Dynamic User Equilibrium based on a Hydrodynamic Model[☆]

Terry L. Friesz[a,*], Ke Han[b], Pedro A. Neto[a], Amir Meimand[a], Tao Yao[a]

[a]*Department of Industrial and Manufacturing Engineering, Pennsylvania State University, PA 16802, USA.*
[b]*Department of Mathematics, Pennsylvania State University, PA 16802, USA.*

Abstract

In this paper we present a continuous-time network loading procedure based on Lighthill-Whitham-Richards model proposed by Lighthill and Whitham (1955); Richards (1956). A system of *differential algebraic equations* (DAEs) is proposed for describing traffic flow propagation, travel delay and route choices. We employ a novel numerical apparatus to reformulate the scalar conservation law as a flow-based PDE, which is then solved semi-analytically with Lax-Hopf formula. This approach allows for an efficient computational scheme for large-scale applications. We will embed this network loading procedure into the *dynamic user equilibrium* (DUE) model proposed by Friesz et al. (1993). The DUE is solved as a *differential variational inequality* (DVI) and with the fixed-point algorithm. Several numerical examples of DUE on networks of varying sizes will be presented, including the Sioux Falls network with a significant number of paths and origin-destination pairs.

The DUE model presented in this article can be formulated as a *variational inequality* (VI) as reported in Friesz et al. (1993). We will present the *Kuhn-Tucker* (KT) conditions for the VI, which is systematically formulated as a linear system. In order to solve the system with reasonable time and memory usage, we present a decomposition of the linear system that allows for the efficient computation of the dual variables. The numerical solutions of DUE obtained from the fixed-point iterations will be tested against the KT condition and validated to be in fact the solutions to the VIs.

Keywords:
LWR model, dynamic user equilibrium, dynamic network loading, Lax-Hopf formula

1. Introduction

Dynamic traffic assignment (DTA) models determine departure rates, departure times and route choices over a given planning horizon. It seeks to describe the dynamic evolution of traffic in networks in a fashion consistent with the fundamental notions of traffic flow and travel demand. This paper is concerned with a specific type of dynamic traffic assignment known as the *dynamic user equilibrium* (DUE) for which generalized travel cost is identical for route and departure time choices associated with any given origin-destination (O-D) pair.

[☆]This work is partially supported by NSF through grant EFRI-1024707, "A theory of complex transportation network design", and by the NSF Graduate Research Fellowship under Grant No. DGE-0750756.

[*]Corresponding author
Email addresses: tfriesz@psu.edu (Terry L. Friesz), kxh323@psu.edu (Ke Han), pan144@psu.edu (Pedro A. Neto), axm512@psu.edu (Amir Meimand), tyy1@engr.psu.edu (Tao Yao)

1.1. *Dynamic user equilibrium (DUE): a brief review*

DUE is the primary form of DTA that this paper will focus on. The last two decades have seen many efforts to develop a theoretically sound formulation of dynamic network user equilibrium that is also in a canonical form acceptable to scholars and practitioners alike. The DUE models tend to be comprised of four essential sub-problems:

1. a model of path delay;
2. flow dynamics;
3. flow propagation constraints; and
4. a route and departure-time choice model.

Furthermore, there are two major components within the analytical DUE: (a) the mathematical expression of Nash-like equilibrium conditions; and (b) a network performance model which is an embedded *dynamic network loading* (DNL) problem. In the DNL procedure arc-specific volumes, arc-specific exit rates and experienced path delay are determined when departure rates are known for each path, see Friesz et al. (2011). It is the embedded dynamic network loading (DNL) problem that makes DUE so difficult. An unfortunate and yet popular misconception is that devising a separate algorithm for the DNL problem means one has chosen a sequential approach to DUE, whereby one first solves the DNL problem then solves the DUE problem, without requiring or attaining consistency. Such a perspective is not even remotely correct. Instead, the DNL process should be understood as the embodiment of the state operator introduced in Friesz et al. (2011) and Friesz (2010) and not as some sort of approximation.[1]

Some early analytical DUE models were greatly influenced by the dynamic system optimal models presented in Merchant and Nemhauser (1978a) and Merchant and Nemhauser (1978b) as well as a desire to find an equivalent optimization problem. In Friesz et al. (1989), and later in Ran et al. (1993), a Beckmann-type objective function is employed to create an optimal control problem whose solutions are of the DUE type. Friesz et al. (1993) showed a variational inequality may be used to represent dynamic user equilibrium when there are both route and departure time decisions. Following suit, Ran and Boyce (1996) suggested other *variational inequality* (VI) formulations of DUE with somewhat unusual flow propagation constraints that are meant to avoid the imposition of state constraints while also assuring physically realistic flows; unfortunately their constraints do not reflect the contraction and expansion of vehicle platoons and are intrinsically inconsistent with other aspects of their model as pointed out by Friesz et al. (2011). In contrast Friesz et al. (2011) utilized flow propagation constraints that involved both state and control variables as well as state-dependent time shifts to account for expansion and contraction of platoons while also maintaining model consistency. A dual-time-scale formulation of DUE with demand evolution was also presented in Friesz et al. (2011) where a *differential variational inequality* (DVI) formulation was used to model dynamic user equilibrium. The authors implemented a fixed-point algorithm to solve the DVI. The DNL phase of DUE is expressed in Friesz et al. (2011) as a system of *differential algebraic equations* (DAE) and approximated as a system

[1]Note that, by referring to the network loading procedure, we are neither employing nor suggesting a sequential approach to the study and computation of DUE. Rather a subset of the equations and inequalities comprising a complete DUE model may be grouped in a way that identifies a traffic assignment subproblem and a network loading subproblem. Such a grouping and choice of names is merely a matter of convenient language that avoids repetitive reference to the same mathematical expressions. Use of such language does not alter the need to solve both the assignment and loading problems consistently and, thus, simultaneously. A careful reading of the mathematical presentation made in subsequent sections makes this assertion quite clear.

of *ordinary differential equations* (ODEs). Moreover, Friesz et al. (2011) also showed the compatibility of the proposed DUE formulation with the CTM proposed in Daganzo (1994, 1995). The papers Wu et al. (1998) and Xu et al. (1999) developed algorithms for the model introduced in Friesz et al. (1993) based upon the gradient projection method without any proof of convergence. The CTM was employed by Lo and Szeto (2002) to create link dynamics as well as a path travel time extraction procedure. Their procedure allows the CTM to subsume the role of the effective path delay operator as originally articulated by Friesz et al. (1993).

Zhu and Marcotte (2000) showed that when departure rates have uniform upper bounds, a solution exists for the continuous-time user equilibrium with route choices. Recently in Han et al. (submitted for publication), the existence of a simultaneous route-and-departure choice DUE has been established without assuming *a priori* bounds on the path flows. In that paper, the authors used the *generalized Vickrey's model* developed in Han et al. (in press,i) for the embedded DNL sub-problem.

1.2. Dynamic traffic flow models

An essential component of many DTA models is the DNL procedure, which aims at describing and predicting the time evolution of system states by introducing dynamics to the traffic flows throughout the network. The accurate and efficient computation of the DNL sub-problem is crucial to the overall performance of the DTA model. The primary purpose of the DNL procedure within the DUE model is to numerically evaluate the so-called effective delay operator introduced by Friesz et al. (1993), which is not available in closed-form. The DNL procedure determines path-specific travel times given a set of known routing and departure time choices. The outcome of the DNL sub-problem relies heavily on the type of link model and junction model chosen. In this section, we will review a few macroscopic traffic flow models most commonly used in the current literature, leading up to the Lighthill-Whitham-Richards (LWR) model of Lighthill and Whitham (1955) and Richards (1956) that we will consider in this paper.

1.2.1. Macroscopic link models

We start our discussion with single link loading. Depending on the ways of specifying flow propagation, we distinguish between *delay-function models* and the *exit-flow function models*. The former is described and solved based on an explicit travel delay function given exogenous parameters that are typically estimated by fitting functions to real-world data. One example is the *link delay model* (LDM); see Friesz et al. (1993), Wu et al. (1998) and Xu et al. (1999). In contrast, exit-flow function models are based on explicitly modeling the underlying flow dynamics as demonstrated by the *M-N model* of Merchant and Nemhauser (1978a,b), the *cell transmission model* (CTM) of Daganzo (1994, 1995) and *Vickrey's model* (VM) of Vickrey (1969).

Among these link models mentioned above, the CTM is the most widely used model in the realm of DNL primarily because: (1) it is based on simple rules for the propagation of traffic states over discrete cells; (2) it is easy to extend to a network; and (3) it captures several key phenomena of the real-world traffic such as shock waves and spillback. It has been shown in Daganzo (1994) that the CTM is the discrete version of LWR model on a single arc, given the trapezoidal density-flow relation. One limitation of the CTM lies in the computational inefficiency as the algorithm stores and updates traffic quantities in a two-dimensional cell. In Nie and Zhang (2005), where a comparative study of various link models was conducted, the CTM was shown to be computationally demanding in terms of solution time and memory consumption.

1.2.2. LWR network model

The LWR model is another macroscopic link model that describes traffic dynamics in terms of the formation, propagation and interaction of kinematic waves. It has received increased attention in the field of traffic flow theory in the past several decades due to its capability of capturing key features of vehicular traffic such as shock waves and spillback. The primary mathematical form of LWR model is a *partial differential equation* (PDE) describing the temporal-spatial evolution of average density and flow. The PDE described in equation (1.1) is based on conservation of vehicles and an explicit density-flow relation known as the fundamental diagram.

$$\frac{\partial}{\partial t}\rho(t,\, x) + \frac{\partial}{\partial x} f(\rho(t,\, x)) \;=\; 0 \tag{1.1}$$

Where $\rho(t,\, x)$ denotes the local vehicle density and $f(\rho(t,\, x))$ is the flow. The fundamental diagram $f(\cdot)$ is a continuous and concave function defined on $[0,\, \rho_{jam}]$ where ρ_{jam} is the jam density. The fundamental diagram encodes the speed-density relation and is calibrated using empirical data. Classical mathematical results on the first-order PDEs of the form (1.1) can be found in Bressan (2000). For a detailed discussion of numerical schemes of conservation laws, we refer the readers to Godunov (1959); LeVeque (1992).

The extension of LWR model to network has been studied extensively in the literature, a list of selected references include Coclite et al. (2005); Daganzo (1994, 1995); Garavello and Piccoli (2006); Holden and Risebro (1995); Jin (2010); Jin and Zhang (2003); Lebacque (1996); Lebacque and Khoshyaran (1999). A detailed survey can also be found in Zhang (2001).

The primary difficulty in the LWR-based network loading problem stems from the complex boundary conditions at a junction of arbitrary topology. In one line, Holden and Risebro (1995) were among the first to propose an analytical framework to derive the weak entropy solution of conservation laws at the junction. In particular, they extended the notion of weak solution to a single conservation law to several conservation laws with coupling boundary conditions. In order to isolate a unique solution at the junction, an entropy condition was introduced which amounts to optimizing some abstract objective function. In Coclite et al. (2005), the notion of entropy condition was instantiated by maximizing the flux through the junction over all possible boundary fluxed of links instance to the junction. They solved the Riemann problems at junctions with fixed vehicle turning percentages. However these conditions are insufficient to isolate a unique solution in the case where the number of incoming links exceeds that of outgoing links. To overcome this difficulty the authors introduced a *right-of-way* parameter specifying the priority of incoming vehicles.

In another line, the complex boundary conditions can be relatively easily analyzed using demand-supply functions proposed in Lebacque (1996) and Lebacque and Khoshyaran (1999). In this framework, local traffic demand (sending flow) and supply (receiving flow) are defined as functions of traffic states on the links immediately adjacent to the junction. Depending on the topology of the junction and specific traffic controls and policies, the boundary fluxes through the junction can be expressed in various ways in terms of demand of upstream links and supply of downstream links. The demand-supply functions are often used with CTM. Such a discrete network model is based on simple releasing rules and conform to physical intuition. Therefore, it is a popular choice for numerically computing traffic flows through networks.

There are primarily three classes of numerical schemes for PDE (1.1), given complex boundary conditions including initial, upstream boundary, downstream boundary and even internal boundary conditions. The three classes are:

1. finite difference schemes, e.g. Daganzo (1994); Godunov (1959); LeVeque (1992);
2. wave-front tracking method, e.g. Bressan (2000); Dafermos (1972, 2010); Garavello and Piccoli (2006); Holden and Risebro (2002); and
3. variational method, e.g. Aubin et al. (2008); Claudel and Bayen (2010a,b); Daganzo (2005, 2006); Evans (2010); Lax (1957, 1973); Le Floch (1988).

The finite difference schemes mainly require a two-dimensional grid on the temporal-spatial domain. Modelers implementing such type of schemes need to ensure that the Courant-Friedrchs-Lewy condition is satisfied in order to guarantee that solutions are numerically stable; see Courant et al. (1928). Furthermore, the solutions tend to exhibit numerical viscosity in which shock waves or contact discontinuities are misrepresented as smooth variations. The wave-front tracking is an event-based computational method that resolves the discontinuities (shocks) exactly. The method was originally proposed in Dafermos (1972) for the study of existence and uniqueness of solution to the initial value problem. The wave-front tracking method is built within a class of piecewise constant boundary conditions and a piecewise affine flux function. It has also been applied to system of conservation laws as well as traffic network models, for example, see Coclite et al. (2005) and Holden and Risebro (1995).

The variational method was initially developed in Lax (1957, 1973) (Lax-Oleinik formula) for the scalar conservation laws, using the method of characteristics. This method was also developed for a Hamilton-Jacobi equation using calculus of variations (Lax-Hopf formula). The Lax-type formula expresses the weak entropy solution (viscosity solution) to the conservation laws (H-J equations) semi-analytically as the solution of a minimization problem. Generalizations of this method are made in Aubin et al. (2008); Daganzo (2005, 2006); Le Floch (1988). The application of this method to fluid-based models such as traffic flows were recently investigated in Bressan and Han (2011a,b); Claudel and Bayen (2010a,b); Han et al. (submitted for publication). In contrast to finite difference schemes, the variational method has the distinct advantage of not requiring a two-dimensional grid in the sense that the solution at one point in space-time does not require intermediate computation on other points in the domain. The variational methods also yields solutions with arguably better precision than finite difference schemes. For example, discontinuities in the solution are more accurately obtained by the variational method, see Claudel and Bayen (2010a,b) and LeVeque (1992).

1.2.3. An LWR-based network loading procedure

This paper presents a new dynamic network loading procedure employing the LWR-PDE (1.1). The procedure is fundamentally based upon the Lax-Hopf formula to solve the PDE. In particular, we will introduce a flow-based conservation law instead of density-based conservation law (1.1). This enables us to transform a boundary value problem into an initial value problem. Such a formulation allows us to apply the Lax-Hopf formula directly and obtain a closed-form solution for a single link. Then we will extend the framework to a general network when the vehicle paths and departure rates are given. In order to find the network delay we will invoke an approach similar to that shown by Newell (1993). That is, the implicit travel time defined via the horizontal difference between cumulative vehicle counts. The novelty of our network loading procedure is the utilization of a flow-based conservation law and a grid-free computational method. With these techniques the upstream boundary value problems for a single link can be solved efficiency and accurately. Furthermore, we will show in Section 2.2 that our approach will work for very general flow profiles in which the cumulative vehicle count may not be Lipschitz continuous or not continuous at all. In these cases, the flows can be unbounded and even distributions. Our procedure has the ability to capture the following several key network traffic phenomena:

1. Queues and delay;
2. Density-velocity relationship;
3. First-in-first-out (FIFO); and
4. Route information.

The analytical solution to the PDE allows us to represent the network loading procedure as a system of *differential algebraic equations* (DAE) instead of *partial differential algebraic equations* (PDAE). Our method also has the key distinction of not relying upon the use of a spatial grid for numerical computation. Therefore, large-scale dynamic traffic assignment network problems can be solved with sufficient accuracy and computational efficiency. Contributions of our work include

1. Derivation of an explicit formula for single link loading procedure based on the LWR model with boundary value conditions. To do this, we propose a flow-based scalar conservation law and transform a boundary value problem into an initial value problem. We then apply the Lax-Hopf formula. The resulting explicit formula is proved to hold for very general inflow profiles such as unboundedness and distributions.
2. Formulation of a DAE system for network loading procedure, based on the model mentioned in 1. In particular, the DAE system encodes merging and diverging junction models, route information and travel delay.
3. Implementation the devised DNL procedure within the context of DUE while demonstrating the ability of the DAE system to handle very complex network topology and routing information. We demonstrate the computational efficiency of our proposed procedure through a numerical example of the Sioux Falls network with significant number of origin-destination pairs and paths. The numerical performance will be compared with other DNL procedures such as CTM and Link Delay Model.
4. Articulation of the Kuhn-Tucker conditions for the variational inequality formulation of DUE proposed in Friesz et al. (2011). Our numerical solutions of DUE will be verified against the KT necessary conditions. This is the first effort in current literature to validate the solutions of the DUE using KT conditions.

1.3. Organization

The rest of the article is organized as follows. In Section 2, we discuss in detail the single link loading problem using Lax-Hopf formula. We then extend the framework to a general network, with specific discussion on merging and diverging of vehicle flows as well as formation and dissipation of traffic queues. Furthermore, a DAE system formulation of the network loading procedure will be presented at the end of Section 2. In section 3, we will briefly review various formulations of the simultaneous route-and-departure choice DUE. For completeness of our presentation we will include the fixed-point algorithm for computing the DUE formulated as a *differential variational inequality*. The Kuhn-Tucker condition for the finite dimensional VI is presented and will be used later to verify that the DUE solutions we obtain is indeed the solution to the VIs. Section 4 discusses several numerical examples of the DUE on networks of varying sizes, including the well-known Sioux Falls network.

2. LWR model and dynamic network loading

The aim of this section is to systematically derive the dynamic network loading procedure employing the LWR model within the DUE framework presented in Friesz et al. (2011). We will first introduce the well-known LWR-PDE of Lighthill and Whitham (1955) and Richards (1956) for the dynamics of a single arc, then transform that PDE into a flow-based conservation law under the assumption of no spillback. This novel approach enables us to apply the Lax-Hopf formula and represent the solution of a single arc in closed form. Furthermore, we will discuss how the phenomenon of bottlenecking and queuing can be captured by the Lax-Hopf formula. We discuss an implicit travel time function based on cumulative vehicle counts and extend the single link model to a general network incorporating the different routing and departure time choices. The use of a solution representation independent of spatial variables allows us to remove the PDE constraints from the PDAE system and thus derive a DAE system in its place.

2.1. Generic network model

We first start by introducing the network model and briefly mention the dynamics on each link and interactions of links at junctions. The detailed discussion of network loading procedure will be postponed to Section 2.2 - 2.6. The notion of a vehicular network is made precise using a directed graph.

Definition 2.1. (Network definition) *A traffic network is a directed graph $(\mathcal{A}, \mathcal{V})$ where each arc $e \in \mathcal{A}$ corresponds to a homogeneous road, each vertex $v \in \mathcal{V}$ represents the junction. In addition, we assume*

1. *For each arc $e \in \mathcal{A}$, define the free-flow speed v_0^e, the jam density ρ_{jam}^e, the flow capacity M^e and the length L^e. Each arc is mapped onto a spatial interval $[a^e, b^e]$ with $b^e - a^e = L^e$;*
2. *Each road $e \in \mathcal{A}$ possesses a queue with zero physical size (point queue), which is located at the vertex in front of the road, i.e. at $x = a^e$;*
3. *Denote for each $v \in \mathcal{V}$ the set of incoming (upstream) arcs by $\mathcal{I}^v$, the set of outgoing (downstream) arcs by $\mathcal{O}^v$. In the case $|\mathcal{O}^v| > 1$, we call node v dispersive and use the allocation rates (turning percentage) $A^{v,e}(t)$ to denote the proportion of total flow coming from $\hat{e} \in \mathcal{I}^v$ that goes to the queue of downstream arc e.*

For each $e \in \mathcal{A}$, let $\rho^e(t, x)$ be the vehicle density at time t and location $x \in [a^e, b^e]$, $q^e(t)$ denotes the queue volume. The well-known LWR model describes the evolution of vehicle density using a scalar conservation law

$$\partial_t \rho^e + \partial_x(\rho^e v^e(\rho^e)) = 0 \tag{2.2}$$

where the dependence on (t, x) is dropped for brevity. The function $v^e(\cdot) : [0, \rho_{jam}^e] \to [0, v_0^e]$ expresses the vehicle velocity as a function of local density and defines the fundamental diagram $f^e(\rho^e) \doteq \rho^e v^e(\rho^e)$. Throughout this article, we assume the following for the fundamental diagram

(A1) The fundamental diagram $f^e(\cdot) : [0, \rho_{jam}^e] \to [0, M^e]$ is continuous and concave, with a unique $\rho^{e,*}$ such that $f^e(\rho^{e,*}) = M^e$

Now the dynamics of the network can be described by a system of coupling PDEs and ODEs, i.e. for $e \in \mathcal{O}^v$,

$$\begin{cases} \partial_t \rho^e(t, x) + \partial_x f^e(\rho^e(t, x)) = 0, & (t, x) \in [t_0, t_f] \times [a^e, b^e] \\ \rho^e(0, x) = 0, & x \in [a^e, b^e] \end{cases} \tag{2.3}$$

$$\frac{d}{dt}q^e(t) \;=\; A^{v,e}(t)\sum_{\hat{e}\in\mathcal{I}^v} f^{\hat{e}}(\rho^{\hat{e}}(t,\,b^{\hat{e}})) - f^e(\rho^e(t,\,a^e)) \tag{2.4}$$

$$f^e(\rho^e(t,\,a^e)) \;=\; \begin{cases} \min\left\{A^{v,e}(t)\sum_{\hat{e}\in\mathcal{I}^v} f^{\hat{e}}(\rho^{\hat{e}}(t,\,b^{\hat{e}})),\; M^e\right\}, & q^e(t) \;=\; 0 \\ M^e, & q^e(t) \;>\; 0 \end{cases} \tag{2.5}$$

where $f^e(\rho^e(t,\,a^e))$ is the service rate, i.e., rate at which cars in the queue enters the road in front, and the rule for the service rate is described as follows: if the queue is empty, the service rate will be the portion of the demand allowed by the capacity; if the queue is non-empty, the service rate remains maximum, see (2.5).

The above PDAE system describes the flow/density propagation on each road and at junctions, but it is incomplete in the sense that a travel time operator needs to be defined and the turning percentage $A^{v,e}(t)$ needs to be extracted from the routing information. In the following section, we will insert details to this PDAE system (2.3)-(2.5) and in fact replace it with a DAE system.

2.2. A flow-based conservation law

In consistency with the assumption of dynamic network loading, we assume the network is initially empty. Consider a homogeneous arc $e \in \mathcal{A}$ in the network characterized by the free-flow speed v_0, the jam density ρ_{jam}, the flow capacity M and the length L. The superscript e is dropped here for brevity. Denote the right hand side of (2.5) by $\overline{u}(t)$, the conservation law (2.3) is in fact a problem with initial and boundary values:

$$\begin{cases} \partial_t\rho(t,\,x) + \partial_x f(\rho(t,\,x)) \;=\; 0 \qquad (t,\,x) \in [t_0,\,t_f] \times [a,\,b]; \\ \rho(t_0,\,x) \;=\; 0; \\ f(\rho(t,\,a)) \;=\; \overline{u}(t). \end{cases} \tag{2.6}$$

In view of the boundary condition above and in (2.4)-(2.5), which are all prescribed in terms of flow, it would be convenient to modify the density-based conservation law and derive a new one whose primary variable is the flow. To this end, denote the vehicle flow by $u(t,\,x)$ and consider the function $g(\cdot)$ defined by

$$\phi(u) \;=\; \inf_{\rho\in[0,\,\rho_{jam}]}\{f(\rho) = u\} \tag{2.7}$$

In other words, $\phi(\cdot)$ is the inverse of $f(\cdot)$ on $[0,\,\rho^*]$, where ρ^* is the unique density such that $f(\rho^*) = M$. By assumption **(A1)**, $\phi(\cdot)$ is continuous, convex and strictly increasing, see Figure 1.

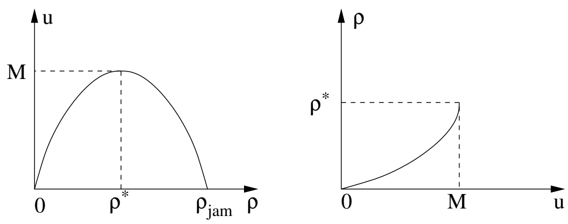


Figure 1: Left: an example of the fundamental diagram $f(\cdot)$. Right: the function $\phi(\cdot)$ defined by (2.7)

Remark 2.2. *By choosing the part of the fundamental diagram with only positive wave speed, we are making the following important modeling assumption; we do not consider backward propagation of traffic densities nor spillback. This assumption can be reflected also in equations (2.3)-(2.5) where the queues (complete jam) will not affect previous arcs.*

In order to apply the classical Lax-Hopf formula, it is convenient to transform the initial/boundary value problem (2.6) into a Cauchy problem, i.e. an initial value problem. To do so, we will next derive a flow-based conservation law and rigorously show its equivalence to (2.6). Let us introduce the notion of a weak solution and entropy admissible solution. For more details on the theory of hyperbolic PDEs, we refer the readers to Bressan (2000).

Definition 2.3. (Weak solution for (2.6)) *Assume $\overline{u} \in \mathcal{L}^1([t_0, t_f])$. A function $\rho : [t_0, t_f] \times [a, b] \to [0, \rho_{jam}]$ is a weak solution to the initial/boundary value problem (2.6) if for every continuously differentiable test function $v(\cdot)$ with compact support contained in $[t_0, t_f] \times [a, b]$, the following holds*

$$\int_{t_0}^{t_f} \int_a^b \{\rho \cdot v_t + f(\rho) \cdot v_x\}\, dxdt + \int_{t_0}^{t_f} \overline{u}(t) \cdot v(t, a)\, dt = 0 \tag{2.8}$$

Definition 2.4. (Entropy admissible solution) *A weak solution $\rho = \rho(t, x)$ to the scalar conservation law (2.6) satisfies the Kruzkov entropy admissibility condition if*

$$\int_{t_0}^{t_f} \int_a^b \{|\rho - k|\, v_t + sign(\rho - k)\left(f(\rho) - f(k)\right) v_x\}\, dxdt \geq 0 \tag{2.9}$$

for every $k \in [0, \rho_{jam}]$ and every continuously differentiable $v \geq 0$ with compact support in $[t_0, t_f] \times [a, b]$.

The boundary value problem (2.6) can now be equivalently written as a conservation law in the variable $u = f(\rho)$:

Proposition 2.5. (Flow-based conservation law) *Given any concave function $f(\cdot) : [0, \rho_{jam}] \to [0, M]$ and function $\overline{u}(\cdot) \in \mathcal{L}^1([t_0, t_f]) : [t_0, t_f] \to [0, M]$, define $\phi(\cdot) : [0, M] \to [0, \rho^*]$ as in (2.7). Then the boundary value problem (2.6) is equivalent to the following Cauchy problem*

$$\begin{cases} \partial_x u(t, x) + \partial_t \phi(u(t, x)) = 0 & (x, t) \in [a, b] \times [t_0, t_f]; \\ u(t, a) = \overline{u}(t). \end{cases} \tag{2.10}$$

where $u(t, x) = f(\rho(t, x))$.

Remark 2.6. *(2.10) can be viewed as an initial value problem by switching the roles of variables t and x. The flow-based PDE is easy to work with in terms of boundary conditions (2.4)-(2.5), and the initial value problem (2.10) immediately leads to the application of the Lax formula discussed in Evans (2010); Lax (1957).*

Proof. The proof is carried out by showing that $\rho(t, x)$ is the entropy admissible weak solution of (2.6) if and only if $u(t, x) \doteq f(\rho(t, x))$ is the entropy admissible weak solution of (2.10).

1. "only if " part. We begin by arguing that the weak entropy solution $\rho(t, x)$ of (2.6) satisfies $\rho(t, x) \in [0, \rho^*]$, where ρ^* is the unique density such that $f(\rho^*) = M$. This is because the characteristics emitting from the left boundary $x = a$ must have non-negative wave speed in order to influence the interior of the domain $[t_0, t_f] \times [a, b]$. Otherwise, if there exists $(\hat{t}, \hat{x}) \in (t_0, t_f] \times (a, b)$ which is connected to the left boundary $[t_0, t_f] \times \{a\}$ by a characteristic line with negative speed, then $\rho(\hat{t}, \hat{x})$ is influenced by the boundary condition at a time $> \hat{t}$, this contradicts the causality principle.

Let $\rho(\cdot, \cdot)$ be the weak entropy solution satisfying (2.8)-(2.9). By previous argument, there holds $u = f(\rho)$, $\rho = \phi(u)$. Then we readily deduce

$$\begin{aligned} 0 &= \int_{t_0}^{t_f} \int_a^b \{\rho \cdot v_t + f(\rho) \cdot v_x\}\, dxdt + \int_{t_0}^{t_f} \overline{u}(t) \cdot v(t, a)\, dt \\ &= \int_{t_0}^{t_f} \int_a^b \{\phi(u) \cdot v_t + u \cdot v_x\}\, dxdt + \int_{t_0}^{t_f} \overline{u}(t) \cdot v(t, a)\, dt \end{aligned}$$

for any continuously differentiable v with compact support. This shows that $u(t, x)$ is the weak solution to the initial value problem (2.10). In addition, let us fix arbitrary $v(t, x) \geq 0$ which is continuously differentiable and compactly supported, for every $s \in [0, M]$, denote $k = \phi(s) \in [0, \rho^*]$. By the entropy condition (2.9),

$$\begin{aligned} 0 &\leq \int_{t_0}^{t_f} \int_a^b \{|\rho - k|\, v_t + \mathrm{sign}(\rho - k)\,(f(\rho) - f(k))\, v_x\}\, dxdt \\ &= \int_{t_0}^{t_f} \int_a^b \{|\phi(u) - \phi(s)|\, v_t + \mathrm{sign}(\phi(u) - \phi(s))\,(u - s)\, v_x\}\, dxdt \\ &= \int_{t_0}^{t_f} \int_a^b \{|u - s|\, v_x + \mathrm{sign}(u - s)\,(\phi(u) - \phi(s))\, v_t\}\, dxdt \end{aligned}$$

this implies $u(\cdot, \cdot)$ is the entropy admissible solution to (2.10). Notice that the last equality above is by observing that $\phi(\cdot)$ is strictly increasing, thus

$$|\phi(u) - \phi(s)| = \mathrm{sign}(u - s)(\phi(u) - \phi(s)), \qquad |u - s| = \mathrm{sign}(\phi(u) - \phi(s))\,(u - s)$$

2. "if" part. The converse is completely similar, the key point is that $\rho(t, x) \leq \rho^*$, therefore $f(\rho) = u$, $\rho = \phi(u)$. The rest of the verification is straightforward. □

It is well-established that a full LWR-based network model requires boundary conditions on both ends of the link, which account for the propagation of upstream/downstream information. In our simplified model, in the absence of influence from downstream, the PDE for the link is constrained by just upstream boundary condition which must admit non-negative wave speeds due to the causality principle.

We remind the reader that so far we've been dealing with the PDE (2.3) with boundary condition $\overline{u}(t)$ given by (2.5). Such a boundary condition depends on the queue $q(t)$ and the upstream flow profiles in a tricky way: (2.5) induces a discontinuous dependence of the ODE (2.4) on the variable $q(t)$. In general, an ODE with discontinuous right hand side is difficult to deal with both theoretically and computationally. For example, the existence of the ODE is not guaranteed. In terms of computation, a finite difference scheme, whether forward or backward, could result in non-physical solutions such as negative queue and negative flow. To overcome the aforementioned difficulties, we propose in the following a Hamilton-Jacobi equation corresponding to the flow-based conservation law (2.10), and a modified Lax-Hopf formula which provides solution to (2.3)-(2.5).

2.3. Hamilton-Jacobi equation and Lax-Hopf formula for a homogeneous link

In this section, we will first introduce the Hamilton-Jacobi equation and the Lax formula for general Cauchy problems, then apply this to the flow-based conservation law (2.10) to obtain a semi-analytical characterization of the dynamics for a homogeneous link.

Initially introduced in Lax (1957, 1973), then extended in Aubin et al. (2008); Bardi and Capuzzo-Dolcetta (1997); Le Floch (1988), and applied to traffic theory in Claudel and Bayen (2010a); Daganzo (2005, 2006), the Lax-Hopf formula provides a new characterization of the solution to the hyperbolic conservation law and Hamilton-Jacobi equation. The Lax formula is derived from the characteristics equations associated with the Hamilton-Jacobi PDE, which arise in the classical calculus of variations and in mechanics, see Evans (2010) for a complete discussion.

Consider the initial value problem for the Hamilton-Jacobi equation

$$\begin{cases}\partial_t \mathcal{N}(t,\,x) + \mathcal{H}(\partial_x \mathcal{N}(t,\,x)) \;=\; 0 & (t,\,x)\in(0,\,\infty)\times\mathbb{R} \\ \mathcal{N}(0,\,x) \;=\; g(x) & x\in\mathbb{R}\end{cases} \tag{2.11}$$

where $N : [0,\,+\infty)\times\mathbb{R}\to\mathbb{R}$ is the unknown, and $\mathcal{H} : \mathbb{R}\to\mathbb{R}$ is the Hamiltonian.

Theorem 2.7. **(Lax-Hopf formula)** *Suppose $\mathcal{H}$ is continuous and convex, $g(\cdot) : \mathbb{R}\to\mathbb{R}$ is Lipschitz continuous, then*

$$\mathcal{N}(t,\,x) \;=\; \inf_{y\in\mathbb{R}}\Big\{t\mathcal{L}\Big(\frac{x-y}{t}\Big)+g(y)\Big\} \tag{2.12}$$

is the unique viscosity solution to the initial-value problem (2.11). Where $\mathcal{L}$ is the Legendre transformation of $\mathcal{H}$:

$$\mathcal{L}(q) \;=\; \inf_p\{\mathcal{H}(p)-qp\} \tag{2.13}$$

Proof. See Evans (2010). □

The Lax-Hopf formula expresses the viscosity solution of the Cauchy problem (2.11) as an optimization problem. In order to transform the flow-based PDE (2.10) into the form (2.11), let us proceed as follows. Denote the cumulative vehicle count $U(t,\,x)$, $\overline{U}(t)$ by

$$U(t,\,x) \;=\; \int_{t_0}^{t} u(s,\,x)\,ds, \qquad \overline{U}(t) \;=\; \int_{t_0}^{t}\overline{u}(s)\,ds \tag{2.14}$$

$U(t,\,x)$ measures the cumulative count of vehicles that have passed the point x by time t, and $\overline{U}(t)$ measures the vehicle count that have entered the link. . In addition, we denote by $Q(t)$ the count of vehicles that have arrived at the entrance of the link (possibly first joining a queue) by time t. In mathematical terms, if the arc is represented by the spatial interval $[a,\,b]$, then

$$Q(t) \;=\; U(t,\,a-), \qquad \overline{U}(t) \;=\; U(t,\,a+)$$

Notice that by definition, $\overline{U}(\cdot)$ must be Lipschitz continuous with constant M, this is because the rate at which cars enter the link is bounded above by the flow capacity, i.e., $\bar{u}(t)\le M$. Regarding $Q(\cdot)$, we assume the following

(**A2**) $Q(\cdot)$ is non-decreasing, left continuous with possibly countably many upward jumps.

Remark 2.8. *In most traffic literature, the function which represent flow is usually assumed to be Lebesgue integrable. Consequently, the cumulative vehicle count – the anti-derivative of flow – is absolutely continuous. Therefore* **(A2)** *generalizes such assumption by allowing the cumulative count to be discontinuous. This corresponds to the situation where a positive amount of cars enter the network at the same time. Although such circumstance is unlikely to happen in reality, it does arise in a mathematical context and demands proper treatment. See Bressan and Han (2011a) for an example where a discontinuity in the cumulative vehicle count is present in the DUE solution in continuous-time.*

According to the rule (2.5), the relation between $Q(\cdot)$ and $\overline{U}(\cdot)$ can be expressed by the following identity

$$\overline{U}(t) \;=\; \inf_{\tau\leq t}\{Q(\tau)+M(t-\tau)\} \tag{2.15}$$

by virtue of (2.15), $\overline{U}(t)\leq Q(t)$ and the difference measure the queue volume; notice that (2.16) is precisely the integral form of (2.4) under the assumption that $q(0)=0$.

$$q(t) \;=\; Q(t)-\overline{U}(t) \tag{2.16}$$

We now introduce the integral form of (2.10), which is following Hamilton-Jacobi equation:

$$\begin{cases}\partial_x U(t,\,x)+\phi\Big(\partial_t\, U(t,\,x)\Big) \;=\; 0 \qquad (x,\,t)\in[a,\,b]\times[t_0,\,t_f];\\[2ex] U(t,\,a) \;=\; \overline{U}(t)\end{cases} \tag{2.17}$$

Note that (2.17) is in the form of a general Cauchy problem for H-J equation in one dimension, i.e. (2.11). The only difference is the notation in which the roles of t and x are switched. With that in mind, we will next proceed to the Lax formula and derive analytical solutions to (2.17). Recall the following Legendre transformation:

Definition 2.9. (Legendre transformation) *The Legendre (convex) transformation of convex function $\phi(\cdot):[0,\,M]\to[0,\,\rho^*]$ is*

$$\psi(u) \;\doteq\; \max_{p}\{pu-\phi(p)\} \tag{2.18}$$

It is easy to verify that the Legendre transform $\psi(\cdot)$ is convex and Lipschitz continuous with Lipschitz constant M. Next, we adapt the Lax-Hopf formula (2.12) to the Cauchy problem (2.17) and readily derive the following

Proposition 2.10. (Lax-Hopf formula) *Given convex function $\phi(\cdot)$ and its Legendre transform $\psi(\cdot)$, the viscosity solution $U(t,\,x)$ to (2.17) satisfies*

$$U(t,\,x) \;=\; \min_{\tau\in\mathbb{R}}\Big\{\overline{U}(\tau)+(x-a)\psi\Big(\frac{t-\tau}{x-a}\Big)\Big\},\qquad x\in[a,\,b] \tag{2.19}$$

In particular, the cumulative count of vehicles exiting the arc is given by

$$U(t,\,b) \;=\; \min_{\tau\in\mathbb{R}}\Big\{\overline{U}(\tau)+L\,\psi\Big(\frac{t-\tau}{L}\Big)\Big\} \tag{2.20}$$

Proof. Adjust the proof in Evans (2010) to the Cauchy problem (2.17). □

Remark 2.11. *The original Lax-Hopf formula used infimum instead of minimum, but since all flows that we consider here are compactly supported in* $[t_0,\, t_f]$*, the infimum can be achieved within this compact set. Therefore from now on, we will use "min" instead of "inf".*

As mentioned before, in a general network, the quantity $Q(t)$ is easy to be recovered from upstream flow profiles. However, one cannot immediately apply formula (2.19) or (2.20) because $\overline{U}(\cdot)$ is not explicitly given. One way to resolve this issue is to utilize identity (2.15), but this would require additional computational effort. Fortunately, the following lemma asserts that this is not necessary and replacing $\overline{U}(\cdot)$ with $Q(\cdot)$ will yield the same solution.

Proposition 2.12. *Given Lipschitz continuous function* $\overline{U}(\cdot)$ *and function* $Q(\cdot)$ *satisfying assumption* **(A2)***, then the solution* $U(t,\, b)$ *satisfies*

$$U(t,\, b) \;=\; \min_{\tau}\Big\{\overline{U}(\tau)+L\psi\Big(\frac{t-\tau}{L}\Big)\Big\} \;=\; \min_{\tau}\Big\{Q(\tau)+L\psi\Big(\frac{t-\tau}{L}\Big)\Big\} \tag{2.21}$$

Proof. Introduce the function

$$\kappa(s) \;\doteq\; -L\psi\Big(\frac{-s}{L}\Big) \tag{2.22}$$

Then one can easily verify that $\kappa(\cdot)$ is a concave and Lipschitz continuous function with Lipschitz constant M. From the formula (2.20) it now follows

$$U(t,\, b) \;=\; \min_{\tau}\Big\{\overline{U}(\tau)-\kappa(\tau-t)\Big\} \tag{2.23}$$

Fix any $t\in[t_0,\, t_f]$, denote

$$\tau_0 \;=\; \inf\Big\{\tau^*:\;\; \overline{U}(\tau^*)-\kappa(\tau^*-t) \;=\; \min_{\tau}\{\overline{U}(\tau)-\kappa(\tau-t)\}\Big\} \tag{2.24}$$

clearly τ_0 exists and satisfies $\overline{U}(\tau_0)-\kappa(\tau_0-t)=\min_\tau\{\overline{U}(\tau)-\kappa(\tau-t)\}$. Next we claim that

$$Q(\tau_0) \;=\; \overline{U}(\tau_0)$$

Otherwise if $Q(\tau_0)>\overline{U}(\tau_0)$, by left-continuity of $Q(\cdot)$, one must have $Q(\tau)>\overline{U}(\tau)$ for a small neighborhood $\tau\in[\tau_0-\delta,\,\tau_0]$. Thus from (2.5) we deduce that $\frac{d}{d\tau}\overline{U}(\tau)=M$. By Lipschitz continuity of $\kappa(\cdot)$, for $\tau\in[\tau_0-\delta,\,\tau_0]$

$$\overline{U}(\tau)-\kappa(\tau-t)-\big(\overline{U}(\tau_0)-\kappa(\tau_0-t)\big) \;\leq\; 0$$

contradicting (2.24). Thereby the claim is substantiated. The claim implies

$$\min_{\tau}\{Q(\tau)-\kappa(\tau-t)\} \;\leq\; Q(\tau_0)-\kappa(\tau_0-t) \;=\; \min_{\tau}\{\overline{U}(\tau)-\kappa(\tau-t)\} \tag{2.25}$$

To show (2.25) with reversed inequality, one only has to notice $Q(\cdot)\geq\overline{U}(\cdot)$. □

Remark 2.13. *Using the notation in the proof, the Lax formula (2.20) asserts that the solution* $U(t,\, b)$ *is the amount by which we can shift upward the graph of* $\kappa(\cdot-t)$ *before hitting the graph of* $\overline{U}(\cdot)$*. What Proposition 2.12 says is that such amount of shift does not change if* $\overline{U}(\cdot)$ *is replaced by* $Q(\cdot)$*. See Figure 2 for a graphical representation.*

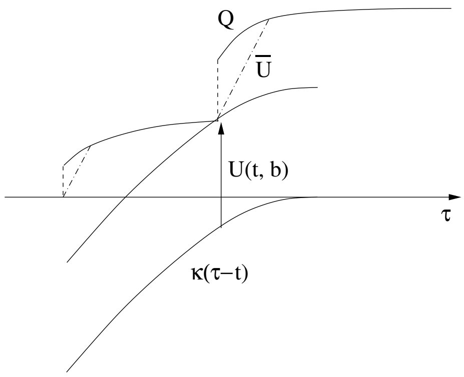


Figure 2: Constructing the profile $t \mapsto U(t, b)$ using the Lax formula. Since κ is Lipschitz continuous with constant M, the amount of shift in (2.21) does not change if $\overline{U}(\cdot)$ is replaced by $Q(\cdot)$.

Remark 2.14. *Although our formulation does not capture between-link congestion, as suggested by (2.3)-(2.5), it does capture within-link congestion. In particular two sources will contribute to the within-link congestion. The first one is the non-affine density-velocity relation, see assumption* **(A1)** *and Figure 1. This means that the vehicle speed will decrease as the local density increases. The second source of congestion is interpreted from Proposition 2.12, which says that when the inflow of a link exceeds the flow capacity vehicles will wait in a non-physical queue (therefore there is no spillback). The waiting time is included in the link traversal time, as we shall see in the next section.*

In this section, we provide a semi-analytical representation of the solution to (2.3)-(2.5), this system of coupling PDEs and ODEs raises several numerical challenges such as the discontinuous dependence on the queue volume for ODE (2.4) and non-physical solution resulting from finite difference discretization. By formulating the system as a Hamilton-Jacobi equation and a modification of the Lax formula, (2.3)-(2.5) is solved by the second identity of (2.21). Next we extend the work to networks and incorporate travel time and route information.

2.4. Link traversal time

An essential component of dynamic traffic assignment models is the travel delay operator. A significant amount of literature has assumed a link travel time expressed explicitly as a function of the state of the link, one example is the *link delay model* (LDM) studied in Friesz et al. (1993, 2011); Wu et al. (1998); Xu et al. (1999). Another approach of computing travel time is via the exit-flow function models. This approach is based on a description of flow dynamics and computation of the exit flow profile. Examples include M-N model of Merchant and Nemhauser (1978a); the CTM of Daganzo (1994, 1995) and Vickrey's model of Vickrey (1969). Notice that neither the M-N model nor the CTM model defines the link traversal time. However, the traversal time can be extracted from the models using flow conservation constraint and the *first-in-first-out* (FIFO) assumption. For example, Newell (1993) proposed a method for computing travel time by measuring the horizontal difference of two cumulative vehicle count curves at the entry and exit of the link. Similar technique has been applied to CTM. Ziliaskopoulos (2000) considered a system-optimal DTA problem in the framework of CTM. The author replaced the travel time function with a cell-delay function in order to obtain a tractable mathematical program.

Since our approach of network loading is based on explicit computation of the exit flow profile, we adapt the approach similar to Newell (1993) by measuring the horizontal difference between the cumulative curves $Q(\cdot)$ and $W(\cdot)$, which represent vehicle count at entrance and exit of a link, respectively. This implicit travel time function is consistent with FIFO and will later be used to construct the turning percentages $A^{v,e}(t)$.

Given the boundary condition $Q(\cdot)$, which measures the count of vehicles that have joined the queue in front of the link, the Lax formula (2.21) uniquely determines the exit profile $W(\cdot)$. The exit time t^E for a car that enters the queue at t^S is given implicitly by

then for a car that enters the queue at t^S, its exit time t^E is given implicitly by

$$Q(t^S) = W(t^E) \tag{2.26}$$

We rewrite (2.26) using a general functional $D(\cdot\,;Q) : t^S \mapsto t^E$, notice that the dependence on $W(\cdot)$ is dropped because $W(\cdot)$ can be uniquely determined by $Q(\cdot)$ via the Lax-Hopf formula. See Figure 3 for a graphical illustration of this operator. Notice that the time spent waiting in the queue is part of the link traversal time.

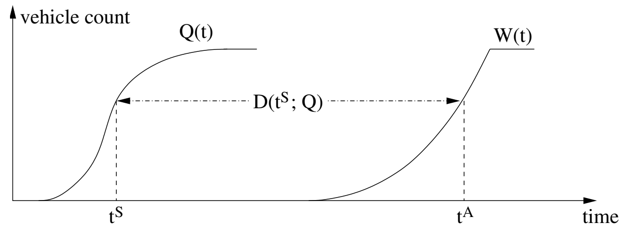


Figure 3: The delay of vehicle departing at t^S is given by the horizontal difference of two cumulative curves $Q(\cdot)$ and $W(\cdot)$.

2.5. *Diverging model*

In this section we consider any node $v \in \mathcal{V}$ with more than one outgoing arcs. We need to determine the time-varying functions $A^{v,e}(t)$ for each $e \in \mathcal{O}^v$ according to route information. Let us first introduce a few notations. Let $\mathcal{P}$ be the set of utilized paths in the network, each path $p \in \mathcal{P}$ is defined as a collection of directed arcs $p = \{e_1, e_2, \ldots, e_{m(p)}\} \subset \mathcal{A}$, where $m(p)$ is the number of arcs traversed by p. For $e \in \mathcal{I}^v$, denote $\mathcal{P}^e = \{p \in \mathcal{P} : e \in p\}$. For $k = 1, 2, \ldots, |\mathcal{P}^e|$ let $q_k^e(t)$ be the rate at which drivers in the k^{th} group arrive at link e. Define functions $\xi_k^e(t)$ according to the following.

$$\xi_k^e(t) \sum_{l=1}^{|\mathcal{P}^e|} q_l^e(t) = q_k^e(t), \qquad k = 1, 2, \ldots, |\mathcal{P}^e|. \tag{2.27}$$

In other words, $\xi_k^e(t)$ represents the percentage of the k^{th} group among all groups that use arc e at time of entry t. Similarly, we define $\eta_k^e(t)$ to be the percentage of the k^{th} group at time of exit t. For every group of drivers k using the link e, the first-in-first-out principle implies that

$$\eta_k^e(t + D(t\,;Q^e)) = \xi_k^e(t), \qquad k = 1, 2, \ldots, |\mathcal{P}^e|. \tag{2.28}$$

With obvious meaning of notations (2.28) can be re-written as

$$
\begin{cases}
Q^e(t) \ = \ \displaystyle\sum_{k=1}^{|\mathcal{P}^e|} Q_k^e(t), \quad \dfrac{d}{dt} Q_k^e(t) \ = \ q_k^e(t); \\
\\
\eta_k^e(t + D(t,\ ; Q^e)) \ = \ \dfrac{q_k^e(t)}{\sum_{l=1}^{|\mathcal{P}^e|} q_l^e(t)}
\end{cases}
\tag{2.29}
$$

(2.29) says that the turning percentages at a dispersive junction can be completely determined by $Q_k^e(t),\ k = 1,\ 2,\ \ldots, |\mathcal{P}^e|$.

2.6. The DAE system

Summing up previous discussion, we present a DAE system for dynamic network loading in this section. We start by introducing some notations

$\mathcal{A}$: the set of arcs in the network;

$\mathcal{V}$: the set of nodes in the network;

$\mathcal{W}$: the set of origin-destination pairs in the network;

$\mathcal{P}$: the set of utilized paths in the network;

$p \ = \{e_1,\ e_2,\ \ldots,\ e_{m(p)}\} \in \mathcal{P},\ e_i \in \mathcal{A}$: a viable path represented as the collection of arcs;

$h_p(t)$: departure rate associated with path $p \in \mathcal{P}$;

$Q_p^e(t)$: cumulative entering vehicle count of arc $e \in \mathcal{A}$ associated with path $p \in \mathcal{P}$;

$q_p^e(t)$: entry flow of arc $e \in \mathcal{A}$ associated with path $p \in \mathcal{P}$;

$W_p^e(t)$: cumulative exiting vehicle count of arc $e \in \mathcal{A}$ associated with path $p \in \mathcal{P}$;

$w_p^e(t)$: exit flow of arc $e \in \mathcal{A}$ associated with path $p \in \mathcal{P}$;

L^e : length of arc $e \in \mathcal{A}$;

ψ^e : Legendre transformation shown in (2.18) and (2.21), associated with $e \in \mathcal{A}$;

$D(t\,; Q^e)$: arc traversal time function of $e \in \mathcal{A}$ when the boundary datum is Q^e, and entry time is t;

$$
Q^e(t) \ \doteq \ \sum_{e \in p} Q_p^e(t), \qquad W^e(t) \ \doteq \ \sum_{e \in p} W_p^{e_i}(t).
$$

By convention we write $q_p^{e_1}(t) \ = \ h_p(t),\ w_p^{e_0}(t) \ = \ h_p(t)$. The following DAE system (2.30)-(2.35) summarizes our network loading procedure based on the LWR partial differential equation.

$$
Q^e(t) \ \doteq \ \sum_{e \in p} Q_p^e(t), \qquad q^e(t) \ \doteq \ \sum_{e \in p} q_p^e(t), \qquad w^e(t) \ \doteq \ \sum_{e \in p} w_p^e(t); \tag{2.30}
$$

$$
\frac{d}{dt} Q_p^e(t) \ = \ q_p^e(t), \qquad \frac{d}{dt} W^e(t) \ = \ w^e(t), \qquad \forall\ p \in \mathcal{P}; \tag{2.31}
$$

$$
q_p^{e_i}(t) \ = \ w_p^{e_{i-1}}(t), \qquad i \in [1,\ m(p)], \qquad p \in \mathcal{P}; \tag{2.32}
$$

$$
W^e(t) \ = \ \min_{\tau} \left\{ Q^e(\tau) + L^e \psi^e \left(\frac{t - \tau}{L^e} \right) \right\}, \qquad \forall\ e \in \mathcal{A}; \tag{2.33}
$$

$$
Q^e(t) \ = \ W^e(t + D(t;\ Q^e)), \qquad \forall e \in \mathcal{A}; \tag{2.34}
$$

$$
w_p^{e_i}(t + D(t;\ Q^{e_i})) \ = \ \frac{q_p^{e_i}(t)}{q^{e_i}(t)}\, w^{e_i}(t + D(t;\ Q^{e_i})), \qquad i \in [1,\ m(p)], \quad p \in \mathcal{P}. \tag{2.35}
$$

We note that equation (2.30) is merely definitional, i.e. the traffic on an arc is disaggregated according to different route choices. Equation (2.32) represents the fundamental recursion, which allows the algorithm to carry forward to the next arc in the path. Equation (2.33) is the Lax formula. (2.34) is often referred to as the flow propagation constraint. It determines the travel time function $D(\cdot\,; Q^e)$ in our case. (2.35) represents the diverging model in Section 2.5.

Remark 2.15. *We provide some insights on the existence and uniqueness of the solution to the above DAE system. Notice that the system does not contain any differential equations, the only equations with differentiation (2.31) are only definitional. Moreover, the Lax-Hopf formula (2.33) provides the unique viscosity solution to the H-J equations. In addition, all variables of interests appear explicit on the left hand side of the equations, so that they can be solved for completely sequentially and in a pseudo cascading fashion. Therefore, the existence and uniqueness of solution to the DAE system is clear.*

Remark 2.16. *When we set out to solve the DNL submodel using a system of DAE in the present paper, we are making a numerical approximation that can be as accurate as the modeler desires given sufficient computational resources. However, that numerical approximation should not be interpreted as an approximate DNL model as presented in the authors' previously published approaches in Friesz et al. (2011).*

3. Formulation of the dynamic user equilibrium and Kuhn-Tucker conditions for the variational inequality

The dynamic network loading discussed in the previous section is viewed as the numerical procedure for evaluating the effective delay operator, which is embedded in the definition of DUE proposed in Friesz et al. (1993). In this section we will re-visit the formulations of DUE as a variational inequality of Friesz et al. (1993), as a differential variational inequality of Friesz et al. (2001) and as a fixed-point problem of Friesz and Mookherjee (2006).

The goal of this section is the following. (1) We wish to provide the readers with a comprehensive review of topics related to the DUE and clear guidance on how to compute the DUE. (2) We will utilize the Kuhn-Tucker conditions for finite-dimensional VIs to verify that the numerical solution we obtain is indeed the solution to the VI, and hence to the DUE problem. To the authors' knowledge this is the first attempt in the literature to test the numerical solutions against necessary conditions of VIs. Thus the work in this section can serve as a valuable side reference to future researchers who are interested in the computational aspect of DUE.

3.1. Dynamic User Equilibrium formulation

We begin by considering a planning horizon $[t_0,\, t_f] \subset \mathbb{R}_+$. Let $\mathcal{P}$ be the set of paths employed by road users. The most crucial ingredient of the DUE model is the path delay operator. Such an operator, denoted by

$$D_p(t, h) \quad \forall p \in \mathcal{P},$$

maps a given vector of departure rates h to the collection of travel times. Each travel time is associated with a particular choice of route $p \in \mathcal{P}$ and departure time $t \in [t_0,\, t_f]$. The path delay operators usually do not take on any closed form, instead they can only be evaluated numerically through the dynamic network loading procedure. On top of the path delay operator we introduce the effective path delay operator which generalizes the notion of travel cost to include early or late arrival penalties. In this paper we consider the effective path delay operators of the following form.

$$\Psi_p(t,h) = D_p(t,h) + F\left[t + D_p(t,h) - T_A\right] \quad \forall p \in P \tag{3.36}$$

where T_A is the target arrival time. In our formulation the target time T_A is allowed to depend on the user classes. We introduce the fixed trip matrix $(Q_{ij} : (i,\ j) \in \mathcal{W})$, where each $Q_{ij} \in \mathbb{R}_+$ is the fixed travel demand between origin-destination pair $(i, j) \in \mathcal{W}$. Note that Q_{ij} represents traffic volume, not flow. Finally we let $\mathcal{P}_{ij} \subset \mathcal{P}$ to be the set of paths connecting origin-destination pair $(i, j) \in \mathcal{W}$.

As mentioned earlier h is the vector of path flows $h = \{h_p : p \in \mathcal{P}\}$. We stipulate that each path flow is square integrable, that is

$$h \in (\mathcal{L}_+^2[t_0,\ t_f])^{|\mathcal{P}|}$$

The set of feasible path flows is defined as

$$\Lambda_0 = \left\{ h \geq 0 : \sum_{p \in \mathcal{P}_{ij}} \int_{t_0}^{t_f} h_p(t)\, dt = Q_{ij} \quad \forall\, (i,j) \in \mathcal{W} \right\} \subseteq \left(L_+^2 \left[t_0, t_f\right] \right)^{|\mathcal{P}|} \tag{3.37}$$

Let us also define the essential infimum of effective travel delays

$$v_{ij} = \text{essinf}\left[\Psi_p(t,h) : p \in \mathcal{P}_{ij}\right] \quad \forall\, (i,j) \in \mathcal{W}$$

The following definition of dynamic user equilibrium was first articulated by Friesz et al. (1993):

Definition 3.1. (Dynamic user equilibrium). *A vector of departure rates (path flows) $h^* \in \Lambda_0$ is a dynamic user equilibrium if*

$$h_p^*(t) > 0, p \in P_{ij} \Longrightarrow \Psi_p\left[t, h^*(t)\right] = v_{ij}$$

We denote this equilibrium by $DUE\left(\Psi, \Lambda_0, \left[t_0, t_f\right]\right)$.

Using measure theoretic arguments, Friesz et al. (1993) established that a dynamic user equilibrium is equivalent to the following variational inequality under suitable regularity conditions:

$$\left.\begin{array}{c} \text{find } h^* \in \Lambda_0 \text{ such that} \\ \sum_{p \in \mathcal{P}} \int_{t_0}^{t_f} \Psi_p(t,h^*)(h_p - h_p^*)dt \geq 0 \\ \forall h \in \Lambda_0 \end{array}\right\} VI(\Psi, \Lambda_0, \left[t_0, t_f\right]) \tag{3.38}$$

It has been noted in Friesz et al. (2001) that (3.38) is equivalent to a differential variational inequality (DVI). This is most easily seen by noting that the flow conservation constraints may be re-stated as

$$\left.\begin{array}{l} \dfrac{dy_{ij}}{dt} = \sum_{p \in \mathcal{P}_{ij}} h_p(t) \\ y_{ij}(t_0) = 0 \\ y_{ij}\left(t_f\right) = Q_{ij} \end{array}\right\} \quad \forall\, (i,j) \in \mathcal{W}$$

which is recognized as a two point boundary value problem. As a consequence (3.38) may be expressed as the following DVI:

$$\left.\begin{array}{c}\text{find } h^* \in \Lambda \text{ such that} \\ \displaystyle\sum_{p\in\mathcal{P}} \int_{t_0}^{t_f} \Psi_p(t, h^*)(h_p - h_p^*)dt \geq 0 \\ \forall h \in \Lambda\end{array}\right\} DVI(\Psi, \Lambda, \left[t_0, t_f\right]) \tag{3.39}$$

where

$$\Lambda = \left\{ h \geq 0 : \frac{dy_{ij}}{dt} = \sum_{p\in P_{ij}} h_p\,(t)\,,\ y_{ij}(t_0) = 0,\ y_{ij}\left(t_f\right) = Q_{ij} \quad \forall\,(i, j) \in \mathcal{W} \right\} \tag{3.40}$$

Finally, we are in a position to state a result that permits the solution of the DVI (3.39) to be obtained by solving a fixed point problem[2]:

Theorem 3.2. (Fixed point re-statement). *Assume that* $\Psi_p(\cdot, h) : \left[t_o,\ t_f\right] \longrightarrow \mathbb{R}_+$ *is measurable for all* $p \in \mathcal{P}$, $h \in \Lambda$. *Then the fixed point problem*

$$h^* \ = \ P_\Lambda\left[h^* - \alpha\Psi\,(t,\ h^*)\right] \tag{3.41}$$

is equivalent to $DVI(\Psi, \Lambda, \Delta)$ *where* $P_\Lambda[\cdot]$ *is the minimum norm projection onto* Λ *and* $\alpha \in \mathbb{R}_+$.

Proof. See Friesz (2010). □

3.2. Solution algorithm based on the fixed-point iterations

The fixed-point problem (3.2) can be instantiated as the following linear-quadratic optimal control problem.

$$h^* = \arg\min_h \left\{ \frac{1}{2}\,\|h^* - \alpha\Psi\,(t,\ h^*) - h\| :\ h \in \Lambda \right\} \tag{3.42}$$

Then we can write the fixed point sub-problem at the k^{th} iteration as follows.

$$h^{k+1} \ = \ \arg\min_h \left\{ \frac{1}{2}\left\| h^k - \alpha\Psi\left(t,\ h^k\right) - h \right\|^2 :\ h \in \Lambda \right\} \tag{3.43}$$

That is, we seek the solution of the optimal control problem

$$\min_h J^k(h) \ = \ \sum_{(i,j)\in\mathcal{W}} v_{ij}\left[Q_{ij} - y_{ij}(t_f)\right] + \sum_{(i,\,j)\in\mathcal{W}} \sum_{p\in\mathcal{P}_{ij}} \int_{t_0}^{t_f} \frac{1}{2}\left[h_p^k - \alpha\Psi_p\,(t,\ h) - h_p\right]^2 \tag{3.44}$$

[2]The fixed-point iteration scheme should not be confused with simulation-based approaches since our iterations are merely approximations that are successively refined until convergence produces a realized system state; that is, iterations prior to convergence have no physical meaning. This is in contrast to simulation where each iteration of a simulation model produces a representation of a system state or scenario at a particular point in time. These states are intended to directly map onto the physical world typically with associated performance measures.

subject to:

$$\frac{dy_{ij}}{dt} = \sum_{p \in \mathcal{P}_{ij}} h_p(t) \qquad \forall (i,j) \in \mathcal{W} \tag{3.45}$$

$$y_{ij}(t_0) = 0 \qquad \forall (i,j) \in \mathcal{W} \tag{3.46}$$

$$h \geq 0 \tag{3.47}$$

Finding dual variables associated with terminal time demand constraints turns out to be relatively easy. Note that the relevant Hamiltonian for (3.44), (3.45), (3.46) and (3.47) is:

$$H^k = \frac{1}{2} \sum_{(i,j) \in \mathcal{W}} \sum_{p \in \mathcal{P}_{ij}} \left[h_p^k - \alpha \Psi_p \left(t, h^k \right) - h_p \right]^2 + \sum_{(i,j) \in \mathcal{W}} \lambda_{ij} \sum_{p \in P_{ij}} h_p \tag{3.48}$$

where each λ_{ij} is an adjoint variable obeying

$$\frac{d\lambda_{ij}}{dt} = (-1) \frac{\partial H^k}{\partial y_{ij}} = 0 \qquad \forall (i,j) \in \mathcal{W}$$

$$\lambda_{ij}(t_f) = \frac{\partial v_{ij} \left[Q_{ij} - y_{ij}(t) \right]}{\partial y_{ij}(t_f)} = -v_{ij} \qquad \forall (i,j) \in \mathcal{W}$$

From the above we determine that

$$\lambda_{ij}(t) \equiv -v_{ij} \qquad \forall \quad t \in [t_0, t_f], \ (i,j) \in \mathcal{W}$$

The minimum principle implies for any $p \in \mathcal{P}$,

$$h_p^{k+1}(t) = \arg \left\{ \frac{\partial H^k}{\partial h_p} = 0 \right\} = \arg \left\{ \left[\left(h_p^k(t) - \alpha \Psi_p \left(t, h^k \right) - h_p(t) \right) \right] (-1) - v_{ij} = 0 \right\}$$

Thus we obtain

$$h_p^{k+1}(t) = \left[h_p^k(t) - \alpha \Psi_p \left(t, h^k \right) + v_{ij} \right]_+ \qquad \forall (i,j) \in \mathcal{W}, p \in \mathcal{P}_{ij} \tag{3.49}$$

Notice that the following flow conservation constraint applies here.

$$\sum_{p \in \mathcal{P}_{ij}} \int_{t_0}^{t_f} h_p^{k+1}(t)\, dt = Q_{ij} \qquad \forall (i,j) \in \mathcal{W}$$

Consequently the dual variable v_{ij} must satisfy

$$\sum_{p \in \mathcal{P}_{ij}} \int_{t_0}^{t_f} \left[h_p^k(t) - \alpha \Psi_p \left(t, h^k \right) + v_{ij} \right]_+ dt = Q_{ij} \qquad \forall (i,j) \in \mathcal{W} \tag{3.50}$$

Recalling that each v_{ij} is time invariant and noticing that the equations of (3.50) are uncoupled, we see that simple line searches will find the value of v_{ij} satisfying the above conditions. Once the v_{ij}'s are determined, the new vector of path flows h^{k+1} for the next iteration may be computed from (3.49).

Remark 3.3. *The fixed-point iteration (3.43) is clearly a particular instance of the abstract algorithm $x^{k+1} = M(x^k)$ for solving the fixed point problem $x^* = M(x^*)$. However such algorithm does not enjoy a theorem assuring strong convergence of $\{x^k\}$, even when $M(\cdot)$ is non-expansive. There are however, instances of convergence result established under various assumptions on the delay operator. In Friesz et al. (2011), the authors showed strong convergence of a modified fixed point algorithm with either strongly monotone or component-wise strongly pseudomonotone delay operators, under minor regularity conditions. However, the delay operator associated with the dynamic network loading procedure proposed in this paper is unlikely to satisfy either one of the above two assumptions. And further analytical properties need to be extracted from the current model to assist the understanding of convergence issue. This has been left as future research. It should be noted that in all the numerical studies we present below, the fixed-point iterations do meet the convergence criteria, that is*

$$\frac{\left\|h^{k+1} - h^k\right\|_{L^2}}{\left\|h^k\right\|_{L^2}} \leq \varepsilon, \qquad \textit{for some } k \tag{3.51}$$

where ε is some user-defined threshold.

3.3. Kuhn-Tucker condition for finite-dimensional VIs

This section focuses on the variational inequality (3.38), we point out that such formulation of DUE problem is a mathematical abstraction and is of no use for computation as it presumes knowledge of the solution h^*. However, a Kuhn-Tucker type necessary condition can be derived for a discrete-time version of (3.38) and utilized to verify numerical solutions obtained from fixed-the point algorithm (3.41).

To derive a finite-dimensional version of VI (3.38), we introduce a uniform time grid with step size Δt

$$t_0 = t_1 < t_2 < \ldots < t_n = t_f, \tag{3.52}$$

Use notation

$$\overline{h}_{p,i} \doteq h_p(t_i) \in \mathbb{R}, \qquad \overline{h}_p \doteq (\overline{h}_{p,i})_{i=1}^n \in \mathbb{R}^n, \qquad \overline{h} \doteq (\overline{h}_p)_{p\in\mathcal{P}} \in \mathbb{R}^{n\times|\mathcal{P}|},$$

$$\overline{\Psi}_{p,i}(\overline{h}) \doteq \Psi_p(t_i, h), \qquad \overline{\Psi}_p(\overline{h}) \doteq (\overline{\Psi}_{p,i}(\overline{h}))_{i=1}^n \in \mathbb{R}^n \qquad \overline{\Psi}(\overline{h}) \doteq (\overline{\Psi}_p(\overline{h}))_{p\in\mathcal{P}} \in \mathbb{R}^{n\times|\mathcal{P}|}$$

$$Q = (Q_{ij})_{(i,j)\in\mathcal{W}} \in \mathbb{R}^{|\mathcal{W}|}$$

Apply rectangular quadrature for the approximation of integrals, then the discrete-time version of inequality in (3.38) becomes

$$\sum_{p\in\mathcal{P}} \Delta t \overline{\Psi}_p(\overline{h}^*)^T(\overline{h}_p - \overline{h}_p^*) \geq 0 \iff \overline{\Psi}(\overline{h}^*)^T(\overline{h} - \overline{h}^*) \geq 0$$

The flow conservation constraints become

$$\sum_{p\in\mathcal{P}_{ij}} \Delta t\, e^T \overline{h}_p = Q_{ij}, \qquad \forall (i, j) \in \mathcal{W}$$

where $e \in \mathbb{R}^n$ is the vector of ones. Introducing compact notation

$$\alpha(\cdot) : \mathbb{R}^{n\times|\mathcal{P}|} \longrightarrow \mathbb{R}^{n\times|\mathcal{P}|}, \quad \alpha(\overline{h}) = -\overline{h} \tag{3.53}$$

$$\beta(\cdot) : \mathbb{R}^{n\times|\mathcal{P}|} \longrightarrow \mathbb{R}^{n\times|\mathcal{P}|}, \quad \beta(\overline{h}) = A \cdot \overline{h} - Q/\Delta t \tag{3.54}$$

where each row of the matrix $A \in \mathbb{R}^{|\mathcal{W}| \times n|\mathcal{P}|}$ has ones located at the spot corresponding to O-D pair $(i,\ j)$ and zeros elsewhere. $\overline{\Lambda}_0$ can be rewritten as

$$\overline{\Lambda}_0 \;=\; \left\{\overline{h} \in \mathbb{R}^{n \times |\mathcal{P}|}: \quad \alpha(\overline{h}) \;\leq\; 0; \quad \beta(\overline{h}) \;=\; 0\right\}$$

Now we are ready to state the discrete-time variational inequality $\overline{VI}(\Psi, \Lambda_0, [t_0,\ t_f])$ and the necessary condition for this finite-dimensional VI.

$$\left.\begin{array}{c} \text{find } \overline{h}^* \in \overline{\Lambda}_0 \text{ such that} \\ \overline{\Psi}(\overline{h}^*)^T(\overline{h} - \overline{h}^*) \;\geq\; 0 \\ \forall \overline{h} \in \overline{\Lambda}_0 \end{array}\right\} \overline{VI}(\overline{\Psi}, \overline{\Lambda}_0, \left[t_0, t_f\right]) \tag{3.55}$$

where

$$\overline{\Lambda}_0 \;=\; \left\{\overline{h} \in \mathbb{R}^{n \times |\mathcal{P}|}: \quad \alpha(\overline{h}) \;\leq\; 0; \quad \beta(\overline{h}) \;=\; 0\right\} \tag{3.56}$$

Theorem 3.4. (Kuhn-Tucker conditions for $\overline{VI}(\overline{\Psi}, \overline{\Lambda}_0, [t_0,\ t_f])$)
Let

$$\overline{h}^* \in \overline{\Lambda}_0 \;=\; \left\{\overline{h} \in \mathbb{R}^{n \times |\mathcal{P}|}: \quad \alpha(\overline{h}) \;\leq\; 0; \quad \beta(\overline{h}) \;=\; 0\right\}$$

be a solution of (3.55). Further assume that $\overline{\Psi}(\cdot)$ and $\alpha(\cdot)$ are continuous on $\overline{\Lambda}_0$, $\alpha(\cdot)$ is differentiable on $\overline{\Lambda}_0$, and $\beta(\cdot)$ is affine on $\overline{\Lambda}_0$. Then if the gradients $\nabla \alpha_i(\overline{h}^*)$ for i such that $g_i(\overline{h}^*) = 0$ together with the gradients $\nabla \beta_i(\overline{h}^*)$ for $i = 1, \ldots, n \times |\mathcal{P}|$ are linearly independent, there exists multipliers $\pi \in \mathbb{R}^{n \times |\mathcal{P}|}$ and $\mu \in \mathbb{R}^{n \times |\mathcal{P}|}$ such that

$$\overline{\Psi}(\overline{h}^*) + [\nabla \alpha(\overline{h}^*)]^T \pi + [\nabla \beta(\overline{h}^*)]^T \mu \;=\; 0 \tag{3.57}$$

$$i = 1, \ldots, n \times |\mathcal{P}|, \qquad \pi_i \alpha_i(\overline{h}^*) \;=\; 0 \tag{3.58}$$

$$\pi \;\geq\; 0 \tag{3.59}$$

Proof. See Friesz (2010) □

Remark 3.5. *The functions $\alpha(\cdot)$, $\beta(\cdot)$ clearly satisfy the assumption of the above theorem. Regarding the continuity of the effective delay operator $\Psi(\cdot,\ \cdot)$ in continuous time, we refer the readers to Bressan and Han (submitted for publication) for a proof. The continuity of $\overline{\Psi}(\cdot)$ in Theorem 3.4 follows directly from the continuity of $\Psi(\cdot,\ \cdot)$. We further comment that the necessary condition for such VI becomes sufficient if in addition the constraint functions $\alpha(\cdot)$ are convex, which is again true in our case.*

Next, we will transform condition (3.57)-(3.58) into a linear system. Note

$$\nabla \alpha(\overline{h}^*) \;=\; -I, \qquad \nabla \beta(\overline{h}^*) \;=\; A$$

where I is the $n|\mathcal{P}| \times n|\mathcal{P}|$ identity matrix. Let $\Omega = \text{diag}(-\overline{h}^*) \in \mathbb{R}^{n|\mathcal{P}| \times n|\mathcal{P}|}$, (3.57)-(3.58) is equivalent to the following linear system

$$\begin{bmatrix} -I & A \\ \Omega & 0 \end{bmatrix} \begin{bmatrix} \pi \\ \mu \end{bmatrix} \;=\; \begin{bmatrix} -\overline{\Psi}(\overline{h}^*) \\ 0 \end{bmatrix} \tag{3.60}$$

In order to verify the sufficient and necessary condition (3.57)-(3.59), one needs to solve the above system and verify that the dual variables $\pi \geq 0$. Notice that the size of such system is $2n|\mathcal{P}| \times (n|\mathcal{P}| + |\mathcal{W}|)$ which could be very large, for example, in the Sioux Falls network example presented in Section 4.5. The Sioux Falls network studied in this paper has 496 O-D pairs and 1,941 paths. A chosen time grid of 100 points will lead to a $400,000$-by-$200,500$ linear system. Such a large system will cause the solution procedure to significantly slow down and even crash if the system is not well-conditioned. We present our resolution to this issue with a decomposition of the system (3.60) summarized in the next lemma.

Lemma 3.6. (Decomposition of the linear system) *For $i = 1, \ldots, |\mathcal{W}|$, denote by $\overline{\Psi}^i(\overline{h}^*)$, π^i, μ^i the vectors formed by selecting the rows corresponding to the i^{th} O-D pair of the original vectors; denote by I^i, A^i, Ω^i the submatrices formed by selecting the rows and colomns corresponding to the i^{th} O-D pair of the original matrices. The linear system (3.60) is then equivalent to a collection of smaller systems*

$$\begin{bmatrix} -I^i & A^i \\ \Omega^i & 0 \end{bmatrix} \begin{bmatrix} \pi^i \\ \mu^i \end{bmatrix} = \begin{bmatrix} -\overline{\Psi}^i(\overline{h}^*) \\ 0 \end{bmatrix}, \qquad i = 1, \ldots, |\mathcal{W}| \tag{3.61}$$

Proof. The conclusion follows from a simple block matrix arithmetic. □

Each system in (3.61) is much smaller in size, however, one may have to solve such linear systems as many times as the number of O-D pairs. This strategy is particularly useful when computer memory is insufficient for large-scale computation.

4. Numerical examples

In this section, we present several numerical studies of the proposed DNL procedure formulated as a DAE system (2.30)-(2.35). In particular, we consider four different vehicular networks of varying sizes and topologies, as summarized in Table 1

Network name	Arcs	Nodes	OD pairs	Paths
Simple network	2	2	1	2
Small network	6	5	2	6
Medium network	19	13	4	4
Sioux Falls network	76	24	496	1941

Table 1: Summary of network examples

All computations of the DNL procedure and fixed-point iteration presented in this section were coded in Matlab 2010 and performed on a 2.7 GHz 4 GB of RAM computer. Every attempt was made to avoid the use of numerical tricks that could not be equally applied to all models. For the numerical results of the DNL, the LWR-based DAE system presented in this paper displays different flow propagation patterns from the LDM and CTM link performance models studied in Friesz et al. (2011): the solutions for the latter are moving parabolic wave forms, while the solutions of our DNL procedure displays the non-uniformity of wave speeds − a key feature of kinematic wave models. This is best explained by the flux function $\phi(\cdot)$ defined in (2.7), which implies that the higher the flow, the slower it propagates, such property will be visualized in Section 4.3.

4.1. Numerical context

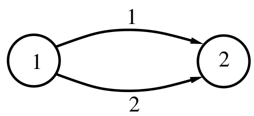
For all the numerical examples resented in this section, we will fix the time horizon [0, 20]. The effective delay takes on the following form

$$\Psi_p(t, h) \; = \; D_p(t, h) + \gamma \times (t + D_p(t, h) - T_A)^2$$

where $D_p(t, h)$ is the network delay operator for path $p \in \mathcal{P}$ and T_A is the target arrival time. We assume that there is a quadratic penalty for arrivals other than the desired time. The parameter $\gamma \in \mathbb{R}_+$ measures the network users' perception of arrival penalties and is allowed to depend on different user classes. As the initial iterate of the fixed-point algorithm, we utilize an inverse parabolic-shaped initial departure rate. Note that an appropriate pre-conditioning such as a different initial guess may improve the performance of the fixed-point algorithm, but this it is beyond the scope of this paper. Finally, the convergence criteria of the fixed-point algorithm is set to be (3.51).

4.2. Simple network

In this section we consider the 2-arc, 2-node network with arc parameters illustrated in Figure 4. This simple network has only one origin-destination pair $\mathcal{W} = \{1, 2\}$ and two paths $p_1 = \{1\}$, $p_2 = \{2\}$.



Arc	Jam density (vehicles / km)	Free flow speed (km / 5min)	Length (km)
1	800	6	4
2	600	3	5

Figure 4: Topology and arc parameters of the simple network.

We assume a fixed travel demand of $Q_{1,2} = 3300$ and set the desired arrival time $T_A = 15$. With the DNL procedure proposed in this paper, it took the fixed-point algorithm 15 iterations to meet the convergence criteria. The equilibrium flows are illustrated in Figure 5 where a clear flat spot in the travel cost curve can be visualized. As the iteration goes on, more traffic volume will switch from the second path to the first path, as shown in Figure 6, this is consistent with the network parameters we choose.

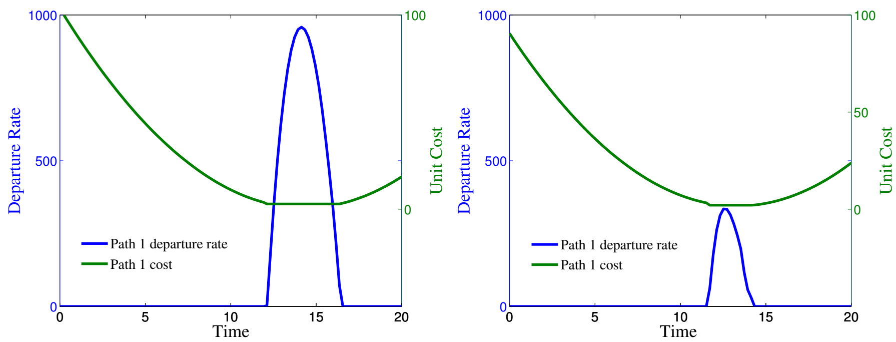


Figure 5: Simple network: path flows and unit costs.

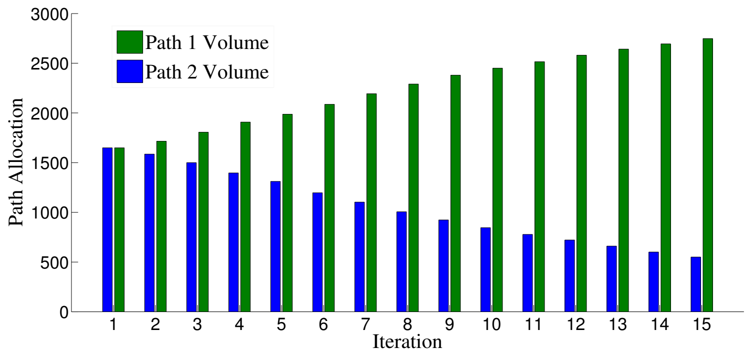


Figure 6: Path allocations during each fixed-point iteration

4.3. Small network

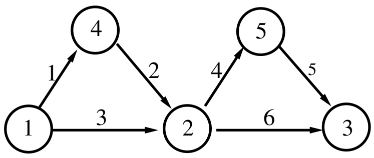
We consider the small network summarized in Figure 7, with 6 arcs and 5 nodes. There are two O-D pairs

$$\mathcal{W} = \{(1, 3), (2, 3)\}$$

among which the following six paths are employed:

$$p_1 = \{3, 6\}, \qquad p_2 = \{1, 2, 6\}, \qquad p_3 = \{1, 2, 4, 5\}, \qquad p_4 = \{3, 4, 5\},$$

$$p_5 = \{6\}, \qquad p_6 = \{4, 5\}$$



Arc	Jam density (vehicles / km)	Free flow speed (km / 5min)	Length (km)
1	800	6	4
2	800	6	8
3	800	8	4
4	800	8	10
5	1000	8	8
6	600	6	10

Figure 7: Topology and arc parameters of the small network.

We assume a fixed demand for O-D pair (1, 3) and (2, 3) such that $Q_{1,3} = 3300$ and $Q_{2,3} = 1650$. The target arrival times of each OD pair are set to be $T_{(1,3)} = 10$ and $T_{(2,3)} = 14$. It took the fixed-point algorithm 12 iterations to terminate and the DUE solutions are shown in Figure 8. We note that the curve corresponding to travel cost is flat where the departure rate is nonzero. Figure 9 shows the propagation of several path flows along each arc. As mentioned at the beginning of this section, the LWR-based models tend to display non-uniform wave speeds, i.e. the part of the flow profile with relatively high value will propagate more slowly than the part with low value.

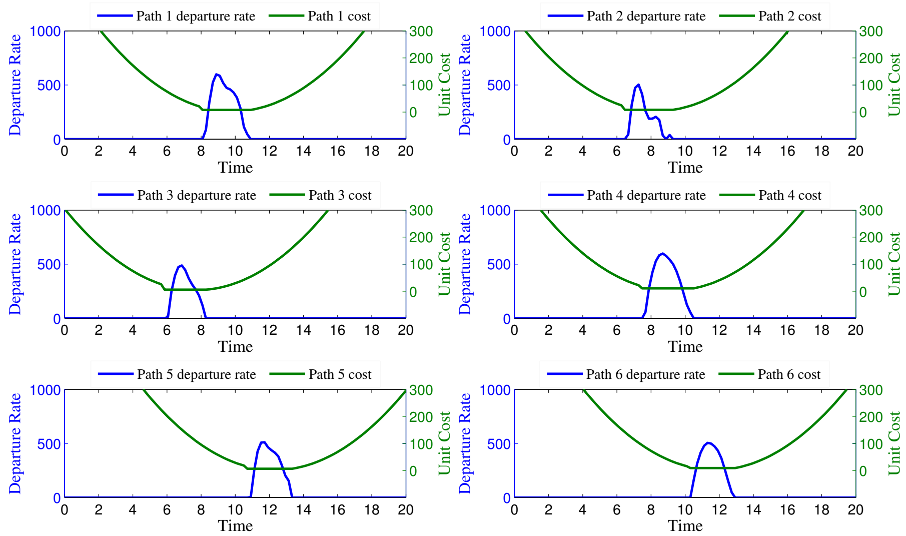


Figure 8: Small network: the six path departure rates and corresponding travel cost in the DUE solution.

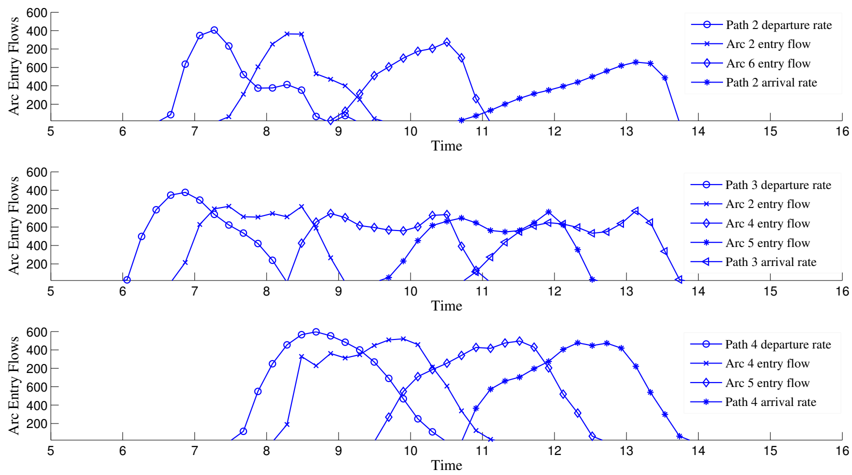


Figure 9: Small network: propagation of several path flows along each arc.

4.4. Medium Network

We now consider the 19 arc 13 node network shown in Figure 10 that was also studied in Nguyen (1984); Nie and Zhang (2008). Four O-D pairs are chosen

$$\mathcal{W} = \{(1, 2), (1, 3), (4, 2), (4, 3)\}$$

each having a fixed travel demand of $Q_{1,2} = Q_{1,3} = Q_{4,2} = Q_{4,3} = 275$. The target arrival times for the four O-D pairs are $T_{(1,2)} = T_{(1,3)} = T_{(4,2)} = T_{(4,3)} = 12$. We employ only one path per O-D pair. Specifically, $\mathcal{P} = \{p_1, p_2, p_3, p_4\}$ and

$$p_1 = \{1, 4, 13\}, \quad p_2 = \{2, 10, 17, 19\}, \quad p_3 = \{9, 14, 15, 16\}, \quad p_4 = \{9, 17, 19\}$$

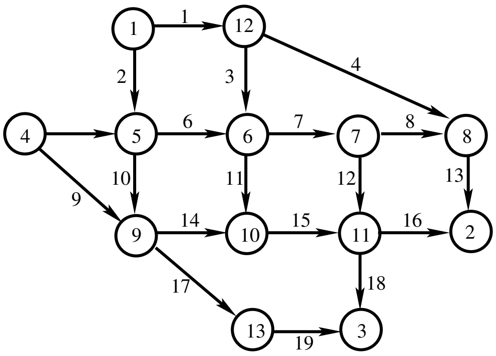


Figure 10: The 19 arc, 13 node medium network

The primary purpose of this network example is to study the effect of bottlenecking on DUE solutions and how the bottleneck is handled by our algorithm. To do so, we set the network parameters in a way such that arc 9 becomes a bottleneck, i.e. it has smaller flow capacity and free flow speed than the rest of the arcs in the network. Under this numerical setting, the DUE solution is obtained after 10 fixed-point iterations and four path flows are shown in Figure 11. We observe from the solution that the paths p_3, p_4 which pass the bottleneck, have higher peak rates than paths p_1 and p_2. This observation implies that equilibrium flows tend to aggravate congestions at bottlenecks as confirmed by Bressan and Han (2011a). Figure 12 demonstrates the equilibrium flows through the bottleneck (arc 9) whose flow capacity is 200. We make the following comments on Figure 12: (1) the flow capacity of arc 9 is captured by the Lax-formula approach as shown in the lower figure; (2) a clear shock wave is represented as a sudden drop of flow around $t = 12.6$; (3) the entry and exit flows of arc 9 are successfully disaggregated by the routing information as well as our diverging model and numerical implementation.

4.5. Sioux Falls network, part I

The aim of this section is to demonstrate the ability of our DNL procedure to compute large-scale problems. We consider the well-known Sioux Falls network (Figure 13) consisting of 76 arcs and 24 nodes. We select a total number of 496 origin-destination pairs and 1941 paths. For each OD pair, we assume a fixed travel demand ranging from 200−2000, and target arrival time ranging from 8−16.

The fixed-point algorithm took 17 iterations to converge. Part of the DUE solutions are presented in Figure 14-19. Figure 14 -17 show departure rates of four paths p_{300}, p_{500}, p_{850}, p_{1902} and the corresponding unit cost. Figure 18 shows the evolution of path allocations within OD pair (18, 20) at each fixed-point iteration. Several arc entry flows in the solution are depicted in Figure 19. We test our numerical solution against the KT condition mentioned in Theorem 3.4, the resulting linear system is decomposed using Lemma 3.6.

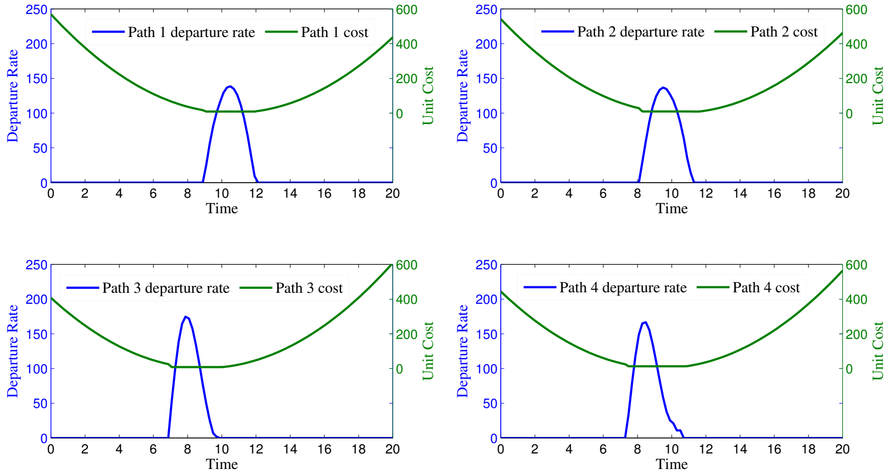


Figure 11: Medium network: DUE departure rates and corresponding travel cost.

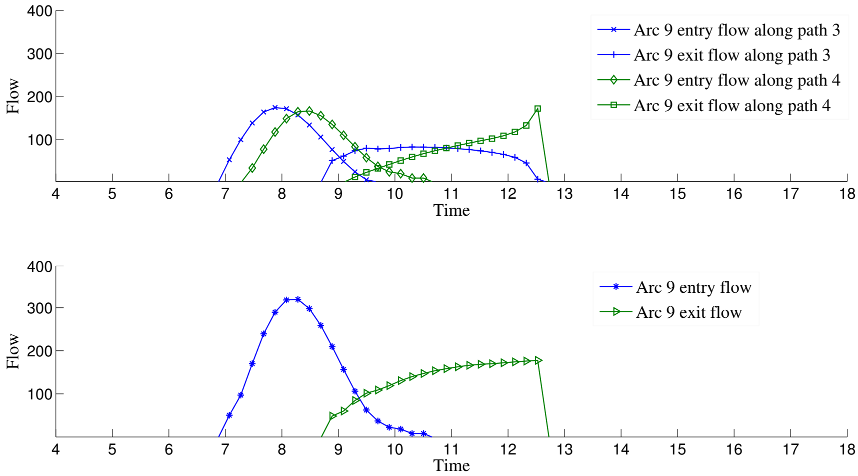


Figure 12: Medium network: DUE solution at bottleneck (arc 9). Upper: path flows through bottleneck. Lower: total flow through bottleneck.

To confirm that our numerical solution indeed solves the DUE problem on Sioux Falls network, we test it against the KKT condition instantiated in Theorem 3.4 and Lemma 3.6. The whole linear system is decomposed into smaller ones according to different O-Ds, and all the resulting dual variables are non-negative.

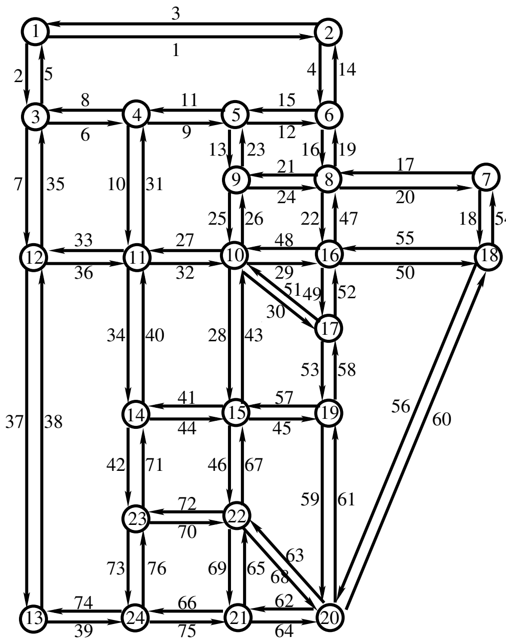


Figure 13: The Sioux Falls network

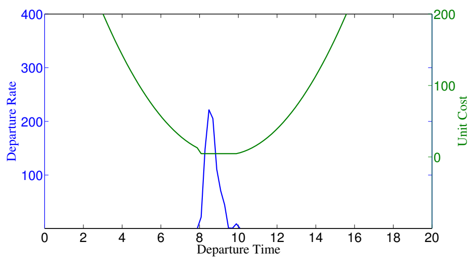


Figure 14: Path 300: departure rate and cost

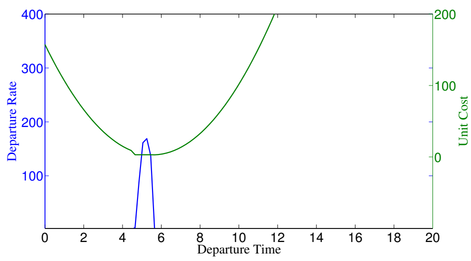


Figure 15: Path 500: departure rate and cost

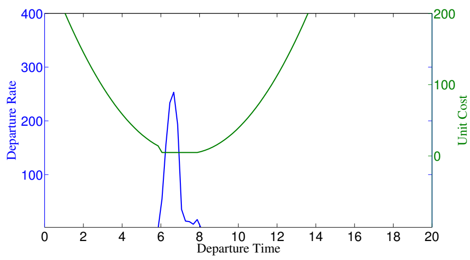


Figure 16: Path 850: departure rate and cost

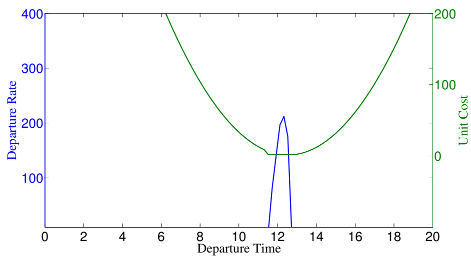


Figure 17: Path 1902: departure rate and cost

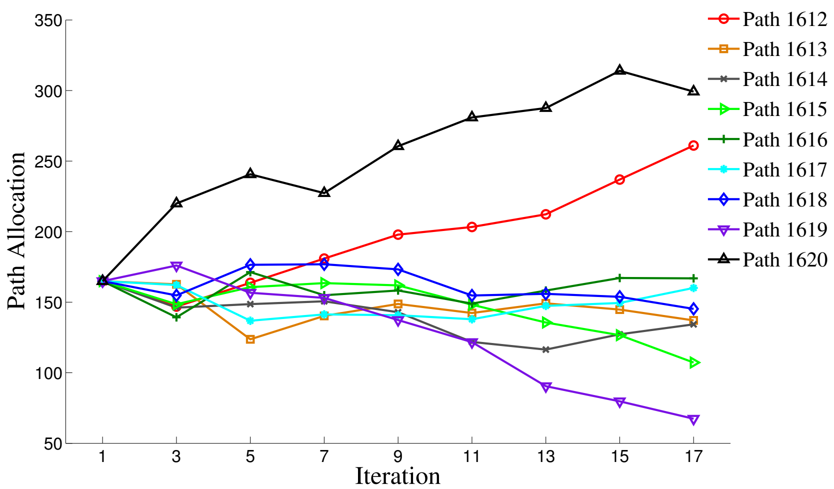


Figure 18: Evolution of path allocations for OD pair (18, 21)

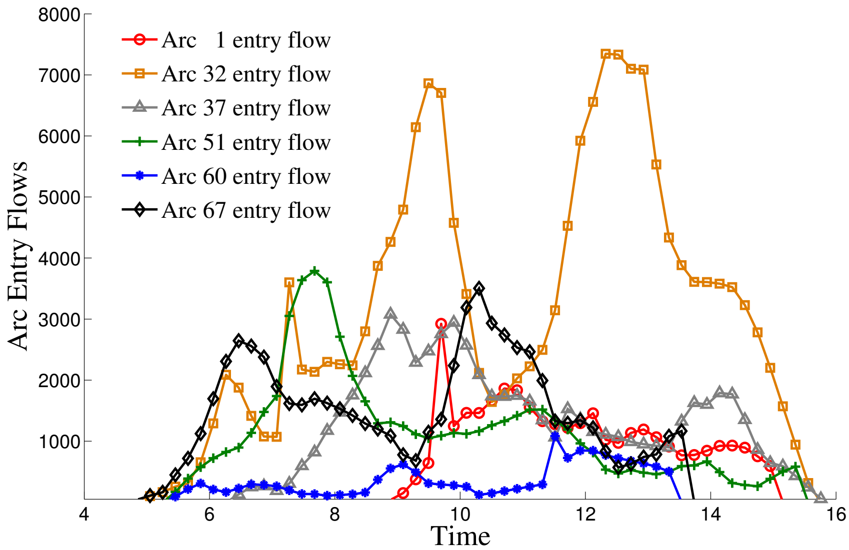


Figure 19: Entry flows on several arcs

4.6. Sioux Falls network, part II

The analytical framework of this paper belongs to the class of path-based DUE models. In other words a (partial) path enumeration is necessary for the analysis and computation of DUE to carry forward. In the first part of the Sioux Falls study, we end up with four paths per O-D pair on average. However, notice that many O-D pairs are nodes right next to each other (e.g. node 4 and 5). It is considered 'unlikely' that users will select too many paths for such an O-D pair.

In order to further illustrate the influence of network congestion and travel demand on the number of utilized paths, and in order to understand the limitation of insufficient path enumeration, we propose this second part of numerical study of the Sioux Falls network. Under the same link parameters as previous section, we only consider one O-D pair (1, 20) with a total of 125 paths. Although the number of paths is still small compared to the total number of viable paths, it is sufficient to serve the purpose of this section.

We perform the computation of the DUE 19 times, each with an increased travel demand ranging from 10^4 to 10^5. The numbers of utilized paths in the corresponding equilibrium solutions are plotted in Figure 20. It is observed that as the demand is relatively low, the increase of utilized paths is faster than when the demand is higher. When the demand is very hight, i.e. $Q_{12} \geq 7 \times 10^4$, the demand needs to increase significantly for the drivers to use a new path.

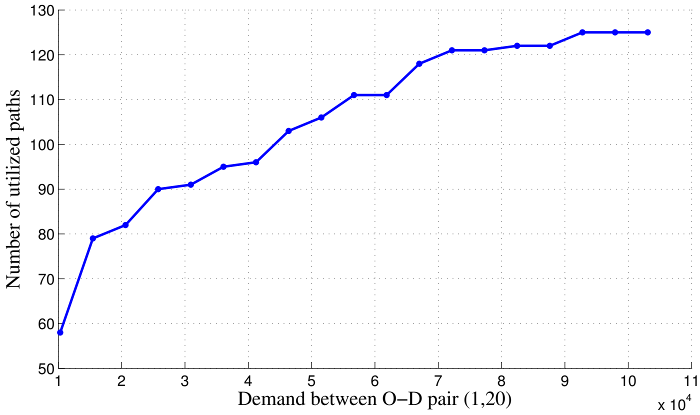


Figure 20: Number of utilized paths in the DUE against the total travel demand between (1, 20)

4.7. Performance of the network loading procedure

The fixed-point algorithm presented in this paper are implemented with several network loading models. The *link delay model* (LDM) is based on an explicit travel time function and the corresponding DNL procedure was approximated by a system of ODEs in Friesz et al. (2011). The *cell transmission model* (CTM) proposed in Daganzo (1994, 1995) is a widely used discrete-time traffic network model. These two network loading procedures, together with the one presented in this paper, are tested against different networks. The numerical performances are summarized in Table 2. We name our procedure "LWR-LH" with "LH" standing for Lax-Hopf formula. Notice that in Table 2, the results in Sioux Falls network is based on a choice of 10 O-D pairs and 200 paths for all three DNL procedure.

We observe from these numerical examples that the LWR-based network loading algorithm presented here is more efficient than its opponents for large-scale network. We interpret such result as follows: (1) the CTM model is a discrete model that requires two-dimensional space-time grid, this could be potentially inefficient in terms of number of operations performed and memory usage; (2) the LDM is based on solving a system of ODEs, the size of such system grows linearly with path numbers, when dealing with large-scale networks with over 1000 paths,

	Iterations			Computational time		
Network	**LDM**	**CTM**	**LWR-LH**	**LDM**	**CTM**	**LWR-LH**
2 arc, 2 nodes	14	14	15	18 s	24 s	38 s
6 arc, 5 nodes	11	12	12	26 s	32 s	75 s
19 arc, 13 nodes	14	10	10	152 s	159 s	130 s
76 arc, 24 nodes	14	13	13	33 min	52 min	15 min

Table 2: DUE fixed point iterations and computational time by problem type.

the ODE solver could slow down dramatically; (3) the LWR-based loading algorithm we're presenting here is a link-based algorithm, i.e. the formula (2.33) is performed only once for each link, in addition, the loading procedure is further accelerated by the closed-form representation of the PDE solutions and independence of spatial variables. The fixed-point algorithm with the proposed DNL procedure is performed on the Sioux Falls network with varying problem sizes, see Table 3 for a summary of performance, in particular, a below-linear growth of solution time with respect to the number of paths is reported.

Sioux Falls network	Iteration	Time
10 OD, 200 paths	13	15 min
20 OD, 360 paths	17	24 min
336 OD, 1371 paths	15	60 min
496 OD, 1941 paths	17	90 min

Table 3: Performance of fixed-point algorithm with LWR-based DNL procedure.

5. Conclusion

In this article, we propose a continuous-time network loading procedure based on the LWR model and a novel solution method based on the Lax formula. The DNL sub-problem is formulated as a DAE system without any partial derivatives and the solution to the PDE is expressed semi-analytically. The DNL model captures network queuing, route and departure time choices, as well as path delays. The proposed network loading procedure is implemented within the fixed-point algorithm for computing the dynamic user equilibrium. The numerical performance of the DNL procedure is tested extensively using various network examples, including the Sioux Falls network with nearly 500 OD pairs and 2000 paths.

The variational approach for conservation laws and Hamilton-Jacobi equations has become a recognized analytical and computational tool. It has the full potential to capture the formation and interaction of kinematic waves within a link. One drawback of the network model discussed in this paper is the absence of vehicle spillback. Thus the DNL procedure tends to underestimate path delays when the network is heavily congested. In order to capture the interaction of kinematic waves among links and address spillback, the analytical framework here needs to be extended to consider both upstream and downstream boundary conditions. This is not difficult to achieve using the Hamilton-Jacobi equation and Lax-Hopf formula. The reason is that the solution of the H-J equation with multiple boundary conditions can be viewed as the lower envelop of solutions that are each obtained with one boundary condition. See Aubin et al. (2008); Claudel and Bayen (2010a); Daganzo (2005) for detailed discussion. In Han et al. (submitted for publication), the variational method has been implemented to the network model based on the above discussion. By choosing appropriate Riemann Solvers at different junctions, the system dynamic can be again formulated as a DAE system. In the same paper, binary variables are used to indicate the traffic states (i.e. free-flow or congested) at beginning and end of each link. The phenomenon of spillback is explicitly modeled by these binary variables even in continuous-time.

Further understanding of the above network loading model in terms of solution existence for arbitrary network topology and continuity of the delay operator is the goal of future research.

References

Aubin, J.P., Bayen, A.M., Saint-Pierre, P., 2008. Dirichlet problems for some Hamilton-Jacobi equations with inequality constraints. SIAM Journal on Control and Optimization 47 (5), 2348-2380.

Bardi, M., Capuzzo-Dolcetta, I., 1997. Optimal Control and Viscosity Solutions of Hamilton-Jacobi-Bellman Equations, Birkhäuser.

Bressan, A., 2000. Hyperbolic Systems of Conservation Laws.The One Dimensional Cauchy Problem. Oxford University Press.

Bressan, A., Han, K., 2011. Optima and equilibria for a model of traffic flow. SIAM Journal on Mathematical Analysis 43 (5), 2384-2417.

Bressan, A., Han, K., 2011. Nash equilibria for a model of traffic flow with several groups of drivers. ESAIM: Control, Optimization and Calculus of Variations. DOI:10.1051/cocv/2011198.

Bressan, A., Han, K., submitted for publication. Existence of optima and equilibria for traffic flow on networks. Networks and Heterogeneous Media.

Claudel, C.G., Bayen, A.M., 2010. Lax-Hopf based incorporation of internal boundary conditions into Hamilton-Jacobi equation. Part I: Theory. IEEE Transactions on Automatic Control 55 (5), 1142-1157.

Claudel, C.G., Bayen, A.M., 2010. Lax-Hopf based incorporation of internal boundary conditions into Hamilton-Jacobi equation. Part II: Computational methods. IEEE Transactions on Automatic Control 55 (5), 1158-1174.

Coclite, G.M., Garavello, M., Piccoli, B., 2005. Traffic flow on a road network. SIAM Journal on Mathematical Analysis 36 (6), 1862-1886.

Courant, R., Friedrchs, K., Lewy, H., 1928. On the partial difference equations of mathematical physics. IBM Journal of Research and Development 11 (2), 215-234.

Dafermos, C.M., 1972. Polygonal approximations of solutions of the initial value problem for a conservation law. Journal of Mathematical Analysis and Applications 38 (1), 33-41.

Dafermos, C.M., 2010. Hyperbolic Conservation Laws in Continuum Physics. Springer.

Daganzo, C.F., 1994. The cell transmission model. Part I: A simple dynamic representation of highway traffic. Transportation Research Part B 28 (4), 269-287.

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

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

Daganzo, C.F., 2006. On the variational theory of traffic flow: well-posedness, duality and application. Network and heterogeneous media 1 (4), 601-619.

Evans, L.C., 2010. Partial Differential Equations. Second edition. American Mathematical Society, Providence, RI.

Friesz, T.L., 2010. Dynamic Optimization and Differential Games, Springer, New York.

Friesz, T.L., Bernstein, D., Suo, Z., Tobin, R., 2001. Dynamic network user equilibrium with state-dependent time lags. Network and Spatial Economics 1 (3-4), 319-347.

Friesz, T.L., Bernstein, D., Smith, T., Tobin, R., Wie, B., 1993. A variational inequality formulation of the dynamic network user equilibrium problem. Operations Research 41 (1), 80-91.

Friesz, T.L., Luque, J., Tobin, R.L., Wie, B.W., 1989. Dynamic network traffic assignment considered as a continuous time optimal control problem. Operations Research 37 (6), 893-901.

Friesz, T.L., Kim, T., Kwon, C., Rigdon, M.A., 2011. Approximate network loading and dual-time-scale dynamic user equilibrium. Transportation Research Part B 45 (1), 176-207.

Friesz, T.L., Mookherjee, R., 2006. Solving the dynamic network user equilibrium with state-dependent time shifts. Transportation Research Part B 40 (3), 207-229.

Garavello, M., Piccoli, B., 2006. Traffic Flow on Networks. Conservation Laws Models. AIMS Series on Applied Mathematics, Springfield, Mo..

Godunov, S.K., 1959. A difference scheme for numerical solution of discontinuous solution of hydrodynamic equations. Math Sbornik 47 (3), 271-306.
Han, K., Friesz, T.L., Yao, T., in press. A partial differential equation formulation of the continuous-time Vickrey's model: methodology and theoretical analysis. Transportation Research Part B.
Han, K., Friesz, T.L., Yao, T., in press. A partial differential equation formulation of the continuous-time Vickrey's model: computational method and application. Transportation Research Part B.
Han, K., Friesz, T.L., Yao, T., submitted for publication. Existence of simultaneous route and departure choice dynamic user equilibrium. Transportation Research Part B.
Han, K., Piccoli, B., Friesz, T.L., Yao, T., submitted for publication. A continuous-time link-based kinematic wave model for dynamic traffic networks. arXiv:1208.5141v1.
Han, K., Yao, T., Friesz, T.L., submitted for publication. Lagrangian-based hydrodynamic model: freeway traffic estimation. Transportation Research Part B.
Holden, H., Risebro, N.H., 1995. A mathematical model of traffic flow on a network of unidirectional roads, SIAM Journal on Mathematical Analysis 26 (4), 999-1017.
Holden, H., Risebro, N.H., 2002. Front Tracking for Hyperbolic Conservation Laws. Springer.
Jin, W.-L., 2010. Continuous kinematic wave models of merging traffic flow. Transportation Research Part B 44 (8-9), 1084-1103.
Jin, W.-L., Zhang, H.M., 2003. On the distribution schemes for determining flows through a merge. Transportation Research Part B 37 (6), 521-540.
Lax, P.D., 1957. Hyperbolic systems of conservation laws II. Communications on Pure and Applied Mathematics 10 (4), 537-566.
Lax, P.D., 1973. Hyperbolic systems of conservation laws and the mathematical theory of shock waves, SIAM.
Le Floch, P., 1988. Explicit formula for scalar non-linear conservation laws with boundary condition. Mathematical Models and Methods in Applied Sciences 10 (3), 265-287.
Lebacque, J., 1996. The Godunov scheme and what it means for first order traffic flow models. In: Proceedings of the 13th International Symposium on Transportation and Traffic Theory, 647-678.
Lebacque, J., Khoshyaran, M., 1999. Modeling vehicular traffic flow on networks using macroscopic models, in Finite Volumes for Complex Applications II, 551-558, Hermes Science Publications, Paris.
LeVeque, R.J., 1992. Numerical Methods for Conservation Laws. Birkhäuser.
Lighthill, M., Whitham, G., 1955. On kinematic waves. II. A theory of traffic flow on long crowded roads. Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences 229 (1178), 317-345.
Lo, H., Szeto, W., 2002. A cell-based variational inequality formulation of the dynamic user optimal assignment problem. Transportation Research Part B 36 (5), 421-443.
Merchant, D.K., Nemhauser, G.L., 1978a. A model and an algorithm for the dynamic traffic assignment problem. Transportation Science 12 (3), 183-199.
Merchant, D.K., Nemhauser, G.L., 1978b. Optimality conditions for a dynamic traffic assignment model. Transportation Science 12 (3), 200-207.
Newell, G.F., 1993. A simplified theory of kinematic waves in highway traffic, part I: General theory. Transportation Research Part B 27 (4), 281-287.
Nguyen, S., 1984. Estimating origindestination matrices from observed flows. In: Florian, M. (Ed.), Transportation Planning Models. Elsevier Science Publishers, Amsterdam, 363380.
Nie, X., Zhang, H.M., 2005. A comparative study of some macroscopic link models used in dynamic traffic assignment. Networks and Spatial Economics 5 (1), 89-115.
Nie, Y. and Zhang, H.M., 2008. A variational inequality formulation for inferring dynamic origin-destination travel demands. Transportation Research Part B 42 (7-8), 635-662.
Ran, B., Boyce, D., LeBlanc, L., 1993. A new class of instantaneous dynamic user optimal traffic assignment models. Operations Research 41 (1), 192-202.
Ran, B., Boyce, D., 1996. Modeling Dynamic Transportation Networks: An Intelligent Transportation System Oriented Approach. Springer-Verlag, New York.
Richards, P.I., 1956. Shockwaves on the highway. Operations Research 4 (1), 42-51.
Vickrey, W.S., 1969. Congestion theory and transport investment. The American Economic Review 59 (2), 251-261.
Wu, J.H., Chen, Y., Florian, M., 1998. The continuous dynamic network loading problem: a mathematical formulation and solution method. Transportation Research Part B 32 (3), 173-187.
Xu, Y.W., Wu, J.H., Florian, M., Marcotte, P., Zhu, D.L., 1999. Advances in the continuous dynamic network loading problem. Transportation Science 33 (4), 341-353.
Zhang, H., 2001. Continuum flow models, traffic flow theory, a state-of-the-art report. Transportation Research Board, original text by Reinhart Kuhne and Panos Michalopoulos.
Zhu, D.L., Marcotte, P., 2000. On the existence of solutions to the dynamic user equilibrium problem. Transportation Science 34 (4), 402-414.

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