Emission Modeling and Pricing on Single-Destination Dynamic Traffic Networks

Rui Ma [a], Xuegang (Jeff) Ban [b,*], W.Y. Szeto [c]

[a] *Department of Civil and Environmental Engineering, University of California, 2041 Academic Surge, 1 Shield Avenue, Davis, CA 95616, USA*

[b] *Department of Civil and Environmental Engineering, University of Washington, 121G More Hall, Seattle, WA 98195, USA*

[c] *Department of Civil Engineering, The University of Hong Kong, Rm 622 Haking Wong Building, Pokfulam Road, Hong Kong, China*

Abstract

This paper proposes an emission pricing model for single-destination dynamic traffic networks. The model contains two sub-models derived from the corresponding two sub-problems: a system optimum dynamic traffic assignment problem and a first-best dynamic emission pricing scheme. For the first problem, it proves that under certain conditions, an optimal solution, if exists, must be a free-flow solution to minimize the generalized system cost including the costs of total travel times and total emissions (or fuel consumption). The optimal first-best emission pricing can then be determined by solving an optimal control problem, using the free-flow dynamic system optimal solution as the input. Numerical results are provided to illustrate the proposed models and the solution methods.

Ma, R., Ban, X. J., & Szeto, W. Y. (2017). Emission modeling and pricing on single-destination dynamic traffic networks. Transportation Research Part B: Methodological, 100, 255-283.

Keywords: emission pricing; aggregated emission functions; dynamic system optimum; free-flow optimal solution; dynamic user equilibrium; first-best pricing

1. Introduction and Motivation

Air pollution has been recognized as acute negative effects in many urban areas. Recently, some major urban areas in Eastern Asia have to face reduced visibility and deteriorated air quality frequently, largely due to emissions (such as smug) from industry and urban traffic. Reducing the emissions from urban traffic becomes necessary and urgent for a sustainable urban environment. Traffic emission programs and regulations on vehicle population and emission rates have been implemented by many regions in the last decades (Netherlands Ministry of Housing, Spatial Planning, and the Environment (2004); Fung et al. (2010)). However, the objective to reduce air pollution has not been fully achieved by these programs alone. Besides government regulations, market-based approaches for network-wide emission control have been proposed and studied recently.

As introduced in the marginal cost pricing theory in Pigou (1920) and its later developments in Walters (1961) and Vichrey (1969), road pricing has been recognized as a key market-based approach to efficiently

* Corresponding author. Tel.: +1-518-276-8043 ; fax: +1-518-276-4833.
E-mail address: banx@uw.edu

© Please cite the paper as
Ma, R., Ban, X. J., & Szeto, W. Y. (2017). Emission modeling and pricing on single-destination dynamic traffic networks. Transportation Research Part B: Methodological, 100, 255-283.

internalize the external costs imposed by the drivers on other users in the network, and to maximize the total social welfare. As a special form of road pricing, network-wide emission pricing aims at controlling the emission externalities in traffic networks to minimize the total externalities, including the environmental costs.

Emission pricing studies on different objectives have been performed in the recent decades. Some emission pricing models considered traffic emissions as the single indicator for the externalities (see Yin and Lu (1999); Hizir (2006); Sharma and Mishra (2011)), while more and more studies simultaneously incorporated the system time costs, e.g., travel and waiting times, in addition to the environmental costs, into the emission pricing model. To achieve the system optimal cost including both time and environmental costs, Johansson (1997) suggested that both the congestion effects and the environmental externalities should be considered.

It was found in Rilett and Benedek (1994) and Yin and Lu (1999) that the objectives of reducing time costs and traffic emissions can be conflicting in traffic networks. There are generally two approaches to balance the trade-off between time and environmental costs. One approach is to embed side constraints in the model, such as the environmental capacity constraints; see Xu et al. (2015); Zhong et al. (2012); Li et al. (2012). The other approach is to propose multi-objective formulations by applying the weighted sum method (Yin and Lawphongpanich (2006); Abou-Zeid (2003); Qiu and Chen (2007); Ferguson et al. (2010); Szeto et al. (2014)), so that both types of costs are combined as a single objective. Yin and Lawphongpanich (2006) showed that a Pareto solution set can be constructed by solutions with varied weights of such two types of costs. Other studies incorporated more dimensions such as social costs into the objective functions; see a comprehensive review in Szeto et al. (2012). Most of the previous studies on network-wide emission pricing focused on static traffic networks, with a few exceptions; see Zhong et al. (2012); Friesz et al. (2013); Kickhöfer and Kern (2015); Kickhöfer and Nagel (2016).

Similar to the congestion pricing problems, network-wide emission pricing can be broadly categorized as first-best pricing and second-best pricing. The former assumes that prices (or tolls) can be imposed on any location (e.g., links) of the network at any time, while the latter assumes that tolls can only be imposed on a selected list of locations in the network or during a particular time period. It is well known that the second-best pricing can be formulated and solved as bi-level problems; see Yang and Bell (1997), Patriksson and Rockafellar (2002), Clegg et al. (2001), Liu and Boyce (2002), Abou-Zeid (2003) and Szeto et al. (2015) for static problems, and Ban and Liu (2009) and Friesz et al. (2007) for dynamic problems. First-best pricing for static networks can be modeled as marginal cost problems and solved accordingly, see Yang and Huang (2005). Such marginal cost approach, however, becomes complicated when applying to dynamic networks mainly because the traffic dynamics (such as flow propagation over time and space) introduce extra terms (such as the terms accounting for inter-temporal effects, see Carey and Srinivasan (1993); Shen et al. (2007)) in the marginal cost formulation that makes the marginal cost much harder to compute. However, implementing road pricing on dynamic networks is more realistic, because time-varying tolls can be implemented for traffic dynamics and more realistic behavior in the modeling framework. More importantly, Lo and Szeto (2005) showed that the static and dynamic modeling approaches can produce diametrically opposite results. They found that the impacts of pricing policies under the static approach could be ill-represented. In some cases, the pricing schemes such determined could actually worsen the congestion problem. This finding illustrates the importance of adopting the dynamic modeling approach for pricing, albeit it is more complex and computationally more demanding than the static modeling approach.

Table A.2 in Appendix A summarizes the literature on dynamic road pricing, from which we can conclude the following. First, most studies mainly focused on congestion pricing whereas only in the last few years, a handful studies considered emissions externalities when setting optimal prices. Secondly, the travel choice dimensions considered varied from one study to another, but for those studying emission pricing, at least two types of choices (among route, departure time and mode choices) were considered. Third, both first and second best pricing have been examined in the literature, including those considering emissions externalities. Second-best pricing seems to be more realistic and practical at the first sight, because not all links are allowed to charge a toll due to various physical or political constraints. However, with the recent emerging technologies, such as mobile sensing (Herrera et al., 2010; Ban et al., 2009a; Ban and Gruteser, 2012; Hao et al., 2012) and connected/automated vehicles , it is possible that each vehicle in the near future can be equipped with a tracking device so that, at least in principle, it can be charged a toll anywhere at any

time if needed (assuming other related issues such as privacy and security can be properly addressed). This will make the first-best pricing idea feasible and probably more appealing. Moreover, although using microsimulation models may give better estimates of emissions, the analytical model presented in this paper allows more rigorous analyses of the model properties and thus provides useful insights about how to design and implement emission pricing schemes.

In this paper, we focus on the first-best emission pricing on a dynamic traffic network. We aim to develop a modeling framework to determine the optimal dynamic pricing schemes for all links of the network so as to minimize the generalized system costs, including total travel times and emission costs. The formulation is link-node based as in Ban et al. (2008, 2012b), which avoids the use of path-specific variables. The optimization objective function in the proposed model incorporates both the economic and environmental externalities. More specifically, the economic externalities refer to the total travel and waiting time costs, while the environmental externalities refer to the total emissions (or fuel consumption). In the literature, there are (at least) two types of methods for emission estimation modeling. Microscopic emission models focus on vehicle-specific driving behaviors, including vehicle types as well as acceleration and deceleration patterns. Macroscopic emission models use aggregated functions based on aggregated traffic data, such as average speeds, or speed distributions, to estimate the total emissions. For network-level analysis, macroscopic/aggregated emission models are more commonly used. In this paper, we follow such common practice and use macroscopic models since we focus on network-level emission control. In particular, we apply two types of aggregated functions. The first is based on the macroscopic emission model developed in Wallace et al. (1998) for carbon monoxide (CO). The second is for fuel consumption estimation based on more recent research results in Boriboonsomsin et al. (2012), which can be viewed as a proxy of multiple pollutants.

The proposed first-best emission pricing problem can be decomposed and solved as two separate sub-problems. For the first sub-problem, a dynamic system optimum (DSO) model with a combined objective function is formulated and solved as an optimal control problem. It is proved that when the emission function is monotonic and an optimal solution exists, the optimal solution must be a free-flow solution, i.e., a solution where the travel times of all links are free-flow travel times and the combined objective function is minimized. It is further shown that the free-flow optimal solution can be derived via a specifically designed procedure. The second sub-problem determines the dynamic optimal emission tolls on all network links to sustain the desired DSO solution obtained from the first problem. It is formulated as an optimal control problem with the objective to minimize the total emission tolls, while satisfying the dynamic user equilibrium (DUE) route choice conditions and the related tolling constraints. The fact that the desired solution is a free flow solution greatly simplifies the model formulation and the solution process. Experimental results on multiple testing networks are presented to illustrate the proposed first-best emission pricing model and its solution process.

The desired free-flow DSO solution is crucial for the proposed emission modeling framework. Firstly, the free-flow solution greatly simplifies the traffic dynamics involved in the network models, e.g., it reduces the time-varying, state-dependent delays to constant delays in the DUE problem. This makes the solution process of the proposed emission pricing problem much easier. Secondly and more importantly, focusing on free-flow solutions of traffic networks offers new perspectives of designing network-wide emission pricing schemes, which has not been fully discussed before. Notice that the free-flow solution here is distinctively different from the solution to static traffic assignment. Static assignment ignores dynamics in a traffic network (such as time-varying demands, traffic dynamics, etc.) and considers traffic congestion in a simplified, aggregated manner, e.g., via the BPR (Bureau of Public Roads) type of link travel time functions. The solution to a static assignment should not be a free-flow solution. The free-flow solution to a DSO problem, on the other hand, is a special DSO solution that considers the dynamics of a traffic network as essential components. Calculating a free-flow DSO solution, especially determining the time-varying demand profile at each origin that can lead to DSO, is not trivial as shown in Ma et al. (2014).

The major contributions of the paper are summarized as follows:

- It proves that under certain conditions, a free-flow optimal solution to the DSO problem can be found to minimize the generalized cost including the costs of system travel times and emissions.

- The optimal first-best emission pricing can be determined by solving an optimal control problem, using the obtained free-flow optimal DSO solution as the input. This avoids the use of the marginal cost approach to derive first-best pricing, which can be complicated due to traffic dynamics.
- The results of the paper have important practical implications for the design of pricing and related traffic management strategies. For example, pricing may be designed around the free-flow traffic state in order to minimize the generalized cost including the costs of system travel times and emissions. This can much simplify the computation of the optimal pricing scheme in a dynamic traffic network, while at the same time the derived pricing scheme can be used to sustain the desired free-flow traffic state in the network.

The rest of this paper is organized as follows. Section 2 briefly describes the double-queue link model, the network and notations. Section 3 presents the macroscopic emission framework adopted in the paper to illustrate the proposed network emission control model. Section 4 develops an emission pricing model, which is formulated based on two sub-problems: a dynamic system optimum problem and a first-best emission pricing problem. Section 5 shows the numerical results on three testing networks. Section 6 concludes the paper.

2. Network and Notations

In this section, we first summarize the double-queue link dynamics model, and then describe the traffic network structure and the notations used throughout the paper.

2.1. The double-queue model

The "double-queue" model provides a concise yet accurate way to describe the evolution of traffic flow over time and space, which is consistent with the kinematic wave theory. The concept was originally derived from the link transmission model (LTM) (Yperman, 2007) based on the triangular fundamental diagram of traffic flow in Newell (1993), as shown in Figure 1. The model was first proposed and used in Osorio et al. (2011) for modeling stochastic DUE problems in discrete time. It was extended to continuous-time in Ma et al. (2014, 2015a) to model the continuous-time DSO and DUE problems, respectively. In this section, we provide a brief review and main results of the double queue model; more details can be found in Ma et al. (2014). We start with the triangular fundamental diagram.

The triangularly shaped fundamental diagram of link (i,j) can be defined by three parameters: a fixed free-flow speed (v_{ij}^f), a flow capacity ($\overline{C}_{ij}$) and a jam density (k_{ij}^{jam}). The critical density is defined as $k_{ij}^M \triangleq \overline{C}_{ij}/v_{ij}^f$. On the left branch of the triangular fundamental diagram, where the density $k_{ij} < k_{ij}^M$, the traffic moves in a fixed free-flow speed v_{ij}^f. On the right branch, where the density $k_{ij} > k_{ij}^M$, the traffic is in a congested state. The maximum backward shockwave speed ω_{ij} is presented by the absolute value of the slope of the right branch. Given the link length l_{ij}, the link free flow travel time is defined as $\tau_{ij}^0 \triangleq l_{ij}/v_{ij}^f$. Moreover, $\tau_{ij}^w \triangleq l_{ij}/\omega_{ij}$ is the time that is needed for the queuing shockwave (with speed ω_{ij}) to travel from the end of the link to the beginning of the link at the congested state. It is actually the minimum travel time for the backward shockwave to travel the entire link. We define τ_{ij}^w as 'the shockwave travel time' hereafter in this paper. From Figure 1, it is observed that usually we have $\tau_{ij}^0 < \tau_{ij}^w$.

Specifically, the dynamics of link (i,j) in the double-queue model can be described by a downstream queue $q_{ij}^d(t)$ and an upstream queue $q_{ij}^u(t)$ in the differential form:

For almost all time $t \in [0,T]$,

$$\begin{aligned} \dot{q}_{ij}^d(t) &= p_{ij}(t-\tau_{ij}^0) - v_{ij}(t), \\ \dot{q}_{ij}^u(t) &= p_{ij}(t) - v_{ij}(t-\tau_{ij}^\omega). \end{aligned} \tag{1}$$

Here $\dot{q}(t)$ denotes the derivative of q over time t. $p_{ij}(t)$ is the inflow rate to link (i,j) at time t, and $v_{ij}(t)$ is the exit flow rate from link (i,j) at time t. The downstream queue dynamics are similar to the queue

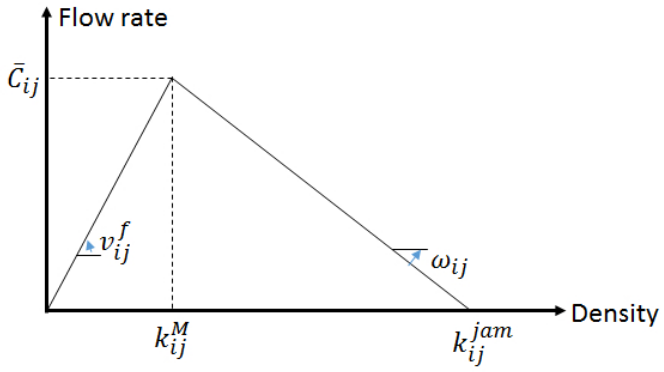


Figure 1: Triangularly shaped fundamental diagram

dynamics of the Linear Complementarity System-based point-queue model in Ban et al. (2012a), while the upstream queue is unique in the double-queue model. In particular, the upstream queue does NOT have a physical meaning; both queues are referred to as "hidden states" in Ma et al. (2014) to regulate flow into or out of the link. Also notice that, $n_{ij}(t)$, the number of vehicles on the link at time t is:

$$n_{ij}(t) = \int_0^t p_{ij}(\xi) - v_{ij}(\xi)d\xi \tag{2}$$

Clearly these two queues (and the number of vehicles on the link) are not independent since they are determined by the link inflow and exit flow rates (but at different time instants). The upstream and downstream queues are the main mechanism that enables the double queue model to properly capture the impact on the upstream flow due to congestion at the downstream. In particular, the two queues function as the "gates" at the entrance and exit of a link to regulate flows in and out of the link. Usually when $q_{ij}^u(t) = \overline{Q}_{ij} = (\tau_{ij}^0 + \tau_{ij}^\omega)\overline{C}_{ij}$ is first reached on a link, it indicates that the queue spillback occurs, the traffic state at the entrance of the link transits from the free-flow state to the congested state, and the inflow to the link may be restricted. If $q_{ij}^d(t) > 0$, on the other hand, queue exists at the end of the link, due to downstream congestion or the inflow exceeding the exit flow capacity of the link. In this case, the exit flow from the link will reach its capacity if there is no queue spillback from the downstream. From their definitions, it is easy to see that $0 \le q_{ij}^d(t) \le n_{ij}(t) \le q_{ij}^u(t) \le \overline{Q}_{ij}$. More discussions of the double queue model and its properties can be found in Ma et al. (2014).

The actual travel time for vehicles entering link (i, j) at time t is implicitly derived from the cumulative curve of inflow and exit flow.

$$\int_0^t p_{ij}(\xi)d\xi = \int_0^{t+\tau_{ij}(t)} v_{ij}(\xi)d\xi \tag{3}$$

The implicit expression of travel time leads to integral forms in modeling and calculation. To simplify the model, we estimate a lower bound of the implicit double-queue link travel time (3), by borrowing the travel time function in point-queue model in Ban et al. (2012a), as shown in (4).

$$\tau_{ij}(t) \approx \tau_{ij}^0 + \overline{C}_{ij}{}^{-1} q_{ij}^d(t + \tau_{ij}^0) \tag{4}$$

The approximated link travel time (4) equals to the actual one, if link (i, j) is free-flowing (i.e., $q_{ij}^d(t+\tau_{ij}^0) = 0$), or the actual exit flow rate $v(t + \tau_{ij}^0) = \overline{C}_{ij}$ when there is a positive downstream queue $q_{ij}^d(t + \tau_{ij}^0) > 0$. When queue spillback from downstream links reduces the actual exit flow rate to be lower than the flow capacity $\overline{C}_{ij}$, the approximated link travel time (4) is normally smaller than the actual one. In this paper, we use the approximated link travel time (4) throughout this paper to simplify the models. More importantly, we will show in later sections that under certain conditions, the optimal solution is free-flowing, for which the approximated link travel time is exact.

2.2. The traffic network

Following the network structure introduced in Ma et al. (2014), we denote a traffic network as $G(N, L)$, where N is the set of nodes and L is the set of links. We consider a traffic network with multiple origins and a single destination (denoted as $\widehat{s}$), i.e., a many-to-one network. All the origins and the destination are nodes in N. For origin o, the total demand is D_o. We denote the number of origins by n.

The regular network G is expanded by adding a dummy origin node $\widetilde{o}_k$ and a corresponding dummy origin link $(\widetilde{o}_k, o_k)$ at each origin node o_k, as described in Ma et al. (2014) in details. The distinction between a dummy node $\widetilde{o}$ and the origin node o is that while there may be incoming links to o with o as the ending node, there is no such incoming link into $\widetilde{o}$. Moreover, there is a unique outgoing link starting at $\widetilde{o}$, namely, the corresponding dummy link. The set of dummy links is L_d, i.e., $(\widetilde{o}_k, o_k) \in L_d$, $k = 1, 2, \cdots, n$. The free-flow and shockwave travel times are both zero for the dummy links; the exit capacity and queue capacity on these links are infinite. In the testing examples, they are set as very large. Since the free-flow and shockwave travel times are both zero, the upstream and downstream queues reduce to a single queue. However, such single queue still behaves differently from a point-queue. This is because the exit flow rate of a dummy link is constrained (actually determined) by adjacent downstream links, and thus can never reach its exit flow capacity (which is infinite), which is very different from a point-queue model. Furthermore, for a dummy origin link $\widetilde{o}, o$, initially, the total demand D_o from the origin node o (which is given and fixed) is queued on the dummy link $(\widetilde{o}, o)$ to be discharged. That is, there is no inflow to a dummy origin link, and the demand on each dummy origin link is presented as a non-increasing queue, whose discharging rate is the demand rate from the corresponding origin node to the regular network.

Since the paper applies the DSO model in Ma et al. (2014), a simple junction model is used here, as shown in Section 4.1. This is very different from DUE models in which a more complicated junction model is usually needed to capture the behavior of traffic at junctions. In the DSO model applied in this paper, we do not specify such behavior at junctions, which is however determined by the DSO principle implicitly. More discussions on this are provided in Section 4.1.1.

2.3. Notations

For a link $(i, j) \in \mathcal{L} \triangleq L \cup L_d$, we define:
(constant) parameters: all positive scalars

$\overline{Q}_{ij}$ upper bound of upstream queue length
$\overline{C}^p_{ij}$ upper bound of inflow rate
$\overline{C}^v_{ij}$ upper bound of exit flow rate
τ^0_{ij} free-flow travel time
τ^ω_{ij} shockwave travel time (congested); in general $\tau^0_{ij} \leq \tau^\omega_{ij}$
l_{ij} link length
ρ_{ij} emission impact coefficient

(time-dependent) trajectories: all non-negative

$\tau_{ij}(t)$ link travel time for vehicles entering the link at time t
$p_{ij}(t)$ inflow rate at time t
$v_{ij}(t)$ exit flow rate at time t
$q^u_{ij}(t)$ upstream queue length at time t
$q^d_{ij}(t)$ downstream queue length at time t
$n_{ij}(t)$ number of vehicles at time t

3. Macroscopic Link Emission Models

Vehicular traffic generates various types of emission pollutants, including carbon monoxide (CO), carbon dioxide (CO_2), hydrocarbons (HC), nitrogen oxide (NO_x), and others. There are at least two types of methods for emission estimation modeling in the traffic network. Microscopic emission models focus on vehicle-specific driving behaviors, including vehicle types as well as acceleration and deceleration patterns. For instance, in Comprehensive Modal Emission Model (CMEM; see Scora and Barth (2006)) and Piccoli et al. (2015), acceleration/deceleration of cars were taken into account in estimating the emissions. Macroscopic emission models use aggregated functions based on aggregated traffic data, such as average speeds, or speed distributions, to estimate the total emissions. Microscopic emission estimation models gain higher degree of accuracy for emissions of a single vehicle with vehicle-specific data. However, they normally require vehicle-specific, high-resolution data of each vehicle in the network and more computationally demanding to model and estimate system-wide emissions. Macroscopic emission models, on the other hand, estimate the total emissions from the aggregated traffic states, so that they are more efficient in terms of computations. Both types of models are widely used in transportation applications. For network-level analysis, macroscopic/aggregated emission models are more commonly used. In this paper, we use macroscopic models since we focus on network (i.e., system) wide emissions. The most commonly used traffic data for macroscopic emission modeling is the link average speed, which are categorized as speed-based in Szeto et al. (2012). A handful of studies have found that for most pollutants, there exist certain nonlinear correlations between the amount of pollutants and the average speed at the link level. A detailed review of different types of emission models can be found in Szeto et al. (2012).

It turns out that in the literature, there are two types of aggregated emission functions. The first can estimate emissions (e.g., grams) per vehicle per unit time (e.g., per hour), while the second estimates emissions per vehicle per unit distance (e.g., per mile) traveled. It is unclear which type is the most appropriate. In practice, it is up to the user to pick which type of emission functions to use. In this paper, to illustrate the proposed method, both types of functions are applied. The first is for CO pollutants based on the work in Wallace et al. (1998), while the second estimates multiple pollutants and fuel consumption based on more recent research results in Boriboonsomsin et al. (2012).

We present the static models and their extensions to dynamic cases in the next two subsections. Notice that the dynamic emission models applied in this paper are simple time-dependent extensions from the static models. Developing a more comprehensive dynamic emission model is not trivial as it will require extensive data collection and analysis, which is beyond the scope of the work. Rather, using the simple dynamic emission models, we aim to show what properties an emission function needs to have in order to be mathematical tractable when applied to a dynamic network modeling framework.

3.1. Static macroscopic emission models

Wallace et al. (1998) proposed a non-linear, static emission model for network links. For link (i, j), the vehicular CO emissions per vehicle per hour is

$$e_{ij}(x_{ij}) = 0.2038\tau_{ij}(x_{ij}) \exp(\frac{0.7962 l_{ij}}{\tau_{ij}(x_{ij})}), \tag{5}$$

where l_{ij} is the link length (in kilometers), x_{ij} is the link state variable, and $\tau_{ij}(x_{ij})$ is the travel time (in minutes) for link (i, j). The emissions or fuel consumption $e_{ij}(x_{ij})$ is in grams per vehicle per hour. For the above static emission function, x_{ij} is usually the link flow.

A more recent study by Boriboonsomsin et al. (2012) developed a series of macroscopic emission models on fuel consumption and multiple pollutants by considering both the average speed and the road grade. For link (i, j), the vehicular emission per vehicle per mile is

$$f_{ij} = \exp\left(\beta_0 + \beta_1 s_{ij} + \beta_2 s_{ij}^2 + \beta_3 s_{ij}^3 + \beta_4 s_{ij}^4 + \beta_5 g_{ij}\right), \tag{6}$$

where s_{ij} is the link average speed (in miles per hour), g_{ij} is road grade (in percentage), and $\beta_i, i = 1, \cdots, 5$ are regression coefficients that vary for different pollutants. The emission or fuel consumption f_{ij} is in grams per vehicle per mile.

It is observed from both types of emission models (5) and (6) that they are non-linear, and monotonicity rarely exists for all the speeds. However, the emission functions above are monotonic functions under certain conditions. Next we present these conditions that will play an important role in analyzing the solution property of the emission pricing model in this paper.

In fact, the monotonicity property of the emission function in (5) only holds when the speed does not exceed 75km/h (46.6 mph). Yin and Lawphongpanich (2006) showed that $e_{ij}(\cdot)$ in (5) is increasing as long as the travel speed $\frac{l_{ij}}{\tau_{ij}(x_{ij})}$ is not too large. Xu et al. (2015) further proved that as long as the link travel time function is strictly increasing ($d\tau_{ij}(x_{ij})/dx_{ij} > 0$) and $0.7962 l_{ij} < \tau_{ij}(x_{ij})$ (time is in minutes, and link length is in kilometers), $e_{ij}(x_{ij})$ is monotonically increasing. Note that $0.7962 l_{ij} < \tau_{ij}(x_{ij})$ means the travel speed is less than $\frac{1}{0.7962}$ km/min ≈ 75 km/h (or 46.6 mph).

For the emission function (6), monotonicity conditions can also be established for specified pollutants. This is calculated and listed in Appendix B. Using the fuel consumption function as an example (used as a proxy of various pollutants in this paper), it can be expressed as (assuming the grade of road is zero):

$$f_{ij} = \exp\left(6.8 - 1.4 \times 10^{-1} s_{ij} + 3.92 \times 10^{-3} s_{ij}^2 - 5.2 \times 10^{-5} s_{ij}^3 + 2.57 \times 10^{-7} s_{ij}^4\right).$$

It is found that for a link with no road grade, when the traffic speed is less than 72 mph, the emission function $f_{ij}(s_{ij})$ is monotonically decreasing with respect to s_{ij}. When the speed is larger than 72 mph, the function is monotonically increasing.

The two types of emission models merit further discussions. As aforementioned, the unit in the emission model (5) is grams per vehicle per hour, while the unit in (6) is grams per vehicle per mile. In the first glance, it may seem trivial to unify the two models. Indeed, if one multiplies speed to the right hand side of equation (5), the unit will become grams per vehicle per mile . Similarly, if one divides speed to the right hand side of (6), the unit will become grams per vehicle per hour. Such conversion, however, is not practically feasible. The fuel consumption models, either in terms of gram per vehicle per mile or gram per vehicle per hour, are essentially estimated from the regression results using measured traffic and emission data. The form of each model (function) is pre-defined and the parameters are calibrated from real data; see Boriboonsomsin et al. (2012). As the pre-defined function form is not necessarily to realistically capture the physical relationship between traffic data (here speed) and emissions, the model may only work well for the assumed unit due to its data-driven nature. Therefore, one may not directly 'create' a new emission model (e.g., in terms of gram per vehicle per hour) by simply manipulating (e.g., by multiplying speed) to a calibrated emission model (e.g., in the unit of gram per vehicle per mile). If one aims to develop such a model (in gram per vehicle per hour), similar data-driven process needs to be followed, i.e., pre-define a functional form, and then collect emission data (e.g., in the unit of gram per vehicle per mile) and speed. The parameters of the function can then be calibrated using measured data. Due to the above considerations, we apply both types of models to test the model and algorithm proposed in the paper.

3.2. Extension to dynamic emission models

The dynamic emission functions applied in this paper extend the above emission models developed in the past by incorporating the time-varying traffic state variable. First, for the static emission model (5) that was based on static link travel times, we use the time-varying link travel times to extend it into dynamic cases. For double-queue link dynamics, we use the time-varying link travel time approximation (4) to calculate the emissions in (5). (Note that such approximation is exact once the link is in free-flow state.) For link (i, j), the vehicular CO emission rate, i.e., grams of CO emissions per vehicle per hour, at time t is:

$$e_{ij}(t) = \mathcal{E}_{ij}^{\tau}(\tau_{ij}(t)) = 0.2038 \rho_{ij} \tau_{ij}(t) \exp\left(\frac{0.7962 l_{ij}}{\tau_{ij}(t)}\right). \tag{7}$$

where the emission can be expressed as a function of the travel time as $\mathcal{E}_{ij}^{\tau}(\tau) = 0.2038\rho_{ij}\tau\exp(\frac{0.7962l_{ij}}{\tau})$ according to (5).

Here $\tau_{ij}(t)$ is the travel time (in minutes) at time t for link (i,j). ρ_{ij} is the emission impact coefficient, which may be different from link to link. The dynamic emission rate $e_{ij}(t)$ is in grams per vehicle per hour. The link travel time for double-queue model (4) is estimated as a function of the downstream queue, i.e., the state variable in the emission function is the downstream queue with a constant time delay $q_{ij}^d(t+\tau_{ij}^0)$. As (4) is usually a lower bound of actual travel time, (7) is a lower bound of actual emission rate. It is clear that when the link is in free-flow state, i.e., the downstream queue is zero, the emission rate is exact.

The link CO emissions over the time period $[t_1, t_2]$ can then be obtained as:

$$E_{ij}|_{t_1}^{t_2} = \int_{t_1}^{t_2} n_{ij}(t)e_{ij}(q_{ij}^d(t+\tau_{ij}^0))dt \tag{8}$$

where $n_{ij}(t)$ is the number of vehicles on link (i,j) at time t. Then we have the network-wide total emission (denoted as TE) over time period $[0,T]$ as:

$$\text{TE} = \sum_{(i,j)\in L} E_{ij}|_0^T = \sum_{(i,j)\in L}\int_0^T n_{ij}(t)e_{ij}(q_{ij}^d(t+\tau_{ij}^0))dt \tag{9}$$

For the fuel consumption and emission functions (6), we use the time-varying link travel time (4) to calculate the speed first. Assuming all links have zero road grade, the dynamic fuel consumption/emission function can be derived and shown in (10):

$$\begin{cases} f_{ij}(t) = F^s(s_{ij}(t)) & = exp\left(\beta_0 + \beta_1 s_{ij}(t) + \beta_2 s_{ij}^2(t) + \beta_3 s_{ij}^3(t) + \beta_4 s_{ij}^4(t)\right) \\ s_{ij}(t) = S_{ij}^q(q_{ij}^d(t+\tau_{ij}^0)) & = \dfrac{l_{ij}}{\tau_{ij}^0 + \overline{C}_{ij}{}^{-1} q_{ij}^d(t+\tau_{ij}^0)} \end{cases} \tag{10}$$

where $f_{ij}(t)$ is the grams of fuel consumption (or emissions of a specified pollutant) per vehicle per mile on link (i,j) at time t, when the corresponding speed is $s_{ij}(t)$, which is calculated by the estimated link travel time. That is, the fuel consumption/emissions can be expressed as a function of the speed as $F^s(s) = exp\left(\beta_0 + \beta_1 s + \beta_2 s^2 + \beta_3 s^3 + \beta_4 s^4\right)$, according to (6). The speed can be expressed as a function of the downstream queue for link (i,j) as $S_{ij}^q(q^d) = \dfrac{l_{ij}}{\tau_{ij}^0 + \overline{C}_{ij}{}^{-1} q^d}$.

Here we can easily show that in order to establish the monotonic increasing property for the emission function $f_{ij}(t)$ respect to the downstream queue $q_{ij}^d(t+\tau_{ij}^0)$, we need two properties: (i) the monotonic decreasing property with respect to speed for the fuel consumption / emission, i.e., $F^s(\cdot)$ is a decreasing function; and (ii) the monotonic decreasing property with respect to the downstream queue for the speed, i.e., i.e., $S_{ij}^q(\cdot)$ is a decreasing function. The second condition is obviously true, given l_{ij}, τ_{ij}^0 and $\overline{C}_{ij}$ are positive.

To summarize, for the two types of emission functions with different units of emissions, we have the following conditions.

(i) If $\mathcal{E}_{ij}^{\tau}(\cdot)$ in equation (7) (emissions per vehicle per hour) is a monotonically *increasing* positive function, then $e_{ij}(t)$ is an increasing function with increasing $q_{ij}^d(t+\tau_{ij}^0)$, the downstream queues with constant time shift;

(ii) If $F^s(\cdot)$ in equation (10) (emissions per vehicle per mile) is a monotonically *decreasing* function, then $f_{ij}(t)$ is an increasing function with increasing $q_{ij}^d(t+\tau_{ij}^0)$, the downstream queues with constant time shifts. (Since $S_{ij}^q(\cdot)$ is a decreasing function respect to $q_{ij}^d(t+\tau_{ij}^0)$.)

The above two statements establish conditions under which emissions or fuel consumption is an increasing function of the downstream queue. They are crucial for the analysis presented in this paper. We also notice that such monotonicity condition only holds for certain speed range for a given pollutant, which varies if different pollutants are concerned. This adds complexity if one aims to control multiple pollutants at the

same time. More discussions can be found in the Conclusion Section of the paper. For the emissions of a single pollutant (as well as the fuel consumption), the above conditions are satisfied for certain ranges of the traffic speed or the queue length, which are listed in details in Appendix C.

From (10), taking link fuel consumption as an example, the link fuel consumption over the time period $[t_1, t_2]$ can then be obtained as:

$$F_{ij}|_{t_1}^{t_2} = \int_{t_1}^{t_2} p_{ij}(t) f_{ij}(t) l_{ij} dt \tag{11}$$

Then we have the network-wide total fuel consumption (denoted as TF) over time period $[0, T]$ as:

$$\text{TF} = \sum_{(i,j)\in L} F_{ij}|_0^T = \sum_{(i,j)\in L} \int_0^T p_{ij}(t) f_{ij}(t) l_{ij} dt \tag{12}$$

Note that the unit of emissions in (7) is grams per vehicle per hour , while the unit of fuel consumption in (10) is grams per vehicle per mile. This is why the number of vehicles $n_{ij}(t)$ and the inflow rate $p_{ij}(t)$ are used in (8) and (11), respectively. Both dynamic emission functions in (7) and (10) are simple extensions from the static emission functions. We note here that these aggregated emission/fuel consumption functions with only one state variable (such as speed or travel times) are coarse. More research is needed in the future to develop more detailed emission/fuel consumption functions that can better capture the dynamic features of vehicles (such as their accelerations / decelerations). See the Conclusion Section for more discussions in this regard.

With the assumptions of the monotonicity properties of $\mathcal{E}_{ij}^{\tau}(\cdot)$ and $F^s(\cdot)$, it is clear that the link CO emissions and the fuel consumption are both monotonically increasing function to the downstream queue. Hereafter in this paper, we will focus on TE in (9) to develop the proposed emission pricing model. Similar analysis and results can also be obtained for TF in (12). Details are omitted in the paper for brevity.

4. Emission Pricing Model

The proposed emission pricing model can be decomposed into two sub-models, each of which is based on one sub-problem. First, a DSO problem with a combined objective function, which considers both the total time and total emission (or fuel consumption) costs, is formulated. Second, the dynamic emission tolls for all links of the network can be obtained by solving an optimal control problem with the DUE problem integrated as constraints and the (dynamic) inflows to each link determined by the DSO problem. It turns out that under the above-discussed monotonicity conditions, using either the CO emission function in (9), or the fuel consumption function as a proxy of multiple emissions in (12), an optimal solution (if exists) is always a free-flow optimal solution for the first sub-problem. This free-flow optimal solution is then used as the input to the pricing model to derive the first-best emission pricing scheme in the second sub-problem. Such an emission pricing framework internalizes the externalities, which includes the emission cost as part of the generalized cost of the drivers. In this way the drivers will consider both the travel time and emission costs while making choice decisions. A flow chart describing the proposed emission pricing framework is shown in Figure 2.

4.1. Sub-problem 1: dynamic system optimum

In this section, we first present a DSO formulation for the first sub-problem, where the objective function takes both emission and time costs into consideration. Then we analyze the model with different objective functions. We first prove in Theorem 2 that for the problem that only minimizes the total emissions, the optimal solution is a free-flow solution under certain conditions. If both costs are considered in the objective function, the same conclusion still hold, i.e., an optimal solution (if exists) is a free-flow solution under the same conditions.

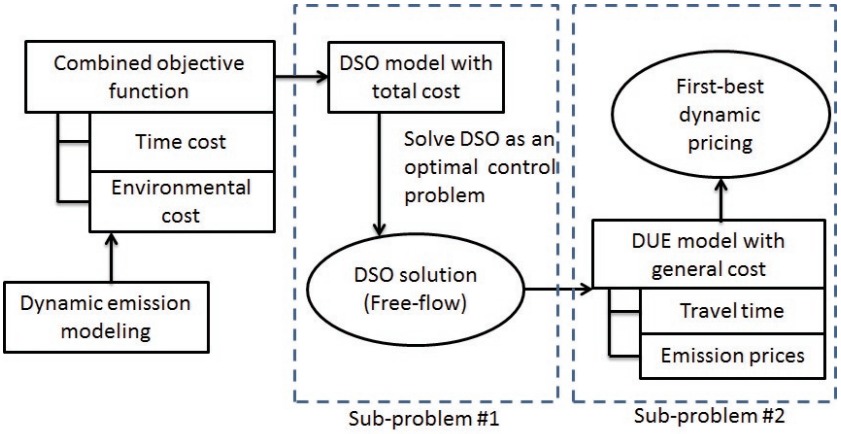


Figure 2: Emission pricing framework

4.1.1. DSO formulations

Recently, a continuous-time DSO framework considering spillback phenomena was proposed in Ma et al. (2014), where the objective function is to minimize the total time cost. As shown in Table A.2, not all existing emission pricing models can consider queue spillbacks. By applying the double-queue model, the model proposed here could capture spillbacks, although we focus on free-flow solutions when designing optimal pricing schemes. The original DSO formulation in Ma et al. (2014) is summarized as in (13), where 'TT' stands for 'total time'.

$$\min_{(p,v,q^d,n)\,\in\,\Omega} \mathrm{TT} \triangleq \sum_{i:(i,\widehat{s})\in L} \int_0^T t\, v_{i\widehat{s}}(t)\, dt. \tag{13}$$

Here Ω is the feasible set of the problem as defined below in equations (15) - (22); p, v, q^d, n are the vectors of inflows, exit flows, downstream queues, and the numbers of vehicles of all links in the network. The DSO formulation for emission control in this paper extends the DSO model in Ma et al. (2014). That is, we aim to minimize the objective function (14) that integrates the time and environmental costs (denoted as 'TC'' in this paper), following the similar idea for the static case in Yin and Lawphongpanich (2006).

$$\min_{(p,v,q^d,n)\,\in\,\Omega} \mathrm{TC}' \triangleq \gamma_e \mathrm{TE} + \gamma_t \mathrm{TT} = \gamma_e \sum_{(i,j)\in L} E_{ij}|_0^T + \gamma_t \sum_{i:(i,\widehat{s})\in L} \int_0^T t\, v_{i\widehat{s}}(t)\, dt. \tag{14}$$

Here γ_e is the value of emissions, which converts the emission (or fuel consumption) into monetary measures; γ_t is the value of time, which is assumed to be constant for all the drivers in the network in this paper. As explained, one can easily replace TE by TF in (12) for an alternative objective function with total time and fuel consumption costs, which is omitted in this paper.

Notice that the total emission function defined in (9) or (12) brings nonlinearity. By including the same constraints from the DSO model in Ma et al. (2014), as shown below, a nonlinear optimal control problem can be formulated. These constraints define the feasible set Ω, i.e., tuples (p, v, q^d, n), of the problem:

1. Flow conservation at all intermediate nodes: for node $i \in N\backslash\{\widehat{s}\}$ and $t \in [0, T]$,

$$\sum_{j:(i,j)\in L} p_{ij}(t) \;=\; \sum_{\ell:(\ell,i)\in\mathcal{L}} v_{\ell i}(t); \tag{15}$$

2. Bounds of the double-queues on regular links: For link $(i, j) \in L$ and $t \in [0, T]$

$$q_{ij}^u(t) \;\le\; \overline{Q}_{ij}; q_{ij}^d(t) \;\ge\; 0; \tag{16}$$

3. Bounds of exit flow rates from all links: For link $(i,j) \in \mathcal{L}$, $t \in [0,T]$,

$$0 \le v_{ij}(t) \le \overline{C}^v_{ij}, \tag{17}$$

with $\overline{C}^v_{\widetilde{o}o} = \infty$ for $(\widetilde{o},o) \in L_d$.

Bounds of inflow rates to regular links: For link $(i,j) \in L$, $t \in [0,T]$,

$$0 \le p_{ij}(t) \le \overline{C}^p_{ij},\ t \in [0, T-\tau^0_{ij}], \tag{18}$$

with $p_{ij}(t) = 0$, $t \in (T-\tau^0_{ij}, T]$;

4. Double-queue dynamics on regular links: For link $(i,j) \in L$ and $t \in [0,T]$

$$\begin{array}{l} \dot{q}^u_{ij}(t) = p_{ij}(t) - v_{ij}(t-\tau^\omega_{ij}) \\ \dot{q}^d_{ij}(t) = p_{ij}(t-\tau^0_{ij}) - v_{ij}(t). \\ n(t) \ = \int_0^t p_{ij}(\xi) - v_{ij}(\xi) d\xi. \end{array} \tag{19}$$

Demand discharging queue dynamics on dummy origin links: For link $(\widetilde{o},o) \in L_d$ and $t \in [0,T]$

$$\begin{array}{l} \dot{q}^d_{\widetilde{o}o}(t) = -v_{\widetilde{o}o}(t). \\ q^d_{\widetilde{o}o}(t) = D_o - \int_0^t v_{\widetilde{o}o}(\xi) d\xi. \end{array} \tag{20}$$

5. Initial conditions:

For $(\widetilde{o},o) \in L_d$,

$$q^d_{\widetilde{o}o}(0) = D_o; \tag{21}$$

For $(i,j) \in L$,

$$q^u_{ij}(0) = q^d_{ij}(0) = 0; \tag{22}$$

6. Other boundary conditions:

For $(i,j) \in L$ and $t \in (-\infty, 0)$, $p_{ij}(t) = v_{ij}(t) = 0$;

For link $(i,j) \in L$, $q^u_{ij}(T) = q^d_{ij}(T) = 0$;

For link $(\widetilde{o},o) \in L^d$, $q^d_{\widetilde{o}o}(T) = 0$.

Equation (19) above describes the traffic dynamics via the double queue model, while equations (15) - (18) defines the simple junction model applied in this paper, which describes the flow conservation and nonnegativity/boundedness constraints at junctions. The double queue model at the link level, along with the simple junction model, connects the inflow rates and exit flow rates of adjacent links, which also extend the double queue model at a link level to a network level. It is obvious that the inflow and exit flow rates of a regular link $(i,j) \in L$ are bounded.

Equation (20) defines the queue dynamics of the dummy origin links. It can be seen (also from the discussions in Section 2.2) that such a queuing model for dummy origin links is different from the point-queue model, since the exit flow is constrained by the downstream links' inflow capacities. It is different from the double-queue model as well, since there is only one queue variable instead of two, with no inflow to the link.

The boundedness of the exit flow from the dummy origin links can also be shown. According to the flow conservation at the origin node o, for the dummy link $(\widetilde{o},o) \in L_d$, the exit flow rate $v_{\widetilde{o}o} = \sum_{\ell:(\ell,o),\, \ell \neq \widetilde{o} \in \mathcal{L}} v_{\ell o}(t) - \sum_{j:(o,j)\in L} p_{oj}(t) \le S \cdot C_{max;o}$, where $C_{max;o} \triangleq \max_{j:(o,j)\in L} \overline{C}^p_{oj}$, and S is the number of outgoing links from node o, i.e., $S = card\{j : (o,j) \in L\}$. As long as S is finite (which is the case for real-world networks), the exit flow rates of a dummy link is also bounded. (The capacity of each dummy origin link does not matter as it must be larger than the corresponding exit flow rate.) As defined by the queue dynamics of dummy origin links in (20), there is no inflow to a dummy origin link, and the demand on each dummy origin link is presented as a non-increasing queue, whose discharging rate is the demand rate from the corresponding origin node.

Overall, since the exit flow rates from dummy origin links, the inflow rate to and the exit flow from each regular link are all bounded, it is ensured that all the flow rates are bounded in the solution set.

The simple junction model (15) - (18) merits further discussions. First, a junction model, sometimes called a Riemann Solver (Han et al., 2016b), is often needed to determine the entering and exiting flows of links adjacent to the same junction. Demand and supply are normally needed for these junction models. However, different from these models, in this paper, we do not need explicit definitions or calculation of demand and supply. This is because the double-queue dynamics can reflect the demand at its downstream queue (q^d), and the supply at its upstream queue (q^u). If the downstream queue reduces to zero, then the exit flow (as the demand of sending flow) is no more than the amount of inflow of such link (with a time delay of free-flow travel time τ^0). If the upstream queue reaches to the queue capacity, then the inflow (as the supply for receiving flow) is no more than the amount of exit flow of the link (again with a time delay of shockwave travel time τ^w). These constraints are guaranteed by the double-queue dynamics in (19), and the boundedness of double queue in (16). Notice that the double queue model applied here is a continuous-time reformulation of the Link Transmission Model (LTM). Detailed discussion on the sending and receiving flows in the LTM can be found in Yperman (2007) [Chapter 4.6]. There are also other equivalent representations of the continuous-time LTM, such as the link-based kinematic wave model in Han et al. (2016b) and Jin (2015). In particular, the difference of the queue storage capacity and the upstream queue in the double-queue model is defined as 'vacancy size' in Jin (2015), which can be used to indicate whether spillbacks occur at the entrance of a link.

Secondly, existing junction models usually need to pre-define or derive some merging / diverging rules at a junction (such as turning ratios or maximizing flows or others). In this paper, we intend not to specify any pre-defined rules for that. Instead, we aim to determine the optimal inflow rates at every junction (i.e., at the entrance of every link) by the DSO principle in order to optimize the specific objective function. Thus the turning ratio is not pre-defined in the formulation, while a set of optimized turning ratios at junctions can be calculated after an optimal solution is obtained. In this sense, one may consider that the turning ratio is calculated by a more complex rule, as implicitly determined by the DSO model itself. These optimized turning ratios can be treated as a target or benchmark to control / manage traffic to achieve the desired DSO objective. How to design control schemes at junctions to achieve the desired turning ratios or objectives is, of course, not a trivial task, which however is not the focus of the paper.

Here we also need to point out that if one intends to model DUE or to incorporate existing traffic control devices (such as traffic signals) into the DSO model, then pre-defined merging/diverging rules (such as pre-defined ratios, maximizing flow, etc.) will have to be integrated into the DSO model. In this case, the junction model will be more complicated as discussed in Han et al. (2016b). The authors have developed such junction models for the double queue link model for DUE problems; details can be found in Ma et al. (2015a).

4.1.2. Discussion on the solution set

As shown in Ma et al. (2014), a DSO problem is feasible when a proper finite terminal time is selected. A DSO problem may have multiple *feasible* solutions. One of such feasible solutions is a free-flow feasible solution that does not have any downstream queue on any of the regular network links. We define the set of free-flow feasible solutions as $\Omega^f \subseteq \Omega$. For any free-flow feasible solution $X^f = (p^f, v^f, q^{d;f}, n^f) \in \Omega^f$, the downstream queues of all regular links are zero at any time instant, i.e.,

$$\forall X^f \in \Omega^f,\ q_{ij}^{d;f}(t) = 0, \forall (i,j) \in L, \forall t \in [0,T]$$

Notice that a free-flow feasible solution is not necessarily an *optimal* solution.

Problem (14) combines the total emission and total time costs as the objective. To study the properties of the solution set, we formulate two other optimal control problems based on these two costs individually.

First, it is clear that the original DSO formulation (13) is an optimal control problem that only minimizes the total time cost. The set of optimal solutions to the original DSO is denoted as $\Omega^{TT} \triangleq \text{argmin}_{(p,v,q^d,n)\in\Omega}\ \text{TT} \subseteq \Omega$. As shown in Ma et al. (2014), Ω^{TT} could contain both free-flow solutions and non-free-flow solutions.

We then construct an optimal control problem that only minimizes the total emissions. The two problems share the same feasible solution set with problem (14), i.e., the set Ω.

$$\min_{(p,v,q^d,n)\,\in\,\Omega} \text{TE} \triangleq \sum_{(i,j)\in L} E_{ij}|_0^T \tag{23}$$

We define the set of optimal solution to formulation (23) as $\Omega^{TE} \triangleq \text{argmin}_{(p,v,q^d,n)\,\in\,\Omega}\ \text{TE} \subseteq \Omega$.

Similar to problem (14), the objective function of (23) is also nonlinear. Generally, it is not trivial to solve nonlinear optimal control problems, or their discrete forms as nonlinear programs. However, we prove in Theorem 2 that a free-flow optimal solution to (23) exists if certain monotonicity conditions are satisfied. This property will be used to solve the nonlinear optimal control problems. We start with Proposition 1.

Proposition 1. *For any non-free-flow feasible solution $X^g \in \Omega$ of problem (23), if $e_{ij}(\cdot)$ is a monotonically increasing positive function for any link $(i,j) \in L$, then a free-flow feasible solution $X^{g;f}$ can be constructed from X^g. That is, if $0 < q_1 < q_2 \Rightarrow 0 < e_{ij}(q_1) < e_{ij}(q_2)$, then $\forall X^g = (p^g, v^g, q^{d;g}, n^g) \in \Omega$, $\exists X^{g;f} = (p^{g;f}, v^{g;f}, q^{d;g;f}, n^{g;f}) \in \Omega^f$. Furthermore, the objective value of (23) under $X^{g;f}$ is smaller than that under X^g. That is, $TE(X^{g;f}) < TE(X^g)$.*

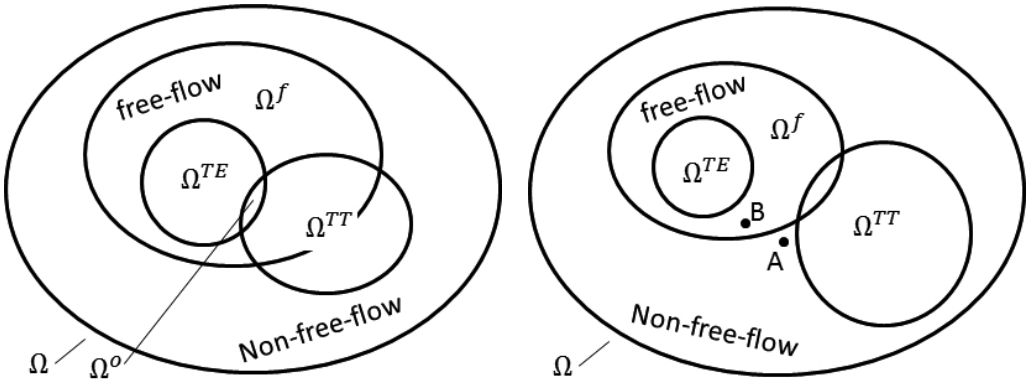
Detailed proof of Proposition 1 can be found in Appendix D, which is based on the construction of a free-flow feasible solution $X^{g;f}$ from a feasible solution X^g. If X^g itself is a free-flow feasible solution, then the proof is trivial since $X^{g;f} = X^g$; Otherwise, $X^{g;f}$ is constructed so that for every link, the total cumulative exit flow, which is equivalent to the total number of vehicles passing through the link over the entire time span, remains the same as in the solution X^g. The constructed solution $X^{g;f}$ reassigns the flow temporally, so that for all regular links, the link dynamics are free-flowing, and the downstream queues are shifted to the dummy origin links.

For problem (23), it is quite different whether vehicles waiting at the regular links or at the dummy origin links. Vehicles waiting at regular links, i.e., downstream queues on regular links, generate emissions (or consume fuel). On the other hand, vehicles waiting at dummy origin links may not generate any emissions since it is reasonable to assume that they do not start their trips and idle at the origins (e.g., at homes). In the constructed free-flow feasible solution $X^{g;f}$, all regular links have zero downstream queues, so that the objective function of problem (23) is reduced, compared to X^g. This is very different from the original DSO problem (13) that only minimizes the time cost. For the different optimal solutions to the original DSO problem (13) that only concerns about travel time, vehicles may wait at the origins or on some intermediate link in the network. In this case, waiting the same time at different locations of the network means the same in terms of objective value. This also explains why the DSO problem may have multiple solutions: even the total system travel times and waiting times (the objective function value of problem (13)) are exactly the same and at the minimum (i.e., the optimal state is obtained), vehicles can still wait at different locations in the network, leading to different optimal DSO solutions. It implies that, while the DSO problem with the time cost objective (13) could have non-free-flow optimal solutions, an optimal solution to (23), if exists, must be a free-flow solution if $e_{ij}(\cdot)$ is a monotonically increasing. This is stated in the following theorem.

Theorem 2. *If $e_{ij}(q_{ij}^d)$ is a monotonically increasing positive function for any link $(i,j) \in L$, an optimal solution to problem (23), if exists, must be a free-flow solution, i.e., $\Omega^{TE} \subset \Omega^f$.*

Proof. Proposition 1 shows that for any non-free-flow feasible solution $X^g \in \Omega \setminus \Omega^f$, we can construct a free-flow feasible solution $X^{g;f} \in \Omega^f$ with a smaller objective value since $e_{ij}(q_{ij}^d)$ is a monotonically increasing positive function for any link $(i,j) \in L$. That is, we have $\text{TE}(X^{g;f}) < \text{TE}(X^g)$, which means a non-free-flow feasible solution is not an optimal solution, i.e., $\left(\Omega \setminus \Omega^f\right) \cap \Omega^{TE} = \emptyset$. Since $\Omega^{TE} \subset \Omega$, we then have $\Omega^{TE} \subset \Omega^f$. □

Theorem 2 shows that if an optimal solution exists for problem (23) that only minimizes total emissions, an optimal solution has to be free-flowing, if the emission function is positive and monotonically increasing. As pointed out in Ma et al. (2014), free-flow optimal solutions indicate that there is no congestion in the network and the drivers wait, if needed, at their origins before departure.



(a) $\Omega^{TE} \cap \Omega^{TT} \neq \emptyset$ (b) $\Omega^{TE} \cap \Omega^{TT} = \emptyset$

Figure 3: Solution sets of formulations (14), (23) and (13)

Figures 2(a) and 2(b) illustrate the relationships of different solution sets of DSO under two scenarios. Here we assume the emission function is a monotonically increasing positive function, so that according to Theorem 2, the set Ω^{TE} is a subset of Ω^f. If the optimal solution sets of formulations (23) and (13) overlap, i.e. $\Omega^o \triangleq \Omega^{TE} \cap \Omega^{TT} \neq \emptyset$, as shown in Figure 3(a), the optimal solution set of formulation (14) is Ω^o. Since $\forall X_o \in \Omega^o$ and $\forall X \notin \Omega^o$, $\mathrm{TE}(X) \geq \mathrm{TE}(X_o)$ and $\mathrm{TT}(X) \geq \mathrm{TT}(X_o)$, we have $\mathrm{TC}'(X) \geq \mathrm{TC}'(X_o)$. We notice that $\Omega^{TE} \subset \Omega^f$, so that $\Omega^o \subset \Omega^f$, i.e., the optimal solution to (14) is a free-flow solution in this case.

If $\Omega^{TE} \cap \Omega^{TT} = \emptyset$, as shown in Figure 3(b), TE and TT are two objective functions that conflict with each other. In other words, there could be no way to minimize one objective without increasing the other objective, which leads to a Pareto optimization. In this case, there might exist a non-free-flow optimal solution to formulation (14). Either a non-free-flow solution (Point A) or a free-flow solution (Point B) could be an optimal solution to formulation (14).

In Theorem 4 below, however, we prove that an optimal solution to problem (14) must be a free-flow solution under the same monotonicity condition as in Theorem 2. That is, feasible solutions like Point A in Figure 3(b) are not optimal for problem (14) under these conditions. Before that, we first show in Proposition 3 that for any non-free-flow feasible solution X, a free-flow feasible solution X^f can be constructed, so that the total emissions of X^f is smaller than the total emissions of X, and the total travel time of X^f is equal to the total travel time of X.

Proposition 3. *For a non-free-flow feasible solution $X \in \Omega$ of problem* (14)*, there exist a free-flow feasible solution $X^f \in \Omega^f$, so that $\mathrm{TT}(X) = \mathrm{TT}(X^f)$; If $e(\cdot)$ is a monotonically increasing positive function, $\mathrm{TE}(X) > \mathrm{TE}(X^f)$.*

Detailed proof of Proposition 3 can be found in Appendix E.

Proposition 3 implies that an optimal solution to the emission DSO problem (14) proposed in this paper must be a free-flow solution, as stated in the following theorem.

Theorem 4. *An optimal solution to the emission DSO problem* (14) *is a free-flow optimal solution, if $e(\cdot)$ is a monotonically increasing positive function and the optimal solution set of* (14) *is non-empty.*

Proof. We prove this by contradiction. Since the optimal solution set is non-empty, i.e., $\Omega^o \neq \emptyset$, assume there exists a non-free-flow optimal solution $X^{NF} \in \Omega^o \setminus \Omega^f$. From Proposition 3, we can construct a free-flow solution $X^F \in \Omega^f$, where $\mathrm{TE}(X^{NF}) > \mathrm{TE}(X^F)$ and $\mathrm{TT}(X^{NF}) = \mathrm{TT}(X^F)$. Then $\mathrm{TC}'(X^{NF}) > \mathrm{TC}'(X^F)$. This contradicts with the fact that X^{NF} is an optimal solution. □

Remarks Theorems 2 and 4 assume that there exists an optimal solution to the emission DSO problem (14). The existence of an optimal solution to the discretized DSO problem is relatively easier to show

by applying the Weierstrass' extreme value theorem, since the feasible set is compact and the objective function is continuous with respect to the state variables. However, the existence of an optimal solution to the continuous-time DSO problem (14) is not that obvious, since the variables for the optimal control problem is infinite dimensional. In Burger (2003) (Theorem 2.3), it was noted that in order to establish the solution existence of a general optimization problem in infinite dimensions, two basic properties are needed: compactness and lower semicontinuity. Eberlein-Smulian theorem shows that compactness is caused by boundedness for Hilbert spaces. However, it is quite challenging and not straightforward to provide an exact definition of a Hilbert space of the tuple $\{p, v\}$ of all links with multiple time-delays as well as the two point boundaries. Thus it is not trivial to rigorously show its solution existence, which is left for future research.

4.2. Sub-problem 2: first-best emission pricing

To determine the optimal emission pricing, we have to consider drivers' choice behaviors, which are modeled as DUE in this paper. The link-node based DUE model presented in Ma et al. (2015b) is used here for the single-destination DUE problem. In general, the flow pattern in a DUE solution is not necessarily a DSO solution. This is where the emission pricing can play a critical role. That is, by imposing the emission tolls, drivers will have to consider this extra cost (in addition to their travel times). This will change their choices and as a result, the flow pattern of the network. By properly selecting the pricing scheme, the DSO flow pattern can be achieved, which is the first-best emission pricing scheme in this paper.

By solving the emission DSO problem (14) in the previous subsection, a free-flow optimal solution can be obtained. This will produce the inflow and exit flow of all links. The demand rate of an origin can also be calculated from the inflow and exit flow rates. The travel time at regular link (i, j) is reduced to the free-flow travel time under a free-flow optimal solution, i.e., $\tau_{ij}(t) = \tau^0_{ij}$.

The Wardrop's route choice in the dynamic user equilibrium, was introduced in Ma et al. (2015b) in the complementarity form:

$$0 \leq p_{ij}(t) \perp \tau^0_{ij} + \pi_j(t + \tau^0_{ij}) - \pi_i(t) \geq 0. \tag{24}$$

In (24), no monetary cost is included. Therefore, $\pi_i(t)$, the minimum cost from node i to the destination at time t, only represents the time cost.

In the Sub-problem 2, we apply the following modifications to the complementarity form (24).

1) For link (i, j), we add the dynamic toll at time t, denoted as $y_{ij}(t)$, into the link cost. The toll is converted into time cost by the value-of-time constant γ_t;

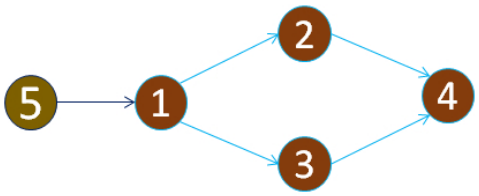
2) The minimum time cost from node i to the destination at time t, $\pi_i(t)$, is replaced by $\tilde{\pi}_i(t)$, which is the minimum travel cost (including travel time and the tolls) from node i to the single destination.

3) With the inflow rates $p_{ij}(t)$ known as the results of Sub-problem 1, the complementarity form of the Wardrop's route choice is reduced to a set of linear constraints. The reduced constraints are either equations (for the case $p_{ij}(t) > 0$) or inequalities (for the case $p_{ij}(t) = 0$).

By incorporating these modifications, an optimal control problem can formulated to calculate the first-best tolling as follows.

$$\begin{aligned} & \min_{y,\tilde{\pi}} \int_0^T \sum_{(i,j)\in L} y_{ij}(t)dt \\ s.t. \quad & y_{ij}(t) \geq 0, \\ & \tilde{\pi}_i(t) \geq \pi^0_i \\ & \text{If } p_{ij}(t) > 0,\ \tau^0_{ij} + y_{ij}(t)/\gamma_t + \tilde{\pi}_j(t + \tau^0_{ij}) - \tilde{\pi}_i(t) = 0; \\ & \text{If } p_{ij}(t) = 0,\ \tau^0_{ij} + y_{ij}(t)/\gamma_t + \tilde{\pi}_j(t + \tau^0_{ij}) - \tilde{\pi}_i(t) \geq 0. \end{aligned} \tag{25}$$

The minimization problem finds the minimum total tolls to satisfy all the constraints, including the non-negativity of the tolls. Notice that since $p_{ij}(t)$ is known from the first sub-problem, all the above constraints are linear. Moreover, the resulting model is a continuous-time optimal control problem with linear constraints and constant time delays. The continuous-time problem can be solved by discretization and then solving the resulting discrete-time linear program. Details are omitted here. We note that the



(a) Network topology

link	τ^0_{ij} (min)	τ^ω_{ij} (min)	$\overline{C_{ij}}$ (veh/min)	$\overline{Q_{ij}}$ (veh)	L_{ij} (mile)	ρ_{ij}
5-1	0	0	108	400	-	0
1-2	1	2	40	120	0.5	1
2-4	2	4	16	96	1	1
1-3	1	2	40	120	0.5	1
3-4	3	6	12	108	1.5	1

(b) Network parameters

Figure 4: Two-route network

optimal toll (denoted as y^*) by solving (25), together with the known free-flow optimal solution to the emission DSO problem (14), guarantees to be an optimal solution to the DUE problem under toll y^*. This would be fine if DUE has a unique solution. However, if DUE has multiple solutions (which is very likely), implementing the obtained optimal toll y^* may or may not produce the same free-flow optimal solution used as an input to (25). This is similar to the bi-level network design problem under multiple UE solutions; see Ban et al. (2013, 2009). More in-depth discussions about this is beyond the scope of this paper and may be pursued in future research.

5. Numerical Results

We test the proposed dynamic emission pricing model using several testing networks. For the simple network with only one origin, we use aggregated fuel consumption as a proxy of overall emissions, and Pareto solutions are shown for various values of fuel consumption. For the network with six nodes, the CO emissions are used. The model is further tested on the well-known Sioux Falls network using CO emissions.

5.1. The simple network

First we present the numerical results of a simple network, as shown in Figure 4(a). This network contains two routes between one origin-destination pair. The free-flow travel time and exit flow capacity of each link is also shown in Figure 4(b). The emission (i.e., fuel in this case) impact coefficients of all regular links are set the same as 1. The total demand from the origin node 1 is 400 (node 5 is the dummy origin node). In this section, we set the upper bounds of the inflow and exit flow as the same for the same link, and denote it as $\overline{C}_{ij}$, i.e., $\overline{C}^p_{ij} = \overline{C}^v_{ij} = \overline{C}_{ij}$. All links have a free-flow travel speed as 30 mph, i.e., 0.5 mile per minute.

For this particular network, it is observed that the optimal solution to the DSO problem (14) is free-flowing, and may not minimize the total time or total fuel consumption simultaneously, which is the Case B as shown in Figure 3(b). From Figure 5, we can find that for different ratios of the value of fuel consumption versus value of time (i.e., γ_e/γ_t), it leads to different solutions to the DSO problem (14) on the efficient frontier of a Pareto efficiency curve. When γ_e/γ_t is equal or less than 0.002, the fuel consumption reaches its highest value as 5.698×10^5 grams; At the same time, the total time spent by all the drivers reaches its minimum value as 4828 minutes. If such ratio is greater than 0.002, the total travel time is not at its minimum, while the total fuel consumptions can be further reduced, until the ratio is large enough (0.055 in this case). As the ratio increases, the total fuel consumption reduces and the total time increases. With different ratios, the optimal solutions form a Pareto solution set. Similar Pareto (or non-dominated) solution sets were also found in Yin and Lawphongpanich (2006) and Miandoabchi et al. (2015). As proved by Theorem 4, the optimal solution (if exists) should be free-flowing, regardless the setting of values of

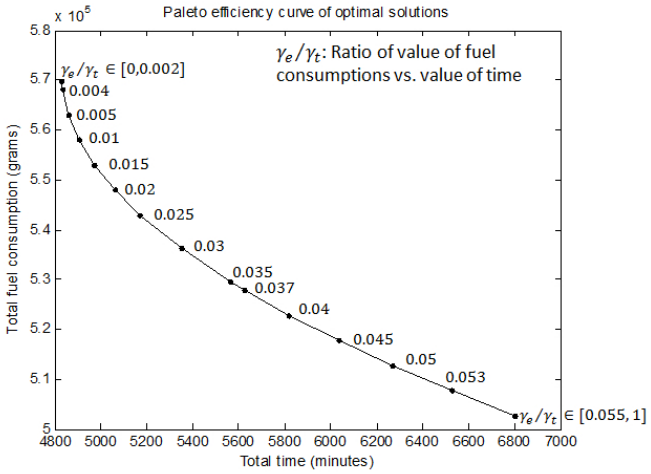


Figure 5: Pareto solutions with different ratio of values of fuel consumption vs. value of time

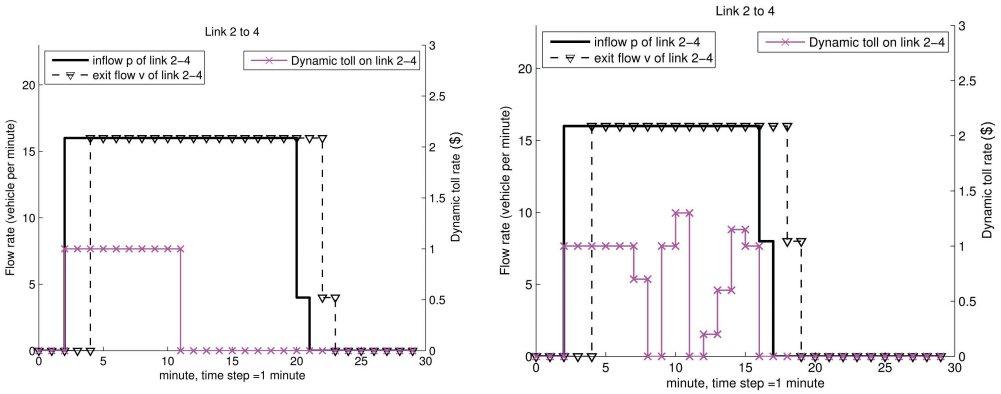
time and fuel consumption, as long as the emission/fuel consumption function used for this network has the monotone property.

We can also observe that for different ratios γ_e/γ_t, the optimal first-best pricing solutions can be very different. Figure 6(b) shows that by setting $\gamma_e/\gamma_t = 0$, i.e., only minimizing the total travel times with no consideration of emissions (fuel consumption), the inflow to link (2,4) reduces to zero at time 17 minutes, which is different from the case $\gamma_e/\gamma_t = 0.02$ (for which both travel time and emissions are considered), where the inflow rate keeps positive until at the time 21 minutes, as shown in Figure 6(a). Furthermore, the resulting dynamic pricing schemes for the system to achieving the optimum are quite different for different γ_e/γ_t ratios. For the case that $\gamma_e/\gamma_t = 0.02$, we can calculate the first-best pricing schemes (tolls). As shown in Figure 6(a), the tolls are only imposed on link $(2,4)$ during time 2 to 11 minutes. On other links there is no toll. The toll happens to be constant for this period of time. For the case that $\gamma_e/\gamma_t = 0$, the calculated tolls vary more dramatically with time, showing a more "dynamic" nature, as shown in Figure 6(b). These figures clearly show that with (e.g., Figure 6(a) with $\gamma_e/\gamma_t = 0.02$) or without (e.g., Figure 6(b) with $\gamma_e/\gamma_t = 0$) considering emissions/fuel consumption in the objective when designing network wide emission pricing schemes, the resulting network flow patterns and the optimal pricing schemes can be very different. This also implies the significance of incorporating emission/fuel consumption objectives in the pricing and control of dynamic traffic networks.

Without the toll, links $(1,2)$ and $(2,4)$ would be the shortest route for drivers departing at node 1 to the destination. As a result, there would be more vehicles traveling on links $(1,2)$ and $(2,4)$, so that the link would be no longer free-flowing and vehicles on it would consume more fuel, thus increasing the total system costs. The toll on link $(2,4)$ increases the travel cost for a driver traveling through link $(2,4)$, but decreases the total system cost, by switching part of the flow to the other alternative, i.e., links $(1,3)$ and $(3,4)$. As shown in Figure 7, the route choice of the vehicles departing at node 1 follow Wardrop's principle regarding the generalized cost (i.e., travel time plus toll costs) for the drivers. During time 2 to 10 minutes, the two routes share the same minimum travel cost, and the inflow splits onto both routes. During other times, traveling via link $(1,2)$ takes smaller cost for the drivers, and all inflow is assigned to link $(1,2)$.

5.2. The network with six nodes

The second testing network has six nodes and was also used in Yin and Lawphongpanich (2006). Two dummy origins (node 8 and node 9) and corresponding dummy links are added before the origin nodes, and a single destination node (node 7) is also added. The network and its parameters are shown in Figure 8 and Table 1. Another important modification of the network is that the emission impact coefficient of link $(1,3)$



(a) Flows and pricing on link $(2,4)$, $\gamma_e/\gamma_t = 0.02$ (b) Flows and pricing on link $(2,4)$, $\gamma_e/\gamma_t = 0$

Figure 6: Flows and pricing on link $(2,4)$ with different γ_e/γ_t ratios ($\gamma_t = 1$)

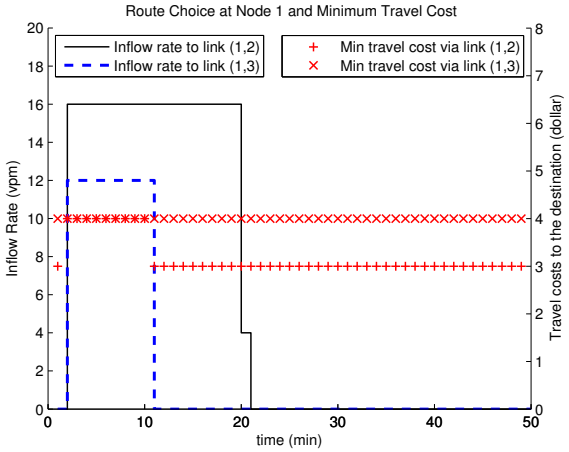


Figure 7: Route choice and minimum travel cost (time and tolls)

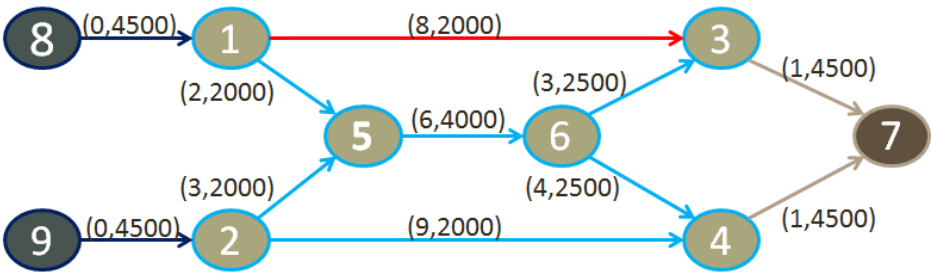
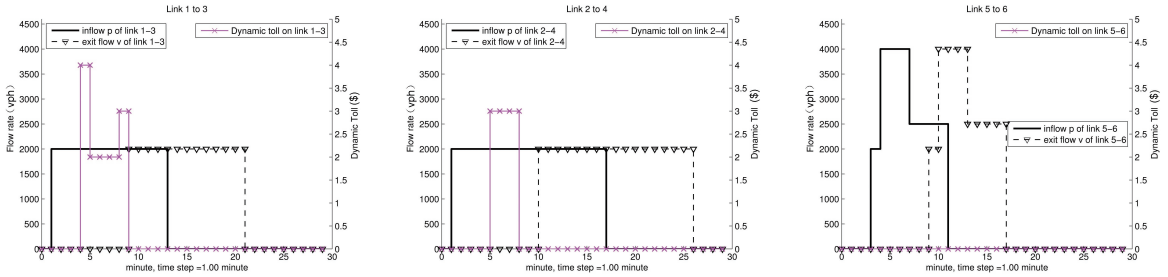


Figure 8: Six-node network

is set to be higher than the coefficients of other links. In order to minimize the total cost, the dynamic toll is expected to be high on link $(1,3)$ in order to shift part of the flow on this link to other alternatives. The objective function is a combination of total CO emissions and total travel time as shown in (14). The value of emission $\gamma_e = 200$, the value of time $\gamma_t = 1$.

Table 1: Six-node network parameters

link	8-1	9-2	1-3	1-5	2-4	2-5	3-7	4-7	5-6	6-3	6-4
τ_{ij}^0 (min)	0	0	8	2	9	3	1	1	6	3	4
τ_{ij}^ω (min)	0	0	16	4	18	6	2	2	12	6	8
$\overline{C}_{ij}$ (vph)	4500	4500	2000	2000	2000	2000	4500	4500	4000	2500	2500
$\overline{Q}_{ij}$ (veh)	10000	10000	800	200	900	300	225	225	1200	375	500
ρ_{ij}	0	0	1.3	1	1	1	1	1	1	1	1



(a) Link $(1, 3)$ (b) Link $(2, 4)$ (c) Link $(5, 6)$

Figure 9: Dynamic link flows and tolls for selected links in six-node network

The numerical results in Figure 9 support our expectation. Figure 9(a) shows that the dynamic tolls on link $(1, 3)$ is positive during time 4 to 9 minutes, and reaches 4 dollars as its maximum at time 4 minutes. In Figure 9(b), there are also tolls imposed on link $(2, 4)$ during time 5 to 8 minutes, which have a shorter duration and the peak value is smaller than those on link $(1, 3)$. The inflow to link $(5, 6)$ reaches 4000 vph during time 4 to 7 minutes as shown in Figure 9(c), and there is no toll for link $(3, 6)$. In fact, the traffic flow that would have entered links $(1, 3)$ and $(2, 4)$ is pushed to link $(5, 6)$ by the dynamic pricing mechanism. With the dynamic tolls, from the network perspective, all the regular links maintain in their free-flow state, and the total cost (time and emissions) is minimized. From the drivers' perspective, at each node, the route choice decision at each time instant is consistent with Wardrop's first principle since the links with the minimum generalized cost (travel time and tolls) are selected. The observation here is similar to that in Figure 7 for the simple network.

5.3. Sioux Falls network

We further test the emission pricing framework on the Sioux Falls network as shown in Figure 10, which is much more complicated than the previous testing networks. Sioux Falls network is abstracted from the City of Sioux Falls, and firstly used by LeBlanc et al. (1975). The network configurations and associated free-flow travel time τ_{ij}^0 and flow capacity can be found in Suwansirikul et al. (1987). The shockwave travel time τ_{ij}^ω is twice of the free-flow travel time for each regular link. Since only two OD pairs are selected (see below), the exit flow capacities $\overline{C}_{ij}$ of all regular links are rescaled as 1/6000 of the original capacities. The queue storage capacity can be calculated as $\overline{Q}_{ij} = (\tau_{ij}^0 + \tau_{ij}^\omega)\overline{C}$ as shown in Ma et al. (2014). The parameters of the dummy links are set similarly as in the six-node network. Traditionally, Sioux Falls network serves as a testbed in static modelling. Recent developed dynamic models also used Sioux Falls network as a testing network (e.g., in Han et al. (2015) and Ma et al. (2015a)). As shown in Figure 10, this network has 24 nodes and 76 links. The entire Sioux Fall network allows up to 528 OD pairs. However, in this numerical test, we only set two OD pairs, namely Node 1 to Node 15 and Node 16 to Node 15, respectively. The demand for each OD pair is 30 vehicles.

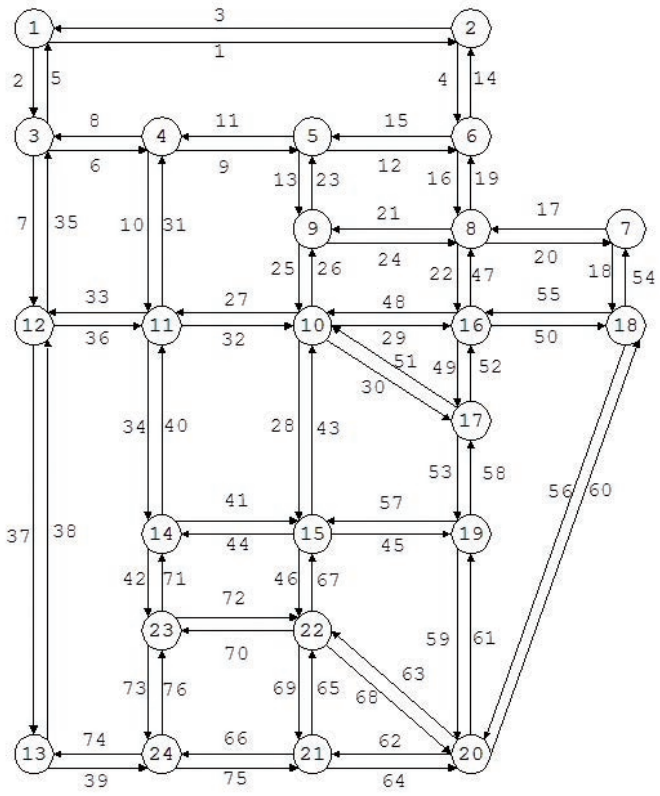
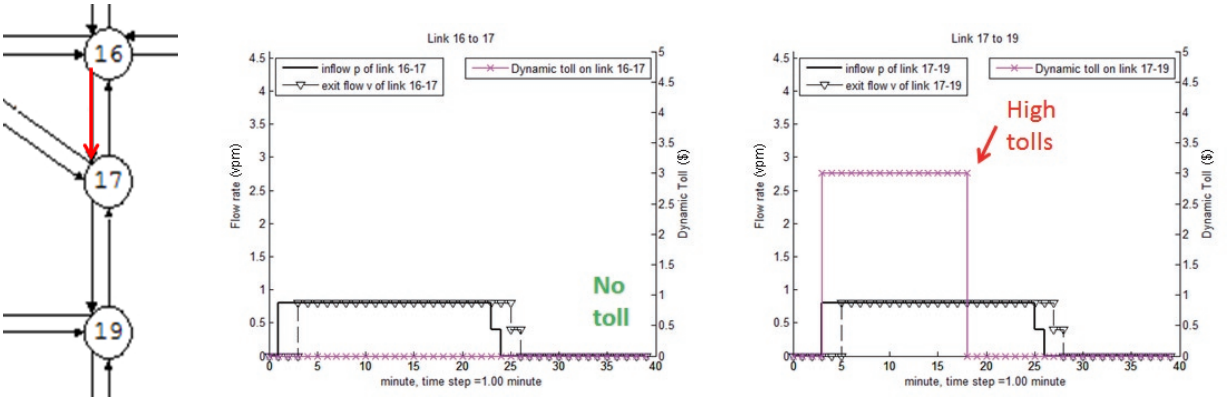


Figure 10: Sioux Falls network

The objective is to minimize the weighted summation of total travel time and total CO emissions (i.e., TE) for this network. In particular, the emission coefficient of link $(16, 17)$ is set higher than the coefficients of other links. Since the Sioux Falls network has been studied very extensively in the literature, we omit the other details of the network, such as its geometry, demands, and link parameters. Through the numerical tests, it is observed that the dynamic tolls are not necessary on the links with the high emission coefficients, which seems to be particularly true for complicated networks. For example, a small portion of the Sioux Falls network is shown in Figure 11(a). Link $(16, 17)$ has a higher emission coefficient $\rho_{16,17}$. In order to decrease the flows on this link, one would assume that dynamic tolls on this link should be high. However, Figure 11(b) shows no toll on link $(16, 17)$. In fact, Figure 11(c) shows that high tolls are observed on link $(17, 19)$, which is adjacent to link $(16, 17)$. The results suggest that the dynamic tolls may be imposed on other links (e.g., adjacent links) as long as the total costs are minimized by the tolling mechanism.

6. Conclusions, Discussions, and Future Research Directions

This paper proposed a dynamic emission pricing framework for traffic networks with a single destination. It includes the development of two sub-problems, and the analysis of the solution sets and the solution techniques. The first sub-problem was a system optimum dynamic traffic assignment, or DSO problem, where the generalized cost, including travel time and environmental costs, was minimized. It was proved that an optimal solution to the DSO problem, if exists, must be a free-flow solution, under certain monotonicity assumptions. The second sub-problem determines the optimal emission prices of all links in the network, which was formulated as an optimal control problem to minimize the total tolls. Wardrop's route choice principle was integrated to the model as part of the constraints. The free-flow optimal solution obtained



(a) Partial Sioux Falls network (b) Link $(16, 17)$ (c) Link $(17, 19)$

Figure 11: Dynamic link flows and tolls of selected links in Sioux Falls network

from the first sub-problem was used as an input to the second sub-problem. Numerical results on multiple testing networks were reported.

The results in this paper indicate that the monotonicity property of the emission (or fuel consumption) functions is crucial for the network emission modeling and control. They further indicate that the monotonicity properties of different pollutants hold at different speed ranges. This implies that there are tradeoffs among the different types of pollutants, which may not be simultaneously reduced. In practice, therefore, one may want to focus on one specific pollutant (such as CO_2, CO, or others). In this paper, we also propose to use the fuel consumption as a proxy to capture all pollutants. This, of course, may not be the best way to address the issue and thus merits further investigations. The results also show the importance of the free-flow traffic state in developing network pricing schemes and managing traffic congestion, emissions/fuel consumption, and other related issues. One may focus on the free-flow traffic state in a network and devise pricing schemes and other management strategies to sustain such free-flow state. This will not only be beneficial to the network as a whole from the perspective of congestion and emission control, but also lead to simplified dynamic network models (since the traffic dynamics are much simpler under the free-flow traffic state) so that the optimal pricing scheme or other strategies can be much easily calculated.

For future research, firstly, we notice that in the current situation, the real world traffic state in a network is rarely free flowing especially during peak periods. The results here provide an 'ideal' situation where one may fully control / manage the system (using pricing and possibly other emerging strategies) to possibly achieve certain desirable state (the free flow state in this case) which can hopefully provide guidance / insights for real world traffic management and control. Arguably this is what most DSO studies try to achieve, i.e., they aim to provide a benchmark or target so that real world traffic control / management can rely on or compare with. The first step in our case here is to determine whether such a free-flow state exists and if so, how to obtain it. Our previous work in Ma et al. (2014) indicated that if only travel times are considered (as the objective), one can find a free-flow optimal solution if it exists. In this paper, we extended the results to state that if both travel times and emissions (or fuel consumption) are considered, an optimal solution must be a free-flow solution, if exists. We also proposed methods to obtain it. In practice, how to implement the full control of the system is a huge challenge as it has to involve both traffic control and demand management (with the guidance of the desirable DSO solution). This might be done for specific situations (such as emergency evacuation, special events, etc.). With emerging new technologies and systems such as connected and automated vehicles, new opportunities may also emerge. Investigating the challenges of how to achieve a certain DSO state is beyond the scope of the current paper and will be certainly studied in the future.

Secondly, the link flow dynamics, the junction model, and the above results in this paper are applicable for a network with a single destination, and the first-in-first-out (FIFO) condition is easily assured. However, for a general multi-destination network, the FIFO condition may not be guaranteed without proper modifications of the proposed model. Studying on the DSO with the double-queue model for general multi-destination networks would be an interesting topic for future research. Furthermore, for a multi-destination network, it is a question whether a free-flow solution exists. This needs to be studied thoroughly in future research. Even if a free flow solution does not exist, the idea proposed in the paper may still be useful. That is, one can try to explore specific DSO solutions that are both beneficial to system control objectives (travel times, emissions, fuel consumption, etc.) and easy to implement. The system management can then focus on how to achieve and sustain such optimal state, and to design specific control/management strategies. In practice, the desirable DSO solution may not be perfectly achieved and maintained; however, it can provide the benchmark (like most DSO studies do) that can guide practical traffic system management and control.

Thirdly, the emission/fuel consumption functions adopted in this paper were originally proposed for static assignment problems. We extended them to the dynamic case with simple schemes. They are aggregated function based on one traffic state variable such as travel time or speed, which are coarse to capture detailed vehicle dynamics such as their accelerations and decelerations. More research is needed in the future to develop more detailed emission/fuel consumption functions that can better capture the dynamic features of vehicles. We believe, however, that even with the current coarse emission functions, the model and results in the paper should be still useful for network wide emission control, such as the monotonicity properties of the emission functions and the design of the control strategies around free flow solutions. On the other hand, there are recent developments in the literature regarding modeling emissions in a more detailed manner, such as via microscopic traffic simulation and via dynamic emission functions that consider vehicle acceleration and vehicle specific parameters; see Szeto et al. (2012). Especially, recent studies such as Piccoli et al. (2015) and Han et al. (2016a) can provide very insightful directions. For example, one may apply data-driven techniques as did in Han et al. (2016a) to fit an aggregated emission or fuel consumption function over multiple variables instead of speed or travel time only. The results in this paper can be helpful in designing such new dynamic emission models. For instance, since ideally the emission function should have certain monotonicity properties, a monotone function may be designed and calibrated as the desired emission/fuel consumption function. In this way, the emission models can better capture vehicle dynamic features, while at the same time also have desired characteristics such as monotonicity to facilitate rigorous analysis.

Fourthly, as aforementioned, the monotonicity of the emission function is crucial for the proposed methodology. As indicated in the literature and in this paper, the emission functions are monotonic under specific conditions when the speed is within certain ranges. Notice that these monotonicity properties are sufficient conditions, which means even if the monotonicity properties are not maintained for certain functions, the results in this paper may still be valid (e.g., the free-flow solution may still exist). In the future, sharper conditions (weaker than monotonicity) should be developed for the methods proposed in this paper to apply. It is our understanding that to derive conclusive results, the emission function and the modeling methodology used in dynamic network analysis needs to balance between its mathematical rigor and its ability to capture realism. In future research, we will investigate issues of using a dynamic or more advanced emission function that can better represent dynamic traffic, which at the same time, has desirable properties to be used in dynamic network analysis.

Last but not least, the results in this research also imply that to properly design and implement the first-best dynamic pricing scheme in a network level, both technical and economic/policy advances are critical. For the former, recent advances in connected/automated vehicles and mobile sensing can play an important role in order to track individual vehicles and enable communications and data sharing among vehicles and between vehicles and the infrastructure. For the latter, new pricing schemes (such as emission pricing as we study in this paper) and recent research works in designing tradable credits in traffic networks (see Yang and Wang (2011); Nie and Yin (2013)) are crucial for network-wide congestion/emission management and control. For this, future research may be conducted to study how the pricing and tradable credit schemes can be properly combined in a dynamic traffic network to control both traffic congestion and emissions.

Acknowledgements

The authors are grateful to the three anonymous reviewers for their constructive comments to the earlier versions of the paper. The research described in this paper was partially funded by grants from the National Science Foundation (NSF) (CMMI-1055555), the Research Grants Council of the Hong Kong Special Administrative Region, China (HKU 17201915), the University Research Committee from the University of Hong Kong (201411159063), and the National Natural Science Foundation of China (71271183). The second author is also supported by a grant from the State Key Laboratory of Automotive Safety and Energy, Tsinghua University, China. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the NSF, the Research Grants Council of the Hong Kong, the National Natural Science Foundation of China, or the State Key Laboratory of Automotive Safety and Energy.

Appendix A. Summary of the literature on dynamic road pricing

Table A.2: Summary of the literature on dynamic road pricing.

	Pricing focus (congestion /emission /both)	Choice (route /departure time /mode) and demand consideration	Pricing scheme	Traffic flow modeling approach	Queue spillback
Henderson (1974)	congestion	departure time	first	flow density funciton	no
Agnew (1977)	congestion	departure time	first	flow density funciton	no
Ben-Akiva et al. (1986)	congestion	route and departure time / elastic demand	second	deterministic queuing model	no
Braid (1989)	congestion	departure time/elastic demand	first and second	deterministic queuing model	no
Arnott et al. (1990)	congestion	route and departure time	first and second	deterministic queuing model	no
Amott et al. (1993)	congestion	departure time/ elastic demand	second	deterministic queuing model	no
Carey and Srinivasan (1993)	congestion	route	first	exit flow model	no
Laih (1994)	congestion	departure time	second	deterministic queuing model	no
Chu (1995)	congestion	elastic demand	first	flow density function	no
Braid (1996)	congestion	route and departure time/elastic demand	first and second	deterministic queuing model	no
Yang and Huang (1997)	congestion	departure time/elastic demand	first	improved exit flow model	no
Wie and Tobin (1998)	congestion	route	first	modified exit flow model	no
Arnott and Kraus (1998)	congestion	departure time	first and second	deterministic queuing model	no
Yang and Meng (1998)	congestion	route and departure time/elastic demand	first	point queue on space time expanded network	no
Daganzo and Garcia (2000)	congestion	departure time	second	deterministic queuing model	no
de Palma and Lindsey (2000)	congestion	route and departure time/elastic demand	Profit maximization pricing and competition	deterministic queuing model	no
de Palma et al. (2004)	congestion	route and departure time and mode	second	deterministic queuing model	no
de Palma et al. (2005b)	congestion	route, mode and departure time	first and second	METROPOLIS - microsimulation model	yes
de Palma et al. (2005a)	congestion	route, mode and departure time	second and third (no queue tolling)	METROPOLIS - microsimulation model	yes
Lo and Szeto (2005)	congestion	route	first	CTM	yes

Joksimovic et al. (2005)	congestion	route and departure time / elastic demand	second	arc delay model/ link performance function	no
Mahmassani et al. (2005)	congestion	route and departure time	second	DYNASMART	yes
Wie (2007)	congestion	route and departure time / elastic demand	second	arc delay model/ link performance function	no
Lu et al. (2008)	congestion	route	second	DYNASMART	yes
Chow (2009b)	congestion	departure time	first	arc delay model, deterministic queuing model, and Whole-link traffic model	no
Chow (2009a)	congestion	departure time	second	arc delay model, deterministic queuing model, and Whole-link traffic model	no
Geroliminis and Levinson (2009)	congestion	departure time	second	combining deterministic queuing model and Macroscopic Fundamental Diagram	yes
Shen and Zhang (2009)	congestion	route and departure time	second	deterministic queuing model	no
Lin et al. (2011)	congestion	route	first and second	CTM	yes
Lu and Mahmassani (2011)	congestion	route and departure time	second	DYNASMART	yes
Chung et al. (2012)	congestion	route and departure time	second	arc delay model	no
Zheng et al. (2012)	congestion	departure time, mode, route, activity sequence	second	combining macroscopic fundamental diagram and an agent-based traffic model	yes
Zhong et al. (2012)	congestion and emissions	route and departure time	first	the whole link model and deterministic queuing model	no
Michalaka and Hale (2013)	congestion	lane	second	CORSIM	yes
Friesz et al. (2013)	congestion and emissions	route and departure time	second	the LWR-Lax model	no
Zhang et al. (2013)	congestion	route	second	DYNASMART	yes
Chen et al. (2014)	congestion	route	second	Anisotropic Mesoscopic Simulation (AMS) model	yes
Zhu and Ukkusuri (2015)	congestion	route	second	CTM	yes
Kickhöfer and Kern (2015)	congestion and emissions	route and mode	first	large-scale agent-based traffic microsimulation	yes
Kickhöfer and Nagel (2016)	emissions	route and mode	first	large-scale agent-based traffic microsimulation	yes
This paper (2016)	congestion and emissions	route and departure time	first	double queue model	yes

Appendix B. Monotonicity conditions for emission functions (6)

Coefficients of emission functions for fuel consumption and different types of pollutants in (6) are listed in Table B.3.

By assuming the road grade to be zero, the emission function is a nonlinear function on the average speed s_{ij}.

$$f_{ij} = \exp\left(\beta_0 + \beta_1 s_{ij} + \beta_2 s_{ij}^2 + \beta_3 s_{ij}^3 + \beta_4 s_{ij}^4\right). \tag{B.1}$$

Piecewise monotonicity conditions over the variable s_{ij} can be observed from these emission functions, as shown in Table B.4.

Table B.3: Coefficients of emission functions (Source: Boriboonsomsin et al. (2012))

	Fuel	CO_2	CO	HC	NO_x
β_0	6.80e+0	7.96e+0	-1.57e-1	-2.12e+0	-8.14e-1
β_1	-1.40e-1	-1.40e-1	-1.36e-1	-1.45e-1	-1.07e-1
β_2	3.92e-3	3.92e-3	4.70e-3	4.56e-3	4.40e-3
β_3	-5.20e-5	-5.20e-5	-6.96e-5	-6.50e-5	-7.29e-5
β_4	2.57e-7	2.57e-7	3.70e-7	3.35e-7	4.17e-7
β_5	1.37e-1	1.37e-1	2.67e-1	1.65e-1	4.02e-1

Table B.4: Piecewise monotonoicity of emission functions

	Increasing intervals (mph)	Decreasing intervals (mph)
Fuel	$(72.38, +\infty)$	$[0, 72.38)$
CO_2	$(72.38, +\infty)$	$(23.00, 42.52) \cup (65.59, +\infty)$
CO	$(32.99, 42.40) \cup (65.69, +\infty)$	$[0, 32.99) \cup (42.40, 65.69)$
HC	$(71.50, +\infty)$	$[0, 71.50)$
NO_x	$(23.00, 42.52) \cup (65.59, +\infty)$	$[0, 23.00) \cup (42.52, 65.59)$

Appendix C. Applicability on the state variable

The monotonicity of the emissions/fuel consumption functions rely on the proper monotonic regions of the traffic speed. On the other hand, since the monotonicity conditions are based on the downstream queues (i.e., the link state variables), in this Appendix we list the regions of the queue length in which the desired monotonicity condition can be satisfied, given the link length and the free-flow travel time as known parameters.

1) The CO emissions with the unit of gram per vehicle per hour

As mentioned, $\mathcal{E}^{\tau}_{ij}(q^d_{ij})$ is an increasing monotonic function when the speed is no more than 46.6 mph, i.e., $s = \dfrac{l_{ij}}{\tau^0_{ij} + \frac{q^d_{ij}}{\overline{C}_{ij}}} \in [0, 72.38)$ mph. So the desired queue length on link (i, j) should be

$$q^d_{ij} \geq \left(\frac{l_{ij}}{46.6} - \tau^0_{ij}\right) \overline{C}_{ij}.$$

• If the free-flow speed is greater than 46.6 mph, i.e., , then a shorter queue $q^d_{ij} < \left(\dfrac{l_{ij}}{46.6} - \tau^0_{ij}\right) \overline{C}_{ij}$ would not satisfy the desired condition.

• If the free-flow speed is less than or equal to 46.6 mph, then any non-negative queue length $q^d_{ij} > 0$ would satisfy the desired condition.

Overall the monotonicity condition of such emission function is applicable for all queuing condition if the free-flow speed of any link in the study network does not exceed 46.6 mph.

2a) The fuel consumption with the unit of gram per length per vehicle

$F^s(s)$ should be a monotonic decreasing function, which means the speed $s \in [0, 72.38)$ mph according to Table B.3. Similarly, the desired queue length on link (i, j) is

$$q^d_{ij} \geq \left(\frac{l_{ij}}{72.38} - \tau^0_{ij}\right) \overline{C}_{ij}.$$

• If the free-flow speed is greater than 72.38 mph, i.e., , then a shorter queue $q^d_{ij} < \left(\dfrac{l_{ij}}{46.6} - \tau^0_{ij}\right) \overline{C}_{ij}$ would not satisfy the desired condition.

• If the free-flow speed is less than or equal to 72.38 mph, then any non-negative queue length $q_{ij}^d > 0$ would satisfy the desired condition.

Overall the monotonicity condition of such emission function is applicable for all queuing condition if the free-flow speed of any link in the study network does not exceed 72.38 mph.

Table C.5: Applicable queue length

$\frac{l_{ij}}{\tau_{ij}^0}$	(0, 23]	(23, 42.52]	(42.52, 65.59]	(65.59, $+\infty$)
q_{ij}^d	0	$\left(0, (\frac{l_{ij}}{23} - \tau_{ij}^0)\overline{C}_{ij}\right)$	$\left((\frac{l_{ij}}{42.52} - \tau_{ij}^0)\overline{C}_{ij}, (\frac{l_{ij}}{23} - \tau_{ij}^0)\overline{C}_{ij}\right)$	$\left(0, (\frac{l_{ij}}{65.59} - \tau_{ij}^0)\overline{C}_{ij}\right) \cup \left((\frac{l_{ij}}{42.52} - \tau_{ij}^0)\overline{C}_{ij}, (\frac{l_{ij}}{23} - \tau_{ij}^0)$

2b) The CO_2 emissions with the unit of gram per length per vehicle

$F^s(s)$ should be a monotonic decreasing function, which means the speed $s \in (23, 42.52) \cup (65.59, +\infty)$ mph. The desired queue length on link (i, j) is

$$q_{ij}^d \in \left[-\tau_{ij}^0\overline{C}_{ij}, (\frac{l_{ij}}{65.59} - \tau_{ij}^0)\overline{C}_{ij}\right) \cup \left((\frac{l_{ij}}{42.52} - \tau_{ij}^0)\overline{C}_{ij}, (\frac{l_{ij}}{23} - \tau_{ij}^0)\overline{C}_{ij}\right).$$

Since the queue length is always non-negative, we can simplify the regions of queue length for different possible free-flow speed shown in Table C.5. It is noted that only when the free-flow speed is less than 42.52 mph or greater than 65.59 mph, the monotonic condition is satisfied at the free-flow condition (zero queue).

The conditions for the HC and NO_x emissions with the unit of gram per length per vehicle are omitted here since they can be acquired similarly as for fuel consumption and the CO emissions.

Appendix D. Proof of Proposition 1

Proof.

Assume $X^g \notin \Omega^f$, i.e., X^g is not a free-flow feasible solution. We need to construct a free-flow feasible solution $X^{g;f}$.

From the double queue dynamics and the definition of number of vehicles of link (i, j) at time t, $n_{ij}(t)$, we have

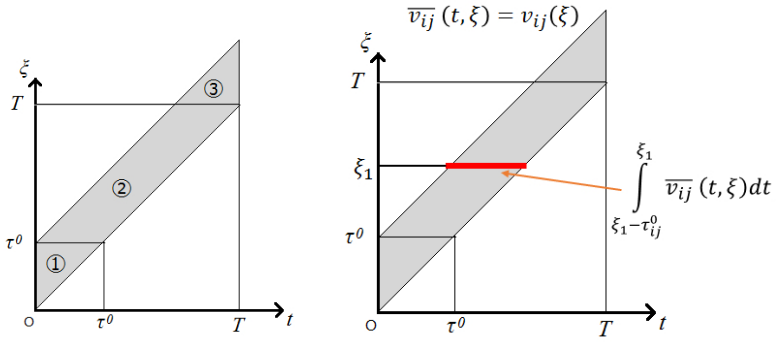
$$\begin{aligned} n_{ij}(t) &\triangleq \textstyle\int_0^t [p_{ij}(\xi) - v_{ij}(\xi)]d\xi \\ &= \textstyle\int_0^t [p_{ij}(\xi) - v_{ij}(\xi + \tau_{ij}^0) + v_{ij}(\xi + \tau_{ij}^0) - v_{ij}(\xi)]d\xi \\ &= q_{ij}^d(t + \tau_{ij}^0) + \textstyle\int_t^{t+\tau_{ij}^0} v_{ij}(\xi)d\xi \end{aligned} \tag{D.1}$$

Substituting (D.1) into (9), it is derived that

$$\begin{aligned} \text{TE} &= \sum_{(i,j)\in L} E_{ij}|_0^T \\ &= \sum_{(i,j)\in L} \int_0^T n_{ij}(t)e_{ij}(q_{ij}^d(t + \tau_{ij}^0))dt \\ &= \sum_{(i,j)\in L} \int_0^T q_{ij}^d(t + \tau_{ij}^0) \cdot e_{ij}(q_{ij}^d(t + \tau_{ij}^0))dt + \sum_{(i,j)\in L} \int_0^T \int_t^{t+\tau_{ij}^0} v_{ij}(\xi)d\xi \cdot e_{ij}(q_{ij}^d(t + \tau_{ij}^0))dt \end{aligned} \tag{D.2}$$

From the double-queue dynamics, boundary conditions and the non-negativity of the downstream queue, it is easily shown that $v_{ij}(t) = 0 \forall t \in \left[0, \tau_{ij}^0\right) \cap (T, +\infty) \forall (i, j) \in L$. We define a dummy term $\overline{v}_{ij}(t; \xi) \triangleq v_{ij}(\xi)$. Then we have

$$\int_0^T \int_t^{t+\tau_{ij}^0} v_{ij}(\xi)d\xi dt = \int_0^T \int_t^{t+\tau_{ij}^0} \overline{v}_{ij}(t; \xi)d\xi dt \tag{D.3}$$



(a) Double integral areas in Equation (D.3)

(b) Integral for Area 2

Figure D.12: Multiple integral of Equation (D.3)

The right-hand-side of Equation (D.3) is a double integral over ξ and t. The region of such double integral is presented in the $t-\xi$ space as the shadow area in Figure 12(a). Such area can be further divided into three sub-areas, i.e., $\int_0^T \int_t^{t+\tau_{ij}^0} \overline{v}_{ij}(t;\xi)d\xi dt = MI_1 + MI_2 + MI_3$, where

$$\begin{aligned} MI_1 &= \int_0^{\tau_{ij}^0} \int_t^{\tau_{ij}^0} \overline{v}_{ij}(t;\xi)d\xi dt \\ MI_2 &= \int_{\tau_{ij}^0}^{T} \int_{\xi-\tau_{ij}^0}^{\xi} \overline{v}_{ij}(t;\xi)dtd\xi \\ MI_3 &= \int_{T-\tau_{ij}^0}^{T} \int_T^{t+\tau_{ij}^0} \overline{v}_{ij}(t;\xi)d\xi dt \end{aligned} \tag{D.4}$$

For Area 1, since $v_{ij}(\xi) = 0\ \forall t \in [0, \tau_{ij}^0)$, $MI_1 = 0$. For Area 3, since $v_{ij}(\xi) = 0\ \forall t \in (T, +\infty)$, we have $MI_3 = 0$. For Area 2, different from the integral in Eq (D.3), the sequence of double integral is switched. Each horizontal line with the same ξ is integrated first, and then the integral over all ξ is made, which is illustrated in Figure 12(b). We then have

$$\begin{aligned} MI_2 &= \int_{\tau_{ij}^0}^{T} \int_{\xi-\tau_{ij}^0}^{\xi} \overline{v}_{ij}(t;\xi)dtd\xi = \int_{\tau_{ij}^0}^{T} \int_{\xi-\tau_{ij}^0}^{\xi} v_{ij}(\xi)dtd\xi \\ &= \int_{\tau_{ij}^0}^{T} \left[\int_{\xi-\tau_{ij}^0}^{\xi} dt\right] v_{ij}(\xi)d\xi \\ &= \int_{\tau_{ij}^0}^{T} \tau_{ij}^0 v_{ij}(\xi)d\xi = \tau_{ij}^0 \int_0^T v_{ij}(t)dt \end{aligned}$$

Then we have

$$\int_0^T \int_t^{t+\tau_{ij}^0} v_{ij}(\xi)d\xi dt = \tau_{ij}^0 \int_0^T v_{ij}(t)dt \tag{D.5}$$

For $\forall X^g \in \Omega \setminus \Omega^f$, $\exists (i,j) \in L\ \exists t \in [0,T]$, $q_{ij}^{d;g}(t) > 0$. Since $e_{ij}(\cdot)$ is monotonically increasing, $q_{ij}^d(t+\tau_{ij}^0)e_{ij}(q_{ij}^d(t+\tau_{ij}^0))$ is also monotonically increasing on $q_{ij}^d(t+\tau_{ij}^0)$. Then we have

$$\begin{aligned} \mathrm{TE}(X^g) &> \sum_{(i,j)\in L} \int_0^T \int_t^{t+\tau_{ij}^0} v_{ij}^g(\xi)d\xi \cdot e_{ij}(0))dt \\ &= \sum_{(i,j)\in L} e_{ij}(0)\tau_{ij}^0 \int_0^T v_{ij}^g(t)dt \end{aligned} \tag{D.6}$$

where the total cumulative exit flow from link (i,j) is defined as $\overline{V}_{ij}^{g} \triangleq \int_0^T v_{ij}^{g}(t)dt$, which is equivalent to the total number of vehicles passing through link (i,j) over the entire time span.

In the following, we construct a free-flow feasible solution $X^{g;f}$ that satisfies all the constraints defining the free-flow feasible solution set, i.e., $X^{g;f} \in \Omega^f$. In the constructed free-flow feasible solution $X^{g;f}$, any link (i,j) shares a common total cumulative exit flow with the non-free-flow feasible solution X^g, i.e., $\int_0^T v_{ij}^{g;f}(t)dt \triangleq \overline{V}_{ij}^{g;f} = \overline{V}_{ij}^{g} \triangleq \int_0^T v_{ij}^{g}(t)dt$.

We construct $X^{g;f}$ by three steps. Step (i) arranges the dummy origin links in an arbitrary order, and defines the cumulative flow to be assigned for each link. We start from the first path starting from dummy origin node $\widetilde{o}_1$, i.e., $k=1$, $r=1$. In Step (ii), from path $P_{\widetilde{o}_k;r}$, i.e., the r-th path from dummy origin $\widetilde{o}_k$ to the destination, we choose the link with the least the cumulative flow to be assigned, subtract such amount of cumulative exit flow from all the links on path $P_{\widetilde{o}_k;r}$, and reassign such amount of flow onto this path as free-flow in $X^{g;f}$, i.e., no downstream queues on the path except on the dummy origin link. The link with the least the cumulative flow to be assigned will not carry any more flow for later time periods after Step (ii) is done. In Step (iii), we check if there is any cumulative flow to be assigned on the dummy origin link $(\widetilde{o}_k, o_k)$. If so, we scroll the time period forward, and redo Step (ii) for the $(r+1)$-th path starting from dummy origin $\widetilde{o}_k$. If there is no more cumulative flow to be assigned on the dummy origin link $(\widetilde{o}_k, o_k)$ and the dummy origin link $(\widetilde{o}_k, o_k)$ is not the last one, we increase k by 1 and set $r=1$, scroll the time period forward, and redo Step (ii) for the first path starting from dummy origin node $\widetilde{o}_{k+1}$. If Step (ii) has been done for all the paths, then there is no cumulative flow to be assigned, and the construction of the free-flow feasible solution is finished.

Step (i) Arrange the dummy links as $(\widetilde{o}_1, o_1), (\widetilde{o}_2, o_2), ..., (\widetilde{o}_n, o_n)$, where n is the number of origin nodes. The ordering of these links can be arbitrary and do not change the way how $X^{g;f}$ is constructed. Define the starting time $t^{\widetilde{o}_k;r;\text{start}}$ for each pair of k and r. Set $t^{\widetilde{o}_1;1;\text{start}} \triangleq 0$. Set $k=1, r=1$.

For link $(i,j) \in \mathcal{L}$, we define the cumulative flow to be assigned as $\overline{F}_{ij}$. Initially, $\overline{F}_{ij} = \overline{V}_{ij}^{g}$.

Step (ii) Let $P_{\widetilde{o}_k;r}$: $\widetilde{o}_k \triangleq i_0 \to i_1 \to i_2 \to \cdots \to i_{m_{k;r}-1} \to i_{m_{k;r}} \triangleq \widehat{s}$ be a path joining the dummy node $\widetilde{o}_k$ to the destination $\widehat{s}$, where $m_{k;r}$ is the number of links on path $P_{\widetilde{o}_k;r}$. For any link $(i_{\ell-1}, i_\ell)$, $1 \le \ell \le m_{k;r}$ on this path, $\overline{F}_{ij} > 0$. We define $F_{k;r} \triangleq \min_{1\le\ell\le m_{k;r}} \overline{F}_{i_{\ell-1},i_\ell}$, and $\overline{C}_{k;r} \triangleq \min_{1\le\ell\le m_{k;r}} \overline{C}_{i_{\ell-1},i_\ell}$.

Define $t_{ij}^{\widetilde{o}_k;r}$ as the time when the flow on path $P_{\widetilde{o}_k;r}$ starts to enter link (i,j); thus $t_{i_0 i_1}^{\widetilde{o}_k;r} = t^{\widetilde{o}_k;r;\text{start}}$; inductively, $t_{i_\ell i_{\ell+1}}^{\widetilde{o}_k;r} = t_{i_{\ell-1} i_\ell}^{\widetilde{o}_k;r} + \tau_{i_{\ell-1} i_\ell}^{0}$ for $\ell = 1, \cdots, m_{k;r}$.

For $\ell = 1, \cdots, m_{k;r}$, set $v_{i_{\ell-1},i_\ell}(t+\tau_{i_{\ell-1},i_\ell}^{0}) = p_{i_{\ell-1},i_\ell}(t) = \overline{C}_{k;r} \le \overline{C}_{i_{\ell-1},i_\ell}$ for $t \in \left[t_{i_{\ell-1},i_\ell}^{\widetilde{o}_k;r}, t_{i_{\ell-1},i_\ell}^{\widetilde{o}_k;r} + \dfrac{F_{k;r}}{\overline{C}_{k;r}} \right]$, which yields $\dot{q}_{i_{\ell-1},i_\ell}^{d}(t) = 0$ for all $t \in \left[t_{i_{\ell-1},i_\ell}^{\widetilde{o}_k;r}, t_{i_{\ell-1},i_\ell}^{\widetilde{o}_k;r} + \dfrac{F_{k;r}}{\overline{C}_{k;r}} \right]$. As $q_{i_{\ell-1},i_\ell}^{d}(t^{\widetilde{o}_k;r;\text{start}}) = 0$, the downstream queue of link $(i_{\ell-1}, i_\ell)$ is always zero, i.e. $q_{i_{\ell-1},i_\ell}^{d}(t) = 0$, which means it takes the free-flow travel time τ_{ij}^{0} for the flow to travel through each link $(i_{\ell-1}, i_\ell)$ on path $P_{\widetilde{o}_k;r}$. The total travel time on path $P_{\widetilde{o}_k;r}$ is thus $\sum_{\ell=1,\cdots,m_{k;r}} \tau_{i_{\ell-1},i_\ell}^{0}$. Define $t^{\widetilde{o}_k;r;\text{end}} \triangleq t^{\widetilde{o}_k;r;\text{start}} + \sum_{\ell=1,\cdots,m_{k;r}} \tau_{i_{\ell-1},i_\ell}^{0} + \dfrac{F_{k;r}}{\overline{C}_{k;r}}$.

Subtract $F_{k;r}$ from $\overline{F}_{i_{\ell-1},i_\ell}$ for all the links $(i_{\ell-1}, i_\ell)$, $\ell = 1, \cdots, m_{k;r}$ on path $P_{\widetilde{o}_k;r}$.

Step (iii) After Step (ii),

• if $\overline{F}_{\widetilde{o}_k,o_k} > 0$, then set $t^{\widetilde{o}_k;r+1;\text{start}} \triangleq t^{\widetilde{o}_k;r;\text{end}}$, increase r by 1, and repeat Step (ii) above;

• if $\overline{F}_{\widetilde{o}_k,o_k} = 0$ and $k < n$, then set $t^{\widetilde{o}_{k+1};1;\text{start}} \triangleq t^{\widetilde{o}_k;r;\text{end}}$, increase k by 1, reset $r=1$, and repeat Step (ii) above;

• otherwise finish the construction of the solution $X^{g;f}$.

Since all cumulative flow, presented as $\overline{F}_{\widetilde{o}_k,o_k}$ for any k, has been re-distributed following the construction before we finish the construction, Step (iii) guarantees the same amounts of cumulative flow are in $X^{g;f}$ and X^{g} for all the links, and all the links have zero downstream queues. Moreover, $X^{g;f}$ has a smaller objective value than that of the non-free-flow feasible solution X^{g}. This is because for the objective of minimizing total emissions, it matters where the vehicles wait. If they wait within the regular network, i.e., forming downstream queues on regular links, they generate emissions. If they wait at the origin, i.e., forming

downstream queues on dummy origin links, they do not generate emissions. Based on Equation (D.6), we then have Equation (D.7) showing that the total emissions of the constructed free-flow solution $X^{g;f}$ is less than that of X^g.

$$\begin{aligned} \mathrm{TE}(X^{g;f}) &= \sum_{(i,j)\in L} \int_0^T \int_t^{t+\tau_{ij}^0} v_{ij}^{g;f}(\xi) d\xi \cdot e_{ij}(0)) dt \\ &= \sum_{(i,j)\in L} e_{ij}(0)\tau_{ij}^0 \int_0^T v_{ij}^{g;f}(t) dt \\ &= \sum_{(i,j)\in L} e_{ij}(0)\tau_{ij}^0 \int_0^T v_{ij}^{g}(t) dt < \mathrm{TE}(X^g) \end{aligned} \tag{D.7}$$

This completes the proof. □

Appendix E. Proof of Proposition 3

Proof. We first show why there could be downstream queues on the regular links, and propose a method to construct a free-flow feasible solution X^f from a non-free-flow feasible solution X, by shifting the downstream queues on the regular links to the dummy origin links, without modifying the arrival flow, i.e., the exit flow from the links connecting to the single destination.

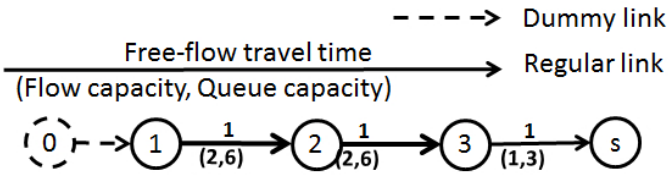
As discussed in Ma et al. (2014), it does not matter for the original DSO problem (13) where the vehicles wait. The vehicles could wait at the origins or any intermediate links, while the total travel time is the same. Because of this, the optimal solution to problem (13) could be free-flow or non-free-flow.

We illustrate this by simple scenarios. First we show the simplest scenario, where the starting node of the congested link is a one-in-one-out node. Figure 13(a) shows a simple chain network with only one OD pair. Link $(0, 1)$ is a dummy origin link, and the destination node is s. The flow capacities of regular links $(1, 2)$, $(2, 3)$ and $(3, s)$ are 2, 2, 1, respectively. Total demand is 2 and is initially stored at the dummy origin link. The inflow rate capacity of link $(3, s)$ is smaller than that of link $(2, 3)$.

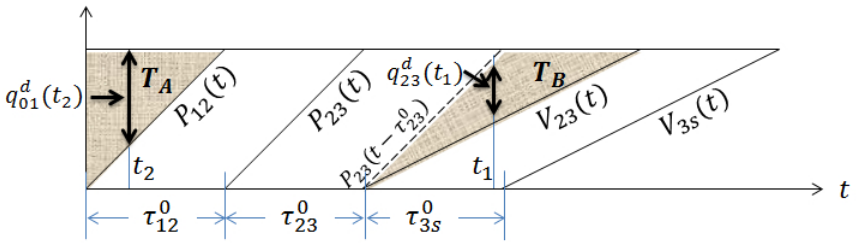
As shown in Figure 13(b), in the non-free-flow solution, because the inflow rate of link $(2, 3)$ is 2 while the exit flow rate is 1, the downstream queue is built up on the regular link $(2, 3)$ for a period of time. For instance, at time t_1, the downstream queue of link $(2, 3)$ is positive. In this case, link $(2, 3)$ is congested, and the starting node of link $(2, 3)$, i.e., node 2, is a one-in-one-out node. The total waiting time caused by this downstream queue is presented as the area of T_B. In this solution, there is another part of waiting time on the dummy origin link. The waiting time caused by the demand waiting to be discharged from the dummy link $(0, 1)$ is presented as the area of T_A.

Figure 13(c) shows a free-flow solution, where there is no downstream queue on any of the regular links. All links discharge their own flow at the bottleneck flow capacity, which is 1, so that all regular links are in their free-flow state. The waiting time is only caused by the demand waiting to be discharged from the dummy link $(0, 1)$. It is obvious from Figures 13(b) and 13(c) that the total waiting time in the non-free-flow solution equals the total waiting time in the free-flow solution. We further notice that the arrival flows in both solutions are exactly the same, i.e., the cumulative flow curves of link $(3, s)$ in both solutions are the same. From the original DSO formulation (13), these two solutions share the same objective value. In fact, these two solutions are both optimal solutions for the original DSO formulation (13) that only minimizes the time cost. This example shows that if the starting node of a congested link (i.e., a link with a positive downstream queue) has only one upstream link and one downstream link (one-in-one-out), the within-network downstream queues can be shifted to the upstream links up to the dummy origin links, by changing the cumulative flow curves of the network links.

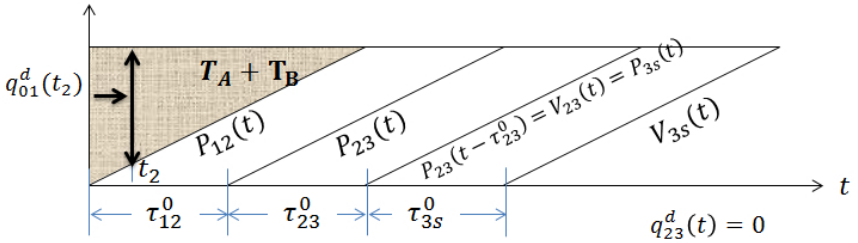
We then show that if the starting node of a congested link has two upstream links and one downstream link (two-in-one-out, merging), the downstream queues can still be shifted to upstream links, and the exit flow of the congested link would not change.



(a) A simple chain network



(b) Cumulative flows of a non-free-flow solution



(c) Cumulative flows of a free-flow solution

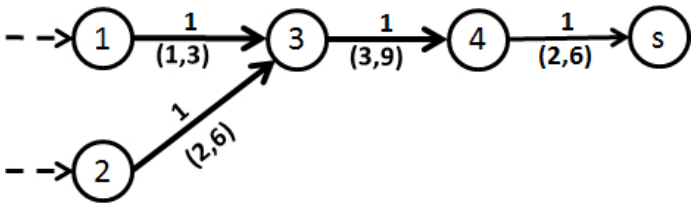
Figure E.13: Shifting the downstream queue on congested link $(2,3)$ to dummy origin link $(0,1)$

Figure 14(a) shows a network with a merging node. For each dummy origin link, the demand is 2. Other network characteristic can be found in Figure 14(a). The inflow rate capacity of link $(4,s)$ is smaller than that of link $(3,4)$.

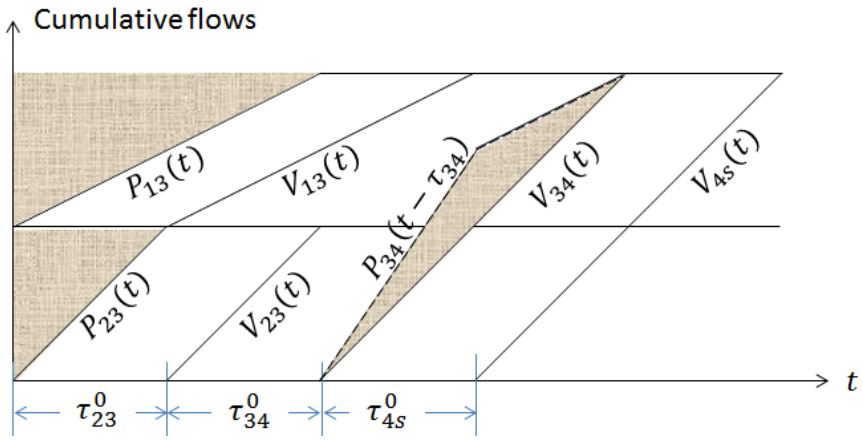
As shown in Figure 14(b), in the non-free-flow solution, because the inflow rate of link $(3,4)$ is 3 for a period of time, while the exit flow rate is 2, the downstream queue is built up on the regular link $(3,4)$. In this case, link $(3,4)$ is congested, and the starting node of link $(3,4)$, i.e., node 3, is a merging node. As indicated by the shaded areas, the total waiting time is caused by the downstream queues on link $(3,4)$ and the demand waiting to be discharged at the dummy links.

Figure 14(c) shows a free-flow solution, where there is no downstream queue on any of the regular links. Links $(1,3)$ and $(2,3)$ both discharge the flow at the flow rate of 1, so that all regular links are in free-flow state. The waiting time is only caused by the demand waiting to be discharged from the dummy links. Same observation can be made that from Figures 14(b) and 14(c) that the total waiting time in the non-free-flow solution equals the total waiting time in the free-flow solution, and the cumulative flow curves of link $(4,s)$ in both solutions are the same. This scenario shows that if the starting node of a congested link is a merging node, the downstream queues can be shifted to the upstream links up to the dummy origin links, by changing the cumulative flow curves of the network links.

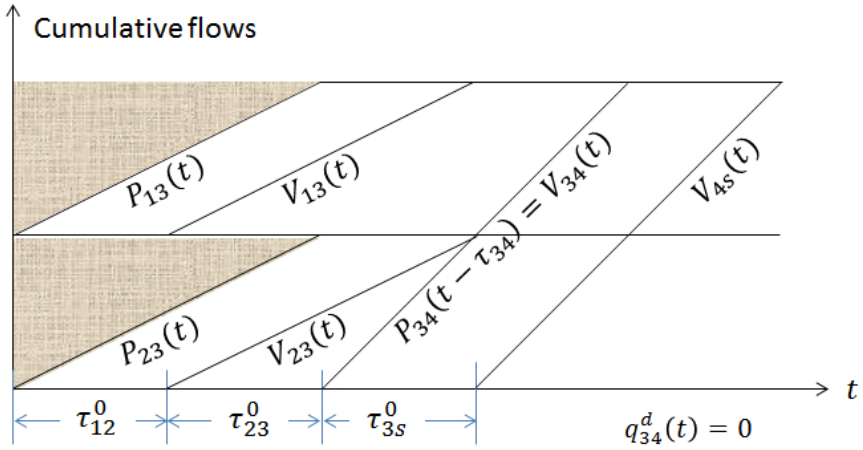
For the scenario where the starting node of a congested link has one upstream links and two downstream link (one-in-two-out, diverging), the downstream queues can be shifted to upstream links as well. Slightly different from the one-in-one-out case, for this case, part of the exit flow of the upstream link is not for the congested link, and should not be modified since it is related to the other downstream link.



(a) Merging node in a network



(b) Cumulative flows of a non-free-flow solution



(c) Cumulative flows of a free-flow solution

Figure E.14: Shifting the downstream queue on congested link $(3, 4)$ to dummy origin links; Node 3 is a merging node

Above we show three scenarios, where the downstream queues can be shifted to upstream links without changing the cumulative exit flow curves of the congested links. We here omit other types of intermediate nodes, since they can be easily transformed into the combinations of one-in-one-out, two-in-one-out and one-in-two-out nodes.

To summarize, the above scenarios show how to shift a non-free-flow feasible solution to a free-flow feasible solution, where the cumulative flow curves need to be modified. More specifically, to shift the downstream queue on a congested link without changing the objective value, the cumulative exit flow curve of the

congested link should not be modified. By decreasing the absolute value of the slopes of cumulative curves for the upstream links, i.e., having them discharge the flow with smaller exit flow rates, the downstream queue on the congested link can be shifted to upstream links in sequence and finally up to the dummy origin links. Compared to the non-free-flow feasible solution, since the absolute values of slopes of cumulative inflow and exit flow curves are smaller in the free-flow feasible solution, all the flow constraints remain satisfied.

Based on the above discussions, a recursive procedure can be developed to obtain a free-flow feasible solution from a general non-free-flow feasible solution.

1. Check the network structure. Network reconstruction may be needed so that the network only contains the above-mentioned three types of nodes, i.e., one-in-one-out, two-in-one-out, and one-in-two-out nodes. See Daganzo (1995) for more details on this.

2. In a non-free-flow feasible solution, choose any regular link (except the dummy origin links) with a positive downstream queue. An operation that shifts the downstream queue to its upstream links can be made as explained above, if such operation does not violate any constraint of the DSO problem. If such operation leads to constraint violation, e.g., the upstream link does not have adequate queue storage, an operation should be made to corresponding upstream link first. Since the dummy origin links do not have queue capacity limits or any upstream link, the recursion will actually start at a link immediately after a dummy origin link, and then proceed to subsequent downstream links until the initially selected link is reached and processed. In the end, this recursive operation can shift the downstream queues on one or multiple links to the dummy origin link.

3. Check if the solution is a free-flow feasible solution. If so, stop. If not, continue on Step 2.

We note that in Shen and Zhang (2014), a method was proposed to construct a holding-free optimal solution to DSO from a non-holding-free solution via creating a space-time expansion network. The recursive method proposed here are directly operated on a continuous-time non-free-flow solution and does not require either time-discrezitation or construction of any space-time network. In particular, the recursive method modifies a non-free-flow feasible solution to a free-flow feasible solution, and does not change the following characteristics:

- Total cumulative flow of each link;
- Cumulative flow curves of the arrival links that connect to the single destination.
- Feasibility, i.e., all constraints are still satisfied.

We thus conclude that any non-free-flow feasible solution $X = (p, v, q^d, n)$ can be modified to a free-flow feasible solution $X^f = (p^f, v^f, q^{d;f}, n^f)$, without violating any constraints, maintain the same total system travel time, and maintain the same total cumulative flow for all the links.

Since in X^f, the total cumulative flow of link (i, j) is not changed, and the emission function is monotonically increasing, we have (see the proof of Proposition 1):

$$\begin{aligned}
\mathrm{TE}(X^f) &= \sum_{(i,j)\in L} \int_0^T \int_t^{t+\tau_{ij}^0} v_{ij}^f(\xi) d\xi \cdot e_{ij}(0)) dt \\
&= \sum_{(i,j)\in L} e_{ij}(0)\tau_{ij}^0 \int_0^T v_{ij}^f(t) dt \\
&= \sum_{(i,j)\in L} e_{ij}(0)\tau_{ij}^0 \int_0^T v_{ij}(t) dt < \mathrm{TE}(X)
\end{aligned}$$

By the original DSO formulation (13), since the cumulative flow curves of the arrival links that connect to the single destination is not changing, we have

$$\mathrm{TT}(X^f) = \mathrm{TT}(X).$$

This completes the proof. □

References

Abou-Zeid, M., 2003. Models and Algorithms for the Optimization of Traffic Flows and Emissions Using Dynamic Routeing and Pricing. Master's thesis. Massachusetts Institite of Technology.

Agnew, C.E., 1977. The theory of congestion tolls. Journal of Regional Science 17, 381–393.

Amott, R., de Palma, A., Lindsey, R., 1993. A structural model of peak-period congestion: a traffic bottleneck with elastic demand. American Economic Review 83, 161–179.

Arnott, A., de Palma, A., Lindsey, R., 1990. Departure time and route choice for the morning commute. Transportation Research Part B 24, 209–228.

Arnott, R., Kraus, M., 1998. When are anonymous congestion charges consistent with marginal cost pricing? Journal of Public Economics 67, 45–64.

Ban, X., Ferris, M., Tang, L., Lu, S., 2013. Risk-neutral second best toll pricing. Transportation Research Part B 48, 67–87.

Ban, X., Herring, R., Hao, P., Bayen, A., 2009a. Delay pattern estimation for signalized intersections using sampled travel times. Transportation Research Record 2130, 109–119.

Ban, X., Liu, H., 2009. A link-node discrete-time dynamic second best toll pricing model with a relaxation solution algorithm. Networks and Spatial Economics 9, 243–267.

Ban, X., Liu, H., Ferris, M., Ran, B., 2008. A link-node complementarity model and solution algorithm for dynamic user equilibria with exact flow propagations. Transportation Research Part B 42, 823–842.

Ban, X., Lu, S., Ferris, M., Liu, H., 2009. Risk averse second best toll pricing, in: Transportation and Traffic Theory, pp. 197–218.

Ban, X., Pang, J.S., Liu, H., Ma, R., 2012a. Continuous-time point-queue models in dynamic network loading. Transportation Research Part B 46, 360–380.

Ban, X., Pang, J.S., Liu, H., Ma, R., 2012b. Modeling and solving continuous-time instantaneous dynamic user equilibria: A differential complementarity systems approach. Transportation Research Part B 46, 389–408.

Ban, X., Gruteser, M., 2012. Towards fine-grained urban traffic knowledge extraction using mobile sensing, in: Proceedings of the ACM SIGKDD International Workshop on Urban Computing, ACM. pp. 111–117.

Ben-Akiva, M., de Palma, A., Kanaroglou, P., 1986. Dynamic model of peak traffic congestion with elastic arrival rates. Transportation Science 20, 164–181.

Boriboonsomsin, K., Barth, M., Zhu, W., Vu, A., 2012. Eco-routing navigation system based on multisource historical and real-time traffic information. Intelligent Transportation Systems, IEEE Transactions on 13, 1694–1704.

Braid, R., 1989. Uniform versus peak-load pricing of a bottleneck with elastic demand. Journal of Urban Economics 26, 320–327.

Braid, R., 1996. Peak-load pricing of a transportation route with an unpriced substitute. Journal of Urban Economics 40, 179–197.

Burger, M., 2003. Infinite-dimensional optimization and optimal design. URL: `ftp://ftp.math.ucla.edu/pub/camreport/cam04-11.pdf`.

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

Chen, X.M., Zhang, L., He, X., Xiong, C., Li, Z., 2014. Surrogate-based optimization of expensive-to-evaluate objective for optimal highway toll charges in transportation network. Computer-Aided Civil and Infrastructure Engineering 29, 359–381.

Chow, A., 2009a. Dynamic system optimal traffic assignment - a state-dependent control theoretic approach. Transportmetrica 5, 85–106.

Chow, A., 2009b. Properties of system optimal traffic assignment with departure time choice and its solution method. Transportation Research Part B 43, 325–344.

Chu, X., 1995. Endogenous trip scheduling: the Henderson approach reformulated and compared with the Vickrey approach. Journal of Urban Economics 37, 324–343.

Chung, B.D., Yao, T., Friesz, T.L., Liu, H., 2012. Dynamic congestion pricing with demand uncertainty: A robust optimization approach. Transportation Research Part B 46, 1504–1518.

Clegg, J., Smith, M., Xiang, Y., Yarrow, R., 2001. Bilevel programming applied to optimising urban transportation. Transportation Research Part B: Methodological 35, 41–70.

Daganzo, C., 1995. The cell transmission model, part II: Network traffic. Transportation Research Part B 29, 79–93.

Daganzo, C., Garcia, R., 2000. A pareto improving strategy for the time-dependent morning commute problem. Transportation Science 34, 303–311.

de Palma, A., Kilani, M., Lindsey, R., 2005a. Comparison of second-best and third-best tolling schemes on a road network. Transportation Research Record 1932, 89–96.

de Palma, A., Kilani, M., Lindsey, R., 2005b. Congestion pricing on a road network: a study using the dynamic equilibrium simulator metropolis. Transportation Research Part A 39, 588–611.

de Palma, A., Lindsey, R., 2000. Private toll road: Competition under various ownership regimes. The Annals of Regional Sciences 34, 13–35.

de Palma, A., Lindsey, R., Quinet, E., 2004. Time-varying road pricing and choice of toll locations. Research in Transportation Economics 9, 107–131.

Ferguson, E.M., Duthie, J., Waller, S.T., 2010. Network methods for project selection based on optimizing environmental impact. Technical Report. University of Texas at Austin. Austin.

Friesz, T.L., Han, K., Liu, H., Yao, T., 2013. Dynamic congestion and tolls with mobile source emission. Procedia - Social and Behavioral Sciences 80, 818–836.

Friesz, T.L., Kwon, C., Mookherjee, R., 2007. A computable theory of dynamic congestion pricing, in: Allsop, R.E., Bell, M.G.H., Benjamin, G.H. (Eds.), Proceedings of 17th international symposium on traffic and transportation theory, London.

Fung, F., He, H., Sharpe, B., Kamakate, F., Blumberg, K., 2010. Overview of China's Vehicle Emission Control Program: Past Successes and Future Prospects. Technical Report. The international council on clean transportation.

Geroliminis, N., Levinson, D., 2009. Cordon pricing consistent with the physics of overcrowding, in: Lam, W.H.K., Wong, S.C., Lo, H.K. (Eds.), Transportation and Traffic Theory 2009: Golden Jubilee, pp. 219–240.

Han, K., Liu, H., Gayah, V., Friesz, T., Yao, T., 2016a. A robust optimization approach for dynamic traffic signal control with emission considerations. Transportation Research Part C: Emerging Technologies 70, 3–26.

Han, K., Piccoli, B., Szeto, W.Y., 2016b. Continuous-time link-based kinematic wave model: Formulation, solution existence, and well-posedness. Transportmetrica B: Transport Dynamics 4(3), 187–222.

Han, K., Szeto, W.Y., Friesz, T.L., 2015. Formulation, existence, and computation of boundedly rational dynamic user equilibrium with fixed or endogenous user tolerance. Transportation Research Part B: Methodological 79, 16–49.

Hao, P., Ban, X., Bennett, K.P., Ji, Q., Sun, Z., 2012. Signal timing estimation using sample intersection travel times. Intelligent Transportation Systems, IEEE Transactions on 13, 792–804.

Henderson, J., 1974. Road congestion: a reconsideration of pricing theory. Journal of Urban Economics 1, 346–365.

Herrera, J.C., Work, D.B., Herring, R., Ban, X.J., Jacobson, Q., Bayen, A.M., 2010. Evaluation of traffic data obtained via gps-enabled mobile phones: The mobile century field experiment. Transportation Research Part C: Emerging Technologies 18, 568–583.

Hizir, A.E., 2006. Using emission functions in mathematical programming models for ustainable urban transportation: an application in bilevel optimization. Master's thesis. Sabanci University. Istanbul, Turkey.

Jin, W.L., 2015. Continuous formulations and analytical properties of the link transmission model. Transportation Research Part B 74, 88–103.

Johansson, O., 1997. Optimal road-pricing: Simultaneous treatment of time losses, increased fuel consumption, and emissions. Transportation Research Part D 2, 77–87.

Joksimovic, D., Bliemer, M., Bovy, P., 2005. Optimal toll design problem in dynamic traffic networks with joint route and departure time choice. Transportation Research Record: Journal of the Transportation Research Board 1923, 61–72.

Kickhöfer, B., Kern, J., 2015. Pricing local emission exposure of road traffic: An agent-based approach. Transportation Research Part D: Transport and Environment 37, 14 – 28.

Kickhöfer, B., Nagel, K., 2016. Towards high-resolution first-best air pollution tolls - an evaluation of regulatory policies and a discussion on long-term user reactions. Networks and Spatial Economics, 16, 175–198.

LeBlanc, L.J., Morlok, E.K., Pierskalla, W.P., 1975. An efficient approach to solving the road network equilibrium traffic assignment problem. Transportation Research 9(5), 309–318.

Laih, C.H., 1994. Queueing at a bottleneck with single and multi-step tolls. Transportation Research Part A 28, 197–208.

Li, Z.C., Lam, W.H.K., Wong, S.C., Sumalee, A., 2012. Environmentally sustainable toll design for congested road networks with uncertain demand. International Journal of Sustainable Transportation 6, 127–155.

Lin, D., Unnikrishnan, A., Waller, T., 2011. A dual variable approximation based heuristic for dynamic congestion pricing. Networks and Spatial Economics 11, 271–293.

Liu, L., Boyce, D.E., 2002. Variational inequality formulation of the system-optimal travel choice problem and efficient congestion tolls for a general transportation network with multiple time periods. Regional Science and Urban Economics 32, 627–650.

Lo, H.K., Szeto, W., 2005. Road pricing modeling for hyper-congestion. Transportation Research Part A 39, 705–722.

Lu, C.C., Mahmassani, H., 2011. Modeling heterogeneous network user route and departure time responses to dynamic pricing. Transportation Research Part C 19, 320–337.

Lu, C.C., Mahmassani, H., Zhou, X., 2008. A bi-criterion dynamic user equilibrium traffic assignment model and solution algorithm for evaluating dynamic road pricing strategies. Transportation Research Part C 16, 371–389.

Ma, R., Ban, X., Pang, J., 2015a. A link-based dynamic complementarity system formulation for continuous-time dynamic user equilibria with queue spillbacks. Transportation Science, under review.

Ma, R., Ban, X., Pang, J.S., 2014. Continuous-time dynamic system optimum for single-destination traffic networks with queue spillbacks. Transportation Research Part B 68, 98–122.

Ma, R., Ban, X.J., Pang, J.S., Liu, H.X., 2015b. Submission to the DTA2012 special issue: Approximating time delays in solving continuous-time dynamic user equilibria. Networks and Spatial Economics 15, 443–463.

Mahmassani, H., Zhou, X., Lu, C.C., 2005. Toll pricing and heterogeneous users: approximation algorithms for finding bi-criterion time-dependent efficient paths in large-scale traffic networks. Transportation Research Record: Journal of Transportation Research Board 1923, 28–36.

Miandoabchi, E., Daneshzand, F., R., Z.F., Szeto, W.Y., 2015. Time-dependent discrete road network design. Journal of the Operational Research Society 66, 894–913.

Michalaka, D.and Yin, Y., Hale, D., 2013. Simulating high-occupancy toll lane operations. Transportation Research Record 2396, 124–132.

Netherlands Ministry of Housing, Spatial Planning, and the Environment, 2004. Traffic emissions policy document: achieving sustainability through cleaner, more efficient and quieter vehicles, and climate-neutral fuels. Technical Report. VROM. Rijnstraat 8, 2515 XP, The Hague.

Newell, G., 1993. A simplified theory of kinematic waves in highway traffic, Part I: General theory, Part II: Queuing at freeway bottlenecks, Part III: Multi-destination flows. Transportation Research Part B 27, 281–313.

Nie, Y., Yin, Y., 2013. Managing rush hour travel choices with tradable credit scheme. Transportation Research Part B 50, 1–19.

Osorio, C., Flotterod, G., Bierlaire, M., 2011. Dynamic network loading: a stochastic differentiable model that derives link state distributions. Transportation Research Part B 45, 1410–1423.

Patriksson, M., Rockafellar, R.T., 2002. A mathematical model and descent algorithm for bilevel traffic management. Transportation Science 36, 271–291.

Piccoli, B., Han, K., Friesz, T., Yao, T., Tang, J., 2015. Second order models and traffic data from mobile sensors. Transportation Research Part C: Emerging Technologies 52, 32–56.

Pigou, A.C., 1920. Wealth and welfare. Macmillan, London.

Qiu, Y., Chen, S., 2007. Bi-level programming for continuous network design of comprehensive transportation system based on external optimization, in: Proceedings of 2007 IEEE International Conference on Grey Systems and Intelligent Services, China.

Rilett, L.R., Benedek, C.M., 1994. Traffic assignment under environmental and equity objective. Transportation Research Record 1443, 92–99.

Scora, G., Barth, N., 2006. Comprehensive Modal Emission Model (CMEM), version 3.01. Technical Report. University of California, Riverside, Center for Environmental Research and Technology.

Sharma, S., Mishra, S., 2011. Optimal emission pricing models for containing carbon footprints due to vehicular pollution in a city network [CD], in: Proceedings of the 90th Transportation Research Board Annual Meeting. Washington, D.C.

Shen, W., Nie, Y., Zhang, M., 2007. On path marginal cost analysis and its relationship to dynamic system-optimal traffic assignment, in: Transportation and Traffic Theory, pp. 327–360.

Shen, W., Zhang, H., 2009. On the morning commute problem in a corridor network with multiple bottlenecks: Its system-optimal traffic flow patterns and the realizing tolling scheme. Transportation Research Part B 43, 267–284.

Shen, W., Zhang, H., 2014. System optimal dynamic traffic assignment: Properties and solution procedures in the case of a many-to-one network. Transportation Research Part B 65, 1–17.

Suwansirikul, C., Friesz, T.L., 1987. Equilibrium decomposed optimization: a heuristic for the continuous equilibrium network design problem. Transportation Science 21, 254–263.

Szeto, W., Jiang, Y., Wang, D., Sumalee, A., 2015. A sustainable road network design problem with land use transportation interaction over time. Networks and Spatial Economics 15, 791–822.

Szeto, W., Wang, Y., Wong, S., 2014. The chemical reaction optimization approach to solving the environmentally sustainable network design problem. Computer-Aided Civil and Infrastructure Engineering 29, 140–158.

Szeto, W.Y., Jaber, X., Wong, S.C., 2012. Road network equilibrium approaches to environmental sustainability. Transport Reviews 32, 491–518.

Vichrey, W., 1969. Congestion theory and transport investment. The American Economic Review 59, 251–260.

Wallace, C., Courage, K., Hadi, M., Gan, A., 1998. TRANSYT-7F user's guide. University of Florida. Gainesville.

Walters, A.A., 1961. The theory and measurement of private and social cost of highway congestion. Econometrica 29, 676–699.

Wie, B., 2007. Dynamic stackelberg equilibrium congestion pricing. Transportation Research Part C 15, 154–174.

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

Xu, X., Chen, A., Cheng, L., 2015. Reformulating environmentally constrained traffic equilibrium via a smooth gap function. International Journal of Sustainable Transportation 9, 419–430.

Yang, H., Bell, M.G., 1997. Traffic restraint, road pricing and network equilibrium. Transportation Research Part B: Methodological 31, 303–314.

Yang, H., Huang, H., 1997. Analysis of the time-varying pricing of a bottleneck with elastic demand using optimal control theory. Transportation Research Part B 31, 425–440.

Yang, H., Huang, H.J., 2005. Mathematical and Economic Theory of Road Pricing. Emerald, Bingley.

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

Yang, H., Wang, X., 2011. Managing network mobility with tradable credits. Transportation Research Part B 45, 580–594.

Yin, Y., Lawphongpanich, S., 2006. Internalizing emission externality on road networks. Transportation Research Part D: Transport and Environment 11, 292–301.

Yin, Y., Lu, H., 1999. Traffic equilibrium problems with environmental concerns. Journal of the Eastern Asia Society for Transportation Studies 3, 195–206.

Yperman, I., 2007. The link transmission model for dynamic network loading. (Doctoral dissertation). Retrieved from `https://lirias.kuleuven.be/bitstream/1979/946/2/phd-final-all.pdf`.

Zhang, K., Mahmassani, H., Lu, C.C., 2013. Dynamic pricing, heterogeneous users and perception error: Probit-based bi-criterion dynamic stochastic user equilibrium assignment. Transportation Research Part C 27, 189–204.

Zheng, N., Waraich, R.A., Axhausen, K.W., Geroliminis, N., 2012. A dynamic cordon pricing scheme combining the macroscopic

fundamental diagram and an agent-based traffic model. Transportation Research Part A 46, 1291–1303.
Zhong, R., Sumalee, A., Maruyama, T., 2012. Dynamic marginal cost, access control, and pollution charge: a comparison of bottleneck and whole link models. Journal of Advanced Transportation 46, 191–221.
Zhu, F., Ukkusuri, S.V., 2015. A reinforcement learning approach for distance-based dynamic tolling in the stochastic network environment. Journal of Advanced Transportation 49, 247–266.