Transportation Research Part B 43 (2009) 267–284

Contents lists available at ScienceDirect

Transportation Research Part B

journal homepage: www.elsevier.com/locate/trb

On the morning commute problem in a corridor network with multiple bottlenecks: Its system-optimal traffic flow patterns and the realizing tolling scheme

Wei Shen, H.M. Zhang *

Department of Civil and Environmental Engineering, University of California, Davis, CA 95616, USA

A R T I C L E I N F O

Article history:
Received 15 November 2006
Received in revised form 10 July 2008
Accepted 10 July 2008

Keywords:
Morning commute
System-optimum
Dynamic congestion pricing
Corridor control

A B S T R A C T

There have been numerous studies of the morning commute problem in a network with a single route or parallel routes with a single bottleneck on each route. Most congested networks, however, often contain more than one congestion spot along each route. In such networks, it is usually difficult to derive analytically their system-optimal traffic patterns and the tolls that realize them. In this paper, we study the morning commute problem in such a network with certain special topological features – a freeway with multiple entry/exit ramps and a surface street grid with large capacities. For this type of networks, we investigated the basic characteristics of their optimal dynamic traffic patterns and the corresponding tolling scheme, for which a graphical solution procedure was also developed. In this network, we found that at system-optimum: (1) the aggregate traffic flow on the freeway has a staircase temporal profile, and piecewise linear dynamic tolls can be imposed on a subset of ramps to achieve it; (2) among all the off-ramps in use, the ones closer to the destination are being tolled longer with higher maximum toll charges than the ones farther away from the destination; and (3) among all the on-ramps in use, the ones with larger cumulative volume-to-capacity ratios are being tolled longer with higher maximum toll charges. Some practical implications of these findings to corridor traffic management were also discussed.

© 2008 Elsevier Ltd. All rights reserved.

1. Introduction

It is well-known that the existence of congestion externality often leads to inefficient use of public roads and this can be remedied by marginal cost pricing. By internalizing the congestion externality, i.e., charging each traveler a toll equal to the additional congestion cost she imposed on all other travelers, the *system-optimal* traffic flow pattern is realized as a *Wardroppian user equilibrium state* in terms of the internalized travel cost (Beckmann, 1965).

The congestion pricing problem has been extensively studied in the context of static transportation networks (i.e., networks with time invariant traffic flow conditions) (e.g., Dafermos and Sparrow, 1971; Netter, 1971; Yang and Huang, 1998). The pricing scheme that realizes a system-optimal state under user-optimal (or selfish) routing decisions can be derived based on a link traffic model capturing the average congestion effect. Unfortunately, such a static analysis is often *inappropriate* and *inadequate* when applied to networks with fluctuating traffic demand over time. First, using the *average congestion effect* to calculate externality is somewhat questionable. As shown by Carey and Srinivasan (1993), congestion

* Corresponding author. Tel.: +1 530 754 9203; fax: +1 530 752 7872.
E-mail address: hmzhang@ucdavis.edu (H.M. Zhang).

0191-2615/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.trb.2008.07.004

externality depends not only on the intensity of congestion, but also on the direction of its change. In networks with fluctuating traffic flow, the externality derived with a static method, averaging out the effect of queue formation and dissipation, can substantially deviate from the true externality. Furthermore, the inefficient *temporal distribution* of traffic is often much more critical than the inefficient *spatial distribution* of traffic in these networks with a limited number of alternative routes (Arnott et al., 1990). A static pricing scheme, without distinguishing the difference of congestion externality in different time periods, cannot influence travelers' departure time choice.

The aforementioned limitations of static congestion pricing schemes call for a time-dependent modeling framework. Unfortunately, deriving the optimal dynamic tolls in a general network is much more challenging than obtaining tolls in the corresponding static network. This is mainly because its underlying problem, the system-optimal dynamic traffic assignment (SO-DTA) problem, is difficult to solve due to the complexities in modeling dynamic traffic flow in a road network. One of the pioneer studies to address the dynamic congestion pricing problem in general networks is due to Carey and Srinivasan (1993). In that work, the SO-DTA problem with given time-dependent travel demands was formulated as a link-based convex mathematical program based on a relaxed[1] exit flow function from the works of Merchant and Nemhauser (1978a,b). The relationship between the externalities and the dual variables of the constraints was identified by analyzing the optimality conditions of the mathematical program. Wie and Tobin (1998) later carried out a similar study with a new traffic flow model in which traffic on each link first travels to the end of the link at free-flow speed, then queues up at the end of the link and discharges according to rates given by the exit flow function, thereby alleviating the *instantaneous flow transition*[2] problem in the initial formulation of Merchant and Nemhauser (1978a,b). Chang et al. (1988) and Yang and Meng (1998) transformed the SO-DTA problem into a static system-optimal traffic assignment problem, using a space–time expansion network (STEN) to endogenously represent the bottleneck model of Vickrey (1969). However, this type of STEN gives rise to some new problems, one of which is the high computational overhead associated with network expansions. Ziliaskopoulos (2000) formulated an SO-DTA model as a link-based linear program based on the relaxed cell transmission model (Daganzo, 1994, 1995) to consider link interactions. The connections between the system marginal cost and the dual variables of the constraints were also explored in that paper. A critical issue concerning all these link-based formulations of the SO-DTA problem is that to derive the optimal traffic pattern and the realizing tolls, they require to solve a large-scale mathematical program or optimal control problem which can easily involve millions of variables and constraints in real-sized applications. To overcome this drawback, some scholars (e.g., Ghali and Smith, 1995; Peeta and Mahmassani, 1995) suggested to use a path-based formulation, which allows the usage of a traffic simulator to: (1) construct the mapping between path flow and path travel cost and (2) evaluate marginal cost approximately. However, as shown by Shen et al. (2007), due to the non-additivity property of and the discontinuity in path marginal cost, deriving a link-based tolling scheme from a path-based formulation is tricky.

As our review indicates, existing SO-DTA formulations and their congestion pricing analyses have a number of deficiencies when applied to general networks. This, however, does not mean that it is not possible to obtain in a precise and efficient manner the optimal dynamic traffic pattern and its corresponding tolling scheme for some specific networks. In practice, many corridor networks, where a freeway and a few alternative arterial routes provide the infrastructure for commuting from suburbs to downtown, share certain common topological characteristics that can be exploited to overcome the difficulties faced in the study of general road networks. This type of network-specific analysis can be traced back to the seminal work of Vickrey (1969) and Hendrickson and Kocur (1981), in which the morning commute problem with departure time choice was studied in a single route with one bottleneck. Among this class of work, Newell (1987) considered heterogeneous travelers in the same network setting; Arnott et al. (1990) generalized the analysis to networks with parallel routes; and recently, Munoz and Laval (2006) proposed a graphical solution method to obtain the optimal route diversion strategies for a corridor network consisting of a freeway and a surface street grid for given time-dependent demand. All of these network-specific analyses provide valuable insights to the design of time-dependent congestion pricing schemes in practice, but are all limited to the case where *only one bottleneck is present on each route*. This restriction makes their results not applicable to corridor networks commonly found in practice where travelers experience more than one congestion spot on their way to the destination.

In this paper, we study the morning commute problem in *a corridor network with multiple bottlenecks along the routes* by characterizing its system-optimal traffic pattern and the tolling scheme that realizes it. Our analysis considers both departure time and route choices with a similar network used in Munoz and Laval (2006), with the following key differences:

- Unlike in Munoz and Laval (2006), our study considers capacity constraints on both on- and off-ramps. This is more realistic since queues often form at ramps due to metering at on-ramps or traffic signals at the downstream end of off-ramps.
- Unlike in Munoz and Laval (2006), our study considers both departure time and route choices. Consequently, the tolling scheme derived in this study is particularly useful for spreading the peak during the morning commute time period to avoid over-concentration of travel demand in any particular time period.

[1] Relaxed in the sense that the *equality* defining the relationship between the flow that *will actually* exit a link and the flow that *can* exit a the link is replaced by an *inequality*. This relaxation makes the model convex, but may also cause flows to be artificially held at links.

[2] This refers to the anomaly that vehicles entering the link tail can instantaneously affect the outflow at the link head.

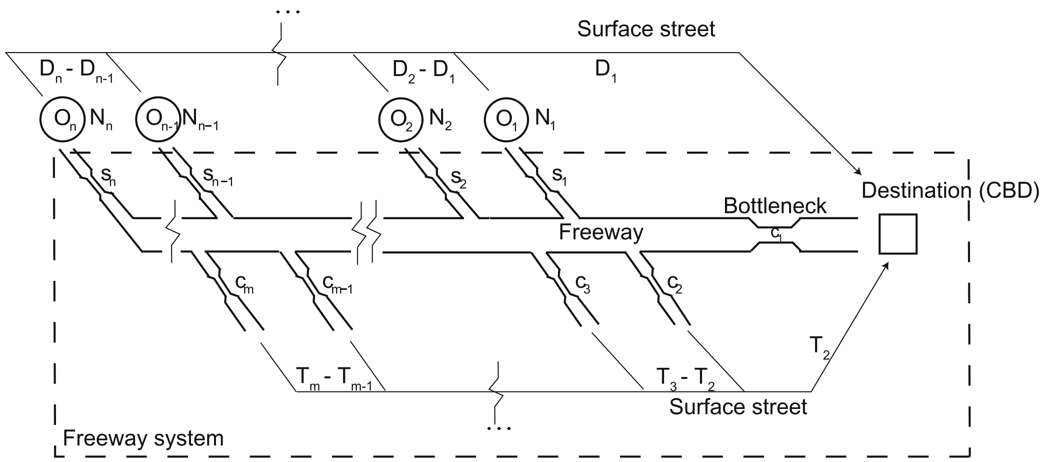

Fig. 1. The study network.

To the best of our knowledge, our study provides the first analytical results on the dynamic system-optimal traffic flow pattern and the corresponding tolling scheme in a network with multiple bottlenecks along each route. Although Kuwahara (1990) and Arnott et al. (1993) performed a network-specific equilibrium queuing analysis considering tandem bottlenecks, their studies focused on dynamic user equilibrium traffic flow patterns with no congestion pricing, and employed substantially different techniques from ours. Moreover, our analysis reveals that the system-optimal traffic flow pattern in the studied corridor network exhibits some interesting topological characteristics that lead to a graphical solution procedure. With this procedure, we can design a tolling scheme that charges time-dependent tolls on a subset of on- and off-ramps to realize the system-optimal dynamic traffic flow pattern.

The rest of the paper is organized as follows: the next section lays out the problem and describes the notations used throughout the paper. Section 3 constructs the optimality conditions, based on which the basic features of the optimal traffic and toll patterns are identified. These features lead to the development of a graphical solution procedure for obtaining the optimal traffic and toll patterns in Section 4. Finally, Section 5 concludes the paper with a summary of key findings and their implications to congestion pricing in practice.

2. The morning commute problem in a corridor with multiple bottlenecks

We consider a corridor network consisting of a freeway with n on-ramps, m off-ramps, a single destination at the end of the freeway, and a surface street grid directly connecting each on-ramp (here we assume trips originated from the tail node of an on-ramp link) and each off-ramp to the destination (CBD) (Fig. 1). For the purpose of exposition, both on-ramps and off-ramps are arranged in an ascending order starting from the one closest to the destination, and the freeway link incident to the destination is regarded as a special off-ramp, and indexed as 1. A bottleneck with capacity c_1 is located on the freeway just before reaching the destination. The capacity of other off-ramps $j = 2, \dots, m$ are assumed to be c_j, respectively. The capacity of on-ramps $i = 1, \dots, n$ are assumed to be s_i, respectively. It is also assumed the surface streets have sufficient capacity so that traffic experiences no congestion on them. Without loss of generality, the sum of the capacities of any subset of the on-ramps is assumed to not equal the sum of capacities of any subset of the off-ramps. Evidently, in practice ramp capacities can always be perturbed slightly to meet this requirement. The travel time from on-ramp i taking the surface street to the destination is denoted as D_i, and the travel time from off-ramp j taking the surface street to the destination is denoted as T_j. In accord with the network topology, the travel time on a surface street connecting a ramp closer to the destination is assumed to be shorter than the travel time on a surface street connecting a ramp farther away from the destination. Compared to the travel times spent on surface streets, the travel times spent on the freeway mainline is assumed to be negligible.[3]

As mentioned earlier, each origin is connected by a single on-ramp where travelers start their trips. During the morning peak period, the total number of travelers from origin i is $N_i, i = 1, \dots, n$. All the travelers are assumed to have the same desired arrival time at the destination, which is taken as $t = 0$. Travel costs consist of two parts: travel time cost and schedule delay cost: arriving at the destination earlier or later than desired. For simplicity, the same assumption used in Arnott et al. (1993) – that late arrival is not permitted – is adopted in this study. For any arrival time $t \leqslant 0$ at the destination, the schedule delay cost (converted into travel time units) is assumed to be linearly decreasing in the rate of $\alpha < 1$. Namely, a traveler's total cost is

[3] It is noted that zero free-flow freeway travel times do not affect the solution in terms of flow distribution.

$$\text{Total travel cost} = \begin{cases} \text{travel time} + \alpha[0 - \text{arrival time tick}] & \text{if arrival time tick} \leqslant 0 \\ \infty & \text{otherwise} \end{cases}$$

It is noted that no late arrival makes it easy to present the analytical results, but do not change the fundamental nature of the problem. A similar analysis incorporating both early and late arrivals can easily be carried out following the steps outlined in this paper. Throughout the paper, the following definition is used to facilitate the discussion.

Definition 1. The subnetwork made up of: (1) all the on-ramps, (2) the freeway mainline, (3) all the off-ramps, and (4) the surface streets connecting the off-ramps with the destination is defined as the freeway system (Fig. 1).

A traveler can choose the surface street or *freeway system* to reach her destination, but once she enters a surface street, we assume that she does not go back to travel on the freeway.

By this definition, each traveler's decision process can be split into two stages:

- First, at the origin, she decides whether to take the surface street directly leading to the destination or to enter the *freeway system*.
- Second, if the *freeway system* is taken, she chooses the departure time and determine which off-ramp to exit. Note that since surface streets are assumed to be always free of congestion, travelers choosing a surface street at the first stage always choose a departure time which can guarantee punctual arrival at the destination.

Moreover, we introduce the following notations to be used in the rest of this paper:

I	The index of the first on-ramp that has flows during the morning peak
J'	The index of the last off-ramp that has flows during the morning peak
J	The index of the last off-ramp downstream of on-ramp I
$v_{ij}(t)$	Number of travelers departing at time t and taking on-ramp i and off-ramp j
$v_{i\bullet}(t)$	Number of travelers departing at time t and taking on-ramp i, $v_{i\bullet}(t) = \sum_{j=1}^{m} v_{ij}(t)$
$v_{\bullet j}(t)$	Number of travelers departing at time t and desiring to take off-ramp j, $v_{\bullet j}(t) = \sum_{i=1}^{n} v_{ij}(t)$
$v_{\bullet}(t)$	Total number of travelers departing at time t, $v_{\bullet}(t) = \sum_{i=1}^{n}\sum_{j=1}^{m} v_{ij}(t)$
V_{ij}	Number of travelers taking on-ramp i and off-ramp j, $V_{ij} = \int_{-\infty}^{0} v_{ij}(t)\mathrm{d}t$
$V_{i\bullet}$	Number of travelers taking on-ramp i, $V_{i\bullet} = \sum_{j=1}^{m} V_{ij}$
$V_{\bullet j}$	Number of travelers taking off-ramp j, $V_{\bullet j} = \sum_{i=1}^{n} V_{ij}$
$V_{\bullet}$	Total number of travelers using the freeway system, $V_{\bullet} = \sum_{i=1}^{n}\sum_{j=1}^{m} V_{ij}$

Note: the number of travelers is always measured at the entrance of the on-ramps.

With these definitions and notations, we can then set up a mathematical program to model the morning commute problem as in (e.g., Vickrey, 1969; Hendrickson and Kocur, 1981; Newell, 1987; Arnott et al., 1990), and derive: (1) the optimal departure time and route choice patterns and (2) the optimal road tolls to realize the optimal traffic flow pattern. These are the topics of the next section.

3. The essential features of the optimal traffic and toll patterns

Corresponding to each traveler's two stage travel decision process, we carry out our analysis in two steps:

(1) At the first step, we decide for each origin i the total number of travelers $V_{i\bullet}$ using the freeway system.
(2) At the second step, for given $\{V_{i\bullet}\}$ we decide the optimal traffic flow distribution on the freeway system.

Since the optimal tolling scheme primarily depends on the optimal traffic flow distribution on the freeway system, the remainder of this section focuses on the analysis of the freeway system, assuming $\{V_{i\bullet}\}$ as given. How to determine the demand split between the surface street and the freeway system for each origin will be explained in Section 4.

3.1. Preliminaries

This section presents a series of propositions describing certain important properties of the optimal traffic flow pattern of the freeway system, which will serve as the building blocks in deriving the analytical representation of this pattern.

Proposition 1. *At dynamic system-optimum with departure time choice, there is no queue within the freeway system.*

Proof. It can be easily shown that if a traffic flow pattern involves queues within the freeway system, we can always have travelers adjust their departure to later times to eliminate the queues without increasing both schedule delay and travel time costs. $\square$

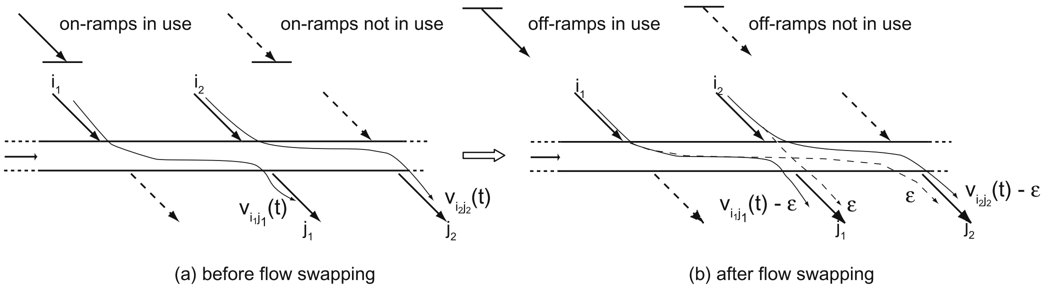

Fig. 2. An illustration of flow swapping.

Proposition 2. *At dynamic system-optimum with departure time choice, swapping the path flows over any two on-ramps at any time t does not change the system cost, i.e., if* $v_{i_1j_1}(t) > 0, v_{i_2j_2}(t) > 0$ *and off-ramp* j_1 *and off-ramp* j_2 *are both downstream of on-ramp* i_1 *and on-ramp* i_2*, then the system cost does not change by a flow swapping of (see Fig. 2)*

$$\begin{cases} v_{i_1j_1}(t) = v_{i_1j_1}(t) - \epsilon \\ v_{i_2j_2}(t) = v_{i_2j_2}(t) - \epsilon \end{cases} \text{ and } \begin{cases} v_{i_1j_2}(t) = v_{i_1j_2}(t) + \epsilon \\ v_{i_2j_1}(t) = v_{i_2j_1}(t) + \epsilon \end{cases}, \quad \epsilon \leqslant \min\{v_{i_1j_1}(t), v_{i_2j_2}(t)\}$$

Proof. The path flow swapping does not change the link flow pattern and thus will not change the system cost. □

The possibility of flow swapping tells us that the optimal path flow pattern in the freeway system is not unique. Consequently, we shall use *link flows* rather than *path flows* to characterize the optimal traffic flow pattern in the freeway system.

Proposition 3. *At dynamic system-optimum with departure time choice, if some off-ramp* j' *has positive flow at time t, all the off-ramps downstream of off-ramp* j' *have flows equal to their bottleneck capacities, i.e.,* $v_{\bullet j'}(t) > 0 \Rightarrow \forall j < j', v_{\bullet j}(t) = s_j$*. Furthermore, if* $v_{\bullet j'}(t_1) > 0$ *and* $v_{\bullet j'}(t_2) = 0$*, then* $v_{\bullet}(t_1) > v_{\bullet}(t_2)$*.*

Proof. For a traffic flow pattern without queues in the freeway system, the total cost for a traveler departing at time t and taking off-ramp j' is $c = T_{j'} + \alpha[0 - (t + T_{j'})] = -\alpha t + (1-\alpha)T_{j'}$. If there is an off-ramp $j < j'$ with $v_{\bullet j}(t) < s_j$, she can instead takes off-ramp j and reduce the cost to $c' = -\alpha t + (1-\alpha)T_j < c$ because $1 - \alpha > 0$ and $T_j < T_{j'}$. Given that $v_{\bullet j}(t) < s_j$, no other travelers' costs will be affected by such a flow shifting. Therefore, the original traffic flow pattern cannot be optimal because the system cost can be further reduced by such a flow shifting. This is equivalent to say that at dynamic system-optimum, if off-ramp j' is being used at time t, all the off-ramps downstream of it must be fully utilized, i.e., $v_{\bullet j'}(t) > 0 \Rightarrow \forall j < j', v_{\bullet j}(t) = s_j$.

Because of this property, given two time ticks t_1 and t_2, $v_{\bullet j'}(t_1) > 0 \Rightarrow v_{\bullet}(t_1) > \sum_{j<j'} s_j$ and $v_{\bullet j'}(t_2) = 0 \Rightarrow v_{\bullet}(t_2) \leqslant \sum_{j<j'} s_j$. Hence, $v_{\bullet}(t_1) > v_{\bullet}(t_2)$. □

This proposition illustrates how off-ramps are utilized in the optimal state: upstream off-ramps are used later and waned off traffic earlier than downstream off-ramps. In other words, upstream off-ramps start to attract traffic later and have shorter duration of usage than downstream off-ramps.

Proposition 4. *At dynamicsystem-optimum with departure time choice, the off-ramps in use are all at the downstream side of the utilized on-ramps (Fig. 3a).*

Proof. Suppose off-ramp j' is upstream of on-ramp i' (Fig. 3b). It suffices to show that $V_{\bullet j'} > 0 \Rightarrow V_{i'\bullet} = 0$. Denote the sets of time intervals when $v_{\bullet j'}(t) > 0$ and when $v_{\bullet j'}(t) = 0$ as P_1 and P_2, respectively.

Obviously, for any time $t \in P_1$, $v_{i'\bullet}(t) = 0$ because otherwise the system cost can be reduced by shifting a small amount of flow in $v_{i'\bullet}(t)$ to use the surface street while letting the same amount of flow originally on off-ramp j' at the same time take the residual capacity on a downstream off-ramp left by the shifting of flow in $v_{i'\bullet}(t)$ to a surface street (Fig. 4).

For any time $t \in P_2$, $v_{\bullet j}(t) = 0\ \forall t \in P_2$ and $v_{\bullet j}(t') > 0\ \forall t' \in P_1$ imply that $\sum_{j=1}^{j'} v_{\bullet j}(t) < \sum_{j=1}^{j'} v_{\bullet j}(t')$ (Proposition 3). In addition, we just know that at time t', all the on-ramps downstream of j' have zero flows, i.e., flows $\sum_{j=1}^{j'} v_{\bullet j}(t')$ are all from

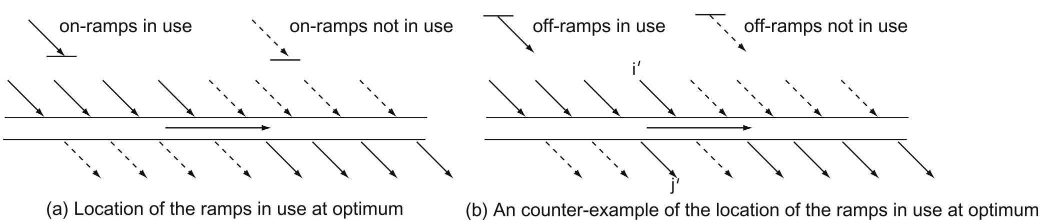

Fig. 3. Location of ramps in use at system-optimum.

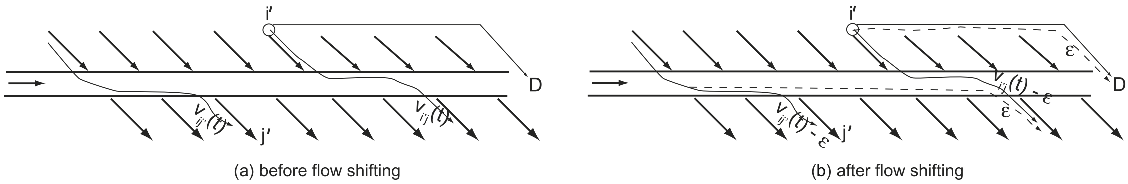

Fig. 4. A flow shifting strategy that reduces the system cost.

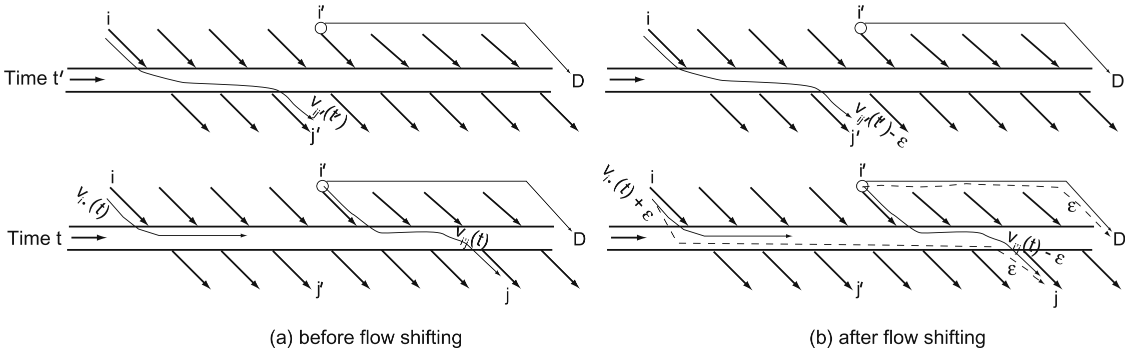

Fig. 5. A flow shifting strategy that reduces the system cost.

on-ramps upstream of j'. Therefore, there must exist at least one on-ramp i upstream of off-ramp j' such that $\sum_{j=1}^{j'} v_{ij}(t) < \sum_{j=1}^{j'} v_{ij}(t') \leqslant s_i$. Without loss of generality, we can assume that $v_{ij'}(t') > 0$. (This can always be attained by swapping path flows according to Proposition 2.) Therefore, if $v_{i'j}(t) > 0$ where j is an off-ramp downstream of on-ramp i', the system cost can be reduced by letting a small amount of flow comprising $v_{i'j}(t)$ use the surface street at time t and the same amount of flow from $v_{ij'}(t')$ originally departing at time t' and taking off-ramp j' take up the capacity slack created at time t (Fig. 5).

Therefore, $v_{i'\bullet}(t) = 0\ \forall t$, i.e., $V_{i'\bullet} = 0$. □

3.2. The formulation of the morning commute problem as a SO-DTA

The propositions previously derived make it possible to formulate the problem of deriving the system-optimal traffic pattern of the freeway system in the morning commute problem as a mathematical program.

Given $\{V_{i\bullet}\}$, Proposition 4 offers a way to simplify the topology of the freeway system subnetwork when searching for the system-optimal traffic flow distribution. That is, since all the utilized off-ramps are downstream of all the utilized on-ramps, it suffices to derive the optimal traffic flow pattern in an abstract network made up of $n - I + 1$ upstream merging branches and J downstream parallel routes (Fig. 6), where the ramp indices I and J can be easily determined for given $\{V_{i\bullet}\}$: $I = \text{argmax}\{i | V_{i\bullet} > 0\}$ and J is the index of the last off-ramp downstream of on-ramp I.

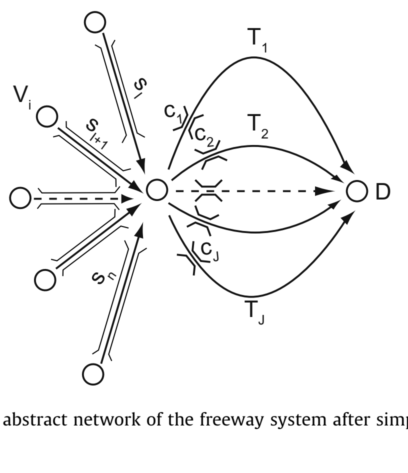

Fig. 6. The abstract network of the freeway system after simplification.

Thanks to the *no queueing* property of Proposition 1, it suffices to search the system-optimal traffic flow pattern among only those without any freeway queues, which, can be further described by the following relationships:

$$\sum_{j=1}^{J}\int_{-\infty}^{-T_j} v_{ij}(t)\mathrm{d}t = V_{i\bullet} \quad i = I,\dots,n \tag{P.1}$$

$$0 \leqslant \sum_{j=1}^{J} v_{ij}(t) \leqslant s_i \quad i = I,\dots,n,\ \ t\in(-\infty,0) \tag{P.2}$$

$$0 \leqslant \sum_{i=I}^{n} v_{ij}(t) \leqslant c_j \quad j = 1,\dots,J,\ \ t\in(-\infty,0) \tag{P.3}$$

where the first equation defines demand conservation and the other two ensure that the traffic flow at any time does not exceed ramp capacities.

For traffic flow patterns defined by (P.1)–(P.3), the total system cost does not contain queueing delay and thus can be represented as the sum of free-flow travel time and schedule delay. Namely

$$\sum_{i=I}^{n}\sum_{j=1}^{J}\int_{-\infty}^{-T_j}[T_j-\alpha(t+T_j)]v_{ij}(t)\mathrm{d}t \tag{P.0}$$

where $T_j-\alpha(t+T_j)$ is the travel cost (i.e., free-flow travel time+schedule delay) experienced by any traveler departing at time t and taking off-ramp j to reach the destination. Hence, $\sum_{i=I}^{n}\sum_{j=1}^{J}\int_{-\infty}^{-T_j}[T_j-\alpha(t+T_j)]v_{ij}(t)\mathrm{d}t$ represents the total travel cost experienced by all the travelers taking the freeway system.

In summary, the SO-DTA problem for morning commute can be formulated as the following mathematical program:

$$\mathrm{FC}(\mathbf{V}) = \min \sum_{i=I}^{n}\sum_{j=1}^{J}\int_{-\infty}^{-T_j}[T_j-\alpha(t+T_j)]v_{ij}(t)\mathrm{d}t \tag{P.0}$$

$$\text{s.t.}\quad (\lambda_i)\qquad \sum_{j=1}^{J}\int_{-\infty}^{-T_j} v_{ij}(t)\mathrm{d}t = V_{i\bullet} \quad i = I,\dots,n \tag{P.1}$$

$$(\omega_i(t))\ \ 0 \leqslant \sum_{j=1}^{J} v_{ij}(t) \leqslant s_i \quad i = I,\dots,n,\ \ t\in(-\infty,0) \tag{P.2}$$

$$(\mu_j(t))\ \ 0 \leqslant \sum_{i=I}^{n} v_{ij}(t) \leqslant c_j \quad j = 1,\dots,J,\ \ t\in(-\infty,0) \tag{P.3}$$

To analyze the fundamental features of the dynamic system-optimal traffic flow pattern, we introduce the respective multipliers λ_i, $\omega_i(t)$, $\mu_j(t)$ associated with (P.1)–(P.3), interpreted as the marginal cost of having one more traveler at on-ramp i, the marginal benefit of expanding the capacity of on-ramp i at time t, and the marginal benefit of expanding the capacity of off-ramp j at time t. The optimality conditions for (P) can thus be written as the feasibility constraints (P.1)–(P.3), together with

$$\omega_i(t) \geqslant 0,\quad s_i-\sum_{j=1}^{J} v_{ij}(t) \geqslant 0 \quad i = I,\dots,n,\ \ t\in(-\infty,0) \tag{1}$$

$$\mu_j(t) \geqslant 0,\quad c_j-\sum_{i=I}^{n} v_{ij}(t) \geqslant 0 \quad j = 1,\dots,J,\ \ t\in(-\infty,0) \tag{2}$$

$$v_{ij}(t) \geqslant 0,\quad -\alpha t+(1-\alpha)T_j+\omega_i(t)+\mu_j(t)-\lambda_i \geqslant 0 \quad i = I,\dots,n,\ \ j=1,\dots,J,\ \ t\in(-\infty,-T_j) \tag{3}$$

and the complementary slackness conditions between the pairs of inequalities in (1)–(3).

The optimality conditions (1)–(3) illustrate the relationship between travelers' experienced costs, externality costs, and their marginal costs, and thus provide a way to design the optimal tolling scheme. More specifically, in the inequality pairs of (3), $-\alpha t+(1-\alpha)T_j+\omega_i(t)+\mu_j(t)$ can be interpreted as the marginal cost for users departing at time t from on-ramp i to take the off-ramp j, and λ_i is the marginal system cost of adding one traveler using on-ramp i. The complementary slackness condition of (3) therefore describes the equilibrium condition in terms of marginal travel costs. Namely,

Lemma 1. *At dynamic system-optimum with departure time choice, for each on-ramp i, the marginal cost of all the paths and all the departure times with positive flows are the same and equal to marginal system cost λ_i, while other paths and departure times have marginal costs no less than that.*

According to Lemma 1, the optimal tolling scheme should be designed in a way such that each traveler departing at time t using on-ramp i and off-ramp j bears the path marginal cost $-\alpha t+(1-\alpha)T_j+\omega_i(t)+\mu_j(t)$. Since a traveler's path marginal cost $-\alpha t+(1-\alpha)T_j+\omega_i(t)+\mu_j(t)$ is made up of the travel cost $-\alpha t+(1-\alpha)T_j$, the externality costs $\omega_i(t)$ and $\mu_j(t)$ that the traveler imposes on other travelers at on-ramp i and off-ramp j, respectively, a link-based tolling scheme that brings the system to its optimal state is simply to charge travelers a time-dependent toll $\omega_i(t)$ at the on-ramp i and a time-dependent toll $\mu_j(t)$ at the off-ramp j.

In addition, the optimality conditions (1) and (2) illustrate a distinct feature of the externality for the bottleneck model. As shown, $\omega_i(t) = 0$ if $\sum_{j=1}^{J} v_{ij}(t) < s_i$ and $\mu_j(t) = 0$ if $\sum_{i=I}^{n} v_{ij}(t) < c_j$, meaning that externality would not arise if the bottlenecks are not saturated (i.e., flow on it reaches its capacity) at the time when the additional traveler traverses it. This is because in the bottleneck model, congestion is localized.

3.3. Some essential properties of the system-optimal flow and toll patterns

We first introduce the following critical time points to aid our discussion of the salient properties of the optimal flow and toll patterns:

t_{j0}	The earliest time when the discharging flow on off-ramp j becomes positive Let $t_{j0} = -T_j$ if the discharging flow is always zero
t_{js}	The earliest time when off-ramp j is saturated (i.e., discharging flow equals its capacity) Let $t_{js} = -T_j$ if it is always unsaturated
τ_{is}	The earliest time when the discharging flow on on-ramp i is positive and the last off-ramp in use is unsaturated Let $\tau_{is} = -T_{J'}$ if all the off-ramps in use are saturated whenever the discharging flow on on-ramp i is positive
$s(i)$	The index of the unsaturated off-ramp corresponding to τ_{is} Let $s(i) = \emptyset$ if $\tau_{is} = -T_{J'}$ (By this definition, $\tau_{is} \in [t_{s(i)0}, t_{s(i)s})$ if $s(i) \neq \emptyset$)
τ_{i0}	The latest time before τ_{is} when the last off-ramp in use is unsaturated Let $\tau_{i0} = t_{10}$ if all the off-ramps in use are saturated when there are flows on them (By this definition, $\tau_{i0} < \tau_{is}$ if $\tau_{is} = t_{s(i)0}$ and $\tau_{i0} = \tau_{is}$ if $\tau_{is} \in (t_{s(i)0}, t_{s(i)s}))$
$o(i)$	The index of the unsaturated off-ramp corresponding to τ_{i0} Let $o(i) = \emptyset$ if $\tau_{i0} = t_{10}$ (By this definition, $o(i) < s(i)$ if $\tau_{is} = t_{s(i)0}$ and $o(i) = s(i)$ if $\tau_{is} \in (t_{s(i)0}, t_{s(i)s}))$

Given any feasible traffic flow pattern in the freeway system, all the above critical time points can be identified accordingly. For example, suppose a freeway system with three on-ramps and two off-ramps has a temporal traffic flow profile as shown in Fig. 7. Both the disaggregate ramp flow (Fig. 7a–e) and aggregate flow (Fig. 7f) are depicted. $t_a, t_b, \ldots, t_f$ are the time

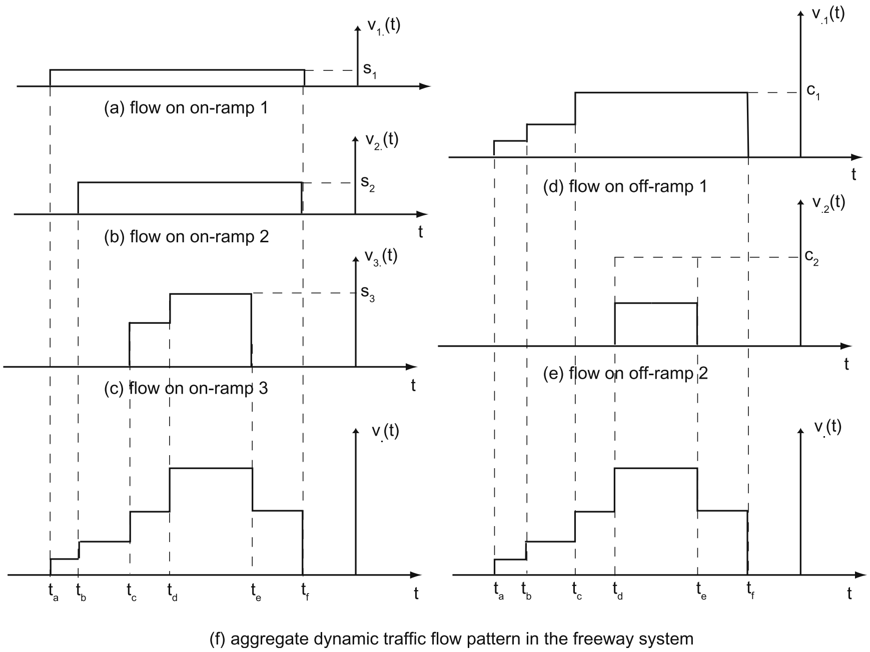

Fig. 7. A traffic flow pattern example in a freeway system with three on-ramps and two off-ramps.

points when the aggregate traffic flow rate changes. According to our definitions, for off-ramp 1, $t_{10} = t_a$, $t_{1s} = t_c$; for off-ramp 2, $t_{20} = t_d$, $t_{2s} = -T_2$ (because the flow on it is always below c_2); for on-ramp 1, $\tau_{1s} = t_a$, $s(1) = 1$, $\tau_{10} = t_a$, $o(1) = 1$; for on-ramp 2, $\tau_{2s} = t_b$, $s(2) = 1$, $\tau_{20} = t_b, o(2) = 1$ (because $\tau_{2s} \in (t_{10}, t_{1s})$; for on-ramp 3, $\tau_{3s} = t_d$, $s(3) = 2$, $\tau_{30} = t_c, o(3) = 2$ (because $\tau_{3s} = t_{20}$).

Based on the propositions and optimality conditions previously derived, we can identify the following important properties of the optimal traffic flow pattern.

Theorem 1

(1) *For any off-ramp j used at system-optimum:*

(1.1) *the flow on it becomes positive exactly* $\frac{1-\alpha}{\alpha}(T_j - T_{j-1})$ *time units after its downstream adjacent off-ramp's flow becomes equal to its capacity, i.e.,*

$$t_{j0} = t_{(j-1)s} + \frac{1-\alpha}{\alpha}(T_j - T_{j-1}) \quad \forall j = 2, \ldots, J'$$

In particular, for off-ramp J' *which is the last off-ramp in use, if* $J' < J$, *then*

$$-T_{J'} - t_{J's} \leqslant \frac{1-\alpha}{\alpha}(T_{J'+1} - T_{J'})$$

(1.2) *the flow on it always ends at time* $-T_j \; \forall j = 1, \ldots, J'$,

(1.3) *the flow on it always equals its capacity from time* t_{js} *to time* $-T_j$, *i.e.,*

$$v_{\bullet j}(t) = c_j \quad \forall t \in (t_{js}, -T_j), \quad j = 1, \ldots, J'$$

(2) *For any on-ramp i used at system-optimum:*

(2.1) *the flow on it always equals its capacity from time* τ_{is} *to time* $-T_{s(i)}$, *i.e.,*

$$v_{i\bullet}(t) = s_i \quad \forall t \in (\tau_{is}, -T_{s(i)})$$

(2.2) *the flow on it is zero before time* τ_{i0} *and after time* $-T_{o(i)}$, *i.e.,*

$$v_{i\bullet}(t) = 0 \quad \forall t \in (-\infty, \tau_{i0}) \cup (-T_{o(i)}, 0)$$

Proof. Theorem 1 can be proven by showing that any traffic flow pattern violating any of the properties in Theorem 1 can be adjusted to further reduce the system cost. For the purpose of exposition, the detailed proof is shown in Appendix. □

A typical dynamic traffic flow pattern $\{v_\bullet(t)\}$, obtained by aggregating flows over all the on-ramps, that has all the properties described in Theorem 1 is depicted in Fig. 8. As shown, the aggregate flow rate $v_\bullet(t)$ has a staircase shape over time, and is monotonically increasing before time $-T_{J'}$ and monotonically decreasing after time $-T_{J'}$.

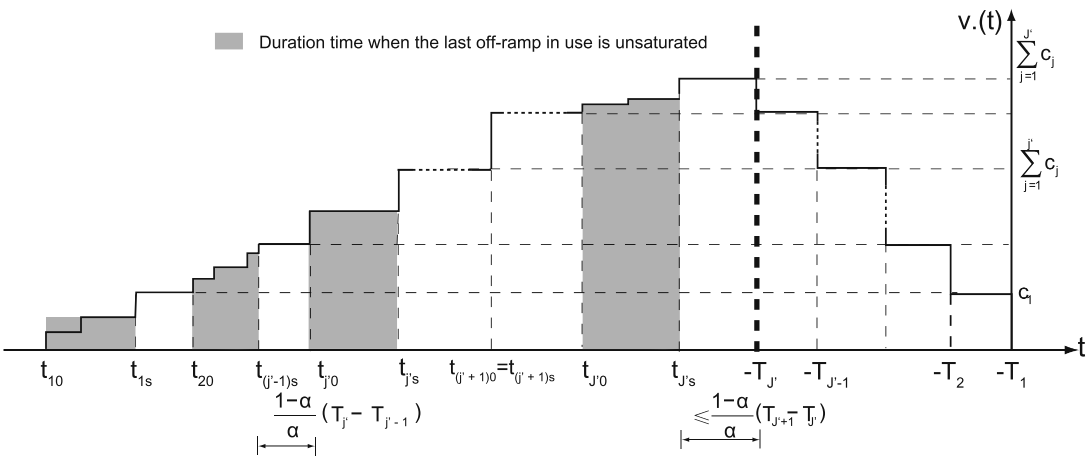

Fig. 8. A typical optimal aggregate dynamic traffic flow pattern in the freeway system.

The entire morning peak can thus be divided into two types of time periods:

Type I: $(t_{js}, t_{(j+1)0}) \cup (-T_{j+1}, -T_j)$, $j = 1, \ldots, J' - 1$ and $(t_{J's}, -T_{J'})$.

During the jth Type I period, the first j off-ramps are in use, and all of them are saturated. In particular, except for the J'th period $(t_{J's}, -T_{J'})$ which lasts no longer than $\frac{1-\alpha}{\alpha}(T_{J'+1} - T_{J'})$, the width of any other jth period ($j = 1, \ldots, J' - 1$) is known and equal to $\frac{1-\alpha}{\alpha}(T_{j+1} - T_j) + (T_{j+1} - T_j) = \frac{T_{j+1} - T_j}{\alpha}$.

Type II: (t_{j0}, t_{js}), $j = 1, \ldots, J'$ (highlighted by the gray area in Fig. 8).

During the jth Type II period, the first j' off-ramps are in use, and all but off-ramp j are saturated. The width of each time interval does not have an analytical form and may be as small as zero.

Fig. 9 shows the typical disaggregate optimal dynamic traffic flow patterns at both on- and off-ramps. As shown, each on-ramp $i = I, \ldots, n$ only has positive flows within time period $(\tau_{i0}, -T_{o(i)})$. In particular, during time $t \in (\tau_{is}, -T_{s(i)}) \subseteq (\tau_{i0}, -T_{o(i)})$, the flow is exactly equal to its capacity s_i; similarly, each off-ramp $j = 1, \ldots, J'$ only has flows within time period $t \in (t_{j0}, -T_j)$, and during $t \in (t_{js}, -T_j) \subseteq (t_{j0}, -T_j)$, the flow is exactly equal to its capacity c_j.

To further explain the basic properties of the optimal dynamic traffic flow pattern, Fig. 10 depicts the typical shapes of the optimal aggregate traffic flow patterns for two special networks. Corresponding to Fig. 10a is a simple merge network representing a freeway system with multiple on-ramps and one off-ramp (i.e. the freeway link incident to the destination), and corresponding to Fig. 10b is a simple diverge network with parallel routes at the downstream side representing a freeway system with multiple off-ramps and one on-ramp.

In Fig. 10a, the entire optimal aggregate traffic flow pattern is made up of a staircase segment before time t_{1s} when on-ramps with $\tau_{is} < t_{1s}$ discharge flows at their capacities and a constant flow segment from time t_{1s} to time 0 when the aggregate flow equals the off-ramp capacity c_1. Depending on the total flow discharged from each on-ramp, the duration time of either of these two segments may be zero.

In Fig. 10b, $t_{j0} = t_{js}\ \forall j = 1, \ldots, J' - 1$, meaning that all the off-ramps $j = 1, \ldots, J' - 1$ always operate at their capacities whenever there are travelers traversing them. $\tau_{1s} = t_{J'0}$, meaning that the on-ramp operates at its capacity from time τ_{1s} to time $-T_{J'}$. Since this duration time may be zero, it is possible that the on-ramp capacity is never saturated during the entire morning peak.

If we denote $V_{i\bullet} = V^{\mathrm{I}}_{i\bullet} + V^{\mathrm{II}}_{i\bullet}$ where $V^{\mathrm{I}}_{i\bullet}$ and $V^{\mathrm{II}}_{i\bullet}$ are the total flow discharged from on-ramp i during all the Type I time periods and during all the Type II periods, respectively. Theorem 1 also leads to the following corollary regarding the usage of on-ramps.

Corollary 1. *Given two on-ramps i' and i''*

$$V^{\mathrm{II}}_{i'\bullet}/s_{i'} \geqslant V^{\mathrm{II}}_{i''\bullet}/s_{i''} \iff \tau_{i's} \leqslant \tau_{i''s} \iff \tau_{i'0} \leqslant \tau_{i''0}$$

Proof. We first show that $V^{\mathrm{II}}_{i'\bullet}/s_{i'} \geqslant V^{\mathrm{II}}_{i''\bullet}/s_{i''} \Rightarrow \tau_{i's} \leqslant \tau_{i''s} \Rightarrow \tau_{i'0} \leqslant \tau_{i''0}$. Property 2 in Theorem 1 implies that any on-ramp i is always saturated during any Type II time period whenever it has positive flows. Therefore, according to the definition of τ_{is}, $V^{\mathrm{II}}_{i'\bullet}/s_{i'} \geqslant V^{\mathrm{II}}_{i''\bullet}/s_{i''} \Rightarrow \tau_{i's} \leqslant \tau_{i''s}$. The definition of τ_{i0} further implies that $\tau_{i's} \leqslant \tau_{i''s} \Rightarrow \tau_{i'0} \leqslant \tau_{i''0}$. The proof for the opposite direction, i.e., $V^{\mathrm{II}}_{i'\bullet}/s_{i'} \geqslant V^{\mathrm{II}}_{i''\bullet}/s_{i''} \Leftarrow \tau_{i's} \leqslant \tau_{i''s} \Leftarrow \tau_{i'0} \leqslant \tau_{i''0}$, is similar since it is equivalent to $V^{\mathrm{II}}_{i'\bullet}/s_{i'} < V^{\mathrm{II}}_{i''\bullet}/s_{i''} \Rightarrow \tau_{i's} > \tau_{i''s} \Rightarrow \tau_{i'0} > \tau_{i''0}$. □

The properties described in Theorem 1 are also sufficient for a feasible traffic flow pattern to be optimal. Moreover, they lead a way to construct the time-dependent toll that realizes the dynamic system-optimal traffic flow pattern.

Theorem 2. *Any feasible traffic assignment pattern having the properties given in Theorem 1 is an optimal flow pattern in the freeway system. Corresponding to the critical time points $\{\tau_{i0}\}, \{\tau_{is}\}, \{t_{j0}\}, \{t_{js}\}$ that characterize the traffic flow pattern, the marginal system cost for travelers from each origin $i = I, \ldots, n$ is*

$$\lambda_i = -\alpha\tau_{i0} + (1 - \alpha)T_{o(i)} \tag{4}$$

and the dynamic tolls to be charged at the ramps to achieve this optimal traffic flow pattern are as follows (Fig. 11):

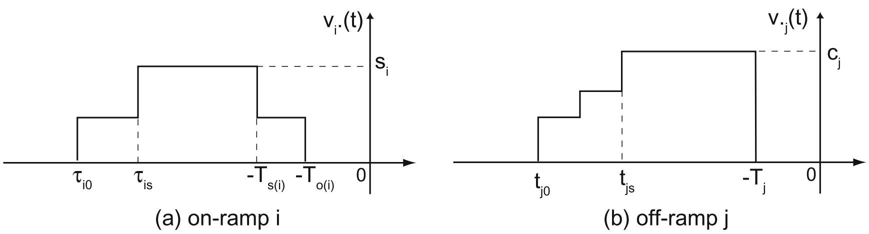

Fig. 9. Typical optimal disaggregate traffic flow patterns on the ramps.

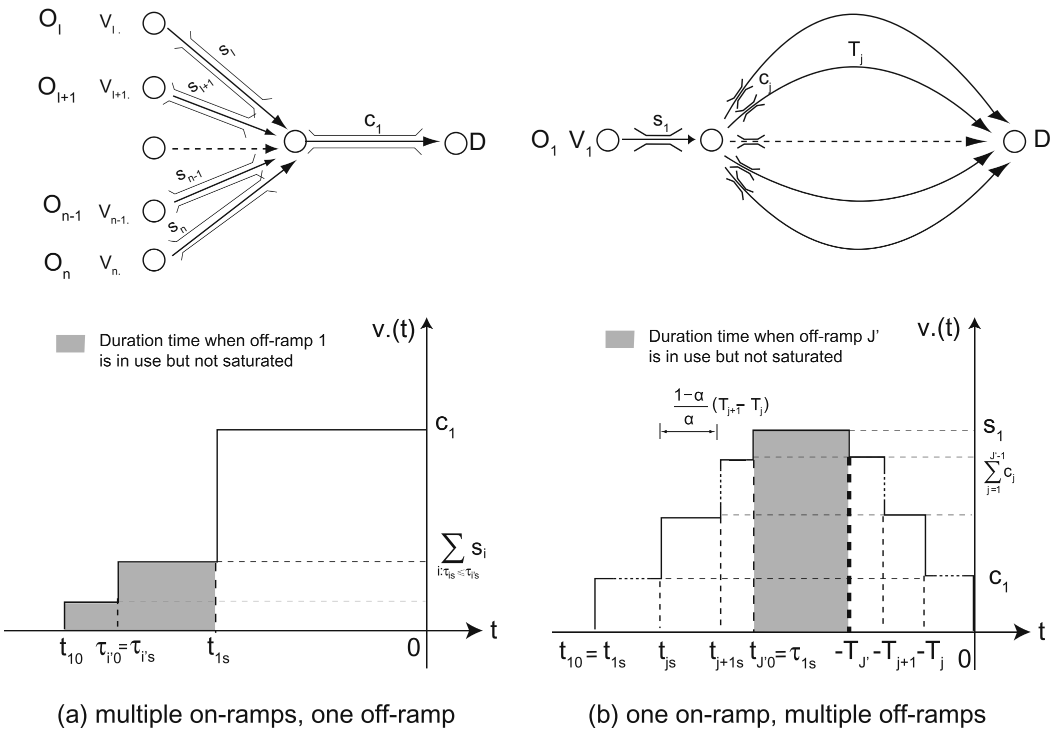

Fig. 10. The aggregate optimal traffic flow patterns for two special networks.

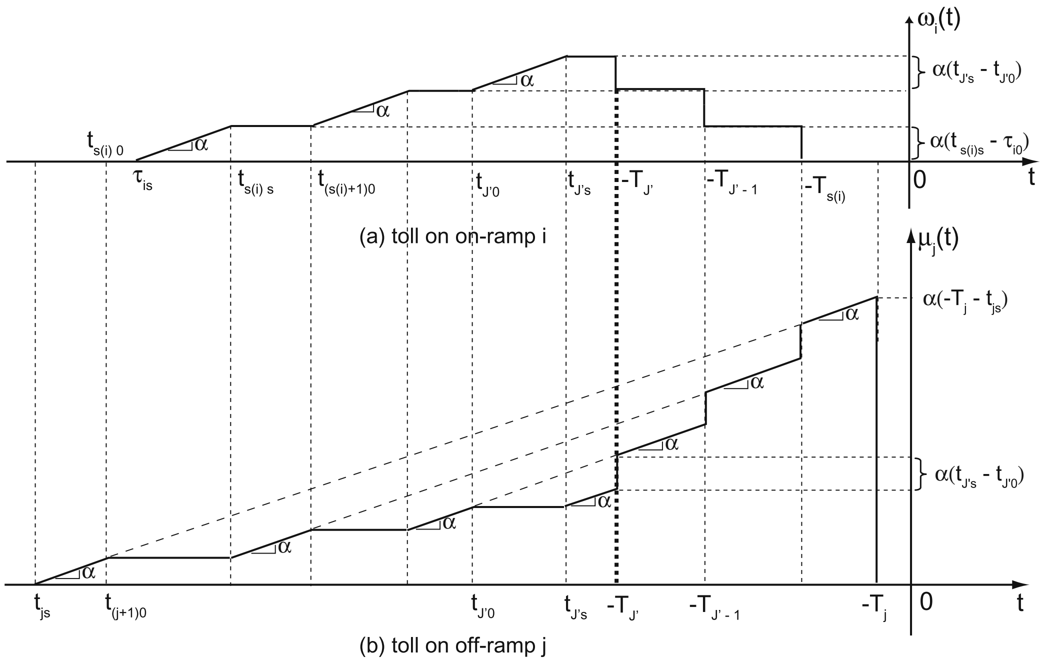

Fig. 11. The optimal dynamic toll on the ramps.

$$\omega_i(t) = \begin{cases} \alpha(t-\tau_{is}) - (1-\alpha)(T_{j'} - T_{s(i)}) & t \in (t_{j'0}, t_{j's}), \quad j' = s(i), \ldots, J' \\ \alpha(t-\tau_{is}) - (1-\alpha)(T_{j'} - T_{s(i)}) - \alpha(t - t_{j's}) & t \in (t_{j's}, t_{(j'+1)0}), \quad j' = s(i), \ldots, J'-1 \\ \alpha(t_{j'0}-\tau_{is}) - (1-\alpha)(T_{j'} - T_{s(i)}) & t \in (-T_{j'}, -T_{j'-1}), \quad j' = J', \ldots, s(i)+1 \end{cases} \tag{5}$$

$$\mu_j(t) = \begin{cases} (1-\alpha)(T_{j'} - T_j) & t \in (t_{j'0}, t_{j's}), \quad j' = j, \ldots, J' \\ (1-\alpha)(T_{j'} - T_j) + \alpha(t - t_{j's}) & t \in (t_{j's}, t_{(j'+1)0}), \quad j' = j, \ldots, J'-1 \\ (1-\alpha)(T_{j'} - T_j) + \alpha(t - t_{j'0}) & t \in (-T_{j'}, -T_{j'-1}), \quad j' = J', \ldots, j+1 \end{cases} \tag{6}$$

Proof. It suffices to show that the multipliers defined by (4)–(6) satisfy the optimality conditions (1)–(3). It is easy to see from Fig. 11 that during the time when both on-ramp i and off-ramp j have positive flows, the sum of the externalities on them, $\omega_i(t) + \mu_j(t)$, increases at the rate of α over time, which is exactly the decreasing rate of the schedule delay at the destination. The reader is referred to Appendix for the complete proof of this theorem. □

As shown in Fig. 11, the dynamic tolling patterns on both on- and off-ramps are piecewise linear, and have two distinct segments divided by time $-T_{J'}$. Interestingly, when both an on-ramp and off-ramp have positive tolls, they take turns to have an increase in the toll at rate α. In other words, whenever the toll on the on-ramp (off-ramp) increases at the rate of α, the toll on the off-ramp (on-ramp) remains constant.

For each on-ramp $i = I, \ldots, n$, the toll starts from time τ_{is} and ends at time $-T_{s(i)}$. During time $t \in (\tau_{is}, -T_{J'})$, $\omega_i(t)$ increases at the rate of α when $t \in (t_{j'0}, t_{j's}), j' = s(i), \ldots, J'$ and remains constant when $t \in (t_{j's}, t_{(j'+1)0}), j' = s(i), \ldots, J'$. During time $t \in (-T_{J'}, -T_{s(i)})$, $\omega_i(t)$ drops $\alpha(t_{j's} - t_{j'0})$ when t passes time $-T_{j'}, j' = J', \ldots, s(i)$ and remain constant otherwise.

For each off-ramp $j = 1, \ldots, J'$, the toll starts from time t_{js} and ends at time $-T_j$. During time $t \in (\tau_{is}, -T_{J'})$, $\mu_j(t)$ increases at the rate of α when $t \in (t_{j's}, t_{(j'+1)0}), j' = j, \ldots, J'$ and remains constant when $t \in (t_{j'0}, t_{j's}), j' = j, \ldots, J'$. During time $t \in (-T_{J'}, -T_j)$, $\mu_j(t)$ jumps $\alpha(t_{j's} - t_{j'0})$ when t passes time $-T_{j'}, j' = J', \ldots, j+1$ and increases at the rate of α otherwise.

Fig. 12 depicts, for the two special networks shown previously in Fig. 10, the optimal dynamic toll profiles, from which it can be easily seen that the total toll that a traveler pays at both the on- and off-ramps she accesses increases at the rate of α over time.

Finally, although the analytical formulae for the optimal time-dependent toll, (5) and (6), are derived based on the assumption that the freeway mainline has negligible free-flow travel time, they can be easily modified to take into account the non-zero free-flow travel time on the freeway mainline. More specifically, suppose the free-flow travel time from any ramp i (either on-ramp or off-ramp) to the destination is fc_i. It suffices to shift the toll profile on any ramp i to the left by fc_i units.

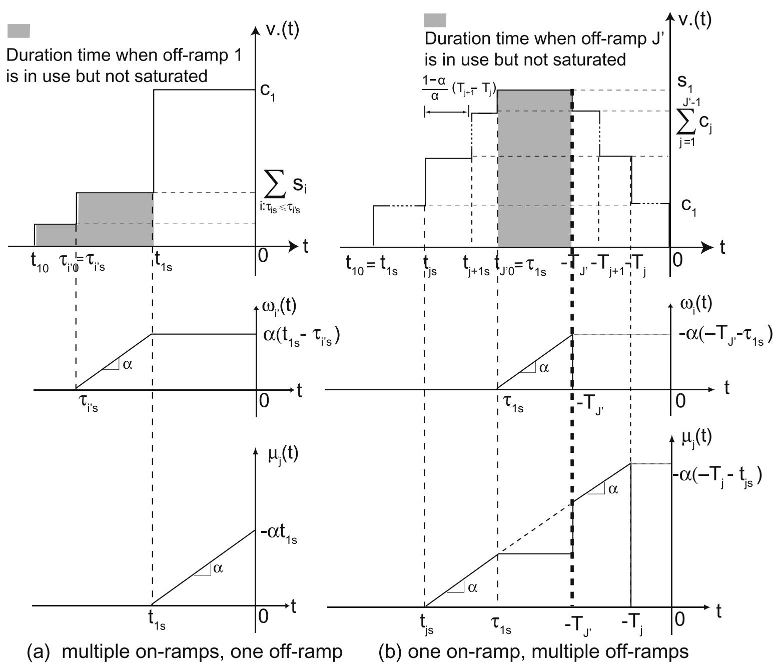

Fig. 12. The optimal dynamic toll profiles for two special networks.

Theorem 2 also leads to the following corollary regarding the marginal cost of each on-ramp i in use at system-optimum:

Corollary 2. *Given any two on-ramps i' and i'' in use at dynamic system optimum*

$$\lambda_{i'} \geqslant \lambda_{i''} \Longleftrightarrow \tau_{i'0} \leqslant \tau_{i''0} \Longleftrightarrow V_{i'\bullet}/s_{i'} \geqslant V_{i''\bullet}/s_{i''}$$

Proof. We first prove for $\lambda_{i''} \geqslant \lambda_{i'} \Rightarrow \tau_{i''0} \leqslant \tau_{i'0} \Rightarrow V_{i''\bullet}/s_{i''} \geqslant V_{i'\bullet}/s_{i'}$:

(1) $\lambda_{i''} \geqslant \lambda_{i'} \Rightarrow \tau_{i''0} \leqslant \tau_{i'0}$:According to Theorem 2, for any on-ramp i in use at system-optimum, $\lambda_i = -\alpha\tau_{i0} + (1-\alpha)T_{o(i)}$. If $\lambda_{i''} \geqslant \lambda_{i'}$, we have the following three possibilities:
  - (i) $o(i') > o(i'')$: obviously, $\tau_{i''0} < \tau_{i'0}$;
  - (ii) $o(i') = o(i'')$: $\tau_{i''0} \leqslant \tau_{i'0}$, since $(1-\alpha)T_{o(i'')} = (1-\alpha)T_{o(i')}$;
  - (iii) $o(i') < o(i'')$: this situation is impossible because otherwise based on the definitions of τ_{i0} and $o(i)$, we have $\tau_{i'0} - \tau_{i''0} < \frac{1-\alpha}{\alpha}(T_{o(i')} - T_{o(i'')})$, i.e., $\lambda_{i'} > \lambda_{i''}$, contradicting with the assumption of $\lambda_{i''} \leqslant \lambda_{i'}$.

(2) $\tau_{i''0} \leqslant \tau_{i'0} \Rightarrow V_{i''\bullet}/s_{i''} \geqslant V_{i'\bullet}/s_{i'}$:

The definitions of $\{\tau_{is}\}$ and $\{\tau_{i0}\}$ imply that $\tau_{i''0} \leqslant \tau_{i'0}$ will lead to $(\tau_{i''s}, -T_{s(i'')}) \supseteq (\tau_{i'0}, -T_{o(i')})$. This means that whenever the flow on on-ramp i' is positive, the flow on on-ramp i'' is equal to its capacity $s_{i''}$. Hence, $\frac{V_{i''\bullet}}{s_{i''}} \geqslant \frac{V_{i'\bullet}}{s_{i'}}$. The proof for the other direction, i.e., $\lambda_{i''} \geqslant \lambda_{i'} \Leftarrow \tau_{i''0} \leqslant \tau_{i'0} \Leftarrow V_{i''\bullet}/s_{i''} \geqslant V_{i'\bullet}/s_{i'}$, is similar and thus omitted here. □

4. A graphical procedure to obtain the optimal traffic flow and toll profiles

According to Theorems 1 and 2, the dynamic system-optimal traffic flow pattern can be found by obtaining a feasible traffic flow profile satisfying all the properties of Theorem 1. We shall show in this section that such a traffic flow profile can be constructed through a graphical solution procedure.

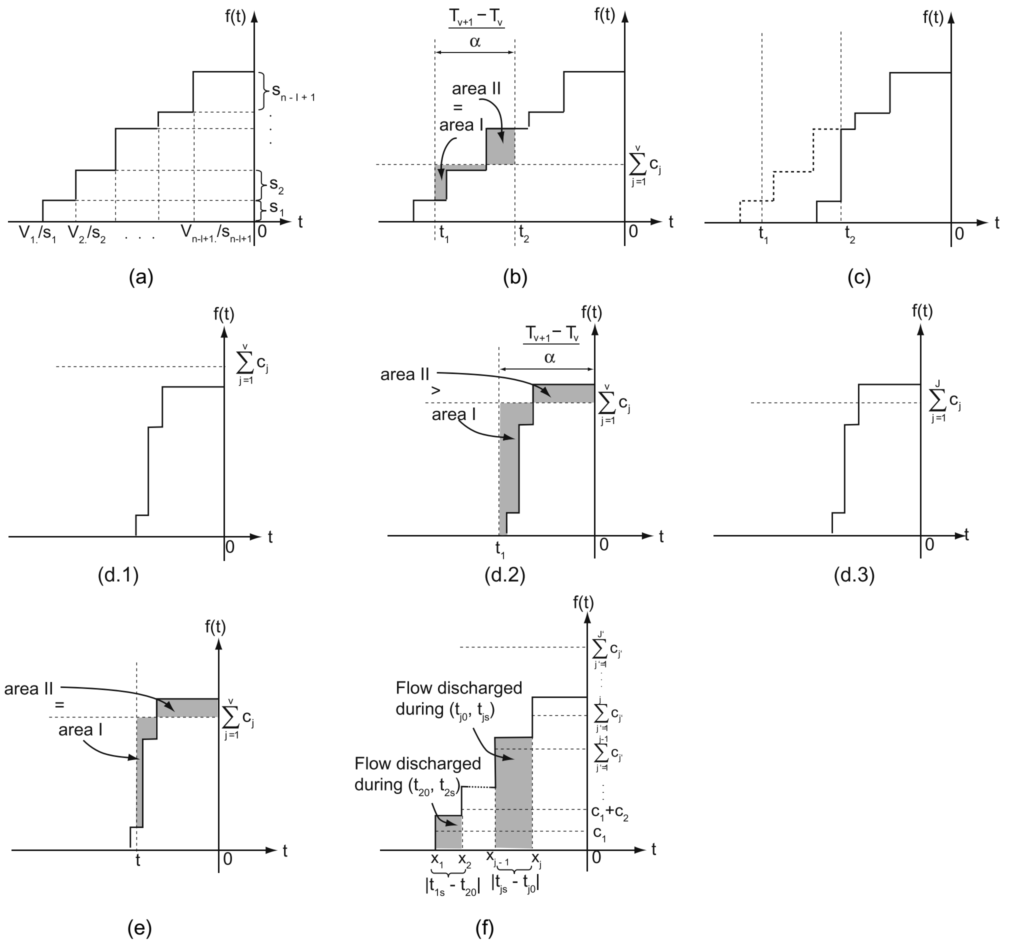

Fig. 13. A graphical illustration of the algorithm to derive the optimal traffic pattern.

For the ease of presentation, the on-ramp indices are relabelled in the descending order of the cumulative volume-to-capacity ratio $\{V_{i\bullet}/s_i\}_{i=I,\dots,n}$. An auxiliary function $f(t)$ is then constructed by drawing horizontal bars with length $V_{i\bullet}/s_i$ and height s_i for each on-ramp $i = 1,\dots,n-I+1$ and stacking them sequentially from the bottom with the same ending point $t = 0$ (Fig. 13a). The whole area below $f(t)$ thus represents the total travel demand to be distributed in the freeway system, and the area under $f(t)$ between $y = \sum_{i'=1}^{i-1} s_{i'}$ and $y = \sum_{i'=1}^{i} s_{i'}$ represents the travel demand discharged into the freeway system from a specific on-ramp i. With this auxiliary function $f(t)$, the process of determining the total and ramp-specific travel demand discharged during a given time period can be represented as *subtracting a certain area from that below* $f(t)$. After each subtraction operation, the remaining demand $\{V_{i\bullet}\}$ is updated and $f(t)$ is re-constructed.

According to Theorem 1, the entire morning peak at dynamic system-optimum can be divided into two types of time periods: Type I periods $j = 1,\dots,J'$ and Type II periods $j = 1,\dots,J'$. For every Type I period, the flow distribution at each off-ramp and the duration time (except for the last Type I period $(t_{J's}, -T_{J'})$) are known, while how flows are discharged from each on-ramp is to be determined; for every Type II period, the flow distribution at each off-ramp is known, while both the duration time and the flow distribution at each on-ramp are to be determined. Corresponding to such a time period partition, our solution procedure is also presented in two steps:

Step 1: Determine both the total and the ramp-specific travel demand discharged during every Type I period $j = 1,\dots,J'$.
Step 2: Determine both the total and the ramp-specific travel demand discharged during every Type II period $j = 1,\dots,J'$.

For narrative convenience, Step 2 is described ahead of Step 1.

4.1. Flow distribution during Type II periods (Step 2)

Suppose the area corresponding to the travel demand discharged during all the Type I periods have been subtracted from that below $f(t)$. The remaining $f(t)$ is thus constructed based on travel demand $\{V^{\mathrm{II}}_{i\bullet}\}$ (Fig. 13f). According to Corollaries 1 and 2, the sequence of $\{V^{\mathrm{II}}_{i\bullet}/s_i\}$ is in the same order of the sequence of $\{V_{i\bullet}/s_i\}$, indicating that this remaining $f(t)$ constructed based on $\{V^{\mathrm{II}}_{i\bullet}\}$ shares a similar shape of a climbing staircase as the original $f(t)$.

To determine the flow distribution during every Type II period, we can first draw a series of horizontal lines $y = c_1,\dots,y = \sum_{j'=1}^{j} c_{j'},\dots,y = \sum_{j'=1}^{J} c_{j'}$ on the plot of the remaining $f(t)$ and denote the x-coordinate of the crossing point between $f(t)$ and any horizontal line $y = \sum_{j'=1}^{j} c_{j'}$ as x_j. According to Property 2 in Theorem 1, each on-ramp i becomes saturated during all the Type II periods once its flows become positive. This implies that the travel demand corresponding to the area below the remaining $f(t)$ between line $y = \sum_{j'=1}^{j-1} c_{j'}$ and line $y = \sum_{j'=1}^{j} c_{j'}$ is discharged during the jth Type II period (t_{j0}, t_{js}) (Fig. 13f), and $|x_{j-1} - x_j|$ represents the duration time $|t_{js} - t_{j0}|$ of the jth Type II period. By this means, the total area below the remaining $f(t)$ is partitioned into every Type II period $j = 1,\dots,J'$.

4.2. Determine the flow distribution during Type I periods (Step 1)

The demand discharged during every Type I period $j = 1,\dots,J'$ can be determined in an iterative way. Namely, during the jth iteration, the flow distribution for the jth Type I period is determined by subtracting the corresponding area from that below $f(t)$.

The following three properties can be utilized to design such an area subtraction operation:

(1) The total demand discharged during the jth Type I period is equal to $\frac{T_{j+1}-T_j}{\alpha}\sum_{j'=1}^{j} c_{j'}$ (Properties 1.2 and 1.3, Theorem 1).
(2) Since the remaining $f(t)$ for Step 2 has a similar shape as the original $f(t)$, it is desirable for the re-constructed $f(t)$ after the jth iteration to maintain a similar shape.
(3) After the jth iteration, if for a certain on-ramp i we have $\sum_{i':V_{i'\bullet}>0, i'\leqslant i} s_{i'} < \sum_{j'=1}^{j} c_{j'}$ (i.e., there is remaining demand potentially to be used during a $j' < j$th Type II time period), this on-ramp i should always be saturated during the jth Type I period (Property 2.1, Theorem 1).

The area subtraction approach satisfying all the above three properties is as follows. Suppose we want to determine the flow distribution during the jth time period $(t_{js}, t_{(j+1)0}) \cup (-T_{j+1}, -T_j)$. On the plot of $f(t)$ constructed based on the remaining demand $\{V_{i\bullet}\}$ from a previous iteration, we draw a horizontal line $y = \sum_{j'=1}^{j} c_{j'}$ and move a vertical band characterized by $t_1 = t^*$ and $t_2 = t^* + \frac{T_{j+1}-T_j}{\alpha}$ around until area I = area II (Fig. 13b). The area below $f(t)$ between $y = t_1$ and $y = t_2$ is the area corresponding to the travel demand to be discharged during this jth Type I time period. By this means, the subtracted area is equal to $\sum_{j'=1}^{j} c_{j'} \times \frac{T_{j+1}-T_j}{\alpha}$; the shape of the re-constructed $f(t)$ based on the remaining demand (Fig. 13c) still has the climbing staircase shape; any on-ramp i with positive remaining demand and satisfying $\sum_{i':V_{i'\bullet}>0, i'\leqslant i} s_{i'} < \sum_{j'=1}^{j} c_{j'}$ is always saturated during the jth Type I period.

The above area subtraction operation can be performed iteratively starting from $j = 1$ till one of the following conditions are satisfied: (1) the line of $y = \sum_{j'=1}^{j} c_{j'}$ is above $f(t)$ (Fig. 13d.1); (2) there is no such $t^* \leqslant 0$ which can make area I = area II (Fig. 13d.2); (3) $j = J$ (Fig. 13d.3). Then, let $J' = j$ and the flow distribution of the J'th Type I period $(t_{J's}, -T_{J'})$ is determined. If termination condition (1) applies, $t_{J's} = -T_{J'}$ since the maximum flow rate that can be discharged from the remaining flow is below $\sum_{j=1}^{J'} c_j$; if termination conditions (2) and (3) applies, the duration time $|-T_{J'} - t_{J's}|$ and the flow distribution is determined by moving the line $t = t^*$ such that area I = area II (Fig. 13e) and $t_{J's} = -T_{J'} + t^*$.

It is easy to verify that the traffic flow pattern constructed by this procedure (Steps 1 and 2) satisfies all the properties in Theorem 1 and hence is an optimal traffic flow pattern. Once the optimal traffic flow pattern is obtained, the time-dependent optimal toll profile can be constructed, according to Theorem 2, from all the characteristic time points $\{t_{j0}\}, \{t_{js}\}, \{\tau_{i0}\}, \{\tau_{is}\}$.

The above algorithm assumes that the demand $V_{i\bullet}$ for each origin i to use the freeway system is known. In fact, the optimal $\{V_{i\bullet}\}$ which leads to the minimal total cost in the corridor network is a solution to the mathematical program optimizing the total cost in the entire corridor network as follows:

$$\min\left\{\sum_{i=1}^{n}(N_i - V_{i\bullet})D_i + \mathrm{FC}(\mathbf{V}) | V_i \geqslant 0 \quad \forall i = 1, \ldots, n\right\}$$

where $\mathrm{FC}(\mathbf{V})$ represents the minimal cost in the freeway system for given $\{V_{i\bullet}\}$. The optimality condition of this program reads

$$V_i \geqslant 0, \quad D_i - \frac{\partial \mathrm{FC}(\mathbf{V})}{\partial V_i} \geqslant 0, \quad V_i\left[D_i - \frac{\partial \mathrm{FC}(\mathbf{V})}{\partial V_i}\right] = 0, \quad i = 1, \ldots, n$$

Note that $\frac{\partial \mathrm{FC}(\mathbf{V})}{\partial V_i}$ is equivalent to the multiplier λ_i of the program (P).

According to this optimality condition, the split between the freeway system and the alternative surface street route for each origin can be determined iteratively in practice. More specifically, starting from the initial scenario in which all the demand from all the origins will use the freeway system, we can perform the graphical procedure to obtain the optimal dynamic traffic flow pattern for this case. Then we check sequentially from the first on-ramp with flows on the freeway system if $\lambda_i \leqslant D_i$. If this condition is violated, reduce $V_{i\bullet}$ by a predetermined amount and perform the graphical procedure again. This iterative process terminates when $\lambda_i \leqslant D_i \ \forall i = 1, \ldots, n$ holds.

5. Concluding remarks

The morning commute problem, considered under the simple setting of a single route with a single bottleneck, brings out the essential elements of traffic congestion and has been extensively studied. Extending this problem to a general network is challenging because of the underlying complexity of modeling traffic flow. Under certain modeling assumptions such an analysis can still be achieved with the help of numerical approximations, which adds realism but tends to lose much of the elegance and clarity found under the simple network setting. For some special networks, such as a corridor network consisting of a freeway and several parallel arterial routes commonly found in cities with a hub and spoke type of road network (i.e., several sparsely connected freeways connecting suburbs to a single downtown), the basic features of the solution to the morning commute problem can still be characterized. These features often provide great insights to the nature of the problem and lead itself to effective remedies for congestion relief. In this paper, we carried out such an analysis on a corridor network with multiple bottlenecks on each route, a problem considered difficult and not addressed in the literature. We first identified the basic properties of the optimal traffic flow pattern in this network under both time-varying demand and route choice, then gave the formulae for a time-dependent toll to realize the optimal traffic flow pattern in the network under user-optimal choice behavior, and finally provided a graphical procedure to obtain both the optimal dynamic traffic flow pattern and the corresponding toll profile.

Our analysis has revealed that at system-optimum:

- The freeway has no queue throughout the entire morning peak. Although there may exist different optimal path/link flow distributions, the total flow in the freeway system is unique and is characterized by Theorem 1. The profile of the total flow has a staircase form and contains a monotonically increasing time period followed by a monotonically decreasing time period.
- All on-ramps have no queues, and the on-ramps in use are all upstream of the off-ramps in use. Moreover, on-ramps with larger cumulative volume-to-capacity ratios (i.e., $V_{i\bullet}/s_i$) are used longer than on-ramps with smaller volume-to-capacity ratios, and on-ramps in use may have a saturated period, preceded and followed by under saturated periods.
- All off-ramps have no queues, and downstream off-ramps are more heavily utilized than upstream off-ramps, i.e., downstream off-ramps get used earlier and longer than upstream off-ramps. Moreover, once an off-ramp becomes saturated, it remains saturated till no traffic accesses this ramp.

To realize the system-optimal traffic flow pattern, it suffices to impose a piecewise linear toll on a subset of on-ramps and off-ramps. Our analysis found that:

- The tolling scheme on both the on-ramps and off-ramps have two distinct time periods divided by the time when the flow on the last off-ramp in use runs out. During the first period, the tolls charged at on-ramps and off-ramps take turns to increase for a certain duration, and the rate of increase is always equal to α. During the second period, only the tolls charged at the off-ramps continue to increase. At time $-T_{j'}, j' = J', \ldots, 1$, there is a jump in the on-ramp tolls and a drop with the same net value in the off-ramp tolls.

- A downstream off-ramp has a longer tolling period and a higher maximum toll than an upstream off-ramp. The maximum toll at an off-ramp is charged at the time when the flow on it runs out.
- The on-ramp with a larger cumulative volume-to-capacity ratio has a longer tolling period and a higher maximum toll than an on-ramp with a smaller ratio. The maximum toll at an on-ramp is charged at the time when the flow on the last off-ramp in use runs out.

Although our analysis is conducted based on some ideal assumptions (vertical queue, infinite capacity on the city streets and no late arrival), the effects of relaxing some of the assumptions are not difficult to capture. Since we have identified that there will be no queue in the freeway at optimum, the optimal solution derived in this analysis actually applies to the situation where queue spill-back is taken into account. Considering late arrival requires minor change of the analysis framework, and is expected to result in a similar staircase optimal traffic flow pattern with different starting/ending time and single-mode piecewise linear ramp toll profiles with a monotonically increasing segment followed by a monotonically decreasing segment. Assuming finite capacity of city streets could impose difficulties for a rigorous analysis, but it can be expected that additional congestion on the city streets is likely to lead to a higher usage of the freeway system.

Although the piecewise linear time-dependent toll may be difficult to implement in practical applications, it does provide a benchmark toll solution to achieve the best system performance. In practice, a piecewise constant time-dependent toll approximating the optimal toll can be used to achieve a "second-best" system performance. How close the system performance given by this second-best toll is to that given by the first-best toll provided in this paper is interesting and worthy of further investigation.

Acknowledgements

This research is supported by a faculty research grant from the Sustainable Transportation Center (STC) at University of California Davis. The views expressed here are those of the authors alone. The authors wish to thank the two anonymous referees for their constructive comments.

Appendix

Proof of Theorem 1

(1.1) Suppose $\exists j' \leqslant J'$, $t_{j'0} \neq t_{(j'-1)s} + \frac{1-\alpha}{\alpha}(T_{j'} - T_{j'-1})$. There are two possibilities: $t_{j'0} > t_{(j'-1)s} + \frac{1-\alpha}{\alpha}(T_{j'} - T_{j'-1})$ and $t_{j'0} < t_{(j'-1)s} + \frac{1-\alpha}{\alpha}(T_{j'} - T_{j'-1})$.

(a) If $t_{j'0} > t_{(j'-1)s} + \frac{1-\alpha}{\alpha}(T_{j'} - T_{j'-1})$.
By Proposition 3, $\forall t^* \in (t_{(j'-1)s} + \frac{1-\alpha}{\alpha}(T_{j'} - T_{j'-1}), t_{j'0}), v_{\bullet}(t^*) = v_{\bullet}(t_{(j'-1)s}) = \sum_{j=1}^{j'} c_j$. Since the sum of the capacities of any subset of the on-ramps is not equal to the sum of capacities of any subset of the off-ramps, $v_{\bullet}(t^*) = v_{\bullet}(t_{(j'-1)s})$ implies that $\exists i', v_{i'\bullet}(t_{(j'-1)s}) > 0$ and $v_{i'\bullet}(t^*) < s_{i'}$.
Since $v_{\bullet(j'-1)}(t_{(j'-1)s}) > 0$ and $v_{i'\bullet}(t_{(j'-1)s}) > 0$, according to Proposition 2, one can always swap the off-ramp flows such that $v_{i'(j'-1)}(t_{(j'-1)s}) > 0$.
Now shift a small amount of flow ϵ originated at on-ramp i' from using off-ramp $j'-1$ at time $t_{(j'-1)s}$ to using off-ramp j' at time t^*. This manipulation can reduce the total system cost by $\epsilon[\alpha(t^* - t_{(j'-1)s}) - (1-\alpha)(T_{j'} - T_{j'-1})] > 0$, contradicting with the optimum assumption.

(b) If $t_{j'0} < t_{(j'-1)s} + \frac{1-\alpha}{\alpha}(T_{j'} - T_{j'-1})$, i.e., $t_{j'0} - \frac{1-\alpha}{\alpha}(T_{j'} - T_{j'-1}) < t_{(j'-1)s}$, by definition of $t_{(j'-1)s}$ $\forall t^* \in (t_{j'0} - \frac{1-\alpha}{\alpha}(T_{j'} - T_{j'-1}), t_{(j'-1)s})$, $v_{\bullet(j'-1)}(t^*) < c_{j'-1}$.
According to Proposition 3, $v_{\bullet}(t_{j'0}) > v_{\bullet}(t^*)$. Hence, there exists one on-ramp i' such that $v_{i'\bullet}(t_{j'0}) > v_{i'\bullet}(t^*)$.
Since $v_{\bullet j'}(t_{j'0}) > 0$ and $v_{i'\bullet}(t_{j'0}) > 0$, one can always make $v_{i'j'}(t_{j'0}) > 0$ by swapping flows.
Now shift a small amount of flow ϵ originated at on-ramp i' from using off-ramp j' at time $t_{j'0}$ to using off-ramp $j'-1$ at time t^*. This manipulation can reduce the total system cost by $\epsilon[(T_{j'} - T_{j'-1}) - \alpha[(t_{j'0} + T_{j'}) - (t^* + T_{j'-1})] > 0$, also contradicting with the optimum assumption.
In summary, $t_{j'0} = t_{(j'-1)s} + \frac{1-\alpha}{\alpha}(T_{j'} - T_{j'-1})$ $\forall j' \leqslant J'$.
For the same reason, for the last off-ramp J', if $-T_{J'} - t_{J's} > \frac{1-\alpha}{\alpha}(T_{J'+1} - T_{J'})$ and $J' < J$, then the flow on off -ramp $J'+1$ will become positive at time $t_{J's} + \frac{1-\alpha}{\alpha}(T_{J'+1} - T_{J'})$, contradicting with the fact that J' is the last off-ramp in use.Hence, $-T_{J'} - t_{J's} \leqslant \frac{1-\alpha}{\alpha}(T_{J'+1} - T_{J'})$.

(1.2) Evidently, the flow on any off-ramp $j' \leqslant J'$ cannot end after time $T_{j'}$. Now suppose $\exists j' \leqslant J$ whose last time interval with positive flow is $t^* < -T_{j'}$.
By Proposition 3, $v_{\bullet}(t^*) > v_{\bullet}(-T_{j'})$. Hence, there exists one on-ramp i' such that $v_{i'\bullet}(t^*) > v_{i'\bullet}(-T_{j'})$.
Since $v_{i'\bullet}(t^*) > 0$ and $v_{\bullet j'}(t^*) > 0$, one can always swap flows such that $v_{i'j'}(t^*) > 0$.
Now shift a small amount of flow ϵ originated at on-ramp i' using off-ramp j' from time t^* to time $-T_{j'}$. This manipulation can reduce the total system cost by $\epsilon\alpha(-T_{j'} - t^*) > 0$, contradicting with the optimum assumption.
In summary, $\forall j' \leqslant J'$, the flow on it ends at time $-T_{j'}$.

(1.3) Suppose $\exists j' \leqslant J', t^* \in (t_{j's}, -T_{j'})$ such that $v_{\bullet j'}(t^*) < s_{j'}$.
By Proposition 3, $v_\bullet(t^*) < v_\bullet(t_{j's})$ hence, there exists one on-ramp i' such that $v_{i'\bullet}(t^*) < v_{i'\bullet}(t_{j's})$.
Since $v_{i'\bullet}(t^*) < v_{i'\bullet}(t_{j's})$ and $v_{\bullet j'}(t^*) < v_{\bullet j'}(t_{j's})$, one can always swap flows such that $v_{i'j'}(t^*) < v_{i'j'}(t_{j's})$.
Now shift a small amount of flow ϵ originated at on-ramp i' using off-ramp j' from time $t_{j's}$ to time t^*. This manipulation can reduce the total system cost by $\epsilon\alpha(t^* - t_{j's}) > 0$, contradicting with the optimum assumption.
In summary, $v_{\bullet j}(t) = s_j \ \forall t \in (t_{js}, -T_j), j = 1, \ldots, J'$.

(2.1) According to (1) $(\tau_{i's}, -T_{s(i')})$ can be divided into two subintervals: $(\tau_{i's}, t_{s(i')s}), [t_{s(i')s}, -T_{s(i')})$.

(a) For $t \in (\tau_{i's}, t_{s(i')s})$, suppose $\exists t^* \in (\tau_{i's}, t_{s(i')s}), v_{i'\bullet}(t^*) < s_{i'}$.Since $v_{i'\bullet}(\tau_{i's}) > 0$ and $v_{\bullet s(i')}(\tau_{i's}) > 0$, one can always swap flows such that $v_{i's(i')}(\tau_{i's}) > 0$.
Now shift a small amount of flow ϵ originated at on-ramp i' using off-ramp $s(i')$ from time $\tau_{i's}$ to time t^*. This manipulation can reduce the total system cost by $\epsilon\alpha(t^* - \tau_{i's})$, contradicting with the optimality assumption.

(b) For $t \in [t_{s(i')s}, -T_{s(i')})$, suppose $\exists t^* \in [t_{s(i')s}, -T_{s(i')}), v_{i'\bullet}(t^*) < s_{i'}$.
According to (a) one can always swap flows such that $v_{i's(i')}(\tau_{i's}) > 0$.
By Proposition 3, $\forall t' \in (\tau_{i's}, t_{s(i')s}), v_\bullet(t') < v_\bullet(t^*)$. Hence, there exists one on-ramp i'' such that $v_{i''\bullet}(t') < v_{i''\bullet}(t^*)$.
Since $v_{i''\bullet}(t^*) > 0$ and $v_{\bullet s(i')}(t^*) > 0$, one can always swap flows such that $v_{i''s(i')}(t^*) > 0$.Now shift a small amount of flow ϵ originated at on-ramp i'' using off-ramp $s(i')$ from time t^* to time t' and the same amount of flow originated at on-ramp i' using off-ramp $s(i)$ from time $\tau_{i's}$ to time t^*. This manipulation can reduce the total system cost by $\epsilon\alpha(t - \tau_{is})$, contradicting with the optimality assumption.
In summary, $\forall t \in (\tau_{is}, -T_{s(i)}), v_{i\bullet}(t) = s_i$.

(2.2) For any on-ramp $i' = I, \ldots, n$, the interval $(-\infty, \tau_{i'0}) \cup (T_{o(i)}, 0)$ can be divided into the following two categories: (1) $t \in (t_{o(i')}, \tau_{i'0}) \cup_{j' < o(i')} (t_{j'0}, t_{j's})$ and (2) $t \in \cup_{j' < o(i')}[(t_{j's}, t_{(j'+1)0}) \cup (-T_{j'}, -T_{j'+1})]$.

(a) Evidently, $v_{i'\bullet}(t) = 0 \ \forall t \in (t_{o(i')}, \tau_{i'0}) \cup_{j' < o(i')} (t_{j'0}, t_{j's})$.

(b) Suppose $\exists t^* \in (t_{j's}, t_{(j'+1)0}) \cup (-T_{j'}, -T_{j'+1}), j' < o(i')$ such that $v_{i'\bullet}(t^*) > 0$.Since $v_{i'\bullet}(t^*) > 0$ and $v_{\bullet j'}(t^*) > 0$, one can always swap flow such that $v_{i'j'}(t^*) > 0$.
By Proposition 3 $\forall t' \in (t_{o(i')0}, \tau_{i0}), v_\bullet(t^*) < v_\bullet(t')$. Hence, there exists one on-ramp i'' such that $v_{i''\bullet}(t^*) < v_{i''\bullet}(t')$.
Since $v_{i''\bullet}(t') > 0$ and $v_{\bullet o(i')}(t') > 0$, one can always swap flows such that $v_{i''o(i')}(t') > 0$. Now shift a small amount of flow ϵ originated at on-ramp i' from using off-ramp j' at time t^* to using off-ramp $o(i')$ at time $\tau_{i'0}$ and at the same time shift the same amount of flow originated at on-ramp i'' from using off-ramp $o(i')$ at time t' to using off-ramp j' at time t^*. This manipulation can reduce the total system cost by $\epsilon\alpha(\tau_{i0} - t') > 0$, contradicting with the optimality assumption.
In summary, $v_{i\bullet}(t) = 0 \ \forall t \in (-\infty, \tau_{i0}) \cup (-T_{o(i)}, +\infty)$. □

Proof of Theorem 2. To prove Theorem 2, it suffices to show that the multipliers defined by (4)–(6) satisfy the optimality conditions (1)–(3).

As $\omega_i(t) > 0$ when $t \in (\tau_{is}, -T_{s(i)})$ and $\omega_i(t) = 0$ when $t \in (-\infty, \tau_{is}) \cup (-T_{s(i)}, +\infty)$, (1) is satisfied.

Similarly, $\mu_j(t) > 0$ when $t \in (t_{js}, -T_j)$ and $\mu_j(t) = 0$ when $t \in (-\infty, t_{js}) \cup (-T_j, +\infty)$, (2) is also satisfied.

Suppose off-ramp j is the last off-ramp in use at time t. Notice that $\mu_{j'}(t) + (1-\alpha)T_{j'} = \mu_j(t) + (1-\alpha)T_j \ \forall j' \leqslant j$ and $\mu_{j'}(t) + (1-\alpha)T_{j'} > \mu_j(t) + (1-\alpha)T_j \ \forall j' > j$. Hence, to show the satisfaction of (3), it suffices to show that if $v_{ij}(t) > 0$ where j is the last off-ramp in use at time t, $-\alpha t + (1-\alpha)T_j + \omega_i(t) + \mu_j(t) = \lambda_i = -\alpha\tau_{i0} + (1-\alpha)T_{o(i)}$.

If $v_{ij}(t) > 0$, there are five possibilities:

(1) $t \in (t_{j's}, t_{(j'+1)0}), o(i) \leqslant j' < s(i)$,
(2) $t \in (t_{j'0}, t_{j's}), s(i) \leqslant j' \leqslant J'$,
(3) $t \in (t_{j's}, t_{(j'+1)0}), s(i) \leqslant j' \leqslant J'$,
(4) $t \in (-T_{j'+1}, -T_{j'}), s(i) \leqslant j' < J'$,
(5) $t \in (-T_{j'+1}, -T_{j'}), o(i) \leqslant j' < s(i)$.

(1) $t \in (t_{j's}, t_{(j'+1)0}), o(i) \leqslant j' < s(i)$
In this case, $\omega_i(t) = 0$. Hence

$$-\alpha t + (1-\alpha)T_j + \omega_i(t) + \mu_j(t) = -\alpha t + (1-\alpha)T_j + \alpha(t - t_{js}) = -\alpha t_{js} + (1-\alpha)T_j$$

(2) $t \in (t_{j'0}, t_{j's}), s(i) \leqslant j' \leqslant J'$
In this case, it is easy to check that $\omega_i(t) + \mu_j(t) = \alpha(t - \tau_{is}) - (1-\alpha)(T_j - T_{s(i)})$. Hence

$$-\alpha t + (1-\alpha)T_j + \omega_i(t) + \mu_j(t) = -\alpha\tau_{is} + (1-\alpha)T_{s(i)}$$

(3) $t \in (t_{j's}, t_{(j'+1)0}), s(i) \leqslant j' \leqslant J'$
In this case, again it is easy to check that $\omega_i(t) + \mu_j(t) = \alpha(t - \tau_{is}) - (1-\alpha)(T_j - T_{s(i)})$. Hence, $-\alpha t + (1-\alpha)T_j + \omega_i(t) + \mu_j(t) = -\alpha\tau_{is} + (1-\alpha)T_{s(i)}$.

(4) $t \in (-T_{j'+1}, -T_{j'}), s(i) \leqslant j' < J'$
In this case, it is easy to check that $-\alpha t + (1-\alpha)T_j + \omega_i(t) + \mu_j(t) = -\alpha\tau_{is} + (1-\alpha)T_{s(i)}$.

(5) $t \in (-T_{j'+1}, -T_{j'}), o(i) \leqslant j' < s(i)$
In this case, $\omega_i(t) = 0$. Hence

$$\begin{aligned} -\alpha t + (1-\alpha)T_j + \omega_i(t) + \mu_j(t) &= -\alpha t + (1-\alpha)T_j + (1-\alpha)(T_{j+1} - T_j) + \alpha(t - t_{(j+1)0}) = -\alpha t_{(j+1)0} + (1-\alpha)T_{j+1} \\ &= -\alpha t_{js} + (1-\alpha)T_j \end{aligned}$$

Finally, $\lambda_i = -\alpha\tau_{i0} + (1-\alpha)T_{o(i)}$. If $o(i) = s(i)$, evidently $\tau_{i0} = \tau_{is}$. Hence, $\lambda_i = -\alpha t_{is} + (1-\alpha)T_{s(i)}$. If $o(i) < s(i)$, then $\tau_{i0} = t_{o(i)s}$ and $t_{j'0} = t_{j's}\ \forall o(i) < j' < s(i)$. Hence, $\lambda_i = -\alpha t_{j's} + (1-\alpha)T_{j'}\ \ \forall o(i) \leqslant j' < s(i)$.

Hence, in all the five cases shown above, $-\alpha t + (1-\alpha)T_j + \omega_i(t) + \mu_j(t) = \lambda_i$.

In summary, any traffic flow pattern satisfying the properties given in Theorem 1 is an optimal traffic flow pattern on the freeway system. □

References

Arnott, R., de Palma, A., Lindsey, R., 1990. Departure time and route choice for the morning commute. Transportation Research Part B 23 (3), 209–228.

Arnott, R., de Palma, A., Lindsey, R., 1993. Properties of dynamic traffic equilibrium involving bottlenecks, including a paradox and metering. Transportation Science 27 (2), 148–160.

Beckmann, M.J., 1965. On optimal tolls for highways, tunnels and bridges. In: Vehicular Traffic Science. American Elsevier, New York, pp. 331–341.

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

Chang, G.-L., Mahmassani, H.S., Engquist, M.L., 1988. System-optimal trip scheduling and routing in commuting networks. Transportation Research Record 1251, 54–65.

Dafermos, S.C., Sparrow, F.T., 1971. Optimal resource allocation and toll patterns in user-optimized transport network. Journal of Transport Economics and Policy 5, 184–200.

Daganzo, C.F., 1994. The cell transmission model: a dynamic representation of highway traffic consistent with the hydrodynamic theory. Transportation Research Part B 28 (4), 269–287.

Daganzo, C.F., 1995. The cell transmission model, part II: network traffic. Transportation Research Part B 29 (2), 79–93.

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

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

Kuwahara, M., 1990. Equilibrium queueing patterns at a two-tandem bottleneck during the morning peak. Transportation Science 24 (3), 217–229.

Merchant, D.K., Nemhauser, G.L., 1978a. A model and an algorithm for the dynamic traffic assignment problems. Transportation Science 12 (3), 183–199.

Merchant, D.K., Nemhauser, G.L., 1978b. Optimality conditions for a dynamic traffic assignment model. Transportation Science 12 (3), 200–207.

Munoz, J.C., Laval, J.A., 2006. System optimum dynamic traffic assignment graphical solution method for a congested freeway and one destination. Transportation Research Part B 40 (1), 1–15.

Netter, M., 1971. Equilibrium and marginal cost pricing on a road network with several traffic flow types. Transportation Research Part A 32 (1), 45–54.

Newell, G.F., 1987. The morning commute for nonidentical travelers. Transportation Science 21 (2), 74–88.

Peeta, S., Mahmassani, H.S., 1995. System optimal and user equilibrium time-dependent traffic assignment in congested networks. Annals of Operations Research 60, 81–113.

Shen, W., Nie, Y., Zhang, H.M., 2007. On path marginal cost analysis and its relation to dynamic system-optimal traffic assignment. In: Proceedings of the 17th International Symposium on Transportation and Traffic Theory, London, pp. 319–352.

Vickrey, W.S., 1969. Congestion theory and transport investment. American Economic Association 59 (2), 251–261.

Wie, B.W., Tobin, R.L., 1998. Dynamic congestion pricing models for general traffic networks. Transportation Research Part B 32 (5), 313–327.

Yang, H., Huang, H.J., 1998. Principle of marginal-cost pricing: how does it work in a general road network? In: Proceedings of the 5th International Symposium on Transportation and Traffic Theory, Berkeley, pp. 331–341.

Yang, H., Meng, Q., 1998. Departure time, route choice and congestion toll in a queuing network with elastic demand. Transportation Research Part B 32 (4), 247–260.

Ziliaskopoulos, A.K., 2000. A linear programming model for the single destination system optimum dynamic traffic assignment problem. Transportation Science 34 (1), 37–49.