Routing on Traffic Networks Incorporating Past Memory up to Real-Time Information on the Network State

Alexander Keimer\textsuperscript{1} Alexandre Bayen\textsuperscript{2}

\textsuperscript{1} Institute of Transportation Studies, UC Berkeley, Berkeley, US, 94720; email: keimer@berkeley.edu
\textsuperscript{2} Institute of Transportation Studies, UC Berkeley, Berkeley, US, 94720; email: bayen@berkeley.edu

Keywords
traffic flow, routing in real time, ordinary delay differential equations, dynamic traffic assignment, real-time information, existence of solutions, uniqueness of solutions

Abstract
We discuss the present routing algorithms for the dynamic traffic assignment (DTA) problem assigning traffic flow in a given road network as realistically as possible. We present a new class of so-called routing operators which route traffic flow at intersections based on real-time information about the status of the network or on historical data. These routing operators thus cover the distribution of traffic flow at all possible intersections. To model traffic flow on the links, we use a macroscopic well-known ordinary delay differential equation. We prove the existence and uniqueness of the solutions of the resulting DTA for a rather broad class of routing operators.

This new routing approach is required and justified by the increased usage of real-time information on the network provided by map services, changing the “laws of routing” significantly. These map and routing services have a huge impact on the infrastructure of cities so that a more precise mathematical description of the emerging new traffic patterns and effects becomes crucial for understanding and improving road and city conditions.
1. Introduction

In recent years, new routing and map services providing (allegedly, almost) real-time information on the status of a traffic network have increased in usage. Based on this information these services suggest “shortest routes” and change and will continue to change the routing behavior of all traffic participants on a given road network all over the world (1, 2, 3, 4, 5).

Due to these changes, well-understood and mathematically elegant models like convex optimization problems for determining Nash-based traffic assignments (6, 7) might be less precise or appropriate, as they were not meant to incorporate changes of routes during the routing process over time.

In this contribution, we will present the current status of mathematical modelling of the routing of traffic flow in networks macroscopically, i.e., as time- (and space-) dependent continuous density. We will explain most of the chosen approaches dealing with the so-called dynamic traffic assignment (DTA). We will then introduce a broad class of routing operators and show their well-posedness on the network for specific macroscopic dynamics, represented by a system of ordinary delay differential equations. In the considered context, well-posedness means that the system of differential equations coupled with the named routing operators admits a (unique) solution. This is not straightforward as the routing depends on the solution in real time, and causes a coupling between flow allocation and state of the system. To our knowledge, such a broad and rigorous approach has not been made in the literature, in particular when the routing itself will depend on the network status in real time, making it necessary to address the well-posedness by means of a fixed-point argument.

1.1. Structure of the article

In Section 2, we present the current state of research with an emphasis on the difference between link and node dynamics (the terminology is explained in the corresponding sections). We start with Section 2.1 and explain macroscopic traffic flow models using ordinary differential equations (ODE) in Section 2.1.1 and partial differential equations (PDE) in Section 2.1.2. In Paragraphs 2.1.1.1 and 2.1.2.1 we discuss the challenges when applying the named models to networks.

After presenting the dynamic models involved in describing traffic flow on links with proper node models, we discuss in Section 2.2 different routing approaches which distribute the flow following specific rules/laws. In Section 2.2.1 we briefly explain the archetypes of probabilistic routing approaches or routing operators located at intersections, followed in Section 2.2.2 by the time-dependent Wardrop’s principles sometimes also called user equilibrium (UE) and by optimal control approaches in Section 2.2.3 to obtain the “best” routing for a specific objective function.

Section 3 introduces a new modelling approach, routing operators, reacting to traffic situations in up to real time. Having defined the needed notation and network structure in Section 3.1 and Section 3.2 we introduce the considered link dynamics, an ordinary delay differential equation in Section 3.3 present the routing operators in Section 3.4 and then, as a whole, the resulting dynamic traffic assignment problem in Section 3.5. In Section 3.6 we investigate the well-posedness of the entire time-continuous routing problem by means of a fixed-point argument and conclude the section with some remarks about shortest path assignments in Section 3.7. We instantiate some well-known routing operators in the literature in Section 3.8 and discuss the applicability of our theory to these
operators.

The contribution is concluded in Section 4 with suggestions on future research directions.

2. State of the art

Many different approaches to the dynamic traffic assignment (DTA) problem have been considered in the literature. Usually, the problem is formulated on a network with links and nodes representing the corresponding roads and intersections of the considered traffic network.

In the following, we distinguish between the link dynamics, meaning the dynamics which traffic flow follows when no intersections are present, and node dynamics, which disperse incoming traffic flow according to predefined policies or “laws of routing.” For the link dynamics, we will only consider macroscopic traffic flow models, that is models describing the flow via density and not single traffic participants.

Depending on the link dynamics, node dynamics become more or less difficult. For instance, consider an intersection where all out-going links are fully congested. Then, flow intending to enter the intersection cannot pass to the leaving links and would thus “spill back” on the entering links. For a rather realistic traffic flow and DTA model this is a crucial requirement as otherwise flow will smoothly pass through the network. Even when travel time increases there will never occur any gridlock situation, although in reality this is one of the most important effects one would like to model and better understand.

but from a modelling point of view it makes the underlying equations more complex and harder to handle. Let us first discuss time-continuous macroscopic modelling of traffic flow on links in Section 2.1.

2.1. Link and node dynamics

In this section, we will describe different time-continuous macroscopic traffic flow models. For a general overview, we refer to [8, 9, 10].

2.1.1. ODE models. ODE models for traffic flow are, for instance, presented in [11, 12, 13] and have the general structure for a given $T \in \mathbb{R}_{>0}$

$$x'(t) = f(x(t), t), \quad t \in [0, T], \quad x(0) = x_0$$

with a proper function $f : \mathbb{R} \to \mathbb{R}$ to be specified, $x(t)$ being the density at time $t \in [0, T]$, and $x_0 \in \mathbb{R}_{\geq 0}$ being the initial density. Usually, the model class is written in greater detail:

$$x'(t) = \text{inflow}[x](t) - \text{outflow}[x](t), \quad t \in [0, T], \quad x(0) = x_0 \quad 1.$$ 

where inflow$[x]$ indicates the time-dependent inflow onto a considered link and outflow$[x]$ the corresponding outflow from a given link. Both inflow and outflow might also depend on the density $x$ of the considered link indicated in the notation. When interpreting the model with regard to an entire network, one might also prescribe additional constraints to make the model more realistic, although mathematically, these constraints can actually cause issues of well-posedness. Assume for instance that one adds box constraints – a capacity constraint on a specific link – and additional flow enters this link. Then, when there is no other modelling taking care of the flow that cannot be processed onto the considered link,
the model is not well-posed, meaning that the underlying equations do not have a solution. This can be overcome by defining a proper node model as discussed in Paragraph 2.1.1.1.

In addition, even though these models have a notion of traffic density, there is no notion of travel time (as one cannot state how long a certain amount of flow takes to pass a link), so one needs to come up with objective functions that compute a “type” of travel time based on the given density (see for instance (7) for how this is done for the stationary traffic assignment). However, this is so far confusing, as the model already contains time $t \in [0, T]$, so travel time should be directly defined in terms of this time.

Another approach is to place queues at the exits/entries of links, where the queuing size is modelled by ordinary differential equations. Several different models exist for this purpose; we refer to (14, 15, 16, 17) and to (18, 19, 20), where different generalization based on the Vickrey model (21) are considered. In addition, now one can impose constraints on links so that if a specific density is reached, the corresponding buffer increases. For short links where real flow dynamics might not develop, this model class might be fortunate. In a simplified version following (15), the model can be written as

$$q'(t) = p(t - t_0) - \begin{cases} \min \{C, p(t - t_0)\} & q(t) = 0 \\ \frac{q(t)}{C} & q(t) > 0 \end{cases} \quad t \in [0, T]$$

$q(0) = q_0$.

Thereby, $q : [0, T] \to \mathbb{R}_{\geq 0}$ denotes the queue size (or density), $t_0 \in \mathbb{R}_{\geq 0}$ the free-flow travel time, $C \in \mathbb{R}_{\geq 0}$ the road capacity, and $q_0 \in \mathbb{R}_{\geq 0}$ the initial queuing size or initial density. In addition, the function $p : [-t_0, T] \to \mathbb{R}_{\geq 0}$ represents the inflow, and according to Eq. (1), the function $\min \{C, p(t - t_0)\} q(t) = 0$ is the outflow. Finally, actual travel time $\tau(t)$ can be computed as $\tau(t) = t_0 + \frac{q(t - t_0)}{C}$.

The next step is to consider ordinary differential equations with delay. One instantiation of these equations is also used for the analysis in the present article; the reader is referred to Section 3.3 [Definition 4]. This model class has been extensively discussed in (22, 23, 24, 25, 26). The advantage is that even though it is still an ODE model, there is now an intrinsic travel time as a function of the state density, and that travel time – the delay of the ordinary differential equation – increases when there is higher density on the link. One disadvantage of this model class is that only when the delay depends affine-linearly on the density (25, Theorem 3.2) can a unique solution for every type of essentially bounded inflow be guaranteed, limiting to some extent its application (25, Theorem 3.2, Chapter 4). Of course, one can then again prescribe state constraints on the dynamics to avoid links with high density; however, as pointed out before, this might have impacts on the well-posedness of the system, as a solution might then be nonexistent after all when inflow into the system is high or too many links merge into one so that flow cannot be processed anymore.

2.1.1.1. Node models. For the ODE models modelling density and for the delayed versions, there is no need for a node model. The leaving density from one link can be directly passed to the following links according to a given routing behavior, as discussed in Section 2.2 as these models do not possess spillback. In case there are additional capacity constraints stated, further adjustments have to be made to have a well-defined system – for instance, introducing additional queues at the intersections or having basically large enough queues in case one already uses queuing models for the links.
2.1.2. PDE models. The second and most of the time more “realistic” approach to modelling traffic flow on roads as a continuous density over time also takes into account the spatial component of the link representing the position on the road. This has been done in (27, 28) following the fundamental theory of conservation of vehicles and a fluid type of approach. The corresponding governing equation, the LWR PDE (after Lighthill, Whitham, and Richards), reads as

\[ \partial_t \rho(t, x) + \partial_x f(\rho(t, x)) = 0 \]

\[ (t, x) \in (0, T) \times X \]

boundary conditions prescribed at \( \partial X \)

where \( \rho(t, x) \) is the traffic density at space-time coordinate \( (t, x) \in (0, T) \times X \), with \( X \subset \mathbb{R} \) being an interval and initial datum \( \rho_0 \) prescribing the density at time \( t = 0 \). The function \( f \) is called the flux function and is usually chosen as \( f(y) := y \cdot v(y) \), \( y \in \mathbb{R} \), with a monotone decreasing velocity function \( v : \mathbb{R} \to \mathbb{R} \). A rather famous flux function is the Greenshields function (29) with \( v(y) = 1 - y \) (when density is scaled to be between 0 and 1), but there exist many more fluxes and corresponding velocity functions dependent on the considered road setup (see (26)). The choice of velocity function is usually made in such a way that flux increases with density until a turning point when the density reaches a critical limit, and after that, the flux decreases in terms of the density until it reaches a minimum or even zero.

Mathematically, the prescribed LWR is a so-called quasi-linear scalar hyperbolic conservation law which can develop shocks and rarefaction waves and also deal with discontinuous initial data. It is thus significantly more appropriate for modelling traffic flow when spillback is important than the ODE models previously described in [Section 2.1.1]. The existence and uniqueness of solutions for these equations are non-trivial even for the Cauchy problem meaning without boundary datum, \( X = \mathbb{R} \), and have been subject to studies and solved in (30, 31, 32, 33, 34) by introducing entropy conditions to single out the physically relevant unique weak solution among the infinitely many possible weak solutions. The presented equation has also been studied as the limit of the well-established microscopic “follow the leader” ordinary differential equations when the number of vehicles approaches infinity (homogenization), see for instance (35). It is worth mentioning that the previously defined model class can be transformed into another, so-called Hamilton-Jacobi PDEs, where semi-explicit solution formula are available based on a finite-dimensional optimization problem for every considered point in space-time (36, 37, 38, 39, 40).

The extension of the previous PDE to nonlocal traffic flow models has been investigated, for instance in (41, 42, 43). Here, the previously named velocity \( v \) depends not on \( \rho(t, x) \) but on the averaged density ahead, that is on \( \frac{1}{\eta} \int_x^{x+\eta} \rho(t, s) \, ds \), \( (t, x) \in (0, T) \times \mathbb{R} \) for a given \( \eta \in \mathbb{R}_{>0} \). The advantages of this type of modelling is that it is more realistic than the previously chosen “local” LWR model and even in its aggregated form – as density – closer to microscopic behavior due to its forward-looking parameter \( \eta \in \mathbb{R}_{>0} \). It is also mathematically interesting: As there is no entropy condition required, a weak solution is by itself unique.

One disadvantage of the previously introduced LWR model class is that velocity behaves purely as a function of the density. For greater realism, one might actually want to model the velocity by its own dynamics for instance, to model phantom shocks, which are very common in high-density traffic flows, etc..
This is why in the literature, second-order traffic flow models are also considered. We only introduce the inhomogeneous ARZ PDE (after Aw, Rascle, and Zhang) \((44, 45)\) for \((t, x) \in (0, T) \times X\) here:

\[
\begin{align*}
\partial_t \rho(t, x) + \partial_x (\rho(t, x) u(t, x)) &= 0 \quad 2. \\
\partial_t (u(t, x) + h(\rho(t, x))) + u(t, x) \partial_x (u(t, x) + h(\rho(t, x))) &= \frac{1}{\tau} (U_{eq}(\rho(t, x)) - u(t, x)) \quad 3.
\end{align*}
\]

\[\rho(0, x) = \rho_0(x)\]
\[u(0, x) = u_0(x)\]

boundary conditions prescribed at \(\partial X\).

The traffic density at time-space coordinate \((t, x) \in (0, T) \times X\) is denoted by \(\rho(t, x)\) and the velocity by \(u(t, x)\). Equation (2) denotes the conservation of vehicles and is inspired by the LWR PDE, with the difference that the velocity function now is not an explicit function of the density but follows its own dynamics in Eq. (3). There, the term \(h: \mathbb{R} \to \mathbb{R}\) with \(h' > 0\) represents the pressure term stemming from fluid dynamics; however, in traffic flow, it is more reasonable to call it a “hesitation function” (compare \((46)\)). \(U_{eq}: \mathbb{R} \to \mathbb{R}\) denotes the equilibrium velocity and \(\tau \in \mathbb{R}_{>0}\) the reaction time of drivers. The considered set of equations can be posed as a system of conservation laws and is thus a quasi-linear system of hyperbolic conservation laws.

For \(X = \mathbb{R}\), meaning again as a pure Cauchy problem without a boundary datum, this model has been studied extensively in \((47, 48, 49, 50)\) for questions on the existence, uniqueness, regularity, and stability of solutions; for locally constrained flow in \((51, 52)\); and for phase transition in \((53)\). Of course, second-order multi-class/commodity models have also been studied in \((54)\) and models with creeping in \((55)\). For general conservation law models, we refer to \((56)\), where questions on existence and uniqueness are discussed in significantly more generality.

2.1.2.1. The node model for PDEs – a (modelling) challenge. In Section 2.1.2 when \(X \subset \mathbb{R}\) is a finite interval, a boundary datum has to be prescribed. However, this boundary datum might not be attained, as the road might already be fully congested, or higher flow might want to enter a given link than the link can take. This is why the boundary datum needs to be described in a weak sense. We refer to \((57, 58, 59, 60)\). For conservation of flow when the boundary datum is not attained, one needs to keep track of the flow not entering the considered link. This can be obtained by placing buffers at the intersections, which make sure that flow is conserved, see \((61)\) for supply chain modelling with buffers and \((62, 63, 64, 65)\) for traffic flow modelling with buffers and the LWR PDE. When these buffers have a finite capacity, they also need to allow spillback to the incoming roads. Thus, it becomes possible for congestion to spread over a node. They also enable the corresponding system of conservation laws to depend in a continuous way on the input parameters.

When not using buffers, one can consider the Riemann problem at the intersection, as has been done in \((66, 67, 68)\) for the LWR PDEs and in \((69)\) for the ARZ model. The problem of the Riemann-solver approach is that the solution is not necessarily continuously dependent on the input datum, and additional equations have to be postulated to obtain a reasonable node model (maximizing throughput; see \((70)\) for maximizing a given objective over the nodes). These dynamics as well as the first-order dynamics have extensions to multi-commodity flows on networks \((71)\).
Having defined the node dynamics, the missing piece for the dynamics to be fully determined is how traffic flow is actually routed at the considered intersections.

### 2.2. Routing

Having defined the corresponding link models via either PDEs or ODEs, and in case the proper node models are needed, we need to discuss how routing is actually realized.

#### 2.2.1. Routing operators

A rather straightforward approach to prescribing routing is to introduce flow ratios at all intersections. These flow ratios might depend on the solution itself, see in particular Section 3, or they might be realized by fixed turning ratios trained on existing data or stochastic modelling. Even though in our analysis we focus on macroscopic modelling, many suggested approaches among route choice models for individuals can also be used for macroscopic modelling. We refer for an overview to (74), to (75) for a fuzzy route choice model, to (76, 77, 78, 79, 80, 81) for general stochastic route choice models, and to (82) for a cognitive cost route choice model. Some of the route choice models also take into account the state of the network in real time (83, 84). For the validation of route-choice models, we refer to (85, 86, 87). Route choice for stationary models has also been considered in (88, 89, 90, 91).

Shortest path–based routing models in real time are quite often used; however, in continuous dynamics, these shortest path–based models might not be well posed, as explained later in [Remark 1](#) for a specific class of models.

#### 2.2.2. Time-dependent Wardrop

Wardrop’s principles are rather famous for describing how flow might get routed in a road network. We introduce them in the following (see (92, p. 345)):

**Wardrop’s first principle:** The journey times on all the routes actually used are equal to each other and less than those which would be experienced by a single vehicle on any unused route.

**Wardrop’s second principle:** The average journey time is a minimum.

The first principle defines a rule on how individuals might route themselves in a network and can be interpreted as a “type” of Nash equilibrium. Individuals chose their route so that their travel time is minimal taking into account that everyone else does the same. This requires that everyone knows about all origin destination demand. It can be motivated by the argument that over a series of days, etc., drivers figure out the proper routes as they obtain more knowledge about the state of the network and their impact on it when diverting from specific routes. In contrast, the second principle suggests a routing based on a “social optimum.” Clearly, the second problem can be approached by an optimal control approach in the routing, as also suggested in Section 2.2.3. However, both principles allow some interpretations when it comes to time dynamics. Also the reasonability of these approaches has to be discussed, as they require one to possess quite precise information about the state of the network to make proper routing decisions. Usually, routing based on a time-dependent Wardrop’s condition is called dynamic user equilibrium. Mathematically, this can be cast as a variational inequality or as a complementary condition. This has been investigated in detail for different user equilibria in (93, 94, 95, 96, 97, 98, 99), where variational formulations for different dynamic user equilibria are given and variational inequalities
obtained. In (100, 101), similar approaches are undertaken with a solid analysis on the well-posedness of solutions and numerical studies. For general variational inequalities, we refer to (18).

2.2.3. Optimal Control. The famous M-N model (11, 12) (after Merchant and Nemhauser) was one of the first models used in (102) to model the dynamic traffic assignment by using an optimal control framework. Having defined link models via ODEs, the authors present an optimal control framework similar to that for the stationary traffic assignment (7), where the routes are optimized over the entire considered time horizon, minimizing a given objective function. Two drawbacks of this approach are evident: First, this class of link models does not explicitly possess a travel time, so one has to come up with an objective which realizes this travel time in a reasonable way. The other drawback consists of the fact that an optimal control problem over the entire considered time horizon is stated, although for the purpose of real time application, such information about when people leave might not be available. The presented formulation is path-based, so the optimal paths are chosen when flows enter the network, and no rerouting will occur. A similar approach is considered in (103). When the link dynamics are realized via PDEs, an optimal routing is considered in (104, 105, 106) for the LWR (27, 28) traffic flow model (also with instantaneous control) and in (107) for a specific class of nonlocal conservation laws.

3. Routing operators based on the network state

As we aim to propose a mathematical framework capable of handling the change of routing due to routing applications using real time information (compare Section 1), we require the following properties on routings in the network: The routing should be

1. presented in a very broad way, keeping as much generality as possible and allowing the results to be applicable to very different routing scenarios,
2. able to distinguish between different types of flow (flow with more information, different types of vehicles, etc.),
3. capable of reacting to a change in the network in real time,
4. instantiated so that it can depend in up to real time on the solution of the entire network,
5. dependent on already passed information or on delayed information,
6. dependent on external factors, such as the closing down of specific roads at specific times,
7. located at the different intersections explicitly as a functional expression.

All these mentioned properties will be satisfied by the model introduced and exploited in Section 3

Although our aim and interest lie mainly in the routing, for a mathematical analysis, we have to consider the full system consisting of the link model and node model, as described in Section 2.1 and routing, as described in Section 2.2 From the node model, we must have a travel time and an increase in travel time when the road is more congested. In addition, we require a multi-commodity model capable of handling different types of flow, different types of destinations, and more. As we do not want to restrict our routing operator, we require that the node model be able to assign – at every time considered – any flow on all outgoing links and to distinguish between different commodities.

This is why we chose for the link dynamics an ordinary delay differential equation, as for instance considered in (25) and mentioned in Section 2.1.1 This equation does not
model spillback, so we can use as the node model the trivial model of assigning just the incoming flow according to the routing. This missing feature is acceptable, as we mainly focus on routing. It still has the property that travel time changes with respect to link density, so it is significantly complex for routing problems. Of course, a similar framework taking the proper node dynamics can be also introduced for more advanced link dynamics, as prescribed in Section 2.1.2. However, in this case, the node model will not be trivial and needs to satisfy a Lipschitz-continuity with respect to the input parameters, making the analysis more complex.

Most of the content in this chapter has been discussed and explored in detail in (72).

3.1. Notation and more

We will require the following function spaces:

**Definition 1 (Function spaces).** For \( p \in [1, \infty] \) and \( T \in \mathbb{R}_{>0} \) we define the following Banach spaces

\[
L^p((0, T)) := \{ f : (0, T) \to \mathbb{R} \text{ Lebesgue-measurable} : \|f\|_{L^p((0, T))} < \infty \}
\]

with

\[
\|f\|_{L^p((0, T))} := \left( \int_{(0,T)} |f(s)|^p \, ds \right)^{\frac{1}{p}}, \quad p \in [1, \infty), \quad \|f\|_{L^\infty((0, T))} := \text{ess-\sup}_{t \in (0, T)} |f(t)|
\]

the space of continuous functions on the interval \([0, T]\) as

\[
C([0, T]) := \{ f : [0, T] \to \mathbb{R} : f \text{ is continuous} \}
\]

with the corresponding norm

\[
\|f\|_{C([0, T])} := \max_{t \in [0, T]} |f(t)|, \quad f \in C([0, T])
\]

and the space of Lipschitz-continuous functions as

\[
W^{1,\infty}((0, T)) := \left\{ f \in L^\infty((0, T)) : \sup_{x, y \in [0, T], x \neq y} \left| \frac{f(x) - f(y)}{x - y} \right| < \infty \right\}
\]

with norm

\[
\|f\|_{W^{1,\infty}((0, T))} := \|f\|_{L^\infty((0, T))} + \|f'\|_{L^\infty((0, T))}
\]

where \( f' \) exists almost everywhere by Rademacher’s theorem so that the previous definition makes sense. Of course, vectorized versions of the previously defined function spaces are defined appropriately.

3.2. Network

As we will need to describe the dynamic traffic assignment problem with regard to the underlying network, we start by defining a network and related attributes, like paths, destination, origin nodes, and more.
**Definition 2 (Network and Paths).** We call a directed graph $G = (V, A)$ where $V$ is a finite set with $|V| \in \mathbb{N}_{\geq 1}$, with nodes $v \in V$ and arcs $a \in A \subseteq V \times V$ a network. Furthermore, we define for $v \in V$ the sets of incoming arcs and outgoing arcs

$$A_{\text{in}}(v) := \{(\tilde{v}, v) \in A : \tilde{v} \in V\}, \quad A_{\text{out}}(v) := \{(v, \tilde{v}) \in A : \tilde{v} \in V\}.$$  

We define the set of paths $P^{v,d}$ between two nodes $(v, d) \in V^2$ without cycles as:

$$P^{v,d} := |V|^2 \bigcup_{k=1} \left\{ p \in A^k : \forall i, j \in \{1, \ldots, k+1\}, i \neq j : v_i, v_j \in V : v_i \neq v_j, v_1 = v, v_{k+1} = d, p = ((v_i, v_{i+1}))_{i \in \{1, \ldots, k\}} \right\}.$$  

For every path $p := (p_1, \ldots, p_k)^T \in P^{v,d}$ with $k \in \mathbb{N}_{\geq 1}$ we define its length by $\text{len}(p) := k$.

In the network we specify source nodes as

$$O \subset V$$

and destination nodes as

$$D \subset \left\{ d \in V : \exists v \in O \text{ s.t. } P^{v,d} \neq \emptyset \right\}.$$  

We define $OD$ as the set of origin/source-destination pairs by

$$OD := \left\{ (v, d) \in O \times D : P^{v,d} \neq \emptyset \right\}.$$  

Finally, we define for $d \in D$ the set $OP^d$ of arcs which are part of paths starting from an arbitrary node $\tilde{v} \in O$ and ending in $d$ if $(\tilde{v}, d) \in OD$:

$$OP^d := \bigcup_{\ell=1}^{\lfloor |A|/\ell \rfloor} \left\{ p_\ell \in A : \exists \tilde{v} \in O \text{ s.t. } (\tilde{v}, d) \in OD, p_\ell \in P^{\tilde{v},d}, \ell \leq \text{len}(p) \right\}.$$  

and the outgoing arcs for a node $v \in V$ from which destination $d \in D$ is reachable as

$$A_{\text{out}}^d(v) := \left\{ a \in A : \exists \tilde{v} \in V \text{ with } a = (v, \tilde{v}) \land a \in OP^d \right\}.$$  

**Assumption 1 (Feasible network).** The network $G$ in Definition 2 contains at least one $OD$-pair, i.e. $|OD| > 0$.

### 3.3. The link model

In this section, we introduce the link dynamics which we consider to use throughout this article and which have been mentioned in Section 2.1.1. To this end we require different commodities representing the mathematical way to express/model different types of traffic flow, different types of “information aware” flow, using different routing systems or no routing suggestions at all, etc. As we want to distinguish between different commodities, we also define

**Definition 3 (Multi-commodity).** For a given $n \in \mathbb{N}_{\geq 1}$ we define the multi-commodity-set as $C := \{1, \ldots, n\}$. 

---

10 A. Keimer and A. Bayen
The link dynamics read as follows:

**Definition 4** (Link delay model for multi-commodity with affine linear delay on a single link). Denote by the superscript \((c,d) \in C \times D\) a tuple containing commodity type and destination node as given in Definition 3. Assume that \(x : [0, T] \rightarrow \mathbb{R}_{\geq 0}^{1 \times |D|}\) represents the flow and let \(u : [0, T] \rightarrow \mathbb{R}_{\geq 0}^{1 \times |D|}\) be the inflow. Then, for a free flow travel time \(b \in \mathbb{R}_{>0}\) and congestion factor \(h \in \mathbb{R}_{\geq 0}\) we consider the following system of delay ODE

\[
\begin{align*}
\dot{x}(t) &= u(t) - g[u, x, \dot{x}](t) & t \in [0, T] \\
x(0) &= 0 \\
x(t) &= \sum_{c \in C} \sum_{d \in D} x^{c,d}(t) & t \in [0, T] \\
\tau[x](t) &= b + hx(t) & t \in [0, T] \\
\rho[x](t) &= t + \tau[x](t) & t \in [0, T] \\
g[u, x, \dot{x}](t) &= \begin{cases} 
0, & \text{a.e. } t \in [0, b) \\
\frac{w[\rho[x]^{-1}(t)]}{1 + hx[\rho[x]^{-1}(t)]}, & \text{a.e. } t \in [b, T + \tau[x](T)]
\end{cases}
\end{align*}
\]

where \(\rho[x]^{-1}\) is the inverse function of \([0, T] \ni t \mapsto \rho[x](t)\).

Some comments are in order to give a better understanding of the previously defined model. **Definition 4** presents the link dynamics. We need to have it in a vectorized form as we will need to keep track of different types of flow: On the one hand, flow heading to different destinations \(d \in D\) has to be kept track of individually, and also flow which follows different routing policies, for instance, flow using information about the traffic system in real time vs. flow not using it, etc.

The dynamical process is presented in a coupled system of ordinary delay differential equations in **Equation (5)**. Thereby, \(u\) denotes the inflow onto the link (vectorized with different commodities and destinations), and \(g\) the outflow which is detailed below and will be a function of the density of the link. **Equation (6)** states that the link is empty when starting. **Equation (7)** denotes the cumulative density of the link, that is the density of all flows on the link, summarizing over all different commodities and destinations. **Equation (8)** denotes the delay caused by the traffic density. The higher the cumulative density \(x\) is, the larger the delay becomes. \(b\) can be seen as the free flow delay or travel time, when the road is empty and \(h\) is a tuning parameter impacting the influence of the travel time or delay with regard to the cumulative density. Based on this, **Eq. (9)** gives the time when an inflow entering at time \(t \in [0, T]\) actually exists the link, and **g** in **Eq. (10)** shows how the outflow is determined by means of the inflow and the travel time. The denominator stands for a spreading or concentration of flow due to higher or lower density and is needed for keeping the flow conserved.

As a solution to an ordinary delay differential equation with a delay depending on the solution itself, does not necessarily exist, we give the following

**Theorem 1** (Existence and Uniqueness of the link delay model with affine linear delay as presented in **Definition 4**). Let \(T \in \mathbb{R}_{>0}\) and the link-delay model with affine linear delay as in **Definition 4** be given and assume that \(u \in L^\infty\left([0, T]; \mathbb{R}_{\geq 0}^{1 \times |D|}\right)\) is given as well as \(b \in \mathbb{R}_{>0}\) and \(h \in \mathbb{R}_{\geq 0}\). Then, the delay ODE system in **Eqs. (5) to (10)** is well-posed and admits a unique Carathéodory solution \(x \in W^{1,\infty}\left([0, T]; \mathbb{R}_{\geq 0}^{1 \times |D|}\right)\).
Proof. The proof can be found in [29, Theorem 3.1, Theorem 3.2] and takes advantages of the delay character so that one can use an iterative approach in time for solving a series of delay equations on small time horizon and sticking them together.

As already pointed out, the advantage of this model is that there is no spill-back so that at every intersection one can assign as much flow to any of the out-going links as needed. This basically makes the normally required node model discussed in Paragraph 2.1.1.1 obsolete or trivial, and simplifies our analysis about routing as we can directly continue to the definition of a routing operator.

3.4. The routing operator

In this section, we define the routing operator. We start with the most general Definition 5 possible and restrict the routing operator later in Definition 7.

**Definition 5** (General routing operator). Let \( T \in \mathbb{R}_{>0} \) and let for \( v \in V \) be \( d \in D, c \in C \) and \( a \in A_{\text{out}}(v) \) be given as in Definition 2. Let \( \text{ext} \in L^\infty((0,T);\mathbb{R}^n_{\text{ext}}) \). Then, we call \( R_{c,d,a} \) routing operator iff

\[
R_{c,d,a} : L^\infty((0,T);\mathbb{R}^{|A||C||D|}_{\geq 0}) \times L^\infty((0,T);\mathbb{R}^n_{\text{ext}}) \to L^\infty((0,T);[0,1])
\]

such that for every \((x, \text{ext}) \in L^\infty((0,T);\mathbb{R}^{|A||C||D|}_{\geq 0}) \times L^\infty((0,T);\mathbb{R}^n_{\text{ext}})\)

\[
\sum_{a \in A_{\text{out}}(v)} \mathcal{R}_{c,d,a}^v(x, \text{ext})(t) \equiv 1 \quad t \in [0,T] \text{ a.e.}.
\]

As one can see from Definition 5, particular from Eq. (11), the routing operator is actually an operator taking into account the full solution on the network over the full time horizon considered. The routing operator carries a sub-index assigning flows to all exiting links \( a \in A_{\text{out}}(v) \) from a intersection node \( v \in V \). Even more, it can route differently for different types of commodities \( c \in C \) and destinations \( d \in D \), a reasonable assumptions when recalling that the OD pairs might vary significantly for different destinations \( d \in D \) and different types of flow might behave/getting routed differently. Finally, the routing can also be influenced by so called externalities. Indicated by Eq. (12) the routing operator has to conserve flow and thus routes all incoming flows with their different destinations and commodities to the out-going links at every time. The function values of the routing operator are assumed to be between 0 and 1 in Eq. (11).

Next, we specify the routing operators as the generality of routing operators provided in Definition 5 is too broad to obtain any results on existence or uniqueness of the system on the network. As it makes sense to consider routing operators, which use at time \( t \in [0,T] \) at most the traffic state \( x \) up to time \( t \) (see Section 1) we distinguish a Lipschitz-continuous routing which is capable of incorporating routing decisions made in up to real time from a delay-type routing operator, where only at finitely many points in time the decision can be made in real time and is otherwise delayed and no Lipschitz-continuity or only continuity is required.

Definition 7 requires the following projection operator which will be helpful in defining delayed routing.
Definition 6 (Projection mapping). Define for \( \alpha \in \mathbb{R} \) the projection mapping \( P_{[0, \alpha]} : \mathbb{R} \to \mathbb{R} \) on \([0, \alpha]\) by: \( \forall t \in \mathbb{R} : \quad P_{[0, \alpha]}(t) := \begin{cases} \min \{ \max \{0, t\}, \alpha \} & \text{if } \alpha \in \mathbb{R}_{\geq 0} \\ 0 & \text{else.} \end{cases} \)

Definition 7 (Routing Operators). Let the assumptions of Definition 5 be given. Then, we define under the assumption that all introduced routing operators satisfy Definition 5:

(A) L-Continuous Routing: We call \( \mathcal{R}L^{c,d}_a \) Lipschitz-continuous routing operator iff it is Lipschitz-continuous in the following sense: For a \( p \in (1, \infty) \) it holds

\[
\forall \text{ext} \in L^\infty((0,T);\mathbb{R}^{\text{ext}}) \quad \forall t \in [0,T] \quad \exists L \in \mathbb{R}_{\geq 0} \quad \forall x, \hat{x} \in C([0,T];\mathbb{R}^{|A||C||D|}) : \\
\left\| \mathcal{R}L^{c,d}_a \left[ x, \text{ext} \right] - \mathcal{R}L^{c,d}_a \left[ \hat{x}, \text{ext} \right] \right\|_{L^p((0,t))} \leq L \|x - \hat{x}\|_{C([0,T];\mathbb{R}^{|A||C||D|})}
\]

(B) Delay-type Routing: We call \( \mathcal{R}D^{c,d}_a \) a delay-type routing operator iff there exists for \( N^{c,d}_a \in \mathbb{N}_{\geq 1} \) a time vector \( t^{c,d}_a := \left( (t^{c,d}_a)_1, \ldots, (t^{c,d}_a)_{N^{c,d}_a} \right)^T \in (0,T)^{N^{c,d}_a} \) with \( (t^{c,d}_a)_{N^{c,d}_a} := T \) and

\[
(t^{c,d}_a)_i < (t^{c,d}_a)_{i+1} \quad \forall i \in \{1, \ldots, N^{c,d}_a - 1\}.
\]

a delay function \( \delta^{c,d}_a \in L^\infty((0,T);[0,T]) \) such that for every \( i \in \{1, \ldots, N^{c,d}_a - 1\} \) the following regularity and estimate on the delay \( \delta^{c,d}_a \) hold

\[
\delta^{c,d}_a \bigg|_{[0,(t^{c,d}_a)_i]} : = 0,
\]

\[
\delta^{c,d}_a \bigg|_{\left((t^{c,d}_a)_i, (t^{c,d}_a)_{i+1}\right)} \in C \left( \left( (t^{c,d}_a)_i, (t^{c,d}_a)_{i+1}\right) \right),
\]

\[
\left\| \delta^{c,d}_a \right\|_{C(\left( (t^{c,d}_a)_i, (t^{c,d}_a)_{i+1}\right))} \leq (t^{c,d}_a)_i
\]

and, finally, for a.e. \( t \in [0,T] \), \( \forall x \in C\left([0,T];\mathbb{R}^{|A||C||D|}\right) \) recalling Definition 6

\[
\mathcal{R}D^{c,d}_a \left[ x, \text{ext} \right](t) = \mathcal{R}D^{c,d}_a \left[ x \circ P_{[0,\delta^{c,d}_a(t)]}(\text{ext}) \right](t).
\]

Again, some explanations are in order. The Lipschitz-continuous operator is meant to be Lipschitz-continuous in \( L^p \), \( p \in (1, \infty) \) when measuring the network state in the uniform topology. As one can see from Item (A) the Lipschitz-continuity has to hold for all \( t \in [0,T] \) guaranteeing that it can only depend on the network state at time \( t \in [0,T] \) at most at time \( t \). As the solutions \( x \) on all links are at least Lipschitz-continuous – as long as the inflows are essentially bounded – we know that \( x \) is continuous and can actually measure in the continuous topology. As we will illustrate later, this Lipschitz-continuous dependency is not a too strong condition and is satisfied by a variety of different routing operators (see Section 3.8).

The Lipschitz-continuous dependency is needed as the real time dependency of the routing operator gives another coupling in any of the considered ODEs on the links. The inflow onto these links is a function of the routing operator and the routing operator by itself
Figure 1
Illustration of a delay function. The thick diagonal line represents the identity, the red graph shows an example of a general feasible delay function $\delta$. Furthermore, the green graph shows the case when a piecewise vanishing delay and the blue graph when a constant delay $\varepsilon \in \mathbb{R} > 0$ occurs.

Figure 2
Illustration of the flow information used by the routing operator. Every horizontal line represents the time line. The delay-type routing operator $R$ maps time points – which are represented on the upper horizontal lines – to time periods that do not occur later in time. This produced time periods – which are part of the lower horizontal lines – build the possible evaluation time periods for the flow $x$. Depicted are the constant $\varepsilon$-delay (left) and the piecewise vanishing delay (right).

again on the state of these links up to the present time. This is the reason why the problem has to be addressed by means of a fixed-point argument, for which – in case one wants to obtain uniqueness of solutions – the contraction mapping principle, Banach’s fixed-point theorem – is essential [108, Theorem 3.1]. We refer to Theorem 3. Also note that all routing operators are coupled with each other as the entire traffic state will change according to the routing operators and the routing operators will change their routing assignment based on the traffic state.

We also want to mention that a weaker form of routing operators is available in [72, Definition 2.12, C], where we only assume continuity. However, as this result is based on Schauder’s fixed-point theorem [108, Corollary 2.13], the solution cannot be expected to be unique, even more we can give examples where there would be infinitely many solutions. We thus renounce the presentation of these classes of routing operators.

The delayed routing in Item (B) does not need any continuity or even Lipschitz-continuity as it mainly acts or makes decisions based on the state of the network in a delayed sense. The additional complexity in Eq. (13) comes from the fact that we also want the possibility that the routing operator depends on the real time solution at at most finitely many points. This is also illustrated in Figs. 1 and 2 which has been borrowed from [72, Remark 2.14].

It is also worth mentioning that the presented approach does not need to know any inflow – origin-destination demand – into the network as it will only need information at time $t \in [0, T]$ and will then react properly on the state of the traffic state.

The previously defined routing operators will now be implemented into the DTA framework in Section 3.5.
3.5. The DTA framework incorporating information of the network state

In this Section 3.5, we present the entire DTA problem when the routing is performed by the routing operators introduced in Section 3.4.

**Definition 8** (Network formulation of the link-delay model with routing operators: DTA). Consider for a time horizon \( T \in \mathbb{R}_{>0} \) a network \( G = (V, A) \) as in Definition 3 with origins \( O \subseteq V \) and destinations \( D \subseteq V \) as in Definition 2. Let the inflow into the network for \( \hat{v} \in O \) be given as \( \hat{s}_v \in L^\infty ((0, T); \mathbb{R}_{\geq 0}^{\mid C\mid \mid D\mid}) \). Then, we pose the dynamics subject to the link delay model for multi-destination with routing as proposed in Definition 5 for all \( t \in [0, T] \) and \( a \in A \) with \( b_a \in \mathbb{R}_{>0} \) and \( h_a \in \mathbb{R}_{\geq 0} \) as:

\[
\dot{x}_a(t) = u_a(t) - g_a[u_a, x_a, \dot{x}_a](t) \quad t \in [0, T]
\]

\[
x_a(0) = 0
\]

\[
x_a(t) := \sum_{c \in C} \sum_{d \in D} x_a(c,d)(t) \quad t \in [0, T]
\]

\[
\tau_a[x_a](t) := b_a + h_a x_a(t) \quad t \in [0, T]
\]

\[
\rho_a[x_a](t) := t + \tau_a[x_a](t) \quad t \in [0, T]
\]

\[
g_a[u_a, x_a, \dot{x}_a](t) := \begin{cases} 
0, & \text{a.e. } t \in [0, b_a) \\
\frac{u_a(\rho_a[x_a]^{-1}(t))}{1 + h_a x_a[\rho_a[x_a]^{-1}(t)]}, & \text{a.e. } t \in [b_a, T + \tau_a[x_a](T)] 
\end{cases}
\]

for a.e. \( t \in [0, T] \). In addition, we define the summarized load at a junction – dependent on if there is an external source or not – for \( (c, d, v) \in C \times D \times V \)

\[
r_v^{c,d} := \sum_{a \in A_{in}(v)} g_a^{c,d}[u_a, x_a, \dot{x}_a] + \begin{cases} 
s_v^{c,d}(v, d) & (v, d) \in OD, c \in C \\
0 & v \in V \setminus O, c \in C 
\end{cases} \text{ on } [0, T].
\]

Finally, recalling the routing operator \( \mathcal{R}_d^{c,d} \) given in Definition 5 with \( d \in D \), the coupling condition for the connecting nodes are given for \( (v, c, d) \in V \times C \times D \), \( a \in A^d_{out}(v) \)

\[
u_a^{c,d} := \begin{cases} 
\mathcal{R}_d^{c,d}[x, \text{ext}] & a \in OP^d \\
0, & \text{else}
\end{cases} \text{ on } [0, T]
\]

with \( OP^d \) as in Eq. (4). The presented set of equations will be called the dynamic traffic assignment (DTA) considered.

Eqs. (15) to (20) represent the link dynamics on every link \( a \in A \) in the network (see Section 3.3), while Eq. (21) the cumulative inflow onto one node. Dependent on if the considered node \( v \in V \) is a source node, additional source \( s_v^{c,d} \) is added or not added. Finally, Eq. (22) represents the routing on the links exiting a node. This is why we prescribe the inflow \( u_a^{c,d} \) for \( a \in A^d_{out}(v) \) as all the flow having entered the node \( v \) and the source flow \( s \), i.e. \( r_v^{c,d} \), and apply the routing operator \( \mathcal{R} \). In case, flow cannot be routed on the considered link as there is no connection, no path, to the destination, we instantiate the inflow as zero. This entire system is called the dynamic traffic assignment subject to routing operators, and in Section 3.6 we will present results on existence and uniqueness of solutions.

www.annualreviews.org • Routing on Traffic Networks 15
3.6. Well-posedness of the DTA with routing operators

In this section, we present the well-posedness of the DTA system introduced in Definition 8 before when performing the routing by the suggested Lipschitz-continuous or delayed routing operators. The study of existence and uniqueness of solutions on the network is crucial. This is due to the fact that the routing operators itself depend on the state of the network so that another coupling between all equations emerges.

Theorem 2 (Existence/uniqueness of the network model for routings with delay-property). Recall the setting described in Definition 8 where every routing operator $R_{c,d}^a$ fulfills the conditions according to Item (B) in Definition 7 with time vectors $t_{c,d}^a \in [0,T]^{N_{c,d}}$. Then there is a unique solution of the system presented in Definition 8 and the solution satisfies

$$x_a \in W^{1,\infty} \left( (0,T); \mathbb{R}_{\geq 0}^{[C][D]} \right) \quad \forall a \in A.$$

Proof. The proof can be found in (72, Theorem 3.1). The basic idea consists of taking advantage of the delay. Using the delay in the routing, the routing choices depend except for finitely many points where the routing might actually happen in real time on the network solution of previous times. As this solution is already well-known there is no problem with any coupling of the routing operator w.r.t. the state of the network. For the time points where the routing might actually use real time information, we can approach these points from previous times and take advantage of the Lipschitz-continuity of the solution to identify the proper solution at the considered time-point.

Theorem 3 (Existence/uniqueness of the network model for (L-)Continuous routing op.). Let the network with the link-delay dynamics as in Definition 8 be given and assume that $R_{c,d}^a$ is a (L-)Continuous routing operator satisfying Item (A) in Definition 7. Then there exists a unique solution $x_a, a \in A$ of link-delay ODEs on the network and the solution satisfies

$$x_a \in W^{1,\infty} \left( (0,T); \mathbb{R}_{\geq 0}^{[C][D]} \right) \quad \forall a \in A.$$

Proof. The proof can be found in (72, Theorem 3.4). The proof is significantly more advanced than the proof of Theorem 2. Here, we take advantage of the result in [25], where the solution of the link-delay ODE is constructed over a sequence of time steps, taking advantage of the delay property of the ODE. Combining this with the Lipschitz-continuous routing operators we can define a self-mapping which is contractive in the uniform topology and obtain a unique fixed-point on the entire network, that is a solution to the DTA subject to Lipschitz-continuous routing operators.

3.7. Some remarks

Remark 1 (Instantaneous shortest path routing and instability). As claimed in Section 2.2.1 often a shortest path routing assignment in real time is posed. However, consider only two possible routes and assume that travel time is identical for both routes at a specific time $t \in [0,T]$. Then, flow has to be distributed on both links simultaneously, however in which ratio is fully model dependent as well as on other inflows of the links part of the considered paths. In case the specific ratio is not met, this directly contradicts shortest paths assignments. Also when discretizing the composed model this problem becomes evident: As
there will be never the same travel time at least numerically, in every time step all or nothing is sent on the out-going links. This “all or nothing” assignment might switch from time step to time step and is sensitive with respect to the discretization. For details and a counter example why a shortest-path assignment in continuous modelling is rather difficult to well-posed, we refer to [72, Remark 4.24].

Remark 2 (Forecasting). Even though our approach does not allow a forecast in time as we only assume to have real time information available, one can do a statistical forecast with the proper model or a simplified forecast and can implement this as \text{ext}. As long as only information in delayed form as suggested in Item (B) is considered, nothing in addition has to be prescribed in order to obtain an unique solution on the network. If information in real time is used and based on this a forecast is run, this forecast needs to depend in a Lipschitz-continuous way on the input datum and also the routing operator would be required to depend in a Lipschitz-continuous way on \text{ext}. We do not go into details.

3.8. Some instantiations of routing operators

In this section, we will present some routing operators for which our assumptions hold. Some of the examples are borrowed from [72]. To formulate the routing operators mathematically, we need a notation of path-travel time which we will present in Definition 9:

Definition 9 (Path travel time). Let \((v, d) \in V \times D\) as in Definition 2 be given. Then, we define for \(t \in [0, T]\) the set \(X_{v,d}(t)\) of the involved path flows from node \(v\) to destination \(d\), where \(x_a \in C \left( [0, T]; \mathbb{R}^{C(D)} \right)\), \(a \in A\), as

\[
X_{v,d}(t) := \bigcup_{p \in P_{v,d}} \left\{ \left( x_{p_1}(t), \ldots, x_{p_{\text{len}(p)}}(t) \right) \right\} \quad t \in [0, T].
\]

Let the travel time \(\tau_a[x_a]\) on every arc \(a \in A\) and \(\ell := |P_{v,d}|\) be given. After redefining its components – for a better readability – respectively to \(\tilde{p}_1, \ldots, \tilde{p}_\ell\), we define the vector of possible travel times \(\tau_{v,d}\) from node \(v\) to destination \(d\) by

\[
\tau_{v,d} \left[ X_{v,d} \right] (t) := \left( \frac{\sum_{i=1}^{\text{len}(\tilde{p}_1)} \tau_{\tilde{p}_1} \left[ x_{\tilde{p}_1}(t) \right]}{\sum_{i=1}^{\text{len}(\tilde{p}_\ell)} \tau_{\tilde{p}_\ell} \left[ x_{\tilde{p}_\ell}(t) \right]} \right) \quad t \in [0, T].
\]

Routing 1 (Routing with a logit-function). Given \(v \in V\) and \(d \in D\) we define weighted path distribution routing as

\[
\mathcal{A}_{v,d}^{x, \text{ext}}(t) := \frac{\sum_{p \in P_{v,d}} e^{-\left( \tau_{v,d} \left[ X_{v,d} \right](t) \right)_p}}{\sum_{p \in P_{v,d}} e^{-\left( \tau_{v,d} \left[ X_{v,d} \right](t) \right)_p}} \quad t \in [0, T].
\]

This routing operator is in the literature sometimes called nested-logit model and assigns on all exiting links at node \(v \in V\) a percentage of the incoming flow to the outgoing links \(a \in \mathcal{A}_{v,d}(v)\) based on the needed travel time to reach the destination. The exponential terms
lead to the fact that most flow is sent onto the shortest route and on all possible routes a flow is assigned, even a very small flow.

The routing operator satisfies Definition 7, Item (A) so that we obtain by Theorem 3 the existence and uniqueness of the solution on the network. Clearly, the operator can be enriched by tuning parameters.

In case of delay, a shortest path assignment can contrary to Remark 1 be considered and yields a unique solution on the network:

Routing 2 (Shortest path with delay). Given \( v \in V \) and \( d \in D \) we define shortest path routing for \( t \in [0, T] \) and \( a \in A_{\text{out}}(v)^d \) with delay \( \varepsilon \in \mathbb{R}_{>0} \) as

\[
R_{a}^{v,d}[x,\text{ext}](t) := \frac{1}{\left\{ p_1 : p \in \arg\min_{p \in P^v,d} \tau^{v,d}[X^{v,d} \circ P_{[0,T-\varepsilon]}](t) \right\}}.
\]

This operator assigns flow only on leaving nodes \( a \in A_{\text{out}}(v) \) with minimal travel time and in case that there is more than one shortest route it assigns flow equally. Due to the delay character it satisfies Definition 7, Item (B) so that we obtain by Theorem 2 the existence and uniqueness of the solution on the network. We emphasize that a real time shortest path delay might not be well-posed as detailed in Remark 1.

Routing 3 (Variational inequalities as routing operators). Having Definition 9 we can formulate the variational inequalities for routing as investigated in Section 2.2.2 as time-dependent Wardrop’s conditions (92), recalling that for every origin destination pair \((v,d)\) \(\in V \times D\), for every commodity \(c \in C\) and for every time \(t \in [0,T]\), the used paths’ travel times have to be smaller or equal to the travel time of the minimal paths, or – in formulae – when the cumulative assignment \(x^*_a, a \in A\) is an optimal assignment: \(\forall t \in [0,T], \forall(v,d) \in V \times D \forall c \in C\)

\[
\sum_{p \in P^v,d} \sum_{a \in \{p_1, \ldots, p_{\text{len}(p)}\}} \tau_a[x^*_a](t) (x_a(t) - x^*_a(t)) \geq 0
\]

for every \(x_a\) cumulative admissible flow. Compare also (13, Chapter 7.3). Altogether, also variational inequalities can be interpreted as routing operators in the sense of Definition 7, however our results are not applicable as there is no Lipschitz-continuity or delay property holding and also only a continuous dependency of the so obtained routing operator is questionable.

For a more comprehensive list of routing operators fitting in the proposed framework we refer to (72, Section 4). We want to emphasize once more that the introduced class of routing operators is very broad so that many in the literature already used routing operators fit in the proposed framework.

4. Conclusion and future research

In this work, we have – for an ordinary delay differential equation on the links – presented a rigorous framework for up to real time routing based on the state of the traffic in the network and how it connects to already existing research in the literature. We have shown the existence and uniqueness of solutions in the network and have considered so-called routing operators that are rather general and broadly applicable. Future research should
address the following related problems: (1) *Replacing the link dynamics instantiated here by ordinary delay differential equations with more advanced traffic flow models, as for instance the LWR model, the ARZ model, or the recent nonlocal models presented in Section 2.1.2.* The additional complexity – as mentioned before – will come from the need to define the proper node models in [Paragraph 2.1.2.1](#). For simple PDE models, a first approach has been made in [73]. (2) *Numerical testing.* Due to the structure of the routing operators potentially depending on the status of the network at any time, every routing operator has to know about this status at all times, requiring on the computational side a clever distribution of this information. Additionally, due to the fact that every destination has to be realized as a commodity, the density on each link is at least of the dimensionality of the number of destinations (multiplied by the number of different commodities) at every time step. As there might be many links in a given network where specific destinations cannot be reached, it might be worthwhile to analyze the network first for its connectivity. (3) *Testing of specific routing operators.* Once the specific structure for the routing operator has been chosen, there are many parameters that can still be chosen in the routing. Therefore, a study on real data optimizing the routing parameters accordingly would be insightful. (4) *Stability analysis.* For the Lipschitz-continuous routing, a stability analysis should be carried out. Changing the Lipschitz routing in the proper topology and all demand slightly, one can conclude that the solution in the full network is also close in the proper topology. Clearly, for a specific link, this question of the mentioned stability can easily be answered positively.

**Acknowledgements**

This work has been supported under the DARE program of the Philippine Commission on Higher Education.

**LITERATURE CITED**


96. H. S. Mahmassani and G.-L. Chang. On boundedly rational user equilibrium in transportation


