The Anticipatory Route Guidance Problem: Formulations, Analysis and Computational Results*

Jon Bottom† Soulaymane Kachani‡ Georgia Perakis§

January 2006¶

Abstract

The anticipatory route guidance problem (ARG), an extension of the dynamic traffic user-equilibrium problem, consists of providing messages, based on forecasts of traffic conditions, to assist drivers in their path choice decisions. Guidance becomes inconsistent when the forecasts on which it is based are invalidated by drivers' reactions to the provided messages. In this paper, we consider the problem of generating consistent anticipatory guidance that ensures that the messages based on dynamic shortest path criteria do not become self-defeating prophecies. We design a framework for the analysis of the ARG problem based on a fixed-point formulation of the problem. We also provide an infinite-dimensional variational inequality (VI) formulation. These equivalent formulations are, to the best of our knowledge, the first general analytical formulations of this problem. We establish, under weak assumptions, the existence of a solution to the ARG problem. Furthermore, we describe a solution approach based on averaging methods. Finally, we provide some computational results.

1 Introduction

1.1 Motivation

An important characteristic of road traffic congestion is its randomness. Data suggest that roughly 60% of congestion-related delays on urban freeways in the U.S. are due to specific random incidents such as accidents, vehicle breakdowns and the like (see Lindley [22] for more details). Even without such incidents, congestion has a random component that derives from variability in demand patterns and in network performance. Because of this randomness, a driver's past experience can be an unreliable basis for predicting the conditions associated with various travel options, and as a result, for making good travel choices.

Advanced traveler information systems (ATIS) attempt to provide tripmakers with data intended to help them make better travel decisions. In this paper, such data will be referred to as *messages*. Messages may have an arbitrary content. They may be available to all tripmakers (for example by radio or television broadcasts) or only to some: for example, those who pass near a particular infrastructure (such as variable message signs or VMS) or who have special receivers in their vehicles. Tripmakers, of course, may react to the messages in any way they choose.

*This paper has been made possible by a grant funded by the Region 2, University Transportation Research Center, City College of NY and the New York State Department of Transportation

†Charles River Associates, Inc.; jbottom@crai.com

‡IEOR Department, Columbia University; kachani@ieor.columbia.edu

§Operations Research Center, M.I.T.; georgiap@mit.edu

¶Preliminary version. Do not distribute without authors' permission

Traveler information systems may be distinguished on the basis of the type of information they provide in messages. Static systems furnish information that changes only infrequently: for example, locations of and directions to trip attractions such as cultural centers or restaurants. *Reactive* systems estimate prevailing travel conditions from real-time measurements and provide messages directly based on these estimates: for example, information about current travel times. *Predictive* or *anticipatory* systems use real-time measurements to forecast travel conditions in the near-term future (up to a few hours), and present messages based on these predictions. As a result, a tripmaker can make a decision based on what conditions are likely to be at network locations at the time he/she will actually be there, rather than on (possibly very different) currently prevailing conditions.

This paper focuses on predictive traveler information systems that provide messages intended to facilitate drivers' path choice decisions before and during a trip. In general, the messages may inform drivers about anticipated traffic conditions on different available paths, or (based on these conditions) recommend a specific path to follow, or both. Such systems are called route guidance systems.

If only a few drivers receive route guidance messages, they may benefit from it by making better path choice decisions. Nevertheless, the choices they make will likely have only a marginal impact on overall network traffic conditions. When more drivers receive guidance, on the other hand, their reactions to the guidance may affect traffic conditions significantly. The key issue in generating guidance messages based on traffic condition forecasts is to ensure that drivers' reactions to the guidance do not invalidate the forecasts that the guidance is based on, and so render the guidance irrelevant or worse. Messages predicting impending congestion on one road, for example, may cause drivers to switch *en masse* to a parallel road less able to accommodate them, leaving the original road free flowing and producing overall worse traffic conditions than if no guidance had been issued. Predictive guidance is said to be *consistent* when the forecasts on which it is based are indeed experienced by drivers after they react to it. Generating consistent guidance is called the Anticipatory Route Guidance (ARG) problem.

Generation of anticipatory guidance clearly requires the application of some kind of traffic prediction model. All such models embody assumptions about the information basis of driver decisions, and about drivers' response to such information. Deterministic traffic assignment models assume that drivers departing from their origin have full information about actual network conditions on the paths available to them, choose one accordingly, and follow it unswervingly to their destination. Stochastic assignment models account for errors in modeling or driver perception by assuming that drivers at the origin choose a path based on a randomly perturbed version of actual network conditions, and again follow it to the destination. A full information assumption underlies even these models in the sense that, as the magnitude of the perception error decreases, the stochastic equilibrium path choices increasingly resemble those that would be made in a full information deterministic setting.

However, the applicability of the full information assumption to general route guidance modeling is questionable. Consider a guidance system consisting of a VMS located somewhere on a network. Drivers leaving their origins make a path choice based on assumptions about traffic conditions that may not be accurate. Drivers whose path choice happens to take them by the VMS receive guidance messages and may decide to switch to another path for the remainder of their trip; those who do not presumably pursue to their destination the path they chose earlier. The information available to a driver, and the resulting en route path switches, depend on the particular path taken through the network and on the guidance information that is available at locations along that path. This path dependency of information availability and driver behavior has no counterpart in full information models.

1.2 Literature review

Some network-level analyses of route guidance systems have assumed that the effect of guidance will be to establish traffic equilibrium conditions (see Kaufman *et al.* [18] and Engelson [8] for more details). Others have modeled guidance effects via a reduction in the perception error of guided drivers using a stochastic assignment model (see Lotan and Koutsopoulos [21] and Hamerslag and van Berkum [12] for more details). These approaches may be appropriate if perfect information is available to drivers at all locations. More generally, however, when a guidance system provides information only at a limited number of network locations, there is no reason to expect that the resulting flow patterns and traffic conditions would correspond to those of a conventional network equilibrium. There is even less reason to expect this if, as is likely to be the case in reality, the messages provided by the guidance system are very succinct network condition summaries or path recommendations.

Some traffic simulation models, such as DYNASMART by Mahmassani and co-workers (Mahmassani *et al.* [26]), MITSIM and DynaMIT by Ben-Akiva and co-workers (Ben-Akiva *et al.* [1] and Bottom [3]), can represent a variety of route guidance technologies, including limited-range systems such as VMS. These models employ various heuristics in an attempt to achieve anticipatory guidance consistency. However, to the best of our knowledge, no analytical results are available regarding these methods, in part perhaps because they did not develop the guidance generation algorithms from a clear analytical formulation of the problem.

Bovy and van der Zijpp [4] and Bottom [3] have proposed modeling frameworks for the ARG problem. In Bottom's framework, for example, the (time-dependent) modeling variables are the network conditions, the path splitting rates at control nodes (i.e. the fractions of flow continuing to their destination via each available path at control nodes) and the guidance messages. The modeling relationships are the dynamic network loading map, which transforms the path splits into network conditions; the guidance map, which transforms the network conditions into guidance messages; and the routing map, which transforms guidance messages into path splits. These three relationships can then be combined into three alternative composite maps that model the ARG problem and lead to three equivalent fixed-point formulations. Finding consistent guidance is equivalent to finding a fixed point of a composite map. When these composite maps are continuous, standard fixed point existence results may be applied to derive existence results. However, this approach requires assuming key properties of the dynamic network loading map, and stronger results (for example, characterizations of the feasible region and/or solution set) are difficult to obtain.

1.3 Goals and contributions

One of our main goals in this paper is to establish, under relatively weak assumptions, the existence of a solution to the ARG problem. To achieve this, we propose what we believe is the first analytical formulation of the ARG problem.

The development and and analysis of this formulation lead to a number of contributions, including:

- development of a network transformation well-suited to represent the ARG problem;
- specification and detailed analysis of an important generalization of the standard dynamic network loading (DNL) problem, in which origin-destination (O-D) flows may change paths at intermediate nodes. This feature is perhaps the major characteristic that distinguishes the ARG problem from the dynamic traffic assignment (DTA) problem;
- comparison of the ARG problem to the DTA and static traffic equilibrium problems, and to hyperpath-based network flow formulations;

- two distinct formulations of the ARG problem, one as a fixed point problem and the other as an infinite-dimensional variational inequality. We show that these are equivalent;
- rigorous analysis of the existence of solutions to the ARG problem; and
- identification of practical solution algorithms and some computational experiments.

In this paper, we will consider only the case where the messages provided to drivers by the guidance infrastructure at control nodes consist of complete information on network travel times. Extension of the analysis in this paper to more general message types (for example, limited summaries of network condition information) is a subject for future work.

1.4 Structure of the paper

The paper is organized as follows. In Section 2, we start by introducing the notation and the feasibility conditions of the ARG problem. We then provide a variational inequality (VI) formulation of this problem. We also present a fixed-point formulation of the problem and establish equivalence of the two formulations. We discuss two special cases – the static ARG problem and the Dynamic User-Equilibrium problem – and also consider the relationships between our approach and hyperpath-based formulations of network equilibrium problems. In Section 3, we study the mathematical properties of the problem. Under sufficient conditions on the path flow rate functions and the travel time functions, we establish that the feasible region $F(ARG)$ of the Anticipatory Route Guidance problem is non-empty, and that the FIFO property holds. We provide a generic counterexample illustrating that the assumptions that we impose to ensure that FIFO holds are the tightest possible. We establish key properties of the feasible region, such as boundedness, closedness and convexity. Furthermore, we establish the existence of a solution to the ARG problem. In Section 4, we propose a solution algorithm that solves the ARG problem. This algorithm is based on averaging methods. We also report on some computational results. Finally, in Section 5, we provide some conclusions and future steps in the study of this problem.

2 Problem Formulations

2.1 Notation

In this section, we introduce the notation necessary to formulate the ARG problem. The notation is somewhat tedious. However, part of the contribution of this work lies in identifying the minimum set of notation that enables us to formulate the ARG problem analytically.

The physical traffic network is represented conceptually by a directed network $G = (N, A)$, where N is the set of nodes and A is the set of directed links. N_1 denotes the set of origin nodes, N_2 the set of control nodes (e.g. nodes where vehicles may receive guidance messages), and P the set of all origin-destination (O-D) paths. In the following, p denotes a path between origin node r and destination node s, and K_{rs} denotes the subset of paths between O-D pair (r, s). In addition, $q^{rs}(.)$ denotes the flow rate from origin r to destination s.

Path variables:

$\lvert P \rvert$	:	number of paths in the network;
$RS(p)$	:	(O-D) pair associated with path p;
p^1	:	first link of path p;
p^l	:	last link of path p;
$f_p(t)$	:	departure flow rate on path p at time t;
$f^{n,RS(p)}(t)$	:	flow rate of O-D pair $RS(p)$ exiting node n at time t;
f	:	vector of path departure flow rate functions $f_p(\cdot)_{p \in P}$ for all times t;
M_p	:	upper bound on the path departure flow rate function $f_p(\cdot)$;
$\beta_{np}(t)$	:	path splitting rate, i.e., the fraction of flow exiting node n via path p at time t;
β	:	vector of path splitting rate functions $\beta_{np}(\cdot)$;
$S_p(t, f)$	:	travel time on path p departing at time t for a vector of flows $f(\cdot)$.

Link variables:

$head(a)$	:	head node of link a; if $a = (m, n)$, $head(a) = m$
$tail(a)$	:	tail node of link a; if $a = (m, n)$, $tail(a) = n$
M_a	:	upper bound on the link entrance flow rate function $u_a(\cdot)$;
$u_a(t)$	:	entrance flow rate of link a at time t;
$v_a(t)$	:	exit flow rate of link a at time t;
$U_a(t)$	:	cumulative entrance flow of link a during interval $[0, t]$;
$V_a(t)$	:	cumulative exit flow of link a during interval $[0, t]$;
$X_a(t)$	:	number of vehicles on link a at time t;
$D_a(y)$	:	traversal time function (or link performance function) of link a, where y is the number of vehicles on link a;
$s_a(t)$	:	exit time of a flow entering link a at time t
	=	$t + D_a(X_a(t))$;
B_{1a}	:	lower bound on the derivative $D'_a(\cdot)$ of the travel time function $D_a(\cdot)$;
B_{2a}	:	upper bound on the derivative $D'_a(\cdot)$ of the travel time function $D_a(\cdot)$.

Link-path flow variables:

(a, p)	:	a link-path pair;
δ_{ap}	=	1 if link a belongs to path p, and 0 otherwise;
$u_{ap}(t)$	:	entrance flow rate on link a traveling on path p at time t;
$v_{ap}(t)$	:	exit flow rate on link a traveling on path p at time t;
$U_{ap}(t)$	:	cumulative entrance flow on link a traveling on path p during interval $[0, t]$;
$V_{ap}(t)$	:	cumulative exit flow on link a traveling on path p during interval $[0, t]$;
$X_{ap}(t)$	:	partial load on link a at time t induced by flow on path p.

Time variables:

t	:	index for continuous time;
$[0, T]$	:	O-D traffic demand period. After time T, the flow rate functions are zero;
$[0, T_\infty]$	:	analysis period. It is the interval of time from the instant when flows enter the network to the first instant when all flows exit the network.

Below, we perform a network transformation that is the basis of our approach and analytical results in this paper.

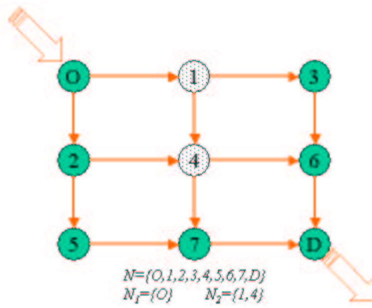

Figure 1: Network Example

We divide every path $p \in K_{rs}$ that goes through a control node into subpaths in the following manner. Subpaths can either (i) originate at r and end at the first control node on path p, or (ii) originate at a control node and end at the following control node on path p, or (iii) originate at the last control node on path p and end at s. Let P_1 denote the set of subpaths we have just described. Note that this definition of subpaths allows P_1 to contain several copies of the same subpath. However, these copies come from different paths. Let P_2 denote the set of paths that do not go through control nodes (however, a path in P_2 might originate at a control node).

Let $\overline{P} = P_1 \cup P_2$. Also, let $\overline{P}_1$ be the set of subpaths that originate at control nodes and hence receive information about the network conditions. It follows that $\overline{P}_1 = \{p \in \overline{P} | head(p^1) \in N_2\}$, where $head(p^1)$ denotes the head of the first link of path p (i.e. the origin node of path p), and N_2 is the set of control nodes. Let $\overline{P}_2 = \overline{P} \setminus \overline{P}_1$.

Let $\widehat{p}(p)$ denote the path containing subpath p (this could be path p itself if $p \in P_2$). $s(p)$ will denote the first subpath on path p (again, this could be path p itself if $p \in P_2$). Finally, $f_{p\widehat{p}(p)}(t)$ denotes the subpath flow rate at time t on subpath p of path $\widehat{p}(p)$.

Example: In Figure 1, we consider an example network of nine nodes (including two control nodes) and one O-D pair (O, D). Applying the network transformation defined above on this network, we obtain:

$$\overline{P}_1 = \{(1,3,6,D),(1,4),(1,4),(4,6,D),(4,6,D),(4,7,D),(4,7,D)\},$$
$$\overline{P}_2 = \{(O,1),(O,1),(O,1),(O,2,4),(O,2,4),(O,2,5,7,D)\}.$$

Notice for instance that there are two copies of subpath $(4,6,D)$ in $\overline{P}_1$ since both path $(O,1,4,6,D)$ and path $(O,2,4,6,D)$ contain this subpath.

2.2 Feasibility Conditions of the ARG Problem

The objective of this subsection is to mathematically define the feasible region of the ARG problem. The region corresponds to the set of feasible network flows and travel times, as established by solutions to dynamic network loading (DNL) problems. In the context of the ARG problem, the DNL problem consists of determining the time-varying link, subpath and path flows and travel times that result from the movement of given O-D flows over the network in accordance with particular path or subpath splitting rates at the origins and at intermediate control nodes. For a fixed network structure and set of O-D flows, the ARG DNL problem can thus be viewed as a map from the domain of path splitting rates to the range of network flows and travel times.

Similar to the DNL maps proposed by Friesz *et al.* [9], by Wu *et al.* [34] and by Kachani [13] in the context of the Dynamic User-Equilibrium (DUE) problem, the ARG DNL map is formulated as a system of equations expressing link dynamics, flow conservation, flow propagation, non-negativity and boundary constraints. Unlike the DUE DNL problem, in which flows departing from their origin on a particular path follow that path without change to their destination, the ARG DNL problem allows flow to switch from one subpath to another at intermediate locations. As a result, the ARG DNL map has two added features: (i) the model is formulated in terms of subpath flow rates instead of path flow rates, and (ii) the path-node splitting rates are explicitly used.

Note that, at every node and every time instant, the set of feasible splitting rates is a simplex. Therefore, the set of all feasible splitting rates is a product of simplices. The ARG feasible region, $F(ARG)$, is then the set of flows and travel times that result as the splitting rates vary across their feasible region. Each point in $F(ARG)$ is obtained as a solution of some particular DNL problem.

Link dynamics equations
The link dynamics equations express the relationship between the flow variables of a link. They are given by:

$$\frac{dX_{ap}(t)}{dt} = u_{ap}(t) - v_{ap}(t), \qquad \forall (r,s), \forall p \in K_{rs}, \forall a \in p. \tag{1}$$

Flow conservation equations
For every link a that has a head node that is neither an origin node nor a control node (i.e. $head(a) \in N \setminus (N_1 \cup N_2)$), the flow conservation equations can be expressed as

$$u_{ap}(t) = v_{a'p}(t), \tag{2}$$

where a' is the link preceding link a on path p.

For every link a that has a head node n that is an origin node (i.e. $n \in N_1$) and for all paths p originating at n, the flow conservation equations can be expressed as

$$\begin{aligned} u_{ap}(t) &= f_{s(p)p}(t), \\ &= \beta_{np}(t) q^{RS(p)}(t), \end{aligned} \tag{3}$$

where the (O-D) pair flow rates $q^{RS(p)}(t)$ are given.

For every link a that has a head node n that is a control node (i.e. $n \in N_2$) and for all paths p that do not originate at n, the flow conservation equations can be expressed as

$$\begin{aligned} u_{ap}(t) &= f_{p'p}(t) \\ &= \beta_{np}(t) f^{n,RS(p)}(t), \end{aligned} \tag{4}$$

where $p^{'}$ is the subpath of p originating at n and $f^{n,RS(p)}(t) = \sum_{\overline{p} \in RS(p), \overline{a} \in \overline{p} | tail(\overline{a}) = n} v_{\overline{a}\overline{p}}(t)$.

Link-path flow relationships

The following relationships express the fact that at each time t the link flow variables are the sum of their corresponding link-path variables:

$$u_a(t) = \sum_{p|a\in p} u_{ap}(t), \qquad v_a(t) = \sum_{p|a\in p} v_{ap}(t),$$

$$U_a(t) = \sum_{p|a\in p} U_{ap}(t), \qquad V_a(t) = \sum_{p|a\in p} V_{ap}(t), \tag{5}$$

$$X_a(t) = \sum_{p|a\in p} X_{ap}(t), \qquad \forall (r,s), \forall p \in K_{rs}, \forall a \in p.$$

Flow propagation equations

Flow propagation equations are used to describe the flow progression over time. Note that a flow entering link a at time t will exit the link at time $s_a(t)$. Therefore, at time t, the cumulative exit flow of link a should be equal to the integral of all inflow rates which would have entered link a at some earlier time ω and exited link a by time t. This relationship is expressed by the following equation:

$$V_{ap}(t) = \int_{\omega\in W} u_{ap}(\omega)d\omega, \qquad \forall (r,s), \forall p \in K_{rs}, \forall a \in p, \tag{6}$$

where $W = \{\omega : s_a(\omega) \leq t\}$.

If the link exit time functions $s_a(\cdot)$ are continuous and satisfy the strict FIFO property, then the flow propagation equations (6) can be equivalently rewritten as

$$V_{ap}(t) = \int_0^{s_a^{-1}(t)} u_{ap}(\omega)d\omega, \qquad \forall (r,s), \forall p \in K_{rs}, \forall a \in p. \tag{7}$$

The strict FIFO condition implies that a car entering link a at time t will exit only after the cars that entered link a before it have all exited. In mathematical terms, this is equivalent to the link exit time functions $s_a(\cdot)$ being strictly increasing, guaranteeing invertibility.

Link exit time functions $s_a(t)$ are obtained from link travel time functions using the following definitional constraint:

$$s_a(t) = t + D_a(X_a(t)).$$

Valid expressions for $D_a(X_a(t))$ can be found in Kachani and Perakis [16].

Non-negativity constraints

We further assume that the departure path flow rates are non-negative:

$$f_p(\cdot) \geq 0 \qquad \forall (r,s), \forall p \in K_{rs}. \tag{8}$$

Boundary equations

Since we assume that the network is empty at $t = 0$, the following boundary conditions are required:

$$U_{ap}(0) = 0, \qquad V_{ap}(0) = 0, \qquad X_{ap}(0) = 0, \qquad \forall (r,s), \forall p \in K_{rs}, \forall a \in p. \tag{9}$$

The above formulation of the DNL map is general enough to account for the case where the FIFO property, defined above, is not necessarily verified (notice that Equation (6) does not assume that the FIFO property holds). In Section 3, we investigate conditions under which the FIFO property holds. We examine conditions on the link travel time functions $D_a(\cdot)$ and on the departure path flow rates $f_p(\cdot)$. When FIFO is verified, the model becomes more tractable.

This completes the formulation of the DNL problem in a route guidance context. The set of solutions to this DNL problem, over all feasible values of the splitting rates, constitutes the feasible region for the ARG problem. Relevant properties of $F(ARG)$ are derived in subsection 3.2.

In the next subsection, we provide a variational inequality formulation of the ARG problem itself.

2.3 A Variational Inequality Formulation

Similarly to Friesz *et al.* [9], who developed variational inequality formulations for the Dynamic User-Equilibrium problem in terms of *path flow rates*, we can formulate the ARG problem in terms of *subpath flow rates*.

As discussed in Subsection 2.1, there are two types of drivers in the network:

- "Uninformed" drivers who travel on subpaths $p \in \overline{P}_2$ and, as a result, do not receive any information. These drivers have some estimate $\widehat{f}$ of the vector of time-dependent flows in the network. (Such estimates could be based on observations by these users in prior trips, for example.) We assume that uninformed drivers utilize these flow estimates to determine the vector of path travel times $S(t, \widehat{f})$. We further assume that these drivers use the path time estimate to select paths that minimize their O-D travel times. The simplest case occurs when $\widehat{f}$ corresponds to the vector of free-flow path travel times $S(t, 0)$. In this case, drivers follow the static shortest path from their origin to their destination.
- "Informed" drivers who travel on subpaths $p \in \overline{P}_1$ (that originate at control nodes) and as a result receive full information about the network traffic conditions. We assume that these drivers behave "rationally" in the sense that they select paths that minimize their flow-dependent dynamic travel times.

In what follows, we consider these two categories of drivers traveling in the same network. Note that the behavior of the uninformed drivers is unaffected by anything that the informed drivers do, but not conversely. The above two categories of drivers allow us to solve the ARG problem by equivalently solving the following variational inequality problem: Find a vector of time-dependent flows $f^\star \in F(ARG)$ satisfying

$$\begin{aligned} &\sum_{r,s} \sum_{p \in \overline{P}_2} \int_0^{T_\infty} S_p(w, \widehat{f})(f_{p\widehat{p}(p)}(w) - f^\star_{p\widehat{p}(p)}(w))dw + \qquad (10) \\ &\sum_{r,s} \sum_{p \in \overline{P}_1} \int_0^{T_\infty} S_p(w, f^\star)(f_{p\widehat{p}(p)}(w) - f^\star_{p\widehat{p}(p)}(w))dw \quad \geq 0, \qquad \forall f \in F(ARG). \end{aligned}$$

The first part of this variational inequality formulation describes the uninformed drivers while the second part describes the informed drivers.

To simplify the notation, we will write $< h, g >= \sum_{r,s} \sum_{p \in \overline{P}} \int_0^{T_\infty} h_p(w) g_p(w) dw$. With this notation, the above infinite-dimensional variational inequality formulation can be written more compactly as: Find a vector of flows $f^\star = (f^\star_{|\overline{P}_1}, f^\star_{|\overline{P}_2}) \in F(ARG)$ satisfying

$$< S(\widehat{f}), f_{|\overline{P}_2} - f^\star_{|\overline{P}_2} > + < S(f^\star_{|\overline{P}_1}, f^\star_{|\overline{P}_2}), f_{|\overline{P}_1} - f^\star_{|\overline{P}_1} > \quad \geq 0, \qquad \forall f = (f_{|\overline{P}_1}, f_{|\overline{P}_2}) \in F(ARG). \tag{11}$$

In summary, the continuous-time Anticipatory Route Guidance problem is equivalent to variational inequality formulation (11) subject to constraints (1)-(9). In general, the ARG problem is a continuous-time non-linear optimization problem. The non-linearity of the model comes from the path flow dependence of the path travel times in the variational inequality, as well as integral equation (6). In this formulation, the problem's data consist of the network structure and link travel time (performance) functions $D_a(X_a)$. The input variables are the (O-D) pair departure flow rates $q^{rs}(t)$. The "decision variables", for which variational inequality (11) is in fact solving, are the path splitting rates $\beta_{np}(\cdot)$. In addition, the feasible region $F(ARG)$ should be thought of as a space of path splitting rates. However, for the sake of clarity and brevity, the ARG problem is formulated explicitly in terms of subpath flow rates, which are intermediary variables that are computed from the path splitting rates (given the input parameters $D_a(X_a)$ and $q^{rs}(t)$). In this formulation, the unknown variables that we wish to determine are, in addition to the path splitting rates $\beta_{np}(\cdot)$, the link and path entrance flow rates $u_a(\cdot)$ and $f_p(\cdot)$, the link exit flow rates $v_a(\cdot)$, the link cumulative entrance and exit flows $U_a(\cdot)$ and $V_a(\cdot)$, the link loads $X_a(\cdot)$, and the link and path exit time functions $s_a(\cdot)$ and $S_p(\cdot)$. Notice that, due to the integral equation (6), this problem is hard to solve.

Furthermore, notice that solving variational inequality (11) is equivalent to solving the following two variational inequalities in sequence: Find a vector of flows $f^\star_{|\overline{P}_2} \in F(ARG)_{|\overline{P}_2}$ satisfying

$$< S(\widehat{f}), f_{|\overline{P}_2} - f^\star_{|\overline{P}_2} > \quad \geq 0 \quad \forall f_{|\overline{P}_2} \in F(ARG)_{|\overline{P}_2}. \tag{12}$$

Then, find a vector of flows $f^\star_{|\overline{P}_1} \in F(ARG)_{|\overline{P}_1}$ satisfying

$$< S(f^\star_{|\overline{P}_1}, f^\star_{|\overline{P}_2}), f_{|\overline{P}_1} - f^\star_{|\overline{P}_1} > \quad \geq 0, \quad \forall f_{|\overline{P}_1} \in F(ARG)_{|\overline{P}_1}. \tag{13}$$

In the next subsection, we provide an alternative formulation of the ARG problem based on a fixed-point approach to the problem.

2.4 A Fixed-Point Formulation

In this subsection we argue that the ARG problem can also be formulated as a fixed point problem in the subpath flow rates. Moreover, in the following subsection we will show that the fixed point and variational inequality formulations are equivalent.

The fixed-point approach solves the following two sub-problems in sequence:

Sub-problem 1: Drivers who do not receive any information follow shortest time paths based on travel times determined from some default (and fixed) estimate of the vector $\widehat{f}$ of network flows. For these drivers, the fixed-point approach attempts to find feasible flows that are a fixed-point solution of the travel time minimization problem with $S(t, \widehat{f})$ as the vector of path travel times. In mathematical terms, this sub-problem is equivalent to: Find a vector of flows $f^\star_{|\overline{P}_2} \in F(ARG)_{|\overline{P}_2}$ satisfying

$$f^\star_{|\overline{P}_2} \in \text{argmin}_{\; f_{|\overline{P}_2} \; \in \; F(ARG)_{|\overline{P}_2}} < S(\widehat{f}), f_{|\overline{P}_2} > . \tag{14}$$

Note that there may be multiple minimum subpaths, so the appropriate fixed point formulation involves set membership rather than equality.

Sub-problem 2: Drivers who receive full information about the network traffic conditions follow shortest time paths based on travel times determined from the actual flows of both uninformed and other informed drivers. For these drivers, the fixed-point approach attempts to find feasible flows that are a fixed-point solution of the flow-dependent travel time minimization problem. In mathematical terms, this sub-problem is equivalent to: Find a vector of flows $f^\star_{|\overline{P}_1} \in F(ARG)_{|\overline{P}_1}$ satisfying

$$f^\star_{|\overline{P}_1} \in \text{argmin}_{\ f_{|\overline{P}_1}\ \in\ F(ARG)_{|\overline{P}_1}} < S(f^\star_{|\overline{P}_1}, f^\star_{|\overline{P}_2}), f_{|\overline{P}_1} > . \quad (15)$$

2.5 Relationship between the Variational Inequality and the Fixed-Point Formulations

In this subsection, we establish the equivalence of the variational inequality formulation in Subsection 2.3 and the fixed-point formulation in Subsection 2.4.

Proposition 1 *Variational inequality formulation (11) is equivalent to fixed-point formulation (14-15).*

Proof:
It is easy to see that (14) is equivalent to (12). Furthermore, (15) is also equivalent to (13). Since solving (12) and (13) in sequence is equivalent to solving variational inequality (11), the result of the proposition follows.

□

2.6 Two Special Cases

In this subsection, we examine two instances of the Anticipatory Route Guidance problem: the Dynamic User-Equilibrium problem (DUE) and the static ARG problem.

2.6.1 The Dynamic User-Equilibrium Problem

The ARG problem is a generalization of the Dynamic User-Equilibrium (DUE) problem. As we discussed in Section 4.1, the DUE problem assumes that flows departing the origin on a particular path remain on that path to their destination. However, in the ARG problem, we allow flow to change from one path to another at intermediate control nodes. Furthermore, the DUE problem assumes that, when departing from their origin, all drivers have full information about the network conditions on the paths available to them (i.e the set of control nodes is exactly the set of origin nodes). As a result, all drivers choose a path that minimizes their actual flow-dependent travel time.

In this special case, the ARG problem simplifies significantly. Instead of two types of drivers as discussed in Subsection 2.3, the only type of drivers we have in this case are drivers who receive full information at the origin. As a result, all paths belong to $\overline{P}_1$, the set $\overline{P}_2$ is empty, and the network transformation introduced in Subsection 2.1 is no more needed. Furthermore, of the three types of conservation equations (2), (4) and (5), only equations (2) and (4) apply in this case. Finally, variational inequality (11) reduces in this case to: Find a vector of flows $f^\star \in F(DUE)$ satisfying

$$< S(f^\star), f - f^\star > \quad \geq 0, \qquad \forall f \in F(DUE). \quad (16)$$

Variational inequality (16) can be interpreted as the problem of determining feasible path flows $f^\star$ so that, at every time t, and for every origin-destination pair (r,s), all used paths belonging to (r,s) have equal and minimum travel times. In other words, it consists of determining feasible path flows that satisfy Wardrop's first principle of traffic assignment at each time instant.

2.6.2 The Static ARG Problem

The static ARG problem is similar in many respects to the dynamic ARG problem discussed above, but considers steady state (i.e. single value) rather than time-varying traffic flows and travel times over the analysis period $[0, T_\infty]$. As a result, there is no notion of flow propagation through the network over time, and the network loading map simplifies to little more than a bookkeeping of the flows and travel times on each link. Subpath and path travel times are obtained as the straightforward addition of the travel times of their component links.

With these simplifications, and with a suitable choice of inner product, the variational inequality formulations of the static and dynamic ARG problems can be written in formally similar ways: Find a vector of flows $f^\star_{|\overline{P}_2} \in F(ARG)_{|\overline{P}_2}$ satisfying

$$< S(\widehat{f}), f_{|\overline{P}_2} - f^\star_{|\overline{P}_2} > \quad \geq 0 \quad \forall f_{|\overline{P}_2} \in F(ARG)_{|\overline{P}_2}. \tag{17}$$

Then, find a vector of flows $f^\star_{|\overline{P}_1} \in F(ARG)_{|\overline{P}_1}$ satisfying

$$< S(f^\star_{|\overline{P}_1}, f^\star_{|\overline{P}_2}), f_{|\overline{P}_1} - f^\star_{|\overline{P}_1} > \quad \geq 0, \quad \forall f_{|\overline{P}_1} \in F(ARG)_{|\overline{P}_1}. \tag{18}$$

In static ARG problem formulation, the inner product $< x, y >$ is simply the scalar product of the two finite-dimensional vectors x and y.

2.7 Relationship to Hyperpath-Based Network Equilibrium Problems

Hyperpath-based approaches to network equilibrium problems have been proposed to address a number of transport network modeling situations where en route decisions may be important.

Their original transport modeling application was to transit networks. Nguyen and Pallottino [28] realized that hyperpaths were a natural and powerful graph theoretic construction for representing and analyzing the notion of transit *strategies* originally proposed by Spiess and Florian [33]. (A transit strategy is an adaptive transit path choice decision procedure in which the selection of a transit line at a transfer point depends on the (random) service characteristics of the available lines at the transfer point.) Although strategies were originally defined for uncongested (i.e., flow-independent) transit networks, subsequent work using hyperpaths led to a number of more general formulations and analyses of static equilibrium problems on networks involving flow dependencies of both link travel times and node waiting times.

Marcotte and Nguyen [27] applied hyperpaths to analyze static equilibria in networks with hard link capacity constraints (upper bounds on link flows). They assumed that travelers departing from their origins select a travel strategy: an ordered set of successor nodes at each node. At each node, a traveler selects the first successor node for which the connecting link is available, where a link's probability of being available is directly proportional to its capacity and inversely proportional to the number of travelers who would like to use it. The authors analyzed a number of instances of this general problem and proposed an algorithmic approach for addressing them. (The paper also contains a useful review of hyperpath applications in transit network modeling.)

Cantarella [7] included hyperpath routing choices in his very general analysis of static network equilibrium problems. Among other things, Cantarella pointed out that hyperpaths could be used to model en route path switching decisions resulting from traffic information systems, but did not pursue this thought in his paper.

The hyperpath approach described above clearly bears some resemblance to the approach that we consider in this paper for the anticipatory route guidance problem. In both cases, travelers leaving their origin are not fully informed about traffic conditions relevant to their path choice, but are able to acquire at certain intermediate nodes information that influences their choice of subsequent path to their destination. On the other hand, there are also some clear differences. To begin with, we consider here dynamic networks whereas, to the best of our knowledge, hyperpath-based approaches have focused on static problems. We would argue that dynamic networks are the most appropriate context in which to analyze issues associated with real-time travel information provision. Furthermore, none of the hyperpath analyses that we are aware of has considered the type of information acquisition at nodes that is important here: full information on network flows and travel times, as influenced by both informed and uninformed tripmakers.

3 Mathematical Properties of the ARG Problem

In this section, we examine the mathematical properties of the ARG problem. In Subsection 3.1, we introduce some definitions and preliminary results. In Subsection 3.2, we establish a few properties of the feasible region $F(ARG)$. In Subsection 3.3, we establish the existence of a solution to the ARG problem.

3.1 Definitions and Preliminary Results

3.1.1 Definitions

In this subsection, we present some important definitions. In particular, the following three definitions express three different types of First In First Out (FIFO) properties. The FIFO property will play a key role in the analysis of our model in Subsection 3.2.

A link verifies the FIFO property if and only if the link exit time function is non-decreasing. This means that a car that enters a link cannot exit before the cars that entered earlier. In particular,

Definition 1 *(FIFO 1): A link verifies the FIFO property if and only if:*

$$\forall (t_1, t_2) \in [0, T]^2, \text{ if } t_1 \leq t_2, \text{ then: } s_a(t_1) \leq s_a(t_2). \tag{19}$$

A link verifies the strict FIFO property if and only if the link exit time function is strictly increasing. This means that a car on a link cannot exit before or at the same time as other cars that entered the same link earlier. In particular,

Definition 2 *(FIFO 2): A link verifies the strict FIFO property if and only if:*

$$\forall (t_1, t_2) \in [0, T]^2, \text{ if } t_1 < t_2, \text{ then: } s_a(t_1) < s_a(t_2). \tag{20}$$

Definition 3 *(FIFO 3): A link verifies the strong FIFO property if and only if:*

$$\exists\ \theta > 0 \text{ such that } \forall (t_1, t_2) \in [0, T]^2, \text{ if } t_1 < t_2, \text{ then: } s_a(t_2) - s_a(t_1) \geq \theta (t_2 - t_1). \tag{21}$$

Lemma 1 below ranks these three FIFO properties by order of strength.

In the model presented in Section 2, the path departure flow rate functions $f_p(\cdot)$ are control variables. In an effort to establish general results, we assume that these functions are Lebesgue integrable. A function is said to be Lebesgue integrable if the set of points where this function is discontinuous is Lebesgue negligible. A set is negligible if its Lebesgue measure is 0.

Later in this section, we show that the cumulative flow variables are differentiable and the solution to the problem is unique almost everywhere (a.e.). The next definition makes this notion precise.

Definition 4 *A solution to a problem is said to be unique (resp. differentiable) almost everywhere if and only if the set of points where this solution is not unique (resp. not differentiable) is Lebesgue negligible.*

3.1.2 Preliminary Results

Next, we provide two lemmas. The first lemma is an easy result that establishes a strength order for the three FIFO properties defined by equations (1)-(3).

Lemma 1 *(i) If the strong FIFO property (FIFO 3) is verified, then so is the strict FIFO property (FIFO 2).*
(ii) If the strict FIFO property is verified, then so is the FIFO property (FIFO 1).

This lemma follows from the definitions of these properties.

In this paper, we show that the strong FIFO property holds for the ARG problem. Hence, using Lemma 1, the FIFO and the strict FIFO properties also hold.

A *diffeomorphism* is a continuously differentiable function that has a continuously differentiable inverse. The following lemma gives sufficient conditions for a function to be a diffeomorphism. This result will be used to establish, under certain assumptions, that the ARG problem leads to link exit time functions that are diffeomorphisms.

Lemma 2 *Let $g(\cdot)$ be a continuously differentiable function on $[0,T]$. If for every scalar $x \in [0,T]$ $g'(x) \neq 0$, then $g(\cdot)$ is invertible on $[0,T]$, its inverse function $g^{-1}(\cdot)$ is continuously differentiable on $[min(g(0), g(T)), max(g(0), g(T))]$ and, $(g^{-1})'(x) = \frac{1}{g'(g^{-1}(x))}$.*

Proof: Since $g(\cdot)$ is a continuously differentiable function, then $g'(\cdot)$ is continuous. Since for every $x \in [0,T]$, $g'(x) \neq 0$, then $g'(\cdot)$ is either positive or negative. Hence, $g(\cdot)$ is either strictly increasing or strictly decreasing. Since every strictly monotonous function is invertible, it follows that $g(\cdot)$ is invertible. Let $g^{-1}(\cdot)$ denote the inverse function of $g(\cdot)$. Then, $g(g^{-1}(x)) = x$. If we differentiate both sides of the above equality, we obtain: $g^{-1\prime}(x)g'(g^{-1}(x)) = 1$. Since $g'(x) \neq 0$ on [0,T], $g'(g^{-1}(x)) \neq 0$. It follows that $g^{-1\prime}(x) = \frac{1}{g'(g^{-1}(x))}$.

□

3.2 Properties of the Feasible Region of the ARG Problem

In this subsection, we establish a few properties of the feasible region $F(ARG)$ of the ARG problem. In Subsections 4.2.1 and 4.2.2, we summarize results due to Kachani [13] that determine the tightest assumptions for the $F(ARG)$ region to be non-empty and for the FIFO property to hold. We start with a network of one link. Subsequently, we show how these results extend to general networks. In Subsection 4.2.3, we establish the boundedness, closedness and convexity of the $F(ARG)$ region.

3.2.1 Network of One Link

Unifying Analysis of Non-Linear and Linear travel time Functions

In this subsection, we consider a network of one link a. Below, we report on a result that establishes that the feasible region of the ARG problem is not empty. This result provides a unifying analysis for both linear and non-linear travel time functions. Corollary 1 shows how linear travel time functions can be interpreted as a limit case of non-linear travel time functions and why linear travel time functions lead to stronger results.

Theorem 1 *[13] If the pair $(D_a(\cdot), u_a(\cdot))$ satisfies the following conditions:*
(A1) The link travel time function $D_a(\cdot)$ is continuously differentiable, and there exist two non-negative constants B_{1a} and B_{2a} such that for every link load X_a, $0 \leq B_{1a} \leq D'_a(X_a) < B_{2a}$.
(A2) The link entrance flow rate function $u_a(\cdot)$ is Lebesgue integrable, non negative and bounded from above by a positive real number M_a on $[0, T]$.
(A3) $M_a \leq \frac{1}{B_{2a} - B_{1a}}$.
Then the feasible region $F(ARG)$ has the following properties:
(1) $F(ARG)$ is well defined, that is, the link load $X_a(\cdot)$, the exit flow rate $v_a(\cdot)$, and the cumulative variables are uniquely determined by the link travel time function $D_a(\cdot)$ and the link entrance flow rate $u_a(\cdot)$ over the analysis period $[0, T_\infty]$.
(2) The Strong FIFO property holds.

Remarks:

- Intuitively, Conditions (A1)-(A3) are the minimal conditions to ensure that the FIFO property is verified. Indeed, $B_{2a} - B_{1a}$ represents the maximum variation of travel time in terms of X_a (the total number of vehicles in the link). During a time interval Δt of decrease in the number of vehicles in the link, the variation in the number of vehicles $X_a(t) - X_a(t + \Delta t)$ is bounded by the quantity $M_a \Delta t$. Therefore, the variation of travel time $D_a(X_a(t)) - D_a(X_a(t + \Delta t))$ is bounded by $(B_{2a} - B_{1a})M_a \Delta t$. Using Condition (A3), it is in turn bounded by Δt. Hence, $s_a(t) = t + D_a(X_a(t)) \leq t + \Delta t + D_a(X_a(t + \Delta t)) = s_a(t + \Delta t)$. Therefore, the FIFO property is verified in this case. On the other hand, during a time interval Δt of increase in the number of vehicles, since the travel time functions are non-decreasing, $D_a(X_a(t)) \leq D_a(X_a(t+\Delta t))$. Therefore, $s_a(t) \leq s_a(t + \Delta t)$, and the FIFO property is also verified in this case.

- Notice that if the link travel time function $D_a(\cdot)$ is linear, then conditions (A1)-(A3) of Theorem 1 simplify significantly. Indeed, in this case, $D'_a(\cdot) = const = B_{1a}$. Moreover, for any arbitrarily small positive scalar ϵ, by introducing $B_{2a} = B_{1a} + \epsilon$, Condition (A1) of Theorem 1 is verified. Furthermore, since Condition (A3) can be rewritten as $M_a \leq \frac{1}{\epsilon}$, M_a can be arbitrarily large. Therefore, the following corollary is obtained.

Corollary 1 *[13] If the pair $(D_a(\cdot), u_a(\cdot))$ satisfies the following conditions:*
(B1) The link travel time function $D_a(\cdot)$ is linear and non-negative.
(B2) The link entrance flow rate function $u_a(\cdot)$ is Lebesgue integrable and non-negative.
Then conditions (A1)-(A3) of Theorem 1 also hold.

In summary, Theorem 1 establishes that by constraining the link entrance capacity to the variation of travel time with respect to the number of vehicles in the link, we limit the effect of flow rate changes, and ensure that the FIFO property holds. Furthermore, when the FIFO property holds, the feasible region $F(ARG)$ is non-empty and we can uniquely determine the exit flow rate in terms of the link entrance flow rate.

Tightness of the Conditions of Theorem 1

Below, we illustrate using a counter-example that conditions (A1)-(A3) in Theorem 1 are tight.

Theorem 2 *For any arbitrarily small positive scalar δ, there exist a link travel time function $D_a(\cdot)$ and a production flow rate function $u_a(\cdot)$ that verify the following conditions*
(C1) $D_a(\cdot)$ is continuously differentiable and nondecreasing;
(C2) $u_a(\cdot)$ is non-negative, Lebesgue integrable and bounded from above by M_a;
(C3) $\frac{1}{M_a} < Max\{D'_a(X_a), X_a \in R\}$ - $Min\{D'_a(X_a), X_a \in R\} \leq \frac{1}{M_a} + \delta$,
and that violate the FIFO property.

Proof: To show this, we will construct a link entrance flow rate function $u_a(\cdot)$ and a link travel time function $D_a(\cdot)$ such that vector $(u_a(\cdot), D_a(\cdot))$ verifies conditions (C1)-(C3) of Theorem 2, and violates the FIFO property.

Let δ and M_a be any positive scalars, B_{1a} and β be any non-negative scalars, and ϵ and α be any two positive scalars such that $\epsilon < \alpha$. Let ω be a positive scalar such that $\omega \in (\alpha, 2\alpha - \epsilon)$.

We first construct the link travel time function $D_a(\cdot)$. We define $D_a(\cdot)$ on three contiguous intervals: $[0, (\alpha - \epsilon)M_a]$, $((\alpha - \epsilon)M_a, \alpha M_a)$ and $[\alpha M_a, +\infty)$. On the first and third intervals, $D_a(\cdot)$ is affine with a slope on its first affine piece less than the slope on its second affine piece. On the second interval, $D_a(\cdot)$ is an exponential, nondecreasing and continuously differentiable function. Let X_{a1}, y_{a1}, X_{a2}, y_{a2}, γ_{a1} and γ_{a2} be given by:

$$\begin{array}{lll} X_{a1} = (\alpha - \epsilon)M_a & \text{and,} & y_{a1} = D_a(X_{a1}) = \alpha + B_{1a}(\alpha - \epsilon)M_a, \\ X_{a2} = \alpha M_a & \text{and,} & y_{a2} = D_a(X_{a2}) = \beta + (B_{1a} + \frac{1}{M_a} + \delta)\alpha M_a, \\ \gamma_{a1} = \frac{B_{1a} + \frac{1}{M_a} + \delta}{y_{a2}} - \frac{2}{X_{a2} - X_{a1}} & \text{and,} & \gamma_{a2} = \frac{B_{1a}}{y_{a1}} - \frac{2}{X_{a1} - X_{a2}}. \end{array}$$

Consider the following link travel time function $D_a(\cdot)$:

$$D_a(X_a) = \begin{cases} \alpha + B_{1a}X_a & \text{, if} \quad X_a \in [0, (\alpha - \epsilon)M_a] \\ y_{a2}(\frac{X_a - X_{a1}}{X_{a2} - X_{a1}})^2 e^{\gamma_{a1}(X_a - X_{a2})} + y_{a1}(\frac{X_a - X_{a2}}{X_{a1} - X_{a2}})^2 e^{\gamma_{a2}(X_a - X_{a1})} & \text{, if} \quad X_a \in ((\alpha - \epsilon)M_a, \alpha M_a) \\ \beta + (B_{1a} + \frac{1}{M_a} + \delta)X_a & \text{, if} \quad X_a \in [\alpha M_a, +\infty). \end{cases}$$

Notice that $D_a(\cdot)$ is continuously differentiable and nondecreasing on $[0, +\infty)$.

Consider the production flow rate function $u_a(\cdot)$ given by:

$$u_a(t) = \begin{cases} M_a & \text{, if} \quad t \in [0, \omega), \\ 0 & \text{, if} \quad t \in [\omega, +\infty). \end{cases}$$

Notice that functions $u_a(\cdot)$ and $D_a(\cdot)$, as defined above, verify conditions (C1)-(C3) of Theorem 2.

In what follows, we solve equations (1)-(9) of the $F(ARG)$ on intervals $[0, \alpha - \epsilon)$ and $[\alpha, \omega]$. That is, we express the variables $U_a(\cdot)$, $v_a(\cdot)$, $V_a(\cdot)$, $X_a(\cdot)$, and $s_a(\cdot)$ in terms of the data above. Then, we show that the FIFO property is violated at $t = \omega$.

Notice that for $t \in [0, \alpha - \epsilon)$, $u_a(t) = M_a$. Hence, $U_a(t) = M_a t$. Furthermore, since $\alpha - \epsilon \leq \alpha = D_a(0) = t_1$, it follows that $V_a(t) = 0$. Thus, $X_a(t) = U_a(t) - V_a(t) = M_a t$.

For $t \in [\alpha, \omega]$, there exists $z \in [0, \alpha - \epsilon)$ such that $s_a(z) = t$. Hence $z + D_a(X_a(z)) = t$. Thus, $z + \alpha + B_{1a} M_a z = t$. It follows that $z = s_a^{-1}(t) = \frac{t-\alpha}{1+B_{1a}M_a}$. Using equation (7) of the DNL functional operator, that describes the relationship between the cumulative link exit flow and the link entrance flow rate, we obtain

$$V_a(t) = \int_0^{s_a^{-1}(t)} u_a(w)dw = \int_0^{\frac{t-\alpha}{1+B_{1a}M_a}} M_a dw = \frac{(t-\alpha)}{1+B_{1a}M_a} M_a.$$

Hence, $v_a(t) = \frac{M_a}{1+B_{1a}M_a}$. Therefore, we obtain $X_a(t) = M_a t - \frac{t-\alpha}{1+B_{1a}M_a} M_a = \frac{\alpha M_a + B_{1a} M_a^2 t}{1+B_{1a}M_a}$. Since $t \geq \alpha$, it follows that $X_a(t) \geq \alpha M_a$. Hence, by definition of $D_a(\cdot)$, it follows that $D_a'(X_a(t)) = (B_{1a} + \frac{1}{M_a} + \delta)$.

Next, we show that $s_a'(\omega) < 0$. Indeed, since $s_a'(t) = 1 + D_a^{'}(X_a(t))(u_a(t) - v_a(t))$, it follows that

$$\begin{aligned} s_a'(\omega) &= 1 + D_a'(X_a(\omega))(u_a(\omega) - v_a(\omega)) \\ &= 1 + (B_{1a} + \frac{1}{M_a} + \delta)(0 - \frac{M_a}{1+B_{1a}M_a}) \\ &= -\delta \frac{M_a}{1+B_{1a}M_a} < 0. \end{aligned}$$

This implies that the exit time function $s_a(\cdot)$ is strictly decreasing at $t = \omega$. Hence, the FIFO property is violated for $t = \omega$.

□

3.2.2 Extension to a General Network

In this subsection, we generalize the results we obtained for a single link network to the case of a general network. The following theorem illustrates this generalization.

Theorem 3 *[13]*
Assume that for every (O-D) pair (r, s), the (O-D) pair departure flow rate function $f^{rs}(\cdot)$ is Lebesgue integrable, non-negative, and bounded from above by M_{rs}. Then there exists a vector $(\widetilde{M}_a)_{a \in A} > 0$ such that:
If the link travel time functions $(D_a(\cdot))_{a \in A}$ satisfy the following conditions:
(D1) $D_a(\cdot)$ is continuously differentiable,
(D2) $\forall x \in R$, $0 \leq B_{1a} \leq D_a^{'}(x) < B_{2a}$, and,
(D3) $B_{2a} - B_{1a} \leq \frac{1}{\widetilde{M}_a}$,
then the feasible region $F(ARG)$ has the following properties:
(1) $F(ARG)$ is well defined (that is, the link entrance flow rates $u_a(\cdot)$, the link exit flow rates $v_a(\cdot)$, the link cumulative entrance and exit flows $U_a(\cdot)$ and $V_a(\cdot)$, the link loads $X_a(\cdot)$, and the link and path exit time functions $s_a()$ and $S_p(\cdot)$, can be uniquely determined by the link travel time functions $D_a(\cdot)$ and the path departure flow rates $f_p(\cdot)$ on the analysis period $[0, T_\infty]$),
(2) The Strong FIFO property holds.

3.2.3 Non-Emptiness, Boundedness, Convexity and Closedness of the Feasible Region

In this subsection, we present some properties of the $F(ARG)$ region. These properties are not only useful in understanding the structure of the model, but will also be useful in Subsection 3.3 to establish the existence of a solution to the ARG problem.

Let $D = (D_a)_{a \in A}$, $q = (q^{rs})$ and $\beta = (\beta_{np}(t))_{n \in N_1 \cup N_2, p \in P}$ denote a vector of link travel time functions, a vector of (O-D) departure flow rate functions, and a vector of path splitting rates respectively. Vector $(D(\cdot), q(\cdot), \beta(\cdot))$ is feasible if it verifies conditions (D1)-(D3) of Theorem 3. In this case, using Theorem 3, the link entrance flow rates $u_a(\cdot)$, the link exit flow rates $v_a(\cdot)$, the link cumulative entrance and exit flows $U_a(\cdot)$ and $V_a(\cdot)$, the link loads $X_a(\cdot)$, and the link and path exit time functions $s_a()$ and $S_p(\cdot)$, can be uniquely determined by $(D(\cdot), q(\cdot), \beta(\cdot))$ on the analysis period $[0, T_\infty]$.

Proposition 2 *The feasible region $F(ARG)$ is non-empty and bounded.*

Proof:
This result follows directly from Theorem 3.

□

Proposition 3 *If vectors $(D(\cdot), q(\cdot), \beta_1(\cdot))$ and $(D(\cdot), q(\cdot), \beta_2(\cdot))$ are feasible, then, for every $\lambda \in [0, 1]$, vector $(D(\cdot), q(\cdot), (\lambda\beta_1(\cdot) + (1 - \lambda)\beta_2(\cdot)))$ is also feasible. In this sense, the feasible region $F(ARG)$ is convex.*

Proof: The result of this proposition is immediate. Note that for every node and every time instant, the set of feasible splitting rates form a simplex. Therefore, the set of all feasible splitting rates is a product of simplices, which is convex.

□

Proposition 4 *If a sequence $(\beta^j(\cdot))_{j \in \aleph}$ of vectors of path splitting rates converges to $(\beta(\cdot))$, and, if for every j, vector $(D(\cdot), q(\cdot), \beta^j(\cdot))$ is feasible, then, the limit $(D(\cdot), q(\cdot), \beta(\cdot))$ is also feasible. In this sense, the $F(ARG)$ region is closed.*

Proof:
Let us assume that for all $j \in \aleph$, vectors $(D(\cdot), q(\cdot), \beta^j(\cdot))$ are feasible. Then, it is easy to see that the limit $(D(\cdot), q(\cdot), \beta(\cdot))$ verifies conditions (D1)-(D3) of Theorem 3.

□

3.3 Existence of a Solution to the Anticipatory Route Guidance Problem

In this subsection, we study key properties of the ARG problem that enable us to establish the existence of a solution to the ARG problem.

As established in Subsection 2.3, the ARG problem is equivalent to solving the following two variational inequalities in sequence: Find a vector of flows $f^*_{|\overline{P}_2} \in F(ARG)_{|\overline{P}_2}$ satisfying

$$< S(\widehat{f}), f_{|\overline{P}_2} - f^\star_{|\overline{P}_2} > \quad \geq 0 \quad \forall f_{|\overline{P}_2} \in F(ARG)_{|\overline{P}_2}. \tag{22}$$

Then, find a vector of flows $f^*_{|\overline{P}_1} \in F(ARG)_{|\overline{P}_1}$ satisfying

$$< S(f^*_{|\overline{P}_1}, f^*_{|\overline{P}_2}), f_{|\overline{P}_1} - f^\star_{|\overline{P}_1} > \quad \geq 0, \quad \forall f_{|\overline{P}_1} \in F(ARG)_{|\overline{P}_1}. \tag{23}$$

The existence of a vector of flows $f^*_{|\overline{P}_2} \in F(ARG)_{|\overline{P}_2}$ satisfying variational inequality formulation (22) follows immediately from the continuity of the scalar product $< S(\widehat{f}), . >$ in (22), and the non-emptiness, boundedness, convexity and closedness results of the $F(ARG)$ region established in the previous subsection.

However, the complexity of establishing the existence of a solution to the ARG problem lies in proving the existence of a vector of flows $f^*_{|\overline{P}_1} \in F(ARG)_{|\overline{P}_1}$ satisfying variational inequality formulation (23). We will refer to the functional operator $S(\cdot)_{|\overline{P}_1} = S(., f^*_{|\overline{P}_2})$ as the ARG operator.

3.3.1 Properties of the ARG Operator

In this subsection, we establish some properties of the ARG operator $S(\cdot)_{|\overline{P}_1}$. These properties will be useful in establishing that the ARG problem has a solution. We first introduce a definition from functional analysis (for more details, see Kirillov [19], Kolmogorov and Fomin [20], and Rudin [30]).

Definition 5 *(Weak Continuity):*
(i) A sequence $(f_n)_{n\in\aleph}$ in a normed space is said to converge weakly to f, if, for every bounded linear functional operator $LM(\cdot)$, $(LM(f_n))_{n\in\aleph}$ converges to $LM(f)$.
(ii) A functional operator MP from a normed space to another is said to be weakly continuous if, for every sequence of functions $(f_n)_{n\in\aleph}$ weakly converging to f, the sequence $(||MP(f_n) - MP(f)||)_{n\in\aleph}$ converges to 0.

The proposition below summarizes some results from functional analysis that are useful to establish that the ARG operator is weakly continuous (for more details, see Kirillov [19], and Kolmogorov and Fomin [20]).

Proposition 5 *[19], [20]*
(i) If g and h are two weakly continuous functional operators, then the functional operators $g + h$, $g.h$ and $g(h)$ are weakly continuous.
(ii) If g is a weakly continuous functional operator on the set of real numbers and has a constant sign, then the functional operator $\frac{1}{g}$ is weakly continuous.
(iii) The integral operator from the space of bounded functions on $L^1([0, T_\infty])$ to $L^2([0, T_\infty])$ defined as $f(\cdot) \mapsto \int_0^t f(w)dw$ is weakly continuous.

We establish the weak continuity of the ARG operator.

Theorem 4 *Conditions (D1)-(D3) of Theorem 3 imply that the ARG operator $S(\cdot)_{|\overline{P}_1}$ is weakly continuous.*

Proof: Property (iii) of Proposition 5 implies that $u_a(\cdot) \mapsto U_a(\cdot)$ is weakly continuous. We will prove by induction over the time intervals $[t_j, t_{j+1})$ (where $t_0 = 0$ and $t_{j+1} = s_a(t_j)$), that the functional operators $u_a(\cdot) \mapsto V_a(\cdot)$, $u_a(\cdot) \mapsto X_a(\cdot)$, $u_a(\cdot) \mapsto s_a(\cdot)$, $u_a(\cdot) \mapsto v_a(\cdot)$, $u_a(\cdot) \mapsto s_a^{-1}(\cdot)$ and $u_a(\cdot) \mapsto (s_a^{-1})'(\cdot)$ are weakly continuous.

We first need to establish a preliminary result.

Lemma 3 *Under conditions (A1)-(A3) of Theorem 1, if the link exit time operator* $u_a \mapsto s_a(\cdot)$ *is weakly continuous on the interval* $[t_j, t_{j+1})$*, then its inverse operator* $u_a \mapsto s_a^{-1}(\cdot)$ *is weakly continuous on the interval* $[t_{j+1}, t_{j+2})$*.*

Proof: We assume that the link exit time operator $u_a \mapsto s_a(\cdot)$ is weakly continuous on the interval $[t_j, t_{j+1})$.

From the proof of Theorem 1 in Kachani [13], we know that for every $t \in [t_j, t_{j+1})$, $s_a'(t) \leq \theta_j$, where $\theta_j \in (0, 1)$. Hence, $s_a^{-1}(\cdot)$ is Lipschitz continuous on $[t_{j+1}, t_{j+2})$ with parameter $\frac{1}{\theta_j}$.

Furthermore, for every $t \in [t_j, t_{j+1})$,

$$\begin{aligned} s_a'(t) = 1 + D_a'(X_a(t))(u_a(t) - v_a(t)) &\leq 1 + D_a'(X_a(t))u_a(t), \\ &\leq 1 + B_{2a}M_a. \end{aligned}$$

Hence, $s_a(\cdot)$ is Lipschitz continuous on $[t_j, t_{j+1})$ with parameter $1 + B_{2a}M_a$.

Let $(u_a^k(\cdot))_{k\in\aleph}$ denote a weakly converging sequence of link flow rate functions to $u_a(\cdot)$. Let $s_a^k(\cdot)$ denote the link exit time function corresponding to $u_a^k(\cdot)$.

Furthermore,

$$\begin{aligned} \int_{t_{i+1}}^{t_{i+2}} |(s_a^k)^{-1}(w) - s_a^{-1}(w)|^2 dw &= \int_{s_a(t_a)}^{s_a(t_{i+1})} |(s_a^k)^{-1}(w) - s_a^{-1}(w)|^2 dw \\ &= \int_{t_a}^{t_{i+1}} |(s_a^k)^{-1}(s_a(w)) - s_a^{-1}(s_a(w))|^2 s_a'(w) dw \\ &= \int_{t_a}^{t_{i+1}} |(s_a^k)^{-1}(s_a(w)) - w|^2 s_a'(w) dw \\ &= \int_{t_a}^{t_{i+1}} |(s_a^k)^{-1}(s_a(w)) - (s_a^k)^{-1}(s_a^k(w))|^2 s_a'(w) dw \\ &\leq \frac{1 + B_{2a}M_a}{\theta_j^2} \int_{t_a}^{t_{i+1}} |s_a^k(w) - s_a(w)|^2 dw. \end{aligned}$$

Since $u_a \mapsto s_a(\cdot)$ is weakly continuous on the interval $[t_j, t_{j+1})$, it follows that $u_a \mapsto s_a^{-1}(\cdot)$ is weakly continuous on the interval $[t_{j+1}, t_{j+2})$.

□

Induction Proof:
Base Case: Time interval $[t_0, t_1)$.
On $[t_0, t_1)$, $v_a(t) = V_a(t) = 0$ and $X_a(t) = U_a(t)$. Hence, the operators $u_a(\cdot) \mapsto V_a(\cdot)$, $u_a(\cdot) \mapsto X_a(\cdot)$, and $u_a(\cdot) \mapsto v_a(\cdot)$ are weakly continuous. Furthermore, since $s_a(t) = t + D_a(X_a(t))$, and $D_a(\cdot)$ are continuous functions (and therefore weakly continuous), using property (i) of Proposition 5, it follows that the operator $u_a(\cdot) \mapsto s_a(\cdot)$ is weakly continuous. Using Lemma 3 and property (i) of Proposition 5, it follows that the operator $u_a(\cdot) \mapsto s_a^{-1}(\cdot)$ is also weakly continuous on $[t_1, t_2)$.

Moreover $(s_a^{-1})'(t) = \frac{1}{s_a'(s_a^{-1}(t))} = \frac{1}{1+D_a'(X_a(s_a^{-1}(t)))(u_a(s_a^{-1}(t))-v_a(s_a^{-1}(t)))}$. Using properties (i) and (ii) of Proposition 5, we obtain that $u_a(\cdot) \mapsto (s_a^{-1})'(\cdot)$ is also weakly continuous on $[t_1, t_2)$.

Induction Step: Time interval $[t_{j+1}, t_{j+2})$. From the induction hypothesis, we know that the operators $u_a(\cdot) \mapsto s_a^{-1}(\cdot)$ and $u_a(\cdot) \mapsto (s_a^{-1})'(\cdot)$ are weakly continuous on $[t_{j+1}, t_{j+2})$. Since $v_a(w) =$

$u_a(s_a^{-1}(w)).(s_a^{-1})'(w)$, using property (i) of Proposition 5, it follows that the operator $u_a(\cdot) \mapsto v_a(\cdot)$ is weakly continuous. Property (iii) of Proposition 5 implies that $u_a(\cdot) \mapsto V_a(\cdot)$ is weakly continuous. Since $X_a(\cdot) = U_a(\cdot) - V_a(\cdot)$ and $s_a(t) = t + D_a(X_a(t))$, property (i) of Proposition 5 implies that $u_a(\cdot) \mapsto X_a(\cdot)$ and $u_a(\cdot) \mapsto s_a(\cdot)$ are weakly continuous.

Using Lemma 3, it follows that $u_a(\cdot) \mapsto s_a^{-1}(\cdot)$ is also weakly continuous on interval $[t_{j+2}, t_{j+3})$. Moreover, since $(s_a^{-1})'(t) = \frac{1}{1+D'_a(X_a(s_a^{-1}(t)))(u_a(s_a^{-1}(t))-v_a(s_a^{-1}(t))}$, using properties (i) and (ii) of Proposition 5, we obtain that $u_a(\cdot) \mapsto (s_a^{-1})'(\cdot)$ is also weakly continuous on $[t_{j+2}, t_{j+3})$. The induction proof is now complete.

For all links a, the head of which is an origin node (i.e. $a^1 \in N_1$) and for all paths p originating at a^1, as introduced Subsection 2.2, the flow conservation equations can be expressed as

$$u_{ap}(t) \quad = \quad f_{s(p)p}(t). \tag{24}$$

Furthermore, the path travel times $S_p(t, f_p(t))$ are the difference between the composition of link exit time functions $s_a(\cdot)$ for all links $a \in p$ and the entrance time t. Using property (i) of Proposition 5, it follows that the ARG operator $S(\cdot)_{|\overline{P}_1}$ is weakly continuous.

□

We now define the notion of pseudo-monotonicity introduced by Brezis [5] and show that the ARG operator $S(\cdot)_{|\overline{P}_1}$ is pseudo-monotonous.

Definition 6 *(Pseudo-monotonicity) A bounded functional operator MP is pseudo-monotonous over X if, whenever a sequence $(f^k)_{k\in\aleph} \in X^{\aleph}$ weakly converging to f satisfies limsup $< MP(f^k), f^k - x > \leq 0$, $\forall x \in X$, it also satisfies liminf $< MP(f^k), f^k - x > \geq < MP(f), f - x >$, $\forall x \in X$.*

Lemma 4 *The ARG operator $S(\cdot)_{|\overline{P}_1}$ is pseudo-monotonous over the $F(ARG)$ region.*

Proof: Notice that $S(\cdot)_{|\overline{P}_1}$ is weakly continuous on the $F(ARG)$ region, and from Proposition 2, the $F(ARG)$ region is bounded. Therefore, $S(\cdot)_{|\overline{P}_1}$ is a bounded operator. Let $diam(F(ARG))$ denote the diameter of the $F(ARG)$ region and let $(f_k)_{k\in\aleph}$ denote a sequence of elements of the $F(ARG)_{|\overline{P}_1}$ region weakly converging to f. Then, for $y \in F(ARG)$,

$$\begin{aligned} < S(f_k)_{|\overline{P}_1} - S(f)_{|\overline{P}_1}, f_k - y > &\leq ||S(f_k)_{|\overline{P}_1} - S(f)_{|\overline{P}_1}||.||f_k - y|| \\ &\leq diam(F(ARG)).||S(f_k)_{|\overline{P}_1} - S(f)_{|\overline{P}_1}||. \end{aligned}$$

Since $S(\cdot)_{|\overline{P}_1}$ is weakly continuous on the $F(ARG)$, it follows that the sequence $(||S(f_k)_{|\overline{P}_1} - S(f)_{|\overline{P}_1}||)_{k\in\aleph}$ converges to 0. Hence $lim_{k\to\infty} < S(f_k)_{|\overline{P}_1} - S(f)_{|\overline{P}_1}, f_k - y >= 0$. It follows that:

$$\begin{aligned} lim_{k\to\infty} < S(f_k)_{|\overline{P}_1}, f_k - y > &= lim_{k\to\infty} < S(f)_{|\overline{P}_1}, f_k - y > \\ &= < S(f)_{|\overline{P}_1}, f - y > . \end{aligned}$$

Hence, the ARG operator $S(\cdot)_{|\overline{P}_1}$ is pseudo-monotonous over the $F(ARG)$ region.

□

3.3.2 Existence of a Solution to the ARG problem

In this subsection, we establish one of the fundamental results of this paper. That is, we illustrate that under weak assumptions, the ARG problem has a solution.

Lemma 5 *(Brezis [5], [6])*
Let K be an non-empty, bounded, convex and closed set. Let $A(\cdot)$ denote a functional operator from K to L that is pseudo-monotonous functional operator over K. Then, for every vector $z \in L$, there exists a vector $f^* \in K$ such that $< A(f^*), f - f^* > \geq < z, f - f^* >$ is verified for every vector $f \in K$.

For more details on the above lemma, see [5], [6], [23], [24], and [25].

Theorem 5 *Under conditions (D1)-(D3) of Theorem 3, the ARG problem has a solution.*

Proof: Under conditions (D1)-(D3), Theorem 4 applies. Hence, the ARG operator $S(\cdot)_{|\overline{P}_1}$ is weakly continuous. Using Lemma 4, it follows that the ARG operator $S(\cdot)_{|\overline{P}_1}$ is pseudo-monotonous over the $F(ARG)$ region. From Propositions 2-4, the $F(ARG)$ region is non-empty, bounded, closed and convex. Using Lemma 5 with $K = F(ARG)$, $A(\cdot) = S(\cdot)_{|\overline{P}_1}$ and $z = 0$, and using the variational inequality formulation (23), it follows that the ARG problem has a solution.

□

4 Solution Algorithm and Computational Results

In this section, we illustrate the ideas of the previous section. We describe a dynamic traffic model and a heuristic algorithm to compute consistent anticipatory route guidance. In Subsection 4.1 we describe the traffic model and solution algorithm, and in Subsection 4.2 we present some computational results.

4.1 Description of a Solution Approach

4.1.1 Dynamic Traffic Model

Bottom in [3] develops a software system to operationalize and explore properties of the consistent anticipatory route guidance problem. It resembles a discrete-time traffic simulation model in that increments of flow are individually routed and moved through the network at fixed time steps. However, unlike a simulator, the flow increments are not individual vehicles but rather fractions of a flow unit. With sufficiently small flow increments (in the experiments below, each unit of flow was divided into ten increments) and time steps (one second steps were used), a good approximation to a continuous-flow continuous-time DTA model can be obtained.

The software system's dynamic network loading map implements a store-and-forward protocol (Gazis [11]). The loader's inputs are a network description in terms of nodes, links and (enumerated) paths; a schedule of time-dependent trip (departure) rates by OD pair; and a table of time-dependent path splits at all network decision points (origins or en route points). The loader's output is a table of average link traversal times by link and by time of link entry. Link volume profiles (i.e., the number of flow units that entered, exited and remained on each link in each time period) can also be output on request. To compute these outputs, the loader generates flow increments in accordance with the input OD trip rates, allocates them to paths in accordance with the input path splits, and simulates their movement through the network, tracking their time of entry and exit from each link.

The distinguishing feature of this traffic simulator is its representation of links as deterministic FIFO (first-in-first-out) single-server queues with given exit (service) capacities. When a flow increment enters a link, its earliest possible exit time is calculated from the link's length and fixed speed. The increment is then placed at the tail of the link's queue. As each successive increment at the head of the queue is processed and moves on to a downstream link, each following increment advances in position until it too arrives at the queue head and is able to move in turn to the next link on its path. As a result, the time taken by a flow increment to traverse a link consists of two terms: a constant moving time, depending on the link length and free flow speed, plus a queuing time (possibly zero), depending directly on the number of flow increments in the queue ahead of it and inversely on the link capacity. Thus, when flow can be assumed to be uniformly spread over the link, the link performance function (link traversal time as a function of its volume) consists of two linear segments: a constant, followed by a segment with positive slope inversely proportional to the capacity. (The model is also able to represent link vehicle storage capacity and spillback, but this feature was not used in the computational experiments described here.)

The general idea for this loader is derived from the dynamic network loading component of the QM model described in Simon and Nagel [32] and the FastLane model of Gawron [10], but the implementation and many details are completely different.

Each flow increment is routed according to a minimum travel time criterion. Uninformed drivers use free-flow travel times for their minimum path determination. Drivers who pass a VMS receive from it the latest estimate of actual dynamic travel times, and recompute minimum paths using that information. (Some drivers may also receive dynamic travel time estimates at the origin, but this feature was not applied here.)

To stabilize the behavior of the model and avoid the discontinuous changes in path flows that typically result from strict shortest path routing, drivers' path choice decisions are represented by a logit-form path choice model in which a path's disutility is proportional to the ratio of its travel time to the travel time of the fastest path currently available from the origin to the destination. The coefficient of this ratio was set to a high number (-10.0), resulting in a good approximation to an all-fastest-path route choice model. Thus, if multiple paths to a destination have equal and minimum travel times, flow will be split equally between them; while paths having a slightly longer travel time will receive a small fraction of the flow.

4.1.2 Solution Algorithm

In what follows we develop an algorithm for solving consistent guidance problems. Although we do not have a convergence proof for the algorithm, it works well in all applications attempted; it thus appears to be a reasonable heuristic.

The algorithm is an averaging procedure that formally resembles algorithms originally proposed by Robbins and Monro [29] and by Blum [2] for root finding in a stochastic approximation setting. (Transportation applications of this algorithm frequently refer to it as the Method of Successive Averages (MSA) [31].) However, in the application here the algorithm has a natural interpretation as a solution method for the equivalent fixed point formulation presented in Section 2.4 above.

Recall the re-formulation of the ARG VI problem as two fixed point sub-problems:

For uninformed drivers: Find a vector of flows $f^\star_{|\overline{P}_2} \in F(ARG)_{|\overline{P}_2}$ satisfying

$$f^\star_{|\overline{P}_2} \in \mathrm{argmin}_{f_{|\overline{P}_2} \in F(ARG)_{|\overline{P}_2}} < S(\widehat{f}), f_{|\overline{P}_2} > . \tag{25}$$

For informed drivers: Find a vector of flows $f^\star_{|\overline{P}_1} \in F(ARG)_{|\overline{P}_1}$ satisfying

$$f^\star_{|\overline{P}_1} \in \mathrm{argmin}_{f_{|\overline{P}_1} \in F(ARG)_{|\overline{P}_1}} < S(f^\star_{|\overline{P}_1}, f^\star_{|\overline{P}_2}), f_{|\overline{P}_1} > . \tag{26}$$

The first sub-problem, for the uninformed drivers, is essentially a shortest path problem based on the default (and fixed) flow estimates $\widehat{f}$ and the corresponding travel times $S(\widehat{f})$. It can be solved by a single application of a dynamic shortest path algorithm using the default times. The second sub-problem, for the informed drivers, is again a shortest path problem, but based on the flows of informed and uninformed drivers, and the corresponding travel times, at optimality. The solution can be found using a recursive averaging algorithm.

The algorithm operates in the domain of time-dependent subpath splits $P(t) = P_n(t)$, for all origins and decision nodes n in the network. Here time is discretized using one second steps over an analysis horizon $t = 0 \ldots T_\infty$. As explained above, subpath splits determine the subpath flows f via the dynamic network loader. The dynamic network loader moves the corresponding OD flows through the network according to these splits, and determines the resulting link traversal times (as well as link flows and corresponding path variables). The loader can thus be viewed as a map S from the domain of time-dependent subpath splits P to the domain of time-dependent link traversal times C. Based on these traversal times, the all-fastest-path algorithm described above is applied to determine a new set of time-dependent subpath splits. This can be represented as a routing map D from the domain C of link travel times to the domain P of subpath splits.

In each iteration, the algorithm attempts to determine an improved estimate of the optimal subpath splits (flows) based on the most recent estimate of these splits. The averaging algorithm proceeds as follows in iteration i:

$$P^{i+1} = P^i + \frac{1}{i+1}(D \circ S(P^i) - P^i);\ i = 0 \ldots$$

The correction $D \circ S(P^i) - P^i$ that the algorithm applies to iteration i's estimate P^i of the solution is weighted by the factor $1/(i+1)$, with the overall correction becoming negligible at some point.

Progress of the algorithm toward convergence can be tracked from iteration to iteration via the value of the norm $\|P - D \circ S(P)\|_2$ of the difference between "input" and "output" path splits. The presence of uninformed drivers who do not respond to the updated path splits prevents the norm from reaching 0; however, when the convergence curve becomes approximately horizontal, it can be concluded that convergence has been approximately reached.

4.2 Computational Experiments

4.2.1 Description of the test network

Runs were made using a simple 14-link network with a single OD pair and eleven OD paths. Figure 2 shows this network. All links are single lane and 1 km long except for links 2 (100-201) and 9 (201-200), which are 1.5 km long. Free flow speed on all links is 100 km/hr. All links have a capacity of 3,600 vehicles/hour except for link 6 (102–200), which has a reduced capacity of 900 vehicles/hour.

The flow rate from origin 1 to destination 2 is 10,800 vehicles/hour (3 vehicles/second) during 1200 seconds, after which it drops to 0. There is a source of link traversal time estimates on link 1 (100–101); this may be thought of as a variable message sign (VMS). Uninformed vehicles on this link access the VMS information and use it to reconsider their choice of path downstream from node 101.

4.2.2 Results

Figure 3 is a graph of the path split convergence norm by iteration. It can be seen that the norm decreases rapidly from its initial values, and reaches a stable range of values by iteration 50. However, there is considerable variability in the norm values from one iteration to the next. The variability results from a number of factors: the presence of a significant fraction of drivers who do not receive guidance; abrupt

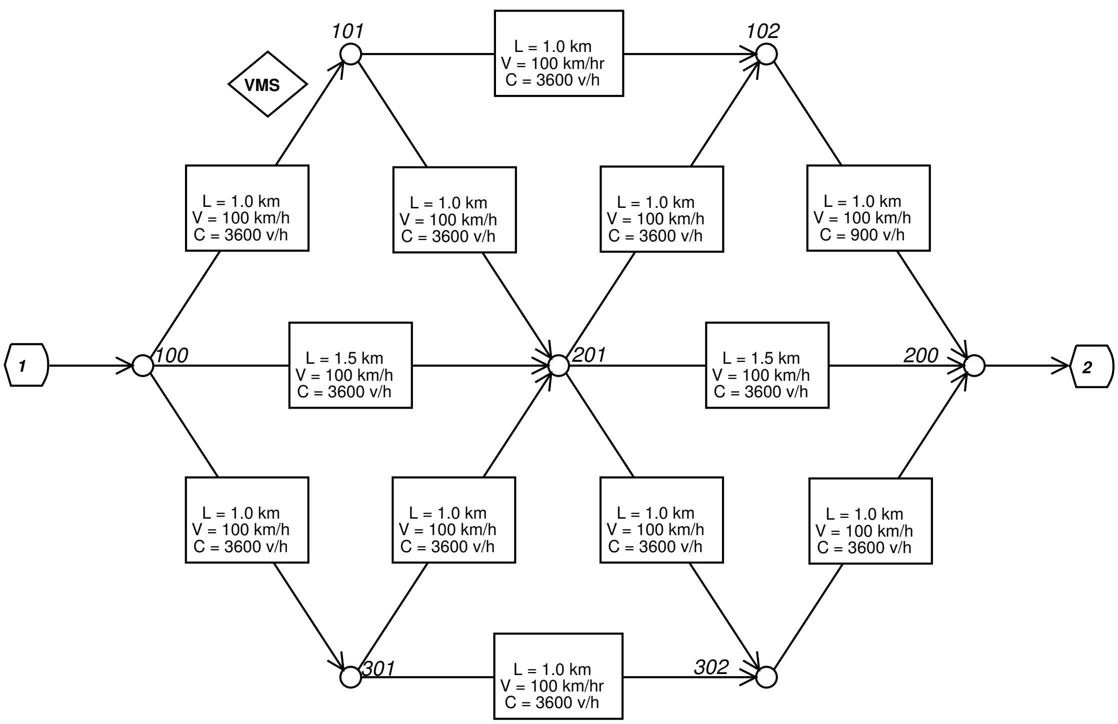

Figure 2: Test network

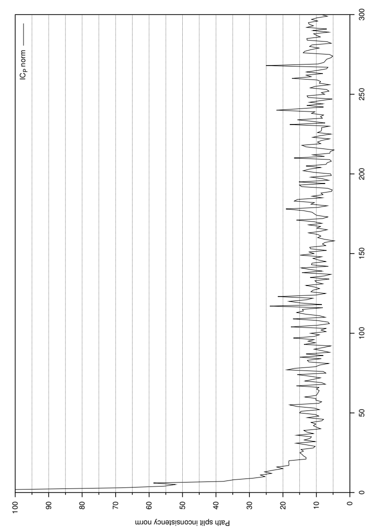

Figure 3: Convergence of fixed point algorithm

changes in path splits inherent in shortest path routing; the fact that $D \circ S(P^i) - P^i$ is not guaranteed to be a descent direction; and minor stochastic effects in the model implementation. Empirically, we have observed that the first of these factors appears to be the most significant: as the fraction of informed drivers increases, the convergence curve both becomes smoother and also has a lower asymptote.

Figures 4 and 5 show the link volumes and traversal times from the computed solution (iteration 300).

5 Conclusions and Future Steps

In this paper, we studied the Anticipatory Route Guidance problem. We proposed a variational inequality (VI) formulation, the first general analytical formulation of this problem. We also presented a fixed-point formulation of the problem and established the equivalence of the two formulations. We provided a unifying analysis for the route guidance dynamic network loading (DNL) problem for both linear and non-linear travel time functions. Under sufficient conditions on the path flow rate functions and the travel time functions, we established that the feasible region $F(ARG)$ of the anticipatory route guidance problem is non-empty, and that the FIFO property holds. We provided a generic counterexample illustrating that the assumptions that we imposed to ensure that FIFO holds are the tightest possible. We established key properties of the feasible region, as a function of the path flow rate functions, such as boundedness, closedness and convexity. Furthermore, we established under weak assumptions the existence of a solution to the ARG problem. Finally, we proposed a fixed-point solution algorithm based on averaging methods and reported on some computational results.

References

[1] M. Ben-Akiva, M. Bierlaire, J. A. Bottom, H. Koutsopoulos, and R. Mishalani. Development of a Route Guidance Generation System for Real-Time Application. In *Proceedings of the 8th IFAC Symposium on Transportation Systems*, pages 433–438, Chania, Greece, 1997.

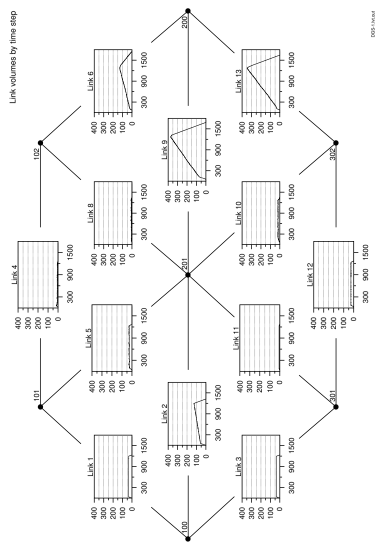

Figure 4: Link volume trajectories (vehicles)

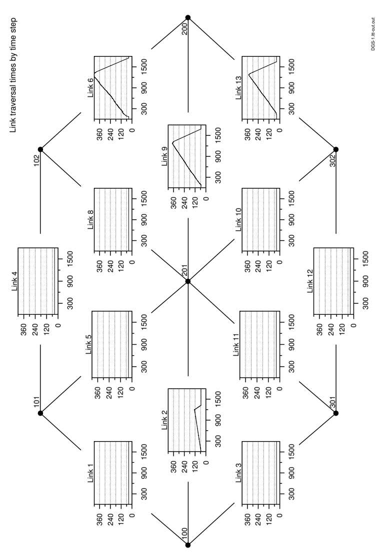

Figure 5: Link time trajectories (seconds)

[2] Julius R. Blum. Multidimensional stochastic approximation methods. *Annals of Mathematical Statistics*, 25(4):737–744, December 1954.

[3] J. A. Bottom. *Consistent Anticipatory Route Guidance*. PhD thesis, Center for Transportation Studies, Massachusetts Institute of Technology, 2000.

[4] P. H. Bovy and N. J. Van Der Zijpp. The Close Connection between Dynamic Traffic Management and Driver Information Systems. In Richard Emmerink and Peter Nijkamp, editors, *Behavioural and Network Impacts of Driver Information Systems*, pages 355–370. Ashgate Publishing Company, 1999.

[5] H. Brezis. Equations and Inequations Non Lineaires dans les Espaces Vectoriels en Dualite. *Annales Institut Fourier*, 18:115–176, 1968.

[6] H. Brezis. Inequations Variationnelles Asssociees a des Operateurs d'Evolution. *NATO Summer School, Venise*, 1968.

[7] G. E. Cantarella. A general fixed-point approach to multi-mode multi-user equilibrium assignment with elastic demand. *Transportation Science*, 31(2):107–128, May 1997.

[8] L. Engelson. Self-Fulfilling and Recursive Forecasts - An Analytical Perspective for Driver Information Systems. In *Proceedings of the 8th IATBR Meeting*, Austin, Texas, 1997.

[9] T. L. Friesz, D. Bernstein, T. E. Smith, R. L. Tobin, and B. W. Wie. A Variational Inequality Formulation of the Dynamic Network User Equilibrium Problem. *Operations Research*, 41(1):179–191, 1993.

[10] Christian Gawron. *Simulation-Based Traffic Assignment*. PhD thesis, University of Köln, 1998.

[11] D. C. Gazis. Modeling and optimal control of congested transportation systems. *Networks*, 4:113–124, 1974.

[12] R. Hamerslag and E. C. van Berkum. Effectiveness of Information Systems in Networks with and without Congestion. In *Transportation Research Record*, volume 1306, pages 14–21, Washington, D.C., 1991. Transportation Research Board, National Research Council.

[13] S. Kachani. Analytical Dynamic Traffic Flow Models: Formulation, Analysis and Solution Algorithms. Master's thesis, Operations Research Center, Massachusetts Institute of Technology, 2000.

[14] S. Kachani and G. Perakis. A Fluid Model of Spillback and Bottleneck Phenomena for Determining Dynamic Travel Times. *Submitted for Publication*, 2001.

[15] S. Kachani and G. Perakis. Modeling Travel Times in Dynamic Transportation Networks; A Fluid Dynamics Approach. *Submitted for Publication*, 2001.

[16] S. Kachani and G. Perakis. Second-Order Fluid Dynamics Models for Travel Times in Dynamic Transportation Networks. In *Proceedings of the 4th International IEEE Conference on Intelligent Transportation Systems*, 2001.

[17] S. Kachani and G. Perakis. A Fluid Model of Pricing and Inventory Management for Make-to-Stock Manufacturing Systems. *Submitted for Publication*, 2002.

[18] D. E. Kaufman, R. L. Smith, and K. E. Wunderlich. An Itererative Routing/Assignment Method for Anticipatory Real-Time Route Guidance. In *Proceedings of the Conference on Vehicle Navigation and Information Systems*, pages 693–700, 1991.

[19] A. A. Kirillov. *Theorems and Problems in Functional Analysis*. Springer-Verlag, New-York, 1982.

[20] A. N. Kolmogorov and S. V. Fomin. *Elements of the Theory of Functions and Functional Analysis*. Graylock Press, Rochester, N.Y., 1957.

[21] Haris N. Koutsopoulos and Tsippy Lotan. Effectiveness of motorist information systems in reducing traffic congestion. In D. H. M. Reekie, E. R. Case, and J. Tsai, editors, *Proceedings of the First Vehicle Navigation and Information Systems Conference (VNIS89)*, pages 275–281, Toronto, Ontario, Canada, 1989.

[22] J. Lindley. Urban Freeway Congestion: Quantification of the Problem and Effectiveness of Potential Solutions. *ITE Journal*, 1987.

[23] J. L. Lions. Sur Quelques Proprietes des Solutions d'Inequations Variationnelles. *C. R. Acad. Sc. Paris, 267*, pages 631–633, 1968.

[24] J. L. Lions. *Quelques Methodes de Resolution des Problemes aux Limites Non Lineaires*. Dunod, Paris, 1969.

[25] J. L. Lions. Some Remarks on Variational Inequalities. *Int. Conference on Functional Analysis, Tokyo*, 1969.

[26] H. Mahmassani and R. Herman. Dynamic User Equilibrium Departure Time and Route Choice on Idealized Traffic Arterials. *Transportation Science*, 18(4):362–384, 1984.

[27] Patrice Marcotte and Sang Nguyen. Hyperpath formulations of traffic assignment problems. In Patrice Marcotte and Sang Nguyen, editors, *Equilibrium and Advanced Transportation Modelling*, chapter 9, pages 175–200. Kluwer, 1998.

[28] S. Nguyen and S. Pallottino. Equilibrium traffic assignment for large scale transit networks. *European Journal of Operations Research*, 37(2):176–186, 1988.

[29] Herbert Robbins and Sutton Monro. A stochastic approximation method. *Annals of Mathematical Statistics*, 22(3):400–407, 1951.

[30] W. Rudin. *Principles of Mathematical Analysis*. McGraw-Hill, New-York, 1976.

[31] Y. Sheffi. *Urban Transportation Networks*. Prentice-Hall, Englewood, NJ, 1985.

[32] Patrice M. Simon and Kai Nagel. Simple queuing model applied to the city of Portland. *International Journal of Modern Physics C*, 10(5):941–960, 1999.

[33] Heinz Spiess and Michael Florian. Optimal strategies: A new assignment model for transit networks. *Transportation Research B*, 23B(2):83–102, 1989.

[34] J. H. Wu, Y. Chen, and M. Florian. The Continuous Dynamic Network Loading Problem: A Mathematical Formulation and Solution Method. *Transportation Research B*, 32(3):173–187, 1995.