Optimal bus transit system considering service network design and route packaging

Nayan, Ashish

2017

Nayan, A. (2017). Optimal bus transit system considering service network design and route packaging. Doctoral thesis, Nanyang Technological University, Singapore.

<http://hdl.handle.net/10356/69546>

<https://doi.org/10.32657/10356/69546>

NANYANG
TECHNOLOGICAL
UNIVERSITY

**OPTIMAL BUS TRANSIT SYSTEM CONSIDERING
SERVICE NETWORK DESIGN AND ROUTE
PACKAGING**

ASHISH NAYAN

SCHOOL OF CIVIL AND ENVIRONMENTAL ENGINEERING

2016

**OPTIMAL BUS TRANSIT SYSTEM CONSIDERING
SERVICE NETWORK DESIGN AND ROUTE
PACKAGING**

ASHISH NAYAN

School of Civil and Environmental Engineering

A thesis submitted to the Nanyang Technological University
in partial fulfilment of the requirement for the degree of
Doctor of Philosophy

2016

Dedicated to my beloved family

“कर्मण्येवाधिकारस्ते मा फलेषु कदाचन”

ACKNOWLEDGEMENTS

I take this opportunity to thank my supervisor, Dr. David Z.W. Wang, Assistant Professor, School of Civil & Environmental Engg., NTU for his continuous support, guidance, patience and belief in me. Without his encouragement and patient supervision, this would not have been possible.

I sincerely thank the distinguished members of my doctoral committee for providing me with in-depth insights into the research topic during the Qualifying Examination. I thank my research colleagues and friends, especially Dr. Liu Haoxiang and Dr. Du Bo for their helpful suggestions and encouragement. I also extend my special gesture of appreciation to the administrative staff members at the Graduate Studies Unit of the School of CEE for their understanding and supportive gesture throughout my period of study. I also feel obliged to the technicians at the Geotechnics, Transport and Geospatial Laboratory for their hearty assistance and the Nanyang Technological University, Singapore for the financial support during my Doctoral candidature.

Last but not the least; I thank my beloved family which has been a constant source of support and encouragement throughout my life.

Ashish Nayan

TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	iii
TABLE OF CONTENTS.....	iv
SUMMARY.....	xiii
LIST OF PUBLICATIONS.....	xv
LIST OF TABLES.....	xvi
LIST OF FIGURES.....	xix
LIST OF ABBREVIATIONS.....	xxi
CHAPTER 1 INTRODUCTION.....	1
1.1 Research Background	1
1.2 Broad objectives.....	3
1.2.1 Specific objectives of the thesis.....	4
1.3 Scope of Research	7
1.4 Outline of the research work.....	8
CHAPTER 2 LITERATURE REVIEW.....	9
2.1 Introduction.....	9

2.2 An overview of important concepts.....	9
2.2.1 Transport modelling and Traffic Theory.....	9
2.2.2 Wardrop's Principles.....	9
2.2.3 Network Design Problems.....	11
2.3 Different mathematical modelling techniques used for Transit Network Design Problems.....	11
2.3.1 Single-level programming.....	11
2.3.2 Linear single-level programming.....	12
2.3.3 Nonlinear single-level programming.....	13
2.3.4 Bi-level programming	13
2.4 Classic solution methods for Transit Network Design Problems.....	15
2.4.1 Solution methods for linear programming.....	15
2.4.2 Solution methods for nonlinear programming.....	16
2.5 Limited stop services.....	16
2.6 Congestion based bus transit network design with common lines.....	22
2.7 Bus Transit Route Packaging.....	24

2.8 Competitive Tendering of Bus Transit Route Packages.....	28
 CHAPTER 3 BUS TRANSIT NETWORK DESIGN WITH LIMITED STOP SERVICES.....	 33
3.1 Introduction.....	33
3.2 Design of limited stop service using DUE	34
3.2.1 Network setting.....	34
3.2.2 Variable definitions	35
3.2.3. Assumptions.....	37
3.2.3.1. Passenger demand	37
3.2.3.2. In-vehicle travel time.....	37
3.2.3.3. Choice behaviour.....	37
3.2.3.4. Total costs.....	38
3.2.3.5. Certain definitions related to transit services.....	38
3.2.4. General mathematical constraint for a limited stop service.....	38
3.2.5. Model formulation for the design of limited stop services.....	40
3.2.5.1. Problem Description	40
3.2.5.2. Model Formulation	41

3.2.5.3. Incorporating the effects of alighting and boarding	45
3.2.6. Solution Method.....	45
3.2.6.1. Linearization with Reformulation-Linearization Technique (RLT).....	46
3.2.6.2. Linear transformation to linearize a product of binary variables.....	47
3.2.6.3. Multidimensional piecewise linearization method	47
3.2.6.4. Treatment for logic constraints.....	51
3.2.6.5. Treatment for the travel time functions incorporating the effects of alighting and boarding.....	52
3.2.7. Reformulated problem.....	53
3.2.8. Additional operational constraints.....	53
3.2.9. Numerical studies for limited stop service design using deterministic user equilibrium.....	54
3.2.9.1. General Parameters.....	55
3.2.9.2. Optimization Results.....	56
3.2.9.2.1. Numerical example with one limited stop service.....	56
3.2.9.2.2. Numerical example with two limited stop services.....	57
3.2.10. Further discussion.....	58

3.3 Design of limited stop service using SUE.....	59
3.3.1. Problem description.....	59
3.3.2. Model Formulation.....	60
3.3.3. Solution Methodology.....	62
3.3.3.1. Piecewise Linearization Techniques.....	63
3.3.4. When transfers are allowed in the SUE model formulation.....	67
3.3.5. Numerical study for limit stop service design using stochastic user equilibrium.....	70
3.3.5.1. Parameters.....	71
3.3.5.2. Optimization Results.....	71
3.3.5.3. Further analysis.....	72
3.4. Conclusion.....	74
 CHAPTER 4 A CONGESTION BASED BUS TRANSIT NETWORK DESIGN WITH COMMON LINES.....	 76
4.1. Introduction and Problem description.....	76
4.2. Methodology.....	77
4.2.1. Notations.....	77

4.2.2. Model Formulation.....	79
4.2.3. Solution Algorithm.....	84
4.2.3.1. Reformulation.....	84
4.2.3.2. One dimensional piecewise linearization.....	85
4.2.3.3. Multi-dimensional piecewise linearization.....	86
4.3. Numerical Study.....	89
4.3.1. Transit Network	89
4.3.2. Optimization results.....	90
4.3.3. Further analysis.....	90
4.4. Conclusion.....	93
CHAPTER 5 A METHODOLOGY FOR TRANSIT ROUTE PACKAGING IN A PRIVATISED CONTRACTING REGIME.....	94
5.1. Introduction and Problem Description.....	94
5.2. Methodology.....	95
5.2.1 Assumptions.....	95
5.2.2. Variable Definitions.....	96

5.2.3. A few important rationale.....	97
5.2.4. Model Formulation.....	98
5.2.5. Solution Methodology.....	100
5.2.5.1. Reformulation linearization technique (RLT).....	101
5.2.5.2. One-dimensional piecewise linearization.....	101
5.2.5.3. Treatment for the absolute value function term in the objective function	102
5.3. Numerical study and discussion.....	103
5.3.1. Parameters.....	104
5.3.2. Optimization results.....	106
5.3.3. Further analysis.....	107
5.3.3.1. Route package design on an expanded transit network.....	109
5.4. Conclusion.....	111
 CHAPTER 6 A METHODOLOGY FOR COMPETITIVE TENDERING OF BUS TRANSIT ROUTE PACKAGES IN A PRIVATISED CONTRACTING REGIME.....	 112
6.1. Introduction and Problem description.....	112
6.2. Methodology.....	114

6.2.1. Assumptions.....	114
6.2.2. Variable definitions.....	115
6.2.3. Model Formulation.....	116
6.2.4. Solution Algorithm.....	119
6.2.4.1. One-dimensional piecewise linearization.....	120
6.2.4.2. Reformulation linearization technique (RLT).....	121
6.2.4.3. Treatment for the absolute value function term in the lower level objective function	122
6.2.4.4. Reformulation using KKT conditions.....	122
6.3. Numerical study.....	123
6.3.1. Parameters.....	124
6.3.2. Optimization Results.....	126
6.3.3 Further analysis.....	128
6.4. Conclusion.....	129
CHAPTER 7 CONCLUSION AND RECOMMENDATIONS.....	132
7.1 Thesis summary.....	132

7.2 Recommendations for future work.....	132
REFERENCES.....	134

SUMMARY

Bus transportation is one of the most important means of public transportation in today's world. All major transit regulators intend to improve the bus operation such that the system functions sustainably with respect to all stakeholders involved viz. authority, operators and the commuters. On a day-to-day basis, several improvements are implemented using consumer satisfaction indices or parameters that are critical and need immediate attention. However, a systematic bus transit system design requires an analytical framework for rational decision making such that the transit network serves the maximum number of commuters at their desired level of service while being financially sustainable for the operators and the regulator. In this study, the mathematical modelling technique is employed to propose strategies for a financially and socially sustainable bus transportation system using the mathematical programming (optimization) approach. A set of problems addressing two important aspects of bus transit operation viz. service/network design and regulatory framework design involving different stakeholders are discussed which represent a diverse range of analytical problems faced by the bus transit industry today.

In the first section of the thesis, problems relating to transit network design are discussed which act as analytical interfaces between the operators and the commuters; helping the operator to come up with optimal strategies in terms of service network design and tackle issues of congestion. Specifically, a comprehensive design of limited stop service is proposed using the deterministic and stochastic user equilibrium. Further, a bus transit network design problem is proposed to include congestion effects and derive optimal strategies of travel on common lines out of a set of lines serving the transit network. In the second section of the thesis, problems relating to transit network and regulatory framework design of bus transit are discussed with respect to the government contracting regime. A few major bus transit friendly cities recently adopted this regime wherein the government/authority owns all infrastructure and tenders out transit route packages (sets of routes) to private transit players through competitive tendering only for routine operation and maintenance. This thesis presents a detailed mathematical

modelling framework of the government contracting regime based on two aspects (i) bus transit route packaging and (ii) competitive tendering of transit route packages to contesting operators.

For each of the mathematical models developed, detailed solution algorithms are developed in order to reformulate, linearize and subsequently convexify the formulation to solve it to global optimality. Various mathematical techniques including linearization for the product of binary variables, reformulation-linearization technique, single and multi-dimensional piecewise linearization method and linear piece-wise approximation method are used in the thesis to propose the solution methodology. Thereafter, the mathematical models are validated on real-time transit networks, and results obtained are discussed with some more discussions.

To conclude, all the model formulations developed in this thesis present a comprehensive decision making framework for a sustainable bus transit system in the long run. Addressing transit network design and the transit regulatory /policy issues, the proposed framework of analytical models in totality would certainly provide transit planners with a good insight into equitable methods of transit operation to promote social welfare.

LIST OF PUBLICATIONS

1) Optimal Bus Transit Route Packaging in a Privatized Contracting Regime

Status: Published

Journal=> Transportation Research Part A: Policy and Practice

2) Bus Transit Network Design with Limited Stop Services

Status: Submitted and under review

Journal=> Transportation Research Part B: Methodological

LIST OF TABLES

Table 2.1 Comparison between the published studies on limited stop service design and this work	20
Table 3.1 Route sections and operating lines	41
Table 3.2 Intrinsic demand at each node for destination node 6 per hour.....	55
Table 3.3 Intrinsic demand at each node for destination node 10 per hour.....	55
Table 3.4 Total operating cost for corresponding candidate transit line setting of the limited stop service l_1	59
Table 3.5 Optimal values of decision variables with changed demand	72
Table 3.6 Optimal values of decision variables with changed fares.....	72
Table 4.1 Route characteristics.....	89
Table 4.2 Route section travel time (mins) on different lines.....	90
Table 4.3 Equilibrium route section flows.....	90
Table 4.4 Sensitivity analysis with respect to demand.....	90
Table 5.1 Network characteristics.....	104
Table 5.2 Route demand (D)/hr. during peak hour (number of commuters) X^r .	105
Table 5.3 Hourly rate of operating buses at a required frequency on route r	

included in package p (F_p^r).....	105
Table 5.4 Hourly rate of maintaining a bus on route r included in package p (B_p^r).....	105
Table 5.5 Hourly rate of bus ownership on route r in package p (O_p^r).....	105
Table 5.6 Travel time on route r in package p including hauling time T_p^r (mins).....	106
Table 5.7 Binary variable for allocation of route r to package p	106
Table 5.8 Fleet size allocated to route r in package p	106
Table 5.9 Frequency of route r in package p in buses/hr.....	106
Table 5.10 Sensitivity analysis with respect to demand.....	107
Table 6.1 Network characteristics.....	124
Table 6.2 Operators competing for different transit packages.....	124
Table 6.3 Route demand/hr during peak hour (number of passengers)	124
Table 6.4 Total fare revenue from packages (\$) per hour during operational hour.....	125
Table 6.5 Cycle time of a route r in package p in minutes.....	125
Table 6.6 Ownership cost/hr (\$) of a bus allocated to a route r in package p	125
Table 6.7 Rate for operating a bus at the desired frequency on route r in	

package p for operator $M1$	125
Table 6.8 Rate for operating a bus at the desired frequency on route r in package p for operator $M2$	126
Table 6.9 Rate for operating a bus at the desired frequency on route r in package p for operator $M3$	126
Table 6.10 Rate for operating a bus at the desired frequency on route r in package p for operator $M4$	126
Table 6.11 Binary variable for allocation of package p to operator m	127
Table 6.12 Fleet size allocated to route r in package p	127
Table 6.13 Frequency of route r in package p in buses/hr	127
Table 6.14 Optimal service fee/hr quoted by an operator m' to operate a transit package p in \$.....	127
Table 6.15 Analysis of regulator cost/hr and packages' optimal fees/hr for different demand patterns.....	128

LIST OF FIGURES

Fig. 1.1 Specific objectives of the thesis.....	7
Fig. 3.1 Transit network.....	34
Fig. 3.2 Possible direct limited stop services between different node pairs in the transit network.....	39
Fig. 3.3 Example transit network with line setting for limited stop service.....	40
Fig. 3.4 Transit network for numerical study.....	54
Fig. 3.5 Optimized line setting for the limited stop service l_1	57
Fig. 3.6 Optimized line setting for the limited stop service l_1	57
Fig. 3.7 Optimized line setting for the limited stop service l_2	58
Fig. 3.8 Optimized line setting for the limited stop service l_1	72
Fig. 3.9 Total cost with different demand and fare for the SUE model.....	73
Fig 4.1 A transit network route section A-B connected by transit lines.....	76
Fig 4.2 Transit network.....	82
Fig 4.3 Route section travel strategy time.....	83
Fig 4.4 Transit Network.....	90
Fig 4.5 Plot of total op. cost vs demand.....	91
Fig 4.6 Percentage increase in total op. cost vs. increase in demand.....	91
Fig 4.7 Plot of equilibrium route section frequency vs. demand.....	92

Fig.5.1 Transit network with numbered routes and bus bays denoted as $P1$, $P2$, $P3$	104
Fig 5.2 Packages' operating cost for different demands.....	108
Fig. 5.3 Plot of regulator costs vs. total packages' operating costs for different demands.....	109
Fig.6.1. Transit network with routes numbered and bus bays denoted as $P1$, $P2$, $P3$	124
Fig.6.2. Packages' optimal fees/hr for different demands.....	128
Fig.6.3 Plot of regulator costs/hr vs. total packages' service fees/hr for different demands.....	128

LIST OF ABBREVIATIONS

CNDP	Continuous network design problem
CT	Competitive Tendering
CTNDP	Continuous transit network design problem
DNDP	Discrete network design problem
DTNDP	Discrete transit network design problem
DUE	Deterministic user equilibrium
KKT	Karush-Kuhn-Tucker
MNDP	Mixed network design problem
MTNDP	Mixed transit network design problem
MILP	Mixed-integer linear programming
MINLP	Mixed-integer nonlinear programming
MPEC	Mathematical Programming with equilibrium constraints
NDP	Network design problem
NP-hard	Non-deterministic polynomial-time hard
O-D	Origin-destination
RLT	Reformulation-linearization technique
SO	System optimum
SUE	Stochastic user equilibrium
VI	Variational inequality

1.1 Research Background

Public transportation is one of the most important aspects of our daily lives. In today's scenario where the world faces some really challenging issues such as that of environmental sustainability and growing demand for travel due to increasing population, the public transportation sector has much to contribute in shaping our future. Public transportation offers several benefits such as reduced use of natural resources, controlled infrastructure, more accessibility to pedestrians, higher carrying capacity and a cheaper means of transport while doing away with congestion, urban space degradation, air and noise pollution and the inefficient use of space and energy. Several governments are trying to match the standard of public transportation modes to that of the private car in terms of level of service and travel time so that more people shift to the former choice. But since bus and train services are very large scale systems designed to cater huge demands, it is difficult to match the commuters' desired service levels which often results into lower ridership especially when the services are not well designed, infrastructure is not sufficiently developed, frequent fare hikes and lack of trust in the system due to frequent breakdowns.

For increasing the attractiveness of transit systems to car users, there is a constant need for strategies to improve the level of service and ensure that the system is financially and socially sustainable. Improvements in the transit network topology and operational management mechanisms are crucial towards revamping the transit situation where the goal is to provide a high performance network composed of robust structures and integrated services. Along with the day-to-day improvements based on survey and operation requirements, the operators and government/authority need to base their decisions with respect to service/network design or regulatory design on a comprehensive, analytical framework that can aptly model the system and propose sustainable strategies. In this thesis, a comprehensive mathematical modelling based analytical framework encompassing transit network design and transit regulatory framework design is proposed in order

to provide a decision making tool aimed towards improvement of the state of bus transit.

Transportation research has often relied on rigorous mathematical techniques to propose decision making methodologies for problems relating to improvement in infrastructure or attaining financial objectives in system operation. Wardrop (1952) first proposed the user equilibrium principles of transportation network and this work laid the foundation for several ground breaking advancements in this area. Several other works such as Kinderlehrer and Stampacchia (2000) then proposed mathematical techniques to model equilibrium and efficiently capture the passengers' choice behaviour to design high performance transit networks. Generally, improvement in bus transit services is either through capacity expansion or by designing the network so as to optimally utilize the existing infrastructure in order to accommodate a growing travel demand. These issues are extensively covered in a large research area called as the Network Design Problem (NDP). A NDP when extended to solve transit systems specifically is known as a Transit Network Design Problem (TNDP). The concept of NDP owes its roots to the basic traffic modelling paradigm comprising four stages called as trip generation, trip distribution, modal split and traffic assignment. With a given demand over the network, TNDP is generally used to design the transit service and network, model the travel demand through modal choice analysis and transit assignment through routing choice analysis. It can also be extended to solve problems with mixed objectives relating to network design and regulatory decisions as ultimately, the analytical technique is being utilized to propose strategies to improve the state of the transit system from an operational or a financial perspective.

NDP has emerged as an advanced, scientific area for progress in handling effective transport planning because of the fast growing travel demand on roads that pose a great challenge to the existing capacity of our urban transport systems and limited available resources to expand the system capacity. Historically, this problem has been presented in two different forms: a discrete form dealing with the addition of new links to an existing road network called as the discrete network design problem (DNDP), and a continuous form dealing with the expansion of optimal capacity of

existing links called as the continuous network design problem (CNDP). A combination of DNDP and CNDP is called as a Mixed Transit Network Design Problem (MNDP). In any form, the objective of the TNDP is to optimize a given transit system performance such as to minimize the total system cost in the best interest of the involved stakeholders. The study by Ceder and Wilson (1986) clearly defines the elements of transit network design as part of the overall operational planning process and was one of the first works in this area of research.

1.2 Broad objectives

This research aims to address the important issue of improving the state of bus transportation in cities so that more people are attracted towards public transportation in general. As mentioned earlier, this can be brought about through various ways such as infrastructure expansion, efficient network design, better contractual regimes etc. Infrastructure expansion generally requires more space and is costlier (Mesbah et al., 2011) especially in congested cities today. Hence, improvement in bus transportation through more efficient network design and better contractual regimes such that the system operates in equity with respect to all involved stakeholders is a more sustainable method to address this issue. In this thesis, a mathematical modelling based comprehensive, analytical framework encompassing transit network design and optimal transit regulatory framework (contracting strategies) incorporating users' choice behaviour is proposed. The selected problems based on the above two themes are formulated as TNDPs and defined through the optimization approach using mathematical programming in chapters 3, 4, 5 and 6 respectively. For each of these model formulations, solution methodologies using various mathematical and linearization techniques are proposed to be able to find the globally optimal solution which is truly the “best possible” solution. These model formulations are thereafter validated on transit networks using a commercial solver and optimal results are presented with certain insights. Hence, the main contribution of this thesis includes a comprehensive mathematical modelling paradigm to improve the overall state of bus transportation based on transit network design and transit contracting strategies.

To summarize, this thesis aims to achieve the following objectives:

- To develop mathematical models in order to aid transit network and regulatory framework design considering operators' cost criteria, passengers' choice behaviour and the authority's macro-economic financial objectives.
- Incorporate transit assignment by using advanced choice behaviour analysis.
- To propose solution methodologies for the formulated models through algorithms and mathematical techniques to attain global optimal solutions.
- Implementation of the mathematical models with Singaporean context based on local characteristics of bus transit.

1.2.1 Specific objectives of the thesis

Specifically, this thesis has the following objectives:

- I To propose a model formulation in the form of a mixed transit network design problem for the design of limited stop services operating in conjugation with the normal service using two different choice behavioural principles viz. deterministic user equilibrium and stochastic user equilibrium. The model formulation is basically a Mixed Integer Nonlinear Programming (MINLP) problem which is solved using various linearization techniques and solution algorithms to ensure a globally optimal solution. The model formulation is thereafter tested for its validity on a transit network using commercial software (YALMIP on MATLAB) and results are presented. The major contributions here include: an explicit design including transit line setting, fleet size, service frequency and a global optimization method to solve the model formulation.
- II To propose a Continuous Transit Network Design Problem (CTNDP) defined through an optimization approach in the form of a Mathematical Program with Equilibrium Constraints (MPEC) wherein service frequencies and fleet size of a set of transit lines operating in a transit network are designed. In this study, the concept of common lines is used while considering the congestion effects. In the MPEC formulation, the objective function is defined as the operator's problem and user equilibrium is

represented by the traditional Wardrop's first principle. The nonlinear, non-convex MPEC is subsequently reformulated using mixed integer conditions and a multidimensional linearization technique into a Mixed Integer Linear Programming (MILP) problem that guarantees global optimality using a routine solver. The model is then evaluated on a transit network to demonstrate its validity. The results determine the optimal line frequencies, fleet sizes and the equilibrium route section flows minimizing the operating and travel costs to the operator and the commuters respectively. This unique approach to design the transit network and attain equilibrium in the transit network using strategy flows due to the common lines operating over various route sections is the essence of this contribution.

- III To propose a Mixed Transit Network Design Problem (MTNDP) with respect to the transit regulator (or a government authority) for designing bus transit route packages to be tendered out to contesting operators through competitive tendering (CT). Routes are packaged taking into account the equity of all stakeholders in a public transportation system- the commuters, operators and the regulator. An optimal transit route packaging would lead to a financially sustainable operation in the long run, benefitting all involved entities. The problem is formulated as a Mixed Integer Nonlinear Programming (MINLP) problem, which is then transformed into Mixed Integer Linear Program (MILP) using linearization techniques so that global optimality of the solution could be guaranteed. A numerical study is then performed on a small section of the Singapore transit network to evaluate the model validity. The model results explicitly demonstrate the optimal packaging strategy, operating frequency of all bus services and the allocated fleet size to each package. This methodology provides a comprehensive decision making framework for the regulator contemplating to contract out bus route packages through CT while ensuring the attractiveness of the bus transit market to contesting operators and commuters' expected service levels.

IV To present a bi-level programming problem modelled in the form of a MTNDP for describing the competitive tendering (CT) process of transit route packages in the form of a Stackelberg game between the transit regulator (master) and the contesting operators (followers). The upper level is the regulator's problem whereas the lower level includes the operators' problem. The bi-level structure of the problem describes the interaction between the regulator and the operators in a CT exercise and attempts to find an optimal solution in terms of regulator's contracting strategies and operators' service fee such that the transit operation is financially sustainable. The regulator's decision variables mainly include route frequency, route fleet size and to decide whether to award a particular package to an operator whereas the decision variable of the operators is their service fee that they quote in the tender proposal. The model formulation, initially a Bi-Level Mixed Integer Nonlinear Programming Problem (MINLP) is subsequently transformed into a single level problem using complementary conditions and linearized using the reformulation-linearization technique. This transforms the problem into a Mixed Integer Linear Programming Problem (MILP) which ensures global optimality of the solution. Numerical studies are then conducted on a small section of the Singapore transit network to test the validity of the model and results are presented. The results demonstrate the optimal contracting strategies for the regulator and the operators in the CT of transit route packages in order to ensure a sustainable and equitable transit operation.

Fig. 1.1 Specific objectives of the thesis

1.3 Scope of Research

- The research would be confined to bus transit services only.
- This research would apply mathematical modelling and mathematical programming approach to develop the analytical framework, while no simulation or survey based approach is to be used.

- In modelling passengers' choice behaviour and transit assignment, the conceptual equilibrium approach would be applied.
- Model formulations developed for the government contracting regime could be applied to any transit market that adopts such a regime with relevant local operational parameters.

1.4 Outline of the research work

This thesis consists of seven chapters. Chapter 1 is the introduction. Chapter 2 presents a comprehensive literature review on bus network design, integrated service network design, common line problem and policy making in bus transit with respect to the government contracting regime. A detailed account of the various mathematical techniques and optimization methods used in this thesis is also provided. Thereafter, an integrated bus transit network design encompassing the limited stop service with the normal service using the deterministic and stochastic user equilibrium is presented in Chapter 3. Chapter 4 presents a congestion based bus transit network design with common lines wherein a queue theoretic approach is applied to model congestion. Chapter 5 presents the methodology for the design of bus transit route packages in a government contracting regime and Chapter 6 extends this study through a mathematical modelling framework for the competitive tendering of the bus transit route packages. Thereafter, Chapter 7 is the conclusion and recommendations for future research.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

An extensive literature review on several subject matters covered in this research thesis is provided in this chapter mainly including limited stop services, transit network design with common lines, bus transit contracting strategies and competitive tendering. For a good understanding of these chapters, a few important concepts relating to traffic theory, network design problems are also discussed followed by a few mathematical modelling methods and solution approaches.

2.2 An overview of important concepts

2.2.1 *Transport modelling and Traffic Theory*

Transport modelling is defined as a branch of science dealing with modelling flows on networks and estimate the optimal and most efficient network pattern. The traditional four step model of transport modelling includes trip generation, trip distribution, modal split and trip assignment. Trip generation involves computing the number of trips generated by households, trip distribution involves the identification of origins and destination for trips, modal split is identifying the mode share for such trips (public transport/private car/para-transit) while trip assignment helps in allocating flows on different routes of the network based on various parameters. There exists a set of basic principles that form the basis of mathematical modelling in transportation research such as the Wardrop's principles which are discussed below.

2.2.2 *Wardrop's Principles*

It is reasonable to understand that all travellers make route choices to minimize their travel times. A state of equilibrium is attained when no traveller can improve his/her travel time by unilaterally changing routes. This is proposed by Wardrop (1952) called as the Wardrop's First Principle or Deterministic User Equilibrium (DUE) which means that the travel time for used routes is lesser than the travel time for unused routes. The UE condition that mathematically describes the First Principle implies that travellers have complete and accurate information about the transit network to support their decision on the best choices and all travellers are

considered to behave identically. However, there is always a possibility of error in judgement between the perceived and actual travel time. A relaxation of UE states that Stochastic User Equilibrium (SUE) is achieved when no traveller believes he/she can reduce his/her perceived travel time by unilaterally changing routes. Beckmann et al. (1956) was the first to propose a mathematical programming model of UE based on the Wardrop's First principle. Consider a traffic network to be represented by a directed graph of N nodes, set of links A , a set of origin-destination (OD) pairs W , q^w representing the travel demand between the O-D pair $w \in W$, $t_a(x_a)$ representing the link performance (e.g. travel time) on link a which is a function of the flow x_a and f_r^w denoting the flow on route r between the O-D pair $w \in W$. The UE traffic assignment pattern can be obtained by solving the following mathematical program:

$$\begin{aligned}
& \min_x \sum_{a \in A} \int_0^{x_a} t_a(w) dw \\
& s.t. \sum_{r \in R^w} f_r^w = q^w, \forall w \in W \\
& x_a = \sum_{w \in W} \sum_{r \in R^w} f_r^w \delta_{ar}^w, \forall a \in A \\
& f_r^w \geq 0, \forall r \in R^w, w \in W
\end{aligned} \tag{2.1}$$

where R^w denotes the set of routes between O-D pair $w \in W$ and δ_{ar}^w is the link-path incidence taking a value of 1 if link a is on path r between O-D pair $w \in W$ and 0 otherwise.

Another principle called as the System Optimum (SO), also known as the Wardrop's Second Principle states that at equilibrium, the average travel time is minimum. This means that all travellers cooperate in choosing routes such that the travel time is minimized over the entire network basically producing the same effects of user equilibrium if congestion effects are ignored. The mathematical formulation for the second principle is the same as the first principle except for the objective function which can be replaced by the expression below:

$$\min_x \sum_{a \in A} x_a t_a(x_a) \tag{2.2}$$

The equilibrium problem is a favourite research area in transportation research wherein several model formulations, mathematical properties and applications of traffic equilibrium models have been comprehensively investigated (Sheffi, 1985; Patriksson, 1994; Bell and Iida, 1997).

2.2.3 Network Design Problems

Network design problems have received great attention in the literature. The main objective of the NDP is to optimize a specific performance index of the transportation system (e.g. minimize the total monetary travel cost), accounting for drivers' routing behaviour or traffic assignment at the same time. A few publications for reference are Boyce (1984), Magnanti and Wong (1984), Friesz (1985), Migdalas (1995), Yang and Bell (1998) and Farahani et al. (2013). The CNDP proposes optimal solutions to link capacity expansions in order to achieve a certain monetary objective subject to a certain flow pattern following the choice behaviour of the passengers. A few examples of such problems can be found in (Abdulaal and LeBlanc, 1979; Suwansirikul et al., 1987; Tan et al., 1979). The DNDP dealing with link additions can be found in (Gao et al., 2005; Poorzahedy and Turnquist, 1982; Solanki et al., 1998), and Luathep et al. (2011) presents the MNDP which is the combination of the CNDP and DNDP.

2.3 Different mathematical modelling techniques used for Transit Network Design Problems

Generally, a transit NDP is formulated as a single-level programming or a bi-level programming model with user equilibrium or system optimization constraints. The different mathematical programming problems possess different properties and hence, each of them has their specific solution algorithms. In this subsection, the most common types of mathematical programming methods used in NDP are discussed along with their respective mathematical properties.

2.3.1 Single-level programming

The term "level" here refers to sets of variables present in the mathematical formulation. A single-level minimization program can be written as:

$$\begin{aligned} & \min_{x \in X} f(x) \\ & \text{subject to:} \end{aligned}$$

$$g_j(x) \geq 0, \forall j \quad (2.3)$$

where x is a vector of variables and X is a vector matrix of domain of variables; $f(x)$ and $g_j(x)$ are functions of variables. There are two types of single-level programming: linear single-level programming and nonlinear single-level programming.

2.3.2 Linear single-level programming

In a linear programming problem, both the constraints and objective function are linear functions. A standard linear program can be formulated as (Sheffi, 1985):

$$\begin{aligned} \min & c^T x \\ \text{s.t.} & Ax \geq b \\ & x \in K \subset R^n \end{aligned} \quad (2.4)$$

where c is a vector of constants, A is a known matrix and b is a vector of parameters.

In certain cases in NDPs, a few variables should be discrete considering practical situations, for example, the decision whether a particular lane is to be expanded or not will be determined through binary integers or the number of lanes to be constructed will always be an integer. Hence, the domain of these variables would belong to an integer set. This type of linear programming is called a Mixed Integer Linear Programming (MILP). The MILP formulation can be described as below:

$$\begin{aligned} \min & c^T x + d^T y \\ \text{s.t.} & Ax + By \geq b \\ & x \in K \subset R^n \\ & y \in Y \subset Z^m \end{aligned} \quad (2.5)$$

where y is a vector of integer variables; c, d are vectors of parameters and A, B are matrices of coefficients. If all the elements in both vector and matrix are zero, then the MILP becomes a linear program. Linear programs including MILPs are extensively used in NDP as they are easy to solve to global optimality.

2.3.3 Nonlinear single-level programming

Convexity plays a major role in nonlinear single-level programming. A set K in R^n is convex if for every pair of vectors $x, y \in K$, the segment joining them is also contained in K :

$$\alpha x + (1 - \alpha)y \in K, \alpha \in [0, 1] \quad (2.6)$$

A function $f : K \rightarrow R^n \rightarrow R$ is convex if its domain is a convex set and

$$f(\alpha x + (1 - \alpha)y) \leq \alpha.f(x) + (1 - \alpha).f(y); x, y \in K, \alpha \in [0, 1] \quad (2.7)$$

If $f(x)$ is a convex function and $g_j(x) \geq 0 \forall j$ is a convex set, then the single-level programming is a convex programming.

A few important definitions related to nonlinear programming are mentioned below.

- (a) Feasible solution: any point which is located in the feasible region of the optimization problem.
- (b) Local optimal solution: A feasible solution which is the optimal solution in a small neighbourhood belonging to the feasible region.
- (c) Global optimal solution: A feasible solution which is no worse than the other feasible solutions for the optimization problem.

Convexity is crucial in nonlinear programming as it can make necessary conditions of a global optimal solution become necessary and sufficient. For instance, a local minimum cannot be guaranteed to be a global minimum for the case of a general nonlinear program; however, a local minimum in a convex programming is certainly a global minimum. Also, in case of a strictly convex mathematical programming, the global minimum is always unique (Sheffi, 1985).

2.3.4 Bi-level programming

Bi-level programming is a special case of multi-level programming wherein there are two levels and is commonly applied to NDPs. Each of these levels has a different set of variables. The upper level of the bi-level programming model demonstrates a control policy and the lower level usually represents the user equilibrium or system optimization.

In general, a bi-level programming problem is defined as follows:

$$\begin{aligned}
(P): & \min_x (U) F(x, y(x)) \\
& s.t. G(x, y(x)) \leq 0
\end{aligned} \tag{2.8}$$

where $y(x)$ is implicitly defined by

$$\begin{aligned}
(L): & \min_y f(x, y) \\
& s.t. g(x, y) \leq 0
\end{aligned} \tag{2.9}$$

In the above model formulation, U is the upper-level problem and L is the lower-level problem. $F(x, y(x))$ is the objective function of the upper-level decision-makers or system managers (called as master), x is their decision vector, $G(x, y(x))$ is the constraint set of the upper-level decision vector, $f(x, y)$ is the objective function of lower-level decision-maker (called follower), y is their decision vector and $g(x, y)$ is their constraint set vector. It should be noted that the decision variable of the lower-level problem is expressed as an implicit function of the decision variable of the upper-level problem and is usually called as the reaction or response function. The master influences the follower by setting x , thus restricting the feasible set for the follower. The master also interacts with the follower through the lower level objective function. Because of their non-convexity and non-differentiability, bi-level programs are inherently hard to solve. Even for a bi-level program wherein both the levels are linear programs, convexity still cannot be assured. Researchers have used different methods to solve NDPs modelled as bi-level programs. Some use iterative heuristic methods while others transform the bi-level program into a single level program by substituting a set of constraints for the lower-level program and incorporating them. Generally, two techniques are often applied in the transformation process namely variational inequality (VI) and the nonlinear complementarity problem (NCP). Gao et al. (2004) presents a descriptive transit NDP and is a good reference to understand how bi-level programming can be used to model multi-party objectives. Let us now try to understand more about the transformation processes used in bilevel programs.

(i) *Variational inequality:*

The VI formulation is useful in describing equivalent constraints of a traffic system such as in the modelling of traffic control and intelligent traffic systems. The

advantage of VI lies in the fact that it can be used to formulate a variety of mathematical problems such as an optimization problem, fixed point problem and the complementarity problem. Also, under some monotonicity condition, the optimal solution of VI exists and is unique in nature. The definition of VI problem can be stated as given below.

A finite-dimensional VI problem is to determine a vector $x^* \in K \subset R^n$ which can satisfy the following constraint:

$$F(x^*). (x - x^*) \geq 0, \forall x \in K \quad (2.10)$$

where F is a continuous function from K to R^n and K is a given closed convex set.

ii. Nonlinear complementarity problem:

For a continuous differentiable function $F : R^{n+} \rightarrow R^n$, a complementarity problem (CP) seeks to determine a vector $x^* \in K \subset R^n$, such that (Facchinei and Soares, 1997):

$$\begin{aligned} F(x^*)^T . x^* &= 0 \\ F(x^*) &\geq 0 \\ x^* &\geq 0 \end{aligned} \quad (2.11)$$

If $F(x^*) = Mx + b$, M is a $n \times n$ matrix and b is a $n \times 1$ vector, the above is not a linear complementarity problem; on the contrary, it is a NCP which is actually a special case of the VI problem.

2.4 Classic solution methods for Transit Network Design Problems

Different mathematical programs require different algorithm methods to solve them to optimality. In this subsection, a few classic solution methods for the general linear program, MILP and nonlinear program are discussed.

2.4.1 Solution methods for linear programming

i. Solving linear programming

The optimal solutions of linear programs always sit at the boundary of the feasible region (Sheffi, 1985), hence, they are easier to solve. The simplex method is one of the most traditional algorithms for solving linear programming problems. The algorithm finds the feasible region by progressing through adjacent corner points. The detailed procedure for solving linear programs using simplex can be found in

several books on optimization. Moreover, with the help of fast developing computer technology, linear programming can be readily solved by the use of efficient commercial software such as CPLEX (commercial), MATLAB (commercial), LPSOLVE (free), etc.

ii. Solving MILP

The complexity of MILP lies in the combinatorial nature of the domain of integer variables. The solution algorithms for MILP can be classified as (Floudas, 1995):

(a) Branch and bound methods: A tree represents the combination of integer variables and the original MILP is separated into many sub-problems. At each level of the tree, bounds of the optimal solution are calculated.

(b) Cutting plane methods: A step wise addition of constraints to reduce a part of the feasible region until an optimal solution is found.

(c) Decomposition methods: These methods explore the structure of MILP models through partitioning, duality, and relaxation methods.

(d) Logic-based methods: These methods are based on disjunctive constraints or symbolic inference techniques which can be demonstrated using integer variables.

Amongst the proposed solution methods, the branch and bound method is the most widely used in solving large-scale MILP. The success of branch and bound method for a specific MILP is evaluated by the percentage of reduced nodes and the computational effort required in solving the sub-problems.

2.4.2 Solution methods for nonlinear programming

The classic nonlinear programming algorithms can be categorized into six groups (Floudas, 1995) such as exterior penalty methods, interior function methods, gradient projection methods, generalized reduced gradient methods, successive linear programming methods, successive quadratic programming methods etc. However, all these classic methods can only find local optimal solution and do not guarantee global optimum.

2.5 Limited stop services

Public transit services are lifelines for daily commute in many major cities in the world. In order to increase the service quality and attract more patronage, constant improvement in operation and design is of paramount importance. In the presence

of increasing daily travel demand, public transit service operators now seek to improve their service quality to efficiently satisfy the travel demand while maintaining operation in a more financially sustainable manner. In many cities, bus transit services have become more convenient with the inclusion of differential services such as normal, express, and limited stop services which are operated to cater for various demand patterns. While a normal service serves all the bus stops present on a route, an express service travels end to end without or with very few intermediate stoppages. A limited stop service serving a selected subset of nodes in a network provides a balanced alternative and helps transit operators in reducing overall passenger travel time by catering to different types of demand and increasing bus service patronage. Hence, a limited stop service is of great financial and social importance to bus transit operation and due academic attention needs to be given to develop methodologies to provide guidelines to bus operators on how to design optimal limited stop services so as to improve their cost-efficiency.

In the literature, transit service network design has attracted much research attention. Ceder and Wilson (1986) discussed the bus route planning problem that minimizes total system operation cost while also addressing the scheduling problem. Since then, a vast body of literature on transit network design has emerged which involves optimal decisions of routing and scheduling, service frequency design, inter-node spacing, fleet size design, etc. Wang and Lo (2008) presented a related work on a multi-fleet ferry routing and scheduling problem that considered ferry services with different operational characteristics. Cortés et al. (2011) presented a methodology to optimise costs while integrating two kinds of services in the transit network with deadheading and short turning services. Yadan et al. (2012) proposed a robust optimization model for the bus route schedule design problem by taking into account the bus travel time uncertainty and the bus drivers' schedule recovery efforts. A few studies discussed the bus dwelling time which is critical towards determining the total travel time of passengers (e.g., Meng and Qu, 2013; Sun et al., 2013). Bus transit generally operates under different market regimes and a few studies in the literature have also contributed towards this aspect (e.g., Li et al., 2010; Li et al., 2008). Liu and Meng (2012) modelled the network

flow equilibrium problem on a multimodal transport network with bus-based park-and-ride system and congestion pricing charges. Li et al. (2011) addressed the design problem of a rail transit line located in a linear urban transportation corridor where the service variables include a combination of rail line length, number and locations of stations, headway, and fare. In addition to the above mentioned studies, there exist many other published works on transit service network design; but unfortunately, very few studies focus on the design of a limited stop service in a bus transit network.

Limited stop services have been in operation in cities like Bogota, Chicago, Montreal, New York City, Santiago, etc. Afanasiev and Liberman (1983) described a limited stop service as a service with stops at intervals of about 0.8 km. Silverman (1998) proposed a few important considerations while designing a limited stop service: wider roadways, not too close to rapid transit corridors, operationally more successful over long distances, etc. Conlon et al. (2001) noted that implementing a limited stop service parallel to a normal bus service drew appreciation from users in Chicago where user satisfaction for both the services increased after the inclusion of the former. El-Geneidy and Surprenant-Legault (2010) observed that a limited stop service is the most preferred choice of passengers as they tend to overestimate their time savings while using this service. Tétreault and El-Geneidy (2010) proposed a stop selection methodology for limited stop services based on archived Automatic Vehicle Location (AVL) and Automatic Vehicle Classification (AVC) data obtained from a travel behaviour survey in Montreal, Canada. This included different scenarios where stops were selected based on passenger activity and transfers. As can be concluded, studies mentioned above mainly focused on the operational aspect of limited stop services which is data-driven and descriptive while no analytical approach was proposed for the service design.

Limited stop bus service design is one typical transit service network design problem. In designing a limited stop service, bus stop selection is the prime decision variable, i.e., to determine which stops the bus service should stop or skip in the transit network. In addition, other operation strategies like optimal fleet size and service frequency should be determined considering the passengers' service

choices. As included in the model formulation proposed in this study, consideration of bus service capacity is also significant because if the capacity is large enough, then minimizing social costs such as passengers' waiting and travelling time (as mentioned in the objective function of the proposed model) would lead to an assignment that is consistent with the passengers' own choice patterns. Some research studies have been conducted to develop methodological frameworks to prescribe guidelines for their operation in terms of optimal service network design. Larrain et al. (2010) proposed the methodology to select optimal express services for a bus corridor with capacity constraints considering various demand criteria whereas Larrain et al. (2015) designed zonal bus services which skip all intermediate nodes over a segment of the transit route while serving all nodes in the initial and final segment. Ulusoy et al. (2010) presented a methodology to optimize the operation of an integrated system of normal, short turn services, and express services. Leiva et al. (2010) presented an optimization approach to design a limited stop service with capacity constraints. However, in both of these works, the selection of the bus stops in the limited stop service is given in priori and the service frequency of limited stop services lines is the only operation strategy determined by the model. In their model, only the optimally selected limited stop service lines are assigned a non-zero frequency and the others are assigned zero frequency. However, using a given subset of bus stops as the candidate service design for limited stop service does not provide the truly "best" design and underappreciates the benefits derived from this service. Chiraphadhanakul and Barnhart (2013) proposed a design of the limited stop service by optimally reassigning certain bus trips rather than providing additional trips. However, this work does not consider transfers or multiple lines operating over common route corridors where passengers can make a choice. Besides, it allows only one limited stop service to be operated over the transit network and the frequency of the limited stop service is not taken into account for passenger assignment onto the respective services. A detailed comparison between the above mentioned studies on limited stop services and this study is illustrated in Table 2.1.

In this chapter, we present a mathematical model formulation to explicitly design a limited stop service with optimal decisions on bus line configurations (the set of bus

stops served by the limited stop service) along with other operation strategies including operating frequencies and the optimal fleet size. The service frequencies need to be decided so that the demand is satisfied and the service continuity is maintained. A dedicated fleet size for each bus service is significant towards estimating the capacity of the service during the peak hour operation and has to be set in coordination with the service frequency. The choice behaviour of passengers is greatly affected by the travel time and hence, the model explicitly computes the travel time for passengers using each of the two services. The user equilibrium principles used in the model formulation provides the rationale that passengers use while choosing a particular service to travel towards their destination node. At the same time, due consideration must be given to the service performance from the perspective of passengers as poor service performance may lead to a drop in demand or the possibility of losing the franchise of operating the routes altogether. Therefore, the objective function also incorporates passengers' travel and waiting time as important factors which are to be adjusted by appropriate weights. Although this study assumes that the bus services are operated in a monopolistic market with fixed total demand, it should be noted that the model framework (which focuses more on how to model and solve the optimal operation strategies for limited stop services) can be easily extended to consider elastic demand or competition with existing alternative bus services.

Table 2.1 Comparison between the published studies on limited stop service design and this work

Factors\Studies	Leiva et al. (2010)	Ulusoy et al. (2010)	Chiraphadhana kul et al. (2013)	This work
Assumptions				
Origin-destination (O-D) matrix	Fixed	Fixed	Fixed	Fixed
Transfers	Allowed	Allowed	Not Allowed	Allowed
Number of limited stop services allowed	Unlimited	Unlimited	1	Unlimited

Objective	Minimize social costs	Minimize social costs	Maximize social welfare	Minimize social costs
Decision variables				
Fleet size	✓	✓	✓	✓
Frequency	✓	✓	-----	✓
Explicit design of a new limited stop service	----- (Given a predefined set of candidate services)	----- (Given a predefined set of candidate services)	✓	✓
Constraints				
Choice behavior	DUE (deterministic user equilibrium)	SUE (stochastic user equilibrium)	System Optimal	DUE (deterministic user equilibrium)
Common line approach	✓	-----	-----	✓
Capacity	✓	✓	✓	✓
Fleet size	✓	✓	✓	✓
Assignment	Proportional to frequency of each attractive line	Logit based model considering wait, transfer, in-vehicle times	A linear function of frequency share and in-vehicle travel time savings	Proportional to frequency of each attractive line

Multiple services along route segments	✓	✓	-----	✓
Shorter running times of limited stop services	✓	✓	-----	✓
Incorporating limited stop service(s) frequency in demand model	✓	✓	-----	✓

In designing the transit network with the normal service and limited stop service, one intrinsic issue to be considered is the travellers' choice behaviour between different services. In this study, such passenger service choices are described by the classical common line problem for the DUE case and the logit model for the SUE. For the DUE, it is assumed that the travellers choose a subset of bus service lines that minimizes the expected total travel time which was defined as the common line problem in Chriqui and Robillard (1975). The common line problem has been investigated in many research works in the literature such as Spiess and Florian (1989), De Cea and Fernandez (1993), Cepeda et al. (2006), etc.

2.6 Congestion based bus transit network design with common lines

Congestion is one of the major problems faced by daily commuters today. Every user perceives the service quality as the ability of the system to accommodate demand while maintaining the minimum desired level of service. Including congestion effects due to passenger flow into transit network design problem is the essence of this work. Transit network design problems can be broadly classified as discrete, continuous and mixed transit network design problems dealing with line setting, service frequency design or both respectively. With a given line setting, transit lines serving a network need to be operated at frequencies such that they cater to the demand efficiently with minimum operating costs and network travel

time. In a transit network design problem modelled as a typical Stackelberg game, the transit operator (leader) minimizes the operating costs whereas the patrons (followers) minimize their network travel time and inconvenience due to congestion effects. Designing transit lines in order to meet the objectives of both the involved entities is the goal of a typical CTNDP.

There exists a vast literature on CTNDP as transit route design modelled as a CTNDP has interested several researchers in the past. The work on bus route design by Ceder and Wilson (1986) discusses the bus route planning problem that minimizes total system operation costs. There also exist studies involving problems on scheduling, service frequency design, inter-node spacing, fleet size etc. Alonso, Moura et al. (2011) proposed a methodology in which a bi-level optimization problem determines the bus stop spacing under different network congestion levels. Chien and Qin (2004) formulates a model for optimal bus stop location to minimize costs considering users' access time and value of time. Kikuchi (1985) showed the relationship between the two most important transit parameters i.e. headway and the number of stops on a transit network. Ceder (1984) described several data collection approaches for the bus operator to set bus frequencies/headways. Apart from elements of route design, a few contributions also focused on the mathematical modelling paradigm of a CTNDP. Constantin and Florian (1995) presented a nonlinear, non-convex mixed integer programming approach to the CTNDP and converted it to a bi-level optimization problem. Gao et al. (2004) then presented a bi-level formulation wherein they designed bus line frequencies with user equilibrium as the lower level problem and the operating costs as the upper level. In this study, the transit network design problem is modelled as a MPEC. A Network Design Problem(NDP) can generally be represented by a MPEC but unfortunately, most of the solution algorithms except for Wang and Lo (2010) do not guarantee global optimality. Wang and Lo (2010) approximated the Continuous Network Design Problem (CNDP) as a mixed-integer linear programming (MILP) problem to attain global optimality.

In this study, the concept of common lines is incorporated while designing the transit network. In the literature, Chriqui (1974) and Chriqui and Robillard (1975)

were the first contributions to define the common line problem which means that travellers will choose a subset of bus service lines that minimizes the expected total travel time. Later, Speiss and Florian (1989) presented their work on an assignment model dealing with optimal travel strategies. They proposed that at every transfer node, a transit commuter travelling between an O-D pair forms a strategy comprised of a set of common lines that he/she could take to travel towards the destination node. The commuter then takes the first vehicle amongst the set of lines and this minimizes his/her overall network travel time. The proposed model computed attractive lines and assigned passenger load onto them on the basis of their known nominal frequencies. Modelling congestion and incorporating its effects into the transit network design has also found some research attention in the past. Gendreau(1984) was the first to model congestion on a transit network using passenger arrival distributions and waiting time through a bulk queue model and hyper paths. However, due to its complexity in dealing with network equilibrium, De Cea and Fernandez (1993) later proposed a new equilibrium model for a congested transit network using a congestion function and the common lines for every route section were computed through a hyperbolic programming problem. But as pointed out by Cominetti and Correa (2001), it has a few drawbacks such as: the function used to exhibit congestion could load the route sections beyond capacities, it is a heuristic formulation etc. Therefore, to improve upon these drawbacks, they proposed a queue theoretic approach to model congestion using common lines. They suggested that at equilibrium, a route section served by a set of transit lines could have multiple travel strategies, i.e. multiple sets of attractive lines that commuters could choose to travel on. Route section travel strategy time was considered to be a function of route section flow and non-linear, non-monotonic in nature. Recently, the study by Shimamoto et al. (2012) proposed that when common lines are used to design a transit network, the optimal network configuration obtained favours direct services when considered over a small network.

2.7 Bus Transit Route Packaging

Bus transit is, by far, the most flexible, accessible and favored mode of public transportation across the globe. Since the advent of organized public transportation

regulatory authorities dedicated towards the systematic provision of transit facilities such as the U.S. Department of Transportation for the USA or the Land Transport Authority (L.T.A.) for Singapore etc., several operative models have been conceptualized and implemented to make the system financially and socially sustainable. Public transportation infrastructure improvisation is one domain where the government interacts with the public in a direct way on a daily basis and therefore, to make it more convenient remains a paramount concern.

In the past, the post-world wars' period in the latter half of the 20th century saw huge financial losses for the public transport industry and the concurrent surge in private vehicle ownership which led to a major downsizing of the public transportation infrastructure. With motorization causing environmental degradation and posing a threat to sustainability, a few nations moved to nationalize the facilities so that the system could operate with minimum government subsidies. Although it was not very beneficial, it still helped keep the industry afloat. The first major breakthrough in transportation policy making came in the form of the Transport Act (1985) passed by the Parliament of the United Kingdom which aimed at the deregulation of public transportation facilities and introducing competition in the transit market through privatization. This was a major development in the transportation industry with a strong intention to structure public transportation policy towards social welfare. It also revolutionized the nature of public transport from a mere service to a competitive market calling for innovation and high levels of service. Many other countries tried to introduce competition through deregulation subsequently.

From the contractual point of view, there exist different regimes under which privatization or deregulation might operate and there is always a debate on the most financially feasible contracting regime for a particular transit system. There have been studies investigating on this topic; however, as since transit facilities are greatly affected by local politics, economy and financial institutions; it is difficult to draw a consensus on a preferred regime. The contribution by Hensher and Stanley (2003) elaborates on the performance based contracts where the operations are based on trust and relationship building. Such a contract is generally negotiated

with the incumbent operators. In recent times, transport authorities of some of the developed cities such as Perth and London observed that their operative models wherein the licensed operators planned, operated transit routes and kept the revenue was a costlier financial model with higher operating costs, duplicated and lower levels of service and lower revenues with higher subsidy requirement. Hence, they gradually transitioned to a government contracting model where the government owns all the assets and the transit routes are tendered out to operators only for the operation through CT. Literature provides certain comparative studies between the two different kinds of operative regimes mentioned above: performance based contracts and CT. The work by Hensher and Stanley (2010) reviews the themes that are crucial in choosing a particular contractual regime suggesting that performance based contracts would deliver a better value for money. Further, Hensher and Wallis (2005) discusses the success and failures of the CT model in pursuit of a better value for money objective and promotes performance based contracts to contain government spending through subsidies. On the contrary, the contribution by Wallis and Bray (2014) provides evidence that CT has greatly reduced the cost of transit provision and improved quality of service in some of the major Australian cities such as Adelaide and Perth as compared to prior regimes such as the performance based negotiated contracts. Hence, choice of an operative model is subjective in nature and largely depends on the local conditions and the financial objectives of all stakeholders involved.

The existing research works mentioned above only provide descriptive discussions using statistical approaches on the qualitative nature of the different kinds of contractual regimes while never addressing analytical frameworks for contracting. Also, studies on CT are only restricted to strategies that could help attain cost efficiency without presenting any mathematical paradigm for decision making in a transit route tendering process. For instance, the work by Iseki (2010) computes the cost efficiency on the basis of degree of contracting i.e. how much of the transit service is contracted out through CT. Another contribution by Iseki (2008) examines the cost efficiency in transit provision through CT by categorizing agencies into different size groups in contrast to past studies where the whole transit industry was assumed to be a set of agencies with similar cost structures. Evidently,

these studies are mostly statistical works that demonstrate how contracting under the CT model could be made cost efficient through the optimal degree of contracting. However, both of these contributions fail to propose any decision making framework for the regulatory authority to contract out transit routes through the CT model.

While contracting out transit routes in a government contracting model through CT, the regulatory authority needs to make decisions on various elements of the contract such as the transit route packages, number of private players to be allowed to operate in the transit market and the methodology of allocating transit route packages to the eligible players. The resulting design of the contracting process should be financially and socially sustainable so that it is equitable for all the involved entities i.e. the regulator, operators and the commuters in the long run. In this study, the transit route packaging aspect is addressed under the CT model. An explicit analytical framework for the regulators' decision making on the optimal transit route packaging for a deregulated transit market is presented. Considering the fact that more countries are now willing to adopt an efficient financial model for public transport operation through bus contracting, this research would be useful for the regulatory authorities in making decisions that promote larger social welfare. In this study, the focus is on the transit route packaging methodology of the regulator, a decision problem that impacts the finances of the operators and the regulator and hence, needs to be acutely studied to ascertain a sustainable operation.

In a more specific context, when the Land Transport Authority, the transportation regulatory authority of the Govt. of Singapore, decided to transition from a privatized to a government contracting scheme of bus operation, they divided the entire bus transit network into 12 route packages (LTA News Release, 2014). Out of these 12 packages, 3 packages are to be contracted out in the first stage and would be operational from 2016 onwards for a period of 5 years. With the remaining packages to be tendered out gradually, the authorities plan to have around 4-5 bus operators operating different route packages by the end of 2021, leading to a competitive transit market in contrast to the present duopoly of SBS and SMRT, both companies with major government shareholding. Apart from the

Singaporean players, the tendering process has reportedly attracted foreign operators from U.K., Australia, Europe leading to stiffer competition for the quality and price bidding of route packages. As per latest developments, Tower Transit, a U.K. based transit operator won the bid to operate the first route package tendered out in October 2014. This is for the first time a foreign player has entered the Singapore bus transit market.

2.8 Competitive Tendering of Bus Transit Route Packages

Bus transportation is today one of the most favoured modes of public transportation in all countries which discourage use of private transport. However, different countries face different challenges depending on the state of their respective bus transit system. These cities could be classified into different categories such as: (i) cities with lack of a bus infrastructure at the primary level (ii) cities with an existing bus infrastructure but a low level of service/operation and (iii) cities with good bus infrastructure and level of service but an unsustainable financial model of operation. This classification is mainly based on the immediacy with which countries perceived the need to promote public transportation after motorization expanded in the post second world war period. In this study, the focus is on the third strata of cities. In many of these cities, although their respective governments tried to nationalize the transit industry, the lack of public funding and the inefficiency of worker unions brought the state of public transportation to a standstill, increasing the rate of motorization. However, since the time countries like the UK stepped up measures to popularize bus transportation through privatization etc., ridership increased and since then, there has been a growing urge to create an equitable bus system in various parts of the world such that the financial objectives of all involved stakeholders viz. regulator (a government authority), operators and the commuters are optimally met. There are cities which, in spite of having the infrastructure and fulfilling commuters' level of service expectations, lack a sustainable financial operative model which poses a constant threat to bus transit. With increasing uncertainty in the financial markets, this issue is on the rise as commuters are averse to fare rise without sufficient improvement in the level of service.

Privatization, in the late 1980's, started with governments awarding licenses to private transit players to operate through negotiated contracts where terms of operation were negotiated between the regulator and the operators. With time, the system proved to be costly as the operators could not keep up with the costs with low fares and low government subsidy. To alleviate this issue, a few other contracting regimes were implemented by various transit markets such as the performance based contracts, CT etc. However, which regime works the best for a particular transit market is a highly subjective issue as each market differs greatly with the local demand pattern, financial scenario, political situation etc. There are several published works addressing this debate over the preferred regime for transit markets. The contribution by Hensher and Stanley (2003) elaborates on the performance based contracts where the operations are based on trust and relationship building. The work by Hensher and Stanley (2010) reviews the themes that are crucial in choosing a particular contractual regime suggesting that performance based contracts would deliver a better value for money. Walters and Cloete (2008) compares the negotiated and CT forms of contracting in South Africa and their respected advantages and disadvantages are highlighted. Mouwen and Rietveld (2013) suggested that tendering slightly improves commuter satisfaction which, however, diminishes with time and especially with the change in operator. Hensher and Stanley (2008) discusses the role of negotiation and CT in a performance based contract proposing that a combination is possibly a good model. The study suggests that while the CT model tends to utilize the public funds in the most feasible manner by reducing costs, however, it stifles communication between the regulator and the operators and deters the regulator from making the most out of the operators' expertise. On the contrary, a combination model provides a way forward when operators fail to comply under reasonable notice. The study suggests a need to develop trusting partnerships which is crucial for a sustainable transit operation. Hensher and Wallis (2005) discusses the success and failures of the CT model in pursuit of a better value for money objective and promotes performance based contracts to contain government spending through subsidies. On the contrary, the contribution by Wallis and Bray (2014) provides evidence that CT has greatly reduced the cost of transit provision and improved quality of service in some of the

major Australian cities such as Adelaide and Perth, as compared to prior regimes such as the performance based negotiated contracts. As regards concerns about implementing the CT model, Schaaffkamp (2014) discusses how several authorities are not fully satisfied with the outcome of their tendered contracts. Although there is no evidence that a certain contracting mechanism ensures success in the transit market but in CT, operators could become insensitive to the commuter's expected level of service as there is no financial incentive involved. This work focuses on the interaction between the regulator and the operators while they optimize their respective financial objectives. Recently, cities such as Perth, London transitioned to a contracting model where the regulator contracts out bus route packages to interested private players through CT. In a bus contracting model, the regulator designs the bus route frequencies, allocates the fleet size, designs the route packages as sets of routes and awards them to contesting operators through the system of CT. Although no studies have yet assessed their performance, but as regards the basic rationale of this system, there is a good chance of a surge in bus ridership because of the following reasons: (i) the regulator who designs the network and draws revenues from fares is more sensitive towards demand pattern and hence changes towards service improvement can be easily implemented (ii) operators are obliged to maintain a minimum level of service or else could be penalized as per the contractual terms (iii) as CT is used to award the contract, the operator with the best operating terms is selected, hence, lowering overall costs to the regulator, maximizing the level of service and ensuring optimal utilization of public funds. In this study, a mathematical framework is proposed for the contracting of transit route packages to contesting operators through CT. This not only helps understanding the intricacies of the contracting process but also defines roles of each stakeholder in contributing towards a sustainable transit operation in the long run. Let us first discuss a few important guidelines for the CT process that directly affect the contractual terms as proposed by Uher (2009).

In a bus transit system based on the CT model, the regulator is the principal and contracts out transit route packages to contesting operators who are the contractors. The regulator advertises the tender for the operation of the route package and all operators are required to submit their proposals within certain duration with their

operating specifications. The regulator is free to decide whether the tender would be an open or a select tender. In the former case, a large number of operators are invited to submit bids. Although this involves high administrative costs, it is advantageous for new market entrants. In a select tender, the regulator selects a few operators with a good and proven track record and invites them to submit bids for the tender. These are few in number and economical for the regulator. For the submission of bids, the operators may or may not be required to pay a surety bond which ensures they do not withdraw their proposal within a stipulated period else the amount could be forfeited. The operators submit a set of documents expressing their intent to operate with the price specifications for various operational aspects. The regulator decides upon the best tender proposal based on the tender qualification criteria which are various quality and price aspects such that the selected proposal is the most equitable for all stakeholders. As the regulator is mostly not-for-profit and is mostly publicly funded, it has a greater responsibility to ensure that the best proposal is selected. Generally, the proposal with the lowest price quote is selected; however, this is not always necessary. A proposal with a low price quote could be possibly overlooking a few aspects such as contingency budgets etc. that would increase cost of operation in the future. However, that might not be fair when the lowest bidder does not secure the contract and would rather be a waste of financial resources. In some cases, even monetary compensations could be provided to such a bidder (Collier, 1969). Sometimes, the regulator selects the proposal which falls within an acceptable range of its own estimate. There exist different practices in different parts of the world regarding selection of the best tender. In operations involving higher risks and an open tender, the proposal with the least quote is not selected as there is a possibility that risk or contingency budgets have not been provided. However, for a select tender, such a proposal could be selected as the bidders have been invited by the regulator based on their past performance record (Eilenberg, 1987). Some regulators also select the best proposal by assigning weights to the technical and commercial value of each submitted proposal where the technical aspects include the operational terms and the commercial aspects include the market share etc. of the company/operator. It must be realized that the costs of tendering are huge for the regulator and the operators.

The winning operator compensates for the costs through service fees and the others try to compensate for the loss in future contracts. To reduce the administrative and tendering costs to the regulator, it is recommended that the final number of total bids from which the best proposal is selected should not exceed seven (Kirkpatrick, 1963).

It is to be noted that although bus transit contracting has been in practice for a long time, however, there have been no published works that provide a methodology for the comprehensive analysis of the tendering process. Modelling the competitive tendering process as an interactive game between the regulator and the operators in the form of a bi-level optimization problem is the essence of this study. This study, therefore, provides a new mathematical modelling paradigm for the transit regulator to decide on optimal bus contracting strategies for a financially sustainable operation.

CHAPTER 3 BUS TRANSIT NETWORK DESIGN WITH LIMITED STOP SERVICES

3.1 Introduction

In summary, this chapter contributes to the literature in two major aspects. First, we develop a mathematical model to fully address the optimal design of limited stop bus services which explicitly determines bus line configurations (the set of bus stops served by the limited stop services), operation frequencies and fleet size assignment while considering passengers' service choices. Previous research studies assumed pre-determined limited stop bus lines which simplified the problem and circumvented complicated model formulation while compromising on the model's capability of determining the truly best limited stop bus services. Second, a global optimal solution method is developed to solve the model formulation as previous studies in transit service network design ignored this aspect due to the inherent non-convexity of the formulated problem.

In this study, the model is formulated into a mixed integer nonlinear programming problem. One may consider this bus network design problem with limited stop services as a bi-level mathematical program; the lower level program describes the passenger assignment problem whereas the upper level program is the bus service network design problem. Alternatively, as in this study, we formulate the lower level passenger assignment problem as equivalent mathematical conditions, thus reducing the bi-level program into a mathematical program with equilibrium constraints (MPEC). The formulated mixed integer nonlinear program is inherently non-convex; hence, we devise a solution algorithm applying various linearization and convexification techniques to find the global optimal solution. A global optimal solution guarantees the best possible operation strategy and therefore, it is necessary for bus service operators to have such a solution if the objective of minimizing the total operation cost is to be achieved while ensuring a financially sustainable bus operation. It should be noted that most of the previous research works on transit network design problem did not guarantee a global optimal solution. One can further note that the constraints in the model formulations of several bus network

design problems in the literature were simply removed or relaxed to facilitate obtaining the solution efficiently but unfortunately, this does affect the solution accuracy to a certain extent.

3.2 Design of limited stop service using DUE

In this section, we present a model formulation to optimally design the limited stop service using the DUE choice behavioural approach. Before doing so, we first define the general network setting and notation, make a few preliminary assumptions and state the general mathematical constraints for a limited stop service.

3.2.1 Network setting

The transit network considered is linear such that the last node number is equal to the total number of nodes in the network. A bus cycle refers to one complete trip from the first node to the last node of the network. A loop service may be considered if the first and the last nodes are the same. For the purpose of illustration, an example transit network of 5 nodes numbered 1 to 5 is shown in Fig. 3.1.

Fig. 3.1 Transit network

In this corridor, passengers can utilize the transit services to travel from any origin node i to any other destination node j (for all values of j greater than i). The normal services cater to every node of the network whereas the limited stop services serve only a certain subset of all the bus stops which are to be determined through optimization and are referred to as *special nodes*. The operator has a fixed fleet of buses which is to be split amongst the operating services. The fixed fleet is split between the services in such a way that the service frequency per operational period satisfies the demand. Each service (limited stop or normal) has its own respective cycle time and hence a minimum dedicated fleet size is to be ensured to ensure continuous operation as per the designed service frequency.

3.2.2. Variable definitions

Sets and parameters

i, j	Any two arbitrary and different nodes in the network
N	Set of nodes in the transit network, $i \in N$
W	Set of O-D pairs
$w(i, j)$	An O-D pair connecting node pair (i, j) , $w(i, j) \in W$
d, e	Indices for origin and destination of any O-D pair
L	Set of all transit lines
L'	Subset of limited stop services, $L' \subset L$
L''	Subset of normal services, $L'' \subset L$
S_{ij}	Set of route sections between any two arbitrary and different nodes i and j
S_i^+	Set of route sections coming out from node i
l_{ij}	Transit line serving route section $ij \in S_{ij}$
S_i^-	Set of route sections going into node i
ij	Route section joining node pair (i, j) , $ij \in S_{ij}$
l_s	Limited stop service
l_r	Normal service
W_c	Value of waiting time
T_c	Value of travel time
k	Parameter whose value depends on the distribution of bus arrival times at stops

T^h	Fixed dwelling time at node h
T_{ij}	Non-stop running time between node pair (i, j)
K_l	Bus ownership cost of any line l
$X^{w(i,j)}$	Exogenous demand for O-D pair $w(i, j)$
F^l	Operating cost per cycle of any transit line l
γ_{ij}^l	Bus fare on line l for node pair (i, j)
B	Available fleet size
Cap_l	Passenger capacity of a bus on any line l
b	Cardinality of the set N (number of nodes in the bus service network)
$t^{l,cycle}$	Full cycle time of the bus service l to return back to the origin node
λ_{trans}	Coefficient to convert the number of transfers to cost terms

Definitional variables

y_p^{i,l_s}	Binary variable, it equals one if the limited stop service l_s leaves node i
y_q^{i,l_s}	Binary variable, it equals one if the limited stop service l_s arrives at node i
t_{ij}^l	Travel time between node pair (i, j) on line l
$V_{ij}^{w(i,j)}$	Passenger flow over route section ij for an O-D pair $w(i, j)$
x_{ij}^l	Line l over route section ij in the common line problem is attractive if it takes a value of one, binary variable
$v_{ij}^{w(i,j),l}$	Passenger flow on any line l over route section ij for an O-D pair $w(i, j)$

j)

$t^{l,cycle}$ Travel time for one bus cycle on any line l

Decision variables

n_l Fleet size allocated to any line l

f^l Operating frequency of any line l

$y_{ij}^{l_s}$ Binary variable, denoting the direct service between node pair (i, j) for a limited stop service l_s exists if it equals one

3.2.3. Assumptions

3.2.3.1. Passenger demand

Exogenous demand between each OD pair is assumed to be given. However, the demand for a particular bus line service depends on its service quality and is determined by the choice behaviour of the passengers.

3.2.3.2. In-vehicle travel time

It is assumed that the running time between bus stops is exogenously given; the standard bus dwelling time at each stop is given. However, the actual dwelling time at each bus stop is affected by the number of passengers alighting from and boarding the service. For a given node pair in a given route, the total in-vehicle travel time is determined by the running time plus dwelling times at the stops located between these nodes. Therefore, the in-vehicle travel time for a limited-stop service is determined by its line setting, i.e., which stops to be served by the limited stop service; whereas for the normal service which serves all intermediate nodes, the in-vehicle travel time is only affected by the number of passengers alighting from and boarding the normal service at each bus stop.

3.2.3.3. Choice behaviour

The model formulation captures choice behaviour by using the attractive/common line approach (Chriqui and Robillard, 1975). It assumes that passengers consider only a subset of lines serving a pair of nodes and the first arriving bus among the subset of lines is chosen by the passengers. The factors primarily affecting bus service choice include service waiting time and in-vehicle travel time of the lines.

The model formulation also takes bus service capacity into account to assign passengers to each of the operating services.

3.2.3.4. Total costs

As mentioned in Leiva et al. (2010), the total costs for the bus corridor comprise of *operator costs* and *user costs*. *Operator costs* in our model formulation are a combination of (i) *bus ownership costs* which is basically the cost of owning or renting a bus for operation on a particular service (the total ownership cost of a particular service is the product of the number of buses and the unit ownership cost of the type of bus given by $K_l n_l$) (ii) *bus operating costs* per unit time which accounts for variable operation costs such as employment costs, taxes, licenses and insurance (total operating cost for a service l can be given by $F^l f^l$). As regards the *user costs*, they are a combination of *passenger waiting time costs* and *passenger travel time costs* over route sections for each O-D pair.

3.2.3.5. Certain definitions related to transit services

Definitions of certain terms related to transit network design are mentioned as in Gao et al. (2004). A “transit line” or a “line” is a group of vehicles that run back and forth between two nodes on the transit network. All vehicles in the same line have identical size, capacity, operating characteristics and travel on the network going always through the same sequence of network links and nodes, referred to as an “itinerary”. A “transit route” is any path that a transit user can follow on the transit network in order to travel between any two nodes. In general, it will be identified by a sequence of nodes, the first one being the origin of the trip, the final being the destination and all the intermediate nodes represent transfer points. A “route section” is the portion of a route between two consecutive transfer nodes. Each route section is associated with a set of “attractive lines” which is called “common lines”.

3.2.4. General mathematical constraint for a limited stop service

In this section, the service network design of the limited stop service is modelled through a mathematical formulation. Consider a transit network schematic in Fig. 3.1 as an example. For the limited stop service l_s , the possible direct service ($y_{ij}^{l_s}$) between any node pair (i, j) can be shown as below in Fig. 3.2:

Fig. 3.2 Possible direct limited stop services between different node pairs in the transit network

The design of the limited stop service can be modelled by the following conditions:

$$y_p^{i,l_s} = \sum_{j>i} y_{ij}^{l_s} \leq 1, \forall i \in N \setminus \{1, b\}, \forall l_s \in L'. \quad (3.1)$$

$$y_q^{j,l_s} = \sum_{i<j} y_{ij}^{l_s} \leq 1, \forall j \in N \setminus \{1, b\}, \forall l_s \in L'. \quad (3.2)$$

$$y_p^{1,l_s} = 1, y_p^{b,l_s} = 0, y_q^{1,l_s} = 0, y_q^{b,l_s} = 1, \forall l_s \in L'. \quad (3.3)$$

$$y_p^{i,l_s} = y_q^{i,l_s}, \forall i \in N \setminus \{1, b\}. \quad (3.4)$$

$$y_{ij}^{l_s} \in \{0, 1\}, \forall l_s \in L'. \quad (3.5)$$

The binary variable $y_{ij}^{l_s}$ represents the direct limited stop service which takes the value 1 if there is a direct limited stop service between node pair (i, j) with no intermediate stoppages in between and 0 otherwise. The variable y_p^{i,l_s} defined in constraint (3.1) describes whether there is an outgoing limited stop service from this particular node i towards any node j located further in the network. Similarly, the variable y_q^{j,l_s} defined in constraint (3.2) states whether there is an incoming limited stop service to any node j . Eq. (3.3) ensures that the limited stop service line starts from the first node 1 and ends its service at the last node b . Hence, at any intermediate node between the first and the last node, the maximum value of the outgoing flow variable y_p^{i,l_s} and incoming flow variable y_q^{i,l_s} is equal to 1 for a certain limited stop service l_s as defined in constraints (3.1) and (3.2). Constraint (3.4) entails the service flow conservation at all intermediate nodes located between

the origin node and the terminal node in the network. It should be noted that although not explicitly defined, y_p^{i,l_s} and y_q^{i,l_s} are indeed binary variables. Constraint (3.5) defines that $y_{ij}^{l_s}$ is a binary variable.

3.2.5. Model formulation for the design of limited stop services

3.2.5.1. Problem Description

In this section, a methodology is presented to design the limited stop services operating in conjugation with the normal service in the network in terms of line setting, operating frequency and the fleet size with the attractive/common line approach. The problem description in this study is similar to that presented in Leiva et al. (2010), however, this study presents a more comprehensive model formulation to design limited stop services. To describe the passengers' service line choices, the concept of route sections (e.g., De Cea and Fernandez (1993)) is applied (each route section connecting nodes i,j denoted as ij). The normal service operates over all route sections throughout the network while a limited stop service operates over certain route sections depending on its line setting. At a particular node, a passenger intending to travel on a route section in the direction towards the destination node only boards a service that is attractive over that route section. As was defined in Chriqui and Robillard (1975), the attractiveness of a particular service over a route section is determined by the total expected travel time including the waiting time and in-vehicle travel time. Allowing for the transfers, the model assumes that passengers have the liberty to travel in stages over route sections towards their respective destinations. Starting from his/her origin node, a passenger boards the first bus from the set of attractive lines over a certain route section and transfers to the next route section until he/she reaches the destination node.

Fig. 3.3 Example transit network with line setting for limited stop service

Consider an example bus transit line setting of a limited stop service as shown in Fig. 3.3, wherein l_0 is the normal service while lines l_1 and l_2 are two limited stop services with line setting as 1->4 and 1->3->4, respectively. Table 3.1 shows the set of lines serving different route sections as per the given transit line settings. A passenger can choose one of these lines to travel over the route section depending on whether it is attractive or not. The attractiveness condition entails that a particular line serving a route section is attractive when the passenger's travel time cost over this line is smaller than the combined travelling and waiting time cost of all other lines serving the same route section.

Table 3.1 Route sections and operating lines

Route section	1-2	1-3	1-4	2-3	2-4	3-4
Lines	l_0	l_0, l_2	l_0, l_1, l_2	l_0	l_0	l_0, l_2

De Cea and Fernandez (1993) presented a hyperbolic programming problem to find the set of common lines operating over a route section. If A_s constitutes the set of transit lines operating over a route section s and (t^l, f^l) defines the in-vehicle travel time and frequency of a bus line l operating over s , then the following optimization problem determines the set of common lines on route section s

$$\min_{\{x^l\}} \frac{1 + \sum_{l=1}^{k'} t^l f^l x^l}{\sum_{l=1}^{k'} f^l x^l} \quad (3.6)$$

subject to $x^l \in \{0,1\}, \forall l \in A_s$, where $k' = |A_s|$ is the number of lines over s and x^l equals one if line l is an attractive line to the passengers and zero otherwise.

3.2.5.2. Model Formulation

The model formulation for the design of limited stop services over a transit network is defined through an optimization problem as given below.

$$\begin{aligned}
\min_{\{n_l, f^l, y_{ij}^{l_s}\}} Z = & \sum_{l \in L} K_l n_l + \sum_{l \in L} F^l f^l + \sum_{w \in W} \sum_{ij \in S_{ij}} V_{ij}^w \frac{W_c k}{\sum_{l \in L} x_{ij}^l f^l} + \sum_{w \in W} \sum_{ij \in S_{ij}} \sum_{l \in L} T_c v_{ij}^{w,l} t_{ij}^l \\
& + \lambda_{trans} \sum_{w \in W} \left(\sum_{ij \in S_{ij}} V_{ij}^w - X^w \right) - \sum_{w \in W} \sum_{ij \in S_{ij}} \sum_{l \in L} v_{ij}^{w,l} \gamma_{ij}^l
\end{aligned} \tag{3.7}$$

Subject to:

$$\sum_{l \in L} n_l \leq B. \tag{3.8}$$

$$t^{l,cycle} f^l \leq n_l, \forall l \in L. \tag{3.9}$$

$$v_{ij}^{w,l_{ij}} = V_{ij}^w \frac{x_{ij}^{l_{ij}} f^{l_{ij}}}{\sum_{l_{ij} \in L} x_{ij}^{l_{ij}} f^{l_{ij}}}, \forall l_{ij} \in L, \forall ij \in S_{ij}, \forall w \in W. \tag{3.10}$$

$$\sum_{ij \in S_i^+} V_{ij}^{w(i,j)} - \sum_{ij \in S_i^-} V_{ij}^{w(i,j)} = \begin{cases} X^{w(i,j)}; i = d \\ -X^{w(i,j)}; i = e \\ 0; otherwise \end{cases}, \forall w(i,j) \in W. \tag{3.11}$$

$$\sum_{i \in N: y_p^{i,l_s} = 1} \sum_{j > i \in N: y_q^{j,l_s} = 1} \sum_{w \in W} v_{ij}^{w,l_s} \leq f^{l_s} Cap_{l_s}, \forall l_s \in L', \forall k \in N. \tag{3.12}$$

$$\sum_{i \in N} \sum_{j > i \in N} \sum_{w \in W} v_{ij}^{w,l_r} \leq f^{l_r} Cap_{l_r}, \forall l_r \in L'', \forall k \in N.$$

$$y_p^{i,l_s} = \sum_{j > i} y_{ij}^{l_s} \leq 1, \forall i \in N \setminus \{1, b\}, \forall l_s \in L'. \tag{3.13}$$

$$y_q^{j,l_s} = \sum_{i < j} y_{ij}^{l_s} \leq 1, \forall j \in N \setminus \{1, b\}, \forall l_s \in L'. \tag{3.14}$$

$$y_p^{1,l_s} = 1, y_p^{b,l_s} = 0, y_q^{1,l_s} = 0, y_q^{b,l_s} = 1, \forall l_s \in L'. \tag{3.15}$$

$$y_p^{i,l_s} = y_q^{i,l_s}, \forall i \in N \setminus \{1, b\}. \tag{3.16}$$

$$t_{ij}^{l_s} = T_{ij} + \sum_{h=i+1}^{h=j-1} y_p^{h,l_s} T^h, \forall l_s \in L', \forall ij \in S_{ij}. \tag{3.17}$$

$$t_{ij}^{l_r} = T_{ij} + \sum_{h=i+1}^{h=j-1} T^h, \forall l_r \in L'', \forall ij \in S_{ij}. \tag{3.18}$$

$$x_{ij}^{l_{ij}} = 1 \Leftrightarrow T_c t_{ij}^{l_{ij}} \leq \frac{k W_c + T_c \sum_{h \in L \neq l_{ij}} t_{ij}^{h_{ij}} f^{k_{ij}}}{\sum_{h_{ij} \in L \neq l_{ij}} f^{h_{ij}}}, \forall l_{ij} \in L, \forall ij \in S_{ij}. \tag{3.19}$$

$$x_{ij}^{l_s} \leq y_p^{i,l_s} y_q^{j,l_s}, \forall l_s \in L', \forall ij \in S_{ij}. \quad (3.20)$$

$$f^l \geq 0, \forall l \in L; V_{ij}^{w(i,j)} \geq 0, \forall ij \in S_{ij}, \forall w \in W. \quad (3.21)$$

$$x_{ij}^{l_s} \in \{0,1\}, \forall l_{ij} \in L, \forall ij \in S_{ij}; y_{ij}^{l_s} \in \{0,1\}, \forall ij \in S_{ij}, \forall l_s \in L'. \quad (3.22)$$

The proposed model is a MINLP and non-convex by nature with the nonlinearity arising from the terms in the objective function and a few constraints. Both binary and continuous decision variables are used in the model which we will now analyse in detail. In the objective function (3.7), the first term is the total ownership cost of the fleet size allocated to all services; the second term is the total operating cost of all services, proportional to the line service frequency; the third term is the waiting time cost for all passengers, the fourth term is the passenger travel time cost, the fifth term is the penalty for transfers and the last term accounts for revenues (as negative cost) in terms of fares. Constraint (3.8) states that the total fleet size of all operating lines is lesser than or equal to the available fleet size. Constraint (3.9) ensures that the fleet size assigned to each service can fulfil the service frequency requirement. It should be noted that the total travel time for the full cycle of the bus service $t^{l,cycle}$ for limited-stop service is determined by the optimal design of the subset of stops to be visited, while for normal service, $t^{l,cycle}$ is fixed, as the in-vehicle traveling time between bus stops and bus dwelling time at bus stops are both assumed to be fixed and exogenously given. Indeed, $t^{l,cycle}$ can be defined by equations (3.17) and (3.18), by letting i and j represent the first and last bus stop nodes, respectively. Equation (3.10) computes the passenger flow on individual lines over a route section depending on whether they are attractive or not. Passenger flow is assigned to a line only if it is attractive over the route section (i.e., when $x_{ij}^l = 1$). Constraint (3.11) defines demand conservation at all nodes of the network. Constraint (3.12) states that for each of the limited stop services and normal services, the total combined capacity of the service on all route sections ij originating before and at any node k and serving various O-D pairs w should be greater than the demand for the service. Constraints (3.13), (3.14), and (3.15) describe the binary variables for transit line setting of the limited stop service lines

as in conditions (3.1)-(3.3). Constraint (3.13) defines the binary variable of y_p^{i,l_s} that is whether there is an outgoing limited stop service from any bus stop i . The same logic applies in constraint (3.14) for an incoming limited stop service into any bus stop node j . Constraint (3.15) states that y_p^{i,l_s} and y_q^{i,l_s} must take values (1,0) and (0,1) for the origin and terminal nodes of the service network, respectively. Constraint (3.16) is the service flow conservation condition as in condition (3.4). Constraints (3.17) and (3.18) compute the travel time over route sections for the limited stop and the normal services respectively. Constraint (3.19) is used to determine the attractive lines over a route section as in the common line problem. It is assumed that the passengers choose a subset of attractive lines with the minimum expected total travel time cost including both waiting time and in-vehicle travel time to travel towards their destination. Constraint (3.19) implies that if a transit line is attractive over a route section, then the travel time cost on that line must be less than or equal to the combined waiting and travel time cost of all other lines serving the same route section, and vice versa. A similar formulation can be found in Leiva et al. (2010). The constraint does not consider the case when the line is not attractive as passenger flow over a route section is assigned to only those lines which are found attractive, and hence, waiting time and travel time costs are computed with respect to only those lines. Constraint (3.20) states that if the bus stops i and j are not served by the limited stop service, it is certain that the limited stop service will not be attractive over route section ij as the limited stop service between the two stops does not even exist. In (3.20), whether the bus stop node pair (i, j) is served by the limited stop service is represented by the bilinear term $y_p^{i,l_s} y_q^{j,l_s}$; only when both y_p^{i,l_s} and y_q^{j,l_s} are equal to 1, the limited stop service can serve the two nodes i and j ; otherwise, passengers cannot take this limited stop service to travel from node i to node j which form route section ij . Hence, a limited stop service line is attractive over a route section only if it serves the route section and its travel time cost is lower than the combined waiting and travelling time costs of all other lines that operate over the same route section. Constraint (3.21) defines the non-negativity condition of the operating frequency of each line and the route section flow respectively. Constraint (3.22) defines the binary variables.

Hence, it should be noted that this model formulation is indeed a bi-level problem wherein the upper level problem is to design an optimal limited stop service operation strategy and the associated lower level problem is to find the attractive set of lines on each route section which describes the passengers' service line choice behaviour and also account for the trip assignment process. The lower-level problem is in essential a common-line problem assuming passengers choose a set of attractive lines so that the expected total travel time is minimized.

3.2.5.3. Incorporating the effects of alighting and boarding

The inclusion of limited stop service in this integrated transit system triggers transfers and hence, it would be interesting to study the effect of alighting and boarding on bus dwell time at nodes. Also, since the limited stop service stops at a few nodes; this concentrates the effect of alighting and boarding at those nodes leading to an increased travel time for the non-transferring passengers. When the demand is low, the effect of alighting and boarding can be conveniently ignored, however, incorporating their effects in case of higher demand would be significant towards a more precise computation of passenger travel time which affects their choice behaviour. Here, we re-define the travel time on the limited stop service and the normal service respectively by incorporating the effects of alighting and boarding.

$$t_{ij}^l = T_{ij} + \sum_{h=i+1}^{h=j-1} y_p^{h,l} T^h \alpha \max \left[\sum_{k < h} (v_{kh}^l), \sum_{h < k} (v_{hk}^l) \right], \alpha > 0 \quad (3.23)$$

$$t_{ij}^l = T_{ij} + \sum_{h=i+1}^{h=j-1} T^h \beta \max \left[\sum_{k < h} (v_{kh}^l), \sum_{h < k} (v_{hk}^l) \right], \beta > 0 \quad (3.24)$$

In the above equations (3.23) and (3.24), the travel time for the two services between a node pair (i, j) takes into account the effect of alighting and boarding. At each intermediate node h , the greater of the number of passengers alighting and the number of passengers boarding would determine the dwell time at h . The two positive coefficients α, β can be set empirically for the two services respectively.

3.2.6. Solution Method

As mentioned, the model formulation in the previous subsection is indeed a MINLP. Due to the inherent nonlinear and nonconvex property, it is very hard to solve the MINLP. In this study, we seek to obtain a global optimal solution of the

problem rather than only a local optimal solution. We first transform the nonlinear terms into linear ones by applying various linearization techniques so that the original MINLP can be transferred to a mixed integer linear program (MILP). Then, many existing solution algorithms like the branch and bound method can be used to solve the MILP which can guarantee a global optimal solution. One can notice that the nonlinearity of the model formulation arises from the objective function (3.7) and the nonlinear constraints (3.9), (3.10), (3.19), and (3.20).

3.2.6.1. Linearization with Reformulation-Linearization Technique (RLT)

Constraint (3.9) is linear when considering the normal service as the cycle time for the normal service is known, but it is nonlinear for the limited stop service due to the product of a continuous variable (service frequency) and the travel time which includes binary variables representing the incoming limited stop service as shown in constraint(3.17). Hence, a RLT as introduced in Sherali and Alameddine (1992) is used to represent this bilinear term through an equivalent set of linear conditions. Denote u_a as binary and $\tilde{x}_a$ as continuous such that $\underline{x}_a \leq \tilde{x}_a \leq \bar{x}_a$, where $\underline{x}_a$ and $\bar{x}_a$ are a sufficiently small positive number and a sufficiently large upper bound on $\tilde{x}_a$ respectively. If $x_a = u_a \tilde{x}_a$, the equivalent linear transformation of the bilinear term can be expressed as:

$$\begin{cases} x_a - u_a \underline{x}_a \geq 0 \\ x_a - u_a \bar{x}_a \leq 0 \\ x_a - \tilde{x}_a + \underline{x}_a - u_a \underline{x}_a \leq 0 \\ x_a - \tilde{x}_a + \bar{x}_a - u_a \bar{x}_a \geq 0. \end{cases} \quad (3.25)$$

Substituting equation (3.17) into constraint (3.9) for the case of limited stop services, the following can be obtained:

$$t^{l_s, cycle} f^{l_s} \leq n_{l_s} \Rightarrow (T_{ij} + \sum_{h=2}^{h=b-1} y_p^{h, l_s} T_s^h) f^{l_s} \leq n_{l_s} \quad (3.26)$$

Here, the nonlinearity of (3.26) arises from the product of service frequency variable f^{l_s} and binary variable y_p^{h, l_s} for the incoming limited stop service. Further substitution can be done as follows to represent this bilinear term $y_p^{h, l_s} f^{l_s}$:

$$g_p^{h,l_s} = y_p^{h,l_s} f^{l_s} \quad (3.27)$$

Using (3.27) in constraint (3.26), the resultant expression becomes:

$$T_{ij} f^{l_s} + \sum_{h=2}^{h=b-1} g_p^{h,l_s} T_s^h \leq n_{l_s} \quad (3.28)$$

Here, the bilinear term g_p^{h,l_s} as the product of service frequency variable f^{l_s} and binary variable y_p^{h,l_s} can be transformed into equivalent linear conditions as in (3.25). That is to say, the nonlinear constraint (3.9) is now completely converted into equivalent linear constraints by applying the RLT method. The same RLT method can be used to linearize the nonlinear travel time function as given in (3.23) and (3.24).

3.2.6.2. Linear transformation to linearize a product of binary variables

For nonlinearity in (3.20) involving the product of two binary variables, the following equivalent linear transformation can be applied for its linearization. Without loss of generality, if $A=cd$ where c and d are binary, then this bilinear term can be expressed equivalently by the following linear constraints which can be easily verified by enumerating all the possible cases as c and d are binary variables.

$$\begin{cases} A \leq c \\ A \leq d \\ A \geq c + d - 1. \end{cases} \quad (3.29)$$

3.2.6.3. Multidimensional piecewise linearization method

This section deals with the nonlinearity in (3.10) and the objective function. To solve the nonlinearity in constraint (3.10), a multidimensional piecewise linear approximation technique as proposed by Misener and Floudas (2009) is adopted. This is a global optimization algorithm based on a cutting constraint method employed to linearize the model into a MILP. Nonlinearity in more than one dimension is often complex to handle and for any nonlinear term, this method adopts a piecewise approximation technique to partition the domain into a number of small rectangles and thereafter finds the rectangle within which the optimal solution lies. This rectangle is referred to as the active rectangle. For instance, the constraint (3.10) is nonlinear in two dimensions and once this is linearized by using

this technique, the third term in the objective function $\sum_{w \in W} \sum_{ij \in S_{ij}} V_{ij}^w \frac{W_c^k}{\sum_{l \in L} x_{ij}^l f^l}$ can be reformulated into a linear form and the travel time cost term $\sum_{w \in W} \sum_{ij \in S_{ij}} \sum_{l \in L} T_c v_{ij}^{w,l} t_{ij}^l$ in the objective function can be further linearized by using the RLT as shown in (3.25).

Considering the objective function:

$$\min Z = \sum_{l \in L} K_l n_l + \sum_{l \in L} F^l f^l + \sum_{w \in W} \sum_{ij \in S_{ij}} V_{ij}^w \frac{W_c^k}{\sum_{l \in L} x_{ij}^l f^l} + \sum_{w \in W} \sum_{ij \in S_{ij}} \sum_{l \in L} T_c v_{ij}^{w,l} t_{ij}^l + \lambda_{trans} \sum_{w \in W} \sum_{ij \in S_{ij}} (\sum_{l \in L} V_{ij}^w - X^w) - \sum_{w \in W} \sum_{ij \in S_{ij}} \sum_{l \in L} v_{ij}^{w,l} \gamma_{ij}^l.$$

This can be reformulated as:

$$\min Z = \sum_{l \in L} K_l n_l + \sum_{l \in L} F^l f^l + \sum_{w \in W} \sum_{ij \in S_{ij}} V_{ij}^w \frac{W_c^k}{q_{ij}} + \sum_{w \in W} \sum_{ij \in S_{ij}} \sum_{l \in L} T_c v_{ij}^{w,l} t_{ij}^l + \lambda_{trans} \sum_{w \in W} \sum_{ij \in S_{ij}} (\sum_{l \in L} V_{ij}^w - X^w) - \sum_{w \in W} \sum_{ij \in S_{ij}} \sum_{l \in L} v_{ij}^{w,l} \gamma_{ij}^l. \quad (3.30)$$

In (3.28), we introduce new terms q_{ij}^l and q_{ij} to represent:

$$q_{ij}^l = x_{ij}^l f^l, \text{ and} \quad (3.31)$$

$$q_{ij} = \sum_{l \in L} q_{ij}^l = \sum_{l \in L} x_{ij}^l f^l. \quad (3.32)$$

Equations (3.31) and (3.32) can be linearized using the RLT approach as shown above in the previous sub-section. Based on (3.31) and (3.32), constraint (3.10) can be written as:

$$v_{ij}^{w,l} = V_{ij}^w \frac{q_{ij}^l}{q_{ij}}, \forall l \in L, \forall ij \in S_{ij}, \forall w \in W. \quad (3.33)$$

Obviously, constraint (3.33) is nonlinear as the right-hand side of this equation involves multiplication and division of multiple variables. Now, we will apply the multi-dimensional piecewise linearization method as proposed by Misener and Floudas (2009) to linearize the reformulated constraint (3.33). Consider the two variables, i.e., V_{ij}^w , passenger flow over route section ij for an O-D pair w , and q_{ij}^l , route section line frequency for line l , both of which fall into bounded intervals partitioned into M_l and N_l smaller segments. These intervals can be explicitly stated as below respectively:

$$V_{ij}^w \in [V_{ij}^{w,m-1}, V_{ij}^{w,m}], m=1, \dots, M_l; q_{ij}^l \in [q_{ij}^{l,n-1}, q_{ij}^{l,n}], n=1, \dots, N_l.$$

Also, the bounds of the feasible domain for the two variables are explicitly shown as follows:

$$\begin{aligned} 0 \leq V_{ij}^w \leq V_{ij}^{w,M_1} \quad \forall ij \in S_{ij}, \forall w \in W, \\ 0 \leq q_{ij}^l \leq q_{ij}^{l,N_1}, \forall ij \in S_{ij}, l \in L. \end{aligned} \quad (3.34)$$

The segments given by $[V_{ij}^{w,m-1}, V_{ij}^{w,m}], [q_{ij}^{l,n-1}, q_{ij}^{l,n}]$ are not necessarily equal in size. If M_l and N_l are sufficiently large such that the distance between any two consecutive points of each segment is very small, the true values of the following functions can be closely approximated by using piecewise linear functions. In this study, we can take the lower bounds of both the variables to be zero and the upper bound values to be equal to the total exogenous demand and maximum allowable route section frequency value respectively. Now, consider the following two nonlinear functions:

$$C_{ij}^w = \frac{V_{ij}^w}{q_{ij}}, \quad \forall ij \in S_{ij}, \forall w \in W, \quad (3.35)$$

$$\overline{C_{ij}^{l,w}} = \frac{V_{ij}^w}{q_{ij}^l}, \quad \forall l \in L, \forall ij \in S_{ij}, \forall w \in W, \quad (3.36)$$

The feasible domains of functions (3.35) and (3.36) cover the bounded intervals of the variables and are divided into $M_1 \times N_1$ rectangles. Each corner point (m, n) of these rectangles is associated with a particular value of variables defined in (3.35) and (3.36) which are explicitly computed by (3.38) and (3.39) below. Consider two sets of SOS1 variables (special ordered set of type 1 of which at most one variable is strictly positive whereas all others are at zero) i.e., S1 and S2 proposed by Beale and Tomlin (1970), to determine the active rectangle where the optimal values of the two variables are located.

$$\begin{aligned} S1: \mu_{ij}^m \in [0,1], m = 1, \dots, M_1, \\ S2: \nu_{ij}^n \in [0,1], n = 1, \dots, N_1. \end{aligned} \quad (3.37)$$

Each candidate rectangle has four corner points denoted by a set of co-ordinates (m,n) and the functions (3.35) and (3.36) can be expressed as :

$$C_{ij}^{w,(m,n)} = \frac{V_{ij}^{w,m}}{q_{ij}^n}, \quad (3.38)$$

$$\overline{C_{ij}^{l,w(m,n)}} = V_{ij}^{w,m} \frac{q_{ij}^{l,n}}{q_{ij}^n}. \quad (3.39)$$

A convex combination of these points is used to determine the value of the two functions within that rectangle. Denoting the coefficient of convex combination ranging between 0 and 1 by:

$$\rho_{ij}^{m,n} : \rho_{ij}^{m,n} \in [0,1], m=0,...,M_1, n=0,...,N_1; \sum_{m=0}^{M_1} \sum_{n=0}^{N_1} \rho_{ij}^{m,n} = 1. \quad (3.40)$$

Hence, with the above described, the following equations are used to conduct the two-dimensional piecewise linearization:

$$C_{ij}^{l,w} = \sum_{m=0}^{M_1} \sum_{n=0}^{N_1} \rho_{ij}^{m,n} C_{ij}^{l,w(m,n)}, \forall l \in L, \forall ij \in S_{ij}, \forall w \in W, \quad (3.41)$$

$$\overline{C_{ij}^{l,w}} = \sum_{m=0}^{M_1} \sum_{n=0}^{N_1} \rho_{ij}^{m,n} \overline{C_{ij}^{l,w(m,n)}}, \forall l \in L, \forall ij \in S_{ij}, \forall w \in W, \quad (3.42)$$

$$V_{ij}^w = \sum_{m=0}^{M_1} \sum_{n=0}^{N_1} \rho_{ij}^{m,n} V_{ij}^{w,m}, \forall ij \in S_{ij}, \forall w \in W, \quad (3.43)$$

$$q_{ij}^l = \sum_{m=0}^{M_1} \sum_{n=0}^{N_1} \rho_{ij}^{m,n} q_{ij}^{l,n}, \forall l \in L, \forall ij \in S_{ij}, \quad (3.44)$$

$$q_{ij}^n = \sum_{l \in L} q_{ij}^{l,n}, \forall ij \in S_{ij}, n=0,...,N_1, \quad (3.45)$$

$$\sum_{m=0}^{M_1} \sum_{n=0}^{N_1} \rho_{ij}^{m,n} = 1, \forall ij \in S_{ij}, \quad (3.46)$$

$$\rho_{ij}^{m,n} \in [0,1], m=1,...,M_1; n=1,...,N_1, \quad (3.47)$$

$$\sum_{n=0}^{N_1} \rho_{ij}^{0,n} \leq \mu_{ij}^1, \forall ij \in S_{ij}, \quad (3.48)$$

$$\sum_{n=0}^{N_1} \rho_{ij}^{m,n} \leq \mu_{ij}^m + \mu_{ij}^{m+1}, m=1,...,M_1-1, \forall ij \in S_{ij}, \quad (3.49)$$

$$\sum_{n=0}^{N_1} \rho_{ij}^{M_1,n} \leq \mu_{ij}^{M_1}, \mu_{ij}^m \in [0,1], m=1,...,M_1, \forall ij \in S_{ij}, \quad (3.50)$$

$$\sum_{m=0}^{M_1} \rho_{ij}^{m,0} \leq v_{ij}^1, \forall ij \in S_{ij}, \quad (3.51)$$

$$\sum_{m=0}^{M_1} \rho_{ij}^{m,n} \leq v_{ij}^n + v_{ij}^{n+1}, n=1,...,N_1-1, \forall ij \in S_{ij}, \quad (3.52)$$

$$\sum_{m=0}^{M_1} \rho_{ij}^{m,N_1} \leq v_{ij}^{N_1}, v_{ij}^n \in [0,1], n = 1, \dots, N_1, \forall ij \in S_{ij}. \quad (3.53)$$

In equations (3.41)-(3.44), the value of each of the variables $C_{ij}^{l,w}, \overline{C_{ij}^{l,w}}, V_{ij}^w, q_{ij}^l$ is computed as the convex combination of its values at the corner points of each of the $M_1 \times N_1$ rectangles that the domain was originally divided into. Equation (3.45) represents the combined frequency of all attractive lines serving a particular route section. Equations (3.48)-(3.53) lay the conditions of interdependence between the SOS1 variables (μ_{ij}^m, v_{ij}^n) and the coefficient for convex combination $\rho_{ij}^{m,n}$. Hence, the two nonlinear functions in (3.35) and (3.36) are linearized with this multidimensional piecewise linearization approach. In this way, the nonlinear constraint (3.10) is converted into linear constraints; the nonlinear terms in the reformulated objective function (3.30) are also linearized.

The fourth term of the reformulated objective function (3.30) involves the product of the service line passenger flow and the travel time. Using (3.36), this term can be expressed as:

$$\chi_{ij}^{l,w} = \overline{C_{ij}^{l,w}} t_{ij}^l \quad (3.54)$$

Since the travel time for the normal service is exogenously known, the product of line flow over normal service and travel time is linear. We only consider the product of the line flow and travel time of the limited stop service over a route section ij as below:

$$\chi_{ij}^{l_s,w} = \overline{C_{ij}^{l_s,w}} t_{ij}^{l_s} = \overline{C_{ij}^{l_s,w}} (T_{ij} + \sum_{h=i+1 \in N}^{h=j-1} y_p^{h,l_s} T_s^h). \quad (3.55)$$

In the above equation, the product of the passenger flow on line l and binary variable is nonlinear, but can be easily represented by equivalent linear conditions as is shown in (3.25).

3.2.6.4. Treatment for logic constraints

Further, constraint (3.19), describing the common line problem, is expressed in the form of logic constraints which cannot be tackled directly in a mathematical programming problem. Constraint (3.19) can also be written as:

$$x_{ij}^l = 1 \Leftrightarrow T_c t_{ij}^l \sum_{k \in L, k \neq l} f^k - (kW_c + T_c \sum_{k \in L, k \neq l} t_{ij}^k f^k) \leq 0, \forall l \in L, \forall ij \in S_{ij}. \quad (3.56)$$

This logical condition can be transformed into the following inequality constraint:

$$Lx_{ij}^l \leq (T_c t_{ij}^l \sum_{k \in L, k \neq l} f^k) - (kW_c + T_c \sum_{k \in L, k \neq l} t_{ij}^k f^k) \leq U(1 - x_{ij}^l) \quad (3.57)$$

In constraint (3.57), L and U are sufficiently large negative and positive constants, respectively. As per this constraint, when the binary variable x_{ij}^l is equal to 1, the left-hand side inequality is always true (as L is a sufficiently large negative constant) while only the right-hand side inequality is effective, entailing the condition for attractiveness of line l as in (3.56). Conversely, when the right-hand side inequality holds, to ensure constraint (3.57) being true, the binary variable has to be equal to 1. In another case, i.e., when the binary variable x_{ij}^l is equal to 0, the right-hand side inequality is always true (as U is a sufficiently large positive constant) while only the left-hand side inequality is effective, indicating the line l is not attractive in accordance with constraint (3.56). Hence, one can observe that constraint (3.57) is the equivalent reformulated inequality for constraint (3.19). The middle term in (3.57) can be easily transformed into equivalent linear constraints by applying the RLT approach as shown in (3.25). By doing so, the logic constraint (3.19) can be cast into an equivalent set of linear inequality constraints.

3.2.6.5. *Treatment for the travel time functions incorporating the effects of alighting and boarding*

The travel time functions for the limited stop service and the normal service incorporating the effects of alighting and boarding are given by:

$$t_{ij}^{l_s} = T_{ij} + \sum_{h=i+1}^{h=j-1} y_p^{h,l_s} T^h \max \left[\sum_{k < h} (v_{kh}^{l_s}), \sum_{h < k} (v_{hk}^{l_s}) \right], \alpha > 0;$$

$$t_{ij}^{l_r} = T_{ij} + \sum_{h=i+1}^{h=j-1} T^h \beta \max \left[\sum_{k < h} (v_{kh}^{l_r}), \sum_{h < k} (v_{hk}^{l_r}) \right], \beta > 0$$

Let $\sum_{k < h} (v_{kh}^{l_s}) = a$ and $\sum_{h < k} (v_{hk}^{l_s}) = b$ respectively. In the above travel time functions, the

“max” function implies $\max[a, b] = a$ if $a > b$ and $\max[a, b] = b$ if $a < b$. Let us

consider a binary variable $\omega \in \{0,1\}$ such that $\omega=1$ when $a > b$ and $\omega=0$ when $a < b$. This can be represented by the following mathematical expressions:

$$\max[a, b] = \omega a + (1 - \omega)b \quad (3.58)$$

$$U\omega + a(1 - \omega) < b < a\omega + L(1 - \omega) \quad (3.59)$$

where U and L are very large negative and positive integers respectively. Now, when ω takes the value of 1 in (3.59), then $U < b < a$ and the “max” function returns ‘ a ’ as the solution in (3.58). Similarly, when ω takes the value of zero, then $a < b < L$ and the “max” function returns ‘ b ’ in (3.58). The nonlinearity in (3.58) and (3.59) can be linearized using the RLT as shown in (3.25).

3.2.7. Reformulated problem

Hence, the original model formulation of this limited stop bus service network design problem has been completely transformed into a mixed integer linear program (MILP) in which the objective function and all the constraints are linear.

Specifically, the reformulated model can be expressed as:

$$\min Z = \sum_{l \in L} K_l n_l + \sum_{l \in L} F^l f^l + \sum_{w \in W} \sum_{ij \in S_{ij}} kW_c C_{ij}^w + \sum_{w \in W} \sum_{ij \in S_{ij}} \sum_{l \in L} T_c \chi_{ij}^{l,w} + \lambda_{trans} \sum_{w \in W} \left(\sum_{ij \in S_{ij}} V_{ij}^w - X^w \right) - \sum_{w \in W} \sum_{ij \in S_{ij}} \sum_{l \in L} \overline{C_{ij}^w} \gamma_{ij}^l \quad (3.60)$$

which is subject to constraints (3.8), (3.9), (3.11)-(3.18), (3.21)-(3.22), (3.28), (3.31)-(3.59).

The reformulated model with all the linearized constraints consists of the reformulated objective function (3.60) subject to 42 constraints along with 7 definitional variables, 3 decision variables and 9 internal variables used in mathematical operations for reformulation and subsequent linearization. As a result, the original model formulation is transformed into a MILP which can be solved by using many efficient solution algorithms like the branch and bound method, etc. Most importantly, the solution property of global optimality of the MILP is guaranteed.

3.2.8. Additional operational constraints

For the sake of operational efficiency, a few more constraints could be added to the presented model formulation. At times, the operator could face a requirement of

optimally selecting up to a fixed number of ‘special’ nodes (say P) for each operating limited stop service. Also, in case of multiple limited stop service lines, the operator might decide that a maximum of one limited stop service line serves a particular ‘special’ node such that the benefit of travelling over limited stop service is equitable over all the nodes in the transit network. Therefore, the following constraints could be added:

$$\sum_i y_p^{i,l_s} \leq P, \forall l_s \in L' \quad (3.61)$$

$$\sum_{l_s} y_p^{i,l_s} \leq 1, \forall i \in N \quad (3.62)$$

Constraint (3.61) states that for each limited stop service, there is a maximum limit of P number of ‘special’ nodes that need to be optimally selected. Constraint (3.62) states that a maximum of one limited stop service line can serve any intermediate node in the transit network.

3.2.9. Numerical studies for limited stop service design using deterministic user equilibrium

Fig. 3.4 Transit network for numerical study

Consider a network of 10 nodes as in Figure 3.4. From the example of the Singapore bus transit network, two kinds of bus services are considered, i.e., normal bus service similar to service 179 and limited stop service similar to service 179A. We assume that the line setting of the limited stop service, i.e., which stops to be served is upon the discretion of the operator. It is also assumed there is currently a bus service line l_0 (similar to 179) providing the normal service and up to two bus lines l_1 and l_2 (similar to 179A) with limited stop services are to be constructed. As stated before, the problem is to determine the optimal line setting for limited stop services l_1 and l_2 , as well as the operating frequencies and fleet size of all the transit lines while minimizing the total costs. We assume that nodes 6 and 10 are major attractors and hence, exogenous demand for these nodes is considered as given in the parameters below. However, a single corridor demand in the direction from

node 1 to node 10 is considered for transit network design in this study. Also, the fare structure for different services differs and hence, the fleet is split and the service frequency is determined accordingly to minimize the total cost in the objective function.

3.2.9.1. General Parameters

1) Bus capacities (number/bus): $l_0 = 60$ passengers/bus, $l_1 = 60$ passengers/bus, $l_2 = 60$ passengers/bus (a capacity of 60 passengers per bus is included to describe a single decker bus service, whereas, a capacity of 100 passengers/bus could be assumed for articulated buses)

2) Intrinsic demands at nodes for destination node 6 and node 10 are shown in Table 3.2 and Table 3.3, respectively:

Table 3.2 Intrinsic demand at each node for destination node 6 per hour

Node	1	2	3	4	5	6	7	8	9	10
Demand	75	65	40	25	15	0	0	0	0	0

Table 3.3 Intrinsic demand at each node for destination node 10 per hour

Node	1	2	3	4	5	6	7	8	9	10
Demand	60	40	20	20	15	30	35	40	35	0

3) Standard running time (minutes): For $j > i$ where (i, j) represent a node pair,
 $T_{ij} = 2(j - i)$

(Assumed to vary linearly with distance and equal inter-node spacing)

4) Operating cost per operational hour: $l_0 = \$280/\text{bus}$, $l_1 = \$200/\text{bus}$, $l_2 = \$250/\text{bus}$.

5) Ownership cost per operational hour: $l_0 = \$160/\text{bus}$, $l_1 = \$160/\text{bus}$, $l_2 = \$160/\text{bus}$.

6) Standard dwelling time at each node T^h (mins.): 1 min

7) Fare for service $l_1: \gamma_{ij}^{l_1} (\$) = 0.6(j - i)$, Fare for service $l_2: \gamma_{ij}^{l_2} = 0.72(j - i)$, Fare for service $l_0: \gamma_{ij}^{l_0} = 0.48(j - i)$ (Fares for limited stop service are assumed higher than those of normal services in general due to reduced travel time; fares between limited stop services could differ due to different on-board services offered).

8) Total available fleet size= 20 buses

9) Assuming Poisson's arrivals, $k = 1$, $\lambda_{trans} = 5\$$, $w_c = \$1/\text{min}$, $T_c = \$1/\text{min}$. (Poisson arrival process has been assumed only as a generalised model, however, the operator can design the operation based on any other probabilistic model as per discretion. The quantification of waiting/travel time is the discretion of the transit

operator and depends on several local factors such as economic condition of the bus transit users, quality of service, availability of alternative mode of transport, political scenario etc. Hence, from a modelling point of view, the quantification in the numerical study is merely a rational assumption.)

10) Maximum number of intermediate ‘special’ nodes: $l_1 = 2, l_2 = 3$

11) Number of nodes in the transit network $b = 10$.

3.2.9.2. Optimization Results

The model was evaluated by using the solver Gurobi on the programming platform YALMIP (Löfberg, 2004) interfaced with MATLAB on a Precision T1650 Dell PC, with a 3.20 GHz processor, 16 GB RAM, and a 64-bit operating system. In the numerical study, we first consider a transit network with one normal service and one limited stop service followed by another example with one normal service and two limited stop services. The results are then analysed and inferences are discussed.

3.2.9.2.1. Numerical example with one limited stop service

In this numerical example, we consider a single limited stop service l_1 operating in conjugation with the normal service l_0 . Using the input parameters from Section 3.2.9.1., the optimization model is solved. We add a realistic constraint as stated in constraint (3.61) that the total number of intermediate stops for the limited stop service l_1 cannot exceed a maximum number predetermined by the operator which is given by the following inequality:

$$\sum_{i=2}^{i=9} y_p^{i,l_1} \leq 2.$$

The optimal solution of this numerical example is listed as follows:

- 1) Optimal limited stop service pattern: 1->2->6->10
- 2) Service frequency: $l_1 = 5$ buses/hr, $l_0 = 10$ buses/hr
- 3) Fleet size: $l_1 = 6$ buses, $l_0 = 10$ buses
- 4) Total operating cost: \$612

Fig. 3.5 Optimized line setting for the limited stop service l_1

3.2.9.2.2. Numerical example with two limited stop services

In this numerical example, we assume that exactly two limited stop bus services l_1 and l_2 are provided along with the normal service l_0 . To further demonstrate that our model formulation is able to capture more realistic stipulations on the practical application of providing multiple limited stop service lines in a transit network, we impose other possible practical requirements to this problem that (1) the number of bus stops served by each limited stop service could not be larger than a certain value; (2) one bus stop would not be served simultaneously by the two limited stop services. This requirement could be entailed by adding two more linear constraints to the original model formulation as already described in constraints (3.61) and (3.62):

$$\sum_{i=2}^{i=9} y_p^{i,l_1} \leq 2; \sum_{i=2}^{i=9} y_p^{i,l_2} \leq 3,$$

$$y_p^{i,l_1} + y_p^{i,l_2} \leq 1, \forall i \in N \setminus \{1,10\}.$$

As per the above mentioned constraint, the limited stop services l_1 and l_2 are restricted to have a maximum of two and three intermediate stops respectively. The detailed model solutions are listed as below:

1) Optimal flow pattern:

$$l_1: 1 \rightarrow 2 \rightarrow 4 \rightarrow 10; l_2: 1 \rightarrow 3 \rightarrow 5 \rightarrow 6 \rightarrow 10.$$

Fig. 3.6 Optimized line setting for the limited stop service l_1

Fig. 3.7 Optimized line setting for the limited stop service l_2

2) Optimal service frequency for the transit lines in buses/hr: $l_1 = 5$ buses/hr, $l_2 = 6$ buses/hr, $l_0 = 5$ buses/hr.

3) Fleet size: $l_1 = 6$ buses, $l_2 = 6$ buses, $l_0 = 5$ buses.

4) Total cost: \$276 (savings)

3.2.10. Further discussion

To further investigate the quality of the solution obtained by the proposed methodology we conduct numerical studies on a few candidate transit line settings and observe the solution gap with the optimal solution. However, we should realize that the number of possible transit line settings can be huge and considering only a small set might not be sufficient to fully illustrate the efficacy of the model formulation. Let us consider the following two cases:

(i) For the case of single limited stop service operating with normal service: If the additional operational constraint (3.61) is considered with a total of 2 allowable stoppages, the possible combinations for the limited stop service l_1 can be 1->2->3->10, 1->2->4->10, 1->2->5->10 and so on which leads to a total of 28 possible combinations.

(ii) For the case of two limited stop services operating with the normal service: If the additional operational constraints (3.61) and (3.62) are considered, the possible combinations for l_1 and l_2 can be $\{(1->2->3->10), (1->4->5->10)\}, \{(1->2->3->10), (1->4->6->10)\}$, and so on. For each candidate transit line setting of l_1 , there exist 15 different possible combinations of l_2 leading to a total of 420 possible combinations.

We should realise that the number of possible combinations is huge even when we consider the additional operational constraints. This number increases manifold with more transit lines added to the network and relaxation of the additional constraints. Hence, we can understand that with just a few predetermined set of

transit line settings, it is not possible to find the best solution as the number of such combinations is fairly huge and cannot be ignored while searching for the global optimal solution.

In the study given below, we consider the numerical study in which a single limited stop service operates with the normal service. We also consider the additional constraint (3.61) with a maximum of two allowable stoppages for the line l_1 as mentioned in the parameters; hence the total operating cost of all candidate transit line settings apart from the optimal line setting is computed.

Optimal total operating cost $Z = \$612$;

Optimal transit line setting for l_1 : 1->2->6->10

Table 3.4 Total operating cost for corresponding candidate transit line setting of the limited stop service l_1

Solutions closest to optimal solution with corresponding transit line setting for l_1	Intermediate solutions with corresponding transit line setting for l_1	Solutions farthest from optimal solution with corresponding transit line setting for l_1
\$644; 1->2->3->10	\$1265; 1->4->5->10	\$1705; 1->6->8->10
\$770; 1->2->4->10	\$1326; 1->3->9->10	\$1585; 1->5->8->10
\$813; 1->4->6->10	\$1339; 1->3->8->10	\$1343; 1->4->8->10

As we can observe, the solutions obtained for the candidate transit line settings range from 644\$ to 1705\$ while the obtained optimal solution is 612\$. The solution gap between the total operating cost values for these candidate line settings and the optimal solution provides evidence for the quality of the solution while strongly advocating for the capability of the proposed model formulation to find the truly best global optimal solution.

3.3 Design of limited stop service using SUE

3.3.1. Problem description

In the previous section 3.2., it is assumed that the bus passengers follow the deterministic route choice behaviour principle. However, in reality, due to the lack of perfect information, the perceived service quality or utility function value of the bus services may be stochastic. In this case, stochastic user equilibrium principle should be applied to capture passengers' bus service choice behaviour. A similar transit network as schematically shown in Fig. 3.1 is considered. An exogenous O-D demand matrix is assumed to be known and given. In this sub-section, the model formulation considers a simplified scenario, wherein one limited stop service line and one normal service line are operating in the network and transfers are not allowed. Hence, passengers are assumed to select a particular service and continue their journey up to their destination node in the same service. The choice behaviour is captured through the logit model. At a node which is not served by the limited stop service, all demand towards further nodes in the network is allocated to the normal service.

3.3.2. Model Formulation

The model formulation for the design of limited stop service over a transit network using the SUE is defined through an optimization problem as given below.

$$\min_{n_l, f^l, y_{ij}^l} Z = \sum_{l \in L} K_l n_l + \sum_{l \in L} F^l f^l + \sum_{w(i,j) \in W} \left\{ \sum_{l \in L} [\xi_{w(i,j)}^l (T_c t_{ij}^l + W_c \frac{k}{f^l} - \gamma_{ij}^l)] + [(1 - y_p^{i,l_s} y_q^{j,l_s}) X^{w(i,j)} (T_c t_{ij}^{l_r} + W_c \frac{k}{f^{l_r}} - \gamma_{ij}^{l_r})] \right\} \quad (3.63)$$

subject to:

$$\sum_{l \in L} n_l \leq B. \quad (3.64)$$

$$y_p^{i,l_s} = \sum_{j>i} y_{ij}^{l_s}, \forall i \in N \setminus \{1, b\}. \quad (3.65)$$

$$y_q^{j,l_s} = \sum_{i<j} y_{ij}^{l_s}, \forall j \in N \setminus \{1, b\}. \quad (3.66)$$

$$y_p^{1,l_s} = 1, y_p^{b,l_s} = 0, y_q^{1,l_s} = 0, y_q^{b,l_s} = 1. \quad (3.67)$$

$$u_{ij}^l = \theta \gamma_{ij}^l + \psi t_{ij}^l + \omega \frac{k}{f^l}, \forall l \in L, \forall i, j \in N. \quad (3.68)$$

$$\xi_{w(i,j)}^l = y_p^{i,l_s} y_q^{j,l_s} \left(\frac{e^{-u_{ij}^l}}{\sum_{k \in L} e^{-u_{ij}^k}} \right) X^{w(i,j)}, \forall l \in L, w(i,j) \in W. \quad (3.69)$$

$$t_{ij}^{l_s} = T_{ij} + \sum_{h=i+1}^{h=j-1} y_p^{h,l_s} T^h, \forall i, j \in N. \quad (3.70)$$

$$t_{ij}^{l_r} = T_{ij} + \sum_{h=i+1}^{h=j-1} T^h, \forall i, j \in N. \quad (3.71)$$

$$\sum_{w(i,j) \in W} \xi_{w(i,j)}^{l_s} \leq Cap_{l_s} f^{l_s}. \quad (3.72)$$

$$\sum_{w(i,j) \in W} [\xi_{w(i,j)}^{l_r} + (1 - y_p^{i,l_s} y_q^{j,l_s}) X^{w(i,j)}] \leq Cap_{l_r} f^{l_r}. \quad (3.73)$$

$$t^{l,cycle} f^l \leq n_l, \forall l \in L. \quad (3.74)$$

$$y_{ij}^{l_s}, y_p^{i,l_s}, y_q^{j,l_s} \in \{0,1\} \forall i, j \in N. \quad (3.75)$$

$$t_{ij}^l \geq 0, \forall l \in L, \forall i, j \in N; \xi_{w(i,j)}^l \geq 0, \forall l \in L, \forall w(i,j) \in W. \quad (3.76)$$

$$n_l \geq 0, f^l \geq 0, \forall l \in L. \quad (3.77)$$

The objective function (3.63) seeks to minimize the total operation cost and the passenger travel time cost. The first term of the objective function is the ownership cost of the total bus fleet and the second term is the total frequency operating cost. The last term within curly brackets consists of two smaller terms denoting travelling time cost, waiting time cost and the revenue from fares (expressed as negative cost) for passengers using the two services for two different scenarios, i.e., first smaller term for the case when O-D pair (i,j) is served by the limited stop service and travel demand is split between the two services, and the second smaller term for the case when O-D pair (i,j) is not served by the limited stop service and the full demand is allocated to the normal service.

Constraint (3.64) states that the sum of fleet sizes of the two services should be lesser than or equal to the total available fleet size. Constraints (3.65)-(3.67) define service network for the limited stop service l_s as explained before in detail. Equation (3.68) defines the utility function for each service over each O-D pair that captures the service cost comprising of travel time, waiting time and fare. Constraint (3.69) is the demand split between the two services depending on whether an O-D pair is served by the limited stop service or not: if y_p^{i,l_s}, y_q^{j,l_s} equal to 1 simultaneously, which means the O-D pair (i,j) is served by the limited stop service, the passengers' service choice behaviour is captured through the logit model and the last term in the objective function is zero as both services would carry passengers between the O-D pair; if one of y_p^{i,l_s}, y_q^{j,l_s} is not equal to 1, which means the O-D pair (i,j) is not served by limited stop service, all the passenger demand between this O-D pair will be allocated to the normal service and hence, travel costs would be computed with regard to the normal service only.

Equation (3.70) computes the travel time between an O-D pair (i, j) served by the limited stop service l_s as the sum of the standard travel time between the O-D pair (i, j) without stopping and the dwelling time spent at the intermediate special nodes. Similarly, equation (3.71) calculates the travel time as per the normal service which would stop at all intermediate nodes between the O-D pair (i, j) . Constraint (3.72) states that the total demand for the limited stop service over all O-D pairs should be lesser than or equal to its total capacity. Similarly, constraint (3.73) is the capacity constraint for the normal service. The constraint (3.74) ensures that the fleet size assigned to each service is able to fulfil the service frequency requirement. Thereafter, constraint (3.75) specifies the binary variables and (3.76)-(3.77), specify the non-negativity of service travel time, service demand, fleet size and operating frequency respectively.

3.3.3. Solution Methodology

The stated model formulation is obviously an MINLP problem. The nonlinearity arises from the objective function (3.63), and constraints (3.68), (3.69), (3.73) and (3.74). In the objective function, the nonlinearities are from the three terms:

$$\xi_{w(i,j)}^l T_c t_{ij}^l (1 - y_p^{i,l_s} y_q^{j,l_s}) X^{(i,j)} W_c \frac{k}{f^{l_r}}, \xi_{w(i,j)}^l W_c \frac{k}{f^l}.$$

The constraint (3.68) is nonlinear due to the waiting time term in utility function (the reciprocal of the service frequency), constraint (3.69) involves the exponential function when describing the users' choice behaviour, constraint (3.73) is nonlinear due to the bilinear limited stop service term $y_p^{i,l_s} y_q^{j,l_s}$ and constraint (3.74) is the minimum fleet size constraint with bilinear term. Next, various linearization techniques are used to firstly transform the MINLP to a MILP problem, so that the solution could possess the property of global optimality. For the objective function, the first term mentioned above is linear when considered for the normal service as travel times for normal service are known; and when used for the limited stop service, it can be linearized by using RLT method as introduced in previous sections, as it involves the product of a binary and continuous variable. To linearize the second term, a piecewise linear approximation method is adopted to first tackle the nonlinear fractional part representing the waiting time and then its product with the binary variable for limited stop service can be linearized using RLT. For the last term, a multi-dimensional piecewise linear approximation method is adopted as it is nonlinear in two dimensions.

The constraint (3.68) can be linearized by using the one-dimensional piecewise linear approximation technique. In constraint (3.69), the nonlinear logit model will be linearized by applying various linear approximation techniques. Constraint (3.73) can be linearized using (3.29) and the constraint (3.74) is linearized by employing the RLT. Each of these linearization processes are detailed in the following sub-section.

3.3.3.1. Piecewise Linearization Techniques

Considering constraint (3.69), let us substitute:

$$R_{ij}^l = \left(\frac{e^{-u_{ij}^l}}{\sum_{k \in L} e^{-u_{ij}^k}} \right). \quad (3.78)$$

Now using (3.78), substitute $y = y_p^{i,l_s} y_q^{j,l_s}$ in (3.69):

$$yR_{ij}^l = Q_{ij}^l. \quad (3.79)$$

Indeed, the RLT method as introduced in previous section (constraint set 3.25) can be used for linearization of (3.79). Hence, constraint (3.69) now becomes:

$$\xi_{w(i,j)}^l = Q_{ij}^l X^{w(i,j)} \text{ which is a linear expression.} \quad (3.80)$$

Further, the exponential component in (3.69) can be linearized by using piecewise linear approximation technique as described below. Inversing both sides of (3.78):

$$\frac{1}{R_{ij}^{l_s}} = 1 + e^{u_{ij}^{l_s} - u_{ij}^{l_r}}, \frac{1}{R_{ij}^{l_r}} = 1 + e^{u_{ij}^{l_r} - u_{ij}^{l_s}}. \quad (3.81)$$

Applying logarithm:

$$\begin{aligned} \ln\left(\frac{1}{R_{ij}^{l_s}} - 1\right) &= \ln(e^{u_{ij}^{l_s} - u_{ij}^{l_r}}) \Rightarrow \ln\left(\frac{1}{R_{ij}^{l_s}} - 1\right) = u_{ij}^{l_s} - u_{ij}^{l_r}, \\ \ln\left(\frac{1}{R_{ij}^{l_r}} - 1\right) &= \ln(e^{u_{ij}^{l_r} - u_{ij}^{l_s}}) \Rightarrow \ln\left(\frac{1}{R_{ij}^{l_r}} - 1\right) = u_{ij}^{l_r} - u_{ij}^{l_s}. \end{aligned} \quad (3.82)-(i,ii)$$

which can be expressed as:

$$\begin{aligned} f(x_s) &= \ln\left(\frac{1}{x_s} - 1\right) = u_{ij}^{l_s} - u_{ij}^{l_r}, \\ f(x_r) &= \ln\left(\frac{1}{x_r} - 1\right) = u_{ij}^{l_r} - u_{ij}^{l_s}. \end{aligned} \quad (3.83)-(i,ii)$$

where $x_s = R_{ij}^{l_s}$, $x_r = R_{ij}^{l_r}$. Divide the range of the modal split (0, 1) into a number of known segments (equally spaced or otherwise) by using break points as shown:

$$(0, \alpha_1], [\alpha_1, \alpha_2], [\alpha_2, \alpha_3] \dots [\alpha_n, \alpha_{n+1}] \dots [\alpha_N, 1).$$

Hence:

$$\ln\left(\frac{1}{\alpha_{n+1}} - 1\right) < u_{ij}^{l_s} - u_{ij}^{l_r} \leq \ln\left(\frac{1}{\alpha_n} - 1\right) : 0 < \alpha_n \leq \alpha < \alpha_{n+1} < 1. \quad (3.84)$$

Therefore, for a particular value of α lying within an interval $0 < \alpha_n \leq \alpha < \alpha_{n+1} < 1$, a linear condition (3.84) has to be satisfied. A set of such constraints as in (3.84)

would cover the entire domain of break points i.e. (0, 1) and hence linearize the curve. A more detailed explanation on the piecewise linearization methodology can be found in Wang and Lo (2008). For the linearization of utility function equation (3.68), which is nonlinear due to the waiting time term (inverse of service frequency), a method of one-dimensional piecewise linear approximation of nonlinear functions proposed by Misener and Floudas (2009) is adopted. Let the domain of the function be defined by E , which is partitioned into an orthogonal grid that spans the domain. Now, any point $e \in E$ within this grid can be expressed as a convex combination of the grid points. However, the convex combination of these grid points is not necessarily unique. Further, every point located within this convex set $E \subset \mathbb{R}^n$ can be expressed as a convex combination of at most $n+1$ points of E . In this one-dimensional case, considering the domain of E as convex, any point in the convex space can be written as a convex combination of two points as $E \subset \mathbb{R}$. Although at most $n+1$ grid points would express any point $e \in E \subset \mathbb{R}^n$, there are more than $n+1$ grid points in any finite domain space. To ensure a unique interpolation of function values using convex combination of grid points, only $n+1$ grid points are activated each time. Hence, when $n+1$ grid points $e_0, e_1, \dots, e_n \in E$ for domain $e \in E \subset \mathbb{R}^n$ are activated, any function $f(e): E \rightarrow R$ can be approximated as:

$$f(e) = \frac{1}{e} = w_0 f(e_0) + w_1 f(e_1) + \dots + w_n f(e_n) \quad (3.85)$$

$$e = w_0 \cdot e_0 + \dots + w_n \cdot e_n, \quad (3.86)$$

$$\sum_{i=0}^n w_i = 1, \quad (3.87)$$

$$w_i \geq 0 \forall i = 0, \dots, n. \quad (3.88)$$

This gives a unique convex combination when the point $e \in E$ is in the interior of $n+1$ grid points; hence, the function $f(e): E \rightarrow R$ is interpolated between the $n+1$ grid points. Hence, the utility function can now be defined as

$u_{ij}^l = \theta \gamma_{ij}^l + \psi t_{ij}^l + k \omega g(f^l)$, such that

$$g(f^l) = \frac{1}{f^l} \quad (3.89)$$

which is expressed as a linear condition in (3.85).

Hence, the nonlinear terms in the utility function equations get linearized. The same procedure followed by RLT is used for linearizing the waiting time function in the objective function as it also involves the product of waiting time with the binary term as can be shown below:

In the term:

$$(1 - y_p^{i,l_s} y_q^{j,l_s}) X^{w(i,j)} W_c \frac{k}{f^{l_r}}, \text{ let } (1 - y_p^{i,l_s} y_q^{j,l_s}) = d \text{ which is obviously a binary variable,}$$

and $M' = dg(f^{l_r})$ which can be linearized using RLT. Hence,

$$(1 - y_p^{i,l_s} y_q^{j,l_s}) X^{w(i,j)} W_c \frac{k}{f^{l_r}} = M' X^{w(i,j)} W_c k. \quad (3.90)$$

For the total travel time involving the limited stop service in $\xi_{w(i,j)}^l t_{ij}^l$, the travel time

$$\text{is defined as } t_{ij}^{l_s} = T_{ij} + \sum_{h=i+1}^{h=j-1} y_p^{h,l_s} T_s^h$$

Substituting:

$$N_{ij}^{l_s} = \xi_{w(i,j)}^{l_s} T_c (T_{ij} + \sum_{h=i+1}^{h=j-1} y_p^{h,l_s} T_s^h) : r = \xi_{w(i,j)}^{l_s} y_p^{h,l_s} \quad (3.91)$$

where r can be easily linearized using RLT. Further, constraint (3.74) is similar to constraint (3.9) in the previous section and is nonlinear for the limited stop service case. Hence, it can be linearized by using the example of (3.26) which is linearized using the RLT.

For the term $\xi_{w(i,j)}^l \frac{k}{f^l}$, a two-dimensional piecewise linear approximation method

can be used for linearization as illustrated in section 3.2.6.3 of this chapter.

Suppose:

$$P_{ij}^l = kW_c \frac{\xi_{w(i,j)}^l}{f^l} \quad (3.92)$$

which is clearly nonlinear in two dimensions, the function (3.92) is similar to (3.33) and can be dealt with similarly. Finally, the objective function can be reformulated as:

$$\min Z = \sum_{l \in L} K_l n_l + \sum_{l \in L} F^l f^l + \sum_{w(i,j) \in W} \left\{ \sum_{l \in L} [N_{ij}^l + P_{ij}^l - \xi_{w(i,j)}^l \gamma_{ij}^l] + [dX^{w(i,j)} T_c^{l,r} + M' X^{w(i,j)} W_c k - dX^{w(i,j)} \gamma_{ij}^{l,r}] \right\} \quad (3.93)$$

subject to constraints (3.64)-(3.67), (3.70)-(3.73), (3.25)-(3.28), (3.75)-(3.77), (3.80), (3.84)-(3.88). The problem now is transformed into a MILP and can be solved by using an efficient solution algorithm like the branch and bound method. This indeed guarantees a globally optimal solution to the problem.

3.3.4. When transfers are allowed in the SUE model formulation

In the case when transfers are allowed, passengers travelling between a node pair (i, j) can complete their journey in different ways viz. using only the normal service, using only the limited stop service if both nodes i and j are served by the limited stop service and using a combination of both the normal and limited service by transferring at an intermediate node.

The objective here is to find the utility of each choice while travelling between a node pair (i, j) and then split the demand between the choices using the logit model. The utility of the normal service and the service involving limited stop service for passengers travelling between the O-D pair (either by only using limited stop service or a combination service with both normal and limited stop service through transfer) are $u_{ij}^{l,r}$ and u_{ij}^{comb} respectively. When transfer is considered in SUE choice behavioural assumption, one can imagine that quite a large number of routes are viable for passengers. To simplify the model formulation, the following assumptions are made:

- 1) A maximum of one transfer is allowed while travelling between a node pair (i, j) .
- 2) When a node pair (i, j) is served by the limited stop service, passengers travelling between this pair choose between two options – either using only normal service or

only a limited stop service similar to the case with no transfers. It is assumed that transfer imposes high disutility to the passengers and if passengers have two options of bus services without transfer, they will not select a travel route with transfer.

3) When only one of the nodes (i, j) is served by the limited stop service, passengers either use the normal service or a combination of services to travel between them.

4) When none of the nodes i and j of the node pair (i, j) are served by the limited stop service, passengers can only take the normal service as it is assumed that using a combination of services with two transfers is unacceptable for passengers. Now, the capacity constraints and the flow conservation are described. The capacity constraints for the normal and the limited stop service can be given by the following respectively:

$$\sum_{w \in W} \xi_{w(i,j)}^{l_r} \leq Cap_{l_r} f^{l_r} \quad (3.94)$$

$$\sum_{w \in W} \xi_{w(i,j)}^{comb} \leq Cap_{l_s} f^{l_s} + Cap_{l_r} f^{l_r}. \quad (3.95)$$

which state that the total demand for the normal service over all O-D pairs should be lesser than or equal to the total capacity of normal service; the total demand for the combination choice over all O-D pairs should be lesser than or equal to the total capacity of the normal service and the limited stop service as the combination choice utilizes both the services. Further, the flow conservation can be stated as below:

$$\xi_{w(i,j)}^{comb} + \xi_{w(i,j)}^{l_r} + \sum_{m>k} X^{w(k,m)} = \xi_{w(k,j)}^{comb} + \xi_{w(k,j)}^{l_r}, \forall i < k < j, \forall w(i, j) \in W \quad (3.96)$$

which means that for a particular O-D pair (i, j) at an intermediate node k , the sum of demands on the combination choice, the normal service and the total exogenous demand at node k is equal to the sum of demands on the combination choice and the normal service for the O-D pair (k, j) where j is any node located further in the network ($j > k$). Let us now consider three scenarios of passengers' route choices:

(i) When both the nodes of the OD pair (i, j) are directly served by the limited stop service with no intermediate special node: In this case, passengers are assumed to use the limited stop service or the normal service directly to reach their destinations

just as the case with no transfers. Therefore, the utility of the combination choice is equivalent to that of the limited stop service; therefore, the demand for combination choice computed through the logit model would be allocated to the limited stop service. Hence,

$$y_{ij}^{l_s} = 1 \Rightarrow u_{ij}^{comb} = u_{ij}^{l_s} : u_{ij}^{l_s} = \theta \gamma_{ij}^{l_s} + \psi t_{ij}^{l_s} + \omega \frac{k}{f_{ij}^{l_s}}, \forall i, j \in N. \quad (3.97)$$

Therefore, the demand split using the logit model can be expressed as below:

$$\begin{aligned} \xi_{w(k,j)}^{comb} &= \left(\frac{e^{-u_{kj}^{comb}}}{e^{-u_{kj}^{comb}} + e^{-u_{kj}^{l_r}}} \right) \left(\sum_{m>k} X^{w(k,m)} + \xi_{w(i,j)}^{comb} + \xi_{w(i,j)}^{l_r} \right), \forall i < k < j, \forall w(i,j) \in W; \\ \xi_{w(k,j)}^{l_r} &= \left(\frac{e^{-u_{kj}^{l_r}}}{e^{-u_{kj}^{comb}} + e^{-u_{kj}^{l_r}}} \right) \left(\sum_{m>k} X^{w(k,m)} + \xi_{w(i,j)}^{comb} + \xi_{w(i,j)}^{l_r} \right), \forall i < k < j, \forall w(i,j) \in W. \end{aligned} \quad (3.98)-(i,ii)$$

where $\xi_{w(k,j)}^{comb}, \xi_{w(k,j)}^{l_r}$ represent the passenger flow between nodes (k, j) using the combination choice and the normal service respectively.

(ii) When one or both the nodes (i, j) are served by the limited stop service with at least one intermediate special node:

$$\begin{aligned} (y_p^i + y_q^j) \sum_{i < k < j} y_p^k \geq 1 \Rightarrow t_{ij}^{comb} &= y_p^i y_q^j t_{ij}^{l_s} + (1 - y_p^i y_q^j) \{ \max_{i < k < j} (y_p^i t_{ik}^{l_s} + y_q^j t_{kj}^{l_s}) \\ &+ \min_{i < k < j} [(1 - y_p^i) y_p^k t_{ik}^{l_r} + (1 - y_q^j) y_q^k t_{kj}^{l_r}] \} \end{aligned} \quad (3.99)$$

This expression states that when both the nodes (i, j) are served by the limited stop service with at least one intermediate special node, passengers use the normal service and the limited stop service only. This is because it is assumed that passengers, apart from the normal service, prefer using the limited stop service over a combination of services when it connects the two nodes. But, in the case when only one of the nodes are served by the limited stop service, the travel route with transfer is selected such that passengers will travel the maximum distance of their journey on a limited stop service and minimum distance on the normal service while taking a maximum of one transfer in order to reach their destination in the least possible time. The utility of the combination can be expressed as:

$$u_{ij}^{comb} = \theta[\gamma_{ij}^{comb}] + \psi[t_{ij}^{comb}] + y_p^i y_q^j \omega \frac{k}{f^{l_s}} + (1 - y_p^i y_q^j) \omega [\frac{k}{f^{l_s}} + \frac{k}{f^{l_r}}] + \beta^t, \quad (3.100)$$

where β^t is the transfer penalty, γ_{ij}^{comb} is derived by replacing the service travel time term with the service fare term in the above logical condition, t_{ij}^{comb} is the travel time using the combination of services and the waiting time terms in the utility function are based on the number of services used depending on whether both nodes are served by the limited stop service. In case both the nodes are served, the waiting time is accounted for the limited stop service only whereas in the other case, the passengers wait for both services as they transfer at an intermediate node. Using the utility function values of the combination choice and the normal service, the logit model is used to compute the demand split.

(iii) When none of the nodes (i, j) are served by the limited stop service with/without intermediate special nodes or only one node is served with no intermediate special node: In such a case, only the normal service is used and hence there is no demand for the combination choice.

$$(y_p^i + y_q^j) \sum_{i < k < j} y_p^k = 0 \Rightarrow \xi_{w(i,j)}^{comb} = 0, \xi_{w(i,j)}^{l_r} = X^{w(i,j)}. \quad (3.101)$$

Hence, the reformulated objective function for the design of limited stop service using SUE formulation with transfers can be stated as below:

$$\begin{aligned} \min Z = & \sum_{l \in L} K_l n_l + \sum_{l \in L} F^l f^l + \sum_{w(i,j) \in W} \{ \xi_{w(i,j)}^{comb} (T_c^{l_{ij}^{comb}} + y_p^i y_q^j W_c \frac{k}{f^{l_s}} + (1 - y_p^i y_q^j) W_c [\frac{k}{f^{l_s}} + \frac{k}{f^{l_r}}] - \gamma_{ij}^{comb}) \\ & + \xi_{w(i,j)}^{l_r} (T_c^{l_{ij}^{l_r}} + W_c \frac{k}{f^{l_r}} - \gamma_{ij}^{l_r}) \}. \end{aligned} \quad (3.102)$$

where the first term is the total ownership cost, second term is the total cost of operating at the desired frequency and the third term is the sum of travelling, waiting costs and revenues over the two mode choices over all O-D pairs. For each O-D pair, the travel time cost, waiting time cost and revenue from fares (negative cost) are computed for each mode choice (combination choice/ normal service) and added to the overall operator's costs in the objective function.

3.3.5. Numerical study for limited stop service design using stochastic user equilibrium

Consider a transit network of 10 nodes as in Fig. 3.4. There are 2 lines operating of which line l_0 is the normal service and line l_1 is the limited stop service whose line setting is to be determined. These lines are loop services similar to 179 and 179 A in the Singaporean transit network wherein buses start at node 1, reach node 10 and then return to node 1. However, in this numerical study, a single corridor network wherein buses start at node 1 and operate until node 10 is considered. The problem is to determine the line setting for limited stop service l_1 , operating frequencies and fleet size of the two transit lines.

3.3.5.1. Parameters

The following set of data is assumed for the idealized transit network where nodes 1 and 10 denote the first and the last node of the network respectively.

- 1) Available total fleet size = 20 buses
- 2) Frequency operating cost per operational hour: $l_0 = \$20/\text{bus}$, $l_1 = \$10/\text{bus}$
- 3) Bus ownership cost per operational hour: $l_0 = \$40/\text{bus}$, $l_1 = \$50/\text{bus}$
- 4) Bus capacities: $l_0 = 50$ passengers/bus, $l_1 = 50$ passengers/bus
- 5) Inter-node travel fare: For $j > i$, $\gamma_{ij}^{l_1}(\$) = 0.3(j-i)$, $\gamma_{ij}^{l_0}(\$) = 0.2(j-i)$.
(Fare assumed to vary linearly with distance with equal inter-node spacing, fare for limited stop service higher than the normal service due to better service levels)
- 6) Standard inter-node running time without stopping: For $j > i$, $T_{ij}(\text{min}) = 2(j-1)$.
(Standard running time assumed to vary linearly with distance and equal inter-node spacing).
- 7) Utility function coefficients: $\theta=0.4$, $\psi=0.3$, $\omega=0.3$
- 8) Assuming Poisson's arrivals $k = 1$, $W_c = \$1/\text{hr}$, $T_c = \$1/\text{min}$.
- 9) Dwelling time at nodes (minutes): 1 minute
- 10) O-D demand per minute during peak hour: For $j > i$, $d_{ij} = 2i(j-i)$
(Assumed that certain nodes provide more patronage than others as per location in an urban setting)

3.3.5.2. Optimization Results

The model was evaluated using the solver Gurobi on the programming platform YALMIP (Löfberg, 2004) interfaced with MATLAB on a Precision T1650 Dell PC, 3.20 GHz processor, 16 GB RAM and 64-bit operating system.

Flow pattern obtained for fleet service l_I :

Fig. 3.8 Optimized line setting for the limited stop service l_I .

- 1) Flow pattern obtained for line l_I : 1->2->3->7->8->9->10
- 2) Fleet size: l_0 = 5 buses, l_I = 3 buses
- 3) Optimal frequency of lines in buses/min: l_0 = 9 buses/hr, l_I =8 buses/hr
- 4) Total operators' costs: \$10,215

3.3.5.3. Further analysis

When the demand and fares are changed, the total operating costs, fleet size and operating frequency of the services also change. Considering different demand patterns and fares, the following results are obtained:

Table 3.5 Optimal values of decision variables with changed demand

Demand w.r.t. original demand	Optimal service frequency	Optimal fleet size	Optimal total operating costs
50%	l_0 : 5 buses/hr, l_I : 4 buses/hr	l_0 : 2 buses, l_I : 3 buses	\$5,107
80%	l_0 : 8 buses/hr, l_I : 7 buses/hr	l_0 : 4 buses, l_I : 3 buses	\$8,172
130%	l_0 : 12 buses/hr, l_I : 10 buses/hr	l_0 : 6 buses, l_I : 4 buses	\$13,279
150%	l_0 : 14 buses/hr, l_I : 12 buses/hr	l_0 : 7 buses, l_I : 5 buses	\$15,322

Table 3.6 Optimal values of decision variables with changed fares

Fare change w.r.t. original	Optimal service	Optimal fleet size	Optimal total operating
--------------------------------	--------------------	-----------------------	----------------------------

fare structure	frequency		costs
50%	l_0 : 9 buses/hr, l_1 : 8 buses/hr,	l_0 : 5 buses, l_1 : 3 buses	\$10,935
80%	l_0 : 9 buses/hr, l_1 : 8 buses/hr,	l_0 : 5 buses, l_1 : 3 buses	\$10,503
130%	l_0 : 9 buses/hr, l_1 : 8 buses/hr,	l_0 : 5 buses, l_1 : 3 buses	\$9,782
150%	l_0 : 9 buses/hr, l_1 : 8 buses/hr	l_0 : 5 buses, l_1 : 3 buses	\$9,494

The findings can be summarized as follows:

- When demand increases, the total operating cost increases. Increasing fares reduce the total operating costs.
- The operating frequency of the limited stop service is lower than that of the normal service signifying the general attractiveness of the normal service.
- Change in demand affects the total operating costs much more than the fare, hence signifying the importance of patronage.

Fig. 3.9 Total cost with different demand and fare for the SUE model

In the above SUE numerical example in section 3.3.5, a single limited stop service is considered with the normal service in order to demonstrate the optimal transit line setting based on choice behaviour of the passengers, i.e. how passengers perceive the limited stop service against the normal service unlike the DUE case, where passengers form preferred choice sets over each route section and board the first vehicle that comes from that choice set. In both of the proposed model

formulations, it is to be noted that along with the user equilibrium being considered as a constraint which assigns the demand onto different routes, the operator incorporates the waiting and travelling time in its own costs along with the revenues earned. This is a comprehensive modelling of the operator's financial situation whilst considering the passengers' choice behaviour that affects demand for a particular service. Further, in the case with SUE wherein transfers are allowed, it must be noted that although there could be several route choices for an O-D pair with a given transit line setting of the limited stop service, a simplified approach with a few assumptions is presented to demonstrate the choice behaviour of passengers based on their rationality. The given numerical examples validate and strengthen confidence in the model formulations to be applied in real-time operation in different transit networks based on the local scenario and parameters which need to be suitably selected.

3.4. Conclusion

This study formulates a transit network design problem in which limited stop services are provided in presence of the normal service. From the transit operators' perspective, this model formulation is used to answer the question that, if limited stop services are offered to cater for more transit service demand, how the operators should determine optimal operations strategies in terms of fleet size assignment between limited stop services and normal services, service frequencies for both service types, and the line setting for the limited services, i.e., which stops are served by the limited stop services while the others are skipped. In capturing travellers' transit routing choice behaviour, the common line approach (DUE) and logit model (SUE) are applied to determine the passenger assignment on the transit network. Mathematically, the models presented are mixed integer nonlinear programs. Solution methods are developed to firstly transform the MINLP into an MILP with various linearization techniques and then solve the approximated mixed integer linear problem to attain a globally optimal solution by existing commercial package.

This research work contributes to the literature in the integrated transit network design with limited stop service by formulating the problem into an analytical

framework and developing a global optimization solution method to solve the problem. Since operating differential services is one viable strategy to improve the transit service quality, increase the transit service patronage, and alleviate traffic congestion, the proposed methodology provides a good decision making tool for transit operations if financially and socially sustainable bus transit services are desired. It can be readily applied to real-life transit networks. The solution method for the model formulations as described in the study leads to global optimality, which ensures the true globally best solution of the transit operation strategies. Besides, the methodology presented in this study can also be extended to solve similar complex system design problems in maritime studies, air-transportation, and various other systems engineering domains.

CHAPTER 4 A CONGESTION BASED BUS TRANSIT NETWORK DESIGN PROBLEM WITH COMMON LINES

4.1. Introduction and Problem description

De Cea and Fernandez (1993) presented a hyperbolic programming problem to find the set of common lines operating over a route section. Let A_s is the set of transit lines operating over a route section s . Suppose (t_l, f_l) defines the in-vehicle travel time and frequency of a bus line l operating over s . Then, an optimization problem defined by:

$$\min_{\{x_l\}} \frac{1 + \sum_{l=1}^k t_l f_l x_l}{\sum_{l=1}^k f_l x_l} \quad (4.1)$$

subject to $x_l \in \{0,1\}, \forall l \in A_s$ finds the set of common lines on the route section. Travel strategies are sets of common lines and are formed out of one or more transit lines considered attractive over a route section. Let us consider the following example in Fig. 4.1 where a route section AB is connected by n lines. There could be many strategies such as $\{1,2,3\}, \{1,2,3,4\}, \{1,2,3,4,5\}$ etc. depending on which lines are considered attractive over the particular route section.

Fig 4.1 A transit network route section A-B connected by transit lines.

Strategy time of a set of common lines is generally expressed as:

$$T_s = \frac{1 + \sum_{l \in A_s} t_l f_l}{\sum_{l \in A_s} f_l} \quad (4.2)$$

where T_s is the strategy time over route section s . Also, once a commuter has computed his/her strategy time, any line whose travel time is lower than this strategy time is also considered attractive. He/she then boards the first arriving bus from the revised set of attractive lines included in his/her strategy. Cominetti and Correa (2001) proposed that there could exist multiple travel strategies at equilibrium if their travel times are minimum and equal. Hence, referring to Fig. 4.1, it implies that at equilibrium, there is a possibility that both strategies $\{1,2,3\}$ and $\{1,2,3,4\}$ consisting of lines (1,2,3) and (1,2,3,4) respectively carry flows if the travel times associated with them are minimum and equal. This is in accordance with the Wardrop's first principle of user equilibrium. If P is the set of strategies over route section s , y_p is the strategy flow of a strategy p , T_p is the strategy travel time over the route section for a strategy p and $v(y)$ is the vector of line flows as per the strategy p , then the equilibrium strategy flow is given by:

$$y_p^* \geq 0 \Rightarrow T_p(v(y^*)) = T(v(y^*)), \text{ where } T(v) := \min_{p \in P} T_p(v) \quad (4.3)$$

The above condition (4.3) is called the *TEW equilibrium* and contributes towards the queue theoretic formulation of equilibrium strategy travel time in a congested transit network. For a network with several O-D pairs, each O-D pair is connected by multiple routes made up of route sections. For any route section s , let there be n bus lines as shown in Fig 4.1. For a particular value of flow over the route section A-B, a set of lines out of those n lines become attractive. In fact, as flow over the route section increases, more lines would become attractive as the waiting time for the faster lines increases to an extent that the slower lines become equally attractive. This is because more passengers would wish to take the faster lines causing more congestion and higher waiting times, thereby reducing their frequency. Hence, at equilibrium, depending on the route section flow value, the strategies are formed and passengers take the first bus from the set of lines included in the equilibrium strategy/strategies.

4.2. Methodology

4.2.1. Notations

Sets and parameters

A	Set of route sections in the transit network $a \in A$
W	Set of origin-destination (OD) pairs $w \in W$
R_w	Set of transit routes $p \in R_w, w \in W$
L	Set of transit lines $l \in L$
N	Set of nodes $i \in N$
δ	Route section/route incidence matrix, $\Delta = [\delta_{ap}^w], \forall a \in A, p \in R_w, w \in W$, where $\delta_{ap}^w = 1$ if route p between O-D pair w includes route section a and $\delta_{ap}^w = 0$ otherwise.
π_p^w	Minimum user travel and waiting cost on a path $p \in R_w$ for O-D pair $w \in W$
D	Vector of exogenous O-D demands $D = [d^w], w \in W$
Ω	Line/route section incidence matrix, $\Omega = [\Omega_a^l], \forall l \in L, \forall a \in A$, where $\Omega_a^l = 1$ if line l operates on route section $a \in A$ and $\Omega_a^l = 0$ otherwise
F^l	Operating cost per bus cycle on line l
U	Total available fleet size
O^l	Ownership cost of a bus on line l
η	Coefficient to convert route section flow to monetary terms
T^l	Full cycle time of a bus on line l
Cap_l	Capacity of a line $l \in L$
t_a^l	Vector of travel time on line l over route section $a \in A$ given by $t_a^l = [t_a^{l_1}, t_a^{l_2}, \dots, t_a^{l_m}, t_a^{l_{m+1}}, \dots, t_a^{l_{n_a^l}}]: t_a^{l_m} < t_a^{l_{m+1}},$
k	Parameter for distribution of bus arrival times
K^l	Vector of capacity of a line l given by $K^l = [K^{l_1}, K^{l_2}, \dots, K^{l_m}, K^{l_{m+1}}, \dots, K^{l_{n_a^l}}]$

n_a^l	Number of lines operating on route section $a \in A$
μ^l	Vector of Poisson's arrival intensity of a line l given by $\mu^l = [\mu^{l_1}, \mu^{l_2}, \dots, \mu^{l_m}, \mu^{l_{m+1}}, \dots, \mu^{l_{n_a^l}}]$

Definitional variables

$\rho_a^{l_m}$	Variable vector for a line l_m on route section $a \in A$ given by $\rho_a^l = [\rho_a^{l_1}, \rho_a^{l_2}, \dots, \rho_a^{l_m}, \rho_a^{l_{m+1}}, \dots, \rho_a^{l_{n_a^l}}]$ such that $\Omega_a^l = 1$
x_a	Total flow on route section a
c_a	Cost due to flow on route section a
$t_a(x_a)$	Strategy travel time at equilibrium on route section a
g_a	Service frequency of a route section a
h_p^w	Flow on a path p serving an O-D pair w
c_p^w	Cost due to flow on path p serving an O-D pair w
$\lambda_a^l, \psi_a^q, \alpha_a^q, r_a^q, s_a^q$	Parameters for determination of strategy travel time w.r.t. different route section flows

Decision variables

x_a^w	Flow on a route section a for O-D pair w
f_l	Frequency of transit line l
n_l	Fleet size allocated to any line l

4.2.2. Model Formulation

$$\min Z = \sum_{l \in L} (O^l n_l + F^l f_l) + \eta \left(\sum_{a \in A} c_a \sum_{w \in W} x_a^w \right) \quad (4.4)$$

subject to:

$$\sum_{l \in L} n_l \leq U \quad (4.5)$$

$$\sum_{a \in A_i^+} x_a^w - \sum_{a \in A_i^-} x_a^w = \begin{cases} d^w; i = Ori. \\ -d^w; i = Des. \\ 0; otherwise \end{cases}, \forall i \in N, \forall w \in W \quad (4.6)$$

$$h_p^w (c_p^w - \pi_p^w) = 0, \forall p \in R_w, \forall w \in W \quad (4.7)$$

$$c_p^w - \pi_p^w \geq 0, \forall p \in R_w, \forall w \in W \quad (4.8)$$

$$c_p^w = \sum_{a \in A} \delta_{ap}^w c_a, \forall p \in R_w, \forall w \in W \quad (4.9)$$

$$c_a = t_a(x_a) + \frac{k}{g_a}, \forall a \in A, \forall w \in W \quad (4.10)$$

$$x_a = \sum_{w \in W} x_a^w, \forall a \in A : x_a^w = \sum_{p \in R_w} \delta_{ap}^w h_p^w, \forall w \in W \quad (4.11)$$

$$\sum_{p \in R_w} h_p^w = d^w, \forall w \in W \quad (4.12)$$

$$g_a = \sum_{l \in L} \Omega_a^l f_l, \forall a \in A \quad (4.13)$$

$$T^l f_l \leq n_l, \forall l \in L \quad (4.14)$$

$$\sum_{w \in W} d^w \leq \sum_{l \in L} f_l Cap_l \quad (4.15)$$

$$0 \leq x_a \leq \sum_{l \in L} K^l \Omega_a^l \mu_a^l, \forall a \in A \quad (4.16)$$

$$\rho_a^{l_m}(x_a) \in [0,1) \Rightarrow \mu_a^{l_m}(\rho_a^{l_m} + \rho_a^{l_m^2} + \dots + \rho_a^{l_m^{k^{l_m}}}) = x_a, \forall \Omega_a^{l_m} = 1, \forall a \in A : n_a^l \geq 2 \quad (4.17)$$

$$t_a(x_a) = t_a^{l_1} + \frac{\rho_a^{l_1}(x_a)}{x_a(1 - \rho_a^{l_1}(x_a))}, \forall x_a \leq r_a^2, \quad (4.18)$$

$$t_a(x_a) = \left\langle \begin{array}{l} t_a^{l_q}, \forall r_a^q < x_a < s_a^q, \\ t_a^{l_i} + \sum_{i=1}^{q-1} \frac{\rho_a^{l_i}(\lambda_a^{l_i} x_a)}{x_a(1 - \rho_a^{l_i}(\lambda_a^{l_i} x_a))}, \forall s_a^q \leq x_a \leq r_a^{q+1}, \\ \text{where } t_a^{l_i} = \sum_{i=1}^q \lambda_a^{l_i} t_a^{l_i} \end{array} \right\rangle, \quad (4.19)$$

$$\forall a \in A: n_a^l \geq 2, q \in [2, \dots, n_a^l]$$

$$f_l \geq 0, n_l \geq 0, \forall l \in L \quad (4.20)$$

Parametric computations:

$$\lambda_a^{l_m} = \frac{\mu_a^{l_m}}{\sum_{j=1}^{n_a^l} \mu_a^{l_j}}, m \in [1, \dots, n_a^l] \quad (4.21)$$

$$\psi_a^q = \left[1 - \frac{1}{\sum_{j=1}^{q-1} \mu_a^{l_j} (t_a^{l_q} - t_a^{l_j})} \right]^{1/K_a^{l_q}} \quad (4.22)$$

$$q \in [2, \dots, n_a^l]$$

$$\forall a \in A: n_a^l \geq 2$$

$$\alpha_a^q = \frac{\psi_a^q}{1 - \psi_a^q}, q \in [2, \dots, n_a^l], \forall a \in A: n_a^l \geq 2 \quad (4.23)$$

$$\begin{aligned} r_a^q &= \sum_{i=1}^{q-1} \mu_a^{l_i} \alpha_a^q \left(1 - \frac{\alpha_a^q}{1 + \alpha_a^q} \right)^{K_a^{l_i}} \\ s_a^q &= \sum_{i=1}^q \mu_a^{l_i} \alpha_a^q \left(1 - \frac{\alpha_a^q}{1 + \alpha_a^q} \right)^{K_a^{l_i}} \end{aligned} \quad q \in [2, \dots, n_a^l] \quad (4.24)\text{-(i,ii)}$$

The objective function (4.4) is the overall costs to the operator. The first term is the combined cost of ownership of buses of all operating lines depending on their fleet size and cost of operating them at the designed frequencies whereas the second term is the cost of commuter flow on transit route sections. The constraint (4.5) states that the sum of fleet size of all services should be within the total available fleet size. Constraint (4.6) is the demand conservation at nodes. Constraint (4.7) and (4.8) are

the user equilibrium constraints expressed as complementary conditions in terms of route flows. It states that flows are assigned onto a route only if the route cost is the least in accordance with Wardrop's first principle. Also route cost is always greater than or equal to a given minimum cost. Constraint (4.9) is the route cost to the commuters expressed as the sum of costs over route sections that are included in the route. Constraint (4.10) is the route section cost in terms of route section travel time and waiting time. Constraint (4.11) states that the route section flow is the sum of flows on all routes in which it is included. Constraint (4.12) states that the demand for an O-D pair is equal to the sum of flows on all routes serving the O-D pair. Constraint (4.13) states that route section frequency is the sum of frequencies of lines serving the route section. Constraint (4.14) states that the fleet size of a line should be more than or equal to the number of buses needed to serve the network during the cycle time of the first leaving bus. Constraint (4.15) states that the frequencies of the bus lines should be sufficient to serve all demand during the operational period. Constraint (4.16) gives the maximum permissible value for the route section flow. For a given route section where two or more lines operate, the lines are initially arranged in a set in the order of increasing in-vehicle travel times over the route section. The model mentions four sets including in-vehicle travel times on the transit lines, their Poisson's arrival intensities, capacity of each line and another parameter for each line serving a particular route section $a \in A$.

Constraint (4.17) calculates the parameter ρ_a^l for a line l serving a route section a . The equations (4.18) and (4.19) are the strategy travel times as functions of route section flow. Consider a network of 5 nodes as shown below:

Fig 4.2 Transit network.

In the above transit network, there are 5 nodes served by three transit lines $L1$, $L2$, $L3$. The line $L1$ serves all the nodes, line $L2$ starts at node 2 and terminates at node 4 whereas line $L3$ starts at node 3 and ends at node 5. We can observe that the route

section 3-4 is served by all the 3 transit lines. The following plot can be constructed for the route section 3-4.

Fig 4.3 Route section travel strategy time

The above plot in Fig. 4.3 is an extrapolation from the results presented by Cominetti and Correa (2001) and mathematically expressed as in (4.18) and (4.19). It shows that the travel strategy time at equilibrium is a non-monotonic and nonlinear curve. When three transit lines $L1$, $L2$, $L3$ operate over the route section 3-4, t_2 and t_3 refer to the in-vehicle travel times over the route section 3-4 on lines 2 and 3, numbered in the order of increasing in-vehicle travel times with line 1 having the lowest in-vehicle travel time value over the route section ($t_1 < t_2 < t_3$). For a particular value of route section flow (less than a value r_a^2), the strategy time is an increasing function, then becomes constant with a value t_2 for the range (r_a^2, s_a^2) , again increases up to t_3 and so on. The route section equilibrium strategy time is included in the route section cost term given by constraint (4.10) along with the waiting time represented by the inverse of the route section frequency which is the sum of frequencies of all lines serving the route section.

Constraint (4.20) states that the service frequency and fleet size of an operating line are non-negative. Thereafter, equations (4.21)-(4.24) are parametric computations used to find the route section equilibrium strategy time which is a function of the equilibrium route section flow. These are taken from the results presented in Cominetti and Correa (2011). A variable q is used whose value lies between two

and the total number of lines operating on any route section a given by n_a^l . Depending on n_a^l , the set of values (r_a^q, s_a^q) are computed for each value of q . For e.g. if $n_a^l = 2$, q takes a value of two and therefore, only (r_a^2, s_a^2) is computed, whereas, if $n_a^l = 3$, q takes values of (2, 3) and therefore, both (r_a^2, s_a^2) and (r_a^3, s_a^3) are computed and so on. These sets determine flow ranges on the x-axis for the curve between the equilibrium route section strategy time and the equilibrium route section flow. If $n_a^l = 1$, there is only one line operating over a route section and hence, there is no need for computation of strategies.

4.2.3. Solution Algorithm

The presented model formulation is non-linear, non-convex. The non-linearity arises from the second term of the objective function (4.4) which depends on (4.10), constraints (4.7), (4.17), (4.18) and (4.19). The constraints (4.18) and (4.19) are linearized using reformulation techniques and piecewise approximation methods. Once these are linearized, the objective function gets linearized which directly depends on the terms included in these constraints. The constraint (4.7) is the user equilibrium complementary condition that is linearized using a set of mixed integer constraints. Once linearized, the model transforms into a Mixed Integer Linear Program (MILP) that can be easily solved using algorithms such as branch and bound and global optimality can be guaranteed.

4.2.3.1. Reformulation

Let us consider the condition (4.17) which is:

$$\rho_a^{l_m}(x_a) \in [0,1) \Rightarrow \mu_a^{l_m}(\rho_a^{l_m} + \rho_a^{l_m^2} + \dots + \rho_a^{l_m^{K^l}}) = x_a, \forall \Omega_a^{l_m} = 1, \forall a \in A : n_a^l \geq 2$$

If K^l is assumed to be large, then the geometric series can be reformulated as follows:

$$\rho_a^{l_m} + \rho_a^{l_m^2} + \dots + \rho_a^{l_m^{K^l}} = \frac{1}{1 - \rho_a^{l_m}} - 1 : \rho_a^{l_m} \in [0,1) \quad (4.25)$$

Therefore, the constraint (4.17) takes the form:

$$\mu_a^{l_m} \left(\frac{1}{1 - \rho_a^{l_m}} - 1 \right) = x_a \Rightarrow \mu_a^{l_m} \left(\frac{\rho_a^{l_m}}{1 - \rho_a^{l_m}} \right) = x_a \quad (4.26)$$

Hence, the nonlinearity in the equation (4.26) is basically due to the fractional term

$\frac{\rho_a^{l_m}}{1 - \rho_a^{l_m}}$ and once this is linearized, the equation (4.26) is linearized. Also using (4.10), the second term of the objective function is given by:

$$c_a x_a = x_a t_a(x_a) + x_a \frac{k}{g_a}, \forall a \in A \quad (4.27)$$

Both of the terms on the right hand side of the above equation are nonlinear. The term $x_a t_a(x_a)$ is the product of the route section flow with the route section travel time. The route section strategy travel time has a different function value for different route section flows as given by equations (4.18) and (4.19). In the route section strategy travel time equations, the travel time of an individual line over the route section is known, hence, the product of route section strategy time with the route section flow results in only one nonlinear fractional term which is $\frac{\rho_a^{l_i}(x_a)}{1 - \rho_a^{l_i}(x_a)}$

or $\frac{\rho_a^{l_i}(\lambda_a^{l_i} x_a)}{1 - \rho_a^{l_i}(\lambda_a^{l_i} x_a)}$ for equations (4.18) and (4.19) respectively. Therefore, the term $x_a t_a(x_a)$ and the equation (4.26) get linearized with the linearization of the abovementioned fractional terms.

4.2.3.2. One dimensional piecewise linearization

The term $\frac{\rho_a^{l_i}(\lambda_a^{l_i} x_a)}{1 - \rho_a^{l_i}(\lambda_a^{l_i} x_a)}$ can be expressed as $f(e) = \frac{e}{1 - e}$ where $e = \rho_a^{l_i}(x_a)$. To

linearize the function $f(e)$ which is nonlinear in one dimension, the one-dimensional piecewise linear approximation method is used. Let the domain of the function be defined by E which is partitioned into an orthogonal grid that spans the domain. Now, any point $e \in E$ within this grid can be expressed as a convex combination of the grid points. However, the convex combination of these grid points is not necessarily unique. Further, every point located within this convex set $E \subset \mathbb{R}^n$ can be expressed as a convex combination of at most $n + 1$ points of E . In

this one-dimensional case, considering the domain of E as convex, any point in the convex space can be written as a convex combination of two points as $E \subset \mathbb{R}$. Although at most $n+1$ grid points would express any point $e \in E \subset \mathbb{R}^n$, there are more than $n+1$ grid points in any finite domain space. To ensure a unique interpolation of function values using convex combination of grid points, only $n+1$ grid points are activated each time. Hence, when $n+1$ grid points $e_0, e_1, \dots, e_n \in E$ for domain $e \in E \subset \mathbb{R}^n$ are activated, any function $f(e): E \rightarrow R$ can be approximated as below. This gives a unique convex combination when the point $e \in E$ is in the interior of $n+1$ grid points; hence, the function $f(e): E \rightarrow R$ is interpolated between the $n+1$ grid points.

$$f(e) = w_0 f(e_0) + w_1 f(e_1) + \dots + w_n f(e_n) \quad (4.28)$$

$$e = w_0 e_0 + \dots + w_n e_n, \quad (4.29)$$

$$\sum_{i=0}^n w_i = 1, \quad (4.30)$$

$$w_i \geq 0 \forall i = 0, \dots, n. \quad (4.31)$$

4.2.3.3. Multi-dimensional piecewise linearization

Now, the nonlinearity in the objective function due to the term $x_a \frac{k}{g_a}$ which is nonlinear in two dimensions is solved. A piecewise multidimensional linear approximation technique as proposed by Luatthep et al. (2011) is used. Consider two variables viz. flow over route section a given by x_a and route section frequency given by g_a to lie within bounded intervals $[x_a^0, x_a^M]$ and $[g_a^0, g_a^N]$ respectively such that these intervals are partitioned into M and N small segments as $[x_a^{m-1}, x_a^m]$, $m=1, \dots, M$ and $[g_a^{n-1}, g_a^n]$, $n=1, \dots, N$. Also the bounds for the variables are as follows:

$$0 \leq x_a \leq x_a^M, 0 \leq g_a \leq g_a^N, \forall a \in A \quad (4.32)$$

These segments $[x_a^{m-1}, x_a^m]$, $m=1, \dots, M$ and $[g_a^{n-1}, g_a^n]$, $n=1, \dots, N$ are not

necessarily equal and if M and N are sufficiently large such that the distance between any two consecutive points of each segment is very small, then the true values of the following function can be closely approximated using piecewise linear functions. Take x_a^0 and g_a^0 to be zero each and the upper bound values x_a^M, g_a^N to be the total demand and maximum allowable route section frequency value respectively. Now consider the function:

$$C_a = \frac{x_a}{g_a} \quad (4.33)$$

The domain of function (4.33) covers bounded intervals $[x_a^0, x_a^M]$, $[g_a^0, g_a^N]$ and is divided into $M \times N$ rectangles. Each rectangle is then associated with values of $C_{ij}^{m,n}$ which is C_{ij} value at each corner point denoted by (m, n) . Consider two sets of SOS1 variables (special ordered set of type 1 of which at most one variable is positive whereas all others are zero) $S1$ and $S2$ proposed by (Beale and Tomlin(1970)) to determine the active rectangle where the optimal values of the variables would be found.

$$\begin{aligned} S1: \mu_{ij}^m &\in [0,1], \forall m = 1, \dots, M \\ S2: \nu_{ij}^n &\in [0,1], \forall n = 1, \dots, N. \end{aligned} \quad (4.34)$$

Each candidate rectangle has four corner points each denoted by a set of co-ordinates (m,n) where the function (4.33) can be expressed as :

$$C_a^{m,n} = \frac{x_a^m}{g_a^n}, \forall a \in A \quad (4.35)$$

A convex combination of these points is used to find the value of the function within that rectangle. Denoting the coefficient of convex combination ranging between 0 and 1 by

$$\gamma_a^{m,n} : \gamma_a^{m,n} \in [0,1], \forall m = 0, \dots, M, n = 0, \dots, N; \sum_{m=0}^M \sum_{n=0}^N \gamma_a^{m,n} = 1, \forall a \in A \quad (4.36)$$

Hence, the following equations for the two-dimensional piecewise linearization of function (4.35) can be formulated:

$$C_a = \sum_{m=0}^M \sum_{n=0}^N \gamma_a^{m,n} C_a^{m,n}, \forall a \in A \quad (4.37)$$

$$x_a = \sum_{m=0}^M \sum_{n=0}^N \gamma_a^{m,n} x_a^m, \quad \forall a \in A \quad (4.38)$$

$$g_a = \sum_{m=0}^M \sum_{n=0}^N \gamma_a^{m,n} g_a^n, \forall a \in A \quad (4.39)$$

$$\sum_{m=0}^M \sum_{n=0}^N \gamma_a^{m,n} = 1, \gamma_a^{m,n} \in [0,1], \quad \forall a \in A \quad (4.40)$$

$$\sum_{n=0}^N \gamma_a^{0,n} \leq \mu_a^1, \quad \forall a \in A \quad (4.41)$$

$$\sum_{n=0}^N \gamma_a^{m,n} \leq \mu_a^m + \mu_a^{m+1}, \quad \forall m=1, \dots, M-1, \quad \forall a \in A \quad (4.42)$$

$$\sum_{n=0}^N \gamma_a^{M,n} \leq \mu_a^M, \mu_a^m \in [0,1], \quad \forall m=1, \dots, M, \quad \forall a \in A \quad (4.43)$$

$$\sum_{m=0}^M \gamma_a^{m,0} \leq \nu_a^1, \quad \forall a \in A \quad (4.44)$$

$$\sum_{m=0}^M \gamma_a^{m,n} \leq \nu_a^n + \nu_a^{n+1}, \quad \forall n=1, \dots, N-1, \quad \forall a \in A \quad (4.45)$$

$$\sum_{m=0}^M \gamma_a^{m,N} \leq \nu_a^N, \nu_a^n \in [0,1], \quad \forall n=1, \dots, N, \quad \forall a \in A \quad (4.46)$$

The nonlinearity in objective function is now solved. For the constraint (4.7), the nonlinearity due to the equilibrium complementary conditions can be linearized using the following mixed integer conditions from Wang and Lo (2010).

$$L\sigma_p^w + \varepsilon \leq h_p^w \leq U(1 - \sigma_p^w) \quad (4.47)-(i,ii,iii,iv)$$

$$L\sigma_p^w \leq c_p^w - \pi_p^w \leq U\sigma_p^w$$

$$c_p^w - \pi_p^w \geq 0$$

$$\sigma_p^w \in \{0,1\}$$

where L and U are very large negative and positive integers, σ_p^w is a binary variable for a particular route p and O-D pair w , ε is a very small positive integer and other symbols are same as described before. Hence, the objective function and constraints are linearized converting the problem to a MILP. The model was evaluated using the solver Gurobi on the programming platform YALMIP (Löfberg, 2004) interfaced with MATLAB on a Precision T1650 Dell PC, 3.20 GHz processor, 16 GB RAM and 64-bit operating system.

4.3. Numerical Study

4.3.1. Transit Network

A section of the Singapore transit network consisting of four transit nodes located at Outram Park, SMU, Broadway and Michael Palace and numbered 1-4 respectively with four operating transit lines 124, 131, 23, 147 denoted as $L1$, $L2$, $L3$, $L4$ respectively is considered as given in Fig. 4.4. In this numerical study, only a single corridor demand in the direction from node 1 towards node 4 is considered. Also, a single O-D pair demand between nodes 1 and 4 is assumed to be exogenously given. Since transfers are allowed, passengers can transfer at any of the intermediate nodes to change to another transit line to ultimately reach the destination node 4. The other parameters are as given below:

1) Table 4.1: Route characteristics

Parameters\Lines	124	131	23	147
Operating cost/bus/hr	10\$	15\$	20\$	30\$
Ownership cost/bus/hr	40\$	50\$	60\$	60\$
Cycle time (mins)	50	60	75	95

2) Peak hourly demand D from node 1 to node 4= 1000 commuters

3) Bus line capacity: 50 commuters on all buses on all lines.

4) Available fleet: 50 buses

Fig 4.4 Transit Network

8) Table 4.2: Route section travel time (mins) on different lines

Lines\route section	1-2	1-3	1-4	2-3	2-4	3-4
<i>L1</i>	25	-	-	-	-	-
<i>L2</i>	-	-	-	20	-	-
<i>L3</i>	-	-	-	-	-	25
<i>L4</i>	15	20	42	15	35	20

9) $\eta = 1, k = 1$

4.3.2. Optimization results

1) Bus frequency vector in buses/min:

L1= 14 buses/hr, *L2*= 1bus/hr, *L3*= 1bus/hr, *L4*=4 buses/hr.

2) Equilibrium flow over route sections:

Table 4.3 Equilibrium route section flows

Route section	1-2	1-3	1-4	2-3	2-4	3-4
Flow	100	450	450	0	100	450

3) Bus fleet sizes: *L1*= 12 buses, *L2*= 1bus, *L3*= 2 buses, *L4*=7 buses

4.3.3. Further analysis

More numerical studies were performed with different demand values and the following results were obtained:

Table 4.4: Sensitivity analysis with respect to demand

Demand	Line frequency $\{L1, L2, L3, L4\}$	Fleet sizes $\{L1, L2, L3, L4\}$	Equilibrium route section flow	Total operating cost(\$)
D=1000	14,1,1,4	12,1,2,7	100, 450,450, 0, 100,450	191,770

D1=DX110%	16,1,1,4	14,1,2,7	200, 450, 450, 0, 200,450	252,880
D2=DX120%	18,1,1,4	15,1,2,7	300, 450, 450, 0, 300, 450	314,000
D3=DX130%	20,1,1,4	17,1,2,7	400, 450, 450, 0, 400, 450	375,110
D4=DX90%	12,1,1,4	10,1,2,7	0, 450, 450, 0, 0, 450	130,650
D5=DX80%	10,1,1,4	9,1,2,7	0,400,400, 0, 0, 400	116,180
D6=DX70%	8,1,1,4	7,1,2,7	233,233, 233, 0, 233,233	101,710

Fig 4.5 Plot of total op. cost vs demand

Fig 4.6 Percentage increase in total op. cost vs. increase in demand

Fig 4.7 Plot of equilibrium route section frequency vs. demand

Fig. 4.5 is the plot of total operating cost vs. the demand which shows that as the demand increases, the total operating cost also increases. However, the rate of increase of operating costs with increasing demand is not consistent. Fig. 4.6 shows the plot of the % increase in total operating costs versus the corresponding increase in demand pattern. It can be observed that there is a sharp rise in the % increase in operating costs when the demand changes from 90%D to D value which then gradually decreases with further increase in demand. This is the inflection demand range, wherein, an increase in demand necessitates the increased allocation of buses and service frequencies and any further increase in demand does not significantly increase the operating costs. Fig. 4.7 shows the equilibrium route section flow pattern for demand values. It can be observed that the equilibrium flow pattern changes with change in demand and also the following: a) The route section 2-3 is never travelled on; b) in case of 80%D and 90%D, only 3 route sections are used. One plausible reason for this could be that as demand increased from 70%D to 80%D, the waiting times to travel over certain route sections increased and hence they are not used. With a further increase in demand, these route sections become attractive at equilibrium as waiting times increase over all route sections. The change in equilibrium route choice with changing demand results in the corresponding change in service frequency and fleet size.

4.4. Conclusion

The model formulation presents a CTNDP wherein line frequencies and fleet size of bus lines are designed considering congestion effects. The common lines approach is used to determine the equilibrium route section strategy time which is non-monotonic, nonlinear and a function of the equilibrium route section flow. The MPEC formulation used is easier to solve mathematically and is computationally efficient. The presented methodology uses complementary constraints as user equilibrium conditions and a mixed integer approach linearizes them transforming the problem to a MILP. The commuters travel on only those paths which are assigned flows at equilibrium, hence, optimizing the overall transit performance. However, with change in demand, the line frequencies change in order to cater to the new demand. Also, due to change in income levels and other associated factors, a scope for changes in overall operation would have to be included in an optimal design framework.

CHAPTER 5 TRANSIT ROUTE PACKAGING FOR A PRIVATIZED BUS CONTRACTING REGIME

5.1. Introduction and Problem Description

This study proposes an analytical framework for the contracting authority to determine the transit routes that should be bundled together in the form of route packages and tendered out to the contesting operators under the bus contracting model through CT. The methodology would equip the authority with a strong analytical foundation for the overall decision making process making it financially and socially equitable for all stakeholders.

A route package can be defined as a set of routes that are allocated to a single or a group of operators under a transit operative model. In this research work, route packages would be designed as per the contracting model considering the overall costs to the authority, operator and the commuters such that each of them benefit from the system optimally. Let us focus on the aspirations of each of these entities individually. The authority planning the route packages and contracting the operation to private/government operators would need to allocate resources such as buses, equipment and also pay for the ownership of the bus depots. The route packages should be so designed such that there is reasonable difference in the packages' operating costs as that would encourage chances of new entrants into the market who could be averse towards bidding for packages incurring high operating costs. Moreover, this would induce more competition for price and quality as more players could contest in the bidding process. The package operating costs here refer to the sum of cost of operating the allocated fleet during the service hours including manpower disbursements, fuel costs etc. and the fleet maintenance costs. As regards commuters, route frequency should be designed such that their transit waiting time is minimized. The authority also needs to ascertain that route packages are designed such that each route is allocated to a single package only for efficient operation and of course, every package is to be allocated a minimum of one route on the basis of the above mentioned design criteria.

This study presents a mathematical modelling approach for bus transit route packaging. In a transit network, there exist several routes connecting different O-D pairs. Each of these routes has its specific demand, fare structure for travel between node pairs on the route, fleet size requirement, service frequency, operation costs etc. Here, a route package is designed as a set of transit routes with each package having an allocated bus bay. A transit package could be identified through the allocated bus bay which behaves as the control point. Under the government contracting model, the regulatory authority designs the package, allocates the fleet, determines the service frequency and hands over the package to the operator who gives the best proposal for the package operation in terms of quality and service fee during the CT process. The presented methodology designs the route packages as sets of routes that should be bundled together while allocating the bus fleet and designing the service frequency for each of these routes. During the contracting process, the operator who quotes the lowest service fee while fulfilling the quality criteria wins the right to operate the package. The methodology also designs route packages such that there is a considerable difference between the different route packages' operating costs as this would encourage entry of new transit operators.

5.2. Methodology

5.2.1 Assumptions

Let us state a few important assumptions involved in the route package design. It is assumed that:

- there is a single bus line operating on each route which is a normal service catering all nodes
- the number of transit routes and transit route packages to be designed are exogenously given
- a single route can be allocated to one route package only
- each route package has a minimum of one route with no upper limit on the number of routes
- each bus bay has a limited capacity to park buses

5.2.2. Variable Definitions

<i>Sets</i>	<i>Definition</i>
P	set of transit route packages $p \in P$
R	set of transit routes $r \in R$
<i>Parameters</i>	<i>Definition</i>
X^r	exogenous demand for route r
F_p^r	price rate of operating buses at a required frequency on route r included in package p per unit operational period
B_p^r	price rate of maintaining a bus on route r included in package p per unit operational period
O_p^r	price rate of bus ownership on route r included in package p per unit operational period
T_p^r	total cycle time for a bus on route r included in package p
w_c	waiting time cost
F	total available fleet size
Cap_b	capacity of a bus
n	number of transit routes
k	number of transit route packages
C_p	capacity of a bus bay allocated to a package p
D^r	distance covered by route r in kilometres
ψ_d	factor to convert unit distance difference in packages' distance spread to monetary terms

ξ_d factor to convert unit difference in total operating cost between packages to monetary terms

Definitional variable *Definition*

γ_p^r total associated costs on a route r included in package p

Decision Variables *Definition*

y_p^r binary decision variable for a route r included in package p

n_p^r number of buses allocated to route r included in package p

f_p^r service frequency of route r included in package p

5.2.3. A few important rationale

Given below are a few rationale on how different cost parameters on a given route r would change after packaging design. It is assumed that all these parameters are already calibrated and exogenously given.

(i) O_p^r : This defines the rate (cost per unit) of ownership of a bus operating on route r in a package p per unit operational period. Buses allocated to each route in a route package would also incur bus bay parking costs to the regulator depending on the location of bus bay (controlling the package operation) in an urban setting.

(ii) B_p^r : This defines the rate (cost per unit) of bus maintenance on a route r in package p per unit operational period. Depending on the packaging strategy, the bus maintenance rates would differ for different bus bays due to the logistical constraints and availability of resources for maintenance.

(iii) F_p^r : This refers to the rate (cost per unit) of operating buses at a required frequency on a route r in a package p per unit operational period. This rate would differ with packages as a high value of route frequency in a congested urban setting can disrupt the traffic in the vicinity. This cost also includes scheduling

charges, crew disbursements as a higher frequency on a route would imply a shorter lay-over time after each cycle.

(iv) T_p^r : Depending on the package to which a router r is allocated, the location of its designated bus bay would change. Hence, the buses serving the route would have different hauling times spent on covering the distance between the bus bay and the route start/end point (which refers to the same node in case of a looped route). The sum of hauling time from the bus bay to the route start point, the travel time on the route (looped) and the hauling time from the route end point back to the bus bay is actually one full cycle time for a bus serving route r in package p .

5.2.4. Model Formulation

The optimization problem to design transit route packages can be given by:

$$\begin{aligned} \min Z_1 = & \sum_{p \in P} \sum_{r \in R} y_p^r (n_p^r O_p^r + w_c \frac{X^r}{f_p^r}) + \sum_{p=1}^{k-1} \sum_{q=p+1}^k \{ \psi_d \left| \sum_r y_p^r D^r - \sum_r y_q^r D^r \right| - \\ & \xi_d \left| \sum_r y_p^r \gamma_p^r - \sum_r y_q^r \gamma_q^r \right| \} \end{aligned} \quad (5.1)$$

Subject to:

$$f_p^r Cap_b \geq X^r, \forall r \in R, \forall p \in P \quad (5.2)$$

$$\sum_{p \in P} \sum_{r \in R} y_p^r n_p^r \leq F \quad (5.3)$$

$$T_p^r f_p^r \leq n_p^r, \forall r \in R, p \in P \quad (5.4)$$

$$\sum_{p \in P} y_p^r = 1, \forall r \in R \quad (5.5)$$

$$\sum_{r \in R} y_p^r \geq 1, \forall p \in P \quad (5.6)$$

$$\sum_{r \in R} y_p^r n_p^r \leq C_p, \forall p \in P \quad (5.7)$$

$$\gamma_p^r = F_p^r f_p^r + B_p^r n_p^r, \forall r \in R, p \in P \quad (5.8)$$

$$f_p^r \geq 0, n_p^r \geq 0, \gamma_p^r \geq 0, y_p^r \in \{0,1\}, \forall r \in R, \forall p \in P \quad (5.9)$$

In (5.1), the objective function of the regulatory authority is minimized. The first term within the round brackets describes the sum of bus ownership and parking costs on a route r allocated to a route package p , while the second term indicates the cost of commuters' waiting time, as the regulator aims to guarantee the commuters' expected service levels and hence include the waiting time cost to its own cost function to be minimized. The first term within the modulus operator minimizes the difference between packages' total distance coverage (sum of distances of all routes included in the package) so that there is not a large disparity in the geographic spread of transit packages, assuming transit routes run through the entire urban setting and connect all important locations. This is included so that there is a higher possibility of a certain transit corridor being operated by multiple operators which would put a check on the level of service. On a given transit route corridor where two different packages operated by two different operators intersect, the preferential ridership on a particular operator's service would be an indicator of inferior services offered by the other which could be suitably penalized or warned to improvise as per the contractual terms. In the objective function, a cost term is associated with every unit difference in packages' total distance coverage. The second term within the modulus operator is the term ensuring the design of route packages such that they are attractive for all kinds of market players. The negative sign before the term is to maximize the difference between the total operation costs of packages such that packages of varying costs are designed to allow smaller players to enter the market and contest for the rights to operate them. The packages involving higher operation costs could anyways be contested upon by the incumbent/bigger players and packages incurring lower costs would encourage market entry of new operators. Hence, this ensures equity for all market players and a check on the market monopoly of the incumbent/bigger operators. A monetary conversion parameter is associated with every unit difference in packages' operating costs which indicates that a higher difference signifies savings for the regulator. Therefore, it can be observed that the objective function ensures that the route packages are designed considering fleet size allocation, service frequency of routes included in each

package to minimize commuter waiting time and also that the bidding for route packages is attractive to all kinds of transit players while maintaining a check on their level of service through comparable geographic spread of transit packages.

Constraint (5.2) states that the capacity of buses on a route should be greater than or equal to the exogenous demand on the route during the operational period. Constraint (5.3) ensures that the total fleet size of all routes across all packages is less than or equal to the fleet size available to the regulator. Constraint (5.4) implies that the number of buses allocated should be able to meet the requirement of service frequency. Constraint (5.5) states that every route must be allocated to a single specific package. Constraint (5.6) entails that each transit package should have a minimum of one transit route. Constraint (5.7) guarantees that the number of buses allocated to a bus package should not exceed the parking capacity of its allocated bus bay. Constraint (5.8) defines the operators' costs on each route r in a package p as the sum of operating frequency costs (product of rate of operating at the required frequency per operational period and the frequency per operational period) and the maintenance costs (product of rate of bus maintenance per operational period and the number of buses serving the route). Constraint (5.9) states the non-negative nature of route frequency, fleet size, operating costs and the binary variables used to allocate transit routes to packages.

5.2.5. Solution Methodology

The given formulation is a Mixed Integer Nonlinear Programming Problem (MINLP). Linearization techniques are adopted to convert its mathematical structure to a Mixed Integer Linear Program (MILP) to attain a globally optimal solution. The nonlinearity in the model arises from the objection function (5.1), constraint (5.3) and constraint (5.7).

Considering the objective function (5.1), the terms $y_p^r(n_p^r O_p^r + w_c \frac{X^r}{f_p^r})$ and $y_p^r \gamma_p^r$

involve the product of a binary (y_p^r) and a continuous variable ($n_p^r, \frac{1}{f_p^r}$ and γ_p^r respectively) which can be linearized using the reformulation linearization

technique (RLT). For the term $y_p^r w_c \frac{X^r}{f_p^r}$ where w_c, X^r are constants, the nonlinearity due to $\frac{1}{f_p^r}$ is first linearized using piecewise linear approximation (shown later) followed by RLT to linearize the product with the binary variable y_p^r .

5.2.5.1. Reformulation linearization technique (RLT)

As regards the RLT for linearizing the term $y_p^r \gamma_p^r$, substituting $\beta_p^r = y_p^r \gamma_p^r$ where y_p^r is binary and γ_p^r is continuous such that $\underline{\gamma_p^r} \leq \gamma_p^r \leq \overline{\gamma_p^r}$, $\underline{\gamma_p^r}$ and $\overline{\gamma_p^r}$ being a sufficiently small positive number(lower bound) and a sufficiently large positive number (upper bound) on γ_p^r respectively, then the equivalent linear transformation of the bilinear term $y_p^r \gamma_p^r$ can be expressed as:

$$\begin{cases} \beta_p^r - y_p^r \underline{\gamma_p^r} \geq 0 \\ \beta_p^r - y_p^r \overline{\gamma_p^r} \leq 0 \\ \beta_p^r - \gamma_p^r + \underline{\gamma_p^r} - y_p^r \underline{\gamma_p^r} \leq 0 \\ \beta_p^r - \gamma_p^r + \overline{\gamma_p^r} - y_p^r \overline{\gamma_p^r} \geq 0 \end{cases} \quad (5.10)(i-iv)$$

The constraints (5.3) and (5.7) can also be linearized using the same procedure.

5.2.5.2. One-dimensional piecewise linearization

Now, considering the waiting time term $w_c \frac{X^r}{f_p^r}$, the nonlinearity is due to the

fractional term $\frac{1}{f_p^r}$ and w_c, X^r are known constants. For the nonlinearity due to the

term $\frac{1}{f_p^r}$, a one-dimensional piecewise linear approximation of nonlinear

functions proposed by Misener and Floudas (2009) is adopted. The nonlinear

function in the model formulation can be expressed as $f(e) = \frac{1}{e}$. Let the domain of

the function be defined by E which is partitioned into an orthogonal grid that spans the domain. Now, any point $e \in E$ within this grid can be expressed as a convex combination of the grid points. However, the convex combination of these grid

points is not necessarily unique. Further, every point located within this convex set $E \subset \mathbb{R}^n$ can be expressed as a convex combination of at most $n+1$ points of E . In this one-dimensional case, considering the domain of E as convex, any point in the convex space can be written as a convex combination of two points as $E \subset \mathbb{R}$. Although at most $n+1$ grid points would express any point $e \in E \subset \mathbb{R}^n$, there are more than $n+1$ grid points in any finite domain space. To ensure a unique interpolation of function values using convex combination of grid points, only $n+1$ grid points are activated each time. Hence, when $n+1$ grid points $e_0, e_1, \dots, e_n \in E$ for domain $e \in E \subset \mathbb{R}^n$ are activated, any function $f(e): E \rightarrow R$ can be approximated as:

$$f(e) = w_0 f(e_0) + w_1 f(e_1) + \dots + w_n f(e_n) \quad (5.11)$$

$$e = w_0 \cdot e_0 + \dots + w_n \cdot e_n, \quad (5.12)$$

$$\sum_{i=0}^n w_i = 1, \quad (5.13)$$

$$w_i \geq 0 \forall i = 0, \dots, n. \quad (5.14)$$

This gives a unique convex combination when the point $e \in E$ is in the interior of $n+1$ grid points; hence, the function $f(e): E \rightarrow R$ is interpolated between the $n+1$ grid points. Hence, the nonlinear term in the objective function is linearized.

5.2.5.3. Treatment for the absolute value function term in the objective function

To treat the nonlinearity due to the absolute function term $\left| \sum_r y_p^r D^r - \sum_r y_q^r D^r \right|$ in the objective function (for a particular value of p and q), the following methodology can be used:

Let us substitute $\sum_r y_p^r D^r - \sum_r y_q^r D^r = a$, then we know that $|a| = \begin{cases} a, a \geq 0 \\ -a, a < 0 \end{cases}$.

Let us now consider a binary variable $\omega \in \{0,1\}$ such that $\omega=1$ when $a \geq 0$ and $\omega=0$ when $a < 0$. Hence, $|a|=a$ when $\omega=1$ and $|a|=-a$ when $\omega=0$. These conditions can be represented by the following linear mathematical expressions:

$$U(1-\omega) \leq a < L\omega \quad (5.15)$$

$$|a| = \omega a + (1-\omega)(-a) \quad (5.16)$$

where U and L are very large negative and positive integers respectively. Now, when ω takes the value of 1 in (5.15), then $0 \leq a < L$ and the “absolute” function

gives ‘ a ’ as the solution in (5.16). Similarly, when ω takes the value of zero in (5.15), then $U \leq a < 0$ and the “absolute” function gives ‘ $-a$ ’ as the solution in (5.16). The nonlinearity on the right hand side of (5.16) can be easily handled using the RLT as shown in (5.10). A similar methodology can be applied to linearize $\left| \sum_r y_p^r \gamma_p^r - \sum_r y_q^r \gamma_q^r \right|$ in the objective function.

Therefore, with the application of above linearization techniques, the original MINLP model formulation is transformed into a MILP which can be easily solved using the branch and bound algorithm or a routine commercial solver ensuring a globally optimal solution which, in this case, would truly be the “best” contracting strategy for the transit regulator and equitable for the operators and the commuters alike.

5.3. Numerical study and discussion

Let us consider a transit network of n routes with a known exogenous demand. It is also assumed that the regulator has a known number of packages k to design. This implies that n transit routes have to be packaged into k transit route packages such that they are financially sustainable for the regulator and the operators while the commuters’ waiting time is also minimized (route travel time is assumed constant and known).

In this numerical example, a segment of the Singapore bus transit network with 4 interchanges and 6 transit routes operating between them is selected as shown in the Fig. 5.1. There are 3 bus bays denoted by $P1$, $P2$ and $P3$ which coordinate the operation of the 3 bus packages to be designed.

The model was evaluated using the solver Gurobi on the programming platform YALMIP (Löfberg, 2004) interfaced with MATLAB on a Precision T1650 Dell PC, 3.20 GHz processor, 16 GB RAM and 64-bit operating system.

Fig.5.1 Transit network with numbered routes and bus bays denoted as $P1$, $P2$, $P3$.

5.3.1. Parameters

(i) Network characteristics:

Table 5.1 Network characteristics:

Origin	Destination	Route/ bus service	Distance (in kms.)
Boon Lay	Toa Payoh	157	22
Boon Lay	Bedok	30	33
Boon Lay	Bukit Merah	198	18
Toa Payoh	Bukit Merah	139	15
Toa Payoh	Bedok	26	10
Bedok	Bukit Merah	16	18

(ii) Number of transit route packages to be designed $k=3$

Given below are a few parameters which a transit regulatory authority would generally possess or could assess as part of the operational details. In this numerical study, a realistic assumption of this data is taken into account for the sake of academic investigation.

(iii) Total route demand per hour during peak hour:

Table 5.2 Route demand (D)/hr. during peak hour (number of commuters) X^r

Route	157	30	198	139	26	16
Demand(commuters)	400	500	500	300	250	300

(iv) Hourly rate of operating buses at a required frequency on route r included in package p :

Table 5.3 Hourly rate of operating buses at a required frequency on route r included in package p (F_p^r) in \$

Package\Route	157	30	198	139	26	16
P1	10	20	15	15	20	20
P2	20	25	20	15	10	15
P3	15	20	30	20	20	15

(v) Hourly rate of maintaining a bus on route r included in package p :

Table 5.4 Hourly rate of maintaining a bus on route r included in package p (B_p^r) in \$

Package\Route	157	30	198	139	26	16
P1	10	15	10	10	15	10
P2	15	10	15	15	20	20
P3	20	25	20	20	25	20

(vi) Hourly rate of bus ownership on route r in package p :

Table 5.5 Hourly rate of bus ownership on route r in package p (O_p^r) in \$

Package\Route	157	30	198	139	26	16
P1	10	20	10	10	10	10
P2	10	20	25	15	15	20
P3	15	20	20	15	10	10

(vii) Total cycle time on route r in package p including hauling time (minutes):

Each bus is assumed to start from its allocated bus bay, reach the route starting node, complete one cycle and return back to the same bus bay. After a certain lay-over period, the same cycle continues.

Table 5.6 Travel time on route r in package p including hauling time T_p^r (mins)

Package\Route	157	30	198	139	26	16
P1	110	120	50	90	100	80
P2	130	160	80	40	60	60
P3	160	130	100	70	50	30

(viii) Waiting time cost $w_c = 1\$/hr$.

(ix) Bus capacity= 100 commuters/bus

(x) Available fleet size= 1000 buses

(xi) Maximum capacity of each bus bay= 150 buses

(xii) $\psi_d = 1\$$, $\xi_d = 0.5\$$

5.3.2. Optimization results

(i) Binary variable for allocation of a route r to package p : y_p^r

Table 5.7 Binary variable for allocation of route r to package p : y_p^r

Package\Route	157	30	198	139	26	16
P1	1	0	0	0	0	0
P2	0	0	0	1	0	0
P3	0	1	1	0	1	1

(ii) Fleet size allocated to route r in package p : n_p^r

Table 5.8 Fleet size allocated to route r in package p : n_p^r

Package\Route	157	30	198	139	26	16
P1	8	0	0	0	0	0
P2	0	0	0	2	0	0
P3	0	11	9	0	3	2

(iii) Frequency of route r in package p in buses/hr.: f_p^r

Table 5.9 Frequency of route r in package p in buses/hr.: f_p^r

Package\Route	157	30	198	139	26	16
P1	4	0	0	0	0	0
P2	0	0	0	3	0	0
P3	0	5	5	0	3	3

The results demonstrate package 1 consists of route 157, package 2 is assigned route 139 and routes 30,198, 26 and 16 are allocated to package 3. The fleet size obtained satisfies demand as per the operating frequency and the schedule. Also, as per the designed route frequencies, the waiting time of commuters along each route is minimized. If costs are compared, the three packages $P1$, $P2$, $P3$ incur operation costs of 230\$,75\$, 570\$ respectively per operational hour which ensures optimal difference in costs between all packages to encourage transit market entry of new operators, also minimizing the ownership and parking costs to the regulator at the

bus bays. The regulator's cost as per this optimal contractual strategy is 6958\$ and the sum of total operating cost of all packages is 875\$.

5.3.3. *Further analysis*

More studies were conducted with change in demand and the results can be stated as below in Table 5.10.

Table 5.10 Sensitivity analysis with respect to demand

Demand	Package design	Regulator cost	Packages' operating cost/hr
Demand D	P1={ 157,198},P2={ 139}, P3={ 30,26,16}	6958\$	\$230(P1), \$75(P2), \$570(P3)
D1: 150% D	P1={ 157},P2={ 139}, P3={ 30, 198,26,16}	18281\$	\$170(P1), \$113(P2), \$1,318(P3)
D2: 120% D	P1={ 157},P2={ 139}, P3={ 30, 198,26,16}	10949\$	\$136(P1), \$90(P2), \$1,050(P3)
D3: 80% D	P1={ 157,198},P2={ 139}, P3={ 30,26,16}	3680\$	\$185(P1), \$60(P2), \$447(P3)
D4:50% D	P1={ 157,198},P2={ 139}, P3={ 30,26,16}	120\$	\$115(P1), \$38(P2), \$50(P3)

Fig 5.2 Packages' operating cost for different demands

The Fig. 5.2 shows the plot of different packages' operating cost for different demands. For each value of demand, the packages' designs change with their associated operating costs. In the Fig. 5.3 plot of the regulator costs vs. total operating costs of all packages for different demand values, it can be observed that the trend changes with changing demand. However, it must be noted that this is not the net cost for the operators but only the package operating costs since revenue is not taken into consideration in this model formulation. In a bus contracting regime, the operators are paid a pre-decided service fee if they are successful in securing the right to operate as per the competitive tendering. This fee is majorly funded from the fare-box collection on the buses and the government authority can contribute an additional amount, in the case, ridership is low resulting in low fare-box revenue collection. It can be observed that for different demands, even if the package designs remain the same such as for D, D3 and D4, their costs vary as the route fleet sizes and the route service frequencies would change to cater to the new demand, affecting the package operating costs.

Fig. 5.3 Plot of regulator costs vs. total packages' operating costs for different demands

Since this model formulation is a decision making framework when the regulator is planning to contract out route packages through CT, only the operating costs are considered which can be assessed by the regulator from experience. Moreover, these costs can be reasonably estimated as they do not change much but revenues from the fare-box collection cannot be pre-estimated as they would totally depend on the ridership which would further depend on the level of service when the operation commences. Hence, revenues are not included in the model formulation and only the regulator costs are plotted against the total packages' operating costs for different demands. Also, it is assumed that all operating services are similar in the sense that they serve the bus stops with similar inter-stop distances, congestion effects and the level of service of available transportation infrastructure. The routes used in the numerical study are real-time routes and are normal services serving bus stops at short spacing. However, if certain routes offer an express or a deadheading or a limited stop service, the transit package design would change on account of different operating costs. These services would have a lower turnaround time (time to complete one cycle) and a higher level of service; therefore, the fare structure for these services would also be different from the normal service.

5.3.3.1. Route package design on an expanded transit network

To further test the validity of the model on a large size transit network, we select an idealized problem with suitable parametric assumptions wherein 30 transit routes are to be packaged into 3 different transit packages. For notation purpose, we consider a two dimensional matrix where i is used to index transit route package and j to index transit route. Hence, the parameter inputs for the model formulation can be stated as follows:

- 1) Transit route distances= $0.2*j*(31-j)$
- 2) Transit route demand= $2*j*(31-j)$
- 3) F_p^r : For $i=1$: $F_p^r=0.25*j*(31-j)$; For $i=2$: $F_p^r=0.3*j*(31-j)$; For $i=3$: $F_p^r=0.20*j*(31-j)$;
- 4) $B_p^r=0.8*(\text{corresponding value of } F_p^r)$
- 5) $O_p^r=0.7*(\text{corresponding value of } F_p^r)$
- 6) $T_p^r=0.5*(\text{corresponding value of } F_p^r)$

The transit route package design resulting from the model solution are demonstrated as follows: Package 1 includes three transit routes {4, 14, 22}, Package 2 has four transit routes {5, 16, 21, 30}, while all the remaining transit routes are bundled into the third package. It is worth noting that, based on our model, the optimal route packaging result does not necessarily bundle the bus routes in a manner of even distribution. One can observe that the majority of the transit routes in this experimental test are grouped into one package. Indeed, this phenomenon can be evidenced by the recent real practice in Singapore. Singapore Land Transport Authority has recently tendered out the first transit route package with 26 bus routes (Bulim Package) to a private transit operator (Tower Transit) for operation. Meanwhile, the two existing bus companies, SBS and SMRT, continue their operation over 196 and 77 routes, respectively, which have now been packaged together. It should be noted that, in our model formulation, the regulator costs related to service levels, barrier entry in terms of package operating cost and the cost due to difference in packages' distance coverage, are not tangible costs, rather, empirical costs formulated to ensure an equitable package design.

5.4. Conclusion

This study presented a decision making analytical framework for transit regulators to package transit routes for contracting them out to operators through the CT process. The model formulation allocates routes to packages while also determining the route frequency and the fleet size for each route such that equity is ensured for all stakeholders of the transit system. Transit packages are so designed such that the ownership costs of the regulator and the travel costs of the commuters are minimized while the packages' operating costs of the operators are equitable. Also, the attractiveness of all packages is ensured to encourage entry of new players while discouraging the evolution of a monopolistic transit market. This methodology is significant as more public transportation friendly economies look towards implementing the privatized contracting regime for greater financial and operational sustainability. Literature provides no mathematical paradigm in public transportation contracting, hence, this work marks a beginning for research on analytical frameworks in bus transit contracting and would be of great utility to regulators adopting the privatized contracting regime.

CHAPTER 6 A METHODOLOGY FOR COMPETITIVE TENDERING OF TRANSIT ROUTE PACKAGES IN A PRIVATIZED BUS CONTRACTING REGIME

6.1. Introduction and Problem description

This chapter presents a methodology to describe the CT process in a privatized bus contracting regime wherein the model formulation defined through an optimization approach aims to optimally allocate transit route packages (a set of transit routes) to contesting operators. As a matter of fact, CT has been a routine process in several industries for long, however, an analytical insight into the intricacies of the bidding process and a good understanding of the role of the transit regulator and the operators to shape a financially sustainable transit market is of great importance. As mentioned, the regulator leads the CT exercise by designing routes, route frequencies, fleet size etc. and subsequently tendering them out to contesting operators. Once the route packages are awarded, the successful operator is only responsible for the routine operation of the transit package and maintenance of buses and bus bays for which it is paid a service fee same as the price quoted in its tender proposal.

For the purpose of CT, a two envelope system including quality and price bidding is generally adopted to decide on the operator who would be awarded the operation of a particular package. In this study, it is assumed that all contesting operators have fulfilled the quality criterion of operation and therefore, only the price criterion is taken into consideration. The quality criterion mostly estimates the overall track record of the operator, details of crew scheduling, provisions to improve level of service, customer feedback mechanism etc. As regards the price criterion, it is the amount each operator quotes in the tender document as the service fee it expects in return for the operation of a particular transit package over a period. In many contracts, the operator is also expected to state the rates for different operational aspects such as maintenance and operating at the required frequency (incurs cost of crew scheduling) to facilitate a detailed examination. However, it is the service fee quote that mainly impacts the chances of a contesting operator to win the right to operate a particular route package. Hence, in order to secure the right to operate, an

operator's price quote must be the minimum of quotes submitted by all other contesting operators. As a result, each contesting operator aims to minimize its operational cost through innovation, use of advanced technology etc. and quotes a service fee which ensures sufficient profits while still being competitive in the tendering process.

In this study, the bus transit CT process is formulated as a mathematical negotiation between the regulator and the contesting operators resembling the Stackelberg game in game theory. The model is presented as a bi-level formulation wherein the transit regulator is the master and the operators are the followers. It is assumed that transit routes are already packaged by the regulator who awards them to contesting operators through CT. The upper level objective is the cost function of the regulator which has to decide whether a particular route package is awarded to a contesting operator. The other decision variables involved in the upper level problem are the fleet size and service frequency of all routes included in every package. In accordance with the decision variables in the upper level problem, the lower level problem is formulated which is an aggregate operators' problem who maximize their total profits while trying to maintain a competitive environment through a narrow gap between their quoted service fees. The lower level decision variable is the service fee to be quoted in case an operator m contests for a package p such that the operator maximizes its own profits while proposing a competitive quote. A transit regulator would generally intend to pay off a package's service fee from the fares, however, if the ridership on the routes included in the package is not high, the regulator would need to put in an additional amount referred to as 'subsidy' in this study. Hence, with known exogenous demand and fare structure for all routes included in a transit package, the successful operator's service fee quote determines the subsidy that the regulator would need to pay and the bi-level formulation is helpful in capturing this inter-dependence of financial resources between the two entities.

It is important to understand that this model formulation lays the foundation to a systematic, analytical prediction of the optimal financing of bus transit operation. In a typical competitive tendering process, the contractual proposals of the operators

have no relation with the regulator's estimates. Each operator contesting for a particular route package would try to make the most accurate estimate of cost of operation and quote a price including the profit margin. Now, because this is not an open tender, it is assumed that the regulator has a fair idea of the competencies of the invited operators and is aware of their operating rates. With this background information, the regulator tries to optimize its own costs while estimating the optimal service fee quote that should be paid to operate a particular transit package in order to ensure equity for both the stakeholders. It is important to estimate the regulator's costs as that money is publicly funded and needs to be carefully utilized. However, once the tendering process is implemented, it is possible that the operators quote service fees which are very different from the regulator's expectations. In this case, it would be binding for the regulator to allocate the transit route packages to the best of the available proposals. Hence, in such a scenario, the optimal contracting strategies and the associated costs computed using the model formulation would help the regulator find the offset between the expected optimal costs and the actual costs it ends up paying. Estimating this amount is crucial as depending on whether the regulator loses or gains financially, it would need to raise this amount by other means or spend on other bus transit improvement projects respectively.

6.2. Methodology

6.2.1. Assumptions

- a) It is assumed that transit routes have already been packaged by the regulator in a way it is financially sustainable for all stakeholders.
- b) It is assumed that a select tender is used to invite tender proposals from transit operators. For tendering transit route packages, the regulator selects a few operators with a proven track record and invites them to submit proposals. However, it is assumed that the set of selected operators are free to bid for one or more of the transit packages being tendered out by the regulator. Hence, it is known in advance which operators are contesting for a particular route package.
- c) No surety bond is needed to be paid as part of the bid.
- d) All contesting operators qualify the quality criteria and the proposal with the lowest service fee quote is selected.

- e) The transit routes and the route in-vehicle travel time remain the same as in the existing transit network. The regulator is sensitive towards commuters' waiting time and minimizes it by setting the route frequency accordingly.
- f) The regulator has enough budgets coming from public funds; however, it tries to minimize its usage.
- g) Tendering costs are already incorporated in the operators' rates for operating at the required route frequency and bus fleet maintenance.
- h) The operators have sufficient financial resources to submit tender proposals for the transit packages they are invited to contest for by the regulator.

6.2.2. Variable definitions

Sets and parameters

P	set of transit route packages $p \in P$
R	set of transit routes $r \in R$
R_p	subset of transit routes r belonging to package $p \in P$, i.e. $R_p \subset R$
M	set of contesting transit operators $m \in M$
M^p	subset of operators m' contesting for a package $p \in P$ i.e. $M^p \subset M, m' \in M^p$
O_p^r	ownership cost of a bus allocated to a route r in package p
w_c	waiting time cost
k	parameter whose value depends on the distribution of bus arrival times
X_p^r	total revenue generated per operational period from a route r in a package p
T_p^r	total cycle time of route r in a package p
U	total fleet size available

D^r	exogenous demand for a route r
C_p	capacity of a bus bay allocated to a package P
Cap_b	capacity of a bus
$F_p^{r,m'}$	price rate for operating a bus at the desired frequency on route r in package p for an operator m'
$B_p^{r,m'}$	price rate for maintaining a bus serving a route r in package p for an operator m'

Lower level decision variables

$\eta_p^{m'}$	service fee quoted by operator m' to operate package p in the tender proposal
---------------	---

Upper level decision variables

$y_p^{m'}$	binary variable defining whether an operator m' is awarded the package p
f_p^r	service frequency of a route r included in a package p
n_p^r	fleet size allocated to a route r included in a package p

6.2.3. Model Formulation

P1:

$$\min_{y_p^{m'}, n_p^r, f_p^r, \eta_p^{m'}} \psi(y_p^{m'}, n_p^r, f_p^r, \eta_p^{m'}) = \sum_{p \in P} \left\{ \sum_{r \in R_p} (O_p^r n_p^r + k w_c \frac{D^r}{f_p^r}) + \sum_{m' \in M^p} y_p^{m'} (\eta_p^{m'} - \sum_{r \in R_p} X_p^r) \right\} \quad (6.1)$$

Subject to:

$$T_p^r f_p^r \leq n_p^r, \forall r \in R_p, \forall p \in P \quad (6.2)$$

$$\sum_{p \in P} \sum_{r \in R_p} n_p^r \leq U \quad (6.3)$$

$$D^r \leq f_p^r Cap_b, \forall r \in R_p, \forall p \in P \quad (6.4)$$

$$\sum_{r \in R_p} n_p^r \leq C_p, \forall p \in P \quad (6.5)$$

$$\sum_{m' \in M^p} y_p^{m'} = 1, \forall p \in P \quad (6.6)$$

$$\sum_{p \in P} y_p^{m'} \geq 0, \forall m' \in M^p \quad (6.7)$$

$$y_p^{m'} = 1 \Leftrightarrow \eta_p^{m'} \leq \eta_p^{m''}, \forall m'' \in M^p \setminus \{m'\}, \forall m' \in M^p, \forall p \in P \quad (6.8)$$

$$n_p^r \geq 0, f_p^r \geq 0, \forall r \in R_p, \forall p \in P \quad (6.9)$$

$$y_p^{m'} \in \{0,1\}, \forall m' \in M^p, \forall p \in P \quad (6.10)$$

$$P2: \min_{\eta_p^{m'}} \phi(y_p^{m'}, \eta_p^{m'}) = \sum_{p \in P} \sum_{m' \in M^p} (-y_p^{m'} \eta_p^{m'} + \sum_{m'' \in M^p \setminus m'} |\eta_p^{m'} - \eta_p^{m''}|) \quad (6.11)$$

Subject to:

$$\sum_{r \in R_p} (F_p^{r,m'} f_p^r + B_p^{r,m'} n_p^r) \leq \eta_p^{m'}, \forall m' \in M^p, \forall p \in P \quad (6.12)$$

$$\eta_p^{m'} \geq 0, \forall m' \in M^p, \forall p \in P \quad (6.13)$$

The objective function (6.1) is the upper level problem of the transit regulator who has to minimize costs. The first term is the cost of ownership of buses on each route included in every package and the second term within the same brackets is the waiting time cost of passengers travelling on each route included in every package. The next term is the subsidy provided by the regulator to the operator who wins the right to operate a particular transit package. Let us understand the cost of subsidy in detail. Each operator submits a price quote during the bidding process and upon being successful, expects the same as payment in return for the services over a

regular period as per the contractual terms. The transit regulator, mostly a non-profit government authority funded by the taxpayers, tries to recover most of this service fee from the revenue generated from the fare-box collection on the buses that the commuters pay. However, in certain cases, there could be a possibility that due to certain unforeseen operational issues or change in the expected ridership pattern due to inferior level of service, there is a low demand and so, the revenue generated from fares is not sufficient to pay the promised service fee. In this situation, the authority pays the remaining amount which is referred to as the subsidy. Hence, minimizing this subsidy is also one of the important objectives of the regulator. The rationale is that the amount paid as subsidy by the regulator is the taxpayers' money in addition to the already paid travel fares. Hence, to avoid such a situation, the transit regulator monitors the level of service of each operator regularly for there is a possibility of the operators slacking on the level of service because of an ensured service fee after being awarded the right to operate. Generally, terms of deterrence are included in the contract which could be in the form of penalties, termination of contract etc. In the other case, the regulator can also modify the route frequency or revise the fares in order to improve ridership if the additional cost of such a change is lower than the revenue earned from an increased demand.

Constraint (6.2) ensures that each route of every package has the minimum fleet size to maintain operation without delay or disruption. Constraint (6.3) states that the total fleet size allocated to all packages should be within the total fleet size available to the regulator. Constraint (6.4) states that the capacity of all buses operating on a route during a particular operational period should be equal to or greater than the route demand. Constraint (6.5) states that the number of buses allocated to each package should be less than or equal to the capacity of the allotted bus bay. Constraint (6.6) states that a particular transit route package can be operated by only a single operator. Constraint (6.7) states that a particular operator could, however, operate multiple transit packages. Constraint (6.8) is a mathematical condition that states that if a particular package is awarded to a contesting operator, the price quote of this operator must be lower than the price quotes of all other operators contesting for the operation of the same transit package. The constraint is valid vice-versa too, i.e. if the price quote of an operator

m' is lower than the quote of all other operators contesting for the same package p ; the package is awarded to the operator m' . Constraints (6.9) and (6.10) define the non-negative variables and binary variables respectively.

The objective function (6.11) is the lower level objective function of all operators which consists of two terms. The first term is to maximize the aggregate revenues in terms of service fees in the event that the operators are successful in securing the operation of the route package they contested for. The second term is to minimize the difference between the service fees quoted by all contesting operators for a particular package to ensure a competitive environment. It must be understood that in a CT process, the price quote of each contesting operator is confidential and is directly submitted to the regulator as part of the contract proposal. Since ensuring a competitive CT of transit route packages is a prerogative for the regulator, it might wish to estimate the combination of price quotes that contesting operators for each transit package would quote with their operating rates exogenously known to the regulator. A successful operator's profit is calculated as the difference between its quoted service fee which it gets paid by the regulator and the total operating cost of the route package it is allocated to operate. Constraint (6.12) states that the quoted service fee should be equal to or greater than the total operating costs of a transit package. Finally, constraint (6.13) defines the non-negativity of the quoted service fee.

6.2.4. Solution Algorithm

The model formulation for the discussed problem is a Mixed Integer Non-Linear Bi-level Programming Problem. The nonlinearity in the model arises from the upper and lower level objective functions given by (6.1) and (6.11) respectively. These are linearized using various linearization techniques. The lower level problem is subsequently reformulated using the Karush-Kuhn-Tucker (KKT) conditions into complementary constraints, transforming the nonlinear bi-level problem into a single level, mixed-integer linear Mathematical Problem with Complementary Constraints (MPCC). This transformed and linearized formulation is basically a Mixed Integer Linear Programming Problem (MILP) which can be easily solved to global optimality using available algorithms such as branch and bound technique etc.

6.2.4.1. One-dimensional piecewise linearization

Let us first linearize the objective function (6.1) in which both the terms are nonlinear. The nonlinearity in the first term is because of the term $\frac{1}{f_p^r}$ which is nonlinear in one dimension and has the route frequency in the denominator. The one-dimensional piecewise linearization methodology proposed by Misener and Floudas (2010) is used for linearization as explained below.

The given nonlinear function accounting for the waiting time as the inverse of bus route frequency can be expressed as $f(e) = \frac{1}{e}$. Let the domain of the function be defined by E which is partitioned into an orthogonal grid that spans the domain. Now, any point $e \in E$ within this grid can be expressed as a convex combination of the grid points. However, the convex combination of these grid points is not necessarily unique. Further, every point located within this convex set $E \subset \mathbb{R}^n$ can be expressed as a convex combination of at most $n+1$ points of E . In this one-dimensional case, considering the domain of E as convex, any point in the convex space can be written as a convex combination of two points as $E \subset \mathbb{R}$. Although at most $n+1$ grid points would express any point $e \in E \subset \mathbb{R}^n$, there are more than $n+1$ grid points in any finite domain space. To ensure a unique interpolation of function values using convex combination of grid points, only $n+1$ grid points are activated each time. Hence, when $n+1$ grid points $e_0, e_1, \dots, e_n \in E$ for domain $e \in E \subset \mathbb{R}^n$ are activated, any function $f(e): E \rightarrow R$ can be approximated as:

$$f(e) = w_0 f(e_0) + w_1 f(e_1) + \dots + w_n f(e_n) \quad (6.14)$$

$$e = w_0 e_0 + \dots + w_n e_n, \quad (6.15)$$

$$\sum_{i=0}^n w_i = 1, \quad (6.16)$$

$$w_i \geq 0 \forall i = 0, \dots, n. \quad (6.17)$$

This gives a unique convex combination when the point $e \in E$ is in the interior of $n+1$ grid points; hence, the function $f(e): E \rightarrow R$ is interpolated between the $n+1$ grid points. Hence, the function can be linearized.

6.2.4.2. Reformulation linearization technique (RLT)

For the linearization of the second term which involves the product of a continuous (quoted service fee) and a binary variable (whether the operator is awarded the package) given by $y_p^{m'} \eta_p^{m'}$, the following reformulation linearization technique as proposed by Sherali and Alameddine (1992) is adopted. Substituting $\beta_p^{m'} = y_p^{m'} \eta_p^{m'}$ where $y_p^{m'}$ is binary and $\eta_p^{m'}$ is continuous such that $\underline{\eta_p^{m'}} \leq \eta_p^{m'} \leq \overline{\eta_p^{m'}}$, $\underline{\eta_p^{m'}}$ and $\overline{\eta_p^{m'}}$ being a sufficiently small positive number(lower bound) and a sufficiently large positive number (upper bound) on $\eta_p^{m'}$ respectively, then the equivalent linear transformation of the bilinear term $y_p^{m'} \eta_p^{m'}$ can be expressed as:

$$\begin{cases} \beta_p^{m'} - y_p^{m'} \underline{\eta_p^{m'}} \geq 0 \\ \beta_p^{m'} - y_p^{m'} \overline{\eta_p^{m'}} \leq 0 \\ \beta_p^{m'} - \eta_p^{m'} + \underline{\eta_p^{m'}} - y_p^{m'} \overline{\eta_p^{m'}} \leq 0 \\ \beta_p^{m'} - \eta_p^{m'} + \overline{\eta_p^{m'}} - y_p^{m'} \underline{\eta_p^{m'}} \geq 0 \end{cases} \quad (6.18)\text{- (i-iv)}$$

The same procedure is adopted to linearize the first term in the objective function of the lower level problem which is nonlinear due to the product of the binary variable $y_p^{m'}$ and the continuous variable $\eta_p^{m'}$. Further, the logical expression (6.8) can be linearly expressed as follows:

$$Ly_p^{m'} \leq \eta_p^{m'} - \eta_p^{m''} \leq U(1 - y_p^{m'}), \forall m'' \in M^p \setminus \{m'\}, \forall m' \in M^p, \forall p \in P \quad (6.19)$$

where L and U are very large negative and positive constants respectively.

6.2.4.3. Treatment for the absolute value function term in the lower level objective function

For a particular value of m' and p , let us substitute $|\eta_p^{m'} - \eta_p^{m''}| = a$. We know that $|a| = \begin{cases} a, a \geq 0 \\ -a, a < 0 \end{cases}$. Let us now consider a binary variable $\omega \in \{0,1\}$ such that $\omega = 1$ when $a \geq 0$ and $\omega = 0$ when $a < 0$. Hence, $|a| = a$ when $\omega = 1$ and $|a| = -a$ when $\omega = 0$. These conditions can be represented by the following linear mathematical expressions:

$$U(1 - \omega) \leq a < L\omega \quad (6.20)$$

$$|a| = \omega a + (1 - \omega)(-a) \quad (6.21)$$

where U and L are very large negative and positive integers respectively. Now, when ω takes the value of 1 in (6.20), then $0 \leq a < L$ and the “absolute” function gives ‘ a ’ as the solution in (6.21). Similarly, when ω takes the value of zero in (6.20), then $U \leq a < 0$ and the “absolute” function gives ‘ $-a$ ’ as the solution in (6.21). The nonlinearity on the right hand side of (6.21) can be easily handled using the RLT as shown in (6.18).

6.2.4.4. Reformulation using KKT conditions

Let us now reformulate the lower level problem into a set of complementary constraints using the famous KKT conditions. Solving the bi-level problem is generally tedious; however, this approach converts the bi-level problem to a single level problem i.e. a MPCC. Hence, the single level problem can subsequently be linearized and solved like a typical linear, single level optimization problem. This method underlines the fact that since $\eta_p^{m'}$ is a global minimizer of the lower level problem P2, the lower level constraint is first relaxed such that $\eta_p^{m'}$ is a local minimizer of P2. The lower level objective function given by (6.11) and the constraint (6.12) are considered. The constraint (6.12) can also be expressed as:

$$\rho(\eta_p^m, y_p^m) : \eta_p^{m'} - \sum_{r \in R_p} (F_p^{r,m'} f_p^r + B_p^{r,m'} n_p^r) \geq 0 \quad (6.22)$$

As discussed in Allende and Still (2013), the lower level problem can be replaced using the Fritz-John and KKT conditions as follows:

$$\lambda_0 \nabla_{\eta_p^m} \phi(\eta_p^m, y_p^m) - \sum_{p \in P} \sum_{m' \in M^p} \lambda_{p,m'} \nabla_{\eta_p^m} [\eta_p^{m'} - \sum_{r \in R_p} (F_p^{r,m'} f_p^r + B_p^{r,m'} n_p^r)] = 0 \quad (6.23)$$

$$\lambda_0 \geq 0, \lambda_{p'}^m \geq 0, \quad (6.24)$$

$$\eta_p^{m'} - \sum_{r \in R_p} (F_p^{r,m'} f_p^r + B_p^{r,m'} n_p^r) \geq 0, \forall m' \in M^p, \forall p \in P \quad (6.25)$$

$$\lambda_{p'}^m [\eta_p^{m'} - \sum_{r \in R_p} (F_p^{r,m'} f_p^r + B_p^{r,m'} n_p^r)] = 0, \forall m' \in M^p, \forall p \in P \quad (6.26)$$

$$\sum_{p \in P} \sum_{m' \in M^p} \lambda_{p'}^m = 1 \quad (6.27)$$

Hence, using the complementary conditions (6.23)-(6.27) with the upper level problem, the model formulation is transformed into a single level MPCC which is basically a MILP for which global optimality can be guaranteed.

6.3. Numerical study

Consider a transit network of five routes with known exogenous demand which have been packaged into three transit packages by the regulator. Four operators *M1*, *M2*, *M3*, *M4* are considered who have been invited by the transit regulator as part of the select tendering process to submit tender proposals for the different transit packages. Hence, the regulator knows which operators would submit tender proposals for any particular transit package.

A segment of the Singapore bus transit network with 4 O-D pairs and 5 operating routes as shown in Fig. 6.1 is selected. There are 3 bus bays denoted by *P1*, *P2*, *P3* which co-ordinate the operation of the 3 bus packages to be designed. The linearized model formulation is evaluated using the solver Gurobi on the programming platform YALMIP (Löfberg, 2004) interfaced with MATLAB on a Precision T1650 Dell PC, 3.20 GHz processor, 16 GB RAM and 64-bit operating system.

Fig.6.1. Transit network with routes numbered and bus bays denoted as $P1$, $P2$, $P3$.

6.3.1. Parameters

(i): Network characteristics:

Table 6.1 Network characteristics

Origin	Destination	Route/ bus service	Package
Boon Lay	Toa Payoh	157	P1
Boon Lay	Bukit Merah	198	P1
Toa Payoh	Bukit Merah	139	P2
Toa Payoh	Bedok	26	P3
Bedok	Bukit Merah	16	P3

(ii) Operators competing for different transit packages:

Table 6.2 Operators competing for different transit packages

Operator	$M1$	$M2$	$M3$	$M4$
Package	$P1, P2, P3$	$P1, P2$	$P2, P3$	$P3$

(iii) Total route demand/hr during peak hour:

Table 6.3 Route demand/hr during peak hour (number of passengers) X^r

Route	157	198	139	26	16
Demand(passengers)	200	300	200	200	300

(iv) Total fare revenue/hr from a package (total fares paid by all passengers travelling between nodes on all routes included in the package)

Table 6.4 Total fare revenue from packages (\$) per hour during operational hour

Package	$P1$	$P2$	$P3$
Revenue (\$)	540	180	720

(v) Cycle time of a route r in package p in minutes:

Table 6.5 Cycle time of a route r in package p in minutes

Package\route	157	198	139	26	16
$P1$	110	120	0	0	0
$P2$	0	0	40	0	0
$P3$	0	0	0	50	30

(vi) Capacity of a bus=80 passengers

(vii) Capacity of each bus bay: $P1$ = 50 buses, $P2$ =50 buses, $P3$ = 50 buses

(viii) Total available fleet=120 buses

(ix) Ownership cost/hr (\$) of a bus allocated to a route r in package

Table 6.6 Ownership cost/hr (\$) of a bus allocated to a route r in package p

Package\route	157	198	139	26	16
$P1$	10	15	0	0	0
$P2$	0	0	20	0	0
$P3$	0	0	0	20	25

(x) Rate for operating a bus at the desired frequency on route r in package p for an operator $M1$:

Table 6.7: Rate for operating a bus at the desired frequency on route r in package p for operator $M1$

Package\route	157	198	139	26	16
$P1$	30	35	0	0	0
$P2$	0	0	25	0	0
$P3$	0	0	0	30	20

(x) Rate for operating a bus at the desired frequency on route r in package p for an operator $M2$:

Table 6.8: Rate for operating a bus at the desired frequency on route r in package p for operator $M2$

Package\route	157	198	139	26	16
$P1$	20	30	0	0	0
$P2$	0	0	35	0	0
$P3$	0	0	0	0	0

(xi) Rate for operating a bus at the desired frequency on route r in package p for an operator $M3$:

Table 6.9: Rate for operating a bus at the desired frequency on route r in package p for operator $M3$

Package\route	157	198	139	26	16
$P1$	0	0	0	0	0
$P2$	0	0	40	0	0
$P3$	0	0	0	30	45

(xii) Rate for operating a bus at the desired frequency on route r in package p for an operator $M4$:

Table 6.10: Rate for operating a bus at the desired frequency on route r in package p for operator $M4$

Package\route	157	198	139	26	16
$P1$	0	0	0	0	0
$P2$	0	0	0	0	0
$P3$	0	0	0	35	40

(xiii) Rate for maintaining a bus serving a route r in package p for operator $M1$: (corresponding values in Table 6.7)*1.2; rate for maintaining a bus serving a route r in package p for operator $M2$:(corresponding values in Table 6.8)*0.9; rate for maintaining a bus serving a route r in package p for operator $M3$: (corresponding values in Table 6.9)*1.3; rate for maintaining a bus serving a route r in package p for operator $M4$: (corresponding values in Table 6.10)*0.95.

6.3.2. Optimization Results

(i) Binary variable for allocation of a package p to an operator m : y_p^m

Table 6.11: Binary variable for allocation of package p to operator m : $y_p^{m'}$

Package/Route	$M1$	$M2$	$M3$	$M4$
$P1$	0	1	0	0
$P2$	1	0	0	0
$P3$	1	0	0	0

(ii) Fleet size allocated to route r in package p : n_p^r

Table 6.12: Fleet size allocated to route r in package p : n_p^r

Package\route	157	198	139	26	16
$P1$	5	8	0	0	0
$P2$	0	0	2	0	0
$P3$	0	0	0	3	3

(iii) Frequency of route r in package p in buses/hr: f_p^r

Table 6.13: Frequency of route r in package p in buses/hr: f_p^r

Package\route	157	198	139	26	16
$P1$	3	4	0	0	0
$P2$	0	0	3	0	0
$P3$	0	0	0	4	5

(iv) Optimal service fee/hr (\$) quoted by an operator m' to operate a transit package p

Table 6.14: Optimal service fee/hr quoted by an operator m' to operate a transit package p in \$: $\eta_p^{m'}$

Package/Route	$M1$	$M2$	$M3$	$M4$
$P1$	686	448	0	0
$P2$	113	187	87	0
$P3$	356	0	571	571

The results demonstrate the optimal contracting strategy for the regulator. They state that the operator $M1$ secures packages 2 and 3, operator $M2$ secures package 1, whereas the operators $M3$ and $M4$ were unsuccessful in their bids. The optimal fleet size and route frequency ensure that the demand is catered efficiently with

minimum waiting time for the passengers. The optimal service fees quoted by the contesting operators ensure that the proposals are highly competitive and also the successful operators maximize their revenues such that the regulator's costs are also minimized. Since there was no limit on the number of bids a selected operator could contest for and operate if selected, operator 1 secures two packages.

6.3.3. Further analysis

More studies were conducted with change in demand and fare revenues and the results can be stated as below:

Table 6.15: Analysis of regulator cost/hr and packages' optimal fees/hr for different demand patterns

Demand	Regulator cost/hr	Packages' optimal service fees/hr
Demand D	61\$	\$448(P1), \$113(P2), \$356(P3)
D1: D x 150%	620\$	\$672(P1), \$220(P2), \$440(P3)
D2: D x 130%	400\$	\$582(P1), \$147(P2), \$427(P3)
D3: D x 80%	-150\$(earnings)	\$358(P1), \$90(P2), \$216(P3)
D4: D x 50%	-501\$(earnings)	\$297(P1), \$63(P2), \$135(P3)

Fig.6.2. Packages' optimal fees/hr for different demands

Fig.6.3. Plot of regulator costs/hr vs. total packages' service fees/hr for different demands

The Fig. 6.2 shows the plot of different packages' optimal fees for different demands. It is observed that with changing demand pattern, the optimal fees also change. In the graph, the optimal fee for the package 1 is the maximum followed by package 3 and 2 respectively. Also in the Fig. 6.3 plot of the regulator costs vs. total packages' optimal fees, it can be observed that for lower demand values; there are earnings for the regulator. Hence, demand pattern affects the overall finances in the contracting process and huge changes in these can disrupt the financial sustainability of the bus transit operation in the long run.

In this study, it is assumed that all operating services are similar in the sense that they serve the bus stops with similar inter-stop distances, congestion effects and the level of service of available transportation infrastructure. The routes used in the numerical study are real-time routes and are normal services serving bus stops at short spacing. However, if certain routes offer an express or a deadheading or a limited stop service, the transit fare structure and the demand pattern would change on account of different choice behaviour which remains a future scope of research. These services would have a lower turnaround time (time to complete one cycle) and a higher level of service; therefore, the fare structure for these services would be different from the normal service. Also, it is important to mention that depending on the local transit market scenario, the regulator may choose to allow or prevent operators from operating multiple transit packages. Further, the usual case of a CT where the proposal with minimum price quote secures the operation rights assuming all quality criteria are satisfied is followed. However, the regulator may choose to select a tender proposal based on any other basis, for e.g. a median quote accounting for risk budgeting etc.

6.4. Conclusion

This study presents an analytical framework to evaluate the optimal contracting strategy for the regulator in terms of how contracts for transit route packages should be awarded to contesting operators in a CC process and also design the route frequency and the fleet size. The framework also predicts the optimal service fee that each contesting operator should quote in the tender proposal to ensure a financially sustainable transit market. Hence, this model formulation is a good

decision making tool for the transit regulator to assess the transit market in advance when bus packages are being tendered out to operators. Although the tender proposals may not necessarily be similar to these results, however, the regulator can get a fair idea of what would constitute an optimal and a sustainable strategy and accordingly determine the gain/loss in the real implementation scenario. A bi-level programming approach is used to model the interaction between the regulator (master) and the operators (followers) in a Stackelberg game. Depending on how the transit regulator fixes its decision variables, the operators (followers) tend to revise their own decision variables (i.e. service fees) in order to maximize their own profits until both parties arrive at an equilibrium beyond which none profits by changing their decision variables unilaterally. Bi-level problems are considered to be one of the most difficult problems in transportation research, however, a solution algorithm is adopted through which the problem is first linearized and then reformulated into a single level problem using complementary constraints. The linearized single level MILP is tested on a real-time transit network to demonstrate its validity and the results are presented. The globally optimal solutions state the optimal contracting strategy and the design parameters which the regulator could use in order to operate sustainably. This helps the regulator to plan finances and ensure the operational aspects are in shape while maintaining attractiveness of bus transit for passengers. In the proposed model formulation, it is assumed that the regulator has a fair idea of the past track record of the operators' performance and hence, their operating rates are known to the regulator. Depending on this known data, the regulator is able to estimate the optimal contracting strategies. However, in the case of an open tender when the regulator does not choose operators who would contest for the operation of transit packages and a larger number of operators can bid, the rates are unknown and hence in such a case, estimation of the operating rates such that the CT leads to an optimal selection of the transit operator for a route package would be important. This could be a future scope of study. Another aspect of this study is that the service fees that operators quote in their respective tender proposals are unknown to each other; however, the operators behave as a combined stakeholder in the transit system and ensure that they maximize their revenues while maintaining a high degree of competitiveness in the transit market. This competition

could also be modelled using game theory as such a problem is a typical example of a non-cooperation game. However, due to the modelling complexity in invoking game theory in a bi-level formulation, this model does not include this aspect which could be another future scope of research.

As mentioned earlier, this modelling framework provides the transit regulator to assess the contracting process and strategically plan resources. With more transit markets transitioning to CT, this model formulation would prove to be a foundation towards decision making in bus transit contracting. With most of the financial issues in bus transit today related to profits of operators and subsidies payable by the regulator, this framework helps the regulator to analyse various aspects of the tendering process for an acute understanding and take good decisions.

CHAPTER 7 CONCLUSION AND RECOMMENDATIONS

This chapter summarizes the research questions addressed and major contributions made by this thesis. A few extensions for future research are also mentioned as recommendations.

7.1 Thesis summary

This research work deals with transit network design problems for a sustainable bus transportation system. A number of interesting transit design and regulatory problems in bus transit operation were considered and mathematical models were formulated along with their corresponding global optimization solution methods. All the models developed in the thesis together contribute towards an equitable decision making for a sustainable bus transit. Today, the state of bus transit in most cities of the world can be summarized as either (i) low ridership due to inferior service levels, or (ii) good service but supply-demand imbalance, or (iii) availability of good bus transit service but issues in contracting regime. This thesis provides transit regulators with a methodology to tackle each of the abovementioned challenges through innovative propositions and design of limited stop services, inclusion of congestion effects in transit network design, designing bus transit route packages and subsequently tendering them competitively to contesting operators through the government contracting regime.

7.2 Recommendations for future work

A few recommendations that could advance the work presented in this thesis are as follows:

- 1) To design the limited stop service in conjugation with normal service considering a competitive market. Principles of game theory and equilibrium problems with equilibrium constraints could be used to formulate such models. Estimation of waiting and travelling time costs is also an important area to investigate.

- 2) The transit route packaging problem could be extended using a bi-level approach by considering the perspective of the regulator and the operators. This would lead to a more comprehensive design that takes a better account of the financial interest of involved stakeholders.
- 3) A methodology that assesses the performance of the government contracting regime based on various key performance indicators would help transit regulators take better decisions in this regard.
- 4) With more transit markets facing financial unsustainability, a pertinent question facing regulators is what portion of the transit market should be contracted out to private players.
- 5) It is also important to decide an optimal competitive scenario and decide on the number of private transit players the regulator would want to operate in the transit market to ensure a premium service level.
- 6) Finally, a comprehensive comparative study of different transit contracting regimes that proposes the extent of suitability of a transit contracting regime for a particular transit market would be a very interesting research area and useful for transit planners across the world.

REFERENCES

- Abdulaal, M., LeBlanc, L.J. (1979), "Continuous equilibrium network design models", Transportation Research Part B: Methodological, Vol. 13, No.1, pp. 19-32.
- Afanasiev, L. L., Liberman, S. Y., (1983) "Principles for organizing express bus services" Transportation Research Part A: General, Vol. 17, No. 5, pp. 343-346.
- Allende, G. and Still, G. (2013) "Solving bilevel programs with the KKT-approach", Mathematical Programming, Vol 138, No. 1, pp. 309-332.
- Alonso, B., Moura, J.L.,Olio, L.D. and Ibeas, A. (2011) "Bus Stop Location under Different Levels of Network Congestion and Elastic Demand", Transport, Vol. 26, No. 2, pp. 141-148.
- Beale, E.M.L. and Tomlin, J.A. (1970) "Special facilities in a general mathematical programming system for non-convex problems using ordered sets of variables", Proceedings of the Fifth International Conference on Operation Research, pp. 447–454.
- Beckmann, M.J., McGuire, C.B. and Winsten, C.B. (1956) Studies in the economics of transportation, Yale University Press.
- Bell, M.G.H. and Iida, Y. (1997), Transportation Network Analysis.
- Boyce, D. (1984) "Urban transportation network-equilibrium and design models: Recent achievements and future prospects", Environment and Planning A, Vol. 16, No.11, pp. 1445-1474.
- Ceder, A. (1984) "Bus frequency determination using passenger count data", Transportation Research Part A: General, Vol. 18, No. 5–6, pp. 439-453.
- Ceder, A. and Wilson, N. H. M (1986) "Bus network design", Transportation Research Part B: Methodological, Vol. 20, No.4, pp. 331-344.

- Chien, S. I. and Qin, Z. (2004) "Optimization of bus stop locations for improving transit accessibility", Transportation Planning and Technology, Vol. 27, No. 3, pp. 211-227.
- Chiraphadanakul,V., Barnhart C., (2013) "Incremental bus service design: combining limited-stop and local bus services", Public Transport, Vol. 5, pp. 53-78.
- Chriqui C. (1974) Réseaux de transports en commun: les problèmes de cheminement et d'accès, Departement d'informatique, Université de Montréal.
- Chriqui, C. and Robillard, P. (1975) "Common Bus Lines", Transportation Science, Vol. 9, No. 2, pp. 115-121.
- Collier, K (1969) Construction Contracts, Reston, New York.
- Cominetti, R. and Correa, J. (2001) "Common-lines and passenger assignment in congested transit networks", Transportation Science, Vol. 35, No.3, pp. 250-267.
- Conlon, M., Foote, P., O' Malley,K. and Stuart, D., (2001) "Successful Arterial Street Limited-Stop Express Bus Service in Chicago", Transportation Research Record: Journal of the Transportation Research Board, Vol. 1760, pp. 74-80.
- Constantin, I. and Florian, M. (1995) "Optimizing frequencies in a transit network: a nonlinear bi-level programming approach", International Transactions in Operational Research, Vol. 2, No. 2, pp. 149-164.
- Cortés, C. E., Jara-Díaz, S. and Tirachini, A., (2011) "Integrating short turning and deadheading in the optimization of transit services", Transportation Research Part A: Policy and Practice, Vol. 45, No. 5, pp. 419-434.
- De Cea, J. and Fernandez, E. (1993) "Transit Assignment for Congested Public Trasnport Systems", Transportation Science, Vol. 27, No. 2, pp. 133-147.
- Eilenberg, I (1987) "Select tendering preferred by all", National Constructor, September, pp. 6-7.

El-Geneidy, A. and Surprenant-Legault, J. (2010) "Limited-stop bus service: an evaluation of an implementation strategy", Public Transport, Vol. 2, No. 4, pp.291-306.

Facchinei, F. and Soares, J. (1997) "A new merit function for nonlinear complementarity problems and a related algorithm" SIAM Journal on Optimization, Vol. 7, No. 1, pp. 225-247.

Farahani, R.Z., Miandoabchi, E., Szeto, W.Y. and Rashidi, H. (2013) "A review of urban transportation network design problems", European Journal of Operational Research, Vol. 229, No. 2, pp. 281-302.

Floudas, C. A. (1995), Nonlinear and Mixed-Integer Optimization: Fundamentals and Applications, Oxford University Press, USA.

Friesz, T.L. (1985) "Transportation network equilibrium, design and aggregation: Key developments and research opportunities", Transportation Research Part A: General, Vol. 19, No. 5–6, pp. 413-427.

Gao, Z., Sun H. and Shan L.L. (2004) "A continuous equilibrium network design model and algorithm for transit systems", Transportation Research Part B: Methodological, Vol. 38, No. 3, pp. 235-250.

Gao, Z., Wu, J. and Sun, H. (2005) "Solution algorithm for the bi-level discrete network design problem", Transportation Research Part B: Methodological, Vol. 39, No. 6, pp. 479-495.

Gendreau, M. (1984) Etude approfondie d'un modele d'équilibre pour l'affectation de passagers dans les réseaux de transports en commun, Publication 384, Centre de Recherche sur les Transports, Univ. de Montréal.

Hensher, D. A. and Stanley, J. (2003) "Performance-based quality contracts in bus service provision", Transportation Research Part A: Policy and Practice, Vol. 37, No. 6, pp. 519-538.

Hensher, D. A. and Stanley, J. (2008) "Transacting under a performance-based contract: The role of negotiation and competitive tendering", Transportation Research Part A: Policy and Practice, Vol. 42, No. 9, pp. 1143-1151.

Hensher, D. A. and Stanley, J. (2010) "Contracting regimes for bus services: What have we learnt after 20 years?", Research in Transportation Economics, Vol. 29, No. 1, pp.140-144.

Hensher, D. A. and Wallis, I. P. (2005) "Competitive Tendering as a Contracting Mechanism for Subsidising Transport: The Bus Experience", Journal of Transport Economics and Policy, Vol. 39, No.3, pp. 295-322.

Iseki, H. (2008) "Economies of scale in bus transit service in the USA: How does cost efficiency vary by agency size and level of contracting?", Transportation Research Part A: Policy and Practice, Vol. 42, No. 8, pp. 1086-1097.

Iseki, H. (2010) "Effects of contracting on cost efficiency in US fixed-route bus transit service", Transportation Research Part A: Policy and Practice, Vol. 44, No. 7, pp. 457-472.

Kikuchi, S. (1985), "Relationship between the number of stops and headway for a fixed-route transit system", Transportation Research Part A: General, Vol. 19, No. 1, pp. 65-71.

Kinderlehrer, D. and Stampacchia, G. (2000) "An Introduction to Variational Inequalities and Their Applications" Society for Industrial and Applied Mathematics.

Kirkpatrick, WG (1963) "Cost of tendering", Building Economist, November, pp. 63–66.

Larrain, H., Giesen R., Muñoz, J.C., 2010. Choosing the right express services for bus corridor with capacity restrictions.Transportation Research Record: Journal of the Transportation Research Board 2197, 63-70.

- Larrain, H., Muñoz, J.C., Giesen, R., 2015. Generation and design heuristics for zonal express services. Transportation Research Part E: Logistics and Transportation Review 79, 201–212.
- Leiva, C., Muñoz, J.C., Giesen, R. and Larrain, H., (2010) "Design of limited-stop services for an urban bus corridor with capacity constraints", Transportation Research Part B: Methodological, Vol. 44, No. 10, pp. 1186-1201.
- Li, Z.C., Lam, W.H.K., Wong, S.C., Sumalee, A., (2011) "Design of a rail transit line for profit maximization in a linear transportation corridor", Procedia - Social and Behavioral Sciences, Vol. 17, pp. 82-112
- Li, Z.C., Lam, W.H.K. and Wong, S.C. (2010) "Optimization of Number of Operators and Allocation of New Lines in an Oligopolistic Transit Market", Networks and Spatial Economics, Vol. 12, No. 1, pp.1-20.
- Li, Z.C., Lam, W.H.K. and Sumalee, A. (2008) "Modeling Impact of Transit Operator Fleet Size Under Various Market Regimes with Uncertainty in Network", Transportation Research Record: Journal of the Transportation Research Board, Vol. 2063, pp. 8-27.
- Liu, Z. and Meng, Q. (2014) "Bus-based park-and-ride system: a stochastic model on multimodal network with congestion pricing schemes", International Journal of Systems Science, Vol.45, No. 5, pp. 994-1006.
- Löfberg, J. (2004) "YALMIP: A Toolbox for Modeling and Optimization in MATLAB", Proceedings of the CACSD Conference, Taipei, Taiwan.
- Luathep, P., Sumalee, A., Lam, W. H. K., Li, Z.C., Lo, H.K. (2011) "Global optimization method for mixed transportation network design problem: A mixed-integer linear programming approach", Transportation Research Part B: Methodological, Vol. 45, No. 5, pp. 808-827.
- Mesbah, M., Sarvi, M., Ouyeyi, I., & Currie, G. (2011). Optimization of transit priority in the transportation network using a decomposition methodology. Transportation Research Part C: Emerging Technologies, Vol. 19, (2), Pp. 363-373.

Misener, R. and Floudas. C.A. (2010) "Piecewise-Linear Approximations of Multidimensional Functions", Journal of Optimization Theory and Applications, Vol.145, No.1, pp. 120-147.

Mouwen, A. and Rietveld, P. (2013) "Does competitive tendering improve customer satisfaction with public transport? A case study for the Netherlands", Transportation Research Part A: Policy and Practice, Vol. 51, pp. 29-45.

News Release (2014) Transition to a government contracting model for the public bus industry. Land Transport Authority.

Patriksson, M. (1994) The Traffic Assignment Problem: Models and Methods. VSP.

Schaaffkamp, C. (2014) "How can customer focus be strengthened in competitive tendering?", Research in Transportation Economics, Vol. 48, pp. 305-314.

Sheffi, Y. (1985), "Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods", Englewood Cliffs, NJ 07632, USA.

Sherali,H. and Alameddine, A.(1992) "A new reformulation-linearization technique for bilinear programming problems", Journal of Global Optimization, Vol. 2, No.4, pp. 379-410.

Shimamoto, H., Schmöcker, J. and Kurauchi, F. (2012) "Optimisation of a Bus Network Configuration and Frequency Considering the Common Lines Problem", Journal of Transportation Technologies, Vol. 2 , No. 3, pp. 220-229.

Silverman, N. C. (1998) "Limited-Stop Bus Service at New York City Transit", Journal of Transportation Engineering, Vol. 124, No. 6, pp. 503-509.

Solanki, R.S., Gorti, J.K. and Southworth, F. (1998) "Using decomposition in largescale highway network design with a quasi-optimization heuristic", Transportation Research Part B: Methodological, Vol. 32, No. 2, pp.127-140.

Speiss, H., Florian, M. (1989) "Optimal Strategies: A New Assignment Model for Transit Networks" Transportation Research Part B: Methodological, Vol. 23, No. 2, pp. 83-102.

- Sun, L., Meng, Q. and Liu, Z. (2013) "Transit Assignment Model Incorporating Bus Dwell Time", Transportation Research Record: Journal of the Transportation Research Board, Vol. 2352, pp. 76–83.
- Suwansirikul, C., Friesz, T.L. and Tobin, R.L. (1987) "Equilibrium decomposed optimization: A heuristic for the continuous equilibrium network design problem", Transportation Science, Vol. 21, No.4, pp. 254-263.
- Tan, H.N., Gershwin, S.B. and Athans, M. (1979) Hybrid optimization in urban traffic networks. Cambridge, MA.
- Tétreault, P. R., El-Geneidy, A. M. (2010) "Estimating bus run times for new limited-stop service using archived AVL and APC data", Transportation Research Part A: Policy and Practice, Vol. 44, No. 6, pp. 390-402.
- Transport Act (1985), The National Archives, Chapter 67.
- Uher, T. E. and Philip, D. (2009), Fundamentals of Building Contract Management (2nd Edition). New South Publishing.
- Ulusoy, Y., Chien, S., Wei, C. (2010) "Optimal all-stop, short-turn, and express transit services under heterogeneous demand" Transportation Research Record: Journal of the Transportation Research Board, Vol. 2197, pp. 8–18.
- Wang, D. Z. W. and Lo, H. K. (2010) "Global optimum of the linearized network design problem with equilibrium flows", Transportation Research Part B: Methodological, Vol. 44, No. 4, pp. 482-492.
- Wang, D.Z.W. and Lo, H.K. (2008) "Multi-fleet ferry service network design with passenger preferences for differential services", Transportation Research Part B: Methodological, Vol. 42, No. 9, pp. 798-822.
- Wallis, I. P. and Bray, D. J. (2014) "The contracting of urban bus services – Recent Australian developments", Research in Transportation Economics, Vol. 48, pp. 48-61.

Walters, J. and Cloete, D. (2008) "The South African experience with negotiated versus competitively tendered bus contracts", Transportation Research Part A: Policy and Practice, Vol. 42, No. 9, pp. 1163-1175.

Wardrop, J.G. (1952) "Some theoretical aspects of road traffic research", Institution of Civil Engineers, pp. 325-362.

Yadan, Y., Meng, Q. ,Wang, S. and Guo, X. (2012) "Robust optimization model of schedule design for a fixed bus route", Transportation Research Part C: Emerging Technologies, Vol. 25, pp.113-121.

Yang, H., Bell, M.G.H. (1998) "Models and algorithms for road network design: A review and some new developments", Transport Reviews, Vol. 18, No. 3, pp. 257-278.