SYSTEM OPTIMAL TRAFFIC ASSIGNMENT

WITH DEPARTURE TIME CHOICE

Ho Fai Andy Chow

A thesis submitted to the University of London for the degree of Doctor of Philosophy

Centre for Transport Studies

University College London

March 2007

DECLARATION

I, *Ho Fai Andy Chow*, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Ho Fai Andy Chow

ABSTRACT

This thesis investigates analytical dynamic system optimal assignment with departure time choice in a rigorous and original way. Dynamic system optimal assignment is formulated here as a state-dependent optimal control problem. A fixed volume of traffic is assigned to departure times and routes such that the total system travel cost is minimized. Although the system optimal assignment is not a realistic representation of traffic, it provides a bound on performance and shows how the transport planner or engineer can make the best use of the road system, and as such it is a useful benchmark for evaluating various transport policy measures. The analysis shows that to operate the transport system optimally, each traveller in the system should consider the dynamic externality that he or she imposes on the system from the time of his or her entry. To capture this dynamic externality, we develop a novel sensitivity analysis of travel cost. Solution algorithms are developed to calculate the dynamic externality and traffic assignments based on the analyses. We also investigate alternative solution strategies and the effect of time discretization on the quality of calculated assignments. Numerical examples are given and the characteristics of the results are discussed.

Calculating dynamic system optimal assignment and the associated optimal toll could be too difficult for practical implementation. We therefore consider some practical tolling strategies for dynamic management of network traffic. The tolling strategies considered in this thesis include both uniform and congestion-based tolling strategies, which are compared with the dynamic system optimal toll so that their performance can be evaluated. In deriving the tolling strategies, it is assumed that we have an exact model for the underlying traffic behaviour. In reality, we do not have such information so that the robustness of a toll calculation method is an important issue to be investigated in practice. It is found that the tolls calculated by using divided linear traffic models can perform well over a wide range of scenarios. The divided linear travel time models thus should receive more attention in the future research on robust dynamic traffic control strategies design. In conclusion, this thesis contributes to the literature on dynamic traffic modelling and management, and to support further analysis and model development in this area.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my supervisor, Professor Benjamin Heydecker, for his supervision and encouragement. The countless hours he spent with me on this research are gratefully appreciated. I would also like to acknowledge my examiners, Professor Frank Kelly (Cambridge) and Dr. Suzanne Evans (Birkbeck, London), who contributed tremendously in the viva. I would like to thank my secondary supervisor, Professor Roger Mackett, for the useful discussion in the early stage of the research. The stimulating discussion, advice, and encouragement from Professor Richard Allsop and Dr J.D. Addison are also gratefully acknowledged. I am grateful to Dr. Hale Cuss from the UCL Language Centre who painstakingly read and commented earlier drafts of the thesis and improved its presentation. I am also grateful to my colleagues at the Centre for Transport Studies at UCL for contributing a stimulating and enjoyable working environment. I would like to particularly thank Chen Cai, Taku Fujiyama, Liselot Hertel, Annji Kinoshita-Bashforth, A.Q. Memon, Dr. Sanjay Rana, and Marcel Rommerts for their support and friendship. The financial support from University College London is also acknowledged.

I am privileged to have various opportunities to work with colleagues around the world from which I have benefited profoundly. In particular, I would like to thank Professor M.J. Smith (York), Professor Malachy Carey (Belfast), Professor David Watling (Leeds), Dr. Simon Shepherd (Leeds), Dr. Narasimha Balijepalli (Leeds), Dr. Agachai Sumalee (Leeds, currently Hong Kong), Dr. Iryo Takamasa (Kobe, Japan), and Professor Terry Friesz (Pennsylvania State, USA) for all the stimulating discussions we have had. I would also like to thank my colleagues in Hong Kong for their support and encouragement. In particular, I am indebted to my MPhil advisor, Dr. Hong K. Lo, whose enlightenment is instrumental to me in succeeding in London. Special thanks are also due to Professor S.C. Wong, Dr. W.Y. Szeto (currently Singapore), and Dr. K.I. Wong (currently Taiwan) for their advice and friendship.

Finally, I owe a great debt to my mother, my brother, and Grace Wong for their love, patience, and support throughout my study. This thesis is dedicated to my mother and *in memoriam* my father for their devotion to their children who have been enjoying the best education and the best quality of life.

Two roads diverged in a wood, and I-
I took the one less travelled by,
And that has made all the difference.

Robert Frost (1920)

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1 GENERAL BACKGROUND

Population growth and economic developments tend to increase the volume of personal and commercial interactions among people. These interactions involve the movement of people or goods from place to place which mostly use the road network for some part of their journey. In 1998, road transport accounted for 44% of all goods traffic and 79% of passenger traffic. In particular, the road share of the goods traffic has been growing constantly, and is expected to reach 47% by 2010 (European Commission, 2001).

Unfortunately, heavy road traffic induces problems of pollution, road congestion, and incident management in every metropolis. The white paper published by European Commission (2001) reported that transport was responsible for 28% of carbon dioxide emissions in Europe, of which road transport alone accounts for 84 %. Heavy road traffic also causes congestion and travel delays. It is not unusual to observe vehicles crawling slowly along busy streets in urban areas. Heavy traffic can also complicate the management of unexpected incidents. An example of this was on 9 May 2005, when three incidents happened in Hong Kong: a fallen tree across Waterloo Road, loose scaffolding at Argyle Street, and fallen scaffolding at Prince Edward Road East. These incidents, together with the heavy daily traffic, induced heavy congestion and hence delays for tens of thousands of travellers (Cheng et al., 2005).

Managing the ever increasing amounts of road traffic is important for the economy. However, simply expanding and improving existing road networks are often restricted by increasingly tight fiscal, physical and environmental constraints (Hau, 1998). Given these constraints, transport scientists, engineers, and planners have to design and implement effective strategies to manage the existing transport facilities. To achieve this, a reliable way to evaluate the travellers' likely response to traffic management measures will be essential: the importance of this has been highlighted by the extreme example of Braess' paradox (Braess, 1968; Murchland, 1970; Kelly, 1991). This paradox refers to a case in which expanding a road network, supposing traffic flows either to be constant or to respond neutrally can lead to decisions that, whilst intended and expected to improve network performance, would cause deterioration. Murchland (1970) also quoted a real experience from Knödel (1969). Knödel (1969) reported that there were major road investments in the city centre of Stuttgart. However, it was found the road construction project failed to yield the expected reduction in travel delay. Eventually, a cross street in the city network had subsequently to be withdrawn from traffic use in order to gain that delay reduction.

1.2 METHODS FOR EVALUATING TRAFFIC MANAGEMENT POLICIES

There are several approaches proposed in the literature to estimate travellers' response to traffic management policies. Certainly, the most direct way is to implement the policy and then to observe the associated effects. However, because of the high costs and risks of implementation and observation, this method is not often considered to be practical, at least in the early stages of the evaluation (Heydecker, 1983).

The second approach is through computer simulation. Simulation methods apply some predefined rules to estimate the resulting effects on traffic after implementing the management policy. The simulation approach is quite popular in practice because it can be sophisticated and can capture fine detail of the real world system that is to be investigated. Some of the most popular simulation models for dynamic traffic assignment include DYNASMART (Mahmassani, 2001), DYNAMIT (Ben-Akiva et al., 2001), and CONTRAM (Taylor, 2003). However, simulation models do not give much information on the underlying mechanisms of the system. Furthermore, calibrating and running simulation models can be computationally demanding due to their complicated nature.

The third approach is through analytical models. In contrast with simulation methods, analytical models are built entirely on mathematical equations and inequalities. These models serve as simplified representations of a part of the real world system, and they only concentrate on certain elements considered to be important for a particular analysis (Ortúzar and Willumsen, 2001). The analytical models have well-defined formulation and properties to analyse. It is also widely recognised that analytical models are more useful for transport planning due to their relative simplicity and lower labour costs for implementation (Friesz and Bernstein, 2000; Ortúzar and Willumsen, 2001). In addition to this, analytical models can offer a common ground for discussing policy and examining the inevitable compromises required in practice with minimum objectivity (Ortúzar and Willumsen, 2001).

1.3 ANALYTICAL TRANSPORT MODELS

In the framework of analytical models, a transport system is often simplified into a form of network and zoning systems. The term *network* refers to a structure in which there are two types of elements: a set of nodes and a set of links that join some pairs of nodes. In a detailed network model, the nodes in a network model represent individual road junctions. Each link corresponds to a section or road; in more aggregate nodes, links can represent collections of roads. The topology of network is specified by the presence or absence of links between nodes which determine the possibility of travel from one place to the other (Heydecker, 2005). There are several attributes associated with each link in the network model to define the characteristics of that link. The most commonly used attributes include link length, free flow travel time, and link capacity. With these attributes, the delay, number of stops and travel time on each link can then be estimated according to the flow of traffic carried by the link. Various functional forms have been proposed to model the relationship between link travel times and traffic flows. In general, because an increase in link traffic flow will normally decrease the travel speed along the link, travel times are usually considered to be positive monotonic increasing functions of traffic flow. Parameters in the travel time functions often include free flow travel times (i.e. link travel times when there is no traffic on links) and link capacities (i.e. maximum values of traffic flow along the link). Some examples of travel time functions can be found in Patriksson (1994, p20) and Mun (2002). In addition to links, the term *route* or *path* is defined to represent a sequence of directed links leading from one node to another. The corresponding travel time along a route can be determined as the sum of the travel times along the links comprising that route, within which each of the link travel time is calculated according to their corresponding time of entry.

The term *zone* in the zoning system refers to a partition of an urban area. Within each of these zones, various data can be collected for calibrating and validating the transport model. These data include demographic features of people in the zone and levels of economic activity including employment, shopping space, educational and recreational facilities (Ortúzar and Willumsen, 2001). Each zone is represented in the network by a special node called a *centroid*. Each centroid can either be an *origin* node from which traffic enters the network, or a *destination* node to which traffic leaves the network.

After building a representation of the transport system, analysis and planning procedure can then be carried out. The classic procedure of analysis and planning in transport practice, known as the *four-stage model*, is shown in Figure 1.1. The four stages are trip generation, trip distribution, modal split, and assignment. The four-stage model starts with estimating the total number of trips generated by each zone based on the data of the levels of economic activity in that zone. The next stage is to distribute these trips from their origins to particular destinations. The following stage, modal split is an estimation of the choice of transport modes, such as car, underground train, or bus, of the trips. The final stage, assignment, is to estimate how the trips travel through the network, the traffic flows generated, the resulting traffic conditions, and the costs of travel for each origin-destination pair. A detailed discussion on the four-stage model can be found in Ortúzar and Willumsen (2001).

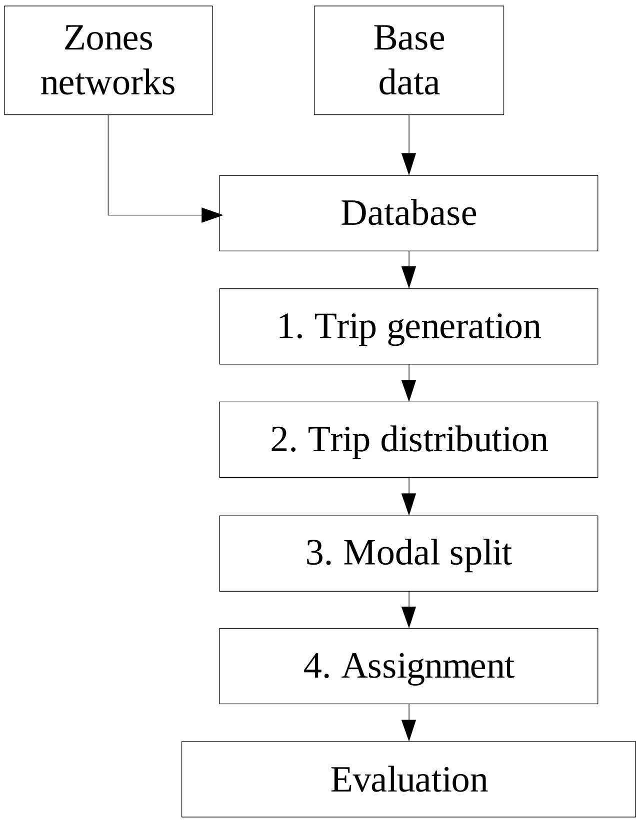


Figure 1.1 The classic four-stage transport planning model

1.4 TRAFFIC ASSIGNMENT

A traffic assignment model aims to estimate how traffic flows through a road system and the associated effects of traffic on the system. These effects can be measured by a number of criteria including distance travelled, travel time, delay, fuel consumption and environmental pollution (Heydecker, 2005). Traffic assignment models can also be used to investigate the responses of traffic to changes in the system (for example, changes in travel demand, travellers' information, road capacities, signal timings, and road tolls).

Formulating and solving a traffic assignment model requires three kinds of information. The two of these are the demand for travel and the characteristics of transport system. The demand for travel, which is estimated by the three earlier stages of the four-stage model, represents the likely travel decisions that travellers would make, given the performance of the transport system. Following the first three steps in the four-stage model, the travel decisions considered include choices of destination, mode, frequency of trip, and even whether to travel at all (IHT, 1997, p91). It should be noted that although population, land-uses, and other factors could vary over time, so does the travel demand. Conventional planning models only consider the travel demand within a particular period of time and the demand is regarded as time-independent throughout that time period. The second component of a traffic assignment formulation is a network model of the characteristics of transport system. The function of this network model is to define the relationship between the travel demand and the performance of the transport system. For example, travel times are modelled as increasing with travel demand, due to the decreases in travel speeds of vehicles (IHT, 1997, p91).

Given the demand for travel and the characteristics of a transport system, the third kind of information is a way of estimating the corresponding distribution of the travel demand over the transport system. The most widely accepted way is through the two principles of traffic assignment proposed by Wardrop (1952). Wardrop adopted the supply-demand equilibrium concept of economics, which suggests that travel demand should be balanced against the performance of the transport system in servicing that level of demand. This gives Wardrop's (1952) first principle, or the *user equilibrium* principle:

> "the journey times on all the routes actually used are equal, and less than those which would be experienced by a single vehicle on any unused route."

The underlying assumption of this principle is that all travellers are supposed to choose their routes of travel through the network according to the common criterion that their individual journey times are minimized. In addition, all travellers will experience the same journey times if they encounter identical traffic conditions. Furthermore, all travellers will have perfect information on all possible routes through network, no matter whether the routes are used or not.

In fact, this concept of equilibrium is found to be a powerful tool for analysing transport system, as Bell and Iida (1997) wrote:

> "While a transport system may never actually be in a state of equilibrium, it is assumed that it is at least near equilibrium, tending towards equilibrium, and only prevented from attaining equilibrium by changes in external factors… At equilibrium, the transport system reduces to a fixed point (the equilibrium costs and flows), and powerful analytical techniques … exist for finding the fixed point. Proponents of equilibrium theory take it as a matter of faith that, given the existence of an equilibrium, there are behavioural mechanisms that push the transport system to this fixed point."

Although user equilibrium may be a good representation of distribution of existing network traffic, such distribution of traffic generally does not lead to the best possible use of the network system. This is because user equilibrium considers that each individual traveller is acting only in their own interests, but not necessarily in the interest of the system as a whole.

In fact, the discrepancy between the behaviour of individual travellers acting on their own interests and the interests of the whole community is known in economic literature as the "divergence between private cost and social cost". It was first raised and discussed by Pigou (1920) and Knight (1924). In accordance with this observation, Wardrop (1952) further proposed his second principle of traffic assignment to describe how travellers could be allocated centrally to minimize the total cost incurred by all travellers. Wardrop's (1952) second principle or the system optimal principle is:

> "the average journey time is a minimum."

Under system optimum, some travellers may be assigned to routes that have costs higher than the minimal that they could travel along. This is because the additional costs incurred by such travellers will be outweighed by the greater savings that accrue to the others. The user equilibrium and system optimal principles will produce identical results when the network is uncongested (Sheffi, 1985, p72). Although the system optimal assignment is not a realistic representation of network traffic, it provides a bound on how we can make the best use of the road system, and as such it is a useful benchmark for evaluating various traffic control policies. Using economic terminology, the user equilibrium and the system optimal assignments respectively represent the descriptive (positive) and normative representations of traffic flow patterns on road networks.

1.4.1 Formulations of traffic assignment

A traffic assignment model should be formulated in mathematical terms before it can be analysed and solved numerically. User equilibrium traffic assignment can be stated equivalently as the following complementary inequality for the route flow e_p:

$$e_p \begin{Bmatrix} > 0 \Rightarrow C_p = C_{od}^* \\ = 0 \Rightarrow C_p \geq C_{od}^* \end{Bmatrix} \qquad \forall p \in P_{od}, \quad \forall od \tag{1-1}$$

where e_p is the flow assigned to route p, P_{od} is the set of all routes from origin o to destination d, C_p is the travel time along route p, and C_{od}^* is the minimum travel time from o to d.

Beckmann et al. (1956) were the first to transform the user equilibrium principle into a mathematical programming problem for link flow v_a.

$$\min_{\mathbf{v}} \sum_{a \in L} \int_{v=0}^{v_a} c_a(v) dv \tag{1-2}$$

subject to

$$\sum_{p\in P_{od}} e_p = E_{od} \qquad \forall od \tag{1-3}$$

$$e_p \geq 0 \qquad \forall p \in P_{od} \quad \forall od\,. \tag{1-4}$$

The notation v_a represents the flow of traffic on link a, $c_a(\cdot)$ is the travel time along link a and is a function of link flow v_a, E_{od} represents the traffic flow between origin o to destination d. It is noted that the objective function is formulated in terms of link flows, while the constraints are formulated in terms of route flows. Hence, the following definitional constraint is required to inter-relate the link flows and route flows

$$v_a = \sum_{od} \sum_{p\in P_{od}} e_p \delta_p^a \qquad \forall a\,, \tag{1-5}$$

where δ_p^a is a indicator variable:

$$\delta_p^a = \begin{cases} 1 & \text{if link } a \text{ is on route } p \\ 0 & \text{otherwise} \end{cases}. \tag{1-6}$$

Constraint (1-5) is also part of the optimization program (1-2).

Beckmann (1956) showed that solving this mathematical programming formulation is equivalent to solving the static user equilibrium assignment problem (1-1). The equivalency can be proven by verifying that the Karush-Kuhn-Tucker (KKT) necessary conditions for a minimum point of the problem (see Sheffi, 1985, pp63 – 66; Patriksson, 1994, pp35-36) are exactly the conditions of user equilibrium. Since its introduction, the transformation technique by Beckmann (1956) is now standard and well-known in transport literature, and hence its mathematical details are not shown here for brevity. A range of efficient solution algorithms were later developed, and they can be employed to solve Beckmann et al.'s (1956) mathematical programming formulation and its extensions effectively. Examples of the algorithms can be found in Evans (1976), Lee (1995), and Bar-Gera (2002).

The system optimal assignment can also be formulated mathematically as a static minimization problem of the total system journey time spent in the network:

$$\min_{\mathbf{v}} \sum_{a \in L} v_a c_a(v_a) \tag{1-7}$$

subject to constraints (1-2) – (1-4).

The optimality conditions of the system optimal assignment are given in Sheffi (1985, pp69 – 72) as

$$e_p \begin{cases} > 0 \Rightarrow C_p + \sum_{\forall a} \delta_p^a v_a \dfrac{\partial c_a(v_a)}{\partial v_a} = C_{od}^* \\ = 0 \Rightarrow C_p + \sum_{\forall a} \delta_p^a v_a \dfrac{\partial c_a(v_a)}{\partial v_a} \geq C_{od}^* \end{cases} \qquad \forall p \in P_{od}, \quad \forall od\,. \tag{1-8}$$

The quantity $C_p + \sum_{\forall a} \delta_p^a v_a \dfrac{\partial c_a(v_a)}{\partial v_a}$ is interpreted as the marginal contribution of an additional traveller on route *p* to the total travel time on that route *p*. The derivative of link travel time with respect to the link flow , $\dfrac{\partial c_a(v_a)}{\partial v_a}$, represents the additional travel time induced by an additional traveller to each of the existing travellers on the link. When the transport network is at system optimum, this *marginal* travel time on all used routes connecting each origin-destination pair in the network is equal. There are also many efficient solution algorithms in the literature that can be employed to solve the system optimization problem. A comprehensive review on these optimization algorithms can be found in Luenberger (1984), and Bazaraa, Sherali and Shetty (1993).

1.4.2 Limitations of traffic assignment models

In addition to the theoretic work, the four-stage model has also been made operational through numerous empirical studies (Small, 1992, p111; Ortúzar and Willumsen, 2001, p23), and has become the core of many kinds of commercial software. The software has been valuable for transport engineers and planners through providing important insights and useful estimations of travellers’ response to various transport policies.

Nevertheless, it should be noted that the traffic assignment model in the traditional four-stage planning procedure adopted a steady-state approximation. On the characteristics of the transport system, the network model considers traffic flows and travel times to be time-invariant and travel demand to be below the physical capacity of the transport network. However, traffic flows and travel times are dynamic in nature. In addition, there is a possibility that during some parts of the day, the travel demand will exceed the capacity of the network. This temporary overloading cannot be represented by static models in a satisfactory manner (Heydecker and Addison, 2005).

On the travel demand side, the steady-state traffic assignment model specifies the demand for travel to a particular time period under consideration, and treats it as constant over that period of the day. This treatment could mask any systematic variation in travel demand over time of the day. Indeed, empirical studies (Hendrickson and Plank, 1984; Small, 1992) confirmed that travellers do change their times of departure subject to the traffic conditions they encounter, especially during morning and evening peak periods. This temporal variation in travel demand, which is known as the *peak spreading* phenomenon, cannot be captured by steady-state traffic assignment models.

Finally, Patriksson (1994, p59) also commented on the steady-state assumption of static traffic assignment models as follows:

> "the fundamental principles underlying the assignment models were stated some forty years ago. The traffic flows in the then relatively uncongested urban networks were probably suitable for approximation by steady-state flows, as Wardrop did. Since those days, the traffic networks have become much more complex and the demand for transportation has become orders of magnitude higher, and the approximation of present traffic flows by steady-state flows is far less realistic."

Accordingly, various dynamic versions of traffic assignment models have been proposed in the literature in which travel demand and travel costs are considered to be varying over time.

1.5 DYNAMIC TRAFFIC ASSIGNMENT MODELS

In accordance with the comments on static traffic assignment models, there is a genuine need for developing more robust and sophisticated traffic assignment models. In the early 1960s, Vickrey (1963) suggested the importance of developing techniques for analysing time-varying road traffic and implementing time-varying traffic management strategies. He wrote (Vickrey, 1963, p452),

> "I will begin with the proposition that in no other major area are pricing practices so irrational, so out of date, and so conducive to waste as in urban transportation. Two aspects are particularly deficient: the absence of adequate peak-off differentials and the gross underpricing of some modes relative to others. In nearly all other operations characterized by peak load problems, at least some attempt is made to differentiate between the rates charged for peak and for off-peak service. Where competition exists, this pattern is enforced by competition: resort hotels have off-season rates; theatres charge more on weekends and less for matinees. Telephone calls are cheaper at night… But in transportation, such differentiation as exists is usually perverse."

Vickrey (1969) later proposed his innovative *bottleneck model*[*] for analysing dynamic traffic pattern. In the bottleneck model, traffic congestion was assumed to take the form of queuing behind a bottleneck of fixed flow capacity, on a single travel link connecting a single origin-destination pair. In the model, each identical traveller was assumed to commute in his or her own car from home (i.e. origin) to work (i.e. destination) along the single travel link. All travellers wish to arrive at work at the same time, which is impractical to achieve because the capacity of the bottleneck is finite. As a result, some travellers have to arrive earlier and some arrive later. The cost of arriving earlier or later than the desired arrival time was called *schedule delay cost*. Each traveller will make his or her choice of time of departure in order to minimize the associated *total travel cost*, which is essentially a cost associated with the time spent on travel plus the schedule delay cost. The equilibrium is achieved if all travel can be made at the same total travel cost. This means that in equilibrium, travellers will trade off changes in schedule delay costs against those in travel time. Those who travel off-peak so as to achieve short journeys do so at the expense of travelling at relatively unfavourable times, which is represented through schedule delay costs. On the other hand, those who arrive close to their desired time do so at the expense of a relatively long journey. Following Vickrey

[*] The bottleneck model is also known as the deterministic queuing model which is the name to be referred in the later parts of the thesis (see Section 2.3.3.1).

(1969), authors including Yagar (1971), Hurdle (1974), and Merchant and Nemhauser (1978a; b) have acknowledged the importance of Vickrey's (1969) work and have contributed to the dynamic transport models.

Nevertheless, the significance of the dynamic models was not widely acknowledged until the inception of intelligent transportation systems and the technological advances in traffic control systems in the early 1990s. The application of intelligent transport systems (ITS) has shown their ability to improve transport networks in many ways by providing information and guidance to travellers. The benefits of ITS include reducing travel times in lightly-loaded conditions, and increasing capacity and hence reducing travel times in more heavily loaded ones. For example, Adler (2001) showed how travel times could be reduced by about 1 minute in a 15-minute journey through providing advanced traffic information and route guidance. Rajamani and Shladover (2001) showed that ITS technologies could be used to provide autonomous adaptive cruise control systems that increase road capacity from about 2,000 to about 3,000 vehicles per lane-hour. A more detailed review on ITS can be referred to Heydecker (2002a). In addition to ITS, designing and implementing innovative traffic control systems and policy also require dynamic traffic assignment models to estimate travellers' likely response. Some examples of these control strategies include network access control (see for example, Smith and Ghali, 1990a; b; Lovell and Daganzo, 2000; Erera et al., 2002), network design and road capacity management (see for example, Ghali and Smith, 1993; Arnott, De Palma and Lindsey, 1993; Heydecker, 2002b), and time-varying road pricing (see for example, Yang and Huang, 1997; Wie and Tobin, 1998; Ettema et al., 2006). Due to these genuine needs, dynamic traffic assignment problems have become a popular and important research topic in both academia and industry in the last two decades.

Following Wardrop's (1952) principles, dynamic traffic assignment can be formulated through two approaches: dynamic user equilibrium and dynamic system optimal assignments. In the literature, dynamic user equilibrium assignment has been being the focus of research. The formulations of dynamic user equilibrium assignment can be grouped into five categories:

1. Mathematical programming (see for example, Janson, 1991; Ran and Boyce, 1996; Han and Heydecker, 2006);

2. Optimal control theory (see for example, Merchant and Nemhauser, 1978a; b; Friesz et al., 1989; Papageorgiou, 1990; Wie et al., 1994; Wie et al., 1995 a, b; Yang and Huang, 1997);
3. Non-linear complementarity problem (see for example, Wie et al., 2002);
4. Fixed point problem (see for example, Addison and Heydecker, 1993; Heydecker and Addison, 1996);
5. Variational inequality (see for example, Friesz et al., 1993; Ran and Boyce, 1996; Lo and Szeto, 2002).

Following the success in tackling static traffic assignment problem, much work on dynamic user equilibrium assignment attempted to use a mathematical programming approach. Janson (1991) proposed a mathematical programming formulation by integrating Beckmann's (1956) equilibrium objective function with respect to time. However, as later pointed out by Lin and Lo (2000), and Boyce, Lee and Ran (2001), the formulation of Janson's (1991) mathematical programme cannot capture the traffic dynamics, the temporally asymmetric nature of dynamic traffic cost functions, and the time-dependent interaction between traffic flows and travel times. Lin and Lo (2000) also showed with simple counter-example that solving Janson's (1991) formulation does not necessarily lead to a solution that satisfies a dynamic user equilibrium condition. Recently, Han and Heydecker (2006) have reformulated Beckmann's (1956) mathematical programme and have addressed the problem raised by Lin and Lo (2000) in Janson's (1991) formulation. However, Han and Heydecker's (2006) formulation can be too cumbersome for practical implementation. In addition, their formulation has yet to be applied to networks in which interactions between flows from different origin-destination pairs are involved.

The optimal control theory is a widely recognised tool for analysing dynamic systems. Following the pioneering work of Merchant and Nemhauser (1978a; b), many researchers (see for example, Carey, 1986, 1987; Friesz et al., 1989; Carey, 1992; Ran and Boyce, 1996; Yang and Huang, 1997; Wie and Tobin, 1998) have adopted their optimal control theoretic formulation and have produced many important insights on the behaviour and management of time-varying network traffic. However, Merchant and Nemhauser (1978a; b) incorporated an outflow traffic model into their analysis, which has later been criticized as being implausible and unrealistic in its representation of traffic dynamics (see Chapter 2).

Friesz et al. (1993) were the first to formulate and analyse the dynamic user equilibrium traffic assignment problem using variational inequalities. As shown by Patriksson (1994) and Nagurney (1993), variational inequalities can be regard as a generalization of mathematical programming, non-linear complementarity problem, and fixed point problem. Due to their generality, variational inequalities have attracted a lot of attention as a means of formulating and analysing dynamic traffic assignment. Detailed discussions on formulation of variational inequality can also be found in Friesz et al. (1996), Ran and Boyce (1996), and Nagurney (1993).

Dynamic user equilibrium is used to represent the distribution of traffic that arises when travellers consider their own interests alone. However, as discussed in Section 1.4, such distribution of traffic generally does not lead to the best possible use of the transport system, because the user equilibrium considers that each individual traveller is acting only in their own interests, rather than those of the community. Dynamic system optimal assignment, in contrast, considers that there is a central *system manager* distributing the traffic over time within a fixed horizon so that the total, rather than individual, benefit of all travellers in the system is maximised.

Analytical dynamic system optimal assignment is an important yet underdeveloped area and indeed it is one of the most challenging areas in transportation research. Different from its static counterpart (1-7), dynamic system optimal assignment is a kind of dynamic optimization problem, which aims to calculate an *optimal time path* for the decision variables instead of a single *optimal value* as in the static case. As noted by Dorfman (1969), such a problem is difficult to solve and “is not for beginners”. In addition, the challenges associated with dynamic system optimal assignment problem are also due to the range of interrelated requirements on their components (i.e. travel demand, characteristics of road system, and the way in which traffic is distributed) to perform in a satisfactory manner. As twelve years ago, Patriksson (1994) wrote:

> “So far, no well-founded dynamic models free from any serious anomaly such as instant propagation of some travellers, infinite cycling, failure to recognize the first-in-first-out principle, etc., have appeared, and their numerical solution most often rely on a time-discretization which brings the dynamic model into a (typically very large) static one.”

Merchant and Nemhauser (1978a; b) were the first to formulate and analyse dynamic system optimal assignment. Merchant and Nemhauser's (1978a; b) formulation was then followed and modified by many others (see for example, Ho, 1980; Carey, 1987; Friesz et al., 1989; Yang and Huang, 1997; Wie and Tobin, 1998). However, these previous studies used an outflow traffic model, whose plausibility was later found to be questionable. Addison and Heydecker (1998) used an alternative calculus of variations technique to analyse and calculate the system optimal assignment with departure time choice. However, the calculus of variations is complicated to use and to implement.

1.6 OBJECTIVE AND OUTLINE OF THE THESIS

This thesis investigates analytical dynamic system optimal assignment with departure time choice in a rigorous and original way. The results achieved in this thesis can be applied to various areas of transport modelling and management including travel activity analysis, transport policy planning, road pricing and network design. This thesis contributes to the literature on dynamic traffic management and supports further development in this area.

The thesis is organized as follows:

In Chapter 2, this thesis starts with giving a comprehensive review on the link traffic flow and travel time models for use in dynamic traffic assignments. Proceeding after the comments made by Patriksson (1994), we summarize the requirements for a traffic or travel time model to be satisfactory for use in dynamic traffic modelling. A review on various traffic models is given and discussed.

In Chapter 3, we investigate the analysis and the solution algorithms for dynamic user equilibrium assignment with departure time choice. Several properties related to the assignments are established. Numerical examples and the characteristics of the assignment results associated with different choices of travel time models and discretizations are discussed.

In Chapter 4, we analyse dynamic system optimal assignment by exploiting a state-dependent optimal control formulation (see for example in Friesz et al., 2001). In the formulation, a fixed volume of traffic is assigned to departure times and routes such that the total system travel cost is minimized. To facilitate the analysis and calculation, we develop a novel sensitivity analysis of travel cost. Solution algorithms are developed to implement this sensitivity analysis and solve dynamic system optimal assignment. Numerical examples are given and the characteristics of the results are discussed. In particular, the results reveal that much study in the literature of dynamic system optimal assignment based on the deterministic queuing model is not generally applicable. In the end of Chapter 4, we also propose some practical tolling strategies for managing dynamic network traffic. These tolling strategies are compared with the dynamic system optimal toll and hence their efficiencies can be evaluated accordingly. Finally, we have an investigation on the robustness of the toll calculation methods which is an important issue to address in practice.

Chapter 5 gives a conclusion of the whole thesis and identifies some possible future extensions in the area.

2. LINK FLOWS AND TRAVEL TIMES

2.1 INTRODUCTION

Temporal variations of link traffic flows and link travel times in dynamic traffic assignment models are represented by *traffic models*. Many different kinds of traffic models have been proposed in the literature (see for example, Vickrey, 1969; Merchant and Nemhauser, 1978a; b; Hendrickson and Kocur, 1981; Mahmassani and Herman, 1984; Newell, 1988; Friesz et al., 1993; Daganzo, 1994, 1995a; Chu, 1995; Ran and Boyce, 1996; Yang and Huang, 1997; Carey et al., 2003). Some of these traffic models are more tractable or convenient to use over the others, while some of the models are more realistic representation of traffic dynamics. Because different traffic models produce different estimations for link flows, travel times, and hence solutions of traffic assignments, it is important to understand the properties, plausibility, and applicability of each traffic model. It is also vital to identify the minimum requirements on a traffic model for it to be used in dynamic traffic assignment formulations.

In general, these traffic models can be summarized in the following general form

$$\tilde{c}_a(s) = \rho_a\left[e_a(s), x_a(s), g_a(s)\right], \tag{2-1}$$

where $\tilde{c}_a(s)$ is the link travel time experienced by traffic enters the link *a* at a time *s*. The rate at which traffic enters and leaves the link at time *s* are denoted by $e_a(s)$ and $g_a(s)$ respectively. The amount of traffic present on each link *a* at time *s* is represented by $x_a(s)$. The link travel time is related with the traffic flow quantities through the traffic model $\rho_a(\cdot)$. Daganzo (1995b) showed that for a traffic model which is dependent of inflow, e_a , a sufficiently fast decline in the link inflows can make the traffic model violate first-in-first-out (FIFO) queue discipline. Likewise, Daganzo (1995b) further showed that the traffic model should also be independent of outflow, g_a , because a sufficiently fast decline in the link outflows can also make the traffic model violate FIFO queue discipline in a similar way. Violation of FIFO queue discipline is considered to be unrealistic in a macroscopic travel time model that considers traffic to be flowing continuously, because it implies that the later and faster vehicles will *jump over* the preceding slower vehicles (Carey, 2004a). Following these observations, Daganzo (1995b) suggested that traffic models should only be a function of amount of link traffic, i.e.

$$\tilde{c}_a(s) = \kappa_a[x_a(s)]. \tag{2-2}$$

Proceeding after Daganzo (1995b), the properties of various kinds of traffic models and their suitability for modelling dynamics of traffic have been investigated widely (see for example, Astarita, 1996; Heydecker and Addison, 1998; 2005; Wu et al., 1998; Xu et al., 1999; Rubio-Ardanaz, 2003; Nie and Zhang, 2005a; b; Carey, 2004a; b).

This chapter gives a comprehensive review and a detailed discussion on the research of traffic models. The chapter is organized as follows. In Section 2.2, the requirements on a link traffic model for use in dynamic traffic assignment formulations are summarized and discussed. Section 2.3 introduces and analyses different kinds of traffic models. For implementation, Section 2.4 describes numerical schemes which transform the continuous time formulation of traffic models into discrete time. Numerical examples are given in Section 2.5 to demonstrate the characteristics of different traffic models. Finally, some concluding remarks are given in Section 2.6.

2.2. REQUIREMENTS ON TRAFFIC MODELS

Following Carey (2004a; b), and Heydecker and Addison (2005), for plausible estimation of traffic flows and travel times, the link traffic model adopted should possess and satisfy the following five properties:

1. non-negativity;
2. first-in-first-out (FIFO) discipline;
3. conservation of flow;
4. consistency between travel time and flow;
5. causality.

In this section, these properties are discussed in detail as follows.

2.2.1 Non-negativity

The non-negativity principle states that if a positive inflow is loaded into a travel link, then each of the resulting traffic, outflow, and the travel times should always also be positive. This condition can be stated as

$$e_a(s) > 0 \quad \Rightarrow \quad g_a[\tau_a(s)] > 0,\; x_a(s) > 0, [\tau_a(s) - s] > 0, \tag{2-3}$$

where $\tau_a(s)$ is the time of exit for a time of entry at time *s*, and hence, $\tau_a(s) - s$ is the corresponding travel time along the link.

2.2.2 First-in-first-out (FIFO) queue discipline

The FIFO queue discipline requires that if a traveller defers his departure time from the origin and join the traffic queue later, then he can expect to arrive at the destination later. That is, the FIFO discipline is satisfied if $s_2 \ge s_1$, $\tau(s_2) \ge \tau(s_1)$ for all times of entry s_1 and s_2. Proposition 2.1 then follows for differentiable functions $\tau(\cdot)$.

Proposition 2.1: If the traffic model satisfies the FIFO queue discipline and the function $\tau(\cdot)$ is differentiable, then the following condition will be satisfied

$$\frac{d\tau}{ds} \ge 0, \tag{2-4}$$

for all times of entry *s* to the link.

Proof:
We first have the condition of link FIFO as $\tau(s_2) \ge \tau(s_1)$ for all s_1 and s_2, $s_2 \ge s_1$. This implies that for $\Delta s > 0$, $\dfrac{\tau(s_2) - \tau(s_1)}{s_2 - s_1} = \dfrac{\Delta\tau(s)}{\Delta s} \ge 0$ because both numerator and denominator are positive. Taking the limit on $\Delta s \to 0$ gives $\dfrac{d\tau}{ds} \ge 0$. ☐

The FIFO queue discipline is an essential property for modelling dynamic traffic. Indeed, Daganzo (1995) and Astarita (1996) have shown that unless the link traffic model respects the FIFO discipline, problems will arise in respect of one or both of non-negativity of traffic and proper propagation of flows. This is further supported by Carey (2004a), who showed that the FIFO discipline is a necessary and sufficient condition to ensure non-negativity of traffic and consistency between traffic flows and corresponding travel times (see proposition 3 in Carey, 2004a). The FIFO condition could be considered to be too strong and unrealistic, but satisfaction of the FIFO discipline is necessary in macroscopic and continuous traffic models. Carey (2004a) explained that FIFO discipline only means to prevent overtaking and passing due to incidental features within the traffic model that do not reflect any real world phenomenon such as a fast vehicle jumps over the preceding slower one.

2.2.3. Conservation of flow

The conservation of flow states that the traffic volume $x_a(s)$, which is the number of vehicles or the occupancy, on a travel link at any time should be equal to the difference between the cumulative inflow and outflow by that time. The underlying assumption of the principle of conservation is that traffic will neither be generated nor dissipated, for example by vehicles entering from and exiting into side links, within the travel link. However, this assumption could in principle be relaxed by introducing origin or destination nodes to the link as noted by Carey (2004a). This conservation of flow can be written as

$$x_a(s) = E_a(s) - G_a(s), \tag{2-5}$$

where $E_a(s)$ and $G_a(s)$ respectively represent the cumulative inflow and outflow by time s. The relationship between the variables in Equation (2-5) is also shown in Figure 2.1.

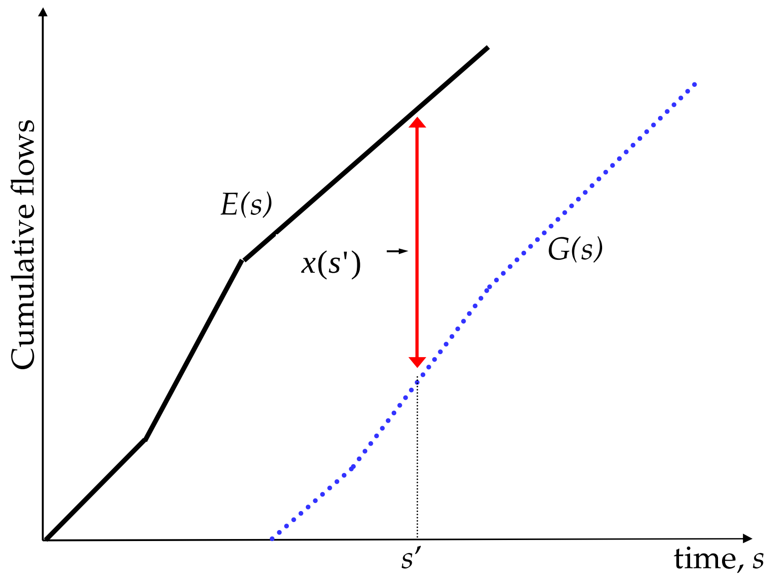


Figure 2.1 Flow conservation

If the variables in Equation (2-5) are differentiable with respect to time *s*, then from differentiating (2-5) we have

$$\frac{dx_a(s)}{ds} = e_a(s) - g_a(s)\,. \tag{2-6}$$

Equation (2-6) states that the rate of change of $x_a(s)$ at any time *s* can be determined as the difference between the inflow and the outflow of at that time.

2.2.4. Consistency between travel time and flow

This travel time-flow consistency is also known as proper propagation of flow (Tobin, 1993; Friesz and Bernstein, 2000; Mun, 2002; Heydecker and Addison, 2005). It states that the cumulative traffic that has entered up to time s must have exited from the link by exactly time $\tau_a(s)$ (see Figure 2.2). This can be expressed as

$$E_a(s) = G_a[\tau_a(s)], \tag{2-7}$$

where $E_a(s)$ and $G_a[\tau_a(s)]$ correspond to the cumulative inflow by s and the cumulative outflow by $\tau_a(s)$ respectively.

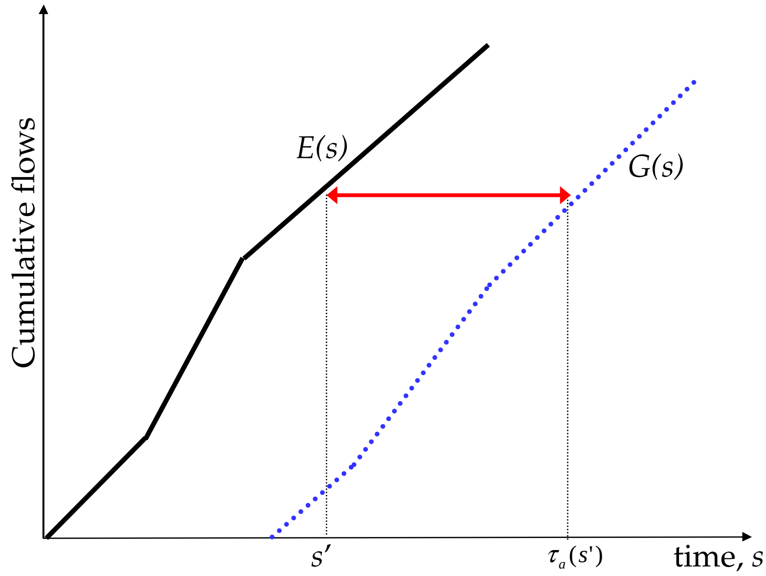


Figure 2.2 Consistency of travel-time and traffic-flow

If the variables in Equation (2-7) are differentiable with respect to s, the we can apply the chain rule and differentiate both sides with respect to time s, and hence Equation (2-7) can be written equivalently as

$$e_a(s) = g_a[\tau_a(s)]\frac{d\tau_a(s)}{ds}. \tag{2-8}$$

Equation (2-8) shows the variation of the flow along the travel link should be based on the rate of change of the link travel time, i.e. $\frac{d\tau_a(s)}{ds}$. Following Equation (2-8), a proposition on the non-negativity of link outflow profile is also deduced.

> ***Proposition 2.2:*** If the traffic model satisfies FIFO queue discipline, and given a positive profile inflow for all time *s*, then the corresponding profile of outflow is also positive.

Proof:

Proposition 2.1 shows that FIFO queue discipline implies positive rate of change of link travel time $\frac{d\tau_a(s)}{ds}$ for all time *s*. Proceeding after this and using Equation (2-8), given the link inflow profile $e_a(s)$ and $\frac{d\tau_a(s)}{ds}$ are positive for all time *s*, then the corresponding link outflow profile $g_a[\tau_a(s)]$ must be also positive. □

2.2.5. Causality

Behaviour of traffic should be affected only by local or conditions downstream, not by traffic conditions upstream. This causal relationship also implies that the outflow profile from a travel link should only depend on the inflow profile at or before the corresponding time of entry but not after.

2.3 LINK TRAFFIC MODELS

This section classifies all link traffic models into three different categories: wave models (for example, Lighthill and Whitham, 1955; Richards, 1956; Payne, 1971; Newell, 1988; Heydecker and Addison, 1996), outflow traffic models (for example, Merchant and Nemhauser, 1978a; b; Ho, 1980; Carey, 1987; Friesz et al., 1989; Yang and Huang, 1997), and travel time models (for example, Vickrey, 1969; Friesz et al., 1993; Mun, 2002; Carey et al. 2003).

2.3.1 Wave model

This class of models was originated by Lighthill and Whitham (1955) and Richards (1956), whose model is now known as the kinematic wave model or simply the LWR model. The LWR model considers traffic stream to be one-dimensional compressible fluid and the model can be stated by the following two conditions:

$$\frac{\partial f}{\partial x} + \frac{\partial k}{\partial t} = 0 \quad \text{and} \quad f = F(k, x, t) \tag{2-9}$$

where f is the traffic flow; k is the density; x and t are space and time variables, respectively, and F is a function relating the traffic flow f and the traffic density k over time and space. The function F is often referred to as the Fundamental Diagram of traffic as shown in Figure 2.3.

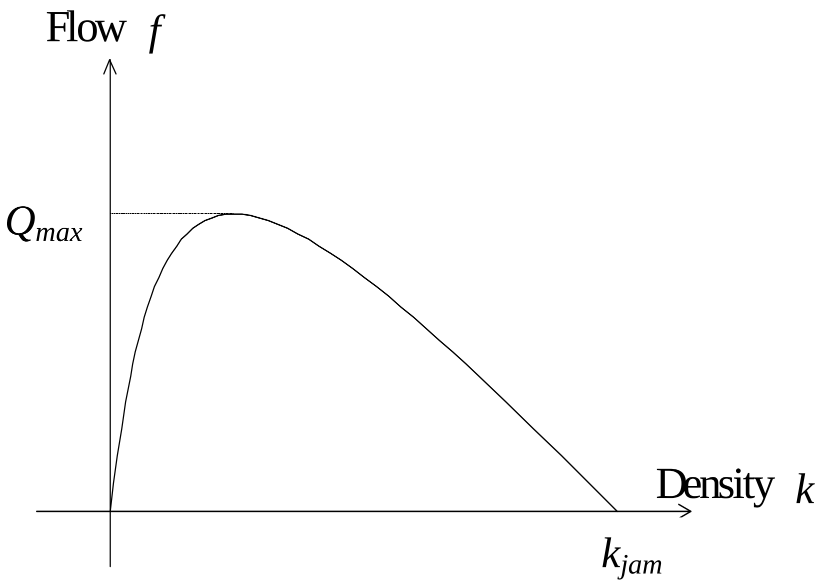


Figure 2.3 The Fundamental Diagram of traffic flow

Supported by the empirical evidence and its highly detailed description of traffic behaviour, the LWR model is arguably one of the most widely accepted models of traffic flow. It takes into account explicitly the macroscopic variables of flow and density and covers the full range of the fundamental density-flow-speed relationships. The traffic model captures the macroscopic features of traffic, including shockwaves, queue formation and queue dissipation, in both congested and uncongested regimes. The wave model has also been applied to dynamic traffic assignment problems (see for example, Newell, 1988; Heydecker and Addison, 1996). Lighthill and Whitham (1955) developed a solution approach based on the method of characteristics, yet they have the disadvantage of being analytically and computationally demanding.

2.3.2. Outflow traffic models

2.3.2.1 Merchant and Nemhauser's (1978) outflow traffic model

This kind of traffic model was first proposed by Merchant and Nemhauser (1978a; b) for solving dynamic system optimal traffic assignment problem. This traffic model is also known in the literature as exit link function (Astarita, 1996), exit-flow model (Friesz and Bernstein, 2000; Carey, 2004b), or simply M-N model (Peeta and Ziliaskopoulos, 2001; Nie and Zhang, 2005a). The model considers outflow from each link in the road network to be a non-decreasing function ψ_a of the traffic volume on the whole link at that time. Thus,

$$g_a(s) = \psi_a[x_a(s)]. \tag{2-10}$$

The link traffic volume $x_a(s)$ can be determined either in continuous time or in discrete time (Carey, 2004b). In continuous time, the traffic volume $x_a(s)$ on the whole link can be determined by conservation of flow (Equation 2-6).

Proceeding after Merchant and Nemhauser (1978a; b), this outflow traffic model has been adopted extensively (see for example, Carey, 1986; 1987; 1990; Friesz et al., 1989; Carey and Srinvasan, 1993; Wie et al., 1994; Wie et al., 1995; Lam and Huang, 1995; Yang and Huang, 1997; Wie and Tobin, 1998). The outflow traffic model has convenient mathematical properties for analysis and has generated some important insights on the properties of time-varying network flows. However, the outflow traffic models have been being criticized for their implausible behaviour since Patriksson (1994). Addison and Heydecker (1995) showed that the traffic models lead to unrealistic flow propagation in which zero travel time could be estimated for some travellers and infinitely long ones for the others. Carey (1992), Janson and Robles (1993), and Astarita (1996) also discovered that the FIFO queue discipline cannot be guaranteed in Merchant and Nemhauser's (1978a; b) outflow traffic model. Hurdle (1986), Astarita (1996), and Heydecker and Addison (1998) also showed that the outflow models structurally violate causality, which is also shown in the following proposition.

> ***Proposition 2.3:*** The outflow traffic model $g_a(s) = \psi_a[x_a(s)]$ structurally violates causality.

Proof (modified from Heydecker and Addison, 1998)*:*
Combining the conditions of flow conservation (2-5) and travel time-flow consistency (2-7) gives $x_a[\tau_a(s)] = E_a[\tau_a(s)] - E_a(s)$.

Following the functional form of outflow models (2-10), the instantaneous outflow $g_a[\tau_a(s)]$ depends on $x_a[\tau_a(s)]$ and hence on the inflow in time interval $(s, \tau_a(s)]$. That is, the outflow $g_a[\tau_a(s)]$ depends on the inflow *after* the departure time *s*, and this represents a violation of causality. □

This acausal behaviour is unrealistic and hence unacceptable for any dynamic model of traffic.

2.3.2.2 The cell transmission model

Exploiting Merchant and Nemhauser's (1978a; b) idea, Daganzo (1994, 1995) developed a Godunov solution scheme[†] called the cell transmission model (CTM) for solving the LWR model of traffic flow. The cell transmission model assumes its Fundamental Diagram to take a trapezoidal form as shown in Figure 2.4. This relationship assumes a constant free-flow speed, v, for lower densities and a constant negative wave speed, w, (always lower than free-flow speed) at higher densities. This simplification was supported with empirical data in Cassidy (1998).

[†] Godunov solution scheme is a finite difference solution scheme for solving partial differential equations. Daganzo adopted this to solve for the LWR model (Equation 2-9) using Merchant and Nemhauser's (1978) formulation of a traffic model.

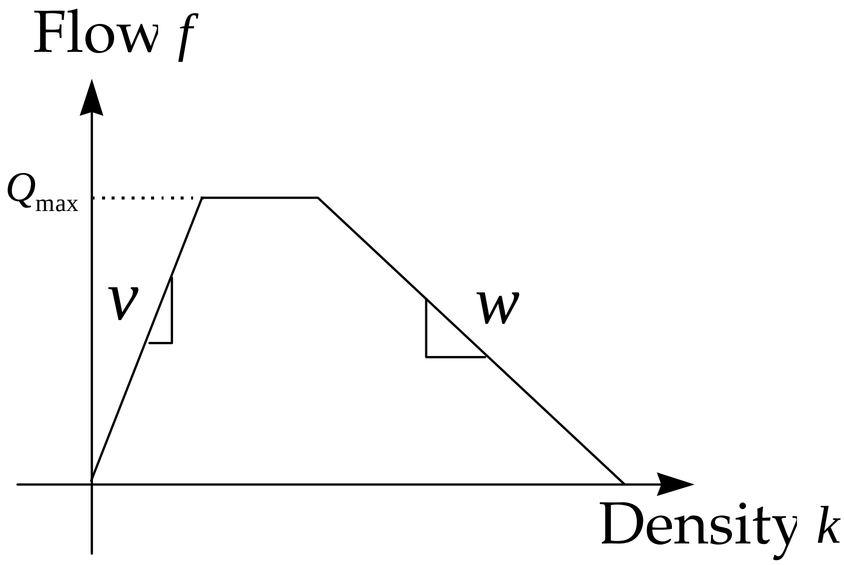


Figure 2.4 The flow-density relationship used in CTM

In the cell transmission model, the road network is represented by a collection of equal-length cells. The length of each cell is equal to the distance that a single vehicle travels in one time step at the free-flow speed. When there is no congestion, it is expected that a vehicle would move from one cell to another at each time step. For a given time interval k, each cell i has a number of vehicles in it, $x_i(k)$, and vehicles ready to enter it, $g_i(k)$. The outflow from each cell i (or the inflow into its downstream cell $i+1$) during the time interval $[k\Delta s, (k+1)\Delta s)$ is governed by the following equation,

$$g_{i+1}(k) = \min\left\{ x_i(k), Q_{i+1}, \frac{w}{v}\left[N_{i+1} - x_{i+1}(k)\right] \right\}, \qquad \text{(2-11)}$$

where Q_{i+1} is the maximum number of vehicles that can enter cell $i+1$ in a single time step; N_i is the spatial capacity of cell i; $\left[N_{i+1} - x_{i+1}(k)\right]$ is the available space in cell $i+1$; and w/v is the ratio of shockwave speed to free-flow speed. This formulation automatically covers both the congested and uncongested regions through the fundamental diagram.

After these flows have been determined for each cell for a specified time step, the traffic conditions in the network at the next time interval, $k+1$, is updated with the following conservation equation:

$$x_i(k+1) = x_i(k) - g_i(k) + g_{i+1}(k) \tag{2-12}$$

Although the cell transmission model is developed as a form of outflow model, causality is preserved in this model. It is because this model considers outflow at one time interval forward $[k\Delta s, (k+1)\Delta s)$, rather than at the current instant $k\Delta s$. The cell transmission model has also been applied to dynamic traffic assignment problems (see for example, Lo, 1999; Ziliaskopoulos, 2000; Lo and Szeto, 2002; 2004). These studies revealed that solving the cell transmission model is computationally expensive. Friesz and Bernstein (2000) also pointed out that the cell transmission model is difficult to analyse because the outflow function (2-12) is piecewise and hence is not differentiable with respect to its state variable $x_i(k)$.

2.3.3. Travel time models

Travel time models consider travel time along each link to be a non-decreasing function of the traffic volume on the link. A key difference between the travel time model and the outflow traffic model is that the outflow traffic model first determines the link outflow profile according to the given outflow function and the current traffic conditions, then back calculates the corresponding link travel time. In contrast, the travel time model first determines the link travel time according to the given travel time function and the current traffic conditions, and then calculates the corresponding link outflow profile.

In general, a travel time function, $\kappa_a(x_a)$, operates in a way that

$$\tau_a(s) = s + \kappa_a(x_a), \tag{2-13}$$

which calculates the corresponding time of exit $\tau_a(s)$ from the travel link for traffic entering the link at time *s*. The travel time functions considered in this thesis has the following properties:

1. $\kappa_a(0) = \phi_a$ when $x_a = 0$, where ϕ_a represents the free flow travel time of the link when the link is empty;

2. $\kappa'_a(x_a) > 0$ for $x_a > 0$, where $\kappa'_a(x_a)$ is the first-order derivative with respect to the state variable x_a;

3. $\kappa'_a(x_a) \to \frac{1}{Q_a}$ when $x_a \to \infty$, where Q_a represents the capacity of the link.

Considering whether the travel time model satisfies FIFO queue discipline, we have the following proposition:

Proposition 2.4: The travel time model, $\kappa_a(x_a)$, satisfies FIFO queue discipline provided that, for all time *s*, the inflow profile satisfies

$$e_a(s) \geq g_a(s) - \frac{1}{\kappa'_a[x_a(s)]}. \tag{2-14}$$

Proof:

Proceeding after proposition 2.1, for the model $\kappa_a[x_a(s)]$ to satisfy FIFO, it requires

$$\begin{aligned} &\frac{d\tau_a(s)}{ds} = 1 + \kappa'_a[x_a(s)]\frac{dx_a(s)}{ds} \geq 0 \\ &\Rightarrow \kappa'_a[x_a(s)][e_a(s) - g_a(s)] \geq -1 \\ &\Rightarrow e_a(s) \geq g_a(s) - \frac{1}{\kappa'_a[x_a(s)]} \end{aligned}.$$

This completes the proof. □

Nie and Zhang (2005b) showed that eliminating the dependence of this condition for FIFO on inflow, then condition (2-14) becomes

$$g_a(s) \leq \frac{1}{\kappa'_a[x_a(s)]}. \tag{2-15}$$

Hence, a travel time model is guaranteed to satisfy FIFO if it satisfies condition (2-15) no matter the inflow profile is. Zhu and Marcotte (2000) conjectured a more convenient criterion to check if FIFO condition is satisfied:

$$\max e_a(s) \leq \frac{1}{\kappa'_a\left[x_a(s)\right]}. \tag{2-16}$$

However, Nie and Zhang (2005b) presented a counter-example in which a piecewise linear travel time model is adopted to disapprove (2-16) and suggested that condition (2-15) should be the correct criterion to use.

If we consider the travel time function $\kappa_a(x_a)$ to be linear with which the function of time of exit, and hence $\tau_a(s) = s + \kappa_a\left[x_a(s)\right] = s + \phi_a + \dfrac{x_a(s)}{Q_a}$ is a linear function of time of entry *s*, then $\dfrac{1}{\kappa'_a(x_a)} = Q_a$ will always be greater than $g_a(s)$ following property 3 above and hence FIFO condition is satisfied for all *s*. Taking this into account, the thesis restricts the attention to travel time models in linear form.

2.3.3.1. Deterministic queuing model

The first linear travel time model that we consider is the deterministic queuing model, which is also known as the bottleneck model (Vickrey, 1969; Kuwahara, 1990; Arnott, de Palma and Lindsey, 1990; 1998). This travel time model considers each link to be freely flowing with a flow-invariant travel time ϕ_a, with a deterministic queue at its downstream end being discharged with a maximum service rate Q_a. In this model, when a traffic queue exists, the link outflow is equal to the capacity and all travellers arriving before the queue dissipates will incur travel delay. Otherwise, if the queue length is zero, the outflow is taken as the inflow at the time of entry and the travellers are unimpeded. Thus,

$$g_a(s) = \begin{cases} e_a(s-\phi_a) & \left(x_a(s)=0,\ e_a(s-\phi_a) < Q_a\right) \\ Q_a & \text{otherwise} \end{cases}. \tag{2-17}$$

The traffic volume in queue, $x_a(s)$, is determined by the following state equation,

$$\frac{dx_a(s)}{ds} = \begin{cases} 0 & \left(x_a(s)=0,\ e_a(s-\phi_a)<Q_a\right) \\ e_a(s-\phi_a)-Q_a & \text{otherwise} \end{cases}, \tag{2-18}$$

which is derived from the conservation of flow. With this deterministic queuing model, the time derivative of the state variable is not continuous with respect to time s and inflow e_a. In particular, there is a corner[‡] on the inflow at $e_a = Q_a$ when $x_a = 0$.

Finally, the time of exit from the link for a time of entry s is calculated as

$$\tau_a(s) = s + \phi_a + \frac{x_a(s+\phi_a)}{Q_a}. \tag{2-19}$$

The deterministic queuing model is the most popular travel time model due to its incisiveness. This travel time model has also been shown to satisfy the requirements summarized in Section 2.2 (see Mun, 2002; Huang and Lam, 2002). In fact, the model has also shown its value in analysing dynamic network traffic and various control policies (for example, Smith and Ghali, 1990 a; b; Ghali and Smith, 1993; Arnott, de Palma and Lindsey, 1998; Akamatsu and Kuwahara, 1999; Han, 2000; Akamatsu, 2003; Akamatsu and Heydecker, 2003; Polak and Heydecker, 2006). However, the deterministic queuing model has been criticized for over-simplifying real traffic behaviour (Arnott et al., 1998). For example, Kimber and Hollis (1979) pointed out that the deterministic queuing model does not give any delay until the link has been over-saturated. This implies that the model fails to estimate any variation in travel time when the road link is in use within its capacity. Chu (1995) commented the fact that the deterministic queuing model cannot capture the change in the period of assignment before and after implementation of a transport policy, which is road pricing in Chu's (1995) example, is unrealistic. In addition, the non-differentiability in the state equation also causes analytical and computational difficulties. Some problems that arise from this non-differentiability are discussed in detail in Chapter 4.

[‡] A *corner* refers to a point of function at which the derivative of the function is discontinuous (Kamien and Schwartz, 1991,p86).

2.3.3.2 Whole link linear traffic model

Friesz et al. (1993) proposed another traffic model that can be used in place of the deterministic queuing model. The model considers the link travel time to be a linear function of the traffic volume on the link. As a result, the time of exit from the link for a time of entry *s* can be calculated as

$$\tau_a(s) = s + \phi_a + \frac{x_a(s)}{Q_a}, \tag{2-20}$$

in which the whole link traffic volume $x_a(s)$ can be determined by the flow conservation condition using Equation (2-6).

Furthermore, the outflow experienced by traffic that enters at time *s* can be established according to correct propagation of flow (Heydecker and Addison, 1998) as

$$g_a[\tau_a(s)] = e_a(s) \Big/ \frac{d\tau_a(s)}{ds} = \frac{Q_a e_a(s)}{Q_a + e_a(s) - g_a(s)}, \tag{2-21}$$

which depends on outflows at time *s* and hence on inflows at earlier times. Incorporating this flow propagation relationship, the state equation for $x_a(s)$ can be re-written as

$$\frac{dx_a(s)}{ds} = e_a(s) - \frac{Q_a e_a[\sigma_a(s)]}{Q_a + e_a[\sigma_a(s)] - g_a[\sigma_a(s)]}, \tag{2-22}$$

where $\sigma_a(\cdot)$ is the inverse function of $\tau_a(\cdot)$ so that $\sigma_a[\tau_a(s)] = s$.

The whole-link traffic model is further investigated by many others (for example, Astarita, 1996; Mun, 2002), and has been shown to satisfy all the requirements listed in Section 2.2. Contrasting with deterministic queuing model, the state equation of the whole-link traffic model is smooth and continuously differentiable with respect to time *s* and inflow e_a.

However, Nie and Zhang (2005b) commented that this whole-link traffic model overstates the actual link travel time and hence underestimates the link outflow rates.

2.3.3.3. Divided linear travel time model

Regarding the properties of the deterministic queuing model and the whole-link traffic model, the divided linear travel time models (see for example, Ran and Boyce, 1996; Mun, 2002; Bliemer, 2006) can be regarded as a hybrid of them. The structure of this class of travel time models is shown in Figure 2.5.

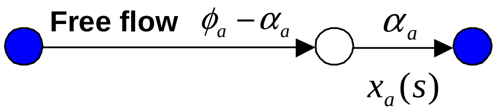


Figure 2.5 Representation of divided link travel time model

Each travel link is considered to be having a freely flowing part and a congestible part. The travel time along the free flow is taken as $\phi_a - \alpha_a$, and that along the congestible part is $\alpha_a + \frac{x_a[s + (\phi_a - \alpha_a)]}{Q_a}$, where α_a is a parameter representing the free flow travel time in the congestible part. Consequently, the time of exit from the link for a time of entry s can be calculated as

$$\begin{aligned} \tau_a(s) &= s + (\phi_a - \alpha_a) + \left\{ \alpha_a + \frac{x_a[s + (\phi_a - \alpha_a)]}{Q_a} \right\} \\ &= s + \phi_a + \frac{x_a[s + (\phi_a - \alpha_a)]}{Q_a} \end{aligned} . \quad (2\text{-}23)$$

This class of travel time models was shown to satisfy all the requirements in Section 2.2 (Mun, 2002). It is noted that the divided travel time model (2-23) includes the deterministic queuing model and the whole-link traffic model as its two extreme cases: the model becomes a deterministic queuing model (2-19) when the parameter α_a is taken as zero; it becomes a whole-link traffic model (2-20) when α_a is equal to the free flow travel time ϕ_a. Mun (2002) adopted a divided linear travel time model with a short congestible part in which $\alpha_a = \Delta s$ for smoothest network loading and assignment results after performing a series of numerical experiments.

2.4 DISCRETIZATION OF LINK TRAVEL TIME MODELS

The analysis in Sections 2.2 and 2.3 is considered in continuous time which enables the exploitation of calculus. To obtain numerical solutions, travel time models have to be transformed into discrete time representation. The continuous time flow quantities, $e_a(s)$ and $g_a(s)$, will be expressed in terms of the amount of traffic, $e_a(k)$ and $g_a(k)$, in a time interval, *k*, where $k = [k\Delta s, (k+1)\Delta s)$, in which Δs is the size of the time interval adopted in discretization (see Figure 2.6). The traffic volume $x_a(k)$ in discrete time interval k will be considered at the *end* of the time interval (i.e. at the instant $(k+1)\Delta s$).

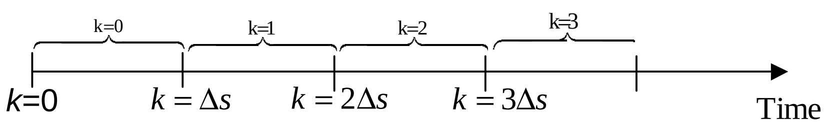


Figure 2.6 Time discretization

Broadly speaking, discretization is a process that transforms the continuous time quantities: $e_a(s)$, $g_a(s)$, $x_a(s)$, and $\tau_a(s)$ into corresponding discrete time ones: $e_a(k)$, $g_a(k)$, $x_a(k)$, and $\tau_a(k)$. In this thesis, a linear (i.e. first order) interpolation technique is adopted to approximate and interpolate the continuous time values in discrete time. Nevertheless, there are still two different approaches to discretize a travel time model which are described in the following sections.

2.4.1 Discretization based on flows

The method of discretization based on flows was first documented in detail in Astarita (1996) and is described as follows:

Step 0 Initialisation

0.1 Set $s := 0$, and $k := \phi_a / \Delta s + 1/2$;

0.2 Set $x_a(0) := 0$ and hence $\tau_a(0) := \phi_a$;

0.3 Set $g_a(k) := 0$ for all $k = 0,1,2,...,\phi_a / \Delta s$.

Step 1 Incremental loading

1.1 Set $s := s + 1$;

1.2 Calculate $x_a(s) := x_a(s-1) + e_a(s-1) - g_a(s-1)$,

and hence $\tau_a(s) := (s-1)\Delta s + \phi_a + \dfrac{x_a(s)}{Q_a}$.

Step 2 Calculating the instantaneous outflow

Calculate $g_a[\tau_a(s)] := \dfrac{e_a(s-1)\Delta s}{[\tau_a(s) - \tau_a(s-1)]}$.

Step 3 Discretizing the outflow

While $\tau_a(s) > (k - 1/2)\Delta s$, continuously interpolate $g_a(k)$ with $g_a[\tau_a(s)]$ and $g_a[\tau_a(s-1)]$ as follows:

3.1. Set $g_a(k) := g_a[\tau_a(s-1)] + \dfrac{g_a[\tau_a(s)] - g_a[\tau_a(s-1)]}{[\tau_a(s) - \tau_a(s-1)]}[(k - 1/2)\Delta s - \tau_a(s-1)]$;

3.2. Set $k := k + 1$.

2.4.2 Discretization based on cumulative flows

Ge and Carey (2002), and Nie and Zhang (2005) later reported separately that discretization scheme using cumulative flows is more efficient in coding and computing and can avoid numerical difficulties such as division by zero. Ge and Carey's (2002) and Nie and Zhang's (2005) discretizing procedure is described as follows:

Step 0: Initialisation

0.1. Set $s := 0$, and $k := \phi_a / \Delta s$;

0.2. Set $x_a(0) := 0$, and hence $\tau_a(0) := \phi_a$;

0.3. Set $g_a(k) := 0$, for all $k = 0,1,2,\ldots,\phi_a / \Delta s$;

0.4. Set $R_e = 0$.

Step 1 Incremental loading

1.1. Set $s := s+1$;

1.2. Calculate $x_a(s) := x_a(s-1) + e_a(s-1) - g_a(s-1)$,

and hence $\tau_a(s) := (s-1)\Delta s + \phi_a + \dfrac{x_a(s)}{Q_a}$.

Step 2 Discretizing the profile of outflow

2.1 Calculate $nk := \dfrac{\tau_a(s)}{\Delta s} - k$;

2.2 If $nk < 1$ then $R_e := R_e + e_a(s-1)$ and go to step 3 directly; else go to step 2.3;

2.3 Set $k := k+1$, calculate $g_a(k) := R_e + \dfrac{[k\Delta s - \tau_a(s-1)]}{[\tau_a(s) - \tau_a(s-1)]} e_a(s-1)$;

2.4 Distribute inflow:

a. Set $j := 2$;

b. Set $k := k+1$;

c. Calculate $g_a(k) := \dfrac{\Delta s}{[\tau_a(s) - \tau_a(s-1)]} e_a(s-1)$;

d. Set $j := j+1$;

e. If $j = nk$ then go the step 2.5; else go to 2.4b;

2.5 $R_e = \dfrac{[\tau_a(s) - k\Delta s]}{[\tau_a(s) - \tau_a(s-1)]} e_a(s-1)$.

Step 3 Stopping criterion

If $s < \left\lfloor \dfrac{T}{\Delta s} \right\rfloor$, go to step 1; otherwise stop.

Discussion

Nie and Zhang (2005) proved that the algorithm can always proceed as long as a travel time model satisfying FIFO is used and the minimum link travel time is greater than the size of discretized time interval, Δs . In addition, Nie and Zhang (2005) also proved that the algorithm converges to the solution of the continuous model as Δs approaches to zero, provided the rate of change of link travel time is bounded above.

2.5 EXAMPLE CALCULATIONS

This section presents some example calculations to show the characteristics of the numerical results of the travel time models presented in Section 2.3.3. We consider a single travel which has a free flow travel time ϕ_a equals to 3 mins and a capacity Q_a equals to 20 veh/min. The size of the discretized time intervals Δs is taken as 1 min. The travel link is initially empty. A parabolic profile of inflow as specified in (2-24) is then loaded into the travel link. This profile has a peak inflow rate of 50 vehs/min, which equals to 2.5 times of the link capacity. Consequently, the travel link will be overloaded for period of time.

$$e_a(s) = \begin{cases} \frac{1}{8}(40-s)s & \text{if } 0 \le s \le 40 \text{ (minutes)} \\ 0 & \text{otherwise} \end{cases} . \tag{2-24}$$

The resulting link outflow and the link travel time on the travel link is estimated by using four different linear travel time models: deterministic queuing model (2-19), whole-link traffic model (2-20), divided linear model (2-23) with $\alpha_a = \Delta s$ (i.e. Mun's (2002) model), and divided linear model (2-23) with $\alpha_a = 2\Delta s$. This example calculation adopts the cumulative flows based algorithm in Section 2.4.2 for discretization.

Given the inflow profile (2-24), Figure 2.7 depicts the resulting link outflow profiles estimated by the travel time models. All travel time models show that the outflow will approach to or equal to, but not exceed, the link capacity for a high inflow rate. With the deterministic queuing model, the outflow equals to either the corresponding inflow when the link is uncongested or the link capacity when the link is congested. Comparing with the deterministic queuing model, the whole-link model and the divided models show a more realistic pattern of queue dispersion in which outflows vary continuously with the inflow over time. It is also observed that as the congestible portion in the travel model increases, the values of the outflow rates approach the link capacity at a faster rate.

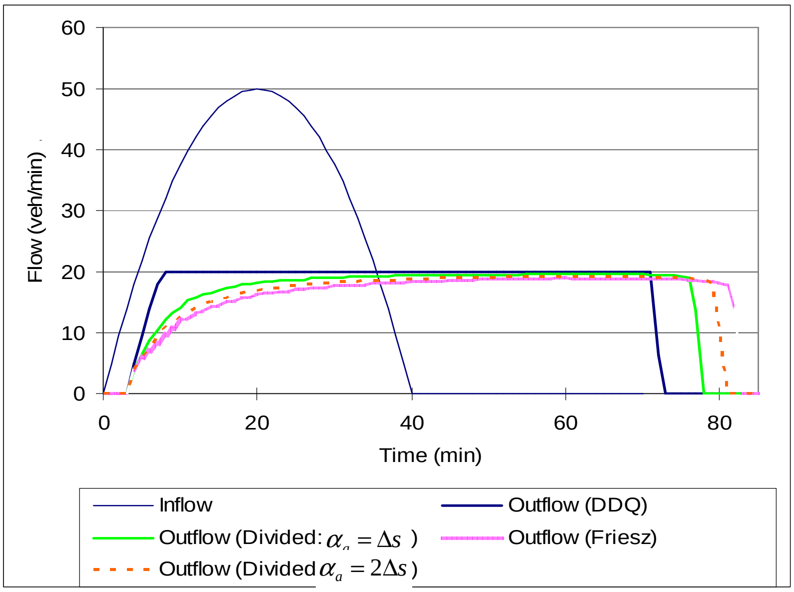


Figure 2.7 Link outflow profiles

Figure 2.8 shows the link travel times estimated by the travel time models. As expected, as the portion of the congestible part taken in the travel time models increases, the corresponding travel time estimated increases. One noteworthy feature of this result, however, is that even we consider a small portion of the link to be congestible (i.e. $\alpha_a = \Delta s$), the travel time will be substantially higher. This difference suggests that choosing an appropriate traffic model to represent the road network system is important, in particular the deterministic queuing model is so predominantly used in the literature for analyzing dynamic network traffic (see for example, Vickrey, 1969; Arnott, de Palma and Lindsey, 1990, 1993, 1998; Akamatsu and Kuwahara, 1999, 2001; Akamatsu and Heydecker, 2003).

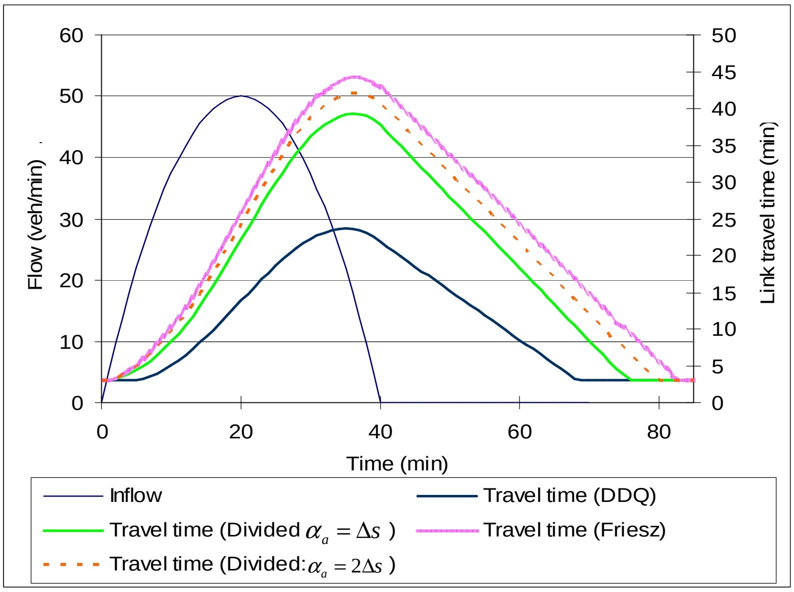


Figure 2.8 Link travel times

2.6. DISCUSSION

This chapter reviews various traffic models that have been used in dynamic traffic assignment formulations. We start with summarizing the requirements for a traffic model to be satisfactory for use in dynamic traffic modelling and assignments. These requirements include non-negativity of traffic, FIFO queuing discipline, conservation of traffic, travel time-flow consistency, and causality. The implications and the relationships between these desirable properties are discussed. In this chapter, the traffic models are classified into three distinct categories: wave models, outflow traffic models, and travel time models. Wave models are the most widely accepted traffic models. However, this class of traffic models is too complex and computationally demanding to use in dynamic traffic modelling and assignments. The outflow traffic models have been used to represent the link flows in dynamic traffic models and assignments. With these traffic models, we have generated some important insights on dynamic network traffic phenomenon and management. However, the outflow traffic models cannot guarantee plausible traffic propagation and causal relationship. Compared with the wave models and the outflow traffic models, the travel time models are more practical to use and they have been shown to be satisfactory with respect to the requirements in Section 2.2, provided the link travel time is a linear function of the associated link traffic volume.

The travel time models considered in this chapter include the deterministic queuing model, the whole link linear traffic model, and the divided linear travel time models. Their associated properties are also analysed and discussed. The deterministic queuing model is the most popular travel time model due to its incisiveness and has also shown its value in analysing dynamic network traffic and various control policies. However, the deterministic queuing model has been criticized for over-simplifying real traffic behaviour. In addition, the non-differentiability in the model also induces difficulties for use in dynamic traffic assignments.

Solution algorithms are presented for discretizing the travel time models. The characteristics of the numerical results are discussed. Given an inflow profile, the deterministic queuing model gives an outflow equal to either the corresponding inflow in uncongested case or the link capacity in congested case. Comparing with the deterministic queuing model, the whole-link model and the divided linear models show a more realistic pattern of queue dispersion in which outflows vary continuously with the inflow over time. In addition, the corresponding travel time estimated increases with the portion of the congestible part considered in the travel time models. One noteworthy feature of this result, however, is that even we consider a small portion of the link to be congestible (such as $\alpha_a = \Delta s$), the estimated travel time is still significantly higher than that of the deterministic queuing model. Chapter 3 and Chapter 4 further investigate the relationship between using different kinds of travel time models and the corresponding results of traffic assignments. The implications of choosing different travel time models for modelling and managing network traffic are also discussed in the following chapters.

3. DYNAMIC USER EQUILIBRIUM ASSIGNMENT WITH DEPARTURE TIME CHOICE

3.1 INTRODUCTION

In this thesis, the response of travellers to the traffic flows and travel times that they encounter can be considered in terms of their choices of departure times and travel routes, which can be represented by a dynamic traffic assignment model. Dynamic traffic assignment models provide important insight into the dynamics of urban network traffic and the sensitivity of travellers' behaviour in response to a range of transport policy measures. Dynamic traffic assignment modelling has been considered in the literature in the context of activity analysis (Zhang et al., 2005; Polak and Heydecker, 2006), transport planning (Yin and Lam, 2002; Heydecker, 2002a), and network management (Smith and Ghali, 1990 a; b; Yang and Meng, 1998; Heydecker, 2002b). The principles of dynamic traffic assignment essentially follow the extensions of Wardrop's (1952) two principles: dynamic user equilibrium and dynamic system optimum. This chapter first reviews and discusses *dynamic user equilibrium* assignment, while *dynamic system optimal* assignment is investigated in Chapter 4.

This chapter is organized as follows. Section 3.2 reviews different specifications of travel demand in dynamic user equilibrium assignment. Section 3.3 introduces different formulations of dynamic user equilibrium assignment with respect to the specifications of travel demand considered. The necessary conditions on traffic flows for each formulation of equilibrium assignment are presented. Such conditions and the associated mathematical analysis make explicit reference to the elements of the traffic model and the travel cost functions. The results of this analysis can hence be applied to any combination of possibilities for these component models, and have substantial generality in terms of the traffic model and the travel cost functions. Section 3.4 presents the requirements on the travel cost functions for dynamic user equilibrium to exist. Section 3.5 shows the analysis of the relationship between the total volume of traffic that is served by the system during a fixed period and the total travel costs associated with this. Section 3.6 describes the solution algorithm for solving the continuous time analysis of dynamic user equilibrium to discrete time solution. Section 3.7 demonstrates the example calculations and the numerical results. We also investigate the effects on the assignments of using different travel time models. Finally, some concluding remarks are given in Section 3.8.

3.2 SPECIFICATIONS OF TRAVEL DEMAND

In dynamic traffic assignment models, travel demand refers to the volume and the temporal profile of traffic that are assigned to each route through the network within a fixed time horizon. Travel demand can be specified according to the dimensions of choices of travellers that are considered. This section presents two different kinds of travel demand specifications for representing travellers' route choice and /or departure time choice.

3.2.1 Specification for modelling route choice

In the dynamic traffic assignment model when only route choice of travellers are considered (see for example, Lam and Huang, 1995; Heydecker and Addison, 1996; Han, 2000), the volume and the profile of travel demand between each origin-destination pair in the network is specified exogenously. Mathematically, this can be expressed as

$$\sum_{p \in P_{od}} e_p(s) = E_{od}(s) \qquad , \forall od, \forall s , \tag{3-1}$$

where $e_p(s)$ is the rate of flow into route p at time s, Pod is the set of all routes connecting origin o and destination d. The amount of travel at each instant s of departure is given by E_{od}(s) exogenously.

3.2.2 Specification for modelling route and departure time choice

The demand specification in Section 3.2.1 confines the dimension of travel choice to route choice only. Moreover, such specification of travel demand requires complete information on the temporal profile of travel demand for the whole network, which could present practical difficulties of data identification and collection (Heydecker and Addison, 2006). In fact, the specification of travel demand in (3-1) can be extended such that the temporal profile of the demand is determined endogenously in the system with a specified total volume within the study time horizon. The specification of travel demand in (3-1) can be reformulated as

$$\sum_{p \in P_{od}} \int_{s} e_p(s)ds = J_{od} \qquad , \forall od \, , \tag{3-2}$$

in which J_{od} represents the specified[§] total amount of travel within the study period.

To formulate the dynamic equilibrium assignment with the specification of travel demand in (3-2), some time-varying components of travel costs is required in additional to the cost of travel time to localise the travel in the time domain, and hence to determine the profile of inflow over time (Heydecker and Addison, 2005, 2006). The detail of those time-varying components of travel cost is discussed in Section 3.3.2.

3.3 DYNAMIC USER EQUILIBRIUM

The interaction between the travel demand, the traffic flows, and the travel times in the transport system can be represented by dynamic user equilibrium assignment. There are a number of ways to define and formulate dynamic user equilibrium, details of which were discussed in Section 1.5 and by Friesz et al. (1993); Boyce et al. (2001); Peeta and Ziliaskopoulos (2001). Following the travel demand specifications introduced in Section 3.2, dynamic user equilibrium assignment can be formulated with route choice, or with combined route and departure time choice.

3.3.1 Dynamic user equilibrium with route choice

Ran and Boyce (1996) gave a definition for dynamic user equilibrium assignment with route choice by extending Wardrop's (1952) user equilibrium principle,

[§] This thesis adopts a fixed demand formulation in which J_{od} is considered to be fixed with respect to travel cost, although J_{od} can also be considered to be a non-increasing function of the associated travel cost in an elastic demand formulation (see for example, Arnott et al., 1993; Yang and Huang, 1997; Wie and Tobin, 1998; Chow, 2007).

"under such equilibrium, the total travel cost is identical for all travellers departing at the same time, irrespective of the routes of travel they have chosen."

Heydecker and Addison (1993; 1996) expressed the dynamic user equilibrium assignment of route choice in a complementary inequality form of inflow as

$$e_p(s) \begin{cases} >0 & \Rightarrow \quad \tilde{C}_p(s) = \tilde{C}^*_{od}(s) \\ =0 & \Rightarrow \quad \tilde{C}_p(s) \geq \tilde{C}^*_{od}(s) \end{cases} \quad \forall p \in P_{od}, \forall od, \forall s, \tag{3-3}$$

where $\tilde{C}_p(s)$ is the cost associated with travel time along route p for traffic entering the route at time s, and $\tilde{C}^*_{od}$ is the minimum cost associated with travel time from origin o to destination d.

Heydecker and Addison (1993, 1996) further analysed and derived the necessary conditions for the dynamic user equilibrium assignment of route choice. They considered the rate of change of the cost of travel on routes that are in use at a particular time s, and then differentiated the first case of (3-3) with respect to time s and obtained

$$e_p(s) > 0 \quad \Rightarrow \quad \frac{d\tilde{C}_p(s)}{ds} = k_{od}(s) \quad \forall p \in P_{od}, \forall od, \forall s, \tag{3-4}$$

where $k_{od}(s) = \dfrac{d\tilde{C}^*_{od}(s)}{ds}$ is the common rate of change of costs for all routes in use between origin-destination pair od at time s. In case when the cost of travel is represented by the travel time alone, we have

$$\tilde{C}_p(s) = \Lambda[\tau_p(s) - s], \tag{3-5}$$

where Λ represents the value of time spent travelling. Using the condition of flow propagation (Equation 2-8) in Chapter 2 to eliminate $\dfrac{d\tau_p(s)}{ds}$ gives the necessary conditions for the dynamic user equilibrium assignment of route choice (Heydecker and Addison, 1996) as

$$e_p(s) = \frac{g_p\left[\tau_p(s)\right]}{\sum_{q \in P_{od}} g_q\left[\tau_q(s)\right]} J_{od}(s) \quad \forall p \in P_{od}, \forall od, \forall s, \tag{3-6}$$

which includes the case of zero inflow since the corresponding outflow will also be zero. Given a causally determinate traffic model, the value of the right-hand-side of the expression is determined by inflow before time *s*, and hence the equilibrium assignment at time *s* is determined by assignment at times before *s* but not after.

3.3.2 Dynamic equilibrium with route and departure time choice

As mentioned in Section 3.2, when both route and departure time choice of travellers are considered, the temporal profile of the demand can be determined endogenously in the system. Hendrickson and Kocur (1981) stated the definition for an assignment to be in dynamic equilibrium in such cases,

> "the total travel cost should be the same for all travellers between each origin-destination pair in the network, no matter what combinations of departure-time and route that the travellers have chosen."

The condition of this equilibrium can be expressed in the form of a complementary inequality in route inflows as (Heydecker and Addison, 2005):

$$e_p(s) \quad \begin{cases} > 0 & \Rightarrow \quad C_p(s) = C_{od}^* \\ = 0 & \Rightarrow \quad C_p(s) \geq C_{od}^* \end{cases} \quad \forall p \in P_{od}, \forall s \tag{3-7}$$

where $C_p(s)$ is the total cost associated with travel, C_{od}^* is the total travel cost at which travel takes place. All travel between each origin-destination pair is achieved at the same cost C_{od}^* throughout the study period.

To complete this extended formulation of equilibrium assignment, some time-varying components of travel cost are required to add to the cost of travel time $\tilde{C}_p(s)$ (see Equation 3-6) to localise the travel in the time domain. Throughout the study we suppose the value of time, Λ, is the same for all routes. For analytical convenience and without loss of generality, we further consider that all additional time-varying components of travel costs are expressed in terms of equivalent time spent travelling. As a result, $\Lambda = 1$.

The first component to be added is a time-specific cost, $h(s)$, associated with the time s of departure of the traveller from the origin. This cost explicitly considers the value of time to travellers at the origin of a journey. We consider that travellers would gain continuing benefit from remaining at their origin but are drawn to their destination by a need to attend there and hence to travel. Consequently, $h(s)$ is considered to be a monotonic non-increasing function of departure time s.

The second component to be added is a time-specific cost, $h(s)$, associated with the time of arrival, s, of the traveller at the destination, so that the arrival cost associated with departure from the origin at time s and using route p is $f[\tau_p(s)]$. Many authors have followed the specification of Vickrey (1969), Hendrickson and Kocur (1981), and Arnott, de Palma and Lindsey (1990) in which piecewise linear functions are adopted for $f(t)$ with constant value throughout an interval surrounding the ideal time of arrival and increasing with increasing deviation from it. Small (1982) introduced the idea of a discontinuous increase in cost at the latest permitted arrival time and reported the empirical finding that the rate of increase of cost for progressively late arrivals is about twice that for progressively early ones. The consideration of the arrival cost in the present formulation is substantially more general than that in the literature, while the choices of the time-specific costs are subject to certain restrictions that are discussed in Section 3.4.

Finally, the total travel cost $C_p(s)$ associated with departure on route p at time s is determined as the sum of all the above three costs:

$$C_p(s) = h(s) + [\tau_p(s) - s] + f[\tau_p(s)]. \tag{3-8}$$

Conditions (3-7) show that in equilibrium the cost $C_p(s)$ that is incurred by the travellers is constant with respect to time. Heydecker and Addison (2005) developed a novel analysis of the equilibrium conditions (3-7), and derived a relationship between route flows and the costs at the origin and destination of a journey that is satisfied by any flows in equilibrium. Consider a route that is in use for travel between a certain origin-destination pair at time *s*. Using the first case of (3-7) together with (3-8), differentiating with respect to departure time *s* gives

$$e_p(s) > 0 \quad \Rightarrow \quad h'(s) + \dot{\tau}_p(s) - 1 + f'[\tau_p(s)]\dot{\tau}_p(s) = 0\,. \tag{3-9}$$

Rearranging this gives the expression

$$e_p(s) > 0 \quad \Rightarrow \quad \dot{\tau}_p(s) = \frac{1 - h'(s)}{1 + f'[\tau_p(s)]}, \tag{3-10}$$

which specifies the rate of change of travel time on any route in use between an origin and destination to achieve dynamic user equilibrium (Heydecker and Addison, 2005). Pursuing the analysis of (3-10) by using the condition of flow propagation (Equation 2-8) to eliminate $\dot{\tau}_p(s)$ gives

$$e_p(s) = \left[\frac{1 - h'(s)}{1 + f'[\tau_p(s)]}\right] g_p[\tau_p(s)]. \tag{3-11}$$

In order to maintain the dynamic user equilibrium, the inflow to a route must satisfy the condition (3-11). The rate of change $h'(s)$ of departure time-specific cost is determined at the time of departure. As discussed earlier, this derivative is negative and values of greater magnitude will increase the initial inflows $e_p(s)$ in equilibrium. The arrival time-specific cost enters in the denominator of the right-hand side of (3-11). The effect of this is that departures will be more intense when they lead to arrival at those times, if any, where the arrival time-specific cost is decreasing and less intense when they lead to arrival at times when it is increasing. Finally, the route inflow is directly proportional to the outflow at the time of arrival. Due to causality, the outflow $g_p[\tau_p(s)]$ is determined by inflows before the time *s*.

3.4 REQUIREMENTS ON THE TIME-SPECIFIC COST FUNCTIONS

Following the analysis in Section 3.3, Heydecker and Addison (2005) summarized certain requirements on the travel cost functions for equilibrium to exist. In order for both inflows and outflows to be positive, it is required that each of the numerator and denominator of the quotient in (3-11) is positive[**]. This gives $h'(s) < 1$ and $f'(s) > -1$. Thus, the departure time-specific costs cannot increase at a rate that exceeds value of time spent travelling[††], otherwise travellers would have an incentive to depart earlier and spend additional time travelling rather than to remain at their origin. Furthermore, the condition $f'(t) > -1$ implies that the arrival cost cannot decrease at a rate that exceeds the value of time spent travelling. If it did, this would imply that travellers would have an incentive either to travel more slowly, or to use a route that is slower but otherwise equivalent as the reduced arrival costs on arrival would more than compensate for the increased travel time. Nevertheless, for reasonable travel behaviour, we usually expect $h'(s) < 0$ and $f'(s) > 0$ in practice, which dominates the requirements above.

Furthermore, if the arrival time-specific cost function is monotonic non-decreasing so that $f'(t) \geq 0$ for all times *t*, then in order for an equilibrium to exist, the departure time-specific cost function has to satisfy

$$h'\left(s_p^0\right) < -f'[s_p^0 + \Phi_p], \tag{3-12}$$

where

$$s_p^0 = \inf \left\{ s \mid h(s) + \Phi_p + f\left(s + \Phi_p\right) \leq C_{od}^* \right\} \tag{3-13}$$

is the time of the first entry to the route *p*. Thus, the departure time-specific cost function is required to decrease at the time of first assignment with greater magnitude than the increase in the arrival time-specific cost at the associated time of arrival. If this condition is not

[**] The case in which both numerator and denominator are negative is excluded as being unrealistic.
[††] Recall that the cost associated with a unit of travel time, Λ, is considered to be equal to 1 (see Section 3.3.2).

satisfied, the total travel cost, $C_p(s)$, will increase, at least initially, so that equilibrium cannot be achieved.

Moreover, suppose that the monotonic non-decreasing arrival time-specific cost function is piecewise continuously differentiable. At time s_p^1, where

$$s_p^1 = \sup \left\{ s \mid h(s) + \Phi_p + f\left(s + \Phi_p\right) \leq C_{od}^* \right\} \tag{3-14}$$

is the time of the last entry to the route *p*, the travel time will decrease because the traffic is being cleared This implies the instantaneous rate of change of travel time at that time is negative (i.e. $\left[\left. \frac{d\tau_p}{ds} \right|_{s_p^1} - 1 \right] < 0$) and gives the third term on the left-hand-side of (3-11) a positive value less than $f'[\tau_p(s_p^1)]$. In order for the equilibrium condition (3-11) to be satisfied, it is required that

$$-h'\left(s_p^1\right) < f'[s_p^1 + \Phi_p]. \tag{3-15}$$

Thus for travel to cease, the cost of remaining at the origin should not decrease at a greater rate than that at which the penalty for late arrival increases.

3.5 COST-THROUGHPUT RELATIONSHIP IN DYNAMIC USER EQUILIBRIUM

The following three quantities:

(a) the total cost C_{od}^* incurred by each traveller,

(b) the time s_p^0 at which each route p is first used, and

(c) total amount of travel E_p that takes place on each route during the study period, are closely inter-related (Heydecker and Addison, 2005). Once the time of first departure s_p^0 on route p is determined, the total cost of travel C_{od}^* can be found directly using (3-8) as $C_{od}^* = C_p\left(s_p^0\right)$. Conversely, if the cost C_{od}^* is specified for an origin-destination pair *od*, the times s_p^0 and s_p^1 of the first and last departures can be found for each route p according to (3-7) and (3-8). Once these times of first and last departure on a route are known, the inflow $e_a(s)$ can be integrated over the intervening interval to give the total amount of traffic E_p that is served by that route during the study period: $E_p = \int_{s=s_p^0}^{s_p^1} e_p(s)\, ds$. Because criterion (3-11) applies separately to each route that is used, the route-specific throughputs calculated according to this procedure can be summed for each origin-destination pair. Heydecker and Addison (2005) further showed how an implicit cost-throughput relationship can be established for each origin-destination pair in a network. In the case that the total volume of traffic is exogenous, the cost at which it will be achieved can be found by searching this implicit relationship. This then establishes the equivalence of the three quantities identified above, so that the specification of any one of them will determine the values of the others.

3.6 ALGORITHM FOR SOLVING DYNAMIC USER EQUILIBRIUM

The analysis in Sections 3.3 – 3.5 is considered in continuous time that facilitates the exploitation of calculus. This section introduces a solution algorithm that transforms the analysis in continuous time into numerical solutions in discrete time. The algorithm is structured as a forward dynamic programme to be solved forward in the order of departure time interval. It is due to the causal property of the travel time models that ensures the travel cost experienced by the traffic that departs from an origin at time s is independent of the traffic departing from that origin after time s. The study period in continuous time, T, is discretized into K intervals each of length Δs. Following this, the instantaneous flow in continuous time formulation is represented as the flow $e(k)$ that is constant through the discrete time interval k: $[k\Delta s, (k+1)\Delta s)$. This flow is tested against the cost $C[(k+1)\Delta s)$ at the late end of the time interval. Within each departure time interval k, the equilibrium inflow is calculated by using Newton method, which converges with an order of convergence at least 2 (Luenberger, 1989, p202).

The algorithmic procedure is described as follows.

Step 0: Initialisation

0.1 Choose an initial equilibrium cost C_{od}^{*}, for all O-D pairs *od*;

0.2 Set the overall iteration counter $n := 1$;

0.3 Set $e_p(k) := 0$ for all *p* between each O-D pair *od*, and for all time $k \in [0, K]$, where $K = T / \Delta s$;

0.4 Set time index $k := 0$;

0.5 Set the origin-destination index $od := 1$;

0.6 Set the route index $p := 1$;

0.7 Set the inner iteration counter $n^i := 1$.

Step 1: Network loading

Find $\tau_p(k+1)$ by loading the travel link using the route inflow $e_p(k)$ at the current iteration. The algorithm in Section 2.4.2 in Chapter 2 is adopted for this purpose.

Step 2: Update the inflow

2.1 Calculate

$$C_p(k+1) = h(k+1) + [\tau_p(k+1) - (k+1)] + f[\tau_p(k+1)];$$

2.2 Calculate $\Omega_p(k) = \dfrac{C_p(k+1) - C_p(k)}{\Delta s}$,

$$\text{and } \frac{\partial \Omega_p(k)}{\partial e_p(k)} = \left(1 + f'\left[\tau_p(k+1)\right]\right) \sum_{\forall a} \delta_p^a \frac{1}{Q_a},$$

$$\text{in which } \delta_p^a = \begin{cases} 1 & \text{if link } a \text{ is on route } p \\ 0 & \text{otherwise} \end{cases}$$ ‡‡.

---

‡‡ In Step 2.2, the derivative $f'\left[\tau_p(k)\right]$ can be estimated using a finite difference approximation as $f'\left[\tau_p(k)\right] \approx \dfrac{f\left[\tau_p(k+1)\right] - f\left[\tau_p(k)\right]}{\tau_p(k+1) - \tau_p(k)}$. The derivation of the expression of $\Omega_a^{'}(k)$ is given in Appendix 3A. It is noted that the equilibrium is achieved if and only if the function $\Omega_p(k) = 0$ for all positive inflow $e_a(k)$.

2.3 If $C_p(k+1) \neq C^*_{od}$, update the inflow as $e_p(k) := \max[(e_p(k) + \pi d_p(k)), 0]$.

The search direction is denoted by $d_p(k) = -\Omega_p(k) \Big/ \dfrac{\partial \Omega_p(k)}{\partial e_p(k)}$ which is second order, and the step size π is interpolated linearly as $\pi = \dfrac{C^*_{od} - C^0_p(k+1)}{C^1_p(k+1) - C^0_p(k+1)}$,

where $C^1_p(k+1)$ and $C^0_p(k+1)$ represent the corresponding values of $C_p(k+1)$ when $e^*_p(k)$ is being updated with π is taken as 1 and 0 respectively. To determine π, we need two network loadings, one before and one after updating the inflow, to obtain the values of $C^1_p(k+1)$ and $C^0_p(k+1)$.

Step 3: Stopping criteria

3.1. Check if $\left|C_p(k+1) - C^*_{od}\right| \leq \varepsilon_i$, where ε_i is a test value, or of n^i is greater than the predefined maximum number of inner iterations, then go to step 3.2; otherwise, set $n^i := n^i + 1$ and go to step 1;

3.2. If $p = P_{od}$, then go to step 3.3; otherwise p:= p + 1 and go to step 0.7;

3.3. If $od = OD$, then go to step 3.4; otherwise od:= od + 1 and go to step 0.6;

3.4. If $k = K$, then go to step 3.5; otherwise k:= k + 1 and go to step 0.5;

3.5 Define $\xi = \dfrac{\sum_{k \in K} \sum_{a \in A} e_a(k) \left|C_a(k+1) - C^*_{od}\right|}{\sum_{k \in K} \sum_{a \in A} e_a(k) C^*_{od}}$ as a measure of disequilibrium. Note that $\xi = 0$ at dynamic user equilibrium. If n is greater than the predefined maximum number of overall iterations or ξ is sufficiently small, i.e. $\xi \leq \varepsilon$ where ε is a test value, then go to step 3.6; otherwise set n:=n+1 and go to step 0.4;

3.6. Check the total throughput $E_{od} = \sum_{\forall p \in P_{od}} \sum_{\forall k} e_p(k)$ of the system against the total demand J_{od} for each *o-d* pair. If $E_{od} = J_{od}$, then *terminate* the algorithm; otherwise update $C^*_{od} := C^*_{od} + \left[\dfrac{J_{od} - E_{od}}{-\dfrac{dE_{od}}{dC^*_{od}}} \right]$ and go to step 0.2. For networks with mutually distinct routes, Heydecker (2002b) established an expression for the derivative

$$\frac{\partial E_{od}}{\partial C_{od}^{*}} = \sum_{p \in P_{od}} \left(\frac{\left[h'(s_p^0) + f'[\tau_p(s_p^0)] \right] - \left[h'(s_p^1) + f'[\tau_p(s_p^1)] \right]}{\left[h'(s_p^0) + f'[\tau_p(s_p^0)] \right] \left[h'(s_p^1) + f'[\tau_p(s_p^1)] \right]} \right) \overline{Q_p} \quad , \quad \text{where} \quad \overline{Q_p} = \min_{\forall a \in p} Q_a \quad \text{is}$$

defined as the critical capacity of the route *p*.

Discussion

The discretization can bring in difficulties in deciding the instant at when the associated costs should be considered. Heydecker and Verlander (1999) showed that a *predictive* manner should be adopted for plausible assignment results. In a predictive discrete time formulation, the travel cost, which is calculated forward in time due to causality, associated with this flow should be considered at the end of the interval (i.e. at the time $(k+1)\Delta s$), rather than at the start of the interval (i.e. at time $k\Delta s$). The consequence of considering the cost at an inappropriate time was illustrated by Heydecker and Verlander (1999).

The inflow at each departure time interval *k* is calculated in Step 2.2 such that the associated value $\Omega_p(k) = 0$. With this inflow, the total travel cost remains constant over time. We further need Step 2.3 to adjust the inflow assigned at the start time of the assignment such that the total travel cost at the start time of assignment is equal to C_{od}^{*}. Once the value of the travel cost at the start of the assignment is calculated correctly, costs thereafter can follow. Furthermore, the magnitude of C_{od}^{*} and the total traffic volume E_{od} are related as discussed in Section 3.5. Consequently, Step 3.6 is used to adjust the magnitude of C_{od}^{*} such that the algorithm can give the same total volume of traffic E_{od} as the predefined one.

Finally, the algorithm above considers networks with multiple origin-destination pairs connected with mutually distinct routes. In case of networks with multiple origin-destination pairs with overlapping routes, traffic entering the network during the journey time of a traveller from other origins downstream can influence the travel time of travellers from its upstream. As a result, some special computational technique, for example Gauss-Seidel relaxation (see for examples in Sheffi, 1985; Patriksson, 1994), is required. The basic idea of such relaxation scheme is to decompose the assignment problem for networks with overlapping routes connecting multiple origin-destination pairs into several sub-problems. In each sub-problem, we calculate the equilibrium flow for one origin-destination pair, and temporarily neglect the influences from the flows between other origin-destination pairs. When equilibrium is reached for the current origin-destination pair, we proceed with calculations for the next pair. The implementation of this relaxation scheme and the numerical experiments on dynamic traffic assignment problems can be found in Han (2000) and Mun (2002).

3.7 EXAMPLE CALCULATIONS

The section shows the example calculations to illustrate the numerical properties and results of dynamic user equilibrium assignment. In particular, we examine the effects of choosing different travel time models and different degrees of discretization on the assignment results.

3.7.1 Problem setting

We compute dynamic user equilibrium inflow in a network with a single origin-destination pair connected with two parallel links as shown in Figure 3.1. Link 1 has free flow time 3 mins and capacity 20 vehs/min, and link 2 has free flow time 4 mins and capacity 30 vehs/min. Four different link travel time models: the deterministic queuing model, two divided linear traffic models with parameter α_a equal to Δs and $2\Delta s$ respectively, and the whole-link linear traffic model are used to represent the link traffic dynamics. The origin-specific cost $h(s)$ is considered to be a monotone linear decreasing function of time with a gradient $h'(s) = -0.4$. The destination cost function $f(t)$ is piecewise linear which has no penalty for arrivals t before the preferred arrival time $t^* = 50$, and increases with a rate $f'(t) = 2$ afterwards. The test values ε_i and ε are set to be 10^{-10}. The length of the study period, T, is set to be 60 minutes, which is long enough such that that all traffic can be cleared. The time incremental step, Δs, in the calculation is taken as 1 min and hence $K = T / \Delta s = T$. The total volume of traffic J_{od} within the period is fixed at 800.

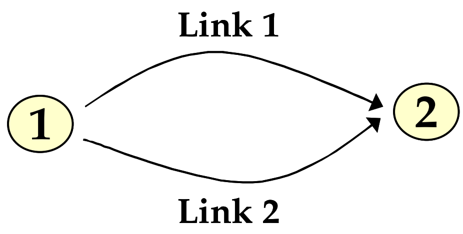
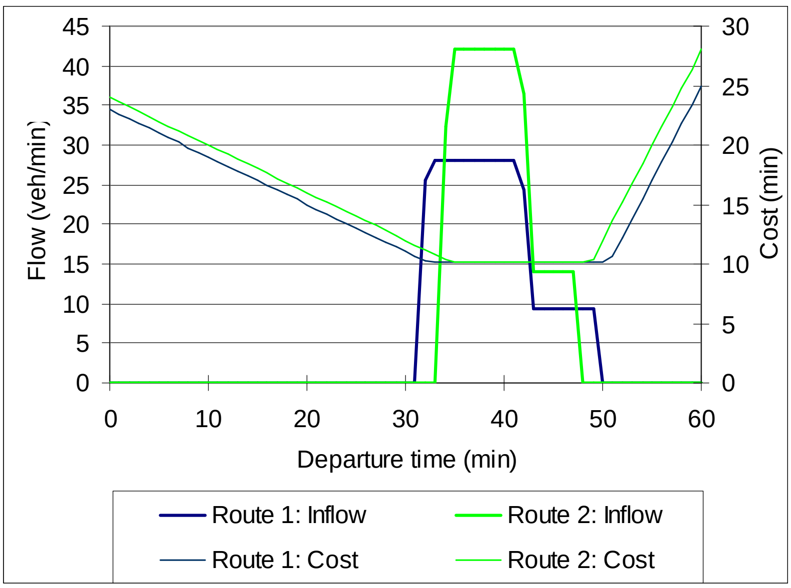


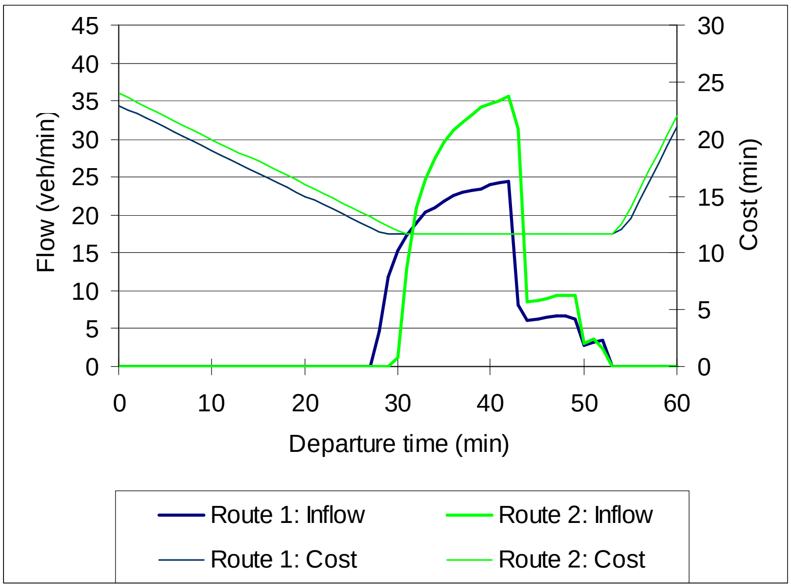
Figure 3.1 Example network

3.7.2 Dynamic user equilibrium assignments

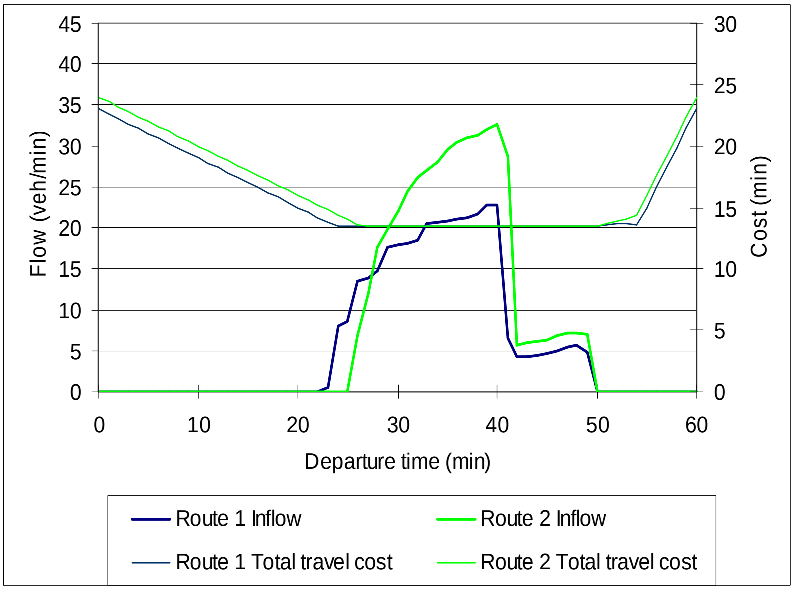
Figure 3.2 shows dynamic user equilibrium assignment results using different travel time models. The assignments show good equilibrations for all travel time models adopted, in which the measure of disequilibrium ξ is below 10^{-17} in all cases. Given the same total volume of traffic, the values of equilibrium cost at which travel take place are estimated to be 10.08 mins, 11.64 mins, 13.43 mins, and 15.58 mins respectively for the deterministic queue, the divided travel time model with $\alpha_a = \Delta s$, the divided travel time model with $\alpha_a = 2\Delta s$, and the whole-link traffic models. These examples show that the larger the congestible portion considered on the link, the higher the resulting travel costs at user equilibrium.



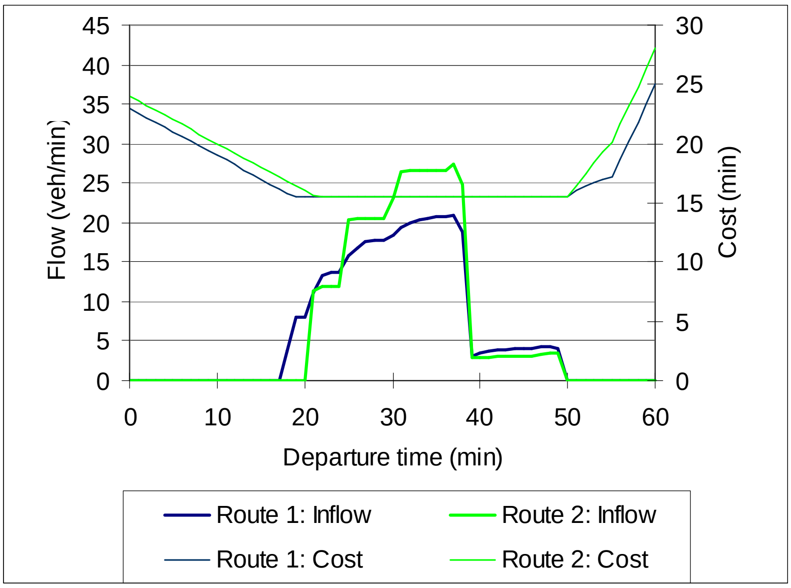
a) Deterministic queuing model



b) Divided linear travel time model ($\alpha_a = \Delta s$)



c) Divided linear travel time model ($\alpha_a = 2\Delta s$)

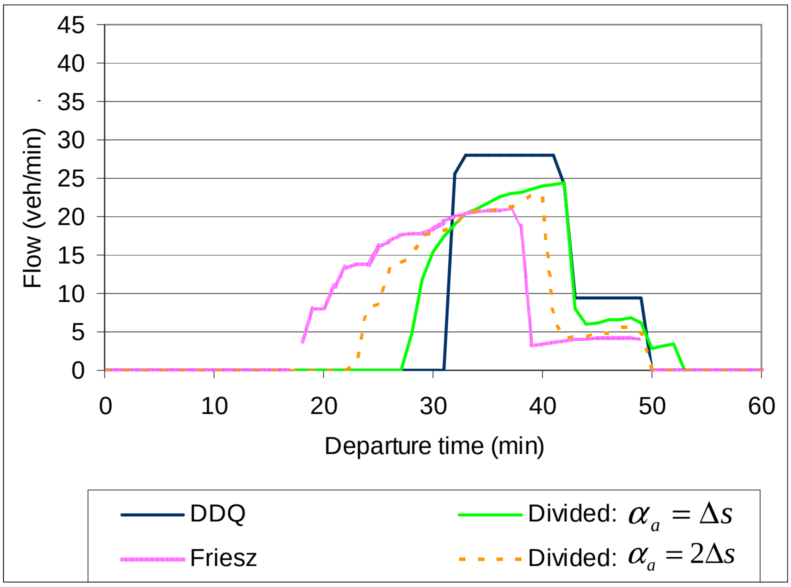


d) Whole-link traffic model

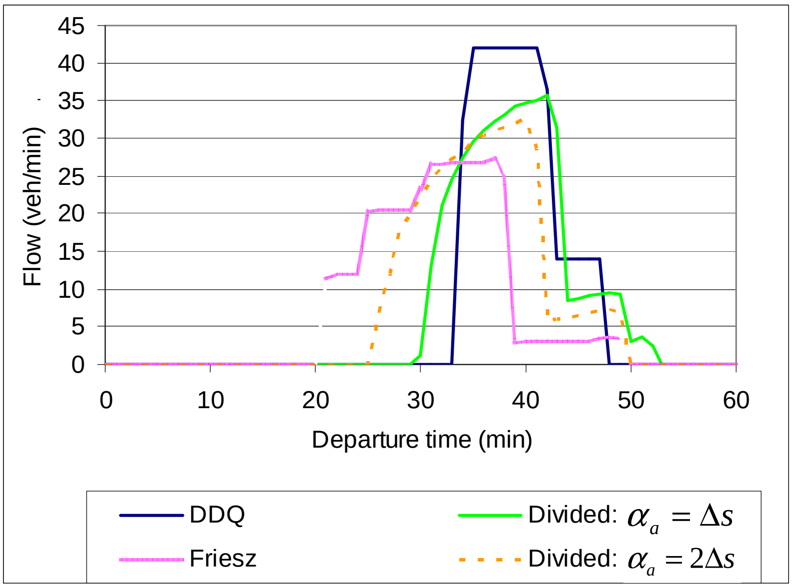
Figure 3.2 Dynamic equilibrium assignments

3.7.3. Effects of choosing different travel time models

We compare the dynamic user equilibrium assignments associated with different travel time models, which are shown in Figure 3.3. With the same amount of travel demand, the assignment inflows spread over longer periods of time for travel time model with a larger congestible portion.



a) Link 1



b) Link 2

Figure 3.3 Dynamic user equilibrium assignments

We also show the associated start times, the end times, and the link volumes of the assignments to each link in Table 3.1. In each of these cases, link 2 serves more traffic than link 1 does due to its higher capacity despite its longer free flow travel time.

Table 3.1 Summary of assignments to each link

	Link 1		
	Start time (min)	End time (min)	Traffic volume (veh)
DDQ	32	49	367.20
Divided: $\alpha_a = \Delta s$	28	52	352.30
Divided: $\alpha_a = 2\Delta s$	23	49	347.81
Friesz	18	49	380.25
	Link 2		
	Start time (min)	End time (min)	Traffic volume (veh)
DDQ	34	47	432.80
Divided: $\alpha_a = \Delta s$	30	52	447.70
Divided: $\alpha_a = 2\Delta s$	26	49	452.19
Friesz	21	49	419.75

Table 3.2 further summarises the start times, the end times, and the durations of assignments to the whole network system. Because all the travel time models commence with the same free flow travel time and associate with the same time-specific costs, following the cost-throughput analysis in Section 3.5, a travel time model with a larger congestible portion

implies a higher total travel cost at user equilibrium which results in a longer duration of assignment as shown in the Table 3.2.

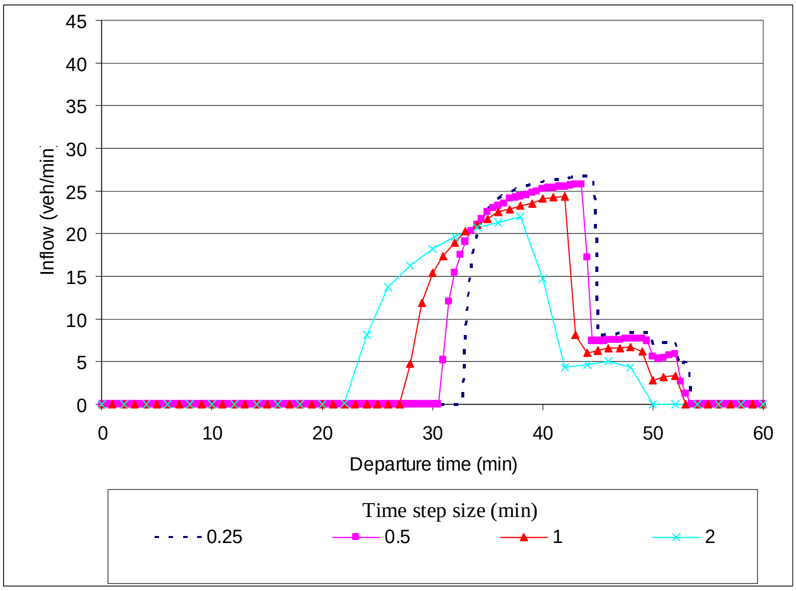
Table 3.2 Summary of assignments

	Start time (min)	End time (min)	Duration (min)
DDQ	32	49	17
Divided: $\alpha_a = \Delta s$	28	52	24
Divided: $\alpha_a = 2\Delta s$	23	49	26
Friesz	18	49	31

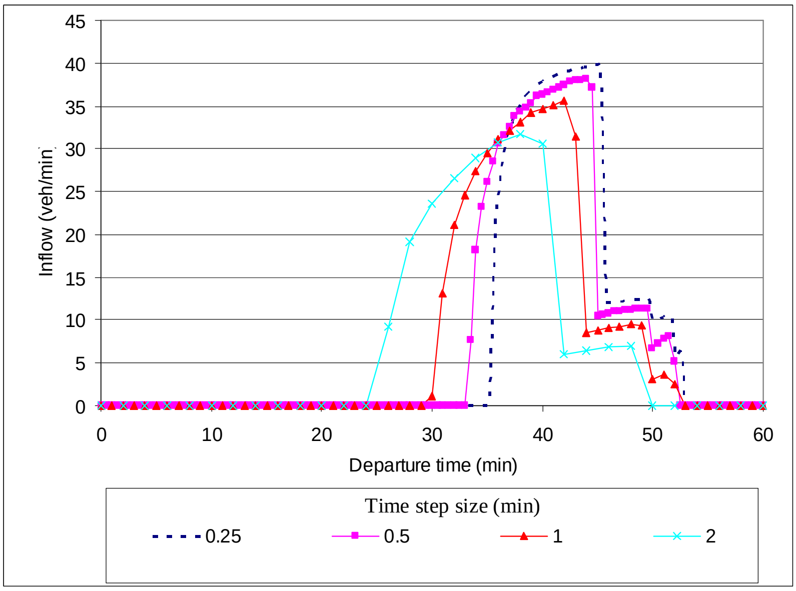
3.7.4. Effects of using different degrees of discretization

Finally, this section investigates the effects of using different degrees of discretization on the assignment solutions. The equilibrium link inflows for the divided linear travel time model with $\alpha_a = \Delta s$ and whole-link traffic model are plotted against different degree of discretization in Figure 3.4 and Figure 3.5 correspondingly. We investigate four different degrees of discretization in which the time incremental step Δs is set to be 0.25 min, 0.5 min, 1 min, and 2 mins respectively.

Similar to the findings in Mun (2002), for the divided linear travel time model, the travel time estimated is dependent on the size of Δs adopted. As a result, the corresponding assignments do differ for different size of discretization as shown in the Figure 3.4. For the whole-link traffic model, the travel time estimated is independent of the value of Δs adopted. Following this, the corresponding assignment shown in Figure 3.5 shows that for the assignment profile converges with respect to the degree of discretization. In general, the solution algorithm gives the assignments with the same start time, end time, and a similar profile for Δs equals to 0.25 min, 0.5 min, 1 min. This experiment also shows that by setting a reasonable discretization Δs = 1 min.



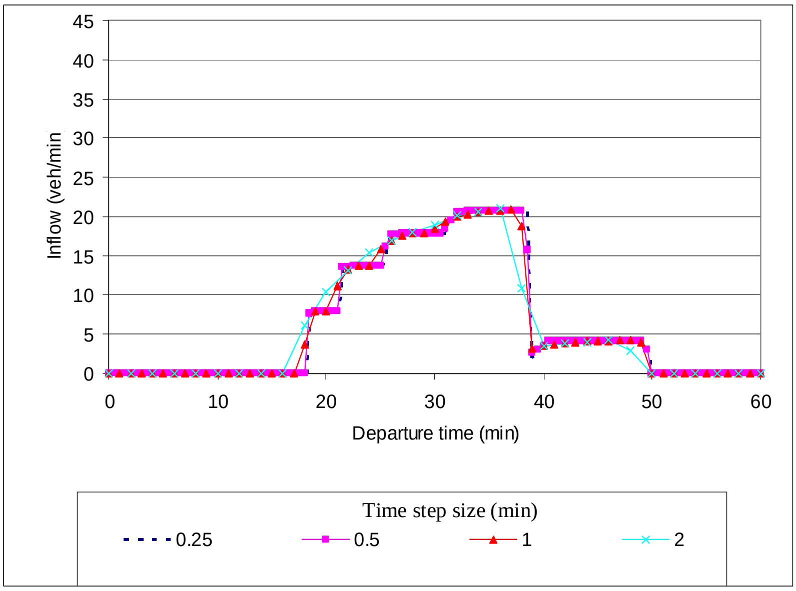
a) Link 1



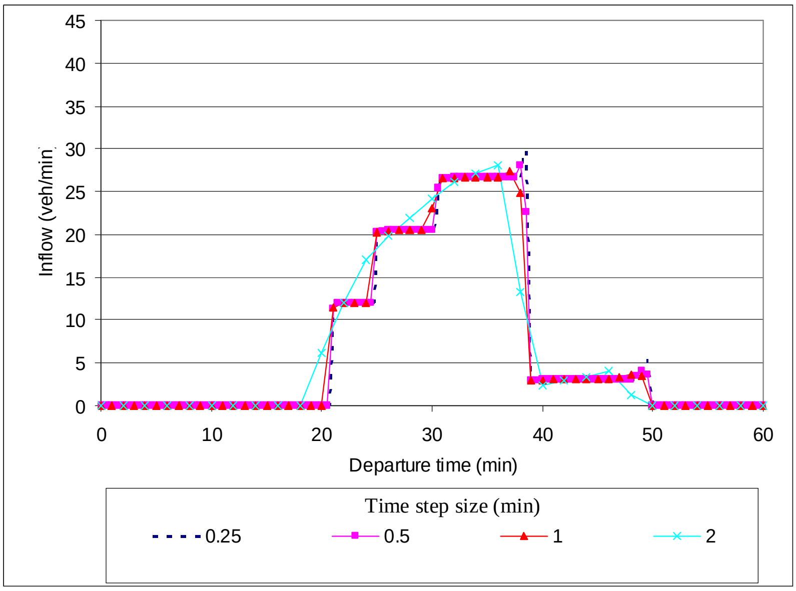
b) Link 2

Figure 3.4 Dynamic user equilibrium assignments with divided linear travel time model

$(\alpha_a = \Delta s)$



a) Link 1



b) Link 2

Figure 3.5 Dynamic user equilibrium assignments with whole-link traffic model

3.8 Discussion

This chapter reviews and discusses dynamic user equilibrium assignment. The formulation, analysis, and calculation of the assignment are illustrated. For dynamic user equilibrium assignment with combined route and departure time choice, the distinct roles of the departure and arrival time-specific cost functions are discussed. Several properties associated with the assignment, including the requirements on the cost functions for an equilibrium solution to exist and the relationship between the travel cost and demand (i.e. cost-throughput relationship), are also established. Based on the analysis, the solution algorithm is proposed using the forward dynamic programming approach. Such approach solves the assignments to a high degree of accuracy. Following Heydecker and Verlander (1999), a predictive assignment is adopted for plausible results. The solution algorithm is applied to numerical examples and the characteristics of the results are discussed. The effects of choosing different travel time models and different degrees of discretization are also investigated.

The deterministic queuing model has been used predominantly in the literature for analysis of this kind. The results presented in this chapter, however, are substantially more general. The analysis developed makes explicit reference to the travel time models and the time-specific cost functions adopted. In addition to the deterministic queuing model, this chapter also analyses and calculates the assignment using other plausible travel time models including the divided linear travel time models and the whole-link traffic model. With different travel time models, it is observed that the corresponding volumes and profiles of equilibrium inflow generated differ substantially. This shows that analyses based on the deterministic queuing model do not apply in general.

Finally, the example calculations shown in this chapter consider networks with mutually distinct routes, i.e. routes without shared bottleneck. In networks with multiple origin-destination pairs connected with overlapping routes, traffic entering the network during the journey time of a traveller from other origins downstream can influence the travel time of traveller from its upstream. In such case, certain relaxation techniques are required to compute the assignment solutions.

Appendix 3A: An expression for the derivative of travel cost with respect to inflow

This appendix derives an approximated expression for the derivative $\frac{\partial \Omega_p}{\partial e_p}$ which is used in the solution algorithm presented in Section 3.6 for dynamic user equilibrium assignment.

Suppose that the travel time is the only component in the total travel cost that is dependent on the inflow e_p. Differentiating the function Ω_p with respect to e_p gives

$$\frac{\partial \Omega_p}{\partial e_p} = \left[1 + f'[\tau_p(k)]\right] \frac{\partial \dot{\tau}_p}{\partial e_p}, \tag{3A-1}$$

in which the route travel time is written as

$$\tau_p(k) = k + \sum_{m=1}^{M(p)} \left[\tau_{a_m}^p(k) - \tau_{a_{m-1}}^p(k)\right] = k + \sum_{m=1}^{M(p)} \tilde{c}_{a_m}^p \left[\tau_{a_{m-1}}^p(k)\right], \tag{3A-2}$$

where $\tau_{a_m}^p(s)$ represents the exit time of the traffic entering the route at time *s* from link a_m on route *p*. The notation a_m represents the *m*-th link on the route, and *M*(*p*) is total number of links on route *p*. This exit time is calculated based on the associated link entry time, $\tau_{a_{m-1}}^p(s)$, as

$$\tau_{a_m}^p(s) = \tau_{a_{m-1}}^p(s) + \tilde{c}_{a_m}^p[\tau_{a_{m-1}}^p(s)], \tag{3A-3}$$

where $\tilde{c}_a[\tau_{a_{m-1}}^p(s)]$ is the link travel time for traffic enters the link at time $\tau_{a_{m-1}}^p(s)$ and $\tau_{a_0}^p(s) = s$ for all routes *p*.

The derivative $\frac{\partial \dot{\tau}_p}{\partial e_p}$ in (3A-1) can be approximated as the following finite differentiation as

$$\frac{\partial \dot{\tau}_p}{\partial e_p} \approx \frac{\Delta \dot{\tau}_p}{\Delta e_p} = \frac{1}{\Delta e_p} \Delta \left\{ 1 + \sum_{m=1}^{M(p)} \frac{d\tilde{c}_{a_m}^{\,p}(\tau_{a_{m-1}}^p)}{d\tau_{a_{m-1}}^p} \dot{\tau}_{a_{m-1}}^p(k) \right\}, \tag{3A-4}$$

in which we assume that the perturbation in the inflow during the time interval *k* only affects the travel time at the final instant of the interval *k* but not at times thereafter.

The travel time models that we adopted in this chapter are all linear in inflow, so we have

$$\frac{d\tilde{c}_{a_m}^{\,p}(\tau_{a_{m-1}}^p)}{d\tau_{a_{m-1}}^p} = \frac{1}{Q_{a_{m-1}}^p} \left(g_{a_{m-1}}^{\,p}[\tau_{a_{m-1}}^p(k)] - g_{a_m}^{\,p}[\tau_{a_{m-1}}^p(k)] \right). \tag{3A-5}$$

where $g_{a_m}^{\,p}(s)$ represents the route-specific flow exiting link a_m on route *p* at time *s*, which also implies that $g_{a_{m-1}}^{\,p}(s)$ is the route-specific flow entering link a_m on route *p*. The notation $Q_{a_{m-1}}^p$ denotes the capacity of link a_{m-1} on route *p*.

Hence, for the linear travel time models adopted in this chapter, we have

$$\begin{aligned} \frac{\Delta \dot{\tau}_p}{\Delta e_p} &= \frac{1}{\Delta e_p} \Delta \left\{ 1 + \sum_{m=1}^{M(p)} \frac{d\tilde{c}_{a_m}^{\,p}(\tau_{a_{m-1}}^p)}{d\tau_{a_{m-1}}^p} \dot{\tau}_{a_{m-1}}^p(k) \right\} \\ &= \frac{1}{\Delta e_p} \Delta \left\{ 1 + \frac{1}{Q_{a_{m-1}}^p} \sum_{m=1}^{M(p)} \left(g_{a_{m-1}}^{\,p}[\tau_{a_{m-1}}^p(k)] - g_{a_m}^{\,p}[\tau_{a_{m-1}}^p(k)] \right) \dot{\tau}_{a_{m-1}}^p(k) \right\}, \\ &= \sum_{m=1}^{M(p)} \left[\frac{1}{Q_{a_{m-1}}^p} \frac{\Delta g_{a_{m-1}}^{\,p}[\tau_{a_{m-1}}^p(k)]}{\Delta e_p} \dot{\tau}_{a_{m-1}}^p(k) \right] \end{aligned} \tag{3A-6}$$

in which we consider that the link outflow $g_{a_m}^{\,p}[\tau_{a_{m-1}}^p(k)]$ is independent of the inflow $e_p(k)$ following causality.

In addition, the flow propagation mechanism gives us that

$$e_p(s) = g_{a_{m-1}}^{\,p}[\tau_{a_{m-1}}^p(s)] \dot{\tau}_{a_{m-1}}^p(s). \tag{3A-7}$$

Consequently,

$$\begin{aligned}\frac{\Delta\dot{\tau}_p}{\Delta e_p} &= \sum_{m=1}^{M(p)}\left[\frac{1}{Q^p_{a_{m-1}}}\frac{\Delta g^p_{a_{m-1}}[\tau^p_{a_{m-1}}(k)]}{\Delta e_p}\dot{\tau}^p_{a_{m-1}}(k)\right] \\ &= \sum_{m=1}^{M(p)}\left[\frac{1}{Q^p_{a_{m-1}}}\frac{\Delta e_p}{\Delta e_p}\frac{1}{\dot{\tau}^p_{a_{m-1}}(k)}\dot{\tau}^p_{a_{m-1}}(k)\right] \\ &= \sum_{m=1}^{M(p)}\frac{1}{Q^p_{a_{m-1}}}\end{aligned}, \qquad (3A\text{-}8)$$

or equivalently,

$$\frac{\Delta\dot{\tau}_p(k)}{\Delta e_p(k)} = \sum_a \delta^a_p\left(\frac{1}{Q_a}\right). \qquad (3A\text{-}9)$$

Finally, it gives the following approximate expression for $\frac{\partial\Omega_p}{\partial e_p}$ as

$$\begin{aligned}\frac{\partial\Omega_p}{\partial e_p} &= [1+f'[\tau_p(k)]]\frac{\Delta\dot{\tau}_p}{\Delta e_p} \\ &\approx [1+f'[\tau_p(k)]]\sum_a \delta^a_p\left(\frac{1}{Q_a}\right)\end{aligned}. \qquad (3A\text{-}10)$$

4. DYNAMIC SYSTEM OPTIMAL ASSIGNMENT WITH DEPARTURE TIME CHOICE

4.1 INTRODUCTION

Dynamic user equilibrium is used to represent the distribution of traffic that arises when travellers consider their own interests alone. However, such distribution of traffic generally does not lead to the best possible use of the transport system, because the user equilibrium considers that each individual traveller is acting only in their own interests, rather than those of the whole system.

This chapter investigates dynamic system optimal assignment, which is an important yet underdeveloped area in the literature. We suppose that there is a central *system manager* distributing the traffic over time within a fixed horizon so that the total, rather than individual, benefit of all travellers in the system is maximised. Although the system optimal assignment is not a realistic representation of traffic, it provides a bound on performance that shows how the transport planner or engineer can make the best use of the road system, and as such it is a useful benchmark for evaluating various transport policy measures. These measures include time-varying pricing (Yang and Huang, 1997; Wie and Tobin, 1998; Polak and Heydecker, 2006), network access control (Smith and Ghali, 1990; Lovell and Daganzo, 2000; Erera et al., 2002), and road capacity management (Ghali and Smith, 1995; Heydecker, 2002b).

Dynamic system optimal assignment, which is a kind of dynamic optimization problem, is difficult to solve. Merchant and Nemhauser (1978a; b) were the first to consider, formulate, and analyse this as an optimal control problem in which traffic is modelled by the outflow model (see Section 2.3.2.1). Merchant and Nemhauser's (1978a; b) formulation was then studied by many others in the past three decades (see for example, Ho, 1980; Carey, 1987; Friesz et al., 1989; Janson, 1991; Carey and Srinivasan, 1993; Wie et al., 1995a; Yang and Huang, 1997; Wie and Tobin, 1998; Friesz et al., 2004). On the one hand, this formulation provides some attractive mathematical properties for analysis. On the other hand, the plausibility of outflow traffic model was later found to be questionable as discussed in Section 2.3.2. In particular, the outflow traffic model ignores the importance of ensuring causality and proper flow propagation as first shown by Tobin (1993), followed by many others such as Astarita (1996); Heydecker and Addison (1998); Friesz and Bernstein (2000); Mun, (2001).

This chapter investigates dynamic system optimal assignment with departure time choice based on plausible travel time models. In Section 4.2, the assignment is reformulated as a *state-dependent* optimal control problem, with which optimal inflow profile is sought to minimize the total system travel cost given a fixed travel demand. The state-dependent control theoretic formulation was investigated by Friesz et al. (2001) and Friesz and Mookherjee (2006) for dynamic user equilibrium assignment, and by Friesz et al. (2004) for dynamic flow routing in data network. This study applies this state-dependent control theoretic formulation to dynamic system optimal assignment. The current formulation considers transport systems with one origin-destination pair connected with mutually distinct routes consisting of one single link. Due to the special properties of the deterministic queuing model as discussed in Section 2.3.3.1, we also particularly show the analysis of dynamic system optimal assignment with such traffic model. Moreover, as any kind of optimization problem, solving dynamic system optimal assignment requires the derivative of the objective function (i.e. total system travel cost) with respect to the control variables (i.e. inflow). In Section 4.3, a novel sensitivity analysis is developed for this. The sensitivity analysis is developed through general flow propagation mechanisms and the analysis is not restricted to any specific travel time model. In Section 4.4, solution algorithms are developed and presented for implementing the sensitivity analyses and solving dynamic system optimal assignments for a range of travel time models. In section 4.5, example calculations are given and the characteristics of the numerical results are discussed. Given the difficulties in solving dynamic system optimal assignment, in Section 4.5, we further suggest an alternative solution algorithm which may be considered to replace the original one for assignments with better quality. Section 4.6 proposes some practical tolling strategies for managing dynamic network traffic flow based on the study of dynamic system optimal assignments. Finally, some concluding remarks are given in Section 4.7.

4.2 DYNAMIC SYSTEM OPTIMAL ASSIGNMENT WITH DEPARTURE TIME CHOICE

Dynamic system optimal assignment with departure time choice is formulated as the following optimal control problem. This seeks an optimal inflow profile $e_a(s)$ that minimizes the total system travel cost within the study period, *T*, given a fixed total amount of traffic, J_{od}:

$$\min_{e_a(s)} Z = \sum_{\forall a} \int_{s=0}^{T} C_a(s) e_a(s) ds \qquad (4\text{-}1)$$

subject to:

$$\frac{dG_a[\tau_a(s)]}{ds} = e_a(s) \quad ,\forall a, \forall s \qquad (4\text{-}2)$$

$$\frac{dx_a(s)}{ds} = e_a(s) - g_a(s) \quad ,\forall a, \forall s \qquad (4\text{-}3)$$

$$\frac{dE_a(s)}{ds} = e_a(s) \quad ,\forall a, \forall s \qquad (4\text{-}4)$$

$$\sum_{\forall a} E_a(T) = J_{od} \qquad (4\text{-}5)$$

$$e_a(s) \geq 0 \quad ,\forall a, \forall s \qquad (4\text{-}6)$$

The objective function (4-1) was first adopted by Merchant and Nemhauser (1978a; b), and by several other researchers since then. The notation $C_a(s)$ there represents the total travel cost associated with a departure time *s*, as defined previously in Section 3.3.2. Equation (4-2) ensures the proper flow propagation along each link. Equation (4-3) is the flow conservation constraint, which serves here as the state equation that governs the evolution of link traffic, $x_a(\cdot)$. Equation (4-4) defines the cumulative link inflow $E_a(\cdot)$. Equation (4-5) specifies the amount of total throughput J_{od} between the origin-destination pair within the time horizon *T*. Condition (4-6) ensures the non-negativity of the control variable, $e_a(\cdot)$. Given a non-negative inflow $e_a(\cdot)$, the corresponding outflow $g_a(\cdot)$ and the link traffic volume $x_a(\cdot)$ are guaranteed to be non-negative (see Proposition 2.2 in Chapter 2). Hence, we do not need any additional constraints to ensure the non-negativity of $g_a(\cdot)$ and $x_a(\cdot)$ as proposed in Friesz et al. (2001). In addition, because the travel time models adopted in the formulation satisfies FIFO discipline structurally, we do not need additional constraint(s) to ensure FIFO as well.

One technical difficulty arise because the duration of the time lag between changes to the control variable, $e_a(s)$, and the corresponding response, $g_a[\tau_a(s)]$, depends on the state x_a. The time lag between the control and the response is the link travel time that is a function of the state variable $x_a(s)$. This state-dependent control theoretic formulation is unorthodox in the control theory literature. Its properties and application to dynamic equilibrium assignment were studied by Friesz et al. (2001). We derive the necessary conditions for dynamic system optimal assignment in the following proposition.

Proposition 4.1: A necessary condition for the solution of the dynamic optimization problem (4-1) – (4-6) is

$$e_a(s)\begin{cases} > 0 \Rightarrow C_a(s) + \Psi_a(s) + \lambda_a(s) - \gamma_a(s) = \\ = 0 \Rightarrow C_a(s) + \Psi_a(s) + \lambda_a(s) - \gamma_a(s) \geq \end{cases}\Bigg\} \mu_a(s) = \nu_{od} \quad ,\forall a, \forall s \in [0,T], \qquad (4\text{-}7)$$

where $\lambda_a(s)$ and $\gamma_a(s) = \lambda_a[\tau_a(s)]$ are the respective costate variables for the flow propagation constraint (4-2) and flow conservation constraint (4-3), and $\mu_a(s) = \nu_{od}$ is the costate associated with constraint (4-4) and is constant with respect to time with magnitude given by ν_{od} which is a multiplier of constraint (4-5). The value of ν_{od} is determined by the total amount of traffic J_{od}. The notation $\Psi_a(s)$ represents the sensitivity of the value of the objective function with respect to a perturbation in the inflow profile, where $\Psi_a(s) = \int_{t=0}^{T} \left[\frac{\partial C_a}{\partial u_s}\bigg|_t e_a(t) \right] dt$ refers to the additional travel cost imposed by an additional amount of traffic, u_s, at time s to existing travellers in the system. This additional cost is also termed as *dynamic externality*. We define the parameters u_s be a perturbation in the inflow profile for which

$$\frac{de_a(t)}{du_s} = \begin{cases} 1 & \text{if } t \in [s, s+ds) \\ 0 & \text{otherwise} \end{cases}, \tag{4-8}$$

in which ds represents the incremental time step[§§].

The value of $\Psi_a(s)$ is equal to the total change in the value of the total system travel cost Z with respect to this change in the inflow profile during the time interval $[s, s+ds)$.

Proof:

The objective function Z is first augmented with the constraints to form the following Lagrangian:

$$\min_{e_a^*(s)} Z^* = \sum_{\forall a} \int_0^T \left\{ \begin{aligned} & C_a(s)e_a(s) + \lambda_a(s)\left\{[e_a(s) - g_a(s)] - \frac{dx_a(s)}{ds}\right\} + \mu_a(s)\left\{\frac{dE_a(s)}{ds} - e_a(s)\right\} \\ & + \gamma_a(s)\left\{\frac{dG_a[\tau_a(s)]}{ds} - e_a(s)\right\} - \rho_a(s)e_a(s) \end{aligned} \right\} ds \\ + \nu_{od}\left(J_{od} - \sum_{\forall a} E_a(T)\right),$$

(4-9)

where $\lambda_a(s)$ and $\gamma_a(s)$ are the respective costate variables for the flow conservation and flow propagation constraints; and $\mu_a(s)$ and $\rho_a(s)$ are the associated multipliers on the cumulative and the non-negativity constraints of the control variables respectively. Finally, ν_{od} is the multiplier associated with the total throughput. Using integration by parts, the terms involving $\frac{dx_a(s)}{ds}$ and $\frac{dE_a(s)}{ds}$ in the integrand over time can be rewritten as follows:

[§§] The inflow $e_a(s)$ is a continuous quantity with respect to time. Strictly speaking, the value of $\partial e_a(s)$ is zero if we refer to only one particular instant, and hence it will not be effective on the cost $C_a(s)$. To validate the analysis, we propose a notation ∂u_s to represent the change in inflow throughout a time interval rather than at a particular instant.

$$\int_0^T \lambda_a(s)\frac{dx_a(s)}{ds}ds = \int_0^T \lambda_a(s)dx_a(s) = \lambda_a(T)x_a(T) - \lambda_a(0)x_a(0) - \int_0^T x_a(s)\frac{d\lambda_a(s)}{ds}ds \text{ ,} \tag{4-10}$$

and

$$\int_0^T \mu_a(s)\frac{dE_a(s)}{ds}ds = \int_0^T \mu_a(s)dE_a(s) = \mu_a(T)E_a(T) - \mu_a(0)E_a(0) - \int_0^T E_a(s)\frac{d\mu_a(s)}{ds}ds \tag{4-11}$$

in which the initial values $x_a(0)$ and $E_a(0)$ are considered to be zero. Consequently, the Lagrangian Z^* becomes

$$\begin{aligned} Z^* &= \sum_{\forall a}\left[-\lambda_a(T)x_a(T) + \mu_a(T)E_a(T)\right] + \nu_{od}\left(J_{od} - \sum_{\forall a}E_a(T)\right) \\ &+ \sum_{\forall a}\int_0^T\left[H_a(s) + \frac{d\lambda_a(s)}{ds}x_a(s) - \frac{d\mu_a(s)}{ds}E_a(s)\right]ds \end{aligned} \text{ ,} \tag{4-12}$$

in which we can identify the Hamiltonian function:

$$\begin{aligned} H_a(s) &= C_a(s)e_a(s) + \lambda_a(s)\left[e_a(s) - g_a(s)\right] - \mu_a(s)e_a(s) \\ &+ \gamma_a(s)\left\{g_a\left[\tau_a(s)\right]\frac{d\tau_a(s)}{ds} - e_a(s)\right\} - \rho_a(s)e_a(s) \end{aligned} \text{ .} \tag{4-13}$$

Before proceeding forward, we define the parameters v_s be a perturbation in the outflow profile for which

$$\frac{dg_a(t)}{dv_s} = \begin{cases} 1 & \text{if } t \in [s, s+ds) \\ 0 & \text{otherwise} \end{cases} \tag{4-14}$$

for consistent with the inflow e_a.

The variation δZ^* of Z^* with respect to its variables is derived as

$$\delta Z^* = \sum_{\forall a}\left[-\lambda_a(T)\delta x_a(T) + \mu_a(T)\delta E_a(T)\right] + \sum_{\forall a}\int_0^T \left(\frac{\partial H_a}{\partial u_s}\right)\delta e_a(s)ds$$

$$+\sum_{\forall a}\int_0^T\left[\frac{\partial H_a}{\partial v_s}\delta g_a(s) + \frac{\partial H_a}{\partial v_{\tau_a(s)}}\delta g_a\left[\tau_a(s)\right] + \left(\frac{\partial H_a}{\partial x_a} + \frac{d\lambda_a(s)}{ds}\right)\delta x_a(s)\right]ds, \qquad (4\text{-}15)$$

$$-\sum_{\forall a}\int_0^T \frac{d\mu_a(s)}{ds}\delta E_a(s)ds - \nu_{od}\sum_{\forall a}\delta E_a(T)$$

in which $\frac{\partial H_a}{\partial u_s}$, $\frac{\partial H_a}{\partial v_s}$, $\frac{\partial H_a}{\partial v_{\tau_a(s)}}$, and $\frac{\partial H_a}{\partial x_a}$ represent the derivatives of Hamiltonian function with respect to its corresponding variables: $e_a(s)$, $g_a(s)$, $g_a[\tau_a(s)]$, and $x_a(s)$ respectively.

Applying the change of variables, $t = \tau_a(s) \Rightarrow dt = \dot{\tau}_a(s)ds$, the bounds of the integral are changed accordingly: $s = 0 \Rightarrow t = \tau_a(0)$ and $s = T \Rightarrow t = \tau_a(T)$.

The variation with respect to $g_a\left[\tau_a(s)\right]$ can now be transformed to

$$\int_0^T\left[\frac{\partial H_a}{\partial v_{\tau_a(s)}}\delta g_a\left[\tau_a(s)\right]\right]ds = \int_{\tau_a(0)}^{\tau_a(T)}\left[\frac{\partial H_a}{\partial v_{\tau_a(s)}}\delta g_a(t)\right]\frac{dt}{\dot{\tau}_a(s)}$$

$$= \int_{\tau_a(0)}^{\tau_a(T)}\frac{\partial H_a}{\partial v_{\tau_a[\sigma_a(t)]}}\frac{1}{\dot{\tau}_a[\sigma_a(t)]}\delta g_a(t)dt \qquad (4\text{-}16)$$

$$= \int_{\tau_a(0)}^{\tau_a(T)}\frac{\partial H_a}{\partial v_{\tau_a[\sigma_a(s)]}}\frac{1}{\dot{\tau}_a[\sigma_a(s)]}\delta g_a(s)ds$$

The time horizon T is taken such that it is long enough for all traffic to be cleared by the end of it, the integral on the right hand side in Equation (4-17) only needs to be calculated up to time T as

$$\int_0^T\left[\frac{\partial H_a}{\partial v_{\tau_a(s)}}\delta g_a\left[\tau_a(s)\right]\right]ds = \int_{\tau_a(0)}^{T}\frac{\partial H_a}{\partial v_{\tau_a[\sigma_a(s)]}}\frac{1}{\dot{\tau}_a[\sigma_a(s)]}\delta g_a(s)ds\,. \qquad (4\text{-}17)$$

Finally, δZ^* becomes

$$\delta Z^* = \sum_{\forall a}\left[-\lambda_a(T)\delta x_a(T) + \mu_a(T)\delta E_a(T)\right] + \sum_{\forall a}\int_0^T \left(\frac{\partial H_a}{\partial u_s}\right)\delta e_a(s)ds$$
$$+\sum_{\forall a}\int_0^{\tau_a(0)}\left(\frac{\partial H_a}{\partial v_s}\right)\delta g_a(s)ds$$
$$+\sum_{\forall a}\int_{\tau_a(0)}^{T}\left[\frac{\partial H_a}{\partial v_s} + \left.\left(\frac{\partial H_a}{\partial v_s}\frac{1}{\dot{\tau}_a}\right)\right|_{\sigma_a(s)}\right]\delta g_a(s)ds$$
$$+\sum_{\forall a}\int_0^T\left[\left(\frac{\partial H_a}{\partial x_a} + \frac{d\lambda_a(s)}{ds}\right)\delta x_a(s)\right]ds - \sum_{\forall a}\int_0^T \frac{d\mu_a(s)}{ds}\delta E_a(s)ds - v_{od}\sum_{\forall a}\delta E_a(T)$$

(4-18)

Or,

$$\delta Z^* = \sum_{\forall a}\left[-\lambda_a(T)\delta x_a(T) + \mu_a(T)\delta E_a(T)\right] + \sum_{\forall a}\int_0^T \left(\frac{\partial H_a}{\partial u_s}\right)\delta e_a(s)ds$$
$$+\sum_{\forall a}\int_{\tau_a(0)}^{T}\left[\frac{\partial H_a}{\partial v_s} + \left.\left(\frac{\partial H_a}{\partial v_s}\frac{1}{\dot{\tau}_a}\right)\right|_{\sigma_a(s)}\right]\delta g_a(s)ds$$
$$+\sum_{\forall a}\int_0^T\left[\left(\frac{\partial H_a}{\partial x_a} + \frac{d\lambda_a(s)}{ds}\right)\delta x_a(s)\right]ds - \sum_{\forall a}\int_0^T \frac{d\mu_a(s)}{ds}\delta E_a(s)ds - v_{od}\sum_{\forall a}\delta E_a(T)$$

(4-19)

because there is no outflow between time 0 and the first arrival time $\tau_a(0)$.

The optimality is achieved when Z^* is stationary (i.e. $\delta Z^* = 0$) with respect to all variations. The stationarity conditions are recognized as:

$$\frac{\partial H_a}{\partial u_s} = C_a(s) + \Psi_a(s) + \lambda_a(s) - \gamma_a(s) - \mu_a(s) - \rho_a(s) = 0 \quad ,\forall a, \forall s; \qquad (4\text{-}20)$$

$$\frac{\partial H_a}{\partial v_s} + \left.\left(\frac{\partial H_a}{\partial v_s}\frac{1}{\dot{\tau}_a}\right)\right|_{\sigma_a(s)} = -\lambda_a(s) + \gamma_a\left[\sigma_a(s)\right] = 0 \quad ,\forall a, \forall s \in [\tau_a(0), T]; \qquad (4\text{-}21)$$

$$\frac{\partial H_a}{\partial x_a} + \frac{d\lambda_a(s)}{ds} = \left(1 + f'[\tau_a(s)]\right)\frac{e_a(s)}{Q_a} + \frac{d\lambda_a(s)}{ds} = 0 \quad ,\forall a, \forall s; \tag{4-22}$$

$$\lambda_a(T) = 0 \quad , \ \forall a; \tag{4-23}$$

$$\frac{d\mu_a(s)}{ds} = 0 \quad ,\forall a, \forall s; \tag{4-24}$$

$$\mu_a(T) - \nu_{od} = 0 \quad ,\forall a. \tag{4-25}$$

We also have the following Karush-Kuhn-Tucker (KKT) conditions hold for the non-negativity constraints on the inflow:

$$e_a(s) \geq 0 \quad ,\forall a, \ \forall s; \tag{4-26}$$

$$e_a(s)\rho_a(s) = 0 \quad ,\forall a, \forall s; \tag{4-27}$$

$$\rho_a(s) \geq 0 \quad ,\forall a, \forall s. \tag{4-28}$$

With equations (4-24) and (4-25), we can deduce that $\mu_a(s)$ will remain constant at $\mu_a(T) = \nu_{od}$ for all s within T because $\frac{d\mu_a(s)}{ds} = 0$.

Furthermore, equation (4-21) can be written equivalently as

$$\gamma_a(s) = \lambda_a[\tau_a(s)] \quad ,\forall a, \forall s \in [\tau_a(0), T], \tag{4-29}$$

and the evolution of the costate variable $\lambda_a(s)$ is governed by (4-22) for all s with the terminal condition (4-23). Finally, combining (4-20) and the KKT conditions (4-26), (4-27), and (4-28), we then get the following conditions for dynamic system optimum:

$$e_a(s)\begin{cases} > 0 \Rightarrow C_a(s) + \Psi_a(s) + \lambda_a(s) - \gamma_a(s) = \\ = 0 \Rightarrow C_a(s) + \Psi_a(s) + \lambda_a(s) - \gamma_a(s) \geq \end{cases}\!\!\Bigg\}\,\mu_a(s) = \nu_{od} \quad ,\forall a, \forall s \in [0, T], \tag{4-30}$$

as is to be shown. □

Moreover, using the costate equation (4-22) and the transversality condition (4-23), the costate variable $\lambda_a(s)$ for any time s can be calculated recursively working backward in time:

$$\begin{aligned}\lambda_a(s) &= -\frac{1}{Q_a}\int_{t=T}^{s}\left(1+f'[\tau_a(t)]\right)e_a(t)dt \\ &= \frac{1}{Q_a}\int_{t=s}^{T}\left(1+f'[\tau_a(t)]\right)e_a(t)dt \qquad ,\forall a,\forall s\end{aligned} \tag{4-31}$$

Similar to their static counterparts (see Sheffi, 1985), Proposition 4.1 shows that the dynamic system optimal assignment can be reduced to an equivalent dynamic user equilibrium assignment formulation in which additional components of the cost, $[\Psi_a(s)+\lambda_a(s)-\gamma_a(s)]$, are introduced to each link and departure time in use. Each of these components is discussed in detail in Section 4.3.

We further investigate on the *sufficiency* of these conditions for a system optimal assignment. It is found that sufficiency cannot be guaranteed with the current formulation and objective function. Further research will be necessary on sufficient conditions for dynamic system optimal assignment, but we nevertheless include the analysis that has been developed at this stage in Appendix 4A for readers' reference.

Discussion: Dynamic system optimal assignment with the deterministic queuing model

The previous analysis requires the state variable, $x_a(s)$, to be continuously differentiable with respect to the inflow. However, this is not the case for the deterministic queuing model, in which the state variable is not differentiable at the point when the inflow equals to capacity (see discussion in Section 2.3.3.1). Arnott, de Palma and Lindsey (1998) derived the dynamic system optimal solution for the deterministic queuing model by intuitive reasoning. They showed that the period of assignment in dynamic system optimum is the same as in dynamic user equilibrium. In addition, the dynamic system optimal inflow profile should equal to the link capacity through the assignment period.

Consider that the frequency of application and simplicity of the deterministic queuing model in the literature of dynamic traffic modeling and management (see for example, Vickrey, 1969; Laih, 1994; Arnott, de Palma, and Lindsey, 1993; 1998; Yang and Huang, 1997; Huang and Lam, 2002), the following proposition derives the optimality conditions of dynamic system optimal assignment for the deterministic queuing model by exploiting the analysis in Proposition 4.1. To the knowledge of the author, such detailed mathematical analysis for dynamic system optimal assignment with deterministic queuing model has not been found in the literature.

Proposition 4.2: The necessary condition for dynamic system optimum with deterministic queuing model is having the inflow profile $e_a(s) = Q_a$ for all links *a* for all times *s*.

Proof:

To derive the system optimality conditions for the deterministic queuing model, we need to consider separately the uncongested case (i.e. $x_a(s) = 0$ and $e_a(s - \phi_a) \le Q_a$) and the congested case (i.e $x_a(s) > 0$ or $e_a(s - \phi_a) \ge Q_a$).

<u>Case 1: Uncongested condition</u>

We consider the travel link be uncongested during the assignment period, i.e. when $x_a(s) = 0$ and $e_a(s - \phi_a) \le Q_a$, and show that the limiting case of $e_a(s) = Q_a$ is preferable. With the deterministic queuing model, the associated link travel time is taken as the link free flow travel time that is independent of the inflow. Under such condition, dynamic system optimal assignment can be considered as minimizing the total system cost (4-1) subject to

$$e_a(s - \phi_a) \le Q_a, \tag{4-32}$$

together with the constraints (4-4) to (4-6). We do not need to include the flow propagation condition (4-2) because the outflow $g_a(s)$ equals to the inflow $e_a(s)$ for all time s and all links a. We also do not need the state equation (4-3) because the state variable $x_a(s)$ equals to zero for all time s. Furthermore, without queuing, the total travel cost $C_a(s)$ is independent of the inflow and is only dependent on the departure time s.

We define $\omega_a(s)$ as the multiplier for the inequality constraint $e_a(s-\phi_a) \le Q_a$. The objective function Z is augmented with the constraints (4-4) - (4-6), and (4-32) to form the following Lagrangian:

$$\begin{aligned}\min_{e_a^*(s)} Z^* = &\sum_{\forall a}\int_0^T \left\{ C_a(s)e_a(s) + \mu_a(s)\left\{\frac{dE_a(s)}{ds} - e_a(s)\right\} - \rho_a(s)e_a(s) - \omega_a(s)[Q_a - e_a(s)]\right\} ds \\ &+ \nu_{od}\left(J_{od} - \sum_{\forall a} E_a(T)\right)\end{aligned} \tag{4-33}$$

which can be transformed using integration by parts as in Proposition 4.1 to

$$\begin{aligned}Z^* = &\sum_{\forall a}\mu_a(T)E_a(T) + \nu_{od}\left(J_{od} - \sum_{\forall a}E_a(T)\right) \\ &+ \sum_{\forall a}\int_0^T \left[H_a(s) - \frac{d\mu_a(s)}{ds}E_a(s)\right] ds\end{aligned}, \tag{4-34}$$

in which the Hamiltonian function in this case is:

$$\begin{aligned}H_a(s) = &C_a(s)e_a(s) - \mu_a(s)e_a(s) \\ &- \rho_a(s)e_a(s) - \omega_a(s)[Q_a - e_a(s)]\end{aligned} \tag{4-35}$$

The variation δZ^* of Z^* with respect to its variables is derived as

$$\delta Z^* = \sum_{\forall a} \mu_a(T)\delta E_a(T) + \sum_{\forall a} \int_0^T \left(\frac{\partial H_a}{\partial u_s} \right) \delta e_a(s) ds$$
$$- \sum_{\forall a} \int_0^T \frac{d\mu_a(s)}{ds} \delta E_a(s) ds - \nu_{od} \sum_{\forall a} \delta E_a(T) \qquad (4\text{-}36)$$

The set of the stationary conditions for this dynamic optimization problem is recognized as:

$$\frac{\partial H_a}{\partial u_s} = C_a(s) - \mu_a(s) + \omega_a(s) - \rho_a(s) = 0 \quad ,\forall a, \forall s; \qquad (4\text{-}37)$$

$$\frac{d\mu_a(s)}{ds} = 0 \quad ,\forall a, \forall s; \qquad (4\text{-}38)$$

$$\mu_a(T) - \nu_{od} = 0 \quad ,\forall a. \qquad (4\text{-}39)$$

We also have the following set of Karush-Kuhn-Tucker (KKT) conditions hold for the non-negativity constraint on the inflow:

$$e_a(s) \geq 0 \quad ,\forall a, \forall s; \qquad (4\text{-}40)$$

$$e_a(s)\rho_a(s) = 0 \quad ,\forall a, \forall s; \qquad (4\text{-}41)$$

$$\rho_a(s) \geq 0 \quad ,\forall a, \forall s. \qquad (4\text{-}42)$$

Combining (4-37) – (4-42), we have the following conditions on inflow to be satisfied at optimality

$$e_a(s) \begin{Bmatrix} > 0 \Rightarrow C_a(s) = \\ = 0 \Rightarrow C_a(s) \geq \end{Bmatrix} \mu_a(s) - \omega_a(s) = \nu_{od} \quad ,\forall a, \forall s \in [0,T]. \qquad (4\text{-}43)$$

We also have another set of Karush-Kuhn-Tucker (KKT) conditions on the inflow due to the inequality constraint (4-32):

$$Q_a - e_a(s) \geq 0 \qquad ,\forall a, \forall s\,; \tag{4-44}$$

$$\omega_a(s)[Q_a - e_a(s)] = 0 \qquad ,\forall a, \forall s\,; \tag{4-45}$$

$$\omega_a(s) \geq 0 \qquad ,\forall a, \forall s\,. \tag{4-46}$$

Following (4-43), consider the case $e_a(s) > 0$, we have the corresponding travel cost as $C_a(s) = \mu_a(s) - \omega_a(s) = \nu_{od}$. Moreover, with (4-44) – (4-46), we have the following condition on positive inflow at optimality as:

$$e_a(s)\begin{Bmatrix} < Q_a \Rightarrow C_a(s) = \\ = Q_a \Rightarrow C_a(s) \leq \end{Bmatrix}\mu_a(s) = \nu_{od} \quad ,\forall a, \forall s \in [0,T]. \tag{4-47}$$

In the uncongested condition, the link travel time is constant at free flow. Following the discussion on time-specific costs in Section 3.4, we also have the derivative of the origin-specific cost $h'(\cdot)$ be negative and the derivative of the destination-specific cost $f'(\cdot)$ be non-negative at all times. Consequently, the total travel cost, $C_a(s) = h(s) + \phi_a + f(s + \phi_a)$, with the deterministic queuing model under uncongested condition will decrease initially over time due to the monotonic decreasing function $h(\cdot)$ of time, and increase in later stage of assignment when the increasing function $f(\cdot)$ starts to dominate due to late arrivals. Consequently, $C_a(s)$ is greater than the cost ν_{od} when the departure time interval s lies outside the assignment period; is equal to ν_{od} at the start and end time intervals of assignment; is smaller than ν_{od} within the assignment period. This gives the dynamic system optimal inflow $e_a(s) = Q_a$ for all link a within the assignment period, and $e_a(s) = 0$ outside the assignment period.

<u>Case 2: Congested condition</u>

In congested case, there is a traffic queue being developed on the link and the associated outflow rate $g_a(\cdot)$ is equal to the link capacity Q_a. The condition for congestion in the deterministic queuing model can be represented by

$$E_a(s) \geq Q_a(s - s_{0a}) \qquad ,\forall a, \forall s, \tag{4-48}$$

where s_{0a} is the first time at which the link is congested. The dynamic system optimal assignment problem is then considered as minimizing the total system cost (4-1) subject to condition (4-48) together with constraints (4-3) to (4-6). We do not need to include the flow propagation constraint (4-2) because the outflow $g_a(s)$ equals to the capacity Q_a for all time *s* and all links *a*.

We define $\varpi(s)$ be the multiplier associated with constraint (4-48). The objective function Z is augmented with the constraints (4-3) - (4-6), and (4-48) to form the following Lagrangian:

$$\begin{aligned}\min_{e_a^*(s)} Z^* = \sum_{\forall a} \int_0^T &\left\{ \begin{aligned} &C_a(s)e_a(s) + \lambda_a(s)\left\{[e_a(s) - g_a(s)] - \frac{dx_a(s)}{ds}\right\} + \mu_a(s)\left\{\frac{dE_a(s)}{ds} - e_a(s)\right\} \\ &- \rho_a(s)e_a(s) - \varpi_a(s)[E_a(s) - Q_a(s - s_0)] \end{aligned} \right\} ds, \\ &+ \nu_{od}\left(J_{od} - \sum_{\forall a} E_a(T)\right) \end{aligned} \tag{4-49}$$

which can be transformed using integration by parts as in Proposition 4.1 to

$$\begin{aligned} Z^* = &\sum_{\forall a} \left[-\lambda_a(T)x_a(T) + \mu_a(T)E_a(T)\right] + \nu_{od}\left(J_{od} - \sum_{\forall a} E_a(T)\right) \\ &+ \sum_{\forall a} \int_0^T \left[H_a(s) + \frac{d\lambda_a(s)}{ds}x_a(s) - \frac{d\mu_a(s)}{ds}E_a(s)\right] ds \end{aligned}, \tag{4-50}$$

in which the Hamiltonian function in this case is:

$$\begin{aligned} H_a(s) = &C_a(s)e_a(s) + \lambda_a(s)\left[e_a(s) - g_a(s)\right] - \mu_a(s)e_a(s) \\ &- \rho_a(s)e_a(s) \ - \varpi_a(s)\left[E_a(s) - Q_a(s - s_0)\right] \end{aligned} \tag{4-51}$$

The variation δZ^* of Z^* with respect to its variables is derived as

$$
\begin{aligned}
\delta Z^* = &\sum_{\forall a}\left[-\lambda_a(T)\delta x_a(T)+\mu_a(T)\delta E_a(T)\right]+\sum_{\forall a}\int_0^T\left(\frac{\partial H_a}{\partial u_s}\right)\delta e_a(s)ds \\
&+\sum_{\forall a}\int_0^T\left[\left(\frac{\partial H_a}{\partial x_a}+\frac{d\lambda_a(s)}{ds}\right)\delta x_a(s)\right]ds-\sum_{\forall a}\int_0^T\left(\frac{\partial H_a}{\partial E_a}+\frac{d\mu_a(s)}{ds}\right)\delta E_a(s)ds \\
&-\nu_{od}\sum_{\forall a}\delta E_a(T)
\end{aligned}
$$
(4-52)

The set of the optimality conditions for the dynamic optimization problem is derived as:

$$\frac{\partial H_a}{\partial u_s}=C_a(s)+\Psi_a(s)+\lambda_a(s)-\mu_a(s)-\rho_a(s)=0 \quad ,\forall a,\forall s; \tag{4-53}$$

$$\frac{\partial H_a}{\partial x_a}+\frac{d\lambda_a(s)}{ds}=\left(1+f'[\tau_a(s)]\right)\frac{e_a(s)}{Q_a}+\frac{d\lambda_a(s)}{ds}=0 \quad ,\forall a,\forall s; \tag{4-54}$$

$$\lambda_a(T)=0 \quad ,\forall a; \tag{4-55}$$

$$\frac{d\mu_a(s)}{ds}-\varpi_a(s)=0 \quad ,\forall a,\forall s; \tag{4-56}$$

$$\mu_a(T)-\nu_{od}=0 \quad ,\forall a. \tag{4-57}$$

We also have the following two sets of Karush-Kuhn-Tucker (KKT) conditions hold for the non-negativity constraint on the inflow:

$$e_a(s)\ge 0 \quad ,\forall a,\forall s; \tag{4-58}$$

$$e_a(s)\rho_a(s)=0 \quad ,\forall a,\forall s; \tag{4-59}$$

$$\rho_a(s)\ge 0 \quad ,\forall a,\forall s. \tag{4-60}$$

Combining (4-53) – (4-60), we have the following conditions on inflow to be satisfied at optimality

$$e_a(s)\left\{\begin{matrix} >0 \Rightarrow C_a(s)+\Psi_a(s)+\lambda_a(s)-\mu_a(s)= \\ =0 \Rightarrow C_a(s)+\Psi_a(s)+\lambda_a(s)-\mu_a(s)\geq \end{matrix}\right\}0 \quad ,\forall a,\forall s\in[0,T]. \tag{4-61}$$

Due to the inequality constraint (4-48), we have another set of KKT conditions on the inflow:

$$E_a(s)-Q_a(s-s_0)\geq 0 \qquad ,\forall a,\forall s; \tag{4-62}$$

$$\varpi_a(s)\left[E_a(s)-Q_a(s-s_0)\right]=0 \qquad ,\forall a,\forall s; \tag{4-63}$$

$$\varpi_a(s)\geq 0 \qquad ,\forall a,\forall s. \tag{4-64}$$

Condition (4-56) gives

$$\varpi_a(s)=-\frac{d\mu_a(s)}{ds} \quad ,\forall a,\forall s. \tag{4-65}$$

Substitute $\varpi_a(s)$ into (4-62) – (4-64), after rearranging, gives

$$E_a(s)\left\{\begin{matrix} >Q_a(s-s_0)\Rightarrow -\omega_a(s)=\dfrac{d\mu_a(s)}{ds}= \\ =Q_a(s-s_0)\Rightarrow -\omega_a(s)=\dfrac{d\mu_a(s)}{ds}\leq \end{matrix}\right\}0 \quad ,\forall a,\forall s\in[0,T]. \tag{4-66}$$

Consider the case when $e_a(s)>0$, we have

$$\mu_a(s)=C_a(s)+\Psi_a(s)+\lambda_a(s) \quad ,\forall a,\forall s\in[0,T], \tag{4-67}$$

and hence

$$\frac{d\mu_a(s)}{ds} = \frac{d}{ds}\left[C_a(s) + \Psi_a(s) + \lambda_a(s)\right] = h'(s) - 1 + \frac{d\Psi_a(s)}{ds} \ ,\forall a, \forall s \in [0,T]. \tag{4-68}$$

In addition, for the deterministic queuing model in the congested state, the rate of change of the sensitivity Ψ_a of the value of the objective function with respect to time is

$$\frac{d\Psi_a(s)}{ds} = -\left(1 + f'[\tau_a(s)]\right)\frac{e_a(s)}{Q_a} \ ,\forall a, \forall s. \tag{4-69}$$

Consequently, we have

$$\frac{d\mu_a(s)}{ds} = h'(s) - 1 - \left(1 + f'[\tau_a(s)]\right)\frac{e_a(s)}{Q_a} \ ,\forall a, \forall s. \tag{4-70}$$

Because $h'(\cdot) < 1$ and $f'(\cdot) > -1$ as discussed in Section 3.4, we have

$$\frac{d\mu_a(s)}{ds} < 0 \Rightarrow \varpi_a(s) > 0 \ ,\forall a, \forall s, \tag{4-71}$$

Together with (4-66), this implies

$$E_a(s) = Q_a(s - s_0) \qquad ,\forall a, \forall s. \tag{4-72}$$

Differentiating both sides with respect to time *s* gives

$$e_a(s) = Q_a \qquad ,\forall a, \forall s. \tag{4-73}$$

which is the solution in both congested and uncongested cases, and hence the solution for dynamic system optimal assignment for the deterministic queuing model. □

Discussion

In case of the deterministic queuing model, congestion is eliminated completely in dynamic system optimal assignment. The consequence of this is that the travel times along each link are constant at the corresponding link free flow travel time all times of departure. One way to decentralize this dynamic system optimal assignment is to impose a dynamic toll to each link in the system. This system optimizing toll should be behaviourally equivalent to the delays that would be incurred in traffic queues under the no-toll dynamic user equilibrium (see Heydecker and Addison, 2005). Such toll can be determined as

$$\beta_a^D(s) = \nu_{od} - h(s) - \phi_a - f[s + \phi_a], \tag{4-74}$$

in which ν_{od} is the cost at which travel takes place at dynamic user equilibrium. The advantage of the toll being incurred as a charge rather than as delay is that the former can be used to communal advantage whereas the latter is a dead loss. More discussion on dynamic tolling strategies is given in Section 4.6.

4.3 COSTATE VARIABLES, DYNAMIC EXTERNALITIES, AND SENSITIVITY ANALYSIS OF TRAVEL COST

Analysing and solving dynamic system optimal assignment requires understanding and determining the costate variables $\lambda_a(s)$ and $\gamma_a(s)$, and the dynamic externality $\Psi_a(s)$. These additional cost components are discussed in detail in this section.

4.3.1. Costate variables

Following Dorfman (1969), Bryson and Ho (1975), and Kamien and Schwartz (1991), the costate variables $\lambda_a(s)$ and $\gamma_a(s)$ in this optimal control formulation represents the sensitivity of the value of the objective function Z with respect to the changes in the state variables $x_a(s)$ and $g_a(s)$ in the corresponding constraints at the associated time s. The costate variable $\lambda_a(s)$, which is given by Equation (4-31), represents the total change in the value of the total system travel cost with respect to a unit change in the link traffic volume

$x_a(s)$ at time s. Likewise, the costate $\gamma_a(s) = \lambda_a[\tau_a(s)]$ represents the magnitude of the change in the total system travel cost with respect to a unit change in the link outflow $g_a(s)$ at time s. The associated minus sign represents that an increase (decrease) in outflow induces a decrease (increase) in the total system cost.

The sum of $\lambda_a(s)$ and $-\gamma_a(s)$ is calculated as

$$\lambda_a(s) - \gamma_a(s) = \lambda_a(s) - \lambda_a[\tau_a(s)] = \frac{1}{Q_a} \int_{t=s}^{\tau_a(t)} (1 + f'[\tau_a(t)]) e_a(t) dt \qquad , \forall a, \forall s, \qquad (4\text{-}75)$$

which represents the *net* change in the total system travel cost with respect to a unit change in the traffic volume that enters the link at time s and exits at time $\tau_a(s)$. Hence, the value of $\lambda_a(s) - \gamma_a(s)$ represents the portion of total system cost which is due to the traffic which stays in the system between times s and $\tau_a(s)$. This quantity can be interpreted as the *external cost* which is charged to the travellers who enter the system at time s for their presence between times s and $\tau_a(s)$.

4.3.2. Dynamic externalities

The notation $\Psi_a(s)$ in the cost components in condition (4-7), where $\Psi_a(s) = \int_0^T \left[\left. \frac{\partial C_a}{\partial u_s} \right|_t e_a(t) \right] dt$, represents *dynamic externality.* Dynamic externality refers to the additional travel cost imposed by an additional amount of traffic, u_s, within an interval s to other existing traffic on the link a.

To determine this externality $\Psi_a(s)$, we need to calculate the sensitivity $\left. \frac{\partial C_a}{\partial u_s} \right|_t$ of the travel cost C_a over time t with respect to the addition amount of traffic u_s. The derivation and the calculation of the derivative $\left. \frac{\partial C_a}{\partial u_s} \right|_t$ are discussed in detail in Section 4.3.3.

4.3.3. Sensitivity of travel cost

To determine $\left.\frac{\partial C_a}{\partial u_s}\right|_t$, we can derive it by differentiating the total travel cost $C_a(s)$ with respect to the inflow perturbation u_s as

$$\left.\frac{\partial C_a}{\partial u_s}\right|_t = \left(1 + f'[\tau_a(t)]\right) \left.\frac{\partial \tau_a}{\partial u_s}\right|_t \qquad ,\forall a, \forall t \qquad . \tag{4-76}$$

The derivative $\left.\frac{\partial C_a}{\partial u_s}\right|_t$ is now expressed in terms of the sensitivity $\left.\frac{\partial \tau_a}{\partial u_s}\right|_t$ of travel time with respect to perturbations in link traffic inflow. The derivation of this derivative $\left.\frac{\partial \tau_a}{\partial u_s}\right|_t$ is given in the Proposition 4.3.

Proposition 4.3: Suppose there is a change of u_s in the link inflow rate at a particular time *s*, the sensitivity of the time of exit at time *t* with respect to this perturbation is

$$\left.\frac{\partial \tau_a}{\partial u_s}\right|_t = \frac{d\tau_a}{dx_a}\left\{ \int_{\kappa=\sigma_a(t)}^{t} \frac{de_a(\kappa)}{du_s} d\kappa + g_a(t) \left.\frac{\partial \tau_a}{\partial u_s}\right|_{\sigma_a(t)} \right\} \qquad ,\forall a, \forall t, \tag{4-77}$$

in which $\sigma_a(t)$ is the time of entry to the link that leads to exit at time *t*. Indeed, $\sigma_a(\cdot)$ is the inverse function of $\tau_a(\cdot)$.

Proof:

The traffic volume on the travel link, $x_a(t)$, at time *t* can be expressed as

$$x_a(t) = E_a(t) - G_a(t) = E_a(t) - E_a[\sigma_a(t)] = \int_{\kappa=\sigma_a(t)}^{t} e_a(\kappa) d\kappa . \tag{4-78}$$

Suppose that there is a small change u_s induced in the profile of inflow at time *s*, the associated change in the value of the function of the time of exit at time *t* can be deduced as

$$
\begin{aligned}
\left.\frac{\partial \tau_a}{\partial u_s}\right|_t &= \frac{d\tau_a}{dx_a}\frac{\partial x_a(t)}{\partial u_s} \\
&= \frac{d\tau_a}{dx_a}\frac{\partial}{\partial u_s}\left(\int_{\kappa=\sigma_a(t)}^{t} e_a(\kappa)d\kappa \right) \\
&= \frac{d\tau_a}{dx_a}\left\{ \int_{\kappa=\sigma_a(t)}^{t} \frac{\partial e_a(\kappa)}{\partial u_s}d\kappa - \frac{\partial \sigma_a(t)}{\partial u_s}e_a\left[\sigma_a(t)\right] \right\}
\end{aligned}
\quad . \tag{4-79}
$$

The first term in the parentheses can be calculated directly.

To calculate the second term in (4-79), we first apply the definitional relationship,

$$\tau_a[\sigma_a(t)] = t \, . \tag{4-80}$$

Differentiating the left hand side with respect to u_s and by using chain rule, the left-hand-side of (4-80) can be written as

$$\left.\frac{d\tau_a}{du_s}\right|_{\sigma_a(t)} = \left.\frac{\partial \tau_a}{\partial u_s}\right|_{\sigma_a(t)} + \frac{\partial \tau_a[\sigma_a(t)]}{\partial \sigma_a(t)}\frac{\partial \sigma_a(t)}{\partial u_s} \, . \tag{4-81}$$

Similarly, differentiating the right hand side with respect to u_s, it gives

$$\left.\frac{d\tau_a}{du_s}\right|_{\sigma_a(t)} = \frac{dt}{du_s} = 0 \, , \tag{4-82}$$

Hence, combining (4-81) and (4-82), it can be deduced that

$$\left.\frac{\partial\tau_a}{\partial u_s}\right|_{\sigma_a(t)}+\frac{\partial\tau_a[\sigma_a(t)]}{\partial\sigma_a(t)}\frac{\partial\sigma_a(t)}{\partial u_s}=0\,. \tag{4-83}$$

Furthermore, because $\sigma_a(\cdot)$ is an inverse function of $\tau_a(\cdot)$, it follows that

$$\frac{\partial\tau_a[\sigma_a(t)]}{\partial\sigma_a(t)}=\left(\frac{\partial\sigma_a(t)}{\partial t}\right)^{-1}. \tag{4-84}$$

Therefore, combining (4-83) and (4-84), and after rearranging terms, it gives

$$\frac{\partial\sigma_a(t)}{\partial u_s}=-\left(\frac{\partial\tau_a[\sigma_a(t)]}{\partial\sigma_a(t)}\right)^{-1}\left.\frac{\partial\tau_a}{\partial u_s}\right|_{\sigma_a(t)}=-\frac{d\sigma_a(t)}{dt}\left.\frac{\partial\tau_a}{\partial u_s}\right|_{\sigma_a(t)}. \tag{4-85}$$

Finally, substituting (4-85) into (4-79) gives

$$\begin{aligned}\left.\frac{\partial\tau_a}{\partial u_s}\right|_s&=\frac{d\tau_a}{dx_a}\left\{\int_{\kappa=\sigma_a(t)}^{t}\frac{de_a(\kappa)}{du_s}d\kappa-\frac{d\sigma_a(t)}{du_s}e_a[\sigma_a(t)]\right\}\\&=\frac{d\tau_a}{dx_a}\left\{\int_{\kappa=\sigma_a(t)}^{t}\frac{de_a(\kappa)}{du_s}d\kappa+e_a[\sigma_a(t)]\frac{d\sigma_a(t)}{dt}\left.\frac{\partial\tau_a}{\partial u_s}\right|_{\sigma_a(t)}\right\}\\&=\frac{d\tau_a}{dx_a}\left\{\int_{\kappa=\sigma_a(t)}^{t}\frac{de_a(\kappa)}{du_s}d\kappa+g_a(t)\left.\frac{\partial\tau_a}{\partial u_s}\right|_{\sigma_a(t)}\right\}\end{aligned} \tag{4-86}$$

as is to be shown. □

The expression in (4-77) is derived from flow propagation mechanism and it is applicable for general travel time models. The derivative $\left.\frac{\partial\tau_a}{\partial u_s}\right|_t$ is expressed in terms of the dependence of the inflow profile $e_a(\kappa)$ in which κ lies between $\sigma_a(t)$ and *t*, the outflow $g_a(t)$ at time *t*, and the value of the derivative at time $\sigma_a(t)$. The derivative $\frac{d\tau_a}{dx_a}$ is the change in the value of $\tau_a(\cdot)$ with respect to the change in the value of the state $x_a(\cdot)$ at the same time. If we consider the linear travel time models: the deterministic queue, divided linear models, and whole-link traffic model, then it gives,

$$\frac{d\tau_a}{dx_a} = \frac{d}{dx_a}\left(t + \phi_a + \frac{x_a}{Q_a} \right) = \frac{1}{Q_a} \qquad ,\forall a . \tag{4-87}$$

In the special case of the deterministic queuing model, the derivative $\left.\frac{\partial \tau_a}{\partial u_s}\right|_t$ is zero when the link is uncongested (i.e. link inflow is less than or equal to the capacity, and amount of traffic in queue is zero) since the link travel time is equal to free flow travel time which is a constant. When the travel link is congested (i.e. amount of traffic queuing greater than zero), $\left.\frac{\partial \tau_a}{\partial u_s}\right|_t$ will be positive and the link traffic will be discharged at the link capacity Q_a. Substituting $g_a(t) = Q_a$ for all times t into (4-77) reduces the equation to

$$\begin{aligned} \left.\frac{\partial \tau_a}{\partial u_s}\right|_t &= \frac{d\tau_a}{dx_a} \int_{\kappa=0}^{t} \frac{de_a(\kappa)}{du_s} d\kappa \\ &= \frac{1}{Q_a} \int_{\kappa=0}^{t} \frac{de_a(\kappa)}{du_s} d\kappa \end{aligned} . \tag{4-88}$$

Equation (4-88) shows that in the particular case of the deterministic queuing model, the derivative $\left.\frac{\partial \tau_a}{\partial u_s}\right|_t$ takes the value of zero for all times *t* before the time of perturbation *s*, and $\left.\frac{\partial \tau_a}{\partial u_s}\right|_t$ equals to $\frac{1}{Q_a}$ for all times *t* after the time *s* of perturbation while the queue persists. This agrees with the previous analyses on the sensitivity of the deterministic queuing model (see for example, Ghali and Smith, 1995; Kuwahara, 2001). However, the sensitivity analysis developed in this thesis allows for other mechanisms of delay and flow propagation, and hence is more general so that it can be applied to other traffic models.

Summary

After determining $\left.\dfrac{\partial \tau_a}{\partial u_s}\right|_t$ as described in Section 4.3.3, the dynamic externality $\Psi_a(s)$ in Section 4.3.2 can be calculated directly. Finally, referring to the necessary condition (4-7) of dynamic system optimum, each traveller in the system who enters the travel link *a* at a time of entry *s* is expected to pay an amount of *toll* equal to $[\Psi_a(s)+\lambda_a(s)-\gamma_a(s)]$ in order for the transport system to operate optimally. This analytical result shows that there is a substantial difference between the traditional analysis on static transport system (for example Sheffi, 1985) and the current analysis of dynamic transport system. To optimize a dynamic transport system, in addition to paying for his/her own externality $\Psi_a(s)$ imposed on others, travellers are also required to be responsible for a toll charge $[\lambda_a(s)-\gamma_a(s)]$, which is charged by the external system manager on the travellers for using the transport system during times *s* and $\tau_a(s)$ (see discussion in Section 4.3.1).

4.4 SOLUTION ALGORITHMS

This section illustrates the algorithms that transform the analysis presented in Sections 4.2 and 4.3 into numerical solutions. Section 4.4.1 first presents an algorithm to calculate the derivatives $\left.\dfrac{\partial \tau_a}{\partial u_s}\right|_t$ derived in Proposition 4.3 and hence the externality $\Psi_a(s)$. Then, Section 4.4.2 introduces an algorithm to solve dynamic system optimal assignment. As dynamic user equilibrium assignment described in Section 3.6, we adopt a modified second-order Newton method to solve dynamic system optimal assignment.

4.4.1 Calculate the sensitivity of travel time

The section presents the solution algorithm which transforms the sensitivity analysis given in Proposition 4.3 into numerical derivatives. The derivatives calculated can be used for solving dynamic system optimal assignment.

Step 1: Initialisation for calculating the derivatives of link exit time

1.1 Set the link index $a := 1$;

1.2 Set the time index $k := 0$, to represent the time interval when the inflow is perturbed;

1.3 Set the time index $\omega := 0$ to refer the time at which we consider the change in exit time due to the perturbation in inflow at time interval *k*;

1.4: Calculate the derivatives of link exit time:

If $\omega < k$, then $\left.\dfrac{d\tau_a}{du_k}\right|_{\omega} := 0$;

else if $k \le \omega \le \lceil \tau_a(k) \rceil$, then $\left.\dfrac{d\tau_a}{du_k}\right|_{\omega} := \dfrac{\Delta s}{Q_a}$;

else $\left.\dfrac{\partial \tau_a}{\partial u_k}\right|_{\omega} := \dfrac{g_a(\omega)}{Q_a}\left.\dfrac{\partial \tau_a}{\partial u_k}\right|_{\sigma_a(\omega)}$ ***.

1.5 If $\omega = K$, then go to step 1.6; otherwise $\omega := \omega + 1$ and go to step 1.4;

1.6 If $k = K$, then go to step 1.7; otherwise *k*:= *k* + 1 and go to step 1.3;

1.7 If $a = A$, then go to step 2; otherwise *a*:= *a* + 1 and go to step 1.2.

<u>Step 2: Calculate the derivatives of total travel cost function</u>

2.1 Set the link index $a := 1$;

2.2 Set the time index $k := 0$;

2.3 Set the time index $\omega := 0$;

2.4 Calculate $\left.\dfrac{dC_a}{du_k}\right|_{\omega} = \left(1 + f'[\tau_a(\omega)]\right)\left.\dfrac{d\tau_a}{du_k}\right|_{\omega}$;

2.5 If $\omega = K$, then go to step 2.6; otherwise $\omega := \omega + 1$ and go to step 2.4;

2.6 If $k = K$, then go to step 2.7; otherwise *k*:= *k* + 1 and go to step 2.3;

2.7 If $a = A$, then go to step 3; otherwise *a*:= *a* + 1 and go to step 2.2.

<u>Step 3: Calculate the externality</u>

3.1 Set the link index $a := 1$;

3.2 Set the time index $k := 0$;

3.3 Initialise $\Psi_a(k) := 0$;

*** The time $\sigma_a(\omega)$ is not necessarily integral. We adopt a linear interpolation to approximate $\left.\dfrac{d\tau_a}{du_k}\right|_{\sigma_a(\omega)}$ as $\left.\dfrac{d\tau_a}{du_k}\right|_{\sigma_a(\omega)} \approx \left.\dfrac{d\tau_a}{du_k}\right|_{\lfloor \sigma_a(\omega) \rfloor} + \left(\left.\dfrac{d\tau_a}{du_k}\right|_{\lceil \sigma_a(\omega) \rceil} - \left.\dfrac{d\tau_a}{du_k}\right|_{\lfloor \sigma_a(\omega) \rfloor}\right)\left(\sigma_a(\omega) - \lfloor \sigma_a(\omega) \rfloor\right)$, where $\lceil \sigma_a(\omega) \rceil$ represents the smallest integer not smaller than $\sigma_a(\omega)$, and $\lfloor \sigma_a(\omega) \rfloor$ is the greatest integer not larger than $\sigma_a(\omega)$.

3.4 Set the time index $\omega := 0$;

3.5 Calculate $\Psi_a(k) = \Psi_a(k) + e_a(\omega) \left. \dfrac{dC_a}{du_k} \right|_{\omega}$;

3.6 If $\omega = K$, then go to step 3.7; otherwise $\omega := \omega + 1$ and go to step 3.5;

3.7 If $k = K$, then go to step 3.8; otherwise $k := k + 1$ and go to step 3.3;

3.8 If $a = A$, then stop; otherwise $a := a + 1$ and go to step 3.2.

4.4.2 Calculate dynamic system optimal assignment

The analysis in Sections 4.2 – 4.3 is considered in continuous time. In addition, except for the very few exceptional examples such as the deterministic queuing model, closed form solutions for dynamic system optimal are generally not available for most of the travel time models. In accordance with this, this section introduces a solution algorithm that transforms the formulation and analysis of dynamic system optimum in continuous time into numerical solutions in discrete time. The algorithm is structured as a combination of forward-backward dynamic programme: to be solved forward in the order of departure time interval for assignment flow profile as the case of dynamic user equilibrium discussed in Section 3.6; solved backward in time for the corresponding externality and response. The study period in continuous time, T, is discretized into K intervals each of length Δs. Following this, the instantaneous flow in continuous time formulation is represented as the flow $e(k)$ that is constant through the discrete time interval k: $[k\Delta s, (k+1)\Delta s)$. This flow is tested against the cost $C[(k+1)\Delta s)$ at the late end of the time interval. Within each departure time interval k, the assignment inflow is calculated by using Newton method, which converges with an order of convergence at least 2 (Luenberger, 1989, p202).

The algorithmic procedure is described as follows.

<u>Step 0: Initialisation</u>

0.1.Choose an initial equilibrium cost C^*_{od} for each origin-destination pair *od*;

0.2 Set the overall iteration counter $n := 1$;

0.3 Set $e_a(k) := 0$ for all links *a* and all time intervals *k*.

0.4 Set costates $\lambda_a(k) := 0$ for all links *a* and all time intervals *k*;

0.5 Set the time index $k := 0$;

0.6 Set the link index $a := 1$;

0.7 Set the inner iteration counter $n^i := 1$.

Step 1: Network loading

Find $\tau_a(k+1)$ by loading the travel link using the inflow $e_a(k)$ at the current iteration. The algorithm described in Section 2.4.2 is adopted.

Step 2: Calculate externality

Calculate the externality $\Psi_a(k)$ associated with each $e_a(k)$ by using the algorithm presented in Section 4.4.1.

Step 3: Determine the auxiliary inflow

3.1 Calculate

$$C_a(k+1) = h(k+1) + [\tau_a(k+1) - (k+1)] + f[\tau_a(k+1)] + \Psi_a(k+1) + \lambda_a(k) - \lambda_a[\tau_a(k)];$$

3.2 Calculate $\Omega(k) = \dfrac{C_a(k+1) - C_a(k)}{\Delta s}$ and $\Omega'(k) = \dfrac{\partial \Omega(k)}{\partial e_a(k)} = (1 + f'[\tau_a(k+1)])\dfrac{1}{Q_a}$;

3.3 Calculate the auxiliary inflow $d_a(k) = -\Omega_a(k) \big/ \Omega_a'(k)$;

3.4. If $a = A$, then go to step 3.5; otherwise $a := a + 1$ and go to step 0.7;

3.5. If $k = K$, then go to step 4; otherwise $k := k + 1$ and go to step 0.6.

Step 4: Determine step size for inflow

Search for θ, for all a and k, by golden section method such that $e_a(k) := \max\{[e_a(k) + \theta[d_a(k) - e_a(k)]], 0\}$ gives the minimum total travel cost.

Step 5: Calculate the associated costate variables

5.1 Set $\lambda_a(K) = 0$ for all links a;

5.2 Set the time index $k := K - 1$;

5.3 Set the link index $a := 1$;

5.4 Compute $\lambda_a(k) = \lambda_a(k+1) + (1 + f'[\tau_a(k)])\dfrac{e_a(k)}{Q_a}\Delta s$;

5.5 Calculate $\lambda_a[\tau_a(k)]$ from $\lambda_a(k)$ and $\tau_a(k)$ using linear interpolation as

$$\lambda_a[\tau_a(k)] \approx \lambda_a|_{\lfloor \tau_a(k) \rfloor} + \left(\lambda_a|_{\lceil \tau_a(k) \rceil} - \lambda_a|_{\lfloor \tau_a(k) \rfloor}\right)\left(\tau_a(k) - \lfloor \tau_a(k) \rfloor\right);$$

5.6. If $a = A$, then go to step 5.7; otherwise a:= a + 1 and go to step 5.4;

5.7. If $k = 0$, then go to step 6; otherwise k:= k - 1 and go to step 5.3.

Step 6: Overall stopping criteria

6.1 Define $\xi = \frac{\sum_{\forall a}\sum_{\forall k} e_a(k)\left|C_a(k+1) - C_{od}^*\right|}{\sum_{\forall a}\sum_{\forall k} e_a(k)C_{od}^*}$ as a measure of disequilibrium, which is equal to zero at system optimum. If n is greater than the predefined maximum number of overall iterations or ξ is sufficiently small, i.e. $\xi \le \varepsilon$ where ε is a test value, then go to Step 6.2; otherwise set n:=n+1 and go to step 0.5;

6.2. Check if the total throughput $E_{od} = \sum_{\forall a}\sum_{\forall k} e_a(k)$ from the system is equal to the predefined total demand J_{od} for the *o-d* pair. If yes, then *terminate* the algorithm; otherwise update $C_{od}^* := C_{od}^* + \left[\frac{J_{od} - E_{od}}{\frac{dE_{od}}{dC_{od}^*}}\right]$, and go back to step 0.2. For networks with mutually distinct routes, following Heydecker (2002b), we can establish an expression for the derivative $\frac{\partial E_{od}}{\partial C_{od}^*} = \sum_{\forall a}\left(\frac{\left[h'(s_a^0) + f'[\tau_a(s_a^0)]\right] - \left[h'(s_a^1) + f'[\tau_a(s_a^1)]\right]}{\left[h'(s_a^0) + f'[\tau_a(s_a^0)]\right]\left[h'(s_a^1) + f'[\tau_a(s_a^1)]\right]}\right)Q_a$, where s_a^0 and s_a^1 respectively represent the first and last times that the link is used.

Discussion

As noted in Section 3.6, a crucial point in solving dynamic traffic assignments is to consider the time-varying variables at the appropriate time. When we calculate the costate variables in Step 3.1 in Section 4.4.2, the values of the costates are considered at the start of the time interval $k\Delta s$ instead of the end of the interval. This is because, contrasting with the travel time which is calculated forward in time, the costate variables are calculated backward in time.

Moreover, the auxiliary flows are calculated based on the traffic conditions at the current iteration up to time step *k*, the costate variables are calculated based on the traffic conditions at the previous iteration after time step *k*. As a result, they are not consistent, and we adopt a step size search (Step 4) as a heuristics to accommodate this.

Finally, recall that in the deterministic queuing model, there is a discontinuity in the derivatives of the state variable with respect to the inflow at $e_a(s) = Q_a$ when $x_a(s) = 0$. For all $x_a(s) = 0$ and $e_a(s) \le Q_a$, the travel link is congestion-free and traffic is flowing at the free flow travel time which is independent of the inflow. It gives the auxiliary inflow $d_a(k) = -\Omega_a(k) \big/ \Omega_a^{'}(k)$ a value of zero while there is no queue, no matter what the current traffic flow is, provided that the inflow does not exceed the system capacity. The consequence of this is that the present solution algorithm cannot achieve an optimal solution or even an improved solution from the initial solution in uncongested condition when the deterministic queuing model is adopted.

4.5 EXAMPLE CALCULATIONS

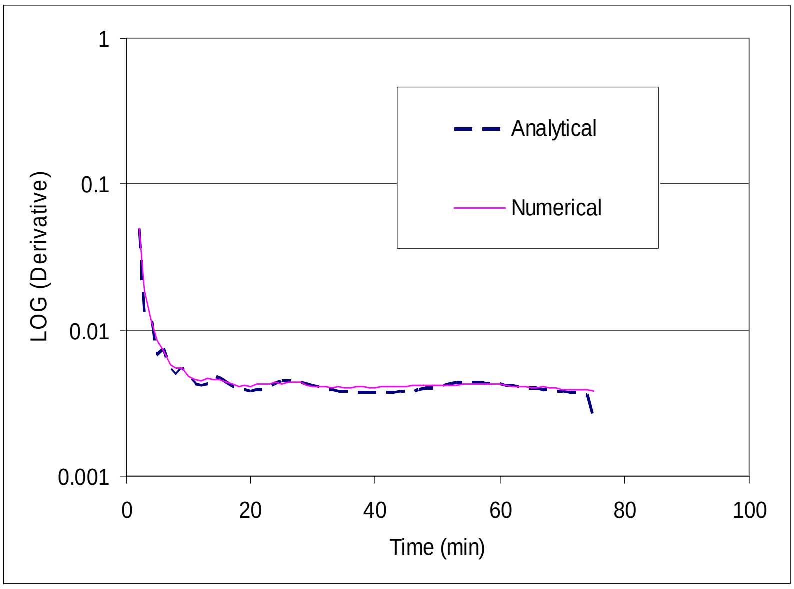
The section presents various example calculations to demonstrate the performance of the solution algorithms and the characteristics of the numerical results of dynamic system optimal assignment.

4.5.1 Sensitivity of travel time

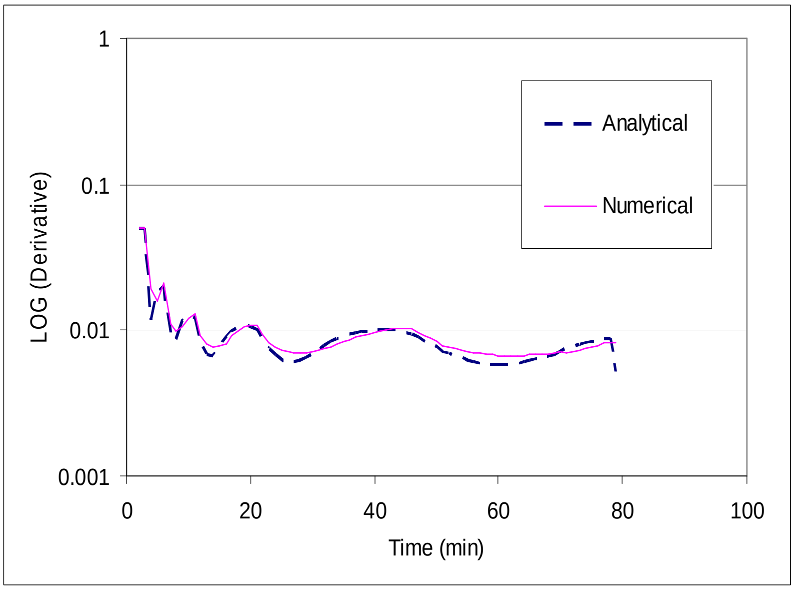
The critical step in determining the externality $\Psi_a(s)$ is calculating the derivative of link exit time $\left.\frac{\partial \tau_a}{\partial u_s}\right|_t$ with respect to perturbation in inflow. Hence, this section tests the numerical accuracy of this derivative as calculated according to the method presented in Proposition 4.3. We consider the single travel link and the parabolic inflow profile that was introduced in Section 2.6. To investigate the accuracy of the sensitivity analysis of the travel time models, we perturb the parabolic inflow profile at time 1, and the associated variations in travel time are plotted in Figure 4.1. The *analytical* variations are calculated according to Equation (4-

77), while the *numerical* variations are determined by using direct numerical finite difference method. Both of these are plotted in *logarithm scale* in Figure 4.1 for comparison. To calculate the finite difference, one extra unit of inflow is added at time 1, while the inflow profile remains unchanged at other times. The numerical variations in travel times are then calculated by subtracting the link travel time loaded by the original inflow profile from that loaded by the perturbed inflow profile. It is noted that we do not include the deterministic queuing model here. For the deterministic queue model, there is no variation in travel time because that traffic model gives a constant estimation of link travel time whenever the inflow rate is less than the link capacity and the volume of traffic in queue is equal to zero.

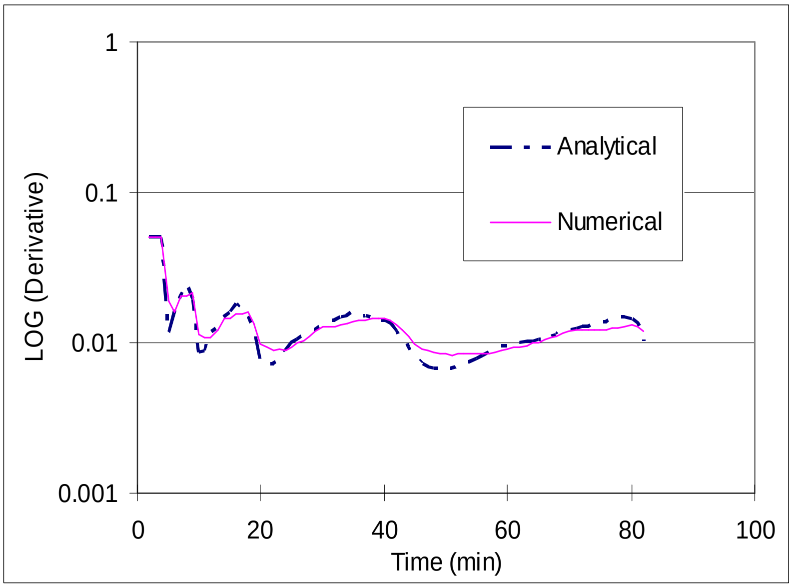
The result shows that the analytical variations given by Equation (4-77) represent the true numerical variations in travel time reasonably well for all travel time models. Both numerical and analytical variations drop to zero at the time when all traffic is cleared from the link. It is also observed that the variations of travel time oscillate for travel time models with longer congestible part, and that the analytical and numerical estimates agree closely on this.



a) Divided linear model ($\alpha_a = \Delta s$)



b) Divided linear model ($\alpha_a = 2\Delta s$)



c) Whole-link traffic model

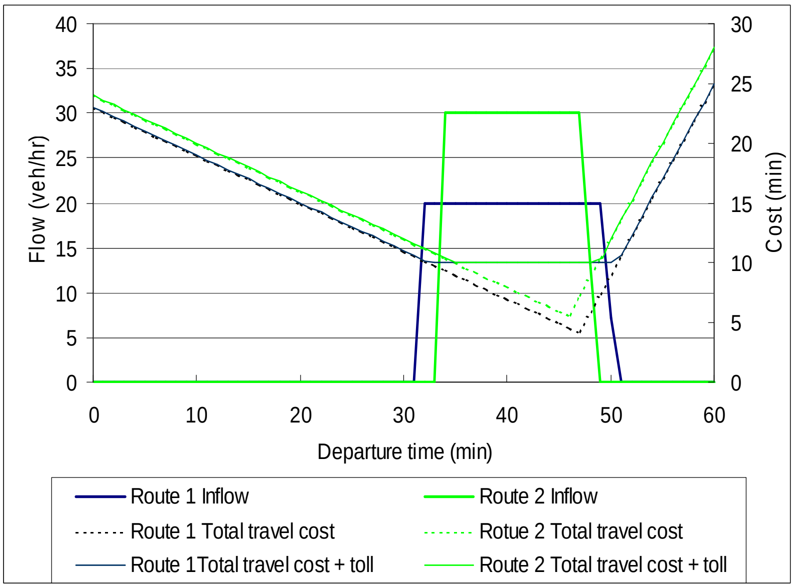
Figure 4.1 Sensitivity of travel times

4.5.2 Dynamic system optimal traffic assignment

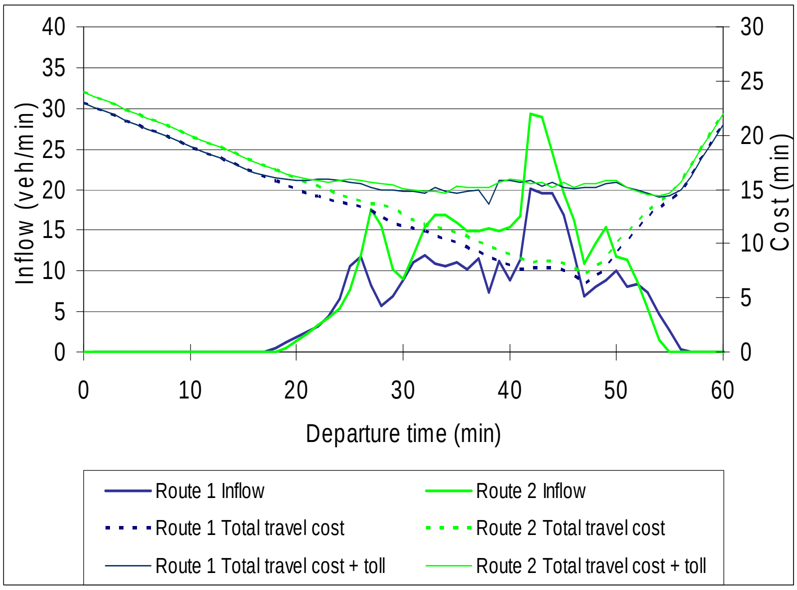
This section shows the dynamic system optimal assignment results. We use the two-link network as shown in Figure 3.1 and in Section 3.7. All assignments are computed numerically by using the solution algorithm described in Section 4.4.2 except for the assignments with the deterministic queuing model. As discussed in Section 4.4.2, the present solution algorithm is not suitable for solving dynamic system optimal assignment with the deterministic queuing model. Solving numerically the dynamic system optimal assignment with such traffic model requires some special heuristics, for example the route-flow swapping technique adopted in Huang and Lam (2002). However, the main focus of this thesis is on the properties of the traffic models and the associated assignments and their implications on dynamic traffic management, rather than on the numerical solution strategies. As a result, this section only presents the analytical solution of the assignments with the deterministic queuing model and compares it with the numerical solutions the assignments of traffic models of other kinds.

Figure 4.2 shows dynamic system optimal assignments using different travel time models. In the figure, the *total travel cost*, i.e. $C_a(s)$, refers to the sum of the cost associated with travel time and the costs associated with the departure and arrival times for a traveller who departs at time s. The *total travel cost + toll* means the total travel cost $C_a(s)$ plus the toll, i.e. $\left[\Psi_a(s)+\lambda_a(s)-\gamma_a(s)\right]$, that the traveller who departs at time s is going to pay. According the optimality condition derived in Proposition 4.1 and Proposition 4.2, the sum of the total travel cost and the toll (i.e. externality) at dynamic system optimum should be constant for all links and all departure times in use. It is clear from the assignment results that the quality of such equilibration drops as the congestion portion considered in the traffic model increases. It implies that solving dynamic system optimal assignment is getting more difficult. The reason of this is that the larger congestible portion implies more interaction of traffic dynamics to be considered in the calculation process and hence solving the dynamic optimization problem becomes more difficult. Further discussion on the quality of dynamic system optimal assignment is given in Section 4.5.3.

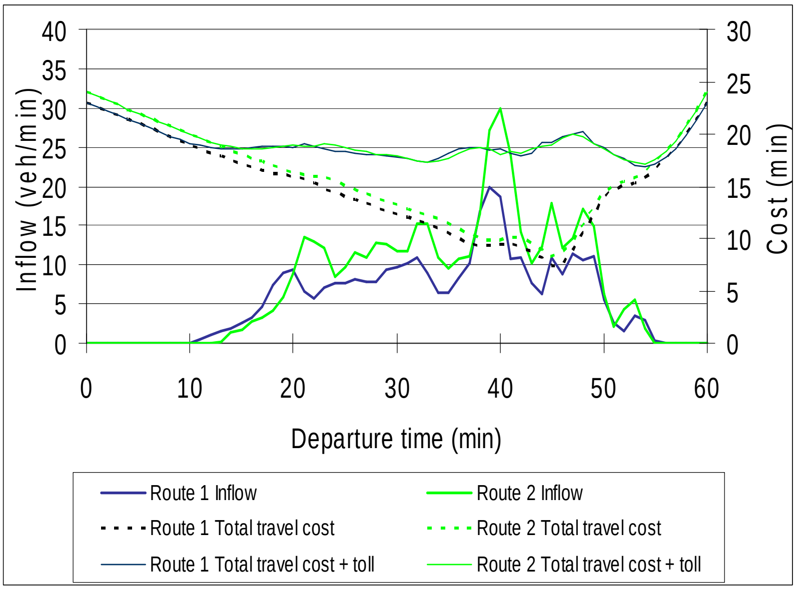
The assignment results also show that the durations of assignments are lengthened and hence the profiles of inflows are spread out at dynamic system optimum as suggested by Chu (1995) in order to reduce the intensity of the congestion. We further observe that the durations of assignments are longer and hence the inflow profile is more spread for travel time models considering a larger congestible portion. The exception is the deterministic queuing model which gives dynamic system optimal assignment with the same duration as its dynamic user equilibrium counterpart, within which the traffic congestion is completely eliminated (see also Arnott et al., 1993; 1998).



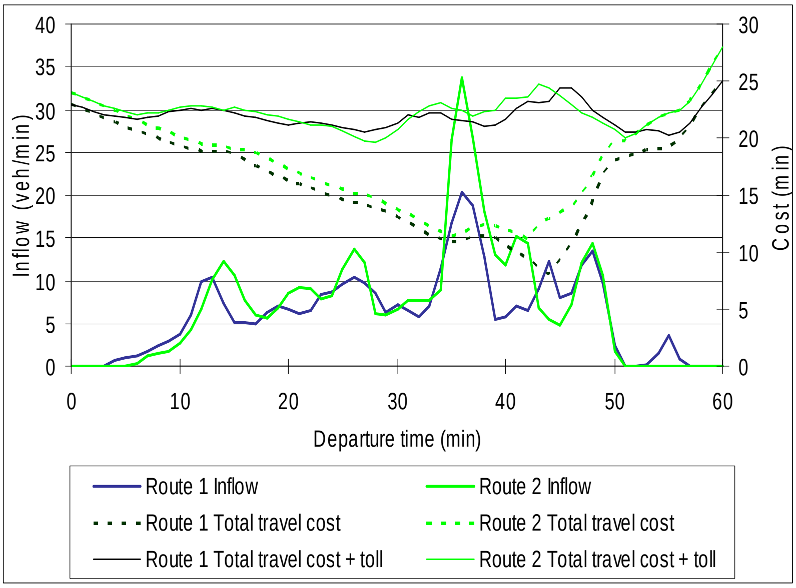
a) Deterministic queuing model



b) Divided linear model ($\alpha_a = \Delta s$)



c) Divided linear model ($\alpha_a = 2\Delta s$)



d) Whole-link traffic model

Figure 4.2 Dynamic system optimal assignments

The associated details for each assignment are also summarized in Table 4.1 for further illustration. In addition to a more spread inflow profile, the table shows that more traffic is assigned to link 2 in dynamic system optimum which can be interpreted as a result of the fact that link 2 has a higher capacity for discharging traffic.

Table 4.1 Summary of dynamic system optimal assignments

	Link 1 Start time (min)	 End time (min)	 Traffic volume (veh)
DDQ	32	49	367.20
Divided: $\alpha_a = \Delta s$	16	55	343.21
Divided: $\alpha_a = 2\Delta s$	10	55	340.50
Friesz	3	56	369.45
	Link 2 Start time (min)	 End time (min)	 Traffic volume (veh)
DDQ	34	47	432.80
Divided: $\alpha_a = \Delta s$	19	54	456.79
Divided: $\alpha_a = 2\Delta s$	13	54	459.50
Friesz	6	50	430.55

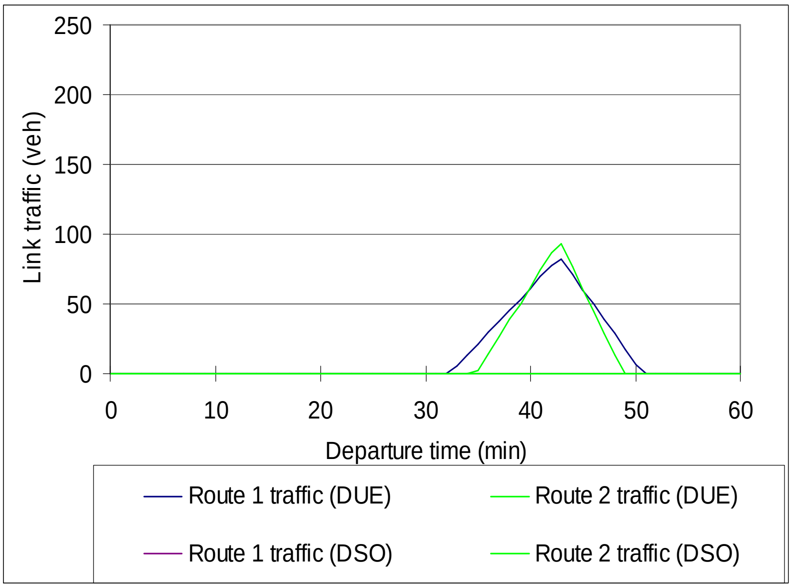
Table 4.2 compares the total system costs and individual costs under dynamic user equilibrium and dynamic system optimal assignments. In general, the total system costs drop when the system transforms from dynamic user equilibrium to dynamic system optimum. We note that this reduction in total system costs decreases as the portion of the congestible part considered on a travel link increases. Considering the two extremes, there can be an estimation of 50% reduction in total system cost when the deterministic queuing model is adopted, but a reduction of only 8.3% when the whole-link model is considered. Moreover, it is observed even if we only consider a small portion of congestible part on the travel link, the improvement in the total system cost in dynamic system optimum will drop significantly compared with the situation when the deterministic queuing model is adopted. As noted earlier and pointed out by Kimber and Hollis (1979), and by Mun (2002), the deterministic queuing model oversimplifies the traffic dynamics and underestimates the travel time before the travel link is saturated (i.e. $e_a(s) > Q_a$). This finding can be a reflection that the traditional analysis of dynamic network system management based on the deterministic queuing model overestimates the efficiency of dynamic system optimal assignment. Equivalently, the traditional analysis underestimates the efficiency of dynamic user equilibrium assignment.

Table 4.2 Comparison of costs under different assignments

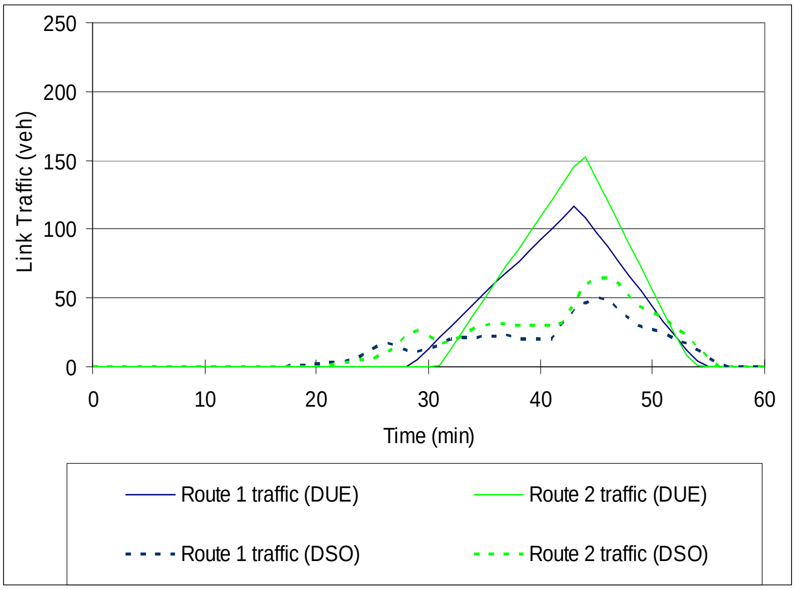
	Total system cost (veh-min)		
	DUE	DSO	Difference
DDQ	8,064.00	4,032.00	-50.00%
Divided: $\alpha_a = \Delta s$	9,310.80	8,076.64	-13.26%
Divided: $\alpha_a = 2\Delta s$	10,741.20	9,627.28	-10.37%
Friesz	12,465.20	11,433.60	-8.28%
	Individual cost (min)		
	DUE	DSO	Difference
DDQ	10.08	10.08	0.00%
Divided: $\alpha_a = \Delta s$	11.64	13.31	14.36%
Divided: $\alpha_a = 2\Delta s$	13.43	18.37	36.82%
Friesz	15.58	21.62	38.75%

The individual cost refers to the total cost (i.e. $[C_a(s) + \Psi_a(s) + \lambda_a(s) - \gamma_a(s)]$, in the unit of minute) that each individual traveller has to be responsible for when he/she makes the trip. We recall from Section 4.2 that this cost is identical for all travellers in dynamic system optimum. Although dynamic system optimal assignment reduces the total system cost of the whole system, the individual cost does increase from dynamic user equilibrium to dynamic system optimum for most travel time models adopted. It is understandable because each traveller has to pay an extra toll in addition to their cost of travel in dynamic system optimum for the good of the whole system. It could be an explanation of why road users would be against road pricing. The only exception is when the deterministic queuing model is adopted with which travellers are estimated to have the same individual cost in dynamic user equilibrium and dynamic system optimum. With the same individual cost incurred, the only difference is that the cost of congestion in dynamic user equilibrium is eliminated and replaced by the equivalent amount of toll which can be used to communal advantage rather than a dead loss (Heydecker and Addison, 2005).

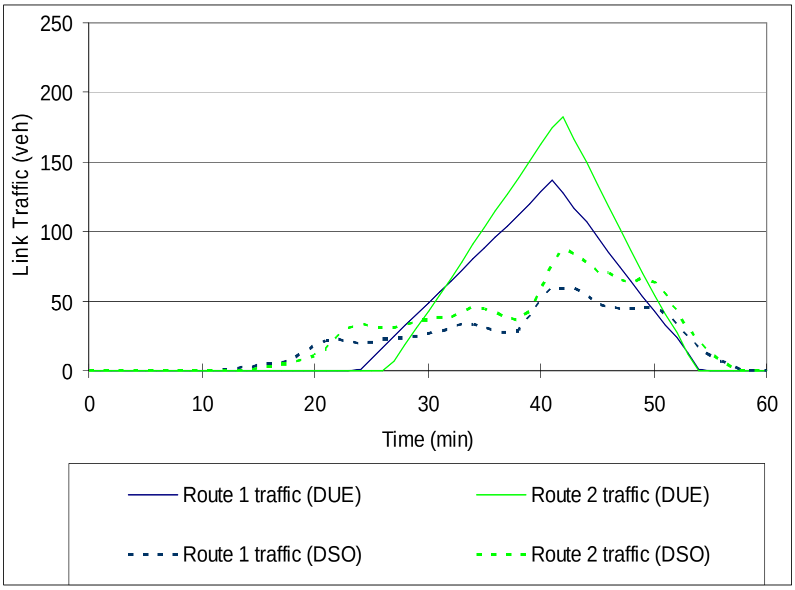
Figure 4.3 plots and compares the profiles of the state variable $x_a(s)$ (i.e. traffic volumes) on each link at dynamic user equilibrium and dynamic system optimum respectively. In the travel time models, the state variables represent the degree of congestion because the link travel time is taken as a monotonic non-decreasing function of it. Interestingly, yet importantly, the results show that dynamic system optimal assignment has to allow congestion for all travel time models except for the deterministic queuing model. According to most models, we can only manage the level of congestion but cannot eliminate it.



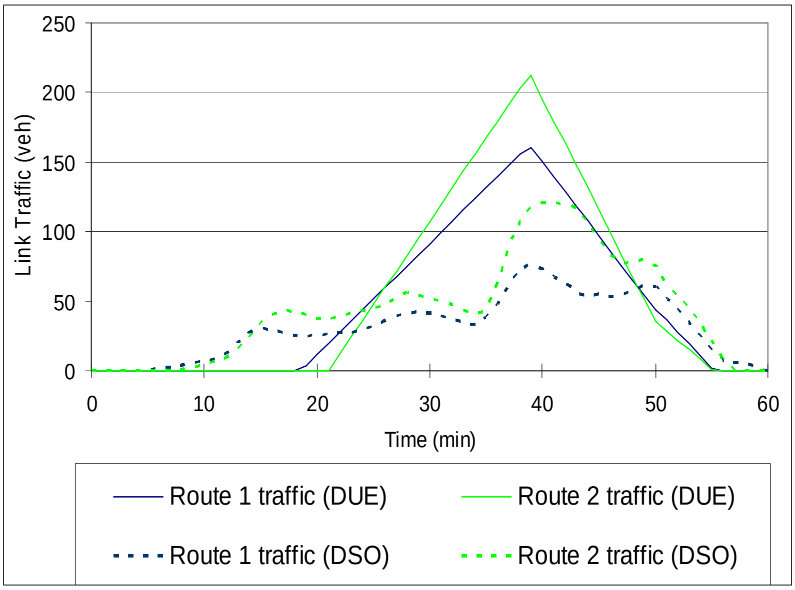
a) Deterministic queuing model



b) Divided linear model ($\alpha_a = \Delta s$)



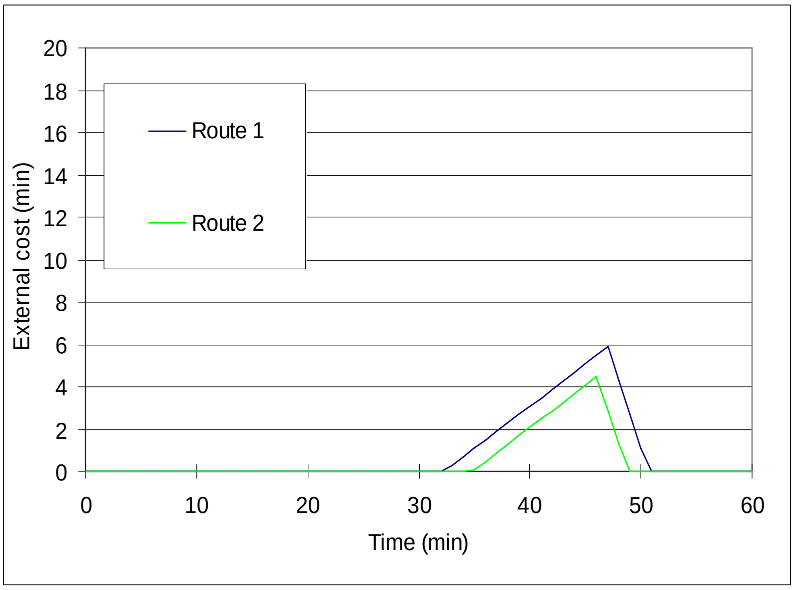
c) Divided linear model ($\alpha_a = 2\Delta s$)



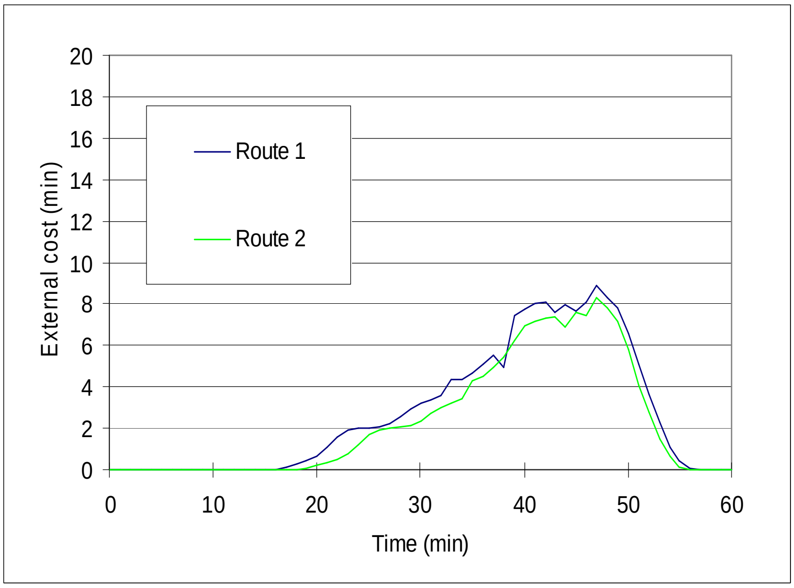
d) Whole-link traffic model

Figure 4.3 Link traffic volumes

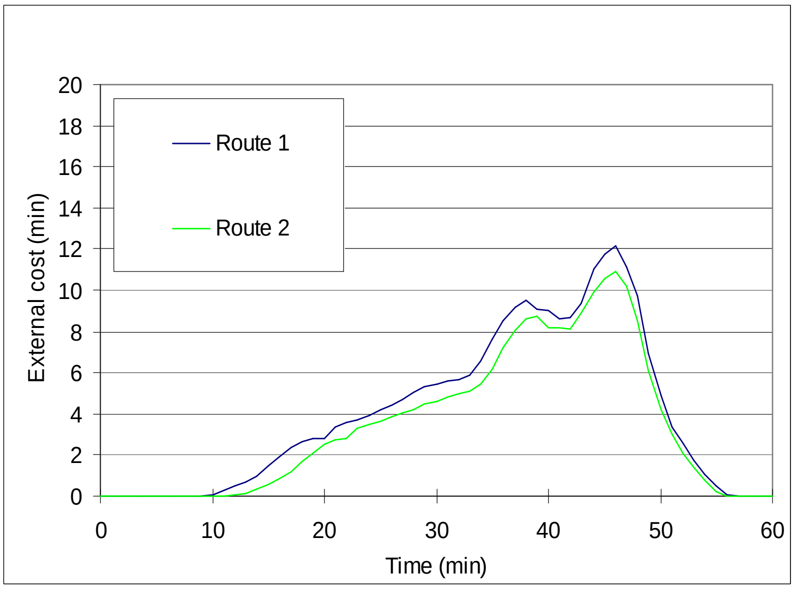
The dynamic tolls (i.e. $[\Psi_a(s)+\lambda_a(s)-\gamma_a(s)]$) which are to be imposed on the travellers are then calculated and plotted in Figure 4.4. The profiles of the tolls are different for different travel time models. In general, the dynamic tolls increase with time for travellers who arrive at the destination before the preferred arrival time, and decrease afterwards. For travel time models that consider a larger portion of congestible part, the charges start earlier and reach a higher maximum value. The reason is that with those travel time models, the traffic congestion is estimated to form earlier and then reach a higher maximum level. As a result, the corresponding dynamic toll has to be implemented earlier and reach a higher magnitude to manage the congestion.



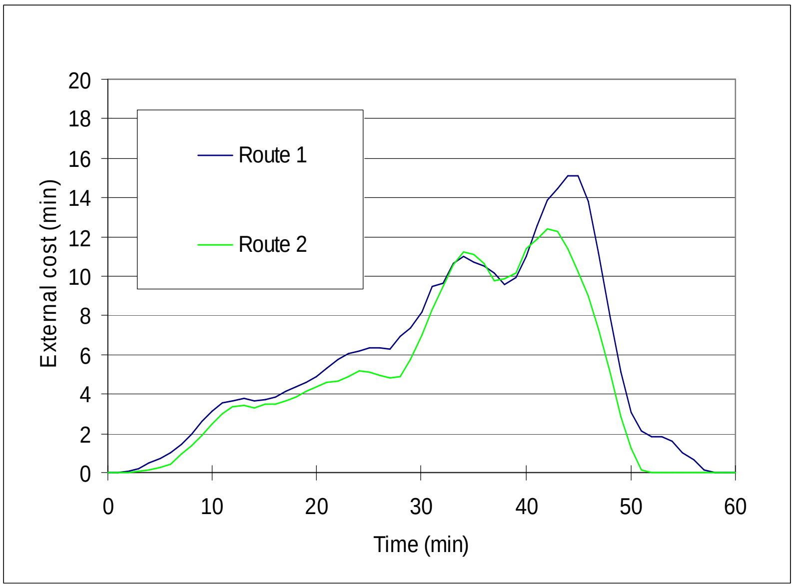
a) Deterministic queuing model



b) Divided linear model ($\alpha_a = \Delta s$)



c) Divided linear model ($\alpha_a = 2\Delta s$)



d) Whole-link traffic model

Figure 4.4 Dynamic system optimal tolls

4.5.3 Performance of the solution algorithm

This section discusses the performance of the solution algorithm for dynamic system optimal assignment. Figure 4.5 and Figure 4.6 illustrate respectively the monotonic reduction of the total system cost and the measure of disequilibrium over iterations. The travel time models considered here include the linear travel time models with $\alpha_a = 1\,\text{min}$, $\alpha_a = 2\,\text{min}$, and the whole link traffic model. The deterministic queuing model is not considered here because such travel time model is not suitable for the solution algorithm to work with due to its non-differentiability as discussed in Section 4.4.2 and Section 4.5.2.

In general, the results agree with the analysis that the total system cost reduces as the measure of disequilibrium drops. The measures of disequilibrium achieved are 0.024, 0.029, and 0.041 respectively for the divided linear models with $\alpha_a = 1\min$, $\alpha_a = 2\min$, and the whole-link traffic model. Such results of equilibrations are much less satisfactory than dynamic user equilibrium assignment which achieves an order of 10^{-17} of disequilibrium measures. The reason of this is that in the solution algorithm for dynamic system optimum, the auxiliary flows are calculated based on the traffic conditions at the last iteration, while the costate variables are calculated based on the traffic conditions at the current iteration. Such procedure gives correct values for costate variables but incorrect ones for assignment flows. The result of this is the improvement in system performance (i.e. reduction in total system cost) while we have to sacrifice the quality of equilibration. The numerical results further show that dynamic system optimal assignment is more difficult to solve for travel time models considering larger congestible portion on the link. The reason is that the larger congestible portion implies more interaction of traffic dynamics to be considered in the calculation process and hence solving the dynamic optimization problem becomes more difficult.

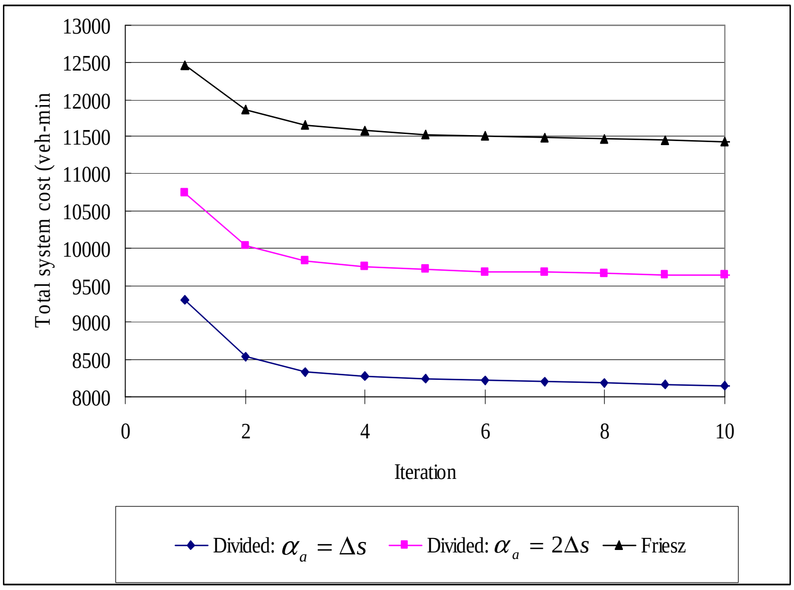


Figure 4.5 Total system cost over iteration

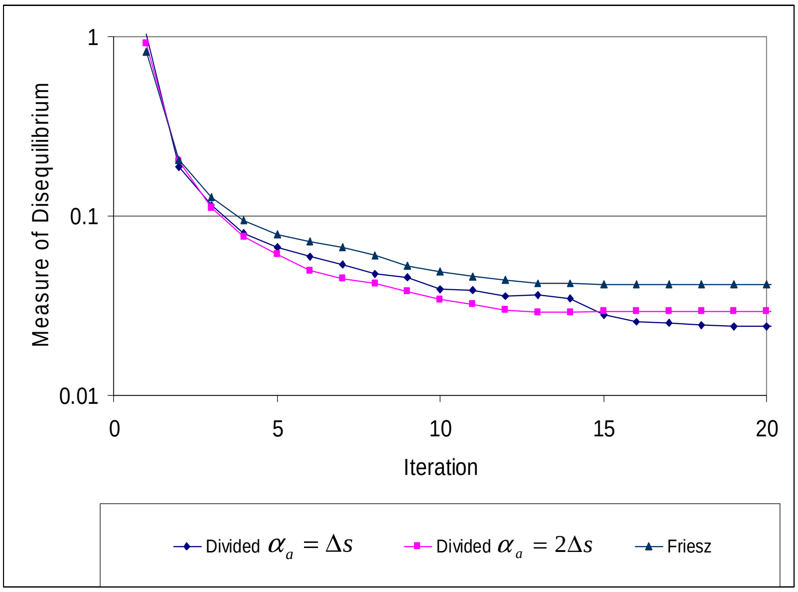
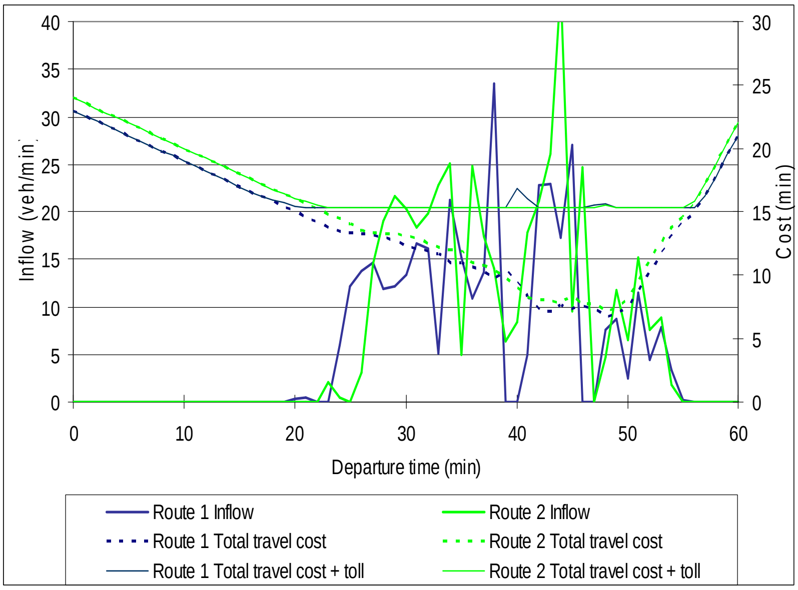


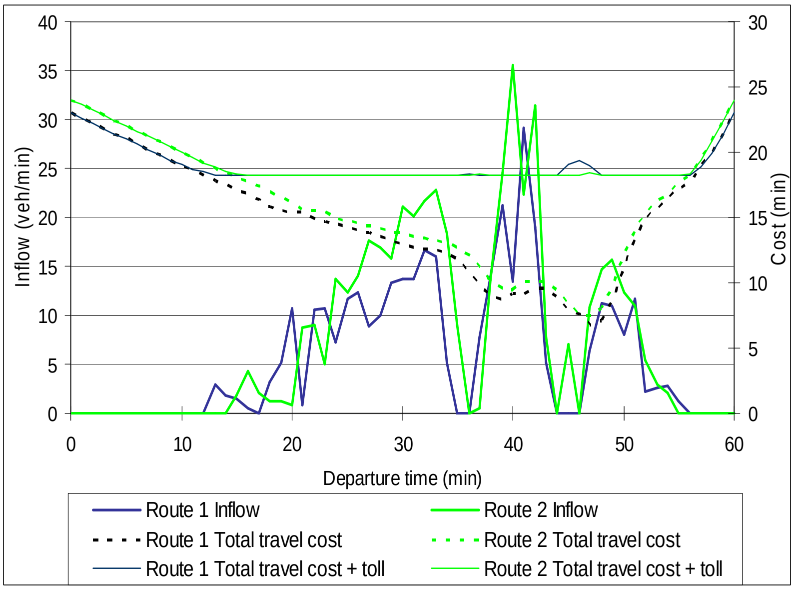
Figure 4.6 Measure of disequilibrium over iteration

4.5.4 An alternative solution algorithm for dynamic system optimum

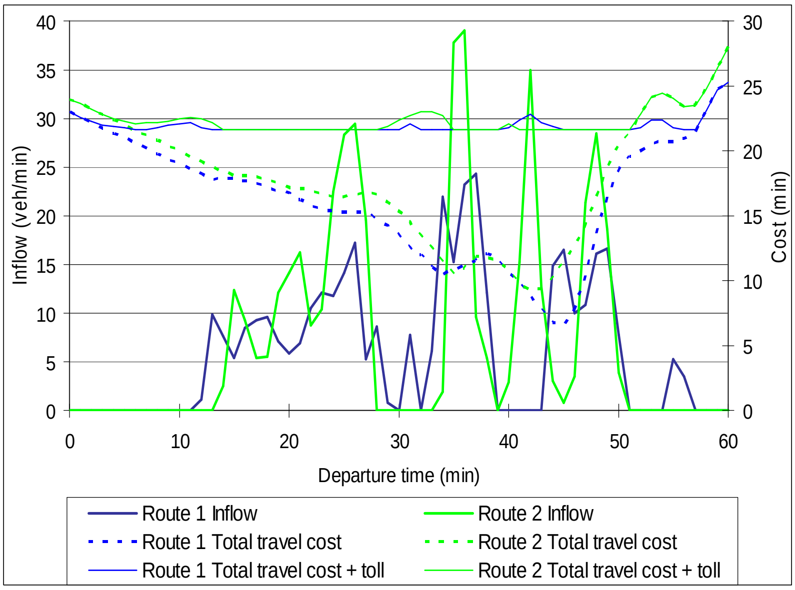
To improve the quality of equilibration in dynamic system optimum, we propose here and test an alternative solution procedure. In this version, we proceed as in Section 4.4.2 except that in the last iteration we do not calculate the optimal step size (i.e. Step 4 in the solution algorithm 4.4.2) and add the costate variables and the externalities (i.e. Step 5 in the solution algorithm in Section 4.4.2). As a result, the last iteration aims to calculate the correct equilibrium assignment to the total travel cost, using the externality and the costate variables from the previous iteration rather than determining them from the final assignment. By contrasting with the solution procedure discussed in Section 4.5.3, this procedure gives correct assignment flows but incorrect costate variables for a system optimal solution. The corresponding assignment results are shown in Figure 4.7. It can be noticed that the assignments are obtained with better quality of equilibration for all travel time models adopted. The measures of disequilibrium reach below 10^{-17} while with the spiky inflow profiles.



a) Divided linear model ($\alpha_a = \Delta s$)



b) Divided linear model ($\alpha_a = 2\Delta s$)



c) Whole-link traffic model

Figure 4.7 Dynamic system optimal assignments solved by the alternative solution strategy

Table 4.3 further shows the performances of these new assignments. The column *DSO** refers to the performances associated with the new assignments calculated by the alternative solution algorithm. This table reveals that with high quality equilibration, the new assignments calculated by the alternative solution method yields 81% - 86% in terms of total system cost reduction compared with the original solution method. It reveals that there is trade-off between the quality of equilibration and system cost reduction.

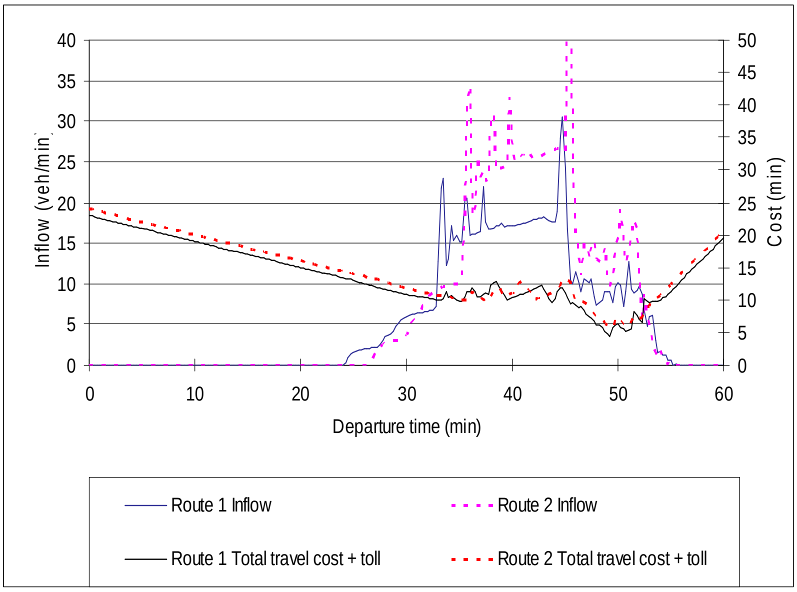
Table 4.3 Performance of dynamic system optimal assignments produced by the alternative solution strategy

	Total system cost (veh-min)			Difference btw.
	DUE	DSO	DSO*	DSO and DSO*
Divided: $\alpha_a = \Delta s$	9,311	8,077	8,310	81%
Divided: $\alpha_a = 2\Delta s$	10,741	9,627	9,778	86%
Friesz	12,465	11,434	11,583	86%

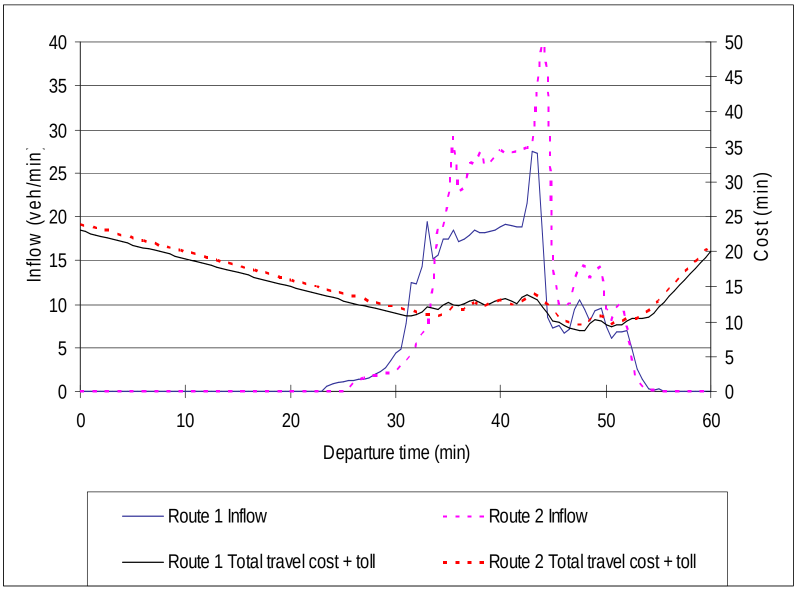
4.5.5 Effect of discretization on the assignment

Finally, this section investigates the effect of using different time discretizations on the quality of dynamic system optimal assignments. The divided linear travel time model with

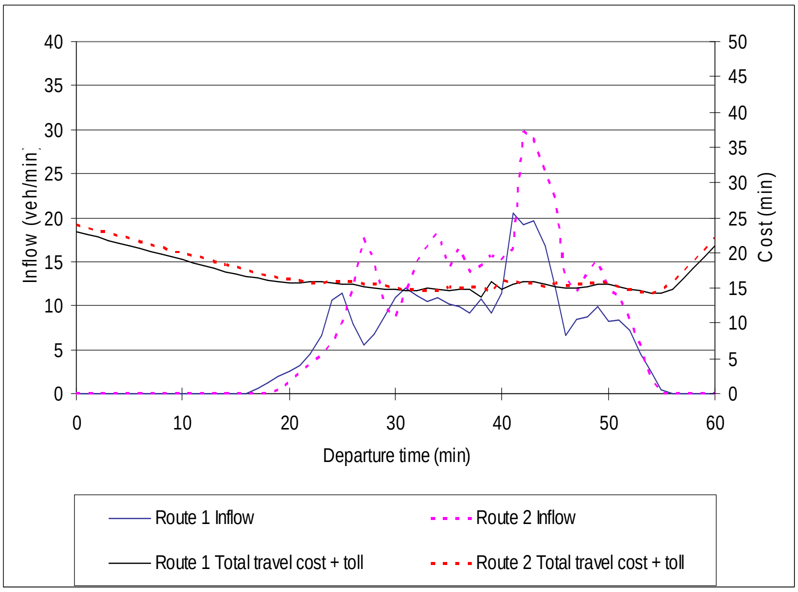
$\alpha_a = \Delta s$ and the whole-link traffic model are chosen for illustration. The main difference between these two travel time models is that the assignment results obtained by using the divided linear travel time model depend on the size of time discretization adopted (see also illustration by Mun, 2002), while for linear travel time model it does not. We investigate four different values of Δs: 0.25 min, 0.5 min, 1 min, and 2 mins. Figure 4.8 shows the results for the divided linear travel time model. As discussed in Section 3.7.4, the assignment profile varies with different degree of discretization due to the nature of the underlying travel time model. It is observed that the assigned inflow profile is smoothened when the size of Δs is large, say 1 min, and 2 mins. It also shows that it gives the best assignment result in terms of equilibration when $\Delta s = 1\text{min}$. This is consistent with the findings by Mun (2002) who studied the effect of choosing the value of Δs on the quality of dynamic user equilibrium assignment.



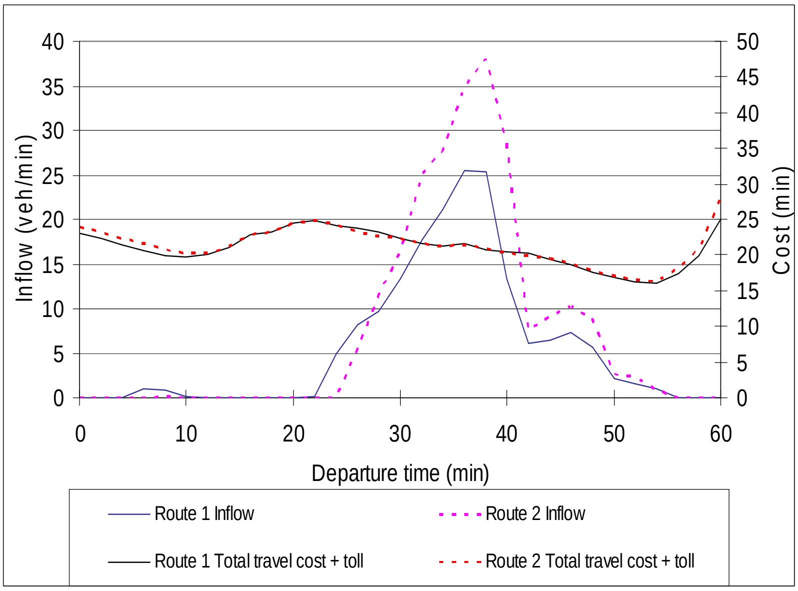
a) $\Delta s = 0.25\,\text{min}$



b) $\Delta s = 0.5\,\text{min}$



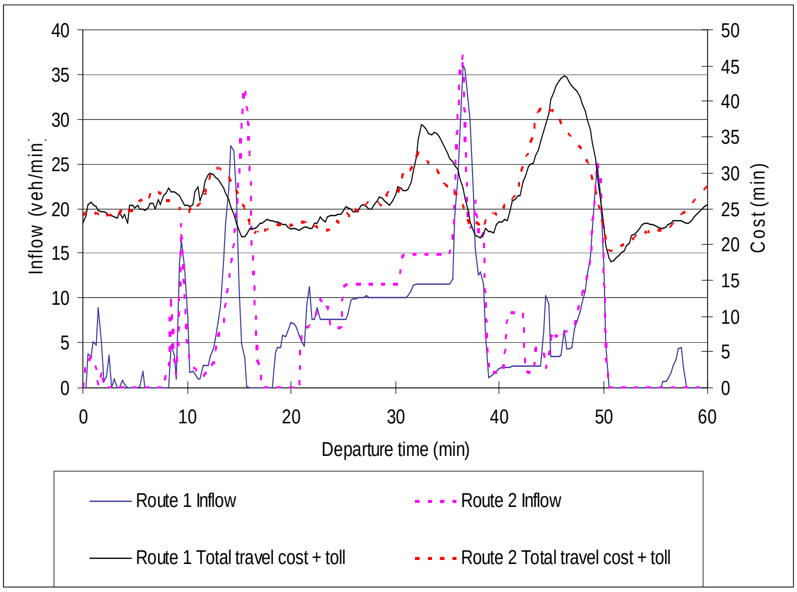
c) $\Delta s = 1\,\text{min}$



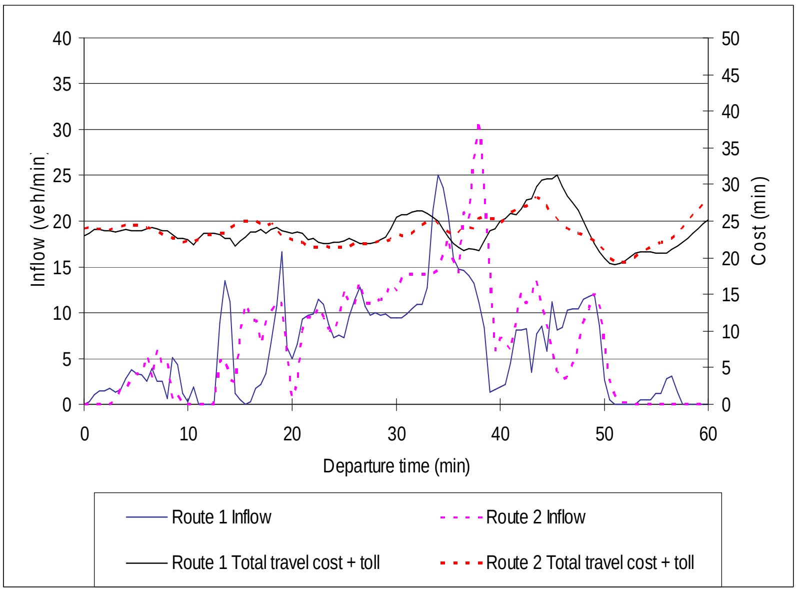
d) $\Delta s = 2\,\text{min}$

Figure 4.8 Dynamic system optimal assignments for divided linear travel time model ($\alpha_a = \Delta s$) against different sizes of discretization

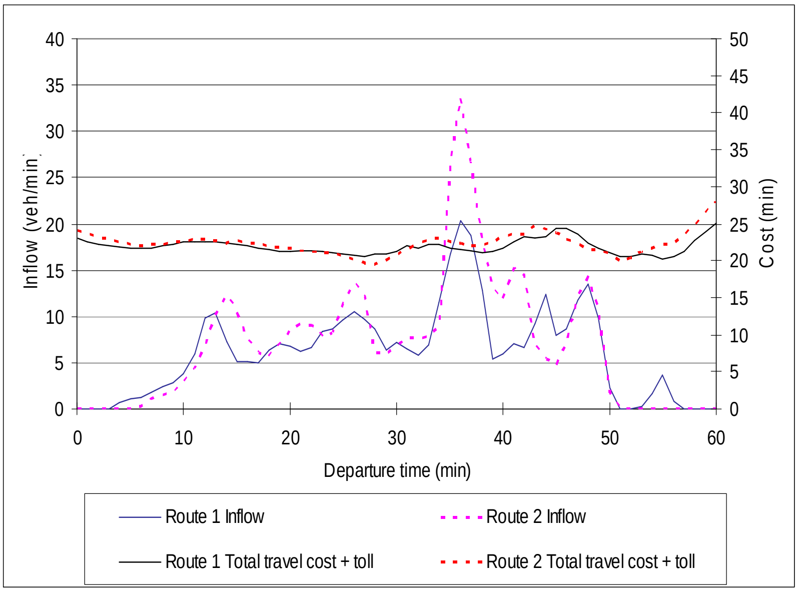
Figure 4.9 shows the results for the whole-link traffic model. Interestingly, the solution algorithm gives better assignment results with coarser discretization. The measures of disequilibrium are 0.18, 0.096, 0.041, and 0.017 respectively for Δs equals to 0.25 min, 0.5 min, 1 min, and 2 mins. It is also observed that the assigned inflow profile is smoothened with the size of Δs increases. Surprisingly, the coarser the discretization is, the better the assignment result in terms of the quality of equilibration is. It can be explained by the fact that finer discretization implies more sub-problems to solve and more interaction of traffic dynamics to capture.



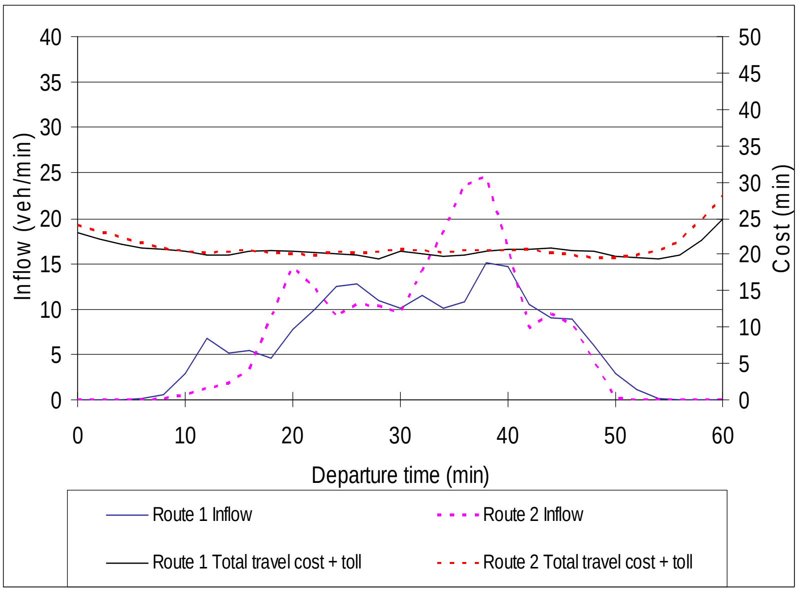
a) $\Delta s = 0.25\,\text{min}$



b) $\Delta s = 0.5\,\text{min}$



c) $\Delta s = 1\,\mathrm{min}$



d) $\Delta s = 2\,\mathrm{min}$

Figure 4.9 Dynamic system optimal assignments for whole-link traffic model against different sizes of discretization

4.6 DYNAMIC TOLLING STRATEGIES

From the experiments that we perform in Section 4.5, it is realized that calculating dynamic system optimal assignment and the associated optimal toll can be too difficult for implementation in practice, or even for research purpose. In the view of this, this section proposes some more practical tolling strategies for managing dynamic network traffic. The tolling strategies are compared with the dynamic system optimal toll and hence their efficiencies can be evaluated. The tolling strategies considered in this section include the uniform tolls and the congestion based tolls.

Given the toll to be imposed at the origin to each link, the corresponding *new* origin time-specific cost can be updated as the sum of two distinct components: one represents the original travellers' personal preferences $h(s)$, and the other is the imposed toll $\beta_a(s)$ for time of entry *s* to the corresponding link *a*. The associated tolled dynamic user equilibrium inflow to each link can be determined by using condition (3-11) as

$$\tilde{e}_a(s) = \left[\frac{1 - h'(s) - \beta'_a(s)}{1 + f'\left[\tau_a(s)\right]}\right]\tilde{g}_a\left[\tau_a(s)\right], \tag{4-89}$$

where $\tilde{e}_a$ and $\tilde{g}_a$ represent the inflow and outflow profiles at the tolled equilibrium respectively. This tolled dynamic user equilibrium assignment, of course, can also be calculated by using the dynamic user equilibrium assignment solver presented in Section 3.6 after adding the toll component, $\beta_a(s)$, to the travel cost, $C_a(s)$.

After obtaining the tolled equilibrium assignment, the corresponding total system travel cost, Z_{toll}, can be calculated. Following this, we can investigate the performance of each of these tolls in terms of *efficiency* by defining the efficiency, η_{toll}, of a toll as

$$\eta_{toll} = \frac{Z_{toll} - Z_{DUE}}{Z_{DSO} - Z_{DUE}}, \tag{4-90}$$

where Z_{DUE} and Z_{DSO} are the values of the total system cost estimated by the travel time model of interest under dynamic user equilibrium and dynamic system optimum respectively. Note that $\eta_{toll} = 100\%$ if the toll associated with dynamic system optimal assignment is implemented.

The following sections aim to evaluate the proportion of efficiency gains from the proposed tolling strategies with respect to the dynamic system optimizing toll.

4.6.1 Uniform toll

A uniform toll, or a time-invariant toll, refers to a toll which is constant while it applies. Such tolling scheme is easy to design and implement in practice. A real-life example of such tolling scheme can be referred to the one being implemented in Central London. A uniform toll of £8 is charged on most vehicles that use the tolling zone during 07:00 and 18:30 on workdays from Monday to Friday (Transport for London, 2007).

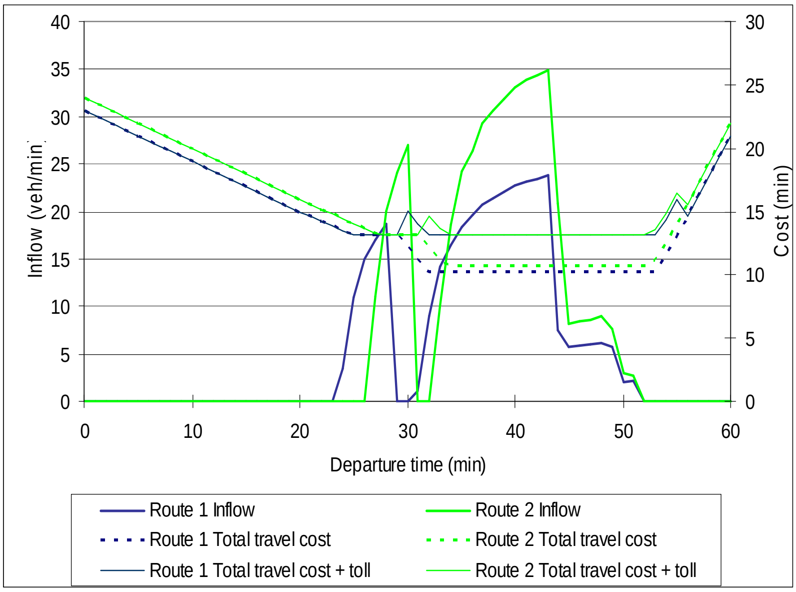
This section aims to determine the optimal uniform tolls which minimize the traffic congestion and compare their performance in terms of congestion reduction with the corresponding dynamic system optimal tolls under each of the travel time model discussed in the thesis. This is also a reflection of the value of using dynamic tolling strategies. Laih (1994) showed by using the bottleneck model (i.e. system with one single travel route modelled by the deterministic queue with a linear schedule delay cost function) that the uniform toll can at most yield 50% efficiency with respect to the optimal time-varying toll. An analytical proof can be found in Proposition 1.1 in Laih (1994).

This study investigates the uniform tolling strategies for the travel time models adopted in this thesis apart from the bottleneck model. We set the tolling period for each toll associated with each travel time model to be the same as the tolling period of corresponding dynamic system optimizing tolls that are calculated in Section 4.5.2. The magnitude, ρ, of the uniform toll to be imposed on each route for each travel time model is then determined by a brute force search such that the corresponding total system travel cost is minimized. The optimal uniform tolls are shown in Table 4.4. The tolls are expressed in the equivalent value of time which is in the unit of minute.

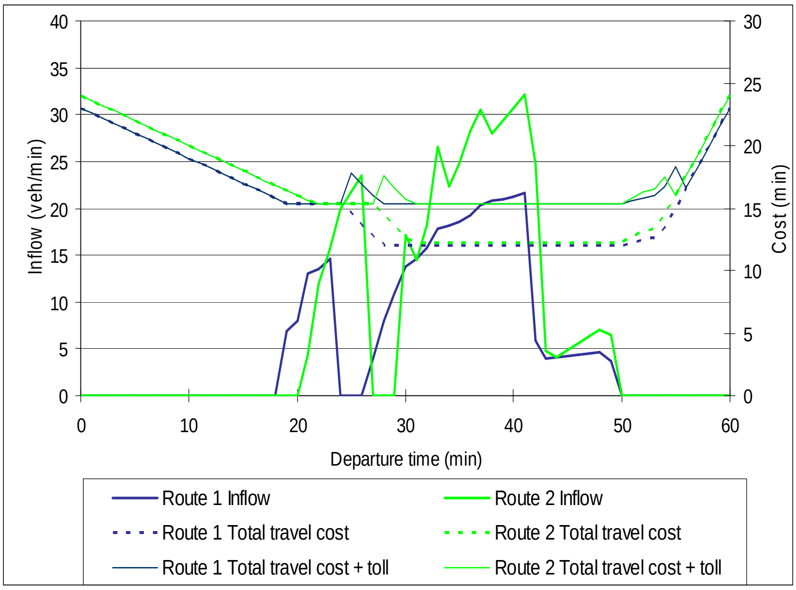
Table 4.4 Optimal uniform tolls

	Link 1		
	Start time (min)	End time (min)	Toll (min)
Divided: $\alpha_a = \Delta s$	29	54	2.92
Divided: $\alpha_a = 2\Delta s$	24	54	3.34
Friesz	19	55	3.95
	Link 2		
	Start time (min)	End time (min)	Toll (min)
Divided: $\alpha_a = \Delta s$	31	53	2.43
Divided: $\alpha_a = 2\Delta s$	27	53	3.09
Friesz	22	55	3.32

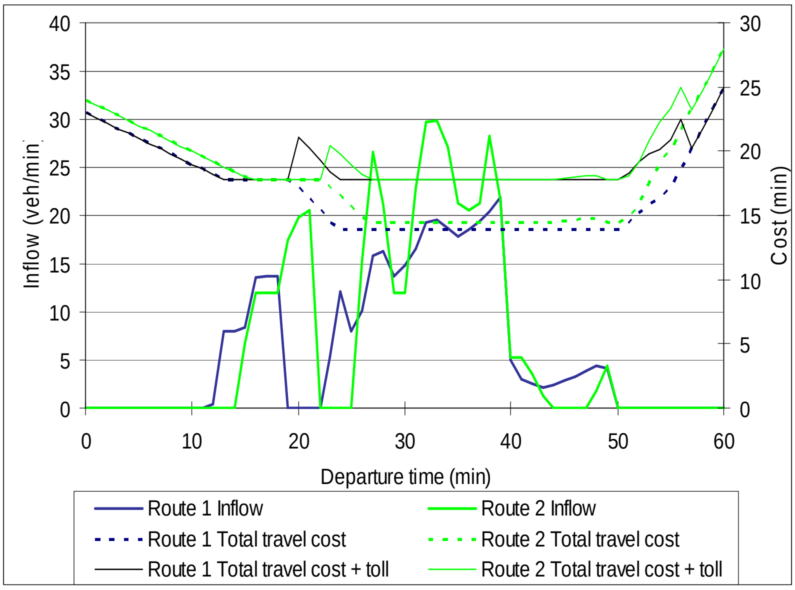
The corresponding tolled equilibrium assignments are also calculated and plotted in Figure 4.10. The dotted lines refer to the total travel costs, $C_a(s)$, while the thin solid lines refer to the total travel cost plus the uniform toll, $C_a(s)+\beta_a(s)$.



a) Divided linear model ($\alpha_a = \Delta s$)



b) Divided linear model ($\alpha_a = 2\Delta s$)



c) Whole link traffic model

Figure 4.10 Tolled assignments with uniform tolls

All assignments are in good equilibration with the tolled total travel cost because the solution procedure only involve solving a forward dynamic programme as in the dynamic user equilibrium solver, rather than solving two dynamic programme (backward and forward) as in the dynamic system optimal assignment solver. There is an interesting observation in the assignment results that, with the uniform toll, there is a mass of traffic flowing into the system just before the toll be effective. It indicates a tendency of travellers that they want to rush into the system within a particular time interval. It is because everyone knows there is an abrupt increase in travel cost after the toll is effective. This observation reveals a shortcoming of traditional time-invariant toll: it can induce traffic disruption during the time period before, and possibly after, the toll charges.

Table 4.5 Performances of uniform tolls

	Total system costs (veh-min)			Efficiency η_U
	DUE	DSO	Uniform tolls	
Divided: $\alpha_a = \Delta s$	9,311	8,077	8,813	40%
Divided: $\alpha_a = 2\Delta s$	10,741	9,627	10,203	48%
Friesz	12,465	11,434	11,943	50%

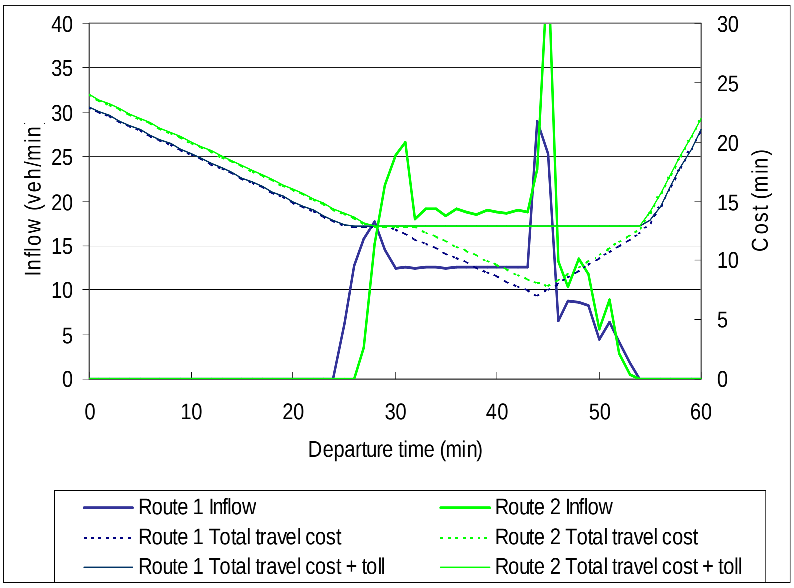
The efficiencies of the uniform tolls calculated with different travel time models are shown in Table 4.5. Similar to the findings by Laih (1994), the optimal uniform tolls yield around 50% efficiency of the dynamic system optimizing toll. In addition to this, it is realized the uniform toll appears to be more effective under travel time model that considers larger portion of congestible part.

4.6.2 Congestion based toll

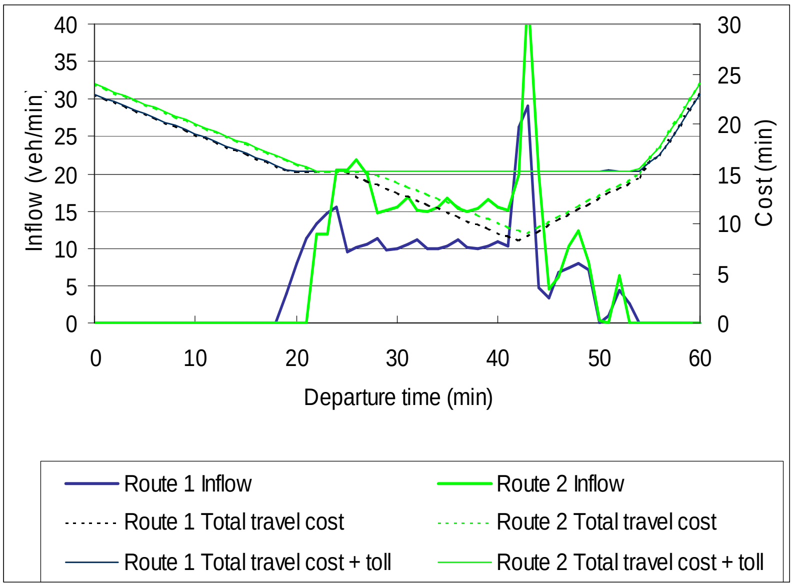
Different from the uniform toll discussed in Section 4.6.1, the congestion based toll is dynamic in nature. The idea of such tolling strategy is simple and it is originated from the deterministic queue based toll. Similar to the deterministic queue based toll, the underlying principle of this congestion based toll is to charge a toll which is equal to the cost associated with congestion (i.e. the total actual travel time minus the total free flow travel time, which is $\frac{x_a}{Q_a}$) that would be incurred in queues in the network system under the untolled dynamic user equilibrium. The advantage of doing this is that the toll collected can be used as a communal advantage rather than the dead loss. In fact, such tolling scheme is similar the one tested in the city of Cambridge, United Kingdom in 1993[†††] (Sharpe, 1993; Small and Gomez-Ibañez,

[†††] Implementation efforts of this tolling scheme in Cambridge ended with a change in the shire government earlier in 1993. Although technically feasible, it was concerned about the potential for public outrage when the road tolls are unpredictable (see Small and Gomez-Ibañez, 1998).

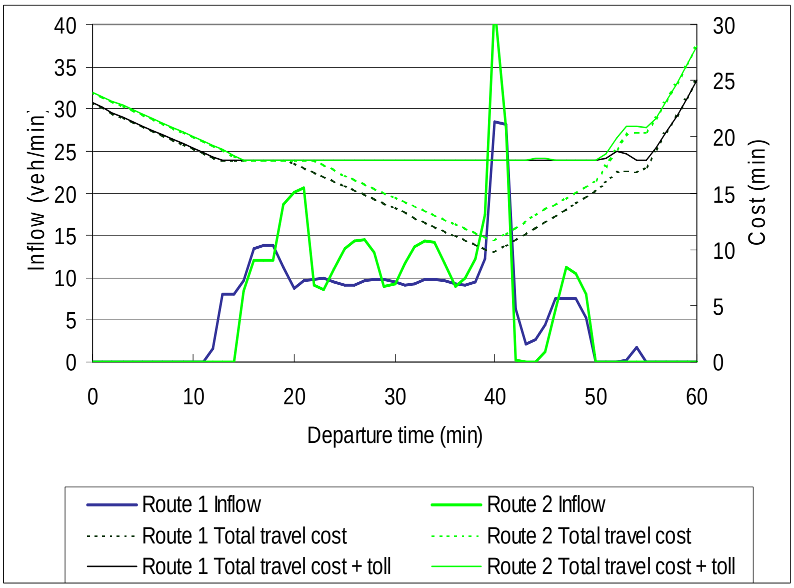
1998). The performance of this congestion based toll is investigated by simulating its performance on the corresponding travel time models. The corresponding assignments are plotted in Figure 4.11.



a) Divided linear model ($\alpha_a = \Delta s$)



b) Divided linear model ($\alpha_a = 2\Delta s$)



c) Whole link traffic model

Figure 4.11 Assignment profiles with the congestion based tolls

It is shown that the assignment is again in good equilibration with the tolled total travel cost due to the straightforward solution procedure as mentioned previously. It is further observed that there are spikes appear in the inflow profile around the times when there is a sharp change in the slope in the time-specific costs. The appearance of spikes can be understood since corners in the cost function can induce corners in the corresponding inflow profiles (Kamien and Schwartz, 1991).

Table 4.6 Performance of congestion based tolls

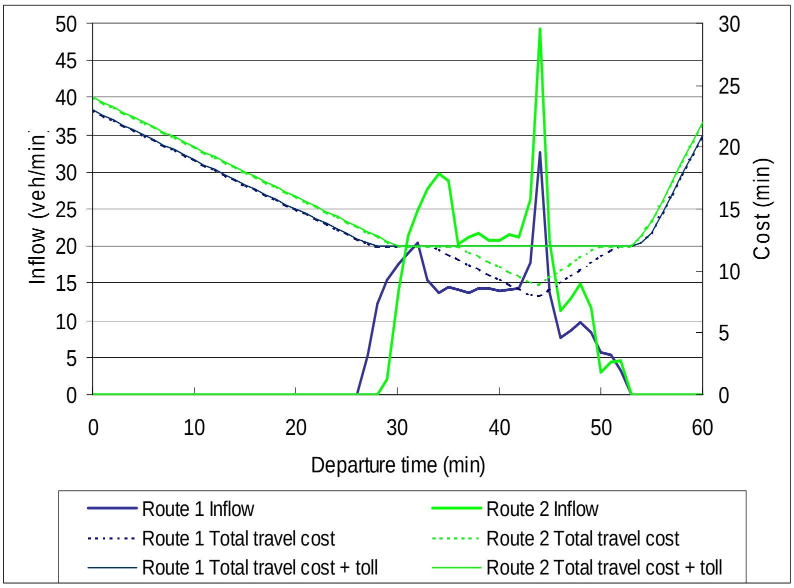
	Total system cost (veh-min)			
	DUE	DSO	Congestion based tolls	Efficiency η_C
Divided: $\alpha_a = \Delta s$	9,311	8,077	8,229	88%
Divided: $\alpha_a = 2\Delta s$	10,741	9,627	9,708	93%
Friesz	12,465	11,434	11,475	96%

To gain further insight on the performance of the assignments, the efficiencies of the congestion based tolls are calculated and shown in Table 4.6. It can be noted that the congestion based toll is reasonably effective in managing dynamic traffic congestion compared to the dynamic system optimizing toll. In general, the congestion based tolls yield a high efficiency of around 90%, although the efficiency drops as the portion of congestible region of the traffic model decreases. The drop in the toll efficiency in traffic models with less congestible portion can be understood as a result of the fact that congestion is less significant in those models. Despite its simplicity, the congestion based toll appears to be a promising strategy for managing dynamic network traffic.

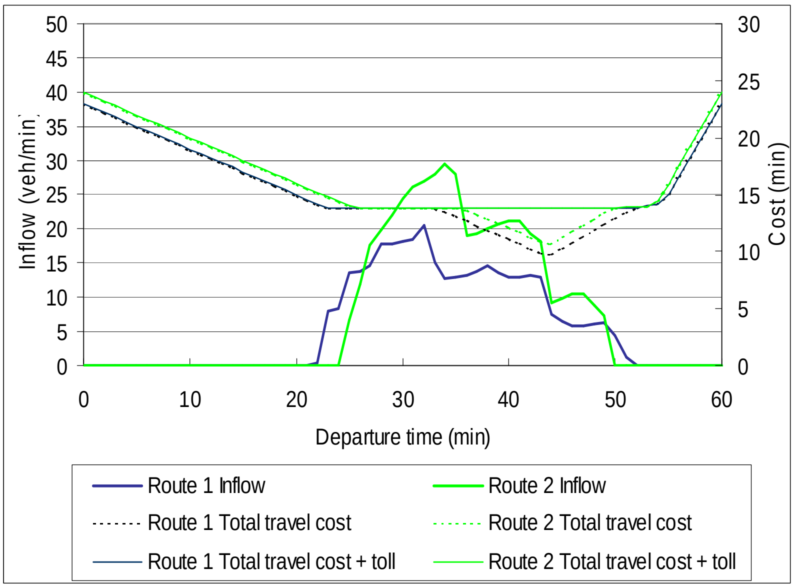
4.6.3 Robustness of toll calculation method

Section 4.6.2 shows that the tolling strategy based on congestion profile is effective on managing dynamic road traffic. In Section 4.6.2, it is assumed that we have an exact model for the underlying traffic behaviour. In reality, we do not have such information and hence it is interesting and important to investigate how *robust* a toll calculation method with respect to the underlying traffic model adopted is for implementation in practice. To our knowledge, such robustness of calculation methods of dynamic tolls has not been studied or documented in the literature.

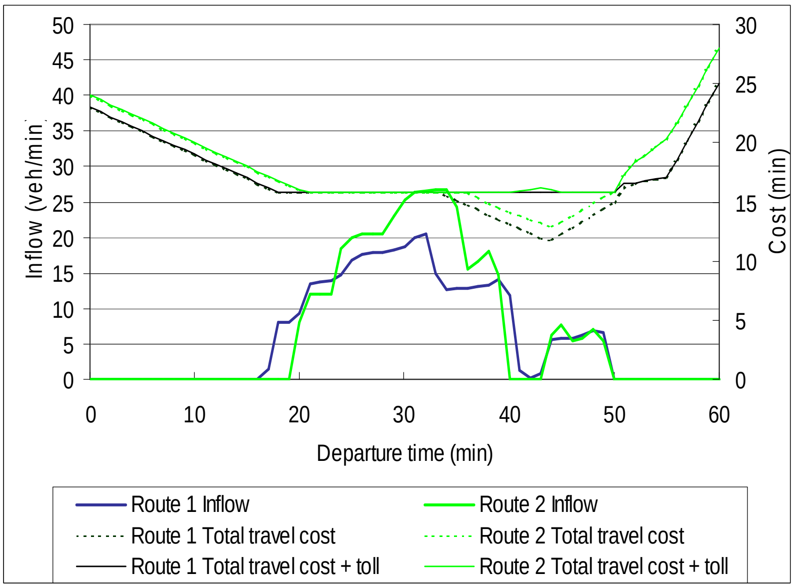
The congestion based tolls associated with each travel time model which are calculated in Section 4.6.2 are now simulated on travel time models of other kinds. The resulting tolled assignments are plotted in Figures 4.12 – 4-15 respectively for tolls based on the deterministic queuing model, the divided linear model ($\alpha_a = \Delta s$), the divided linear model ($\alpha_a = 2\Delta s$), and the whole-link traffic model. It can be seen that all assignments are in good equilibration, which is expected. Similar to the assignments plotted in Section 4.6.2, the spikes in Figures 4.12a, 4.13a, 4.14a, 4.14b, 4.15a, 4.15b, and 4.15c still appear around the times when there is a sharp change in the slope of the time-specific costs.



a) On divided linear model ($\alpha_a = \Delta s$)

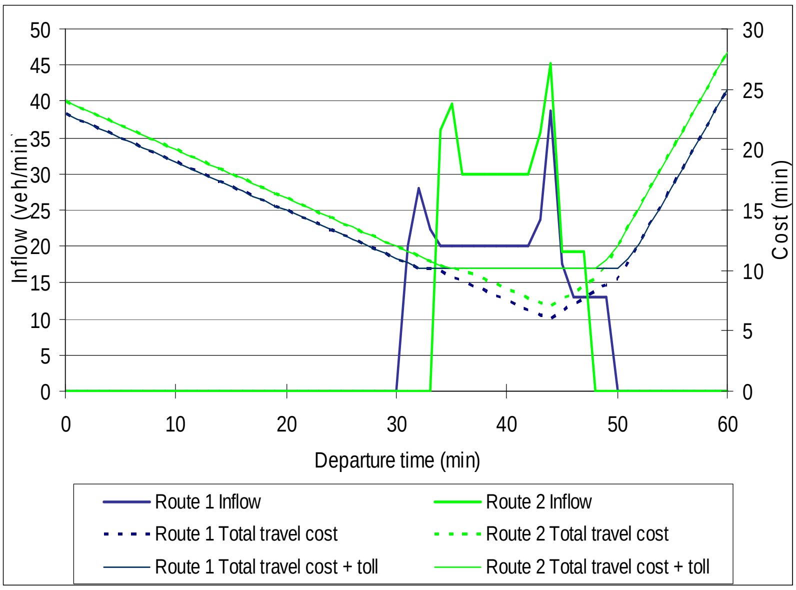


b) On divided linear model ($\alpha_a = 2\Delta s$)

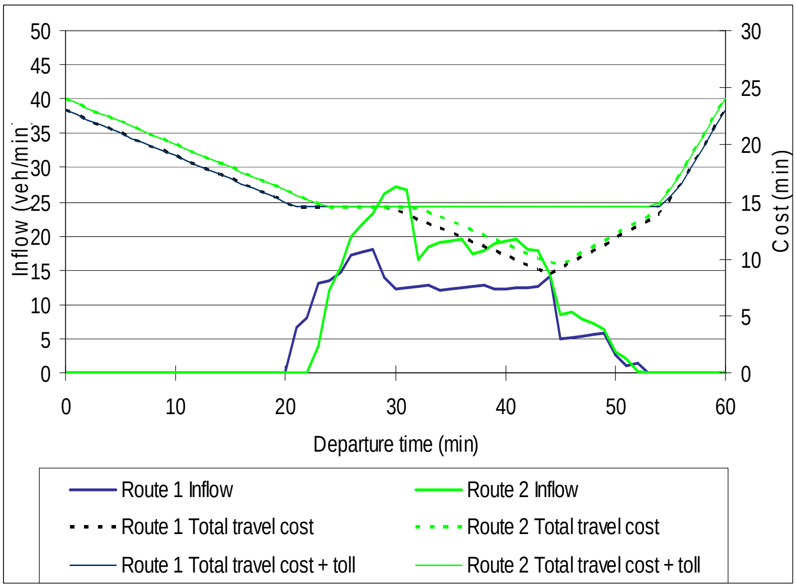


c) On whole link traffic model

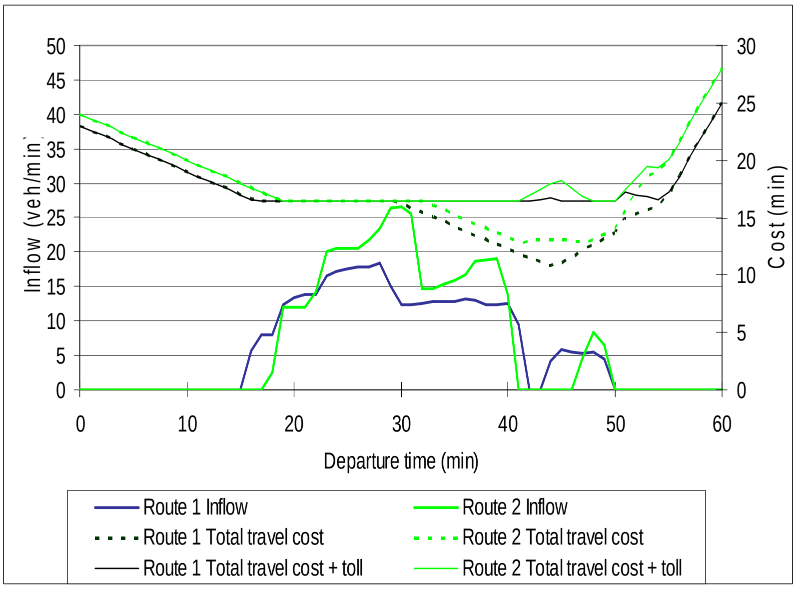
Figure 4.12 Tolled assignments with deterministic queue based toll



a) On deterministic queuing model

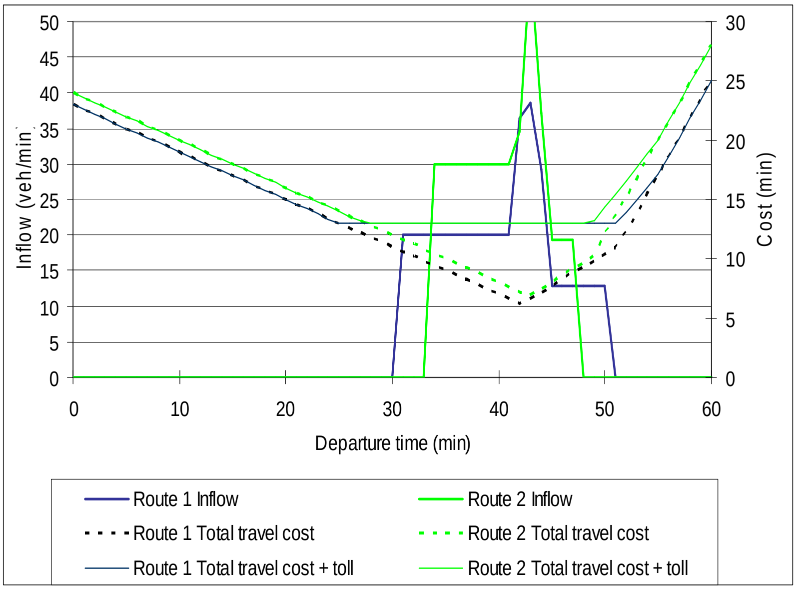


b) On divided linear model ($\alpha_a = 2\Delta s$)

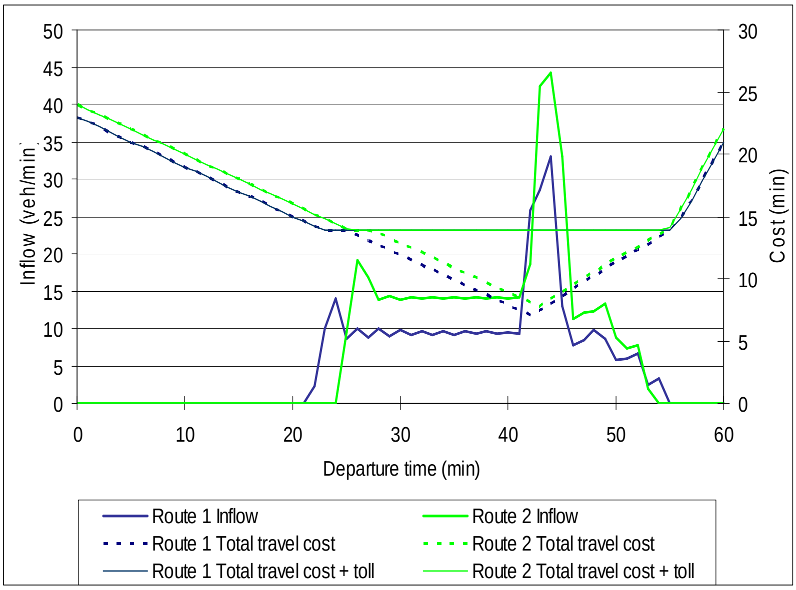


c) On whole link traffic model

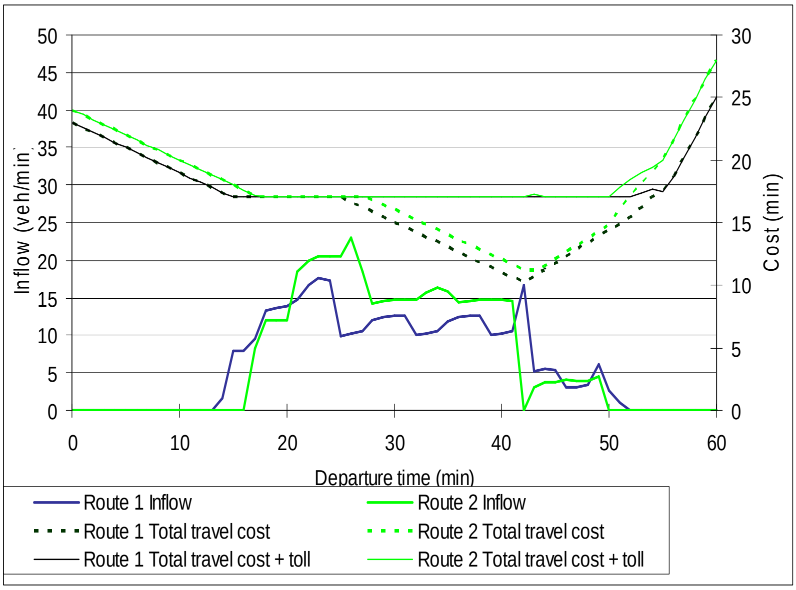
Figure 4.13 Tolled assignments with divided linear model ($\alpha_a = \Delta s$) based toll



a) On deterministic queuing model

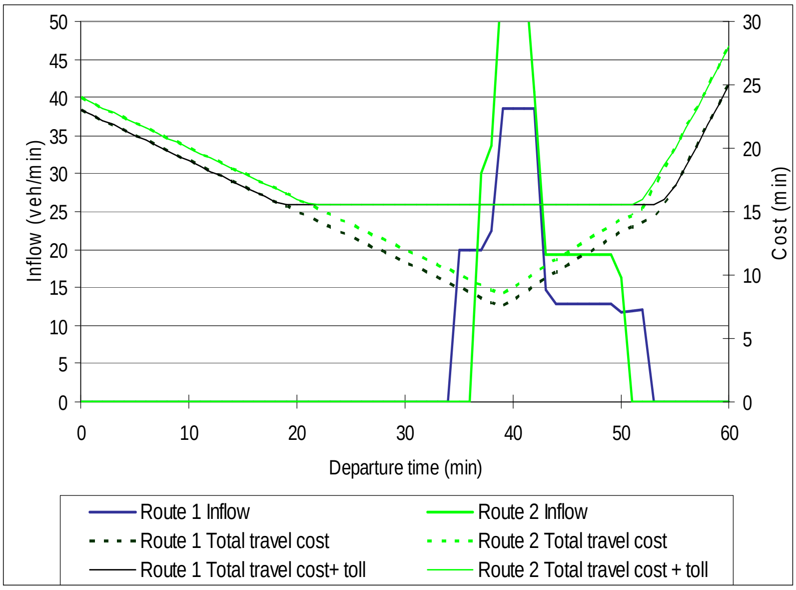


b) On divided linear model ($\alpha_a = \Delta s$)

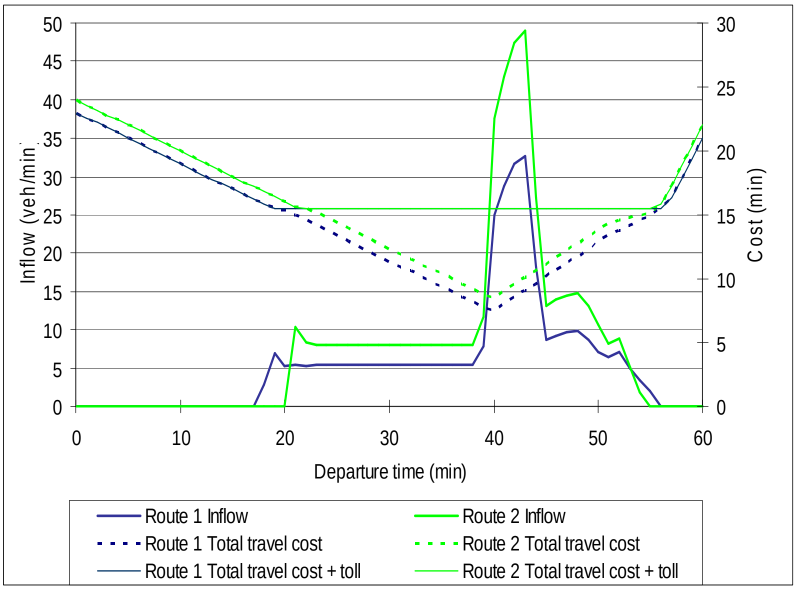


c) On whole link traffic model

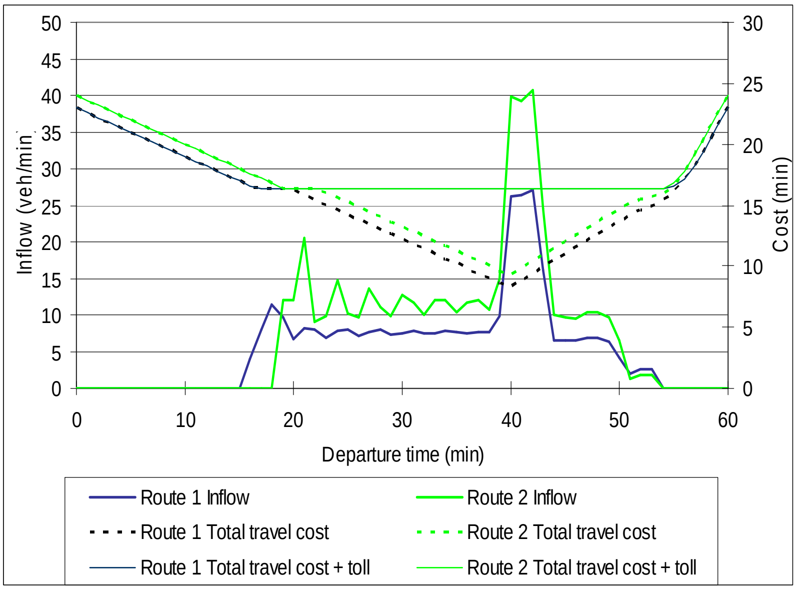
Figure 4.14 Tolled assignments with divided linear model ($\alpha_a = 2\Delta s$) based toll



a) On deterministic queuing model



b) On divided linear model ($\alpha_a = \Delta s$)



c) On divided linear model ($\alpha_a = 2\Delta s$)

Figure 4.15 Tolled assignments of whole link traffic model based toll

To gain further insight, the performance in terms of total system travel costs and efficiencies of each toll with each of the travel time model are shown in Tables 4.7a and 4.7b respectively. In both tables, the traffic model down the column represents the one which the congestion toll is calculated based upon, while the traffic model along the row refers to the one by which the underlying system is evaluated. The entries on the diagonal of the tables, which are bold, represent the performance and the efficiencies of the congestion based tolls that are calculated and implemented by the same traffic model. These numbers represent *perfect* knowledge on the underlying traffic dynamics and the associated results are exactly the same as those we calculated in Section 4.6.2.

For the toll calculated based on the deterministic queuing model, its performance and efficiency decreases along the row. This implies that such toll is more effective for travel time models that consider smaller portion of congestible part (e.g. the divided linear model with $\alpha_a = \Delta s$). The deterministic queue based toll can still achieve an efficiency of 63% under the divided linear travel time model ($\alpha_a = \Delta s$), while the efficiency of the toll drops to 15% when it is being implemented on whole-link traffic model.

Contrasting with the deterministic queuing model, the toll based on the whole-link traffic model has better performance on traffic models with larger portion of congestible part (e.g. the divided linear model with $\alpha_a = 2\Delta s$), while its efficiency drops to -5% (the negative sign means it is actually worse than no-tolled equilibrium condition) when it is implemented on the deterministic queuing model. These findings are due to the similarity between the traffic models upon which the toll is calculated and on which the toll is evaluated.

The tolls calculated using the divided linear traffic models perform best on themselves, but they are still able to give reasonable performance over the two extremes. In particular, the divided linear model ($\alpha_a = 2\Delta s$) achieves efficiencies of 32% and 52% under the deterministic queuing model and the whole-link traffic model respectively. This suggests that the class of divided linear travel time models should receive more attention on designing robust dynamic traffic control strategies in future research.

Table 4.7 Robustness of congestion based tolls

a) Total system travel costs (veh-min)

Traffic model for toll calculation	Traffic model for evaluation: DDQ	Divided: $\alpha_a = \Delta s$	Divided: $\alpha_a = 2\Delta s$	Friesz
DDQ	**4,032**	8,534	10,269	12,313
Divided: $\alpha_a = \Delta s$	6,768	**8,229**	10,060	12,276
Divided: $\alpha_a = 2\Delta s$	6,867	8,298	**9,708**	11,929
Friesz	*8,248*	8,926	9,776	**11,475**

b) Efficiencies

Traffic model for toll calculation	Traffic model for evaluation: DDQ	Divided: $\alpha_a = \Delta s$	Divided: $\alpha_a = 2\Delta s$	Friesz
DDQ	**100%**	63%	42%	15%
Divided: $\alpha_a = \Delta s$	32%	**88%**	61%	18%
Divided: $\alpha_a = 2\Delta s$	30%	82%	**93%**	52%
Friesz	*-5%*	31%	87%	**96%**

4.7 DISCUSSION

This chapter investigates dynamic system optimal assignment with departure time choice in a rigorous and original way. The system optimal assignment is formulated as the optimal control problem. A fixed amount of traffic is assigned over a system with a single origin-destination pair connected by mutually distinct links. The objective of the assignment is to minimize the total system travel cost within a fixed study period. Similar to its static counterparts, it is shown that dynamic system optimal assignment can be reduced to an equivalent dynamic user equilibrium assignment with additional components of the cost introduced. Each traveller who enters a travel link at a time of entry is expected to pay an amount of toll which is equal to his/her own externality imposed to the others, plus a toll that is charged by an external system manager for using the transport system during the time when he/she presents in the system. To capture the dynamic externality, we develop a novel sensitivity analysis of travel cost. Solution algorithms are presented for implementing the sensitivity analysis and solving dynamic system optimal assignments for a range of travel time models. Example calculations are presented and the characteristics of the results are discussed. The results first show that the sensitivity analysis can accurately capture the variations in travel time and travel cost with respect to perturbations in traffic flows. On dynamic system optimal assignment, it is observed that congestion is generally inevitable even in dynamic system optimum. The durations of assignments are lengthened and the inflows are spread out as suggested by Chu (1995) in order to reduce the intensity of traffic congestion. This finding suggests that analysis based on the deterministic queuing model is not generally applicable. With the deterministic queuing model, congestion is eliminated in dynamic system optimum and the period of assignment is identical to that of dynamic user equilibrium.

To improve the quality of equilibration in dynamic system optimum, we propose and test an alternative solution strategy. Contrasting with the original solution method, the alternative solution algorithm does give assignments with better quality equilibration. However, the results also show that we have 15% - 20% drop in delay reduction. It clearly reveals that there is trade-off between the quality of equilibration and system cost reduction. We also investigate the effect of using different time discretizations on the quality of dynamic system optimal assignments. Surprisingly, the coarser the discretization, the better the assignment result in terms of the quality of equilibration. It can be explained by the fact that finer discretization implies more sub-problems to solve and more interaction of traffic dynamics to capture. As a result, the dynamic optimization problem becomes more intricate to solve. These preliminary numerical experiments suggest that future work is still necessary to investigate more efficient solution algorithms for analytical dynamic system optimal assignment.

It is realized that calculating dynamic system optimal assignment and the associated optimal toll can be too difficult for implementation in practice, or even for research purpose. In the view of this, we propose some more practical tolling strategies for managing dynamic network traffic, which are the uniform tolls and the congestion based tolls. These tolling strategies are compared with the dynamic system optimal toll and hence their efficiencies can be evaluated. The uniform toll refers to a toll which is constant while it applies. Such tolling scheme is convenient and popular to design and implement in practice. With the uniform toll, the assignment results show that there is a mass of traffic flowing into the system just before the toll becomes effective. The effect can induce traffic disruption during the time period before and after the toll charges. The uniform tolls can gain around 50% of the delay reduction which would have been achieved by implementing the fully time-varying tolls. This finding is consistent with the earlier one carried out by Laih (1994). The congestion based tolls are more effective in managing dynamic network traffic. In general, the congestion based tolls yield an efficiency of around 90%, although the efficiency drops as the portion of congestible region of the traffic model decreases. The drop in the toll efficiency in traffic models with less congestible portion can be understood as a result of the fact that congestion is less significant in those models.

In deriving the tolling strategies, it is assumed that we have an exact model for the underlying traffic behaviour. In reality, we do not have such information and hence it is interesting and important to investigate how *robust* a designed tolling strategy is for implementation in practice. Given its effectiveness, we further investigate the robustness of the congestion based tolls with respect to the underlying traffic model adopted. It is found that the tolls calculated by using divided linear traffic models can generally give good performances on a wide range of traffic models. The divided linear travel time models thus should receive more attention in the future research on robust dynamic traffic control strategies design.

Appendix 4A: A preliminary study on the sufficiency of the necessary conditions of dynamic system optimal assignment

This appendix shows a preliminary study on the sufficiency of the necessary conditions of the dynamic system optimal assignment derived in Proposition 4.1.

Recall the Lagrangian (4-9) formed by augmenting the objective function and the constraints in the dynamic system optimization formulation (4-1) as:

$$\begin{aligned}\min_{e_a^*(s)} Z^* = \sum_{\forall a} \int_0^T & \left\{ \begin{aligned} & C_a(s)e_a(s) + \lambda_a(s)\left\{[e_a(s) - g_a(s)] - \frac{dx_a(s)}{ds}\right\} + \mu_a(s)\left\{\frac{dE_a(s)}{ds} - e_a(s)\right\} \\ & + \gamma_a(s)\left\{\frac{dG_a[\tau_a(s)]}{ds} - e_a(s)\right\} - \rho_a(s)e_a(s) \end{aligned} \right\} ds \\ & + \nu_{od}\left(J_{od} - \sum_{\forall a} E_a(T)\right)\end{aligned} ,$$

(4A-1)

and its first order variation with respect to its variables in (4-16):

$$\begin{aligned}\delta Z^* = & \sum_{\forall a}[-\lambda_a(T)\delta x_a(T) + \mu_a(T)\delta E_a(T)] + \sum_{\forall a}\int_0^T \left(\frac{\partial H_a}{\partial u_s}\right)\delta e_a(s)ds \\ & + \sum_{\forall a}\int_0^T \left[\frac{\partial H_a}{\partial v_s}\delta g_a(s) + \frac{\partial H_a}{\partial v_{\tau_a(s)}}\delta g_a[\tau_a(s)] + \left(\frac{\partial H_a}{\partial x_a} + \frac{d\lambda_a(s)}{ds}\right)\delta x_a(s)\right]ds , \\ & - \sum_{\forall a}\int_0^T \frac{d\mu_a(s)}{ds}\delta E_a(s)ds - \nu_{od}\sum_{\forall a}\delta E_a(T)\end{aligned}$$

(4A-2)

in which the Hamiltonian function is defined as in (4-13) as:

$$\begin{aligned}H_a(s) = & C_a(s)e_a(s) + \lambda_a(s)[e_a(s) - g_a(s)] - \mu_a(s)e_a(s) \\ & + \gamma_a(s)\left\{g_a[\tau_a(s)]\frac{d\tau_a(s)}{ds} - e_a(s)\right\} - \rho_a(s)e_a(s)\end{aligned} . \qquad (4A-3)$$

The notation $\frac{\partial H_a}{\partial u_s}$, $\frac{\partial H_a}{\partial v_s}$, $\frac{\partial H_a}{\partial v_{\tau_a(s)}}$, and $\frac{\partial H_a}{\partial x_a}$ represent the derivatives of Hamiltonian function with respect to its corresponding variables: $e_a(s)$, $g_a(s)$, $g_a[\tau_a(s)]$, and $x_a(s)$ respectively. The parameters u_s and v_s represent respectively the perturbations at time s in the inflow e_a and outflow g_a as defined in (4-8), (4-14), and (4-15).

To ensure sufficiency, we require that the second order variation of Lagrangian Z^* to be non-negative with respect to all perturbation in the neighbouring paths (see for example, Bryson and Ho, 1975, p181). The second order variation of Z^* is derived as:

$$
\begin{aligned}
\delta^2 Z^* = &\sum_{\forall a}\int_0^T \left[\delta^T e_a(s)\left(\frac{\partial^2 H_a}{\partial u_s^2}\right)\delta e_a(s) + \delta^T g_a(s)\left(\frac{\partial^2 H_a}{\partial v_s \partial u_s}\right)\delta e_a(s) + \delta^T x_a(s)\left(\frac{\partial^2 H_a}{\partial x_a \partial u_s}\right)\delta e_a(s)\right] ds \\
&+\sum_{\forall a}\int_{\tau_a(0)}^T \left[\delta^T e_a(s)\left(\frac{\partial^2 H_a}{\partial u_s \partial v_s}\right)\delta g_a(s) + \delta^T g_a(s)\left(\frac{\partial^2 H_a}{\partial v_s^2}\right)\delta g_a(s) + \delta^T x_a(s)\left(\frac{\partial^2 H_a}{\partial x_a \partial v_s}\right)\delta g_a(s)\right] ds \\
&+\sum_{\forall a}\int_{\tau_a(0)}^T \left[\begin{array}{l} \delta^T e_a(s)\left(\frac{\partial^2 H_a}{\partial u_s \partial v_s}\frac{1}{\dot{\tau}_a}\right)\Bigg|_{\sigma_a(s)}\delta g_a(s) + \delta^T g_a(s)\left(\frac{\partial^2 H_a}{\partial v_s^2}\frac{1}{\dot{\tau}_a}\right)\Bigg|_{\sigma_a(s)}\delta g_a(s) \\ + \delta^T x_a(s)\left(\frac{\partial^2 H_a}{\partial x_a \partial v_s}\frac{1}{\dot{\tau}_a}\right)\Bigg|_{\sigma_a(s)}\delta g_a(s) \end{array}\right] ds \\
&+\sum_{\forall a}\int_0^T \left[\delta^T e_a(s)\left(\frac{\partial^2 H_a}{\partial u_s \partial x_a}\right)\delta x_a(s) + \delta^T g_a(s)\left(\frac{\partial^2 H_a}{\partial v_s \partial x_a}\right)\delta x_a(s) + \delta^T x_a(s)\left(\frac{\partial^2 H_a}{\partial x_a^2}\right)\delta x_a(s)\right] ds
\end{aligned}
$$

(4A-4)

The notation $\frac{\partial^2 H_a}{\partial u_s^2}$, $\frac{\partial^2 H_a}{\partial v_s^2}$, $\frac{\partial^2 H_a}{\partial x_a^2}$, $\frac{\partial^2 H_a}{\partial u_s \partial x_a}$, $\frac{\partial^2 H_a}{\partial u_s \partial v_s}$, $\frac{\partial^2 H_a}{\partial x_a \partial u_s}$, and $\frac{\partial^2 H_a}{\partial x_a \partial v_s}$ are the second order derivatives with respect to the variables e_a, g_a, and x_a, and they are reckoned as

$$
\begin{aligned}
\frac{\partial^2 H_a}{\partial u_s^2} &= \frac{\partial}{\partial u_s}[C_a(s)+\Psi_a(s)+\lambda_a(s)-\gamma_a(s)-\mu_a(s)-\rho_a(s)] \\
&= \frac{\partial}{\partial u_s}[C_a(s)+\Psi_a(s)]
\end{aligned}, \quad (4A\text{-}5)
$$

$$\frac{\partial^2 H_a}{\partial x_a^2} = f''[\tau_a(s)]\frac{e_a(s)}{Q_a^2}, \quad (4A\text{-}6)$$

$$\frac{\partial^2 H_a}{\partial x_a \partial u_s} = \frac{\partial^2 H_a}{\partial u_s \partial x_a} = (1 + f'[\tau_a(s)])\frac{1}{Q_a}, \tag{4A-7}$$

$$\frac{\partial^2 H_a}{\partial v_s^2} = 0, \tag{4A-8}$$

$$\frac{\partial^2 H_a}{\partial u_s \partial v_s} = \frac{\partial^2 H_a}{\partial v_s \partial u_s} = 0, \tag{4A-9}$$

$$\frac{\partial^2 H_a}{\partial x_a \partial v_s} = \frac{\partial^2 H_a}{\partial v_s \partial x_a} = 0, \tag{4A-10}$$

Consequently, the matrix can be reduced to

$$\begin{aligned}\delta^2 Z^* &= \sum_{\forall a} \int_0^T \left[\delta^T e_a(s) \left(\frac{\partial^2 H_a}{\partial u_s^2} \right) \delta e_a(s) + \delta^T x_a(s) \left(\frac{\partial^2 H_a}{\partial x_a \partial u_s} \right) \delta e_a(s) \right] ds \\ &+ \sum_{\forall a} \int_0^T \left[\delta^T e_a(s) \left(\frac{\partial^2 H_a}{\partial u_s \partial x_a} \right) \delta x_a(s) + \delta^T x_a(s) \left(\frac{\partial^2 H_a}{\partial x_a^2} \right) \delta x_a(s) \right] ds\end{aligned}, \tag{4A-11}$$

or further to the matrix form as

$$\delta^2 Z^* = \sum_{\forall a} \int_0^T \left\{ \begin{bmatrix} \delta^T e_a(s) & \delta^T x_a(s) \end{bmatrix} \begin{bmatrix} \frac{\partial^2 H_a}{\partial u_s^2} & \frac{\partial^2 H_a}{\partial x_a \partial u_s} \\ \frac{\partial^2 H_a}{\partial u_s \partial x_a} & \frac{\partial^2 H_a}{\partial x_a^2} \end{bmatrix} \begin{bmatrix} \delta e_a(s) \\ \delta x_a(s) \end{bmatrix} \right\} ds \quad . \tag{4A-12}$$

For the second order variation to be always positive, the matrix $\begin{bmatrix} \frac{\partial^2 H_a}{\partial u_s^2} & \frac{\partial^2 H_a}{\partial u_s \partial x_a} \\ \frac{\partial^2 H_a}{\partial x_a \partial u_s} & \frac{\partial^2 H_a}{\partial x_a^2} \end{bmatrix}$ has to be positive semidefinite. A necessary and sufficient condition for this is that the determinant of the matrix has to be non-negative.

The determinant is calculated as

$$\begin{bmatrix} \frac{\partial^2 H_a}{\partial u_s^2} & \frac{\partial^2 H_a}{\partial u_s \partial x_a} \\ \frac{\partial^2 H_a}{\partial x_a \partial u_s} & \frac{\partial^2 H_a}{\partial x_a^2} \end{bmatrix} = \frac{\partial^2 H_a}{\partial u_s^2}\frac{\partial^2 H_a}{\partial x_a^2} - \frac{\partial^2 H_a}{\partial x_a \partial u_s}\frac{\partial^2 H_a}{\partial u_s \partial x_a} ,$$

$$= f''[\tau_a(s)]\frac{e_a(s)}{Q_a^2}\left\{\frac{\partial}{\partial u_s}[C_a(s)+\Psi_a(s)]\right\} - \left[(1+f'[\tau_a(s)])\frac{1}{Q_a}\right]^2 \tag{4A-13}$$

which has indeterminate sign. As a result, the matrix $\begin{bmatrix} \frac{\partial^2 H_a}{\partial u_s^2} & \frac{\partial^2 H_a}{\partial u_s \partial x_a} \\ \frac{\partial^2 H_a}{\partial x_a \partial u_s} & \frac{\partial^2 H_a}{\partial x_a^2} \end{bmatrix}$ is not positive semi-definite although it is symmetric. Hence, the necessary condition in Proposition 4.1 may not be sufficient for a solution of the optimization problem (4.1), and it may imply that multiple dynamic system optimal solutions can exist. Similar results have been shown by Yang and Huang (2005, p72) in static case. Nevertheless, the issue of sufficiency will still be subject to future investigation.

5. CONCLUSIONS

5.1 SUMMARY

This thesis investigates analytical dynamic system optimal assignment with departure time choice in a rigorous and original way.

In Chapter 2, this thesis starts with giving a comprehensive review on the link traffic flow and travel time models for use in dynamic traffic assignments. We summarize the requirements for a traffic or travel time model to be satisfactory for use in dynamic traffic modelling. The traffic or travel time model has to ensure non-negativity of traffic, FIFO queuing discipline, conservation of traffic, travel time-flow consistency, and causality. A review of various traffic models including the wave models, the outflow traffic models, and the travel time models is given and discussed. This thesis focuses on the linear travel time models because such models have been shown to be satisfactory with respect to all requirements listed above.

In Chapter 3, we investigate the analysis and the solution algorithms for dynamic user equilibrium assignment with departure time choice. Several properties related to the assignments, including the requirements on the travel cost functions for an equilibrium solution to exist and the relationship between the travel cost and demand, are established. Numerical examples and the characteristics of the assignment results associated with different choices of travel time models and discretizations are discussed.

In Chapter 4, we analyse dynamic system optimal assignment by exploiting a state-dependent optimal control formulation. In the formulation, a fixed volume of traffic is assigned to departure times and routes such that the total system travel cost is minimized. The analysis shows that dynamic system optimal assignment can be expressed as an equivalent dynamic user equilibrium assignment with additional components of the travel cost introduced for each traveller. In general, to operate the transport system optimally, each traveller in the system is expected to pay an additional amount of cost or toll which is equal to the externality that he/she imposes on the system. To analyse and calculate this externality, we develop a novel sensitivity analysis of travel cost. Solution algorithms are developed to implement this sensitivity analysis and solve dynamic system optimal assignment. Numerical examples are given and the characteristics of the results are discussed. It is observed that congestion is generally inevitable even in dynamic system optimum in which the durations of assignments are lengthened and the inflows are spread out to minimize the intensity of congestion. This finding suggests that much study in the literature of dynamic system optimal assignment based on the deterministic queuing model is not generally applicable. We also investigate an alternative solution algorithm and the effect of time discretization on the quality of the assignments. Interestingly, it is found that the solution algorithm performs better with coarser discretization. This could be a consequence of the finer discretization giving rise to a greater number of sub-problems, making the assignment more difficult to solve accurately.

Calculating dynamic system optimal assignment and the associated optimal toll can be too difficult for implementation in practice or even for research purpose. In the view of this, we propose some practical tolling strategies for managing dynamic network traffic, which are the uniform tolls and the congestion based tolls. These tolling strategies are compared with the dynamic system optimal toll and hence their efficiencies can be evaluated accordingly. The uniform tolls can gain around 50% of the delay reduction that would have been achieved by implementing the fully time-varying tolls. This finding is consistent with the earlier one reported by Laih (1994). The congestion based tolls are more effective in managing dynamic network traffic, which in general yield an efficiency of around 90%.

In deriving the tolling strategies, it is assumed that we have an exact model for the underlying traffic behaviour. In reality, we do not have such information so that the robustness of a toll calculation method is an important issue to be investigated in practice. Given its effectiveness, we further investigate the robustness of the congestion based tolls with respect to the underlying traffic model adopted. It is found that the tolls calculated by using divided linear traffic models can generally give good performance according to a wide range of traffic models. The divided linear travel time models thus should receive more attention in the future research on robust dynamic traffic control strategies design.

5.2 FUTURE WORK

This section identifies several limitations of the work presented in this thesis and suggests possible future research directions.

In this thesis, the analysis and calculation are restricted to networks with origin-destination pairs that are connected with mutually distinct routes consisting of single links. Extending the current study to general networks is important future research direction for practice. In the case of networks that have origin-destination pairs with overlapping routes, traffic entering the network during the journey time of a traveller from other origins downstream can influence the travel time of travellers from its upstream. As a result, some special computational technique, for example Gauss-Seidel relaxation (see Sheffi, 1985; Patriksson, 1994), seems likely to be required. The basic idea of such relaxation scheme is to decompose the assignment problem for networks with overlapping routes connecting multiple origin-destination pairs into several sub-problems. In each sub-problem, we calculate the assignments for one origin-destination pair, and temporarily neglect the influences from the flows between other origin-destination pairs. When dynamic user equilibrium or dynamic system optimum is reached for the current origin-destination pair, we proceed with calculations for another pair. The procedure is repeated until equilibrium or system optimum is reached in the whole network. The relaxation scheme is not guaranteed to converge, but if it does, the solution will be the final assignment pattern (see Sheffi, 1985, p217). In case of routes with multiple links, difficulties are introduced when we have to calculate the derivatives of *route* exit time (see for example Balijepalli and Watling, 2005). As shown earlier in Proposition 4.3, changing the inflow to a link on the route during one time interval will induce perturbations in the link travel time, the link outflow, and hence the inflow to subsequent link(s) in several succeeding time intervals. Hence, the dimension of time intervals to be considered in calculating the derivatives will expand exponentially along the route. Investigating the strategies to cope with the resulting *curse of dimensionality* will be an important area of future study. Efficient computing methods for system optimal assignments in general networks will also require investigation and the work reported in this thesis provides a foundation for research along this line.

On the travellers' behaviour, this thesis supposes that all travellers have perfect information on the traffic conditions and hence the associated travel costs that they will encounter on their journeys. However, it is understood that travellers do not have such information in reality. Investigating the effects of imperfect information is certainly an important future extension.

One popular way to capture this uncertainty in travel behaviour is through adding stochastic terms in the travel cost functions to represent the uncertainties in travel information obtained by travellers (see for example, Sheffi, 1985; Lim and Heydecker, 2005; Maher, et al., 2005). In addition to the realism, such stochastic traffic assignment models have also been shown in the literature to have certain computational advantage over the deterministic ones. Several studies (see for example, Ying and Yang, 2005; Connors et al., 2007) have shown that incorporating the stochastic terms has a desirable consequence of providing smoothness and convexity to both demand and travel performance functions for analysis and solution algorithms to work with. Furthermore, this thesis considers travellers have same value of travel time and time-specific costs, while it is also not exactly the case in reality. Taking the *heterogeneity* among travellers into account is necessary for implementing equitable transport policy which is shown to be an important social concern. Transport economists revealed that *anonymous*[‡‡‡] control policies tend to benefit disproportionately those road users with a high value of time, who are typically rich (Arnott, De Palma and Lindsey, 1994, 1998). Technically, capturing the effects associated with heterogeneity introduces a number of difficulties (Newell, 1987; Arnott, De Palma and Lindsey, 1988, 1992, 1994; Yang and Meng, 1998; Lindsey, 2004) and it remains as a challenging topic in transportation research.

On capturing the behaviour of traffic flow, this thesis treats traffic as physically dimensionless in which vehicles queue vertically. However, much literature has shown that capturing the physical dimension of traffic is crucial for modelling realistic traffic behaviour (see for example Daganzo, 1998; Lo and Szeto, 2002; Szeto and Lo, 2005; Chow and Lo, 2007; Lago and Daganzo, 2007), although most models that explicitly consider the physical dimension of traffic are also shown to be difficult to apply and solve. Recently, Daganzo (2005 a, b) proposed a variational reformulation for solving kinematic wave model which is a realistic and widely accepted physical queuing model (see discussion in Section 2.3.1). This variational reformulation leads to efficient analytical solution methods for kinematic wave model. Hence, incorporating Daganzo’s (2005 a, b) work into the present framework in the thesis is an interesting future research. The outcome will be a network model that simultaneously represents both the economics of travel behaviour and the physics of traffic flow in a dynamic framework. The resulting model will be more realistic and reliable for use of transport planning, policy implementations, network design, and incident management.

[‡‡‡] Anonymous policy refers to the policy which is imposed identically on all individuals.

REFERENCES

Addison, JD and Heydecker, BG (1995) Traffic models for dynamic assignment. NH Gartner and G Improta, eds. *Urban Traffic Networks: dynamic flow modelling and control*. Springer, London, 213-231.

Addison, JD and Heydecker, BG (1998) System optimal assignment with departure time choice. JD Griffiths, ed. *Mathematics in Transport Planning and Control*. Pergamon, Oxford, 145-154.

Adler, JL (2001) Investigation the learning effects of route guidance and traffic advisories on route choice behaviour. *Transportation Research*, **9C**(1), 1-14.

Akamatsu, T (2001) An efficient algorithm for dynamic traffic equilibrium assignment with queues. *Transportation Science*, **35**, 389-404.

Akamatsu, T and Heydecker, BG (2003) Detecting capacity paradoxes in dynamic traffic assignment in saturated networks. *Transportation Science*, **37,** 123-138.

Akamatsu, T and Kuwahara, M (1999) A capacity increasing paradox for a dynamic traffic assignment with departure time choice. A. Ceder, ed. *Transportation and Traffic Theory*. Elsevier, Oxford, 301-324.

Arnott, R, de Palma, A and Lindsey, R (1988) Schedule delay and departure time decisions with heterogeneous commuters. *Transportation Research Record*, **1197**, 56-67.

Arnott, R, de Palma, A and Lindsey, R (1990) Departure time and route choice for routes in parallel. *Transportation Research*, **24B**(3), 209-228.

Arnott, R, de Palma, A and Lindsey, R (1992) Route choice with heterogeneous drivers and group-specific congestion costs. *Regional Science and Urban Economics*, **22**, 71-102.

Arnott, R, de Palma, A and Lindsey, R (1993) A structural model of peak-period congestion: a traffic bottleneck with elastic demand. *American Economics Review*, **83**(1), 161-179.

Arnott, R, de Palma, A and Lindsey, R (1994) The welfare effects of congestion tolls with heterogeneous commuters. *Journal of Transport Economics and Policy*, **28**(2), 139-161.

Arnott, R, de Palma, A and Lindsey, R (1998) Recent developments in the bottleneck model. KJ Button and ET Verhoef, eds. *Road Pricing, Traffic Congestion and the Environment: Issues of Efficiency and Social Feasibility*. Edward Elgar, Aldershot, Hampshire, UK, 79-110.

Astarita, A (1996) A continuous time link model for dynamic network loading based on travel time functions. JB Lesort, ed. *Transportation and Traffic Theory*. Pergamon, Oxford, 79-102.

Balijepalli, NC and Watling, DP (2005) Doubly Dynamic Equilibrium Distribution Approximation Model for Dynamic Traffic Assignment. H Mahmassani, ed. *Transportation and Traffic Theory*. Pergamon, Oxford, U.K., 741-760.

Bar-Gera, H (2002) Origin-based algorithm for the traffic assignment problem. *Transportation Science,* **36,** 398-417.

Bazaraa, M, Sherali, HD and Shetty, CM (1993) *Nonlinear programming: theory and algorithms.* (2nd edn) John Wiley and Sons, New York.

Beckmann, M, McGuire, C and Winsten, CB (1956) Studies in the Economics of Transportation. Yale University Press, New Haven.

Bell, MGH and Iida, Y (1997) Transportation network analysis. John Wiley & Sons, England.

Ben-Akiva, M, Bierlaire, M, Burton, D, Koutsopoulos, HN and Mishalani, R (2001) Network state estimation and prediction for real-time traffic management. *Network and Spatial Economics*, **1**, 293-318.

Bliemer, M (2006) Dynamic queuing and spillback in analytical multiclass dynamic traffic assignment model. *Proceedings of the 1st international symposium on dynamic traffic assignment,* Leeds, UK. 21-23 June.

Boyce, D, Lee, DH and Ran B (2001) Analytical Models of the Dynamic Traffic Assignment Problem. *Networks and Spatial Economics* **1**, 377-390.

Braess, D (1968) Ube rein paradoxen der verkehrplaning, *Unternehmensforschung*, **12**, 258-268.

Carey, M (1986) A Constraint Qualification for a Dynamic Traffic Assignment Model. *Transportation Science* **20**(1), 55-58.

Carey, M (1987) Optimal time-varying flows on congested networks. *Operations research,* **35**(1), 58-69.

Carey, M (1990) Extending and solving a multi-period congested network flow model. *Computers and Operations Research* **17**(5), 495-507.

Carey, M (1992) Non-convexity of the dynamic traffic assignment problem. *Transportation Research*, **26B**(2), 127-133.

Carey, M, Ge, YE and McCartney, M (2003) A whole-link travel time model with desirable properties. *Transportation Science*, **37**(1), 83-96.

Carey, M (2004a) Link travel times I: desirable properties. *Network and Spatial Economics*, **4**(3), 257-268.

Carey, M (2004b) Link travel times II: properties derived from traffic-flow models. *Network and Spatial Economics*, **4**(4), 379-402.

Carey, M and Srinivasan, A (1993) Externalities, average and marginal costs, and tolls on congested networks with time-varying flows. *Operations research*, **41**(1), 217-231.

Cassidy, MJ (1998) Reproducible bivariate relations in nearly stationary highway traffic. *Transportation Research,* **32B**, 49-49.

Cheng, YWT., Hui, HMH and Lo, H (2005) Report of the task force on emergency transport coordination, *Report to the Secretary for the Environment, Transport and Works,* Hong Kong.

Chow, AHF (2007) Toward a general framework for dynamic road pricing. BG Heydecker, ed. *Mathematics in Transport*. In press.

Chow, AHF and Lo, H (2007) Sensitivity analysis of signal control with physical queuing: delay derivatives and an application. *Transportation Research,* **41B**(3), 462-477.

Chu, X (1995) Endogenous trip scheduling: The Hendrickson approach reformulated and compared with the Vickrey approach. *Journal of Urban Economics*, **37**, 324-342.

Connors, RD, Sumalee, A and Watling, DP (2007) Sensitivity analysis of the variable demand probit stochastic user equilibrium with multiple user-classes. *Transportation Research*, **41B**(6), 593-615.

Daganzo, CF (1994) The cell-transmission model: a simple dynamic representation of highway traffic. *Transportation Research,* **28B**(4), 269-287.

Daganzo, CF (1995a) The cell-transmission model, part II: network traffic. *Transportation Research* **29B**(2), 79-93.

Daganzo, CF (1995b) Properties of link travel time functions under dynamic loads. *Transportation Research* **29B**(2), 95-98.

Daganzo, CF (1998) Queue spillovers in transportation networks with a route choice. *Transportation Science* **32**, 3-11.

Daganzo, CF (2005a) A variational formulation of kinematic waves: basic theory and complex boundary conditions. *Transportation Research,* **39B**(2), 187-196.

Daganzo, CF (2005b) A variational formulation of kinematic waves: solution methods. *Transportation Research,* **39B**(10), 934-950.

Dorfman, R (1969) An economic interpretation of optimal control theory, *The American Economics Review* **59**, 817-831.

Erera, AL, Daganzo, CF and Lovell, DJ (2002) The access control problem on capacitated FIFO networks with unique O-D paths is hard, *Operations Research* **50**, 736-743.

Ettema, D, Ashiru, O and Polak, JW (2006) Modelling timing and duration of activities and trips in response to road pricing policies, *Transportation Research Record* (in press).

European Commission (2001) *White paper - European transport policy for 2010: time to decide*. Office for official publications of the European communities, Luxembourg.

Evans, SP (1976) Derivation and analysis of some models for combining trip distribution and assignment. *Transportation Research* **10**, 37–57.

Friesz, TL, Luque, FJ, Tobin, RL and Wie, BW (1989) Dynamic network assignment considered as a continuous time optimal control problem. *Operations Research* **37**(6), 893-901.

Friesz, TL, Bernstein, D, Smith, TE, Tobin, RL and Wie, BW (1993) A variational inequality formulation of the dynamic network user equilibrium problem. *Operations Research* **41**(1), 179-191.

Friesz, TL, Bernstein, D and Stough, R (1996) Dynamic systems, variational inequalities and control theoretic models for predicting urban network flows. *Transportation Science,* **30**(1), 14-31.

Friesz, TL and Bernstein, D (2000) Analytical dynamic traffic assignment models. D.A. Hensher amd K.J. Button, eds. *Handbook of transport modelling*, 181-195.

Friesz, TL, Bernstein, D, Suo, Z and Tobin, RL (2001) Dynamic network user equilibrium with state-dependent time lags. *Networks and Spatial Economics* **1**(3/4), 319-347.

Friesz, TL, Bernstein, D and Kydes, N (2004) Deterministic Flow Routing and Oligopolistic Competition in Dynamic Data Networks: Modelling and Numerical Solution Using Multigrid Optimization Techniques, *Networks and Spatial Economics*, **4**(1), 55-73.

Friesz, TL and Mookherjee, R (2006) Solving the Dynamic Network User Equilibrium Problem with State-Dependent Time Shifts, *Transportation Research*, **40B**(3), 207 – 229.

Frost, R (1920) Mountain Interval. New York: Henry Holt and Company.

Ghali, MO and Smith, MJ (1995) A model for the dynamic system optimum traffic assignment problem. *Transportation Research,* **29B**(3), 155-170.

Han, SJ (2000) Dynamic traffic assignment techniques for general road networks. PhD thesis, University of London, U.K.

Han, SJ and Heydecker, BG (2006) Consistent objective and solutions of dynamic user equilibrium models. *Transportation Research* **40B**(1), 16-34.

Hau, TD (1998) Congestion pricing and road investment. K-J Button and E-T Verhoef, eds. *Road Pricing, Traffic Congestion and the Environment: Issues of Efficiency and Social Feasibility.* Edward Elgar, Aldershot, Hampshire, U.K., 39-78.

Hendrickson, C and Kocur, G (1981) Schedule delay and departure time decisions in a deterministic model. *Transportation Science* **15**(1), 62-77.

Hendrickson, C and Plank, E (1984) The Flexibility of Departure Times for Work Trips. *Transportation Research* **18A**(1), 25-36.

Heydecker, BG (1983) Mathematical analysis of responsive priority for buses at traffic signals. PhD thesis, University of London, U.K.

Heydecker, BG (2002a) Influence of peak spreading and departure time choice on Evaluation of ITS Measures. *Journal of Advanced Transportation: ITS Special Issue,* **36**(3), 293 - 322.

Heydecker, BG (2002b) Dynamic equilibrium network design. MAP Taylor, ed. *Transportation and Traffic Theory*. Pergamon, Oxford, 349-70.

Heydecker, BG (2005) Equilibrium network modelling. Lecture note T14 of Intercollegiate MSc Course in Transport, University of London (Unpublished).

Heydecker, BG and Addison, JD (1996) An exact solution of dynamic equilibrium. JB Lesort, ed. *Transportation and Traffic Theory*. Pergamon, Oxford, 359-384.

Heydecker, BG and Addison, JD (1998) Analysis of traffic models for dynamic equilibrium traffic assignment. MGH Bell, ed. *Transportation Networks: Recent methodological Advance.* Pergamon, Oxford, 35-49.

Heydecker, BG and Addison, JD (2005) Analysis of dynamic traffic equilibrium with departure time choice. *Transportation Science,* **39**(1), 39–57.

Heydecker, BG and Addison, JD (2006) Analysis of dynamic traffic assignment. *Proceedings of the 1st international symposium on dynamic traffic assignment,* Leeds, UK. 21-23 June.

Heydecker, BG and Verlander, NQ (1999) Calculation of dynamic traffic equilibrium assignments. *Proceedings of the European Transport Conference, Seminar F, P343,* London, 79-91.

Ho, JK (1980) A successive linear optimization approach to the dynamic traffic assignment problem. *Transportation Science,* **14**, 265-305.

Holden, DJ (1989) Wardrop's third principle: urban traffic congestion and traffic policy. *Journal of Transport Economics and Policy,* **23**, 239-262.

Huang, HJ and Lam WHK (2002) Modelling and solving dynamic user equilibrium route and departure time choice problem in network with queues. *Transportation Research*, **36B**(3), 253-273.

Hurdle, VF (1974) The effect of queuing on traffic assignment in a simple road network. DJ Buckley, ed. *Transportation and Traffic Theory*. Elsevier Science, Amsterdam, 519-540.

Hurdle, VF (1986) Technical note on a paper by Andre de Palma, Moshe Ben-Akiva, Claude Lefevre, and Nicolaos Litinas entitled "stochastic equilibrium model of peak period traffic congestion," *Transportation Science*, **20**(4), 287-289.

Institution of Highways and Transportation (1997) *Transport in the Urban Environment*.

Janson, BN (1991) Dynamic traffic assignment for urban road networks. *Transportation Research*, **25B**(2-3), 143-161.

Janson, BN and Robles, J (1993) Dynamic traffic assignment with arrival time costs. CF Daganzo, ed. *Transportation and Traffic Theory*. Pergamon, Oxford, 127-146.

Kamien, MI and Schwartz, NL (1991) *Dynamic Optimization: The Calculus of Variations and Optimal Control in Economics and Management.* Second Edition, North-Holland, New York.

Kelly, FP (1991) Network routing. *Philosophical Transaction: Physical Sciences and Engineering, Royal Society of London,* **337A**, 343-367.

Kimber, RM and Hollis, EM (1979) Traffic queues and delays at road junctions. *TRRL Report LR 909*. Transport and Road Research Laboratory, Crowthorne.

Knight, FH (1924) Some fallacies in the interpretation of social cost. *Quarterly Journal of Economics*, **38**, 582 – 606.

Knödel , W (1969) *Graphentheoretische methoden und ihre Anwendungen*, 56-59. Heidelberg, Berlin.

Kuwahara, M (2001) A theoretical analysis on dynamic marginal cost pricing. *Proceedings of the Sixth Conference of Hong Kong Society for Transportation Studies*, 28-39.

Lago, A and Daganzo, CF (2007) Spillovers, merging traffic and the morning commute. *Transportation Research*, **41B**(6), 670-683.

Laih, CH (1994) Queuing at a bottleneck with single- and multi-step tolls. *Transportation Research*, **28A**(3), 197–208.

Lam, WHK and Huang, HJ (1995) Dynamic user optimal traffic assignment model for many to one travel demand. *Transportation Research*, **29B**(4), 243-259.

Lee, S (1995) Mathematical programming algorithms for equilibrium road traffic assignment. PhD thesis, University of London, U.K.

Lighthill, MJ and Whitham, JB (1955) On kinematic waves: II. A theory of traffic flow on long crowded roads. *Proceedings of the Royal Society* **229A**, 317-345.

Lim, Y and Heydecker, BG (2005) Dynamic departure time and stochastic user equilibrium assignment. *Transportation Research* **39B**(2), 97-118.

Lindsey, R (2004) Existence, uniqueness, and trip cost function properties with multiple user classes. *Transportation Science*, **38**(3), 293 – 314.

Lo, H (1999) A cell-based variational inequality formulation of the dynamic user optimal assignment problem. A Cedar, ed. *Transportation and Traffic Theory*, 327-350.

Lo, H and Szeto, WY (2002) A cell-based variational inequality formulation of the dynamic user optimal assignment problem. *Transportation Research* **36B**(5), 421-443.

Lovell, DJ and Daganzo, CF (2000) Access control on networks with unique origin-destination paths. *Transportation Research* **34B(**3), 185-202.

Luenberger, DG (1984) *Linear and nonlinear programming.* Addison-Wesley, Menlo Park, California.

Maher, MJ, Stewart, K. and Rosa, A (2005) Stochastic social optimal traffic assignment. *Transportation Research* **39B(**8), 753-767.

Mahmassani, H (2001) Dynamic network traffic assignment and simulation methodology for advanced management applications. *Network and Spatial Economics*, **1**, 267-292.

Mahmassani, H and Herman, R (1984) Dynamic user equilibrium departure time and route choice on idealized traffic arterials. *Transportation Research*, **18**, 362-384.

Merchant, DK and Nemhauser, GL (1978a) A model and an algorithm for the dynamic traffic assignment problem. *Transportation Science,***12**(3), 183-199.

Merchant, DK and Nemhauser, GL (1978b) Optimality conditions for a dynamic traffic assignment model. *Transportation Science,***12**(3), 200-207.

Murchland, JD (1970) Braess's paradox of traffic flow. *Transportation Research*, **4**(4), 391-394.

Mun, JS (2002) Dynamic traffic assignment models with departure time choice. PhD thesis, University of London, U.K.

Nagurney, A (1993) Network economics: a variational inequality approach. Kluwer Academic Publishers. Norwell, Massachusetts, USA.

Newell, GF (1987) The morning commute for nonidentical travellers. *Transportation Science*, **21**(1), 74-88.

Newell, GF (1988) Traffic flow for the morning commute. *Transportation Science*, **22**(1), 47-58.

Nie, X and Zhang HM (2005a). A comparative study of some macroscopic link models used in dynamic traffic assignment. *Networks and Spatial Economics*, **5**, 89-115.

Nie, X and Zhang HM (2005b). Delay-function-based link models: their properties and computational issues. *Transportation Research*, **39B**(8), 729-752.

Ortúzar, JD and Willumsen, LG (2001) Modelling transport (3rd Edition). John Wiley & Sons, England.

Papageougiou, M (1990) Dynamic modelling, assignment and route guidance in traffic networks. *Transportation Research* **24B**(6), 471-95.

Patriksson, M (1994) The traffic assignment problem: models and methods. VSP, The Netherlands.

Payne, HJ (1971) Models of freeway traffic and control. *Mathematical Models of Public Systems. Simulation Council Proceedings Series*, **1**(1), 51-61.

Peeta, S and Ziliaskopoulos, AK (2001) Foundations of dynamic traffic assignment: the past, the present and the future. *Networks and Spatial Economics*, **1**(3/4), 233-265.

Pigou, AC (1920) Wealth and Welfare. Macmillan, London, England.

Polak, JW and Heydecker, BG (2006) The dynamic assignment of tours in congested networks with pricing. *Proceedings of the 1st International Symposium on Dynamic Traffic Assignment*, Leeds, 21-23 June.

Rajamani, R and Shladover, SE (2001) An experimental comparative study of autonomous and co-operative vehicle-follower control systems. *Transportation Research*, **9C**(1), 15-31.

Ran, B and Boyce, DE (1996) Modelling dynamic transportation networks: an intelligent transportation system oriented approach (2nd Edition). Springer-Verlag, Heidelberg.

Richards, PI (1956) Shockwaves on the highway. Operations Research, **4**, 42–51.

Sharpe, JM (1993) Demand management: The Cambridge approach. *Transportation Studies*, Cambridgeshire County Council, Cambridge, UK.

Sheffi, Y (1985) Urban transportation networks: equilibrium analysis with mathematical programming methods. Englewood Cliffs, NJ: Prentice-Hall.

Small, K (1982) The scheduling of consumer activities: work trips. *American Economics Review*, **72**(3), 467-479.

Small, K (1992) *Urban Transportation Economics.* Harwood Academic Publishers.

Small, K and Gomez-Ibañez, JA (1998) Road pricing for congestion management: the transition from theory to policy. K-J Button and E-T Verhoef, eds. *Road Pricing, Traffic Congestion and the Environment: Issues of Efficiency and Social Feasibility.* Edward Elgar, Aldershot, Hampshire, U.K., 213-246.

Smith, MJ and Ghali, MO (1990) Dynamic traffic assignment and dynamic traffic control. M Koshi, ed. *Transportation and Traffic Theory*. Elsevier, New York, 223-263.

Szeto, WY and Lo, H (2005) Non-equilibrium dynamic traffic assignment. H. Mahmassani, ed. *Transportation and Traffic Theory*. Pergamon, Oxford, 427-445.

Taylor, NB (2003) The CONTRAM dynamic traffic assignment model. *Network and Spatial Economics*, **3**, 297-322.

Tobin, RL (1993) Notes on flow propagation constraints. *Working paper*, **93-10**, Network Analysis Laboratory, George Mason University, USA.

Transport for London (2007) London congestion charging, http://www.cclondon.com/index.shtml. Accessed 11 February, 2007.

Vickrey, WS (1963) Pricing in urban and suburban transport. *American Economics Review*: *Papers and Proceedings*, **53**(2), 452-65.

Vickrey, WS (1969) Congestion theory and transport investment. *American Economics Review*, **59**, 251-260.

Wardrop, JG (1952) Some theoretical aspects of road traffic research. *Proceedings of the Institute of Civil Engineers Part II*, 325-378.

Wie, BW and Tobin, RL (1998) Dynamic congestion pricing models for general traffic networks. *Transportation Research*, **32B**(5), 313-327.

Wie, BW, Tobin, RL and Friesz, TL (1994) The augmented Lagrangian method for solving dynamic network traffic assignment models in discrete time. *Transportation Science*, **28**(3), 204-220.

Wie, BW, Tobin, RL, Bernstein, D. and Friesz, TL (1995a) A comparison of system optimum and user equilibrium traffic assignments with schedule delay. *Transportation Research*, **3C**(6), 389-411.

Wie, BW, Tobin, RL, Bernstein, D. and Friesz, TL (1995b) A discrete time, nested cost operator approach to the dynamic network user equilibrium problem. *Transportation. Science*, **29**(1), 79-92.

Wu, JH, Chen, Y and Florian, M (1998). The continuous dynamic network loading problem: a mathematical formulation and solution method. *Transportation Research*, **32B**(3)**,** 173-187.

Xu, YW, Wu, JH, Florian, M, Marcotte, P and Zhu, DL (1999), Advances in the continuous dynamic network loading problem, *Transportation Science* **33**(4), 341–353.

Rubio-Ardanaz, JM, Wu, JH, and Florian, M (2003), Two improved numerical algorithms for the continuous dynamic network loading problem, *Transportation Science* **37**, 171–190.

Yagar, S (1971) Dynamic traffic assignment by individual path minimization and queuing. *Transportation Research* **5**(3), 179-196.

Yang, H and Huang, HJ (1997) Analysis of the time-varying pricing of a bottleneck with elastic demand using optimal control theory. *Transportation Research* **31B**(6), 425-440.

Yang, H and Huang, HJ (2005) *Mathematical and Economic Theory of Road Pricing.* Elsevier, The Netherlands.

Yang, H and Meng, Q (1998) Departure time, route choice and congestion toll in a general queuing network with elastic demand. *Transportation Research* **32B**(4), 247-260.

Ying, JQ and Yang, H (2005) Sensitivity analysis of stochastic user equilibrium flow in a bi-modal network with application to optimal pricing. *Transportation Research* **39B**(9), 769-795.

Zhu, D. and Marcotte. P (2000) On the existence of solutions to the dynamic user equilibrium problem. *Transportation Science* **34**(4), 402-414.

Ziliaskopoulos, AK (2000) A linear programming model for the single destination system optimum dynamic traffic assignment problem. *Transportation Science*, **34**(1), 1-12.