by others operators. This leads to a generalized Nash game presented by Harker and Pang (1990), formulated as:

$$\psi^k(G^{k*}, G^{-k*}) \geq \psi^k(G^k, G^{-k*}) \ . \tag{5.33}$$

where $G = (\cdots, g_{l,i}, \cdots)$, is the vector of passenger line schedule. * denotes the optimal line schedule strategy, and $-k$ denotes operators, other than operator k in the market. According to Harker (1991), the above general Nash equilibrium (GNE) can be formulated as a quasi-variational inequality (QVI) problem:

$$\sum_k V^k(G^*)^{\mathrm{T}}(G^k - G^{k*}) \geq 0 \,. \tag{5.34}$$

$V^k(G)$ is the negative gradient of the profit ψ^k:

$$V^k(\mathbf{y}) = -\nabla_{G^k}\psi^k(G) \,, \tag{5.35}$$

where

$$\frac{\partial \psi^k(\mathbf{g},\alpha^k)}{\partial g_{l,j}} = (cf_l - \mu_b co_l)\cdot\frac{\partial E(F_{l,j}(g_{l,j}))}{\partial g_{l,j}} - \mu_l^{cd}\frac{\partial p_{l,j}^d(g_{l,j})}{\partial g_{l,j}}$$

$$-\frac{\phi^{-1}(\alpha)}{2}\cdot\left\{(cf_l^2+\beta_b^2 co_l^2\mu_b^2)+2\sigma_l^{cd\,2}p_{l,j}^d(g_{l,j})\cdot\frac{\partial p_{l,j}^d(g_{l,j})}{\partial g_{l,j}}\right\}\cdot \tag{5.36}$$

$$\cdot\sqrt{\begin{array}{l}\sum_{l\in S_l^k}\{(cf_l^2+\beta_b^2 co_l^2\mu_b^2)\cdot\sum_{j\in O_j^l}E(F_{l,j}(g_{l,j}))\\ +\sum_{j\in O_j^l}\sigma_l^{cd\,2}p_{l,j}^d(g_{l,j})^2+\beta_v co_l^2\cdot E(Tv_l)\}\end{array}}^{\,-\frac{1}{2}}$$

For the full Cartesian product of the individual operator's strategy sets, the equivalent VI problem for the Nash game is simply the summation of the individual operator's first-order conditions (Harker and Pang, 1990):

find G such that

$$\sum_{k\in K}-\nabla_{G^k}\psi^k(G)^T(G-G^*)\geq 0\ . \tag{5.37}$$

The above variational equivalent formulation in Equation (5.37) is equivalent to the following mathematic programming:

$$\min\sum_{l'\in S_l^k,\,j'\in O_j^l}\int_0^{g_{l',j'}}-\frac{\partial\psi^k(u,\alpha^k)}{\partial g_{l,j}}du\ , \tag{5.38}$$

and also equivalent to

$$\max_{G^k}\ \psi^k(G^k,G)\ . \tag{5.39}$$

5.5 SOLUTION ALGORITHM

The diagonalization algorithm is also known as the nonlinear Jocabi method, popularly used in studies and reported in the literature to solve nonlinear and asymmetry equilibrium problems. It is widely used to solve both traffic equilibrium problems (Dafermos, 1982; Friesz et al., 1984) and equilibrium network design problems (Harker, 1984; Friesz and Harker, 1983). From each iterative step of the diagonalization algorithm, an iterative decision variable is produced. The diagonalization of the objective function is performed to obtain the optimal objective function value by finding the optimal diagonalization variable. The algorithm continues until a satisfactory termination criterion is met.

The algorithm starts by selecting an initial G^0, followed by solving the variational inequality by step n $(n = 1, 2, ...)$:

$$\psi^{\mathrm{T}}(G^n, G^{n-1})(G - G^n) \geq 0 \ , \qquad (5.40)$$

The equivalent form of Equations (5.37-5.39), the variational inequality with respect to each individual operator Equation (5.40), is then solved by the diagonalization algorithm.

The steps of the solution algorithm to the equilibrium α - confident schedule design problem are as follows:

Step 0 Choose an initial G^0, set $n = 0$.

Step 1 Solve Equation (5.40) for $G^{k,\ n+1}$ for each $k = 1, 2, ..., m$:

Step 1.1 Initialization for the start of diagonalization.

Step 1.2 Find the temporary optimal line schedule $g_{l,j}^{k,n+1}$ for each line and run, where $g_{l,j}^{k,n+1} \in \{g_{l',j'}^{k',n}, ..., g_{l,j-1}^{k,n}, g_{l,j}^{k,n+1}, g_{l,j+1}^{k,n}, ...\}$, based on the equilibrium passenger flow. The distribution of passenger travel choices is obtained by the

MSA-type algorithm for the lower-level RSUE problem, responding to the updated line schedules.

Step 1.3 Find the temporary optimal line schedules $G^{k,n+1}$ for each operator. $G^{k,n+1}$ is the feasible descent direction of line schedules of operator k at each line and run, but not the whole operation. As the actual relationship between operator's line schedule and their α -confident profit is not obvious, the enumeration with all combination from $G^{k,n}$ to $G^{k,n+1}$ is performed by the change of single minute.

Step 2 If $G^{k,n+1} = G^{k,n}$ for all $k = 1, 2, ..., m$, stop; otherwise, set $n = n + 1$, and return to Step 1.

5.6 NUMERICAL EXAMPLE

A simplified network connecting Tsing Yi New Town and HKIA is portrayed in Figure 5.1. There are two services between these two locations, the Airport express line (AEL) and a bus line. During the study period of 1 hour, each line has 6 runs to dispatch. The expected time of vehicle availability for both lines is $\bar{G} = (2, 12, 22, 32, 42, 52)$, and the latest time constraint for dispatching is $\hat{G} = (9, 19, 29, 39, 49, 59)$. AEL is an express rail line, thus the designed capacity (500 persons/vehicle) is much larger than the bus line (120 persons/vehicle). The service reliability of AEL is also much higher than that of the bus line, but the capital and operation cost is relatively higher. Table 5.1 shows the respective capital cost, operation cost, and the extra cost of dispatching an extra run.

The designed dispatching time should be lager or at least equal to the average vehicle availability time. The initial dispatching time vector, which is the same as the expected vehicle availability time $G^0 = \overline{G}$, is used to initiate the solution algorithm. In the case of the two operators both holding a strong risk-averse attitude towards stochastic profits, $\alpha^{AEL} = \alpha^{Bus} = 95\%$, the convergence of the algorithm and the optimal line schedule schemes are listed in Table 5.2.

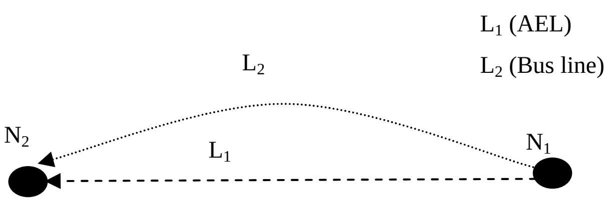


Figure 5.1 Example Transit Network

Table 5.1 Transit System Capital and Operation Costs

Unit: HK$/veh/hr	AEL	Bus
Capital Cost	2000	600
Operation Cost	500	80
Dispatching Penalty: Mean/ Standard Deviation	$2000 \cdot p_{l,i}^d /$ $1000 \cdot \sqrt{p_{l,i}^d (1 - p_{l,i}^d)}$	$600 \cdot p_{l,i}^d /$ $600 \cdot \sqrt{p_{l,i}^d (1 - p_{l,i}^d)}$

The designed dispatching time should be lager or at least equal to the average vehicle availability time. The initial dispatching time vector, which is the same as the expected vehicle availability time $G^0 = \overline{G}$, is used to initiate the solution algorithm.

In the case of the two operators both holding a strong risk-averse attitude towards stochastic profits, $\alpha^{AEL} = \alpha^{Bus} = 95\%$, the convergence of the algorithm and the optimal line schedule schemes are listed in Table 5.2.

It can be observed in Table 5.2, that the algorithm stopped after 6 iterations when the optimal line schedules converged. At equilibrium, both operators' 95% confident profits were considerably improved compared to the initial confident profits. This is because 1) the operation cost, including the dispatching penalty, decreased dramatically by rescheduling the lines; 2) the revenue, generated by passenger patronage, was balanced between the two operators at equilibrium. When equilibrium was reached, the α-confident profit function was still fluctuating (see Iterations 5 and 6). This is because the stochastic network loading resulted in slight passenger loading variation.

Table 5.2 Convergence of the Solution Algorithm

Iteration	Rail operator							Bus operator						
	line schedules (min)						95% confident profit (HK$)	line schedules (min)						95% confident profit (HK$)
0	2	12	22	32	42	52	711	2	12	22	32	42	52	615
1	6	14	27	37	48	57	879	7	14	26	37	49	58	1696
2	7	14	26	38	47	59	937	8	14	27	39	49	58	1757
3	8	15	28	36	48	59	994	9	17	28	38	49	58	2018
4	8	15	27	39	48	59	1231	9	19	28	39	49	59	2013
5	9	14	28	39	48	59	1178	9	18	28	39	49	59	2078
6	9	14	28	39	48	59	1166	9	18	28	39	49	59	2052

Table 5.3 Service Attributes by Operator Risk Preferences and Network Congestion and Uncertainty Levels

	Scenarios As	AI	AII	AIII	AIV
	OD 1.0 BO's SD 2.0	RO RN (50%) BO RN (50%)	RO RN (50%) BO RA (95%)	RO RA (95%) BO RN (50%)	RO RA (95%) BO RA (95%)
RO	Schedule	7 14 27 38 46 58	8 15 28 39 49 58	9 15 26 39 48 59	9 14 28 39 48 59
RO	Patronage	256	259	252	256
RO	Profit Mean/SD (CV)	2566/991 (0.38624)	2891/975 (**0.3373**)	2465/963 (0.3907)	2750/963 (0.3502)
RO	Dispatching cost Mean/SD (% profit)	192/224 (7.482/22.60)	47/121 (1.626/12.41)	53/129 (2.150/13.40)	**8/50** **(0.290/5.192)**
BO	Schedule	7 16 27 37 48 58	9 18 29 39 49 59	8 16 27 38 48 58	9 18 28 39 49 59
BO	Patronage	239	234	**243**	237
BO	Profit Mean/SD (CV)	2756/532 (0.1930)	2734/463 (0.1693)	2819/531 (0.1884)	2817/465 (0.1651)
BO	Dispatching cost Mean/SD (% profit)	168/308 (6.096/57.89)	**50/173** **(1.829/37.37)**	138/280 (4.90/52.73)	**50/173** **(1.775/37.20)**
	Scenarios Bs	**BI**	**BII**	**BIII**	**BIV**
	OD 1.2 BO's SD 2.5	RO RN (50%) BO RN (50%)	RO RN (50%) BO RA (95%)	RO RA (95%) BO RN (50%)	RO RA (95%) BO RA (95%)
RO	Schedule	6 14 28 37 47 57	6 14 26 38 48 57	8 15 29 39 48 57	8 15 28 37 47 57
RO	Patronage	337	**349**	332	344
RO	Profit Mean/SD (CV)	7424/1128 (0.1519)	8138/1148 (0.1411)	7271/1102 (0.1516)	7989/1121 (**0.1403**)
RO	Dispatching cost Mean/SD (% of Profit)	193/237 (2.600/21.01)	199/241 (2.445/20.99)	48/123 (0.6602/11.62)	45/121 (0.5633/10.79)
BO	Schedule	9 **14 25** 37 49 58	9 18 27 37 49 59	9 15 26 38 48 59	9 **19 29** 38 49 59
BO	Patronage	261	252	**262**	255
BO	Profit Mean/SD (CV)	3106/642 (0.2067)	3088/556 (0.1801)	3248/608 (0.1872)	3245/524 (0.1615)
BO	Dispatching cost Mean/SD (% profit)	430/453 (13.73/70.56)	198/332 (6.412/59.71)	315/403 (9.698/66.28)	125/269 (3.852/51.34)

Note: RO=Rail Operator BO=Bus Operator

SD=Standard Deviation CV=Correlated Variation

RN=Risk Neutral ($\alpha = 0.5$) RA=Risk Averse ($\alpha = 0.95$)

The competitions between the rail and bus operators are categorized into four scenarios, each according to one of the four combinations of their risk-neutral and risk-averse attitudes towards the stochastic profit. In Scenario I, both operators are risk-neutral; in Scenarios II and III operators have different risk preferences; in Scenario IV both operators are risk-averse. Both the congested and uncongested network conditions are considered in the model. The four scenarios when the network is not congested are denoted respectively as Scenarios AI-AIV, with parameters of demand multiplier and network congestion: $\theta = 1.0$; $\beta_b = 2.0$. The congested four scenarios are denoted respectively as Scenarios BI-BIV, with parameters of demand multiplier and network congestion: $\theta = 1.2$; $\beta_b = 2.5$. The optimal line schedules and the relevant passenger patronage, the mean, standard deviation and coefficient variation of profit, and the particular dispatching cost in the congested and uncongested network are presented in Tables 5.3. The α -confident profit distributions for different transit network situation are illustrated in Figure 5.2. Unit for the schedule is minute, passenger patronage is passenger per hour, profit and cost is HK$ per hour.

The rail and bus operators having the same risk preferences (either risk-neutral at Scenario AI or both risk-averse at Scenario AIV) are first considered. It is shown in Table 5.3 that the dispatching costs (both mean and standard deviation) are the highest when both operators are risk-neutral and the lowest when both operators are risk-averse. The line schedules are mostly postponed 1-2 min to reach the higher confidence of transit profit. When the transit network became congested and the travel time randomness became more severe for bus, operators' optimal line schedules are delayed even later to compensate for dispatching uncertainty. The dispatching time of the second and third buses is delayed up to 5 min, from 14 and 25 min (at Scenario BI) to 19 and 29 min (at Scenario BIV).

It can also be observed that, if both operators' risk preferences are not considered in the competition model, their average profits are underestimated though the profit stochasticity is higher. Rail operator's average profit increases from 2566 (risk-neutral, Scenario AI) to 2750 (risk-averse, Scenario AIV), and bus from 2756 to 2817 while passenger patronage does not greatly change. Their stochasticity of profit is saved from 991 to 963 and 532 and 465 respectively.

When the road network became congested and the bus mode became unreliable, rail and bus operators' risk-averse preferences (Scenario BIV) lead to an increase in passenger rail patronage (from 337 to 344), but a reduction in bus patronage(from 261 to 255). Rail passenger patronage is seen to increases significantly in Scenario BIV compared to Scenario AIV (from 256 to 344), while bus passenger patronage does not greatly increase (from 237 to 255). The difference in the change of passenger patronage indicates that rail is more attractive to passengers when the transit network environment is adversely congested, with the aggravation of uncertainties, especially when operators' risk preferences are risk-averse. Bus operators, thus have to balance passenger patronage (revenue) and dispatching penalty (cost component) in order to maximize the profit.

It can be observed from the change of CV when the network becomes more congested and vehicle running time becomes more stochastic. The CV of rail operator's profit is too large (higher than 0.3) when the OD demand is small at Scenarios AI-AIV, indicating the high possibility of negative profit. The CV becomes smaller (lower than 0.2) when passenger demand increased at Scenarios B, possibly indicating that ignorance of network uncertainties causes loss of rail mode profit. Based on the data in Table 5.3, short-term planning schemes appear unable to improve the loss of profit for the rail company. Hence the proposal of frequency reduction should be attempted in the long-term planning, provided that the minimum frequency is guaranteed.

For the bus company, dispatching cost saving is important to maintain profit stability. The bus company's average profit at Scenarios AIII and BIII is the highest; however the standard deviation of profit is also high, indicating profit instability. The bus operator's 90-confident profit at Scenario AIV (risk-averse while rail operator risk-neutral) is HK$2221, higher than the 90-confident profit at Scenario AIII (risk-neutral while rail operator risk-neutral). Hence it could be concluded that the bus operator, running an unreliable and low capacity transit service, has to balance operation cost and passenger patronage, in order to increase both the profit itself as well as profit stability.

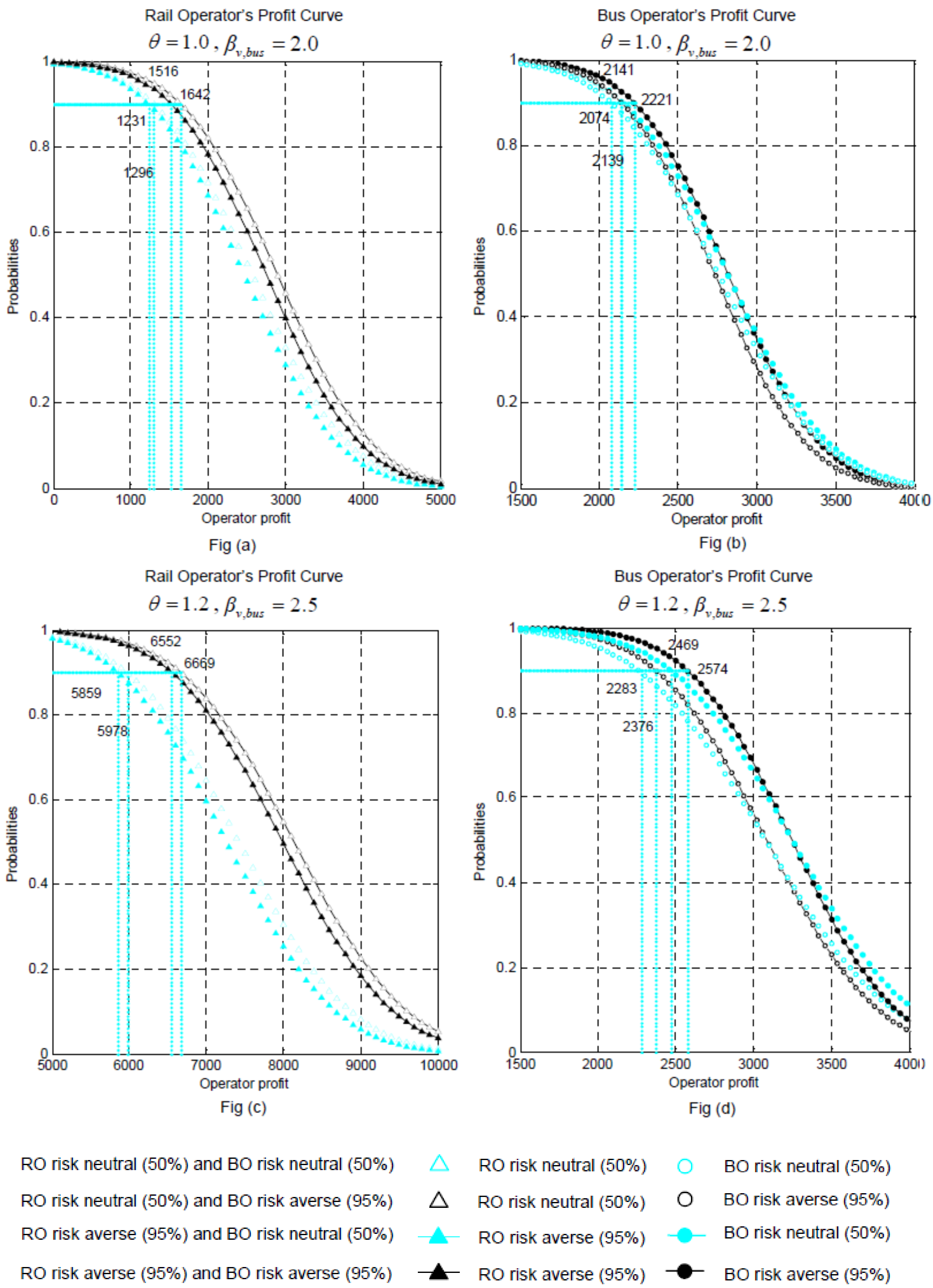


Figure 5.2 Stochastic Profit Curves by Operator Risk Preferences and Network Congestion and Uncertainty Levels

Scenarios AII, AIII, BII and BIII show the influences of market competition when rail and bus operators hold different risk preferences. By changing the line schedule, the rail operator gained the highest passenger patronage by being risk-neutral (259 and 349 respectively in Scenarios AII and BII), while the bus operator had the lowest passenger patronage by being risk-averse (259 and 349 respectively). Thus the most favorable transit market Scenario for the rail company is risk-neutral, and for the bus company, is risk-averse.

The same result can be drawn from Figure 5.2. In Figures 5.2(a) and (c), lines with dark hollow triangles, representing the rail operator's stochastic profit curves when the rail operator risk-neutral and the bus operator risk-averse, have the highest α -confident profit at all confidence levels. These two figures implicate that the rail operator's preference is to be risk-neutral regardless of the risk preference of an opponent. A possible reason relates to the economics of running an expensively built transit mode. Operating at high capacity better ensures a financial return on the capital investment. The greatest concern is to attract more passengers, rather than trying to save operation costs. On the other hand, in Figures 5.2 (b) and (d), lines with dark solid dots, representing the bus operator's stochastic profit curves when the rail and the bus operators both risk-averse, overpass other lines when the confidence level is higher than 0.6. These two figures indicate that the bus company can gain a higher confident profit when risk-averse (α higher than 60 in Figure 5.2 (b) and (d)), but lower when a risk-neutral policy is adopted. Thus, the rail and bus operator reach the equilibrium of the game, Scenario AII or BIV.

The α-confident profit curves of rail and bus operators also shows that if operators' risk-taking behavior is not considered, their profits are both underestimated (lines with solid triangles and dots above lines with hollow triangles and dots). Moreover, the rail and bus operators have very different risk preferences when competing in the same transit market. The rail operator always prefers to be risk-neutral as the risk-neutral profit curves (the lines with hollow triangles) are always above the risk-averse profit curves (lines with solid triangles).

5.7 SUMMARY

In this chapter, the line schedule design problem under network uncertainties has been investigated in a competitive transit market. Operators' risk preferences facing profit uncertainties and the recursive impacts on their optimal transit line schedules have been discussed. Each individual operator's α-confident profit, defined as the stochastic profit within a confidence threshold, has been maximized by the proposed model. The Stackelberg equilibrium between operators and passengers has been formulated as a VI problem and solved by the adapted diagonalization algorithm.

The numerical results give the following important insights:

1) The risk-averse operators tended to postpone the line schedule by a few

minutes to maintain the reliable availability of transit vehicles.

2) If operators' risk preferences were not considered, the average profits were underestimated and the stochasticity of profit was higher for both rail and bus operators.

3) Rail was more attractive to passengers when the transit network environment was adversely congested with the aggravation of uncertainties, especially when both operators were risk-averse. Bus operator had to balance the passenger patronage (revenue) and dispatching penalty (cost component) in order to maximize profit.

4) The rail operator always preferred to be risk-neutral no matter the risk preference held by the opponent. This is because, when operating a reliable, high capacity, and expensively built transit mode, the rail company depended strongly on attracting more passengers, rather than saving operation cost.

5) The rail operator has a high possibility of losing profit when the OD demand is small (though the average profit was positive, the CV of profit was very high, larger than 0.3). The CV became much smaller (lower than 0.2) when passenger demand increased. Thus, the rail operator was recommended to propose long-term planning strategies to the authority, such as to decrease train frequency.

6) When both operators were free to choose the optimal risk preference for transit market competition, the Stackelberg equilibrium was rail operator risk-neutral and bus operator risk-averse.

CHAPTER 6

CONCLUSIONS

6.1 INTRODUCTION

Passenger demand prediction and transit line scheduling are important strategies in improving short-term transit operations. Many models, assuming the study network to be static and deterministic, have been proposed to study these strategies. The lack of recognition of network uncertainties and temporal evolution of transit system, however lead to biased prediction of passenger demand profiles and vehicle operation. Inefficiency of transit schedules, will fail to reflect passenger boarding demand at each stop and time period, causing long passenger waiting time and over-load delay, and finally will lead to the degradation of transit service and a change of passenger mode choice.

The research described in this thesis has explored the transit assignment and short-term planning models for solving passenger flow prediction and transit schedule design problems under network uncertainties. Two RSUE transit assignment models were proposed for passenger flow prediction in networks with single-class risk-averse passengers and multi-class risk preference passengers. Two transit line scheduling models were developed in this research study taken account of passenger travel choices explored in the two proposed transit assignment models.

The main objectives of the research, as stated in Chapter 1, have been to: (a) specify sources of transit network uncertainties, (b) develop dynamic transit assignment models with various passenger risk taking attitudes under transit network uncertainties, and (c) investigate the transit line scheduling problem under different market regimes with uncertainties. The key findings are summarized in Section 6.2.

6.2. KEY FINDINGS

6.2.1 The Sources of Uncertainties

Transit network uncertainties have been determined from passenger demand and transit supply sides, as described in Chapter 2. The generation and influence of uncertainties in transit assignment and line schedule design models were found to differ. For the proposed RSUE transit assignment models, the impacts of uncertainties were mainly based on passenger random travel time (or travel cost/disutility) and random transit vehicle operation time. Uncertainty of vehicle dispatching schedules was another main cause of the degraded transit service or risk of operation profit, as examined by the transit line schedule design model in Chapter 5.

In the proposed RSUE transit assignment models, the given and fixed stochastic properties of transit service and passenger demand patterns were specified as exogenous uncertainties. The deduced stochastic phenomena were endogenous uncertainties, which included passenger waiting time and vehicle dwell time. They were the result of the interaction between the random arrivals of passengers and transit vehicles. The proposed PAB (passenger arriving and boarding) process described such interaction completely and derived the analytical expressions of these endogenous uncertainties. The PAB process duration length was indicative of the prolonged vehicle dwell time and vehicle bunching problem at stops and further low passenger loading proficiency.

The randomness of passenger in-vehicle time and waiting time decreased in the transit network with multi-class risk preferences passengers compared to that with single-class risk preference passengers. Such decrease, however incurred high monetary cost of risk-averse passengers, as to ensure travel time reliability. Those risk-averse passengers may change to other expensive lines or modes with higher reliability services. Transit network uncertainties imposed higher monetary cost on risk-averse passengers. It was found in the multi-class assignment results that the peak period of passenger waiting time at stops was more decentralized and later than that shown in the single-class assignment, as the passenger risk-taking attitudes would greatly impact passenger route and departure time choices.

The usage of vehicle design capacity could be improved by introducing an overload parameter in the RSUE transit assignment models to take account the effects of the stochastic passenger boarding demand. Otherwise, in congested transit networks, the practical number of passengers on-board would be underestimated and lead to severe in-vehicle congestion and passenger over-load delays at stops.

6.2.2 Passenger Behavior Responses

Two novel RSUE transit assignment models, predicting passenger travel decisions, such as travel route and departure time choices given the pre-specified transit service configurations, were proposed in Chapters 2 and 3. Neglecting transit network uncertainties resulted in biased estimation of passenger flow on transit lines or inaccurate estimation of passenger departure during peak periods. Under uncertainties, passengers included additional safety margins in their travel decision (and journey plan) to avoid network uncertainties. Passengers with travel time reliability requirements tended to depart earlier than those without, to ensure on-time arrivals and accommodate for unexpected delays. The downstream passengers shifted from long lines to short lines because of the long-line aggregated randomness. The adverse effects of such estimates can be poor level of transit service and shifts of passenger travel demand to other transport modes.

The risk-averse passengers (confidence level $\alpha > 0.5$) chose the more expensive routes than other risk-neutral ($\alpha = 0.5$) and risk-prone ($\alpha < 0.5$) passengers for the sake of travel time reliability. The extra monetary cost to ensure travel time reliability during peak period was an unfair penalty imposed on risk-averse passengers. The overload congestion at peak periods in the multi-class RSUE model was less severe and the peak period was longer than that in the single-class RSUE model. Passenger travel decisions with various risk preferences were distinguished by different travel time safety margins for different reliability requirements.

In the regulated transit market where transit system was operated by government authorities, passenger behavior responses to the improved transit line schedules caused a more discretized passenger flow distribution and therefore saved passenger waiting time and maintained transit service reliability. In the deregulated transit market fully explored in Chapter 5, it was found that passengers change their transit mode choices while bus and rail operators have different risk preferences and relevant optimized transit line schedules.

6.2.3 Operator Planning Strategies

Two novel transit bi-level transit line scheduling models were proposed under different transit market regimes. The transit assignment models proposed in Chapters 2 and 3 were applied as sub-models in the estimation of passenger behavior responses to alternative transit line schedules. Transit system performance such as service efficiency and reliability could be improved in transit market regulated by government authorities. The total passengers' generalized travel cost was reduced and the fleet size was also balanced among different lines without the extra cost of fleet resources. The passenger loading profile was also more balanced, which implicated a better level of transit service.

The Nash equilibrium between operators was rail operator risk-neutral and bus operator risk-averse in the deregulated transit market running by private bus and rail operators. Both operators' average profits under market equilibrium were underestimated and their profit variations were overestimated if their risk preferences were not considered. This is because they tended to postpone line schedules by a few minutes to maintain the reliable availability of transit vehicles. Rail operators had a high possibility of profit loss in un-congested transit networks, as the stochasticity of profit, particularly when influenced by passenger patronage, was very high. Thus, the rail operator was recommended to propose a decrease of frequency in the uncongested transit network.

6.3 FUTURE RESEARCH

Based on this research study, several areas may merit further study:

1) Extending the proposed model with consideration of elastic demand so as to reflect passenger demand response to transit service improvements which are based on the result of transit line schedule optimization under network uncertainties;

2) Exploring efficient reliability-based transit path finding algorithm and applying this algorithm to the transit assignment and network design models under uncertainties so as to enhance the scope of model applications in practice;
3) carrying out sensitivity analyses to investigate transit operation tactics such as short-turning, stop-skipping, and vehicle holding to promote optimal operation strategies;
4) calibrating stochastic parameters, such as boarding time per passenger, passenger arrival and OD demand variations;
5) incorporating passenger travel strategies into the dynamic reliability-based transit assignment model;
6) designing more effective algorithms to guarantee optimality for the transit network design models;
7) investigating the integration of cooperative and non-cooperative competition between transit operators under network uncertainties;
8) incorporating technological advantages into transit network design and operation, to enable the operator to deal with unexpected schedule variations.

REFERENCES

Adamski, A. (1992) Probabilistic models of passengers service processes at bus stops. *Transportation Research*, Vol. 26B, pp. 253-259.

Baaj, M. H. and Mahmassani, H. (1991) An AI-based approach for transit route system planning and design. *Journal of Advance Transportation*, Vol. 25, pp. 187–210.

Beimborn, E.A., Greenwald, M.J., and Jin, X. (2003) Accessibility, connectivity, and captivity: impacts on transit choice. *Transportation Research Record*, Vol. 1835, pp. 1–9.

Bell M.G.H., Schmöcker J.D., Iida Y., and Lam W.H.K. (2002) Transit Network Reliability: An Application of Absorbing Markov Chains. In: Taylor MAP (ed), *Proceedings of the 15th international symposium on transportation and traffic theory.* Elsevier Science, Amsterdam, pp. 43–62

Black, J. (1981) Urban transport planning: theory and practice. *Taylor & Francis*.

Carey, M. (1998) Optimizing scheduled times, allowing for behavioral response. *Transportation Research*, Vol. 32B, pp. 329-342.

Ceder, A., Colany, B., and Tal, O. (2001) Creating bus timetable with maximal synchronization. *Transportation Research*, Vol. 35A, pp. 913–928.

Ceder, A. (2003) Public transportation timetabling and vehicle scheduling. *Advanced Modeling for Transit Operations and Service Planning*. William, H.K.L, and Michael, G.H.B. Pergamon.

Ceder, A. (2007) Public Transit Planning and Operation: Theory, Modeling and Practice. Butterworth Heinemann, Elsevier.

Chen A., Yang, H., Lo H.K. and Tang W.H. (1999) A capacity related reliability for transportation networks. *Journal of Advanced Transportation,* Vol. 33, pp. 183-200.

Chen, A., Ji, Z., and Recker, W. (2002) Travel time reliability with risk-sensitive travelers. *Transportation Research Board*, Vol. 1873, pp. 27-33.

Chen, A., Subprasom, K., and Ji, Z.G. (2003) Mean-variance model for the build-operate-transfer scheme under demand uncertainty. *Transportation Research Record*. Vol. 1857, pp. 93-101.

Chen, A., Kim, J., Zhou, Z., and Chootinan, P. (2007) Alpha reliable network design problem. *Transportation Research Record*, Vol. 2029, pp. 49-57.

Chootinan, P., Wong, S.C. and Chen, A. (2005) A reliability-based network design problem. *Journal of Advanced Transportation*, Vol. 39, pp. 247-270.

Cominetti, R. and Correa, J. (2001) Common-lines and passenger assignment in congested transit networks. *Transportation Science,* Vol. 35, pp. 250-267.

Dafermos, S.C. (1982) The general multimodal network equilibrium problem with elastic demand. *Networks*, Vol.12, pp.57–72.

De Cea, J. and Fernandez, E. (1993) Transit assignment for congested public transport system: An equilibrium model. *Transportation Science,* Vol. 27, pp. 133–147.

Ding, Y.Q. and Chien, S.I. (2001) Improving transit service quality and headway regularity with real-time control. *Transportation Research Record,* Vol. 2150, pp. 161-170.

Fan, W. and Machemehl, R.B. (2006) Optimal transit route network design problem with variable transit demand: Genetic algorithm approach. *Journal of Transportation Engineering*, Vol. 132, pp. 40-51.

Fernandez, J.E., Cea Ch, J., and Malbran, R.H. (2008) Demand responsive urban public transport system design: Methodology and application. *Transportation Research*, Vol. 42A, pp. 951-972.

Fleurrent, C., Lessard, R., Seguin, L., and Inc, G. (2005) Transit timetable synchronization: evaluation and optimization. *9th International Conference on Computer Aided Scheduling in Public Transportation.*

Friesz, T.L. and Harker, P.T. (1983) Multicriteria spatial price equilibrium network design: theory and computational results. *Transportation Research*, Vol. 17B, pp. 41 l-426.

Friesz, T.L., Harker, P.T., and Tobin, R.L. (1984) Alternative algorithms for the general network spatial equilibrium problem. *Journal of Regional Science*, Vol. 24, pp. 475-507.

Fu, L., Liu, Q., and Calamai, P. (2003) Real-time optimization model for dynamic scheduling of transit operations. *Transportation Research Record*, Vol. 1857, pp. 48-55.

Furth, P.G. and Wilson, N. H. M. (1981). Setting frequencies on bus routes: Theory and practice. *Transportation Research Record,* Vol. 818, pp. 1-7.

Hamdouch, Y., and Lawphongpanich, S. (2008) Schedule-based transit assignment model with travel strategies and capacity constraints. *Transportation Research*, Vol. 42B, pp. 663-684.

Harker, P.T. (1984) A variational inequality approach for the determination of oligopolistic market equilibrium. *Mathematical Programming*, Vol. 30, pp. 105-111.

Harker, P.T. and Pang, J.S. (1990) Finite-dimensional variational inequality and nonlinear complimentarily problems: A survey of theory, algorithms and applications. *Mathematical Programming*, Vol. 48, pp. 161-220.

Harker, P.T. (1991) Generalized Nash games and quasi-variational inequalities. *European Journal of Operational Research*, Vol. 54, pp. 81-94.

Hickman, M.D. (2001). An analytic stochastic model for the transit vehicle holding problem. *Transportation Science,* Vol. 35, pp. 215-237.

Higgins, A., Kozan, E., and Ferreira, L. (1996) Optimal scheduling of trains on a single line track. *Transportation Research*, Vol. 30B, pp. 147–161.

Gasinski, L. and Papageorgiou, N.S. (2005) Nonlinear Analysis. *Chapman & Hall/CRC*.

Gao, Z.Y., Sun, H.J., and Shan, L.L. (2004) A continuous equilibrium network design model and algorithm for transit systems. *Transportation Research,* Vol. 38B, pp. 235-250.

Ghoseiri, K., Szidarovszky, F. and Asgharpour, M.J. (2004) A multi-objective train scheduling model and solution. *Transportation Research,* Vol. 38B, pp.

927-952.

Gomez-Lobo, A. (2007) Why competition does not work in urban bus markets: Some new wheels for some old ideas. *Journal of Transport Economics and Policy*, Vol. 41, pp. 283-308.

Goldberd, D.E (1989) Genetic Algorithms in Search, Optimization and Machine Learning, Addison-Wesley, New York.

Guihaire, V. and Hao, J.K. (2008) Transit network design and scheduling: A global review. *Transportation Research*, Vol. 42B, pp. 1251-1273.

Kidwai, F.A., Marwah, B.R., Deb, K., and Karim, M.R. (2005) A genetic algorithm based bus scheduling model for transit network. *Proceedings of the Eastern Asia Society for Transportation Studies,* Vol. 5, pp. 477-489.

Kurauchi F., Bell, M.G.H., and Schmöcker, J.D. (2004) Capacity constrained transit assignment with common lines. *Journal of Mathematical Modelling and Algorithms*, Vol. 2, pp. 309–327.

Lam, W.H.K., Cheung, C.Y. and Poon, Y.F. (1998) A study of train dwelling time at the Hong Kong Mass Transit Railway system. *Journal of Advanced Transportation*, Vol. 32, pp. 285-296.

Lam, W. H. K., Gao, Z. Y., Chan, K. S., and Yang, H. (1999) A stochastic user equilibrium assignment model for congested transit networks. *Transportation Research,* Vol. 33B, pp. 1-18.

Lam, W. H. K., Shao, H., and Sumalee, A. (2008) Modeling impacts of adverse weather conditions on a road network with uncertainties in demand and supply. *Transportation Research*, Vol. 42B, pp. 890-910.

Larrain, H. and Muñoz, J.C. (2008) Corridor assignment assuming congestion due to passenger boarding and alighting. *Network Spatial Economics*, Vol. 8, pp. 241-256.

Last, A. and Leak, S.E. (1976) TRANSEPT: a bus model. *Traffic Engineering and Control*, Vol. 29, pp. 520–526.

Lee, Y.J. and Vuchic, V.R. (2005) Transit network design with variable demand. *Journal of Transportation Engineering,* Vol. 131, pp. 1-10.

Li, Z.C., Lam, W.H.K., and Wong, S.C. (2009) The optimal transit fare structure under different market regimes with uncertainty in the network. *Networks and Spatial Economics*, Vol. 9, pp. 191-216.

Lo, H., Luo, X.W., and Siu, B.W.Y. (2006) Degradable transport network: travel time budget of travelers with heterogeneous risk aversion. *Transportation Research*, Vol. 40B, pp. 792-806.

Medhi, J. (2003) Stochastic Models in Queuing Theory, Second Edition. Academic Press, Elsevier Science.

Murray, A.T, Davis, R., Stimson, R.J., and Ferreira, L. (1998) Public transportation access. *Transportation Research,* Vol. 3D, pp. 319-328.

Noland, R.B. and Polak, J.W. (2002) Travel time variability: A review of theoretical and empirical issues. *Transport Reviews*, Vol. 22, pp. 39-54.

Nuzzolo,A., Russo, F., and Crisalli, U. (2001) A doubly dynamic schedule-based assignment model for transit networks. *Transportation Science*, Vol. 35, pp.

268-285.

Nuzzolo, A., Russo, F. and Crisalli U. (2003) Transit network modeling: The schedule-based dynamic approach. Franco Angeli s.r.l., Milana, Italy.

Pattnaik, S.B., Mohan, S., and Tom, V.M. (1998) Urban bus transit route network design using genetic algorithm, *Journal of Transportation Engineering*, Vol. 124, pp. 368-375.

Poon, M.H., Wong, S.C., and Tong, C.O. (2004) A dynamic schedule-based model for congested transit networks. *Transportation Research,* Vol. 38B, pp. 343-368.

Powell, W. (1982) Analysis of airline operating strategies under stochastic demand. *Transportation Research*, Vol. 16B, pp. 31-43.

Powell, W.B., and Sheffi, Y. (1983) A probabilistic model of bus route performance. *Transportation Science,* Vol. 17, pp. 376-404.

Powell, W. and Winston, C. (1983) A numerical investigation of the impact of uncertain demand and varying risk preferences on the pricing and capacity decisions of transportation firms: The case of airlines. *Transportation Research,* Vol. 17B, pp. 471-490.

Schmöcher, J.D., Bell, M.G.H., and Kurauchi, F. (2008) A quasi-dynamic capacity constrained frequency-based transit assignment model. *Transportation Research*, Vol. 42B, pp. 925-945.

Sheffi, Y. (1985) Urban Transport Networks: Equilibrium Analysis with Mathematical Programming Methods. *Prentice-Hall Inc.*

Shrivastava, P. and O'Mahony, M. (2006) A model for development of optimized feeder routes and coordinated schedules - A genetic algorithms approach. *Transport Policy*, Vol. 13, pp. 413-425.

Spiess, H. and Florian, M. (1989) Optimal strategies: A new assignment model for transit networks. *Transportation Research*, Vol. 23B, pp. 83-102.

Sumalee A., Watling, D., and Nakayama, S. (2006) Reliable network design problem - case with uncertain demand and total travel time reliability. *Network Modelling*, Vol. 1964, pp. 81-90.

Szeto, W.Y., Solayappan, M., and Wong, K.I. (2009) Reliability-based stochastic transit assignment with capacity constraints. *CD-ROM of 11th International Conference on Advanced Systems for Public Transport*, Hong Kong SAR.

Teklu, F., Watling, D. and Connors, R. (2007) A Markov Process model for capacity constrained transit assignment. *Transportation and Traffic Theory*, pp. 483-506.

Teodorovic, D., Kalic, M., and Pavkovic, G. (1994) The potential for using fuzzy set theory in airline network design. *Transportation Research*, Vol. 28B, pp. 103-121.

Uchida, K., Sumalee, A., Watling, D., and Connors, R. (2007) A study on network design problems for multi-modal networks by probit-based stochastic user equilibrium. *Network Spatial Economy*, Vol. 7, pp. 213–240.

Wilson, N. and Nuzzolo, A. (2004) Schedule-Based Dynamic Transit Modeling - Theory and Application. Kluwer Academic Publishers.

Wirasinghe, S.C. (2003) Initial planning for urban transit systems. *Advanced Modeling for Transit Operations and Services Planning*, *Elsevier Science*.

Yan, S. and Chen, H.L. (2002) A scheduling model and a solution algorithm for inter-city bus carriers. *Transportation Research,* Vol. 36A, pp. 805–825.

Yan, S.Y. and Chen, H.L., and Tang, C.H. (2006) Inter-city bus routing and timetable setting under stochastic demands. *Transportation Research,* Vol. 40A, pp. 572-586.

Yan, S.Y. and Tang, C.H. (2008) An integrated framework for intercity bus scheduling under stochastic bus travel times. *Transportation Science,* Vol. 42, pp. 318-335.

Yang, L. and Lam, W.H.K. (2006) Probit-based reliability-based transit assignment model. *Transportation Research Record*, Vol. 1977, pp. 154-163.

Yin, Y., and Ieda, H. (2001) Assessing performance reliability of road networks under nonrecurrent congestion. *Transportation Research Record*, Vol. 1771, pp. 148-155.

Zhou, J. and Lam, W.H.K. (2001) A bi-level programming approach- optimal transit fare under line capacity constraints. *Journal of Advanced Transportation*, Vol. 35, pp. 105-124.

Zhao, J., Dessouky, M. and Bukkapatnam, S. (2006) Optimal slack time for schedule-based transit operations. *Transportation Science*. Vol. 40, pp. 529-539.

Zhao, F. and Zeng, X.G. (2008) Optimization of transit route network, vehicle, headways and timetables for large-scale transit networks. *European Journal of Operational Research*, Vol.186, pp.841-855.

Zuckerman, D. and Tapiero, C.S. (1980) Random vehicle dispatching with options and optimal fleet size. *Transportation Research*, Vol. 14B, pp. 361–368.