Annals of Operations Research 60(1995)81–113

System optimal and user equilibrium time-dependent traffic assignment in congested networks

Srinivas Peeta
School of Civil Engineering, Purdue University, West Lafayette, IN 47907-1248, USA
E-mail: peeta@ecn.purdue.edu

Hani S. Mahmassani
Center for Transportation Research, The University of Texas at Austin, Austin, TX 78712, USA

This paper formulates two dynamic network traffic assignment models in which O-D desires for the planning horizon are assumed known a priori: the system optimal (SO) and the user equilibrium (UE) time-dependent traffic assignment formulations. Solution algorithms developed and implemented for these models incorporate a traffic simulation model within an overall iterative search framework. Experiments conducted on a test network provide the basis for a comparative analysis of system performance under the SO and UE models.

1. Introduction

The advent of Intelligent Transportation Systems (ITS) has generated considerable interest in the area of dynamic traffic assignment, of time-dependent origin-destination (O-D) trip desires in a traffic network of paths to their respective destinations so as to achieve certain individual or system-wide objectives. Dynamic traffic assignment encompasses a wide variety of problems, each corresponding to different sets of decision variables and underlying behavioral assumptions, and possessing varying data requirements and capabilities in terms of representing the traffic system or control actions. A comprehensive review of the literature in this area is outside the scope of this paper and can be found in Peeta [13].

One of the ITS technologies, Advanced Traveler Information Systems (ATIS), provides travelers real-time information on existing traffic conditions and/or instructions on route selection from their current locations to their destinations. The dynamic traffic assignment capabilities required for ATIS must serve the following principal functions. They must provide a descriptive capability that allows a controller or analyst to determine, for partially or completely known origin-destination (O-D) trip desires, the time-varying link flow patterns that result from the path choice decisions made by motorists, in response to real-time information supplied by the ATIS controller. This capability is needed for the off-line evaluation of alternative information supply strategies, for real-time prediction, as well as for determining what information to provide to motorists.

The other dynamic assignment capability is to address the problem faced by a central controller with partial or complete information on time-dependent O-D trip desires as well as current network conditions (prevailing link traffic descriptors, loading, incidents), that seeks to route suitably equipped vehicles from their current positions (including initial origins and intermediate locations) to their respective destinations so as to achieve certain system-wide objectives, subject to individual routing constraints. In other words, the controller seeks to direct users to paths that "optimize" the overall system performance, subject to fairness and reasonableness constraints for the individual users. Such a capability reflects a normative perspective and would be used on-line for the above purpose, or off-line to determine initial routing assignments for the routine and historically known loading patterns, which would be subsequently updated on-line.

Several categories of problems can be identified under the umbrella of dynamic assignment. Two principal bases for classification in the ITS context include:

(1) Number of user classes: (i) the single class dynamic traffic assignment problem where the network users are homogenous in terms of information availability, the type of information supplied to them, and their response to the supplied information (for example, Friesz et al. [4], Ghali and Smith [5], Janson [7], Ran et al. [15], Mahmassani and Peeta [10, 11]), and (ii) multiple user classes based on information availability, information supply strategy, and response behavior (Peeta [13], Mahmassani et al. [9]).

(2) Information availability on O-D trip desires: (i) full a priori information is available on the time-dependent O-D trips over the entire planning horizon (virtually all existing dynamic traffic assignment models fall in this category), and (ii) information on O-D trips is available in quasi real-time (Peeta [13], Peeta and Mahmassani [14]).

Because assignment decisions can in fact be determined a priori for the entire duration of interest under the full information scenario, we prefer to refer to it as the time-dependent assignment problem, reserving the term "dynamic" to the case where assignment decisions are made in real-time or quasi real-time as O-D information becomes available only progressively over the duration of interest. This paper addresses the single user class time-dependent traffic assignment problems for the SO and UE cases.

2. Problem definition

This section specifies the notation for, and defines the SO and UE single user class time-dependent traffic assignment (TDTA) problems. It then discusses issues relevant to the formulation of these problems.

2.1. DEFINITION OF VARIABLES AND NOTATION

The following notation is used to represent variables in the formulations:

i = subscript for origin node, $i \in I$,
j = subscript for destination node, $j \in J$,
n = node in the network, $n \in N$,
a = superscript for a link (or arc) in the network, $a \in A$,
k = subscript for a path in the network, $k \in K_{ij}$, for $i \in I$ and $j \in J$,
τ = superscript denoting departure (or start) time interval, $\tau = 1,\ldots,T$,
t = superscript denoting current time interval, $t = 1,\ldots,T$,
T' = total duration (peak period) for which assignments are to be made,
Δ = length of a time interval (equal to T'/T),
r_{ij}^{τ} = number of vehicles who wish to depart from i to j in period τ,
r_{ijk}^{τ} = number of vehicles who wish to depart from i to j in period τ assigned to path k,
$\delta_{ijk}^{\tau ta}$ = time-dependent link-path incidence indicator, equal to 1 if vehicles going from i to j assigned to path k at time τ are on link a in period t, i.e.,

$$[\delta_{ijk}^{\tau ta} = 1, \text{ if } r_{ijk}^{\tau} \text{ is on arc } a \text{ during period } t,$$
$$= 0, \text{ if arc } a \text{ does not belong to path } k,$$
$$= 0, \text{ if } \tau > t,$$
$$= 0, \text{ if } r_{ijk}^{\tau} \text{ is not on arc } a \text{ during period } t],$$

T_{ijk}^{τ} = experienced path travel time for vehicles going from i to j that are assigned to path k at time τ,
$x_{ijk}^{\tau ta}$ = number of vehicles going from i to j assigned to path k in period τ that are on link a at the beginning of period t,
$d_{ijk}^{\tau ta}$ = number of vehicles (i to j) assigned to path k in period τ which enter arc a in period t,

$m_{ijk}^{\tau ta}$ = number of vehicles (i to j) assigned to path k in period τ which exit link a in period t,
x^{ta} = total number of vehicles on link a at the beginning of period t,
d^{ta} = total number of vehicles which enter link a in period t,
m^{ta} = total number of vehicles which exit link a in period t,
I_n^t = number of vehicles generated at node n in period t,
O_n^t = number of vehicles exiting the network through node n in period t,
$C(n)$ = set of links incident to node n,
$B(n)$ = set of links incident from node n.

2.2. THE TIME-DEPENDENT TRAFFIC ASSIGNMENT PROBLEM

Consider a traffic network represented by a directed graph $G(N, A)$, where N is the set of nodes and A the set of directed arcs. A node can represent a trip origin and/or a destination and/or a junction of physical links. A network with multiple origins $i \in I$ and destinations $j \in J$ is considered for generality. The analysis period of interest, taken here as the peak period or the planning horizon T', is discretized into small intervals $t = 1,\ldots,T$. Given a set of time-dependent O-D vehicle trip desires for the peak period (or planning horizon), expressed as the number of vehicle trips r_{ij}^{τ} leaving node i for node j in departure time interval τ, $\forall i \in I, j \in J$ and $\tau = 1,\ldots,T$, determine a time-dependent assignment of vehicles to network paths and corresponding arcs so as to achieve certain objectives for the system and/or the users. Hence, the objective is to find the number of vehicles r_{ijk}^{τ} that depart along path $k = 1,\ldots,K_{ij}$ between i and j at time τ, $\forall i \in I, j \in J$ and $\tau = 1,\ldots,T$, as well as the associated numbers of vehicles $x_{ijk}^{\tau ta}$ on each arc $a \in A$ in each time interval t of the duration of interest. The distinction between τ and t is necessary to differentiate between the departure time of a vehicle and the current time. It follows that $\tau \leq t$ in the definition of $x_{ijk}^{\tau ta}$.

2.2.1. *System optimal time-dependent traffic assignment (SOTDTA)*

The SOTDTA problem seeks to minimize the total system travel time over the planning horizon. In the ATIS context, it provides a benchmark against which other assignment or flow patterns can be gauged, thereby yielding an upper bound on the benefits attainable under real-time traffic information.

2.2.2. *User equilibrium time-dependent traffic assignment (UETDTA)*

The UETDTA problem seeks time-dependent user path assignments that satisfy the temporal extension of Wardropian UE condition (Wardrop [17]), namely, no user can improve his/her experienced travel time by unilaterally switching routes (for a given departure time). A time-dependent UE pattern may be considered a useful proxy for a favorable scenario of long-term network performance under real-time descriptive information.

2.3. FORMULATION ISSUES

This section discusses the principal issues involved in formulating time-dependent traffic assignment problems.

2.3.1. Path-based and link-based formulations

In virtually all existing TDTA models, link flows are the decision variables. However, from the perspective of ATIS operations, path-based assignments r_{ijk}^{τ} are ideal because the controller needs to provide paths to tripmakers. The problem with obtaining path flows from link-based formulations using link-path incidence relationships is that uniqueness is not guaranteed. However, the solution of path-based formulations is likely to require partial enumeration of paths for each O-D pair, which can be computationally burdensome. Our formulations are path-based.

2.3.2. Traffic flow modeling and objective function evaluation

Most existing TDTA formulations assume convex, continuous and non-decreasing link performance functions to represent link costs. The use of such functional forms considerably simplifies computation of the objective function, but ignores essential aspects of the time-dependent nature of the problem. The paths followed by future O-D desires may share common links with paths assigned to current trips, thereby influencing the travel times experienced by the latter. The experienced path travel times are the net result of complex nonlinear spatial and temporal interactions among vehicles in the system over a period of time, virtually precluding the ability to evaluate the objective function analytically. Furthermore, analytical evaluation would entail correct representation of the various dynamic traffic flow phenomena (queue formation and discharge, congestion build-up and dissipation), a task that is far from the capability of the state of the art in traffic flow modeling. For these reasons, simulation suggests itself as a plausible candidate for modeling traffic and evaluating the objective function.

2.3.3. Holding of traffic

In a traffic network, it may often be advantageous to favor certain traffic streams or movements over others to minimize system-wide travel delays (e.g., holding back traffic at the minor approach of an intersection in favor of the major approach). Unless otherwise specified, the solution of a SO assignment formulation may entail holding of traffic on one path in favor of traffic on other paths for some significant amount of time at points where the paths overlap or intersect. In other words, vehicles may be artificially delayed on a link for a time that exceeds what may be considered "fair" or "reasonable". Such a solution is probably not acceptable socially nor realistic operationally.

Some of the earlier mathematical programming SO/UE formulations (e.g., Carey [1]) used the concept of a concave exit function $g_a(x^{ta})$, specified as an upper bound on the number of vehicles exiting link a in period t, to model link congestion. Constraints on the vehicles exiting a link were specified as follows to obtain a convex constraint set:

$$m^{ta} \leq g_a(x^{ta}) \qquad \forall t, a, \tag{1}$$

$$m^{ta} = x^{ta} - x^{t+1a} + d^{ta} \qquad \forall t, a. \tag{2}$$

If the number of vehicles that can exit link a in interval t is, in reality, greater than that specified by the bound $g_a(x^{ta})$, vehicles are held back at nodes (or intersections) through artificial control.

When traffic simulation is used to model traffic movements to evaluate network performance for a given assignment, unintended holding is implicitly precluded, and no additional explicit constraints are needed to prevent this problem.

2.3.4. *The first-in, first-out (FIFO) requirement*

The behavior of traffic on a link exhibits the so-called "first-in, first-out" (FIFO) property, creating a particularly vexing difficulty in the solution of mathematical programming formulations of the network assignment problems. The FIFO requirement states that traffic that enters a road at a particular time exits from the facility, on average, before traffic which enters in later periods. While individual vehicles may travel at different speeds and do pass each other, FIFO should not be violated when considering travel time, averaged over a reasonable number of vehicles entering the link, in a given time interval. The problem does not arise in static assignment problems (single or multiple destinations) nor in TDTA with a single destination. However, in TDTA problems with multiple destinations, vehicles on different paths that share one or more common links may be moved across this arc in a manner that violates FIFO; for instance, if the downstream arc along one path is blocked but not for the other path(s).

The problem arises regardless of whether SO or UE models are being considered. For SO problems, total travel costs could be lowered if some commodities (e.g. traffic between given O-D pair) could be temporarily held back on an arc, while allowing some other traffic types to proceed to downstream arcs. This form of holding back would violate FIFO, and is not generally physically possible, especially under congested conditions, as vehicles cannot make such "jumps" over traffic ahead of them.

FIFO is a serious liability from a mathematical programming standpoint. Carey [2] proposed possible additional mathematical constraints to impose FIFO. However, these constraints make the feasible set non-convex, destroying many of the computational and mathematical advantages of the formulation. Smith [16] proposes an UETDTA model which approximately bypasses the FIFO issue by associating a "history" with each vehicle that determines its queueing priority and satisfies a no overtaking condition.

The FIFO issue further highlights the advantages of a traffic simulator in the TDTA context. Simulation moves vehicles based on their current location and speed, and FIFO is implicitly satisfied.

2.3.5. Temporal issues

While TDTA formulations may assume either continuous or discrete time, implementation of solution procedures usually requires discretization of time. In TDTA problems, different time interval sizes, or resolution, will typically be used for different aspects or processes, often reflecting data input availability, accuracy requirements for certain processes, and computational effort considerations. For instance, the size of the time interval for assignment decisions, for which time-dependent O-D information will be available, is typically larger than the time step for the representation (or simulation) of traffic movement.

The formulations developed in the next section update the traffic-related variables such as $x_{ijk}^{\tau t a}$ once every simulation time interval of length Δ (assumed six seconds in all experiments conducted for this study). However, as network traffic conditions do not change appreciably over such small durations, O-D desires r_{ij}^{τ} and associated path assignment decision variables r_{ijk}^{τ} are defined for larger (assignment) time intervals (of the order of several minutes) in the implementation of the solution algorithms.

3. Problem formulations

3.1. SYSTEM OPTIMAL TIME-DEPENDENT TRAFFIC ASSIGNMENT (SOTDTA)

The formulation is mostly a conceptual formulation, and the mathematical expressions are primarily for conceptual clarity. It does not represent a complete formulation from a mathematical standpoint (for example, there are no explicit constraints for ensuring FIFO or precluding holding of vehicles), and is not sufficient for the development of a complete solution algorithm. Section 2 illustrated the advantages of using a traffic simulator to evaluate the objective function and address key formulation issues. Section 4 discusses a simulation-based solution algorithm in which a traffic simulator implicitly satisfies these (FIFO, holding of vehicles, etc.) and other constraints to solve the SOTDTA problem.

The formulation is a nonlinear mixed integer formulation that incorporates 0–1 time-dependent link-path incidence variables, $\delta_{ijk}^{\tau ta}$, to relate the number of vehicles assigned to each path to those on each link. The fundamental difficulty in solving TDTA is that the time-dependent incidence variables are themselves a function of the assignment, giving rise to a complicated fixed-point problem. In the static case, flows assigned to a path exist on all the links along that path simultaneously, leading to a constant link-path incidence matrix. In the dynamic case, vehicles starting along a path will only be present on a certain link at a given time that depends on the travel times along the path.

Essentially, the resulting formulation, which involves nonlinearity in the objective function as well as in the constraints, yields generally undesirable mathematical properties that preclude the guarantee of global optimality. The formulation is as follows:

Given:

$$r_{ij}^{\tau}, \quad \forall i, j \text{ and } \tau = 1, \ldots, T.$$

Objective function:

$$\operatorname{Min} \sum_{\tau} \sum_{i} \sum_{j} \sum_{k} \left(r_{ijk}^{\tau} \cdot T_{ijk}^{\tau} \right) \tag{3a}$$

or

$$\operatorname{Min} \left[T\left(r_{ijk}^{\tau}, \quad \forall i, j, k, \tau \right) \right], \tag{3a'}$$

subject to:

$$r_{ij}^{\tau} = \sum_{k} r_{ijk}^{\tau}, \qquad \forall i, j, \tau, \tag{3b}$$

$$\sum_{b} d^{tb} = \sum_{c} m^{tc} + I_n^t - O_n^t, \qquad \forall t, n, b \in B(n),\ c \in C(n), \tag{3c}$$

$$x^{ta} = x^{t-1a} + d^{t-1a} - m^{t-1a}, \qquad \forall t, a, \tag{3d}$$

$$x^{ta} = \sum_{k} \sum_{\tau} \sum_{i} \sum_{j} \left(r_{ijk}^{\tau} \cdot \delta_{ijk}^{\tau ta} \right), \qquad \forall t, a, \tag{3e}$$

$$T_{ijk}^{\tau} = \sum_{t} \sum_{a} \left[\delta_{ijk}^{\tau ta} \cdot \Delta \right], \qquad \forall i, j, k, \tau, \tag{3f}$$

$$\delta_{ijk}^{\tau ta} = F\left[\left(r_{ijk}^{\tau} \right), \quad \forall i, j, k, \tau \right] \qquad \forall i, j, k, \tau, t, a, \tag{3g}$$

$$d^{ta} = \sum_{k} \sum_{\tau} \sum_{i} \sum_{j} d_{ijk}^{\tau ta}, \qquad \forall t, a, \tag{3h}$$

$$m^{ta} = \sum_k \sum_\tau \sum_i \sum_j m_{ijk}^{\tau ta}, \qquad \forall t, a, \tag{3i}$$

$$I_n^t = \sum_j r_{nj}^t, \qquad \forall t, n \in I, \tag{3j}$$

$$O_n^t = \sum_k \sum_\tau \sum_i \sum_c m_{ink}^{\tau tc}, \qquad \forall t, n \in J,\ c \in C(n), \tag{3k}$$

$$\tau \leq t, \tag{3l}$$

$$\delta_{ijk}^{\tau ta} = 0 \text{ or } 1, \qquad \forall i, j, k, \tau, t, a, \tag{3m}$$

$$\text{all variables } \geq 0. \tag{3n}$$

In contrast to most existing TDTA models, which are link-based, this path-based formulation explicitly seeks vehicle assignments r_{ijk}^{τ} to paths, circumventing the ATIS issues discussed in section 2.3. As noted, the complex time-dependent interactions that characterize these problems preclude the guarantee of convexity and/or differentiability for the objective function. The constraint set is nonconvex, and explicitly incorporates path-based variables using time-dependent link-path incidence variables $\delta_{ijk}^{\tau ta}$.

The objective function (3a) states that the total travel time of the assigned vehicles in the system is the aggregated sum of the product of the number of vehicles assigned to a path and the corresponding experienced path travel time, for every origin, destination and start time. While the interpretation of (3a) is straightforward, it is computationally intractable. Constraints (3f) and (3g) illustrate the nonlinearity of this objective function as the path travel times T_{ijk}^{τ} are themselves a complicated non-explicit function of the assignment decisions r_{ijk}^{τ}. The intractability arises due to (3g) which represent the traffic flow in the network and capture the complex nonlinear time-dependent interactions among vehicles. There exist no known analytical functions $F(r_{ijk}^{\tau})$ that adequately capture these interactions. The problems associated with the computation of (3a) motivate a second form of objective function, (3a′), which states that the total travel time of all vehicles assigned to the various paths during the planning horizon is some function $\boldsymbol{T}(\cdot)$ of the assignment. Like (3a), this objective function also is not well-behaved, and due to the complex dynamic interactions between vehicles implicit in its definition, its properties are not well understood. The solution algorithms described in the next section use a simulation model to evaluate (3a′) for a given path assignment. Reinforcing the advantages of simulation from a traffic flow perspective as discussed in section 2.3, the computation of $\boldsymbol{T}(\cdot)$ using a simulator captures the essential dynamic phenomena and circumvents its analytical intractability.

Constraints (3b) represent the conservation of O-D desires (vehicles) at the origin nodes $i \in I$. Constraints (3c) denote the conservation of vehicles at nodes. They imply that vehicles cannot be stored at nodes, and state that, at any time t on a node n, the number of vehicles entering all links incident from the node should equal the sum of the number of vehicles exiting from all links incident to that node and the net generation. Constraints (3d) represent the conservation of vehicles on links, and state that the number of vehicles on any link a at the beginning of time interval t is the net algebraic sum of the number of vehicles on the link at the beginning of the previous time interval $(t-1)$, vehicles entering the link in the interval $(t-1)$, and vehicles exiting the link in the interval $(t-1)$.

Constraints (3e), (3f) and (3g) incorporate the previously described time-dependent link-path incidence variables $\delta_{ijk}^{\tau ta}$. Constraints (3e) express the number of vehicles on a link, x^{ta}, in terms of the path vehicle assignments r_{ijk}^{τ}. They are nonlinear due to the dependence of the link-path incidence variables on the path assignments as expressed in (3g). Constraints (3f) define the path travel times using the incidence variables. The number of time steps in which $\delta_{ijk}^{\tau ta}$ (for given i, j, k and τ) takes a value 1 implies the number of discrete time steps that the corresponding "packet" of vehicles r_{ijk}^{τ} spend in the system, and multiplying with Δ gives the actual (or experienced) travel time for that packet. Constraints (3f) have several implications for the formulation of dynamic traffic assignment formulations in general. First, the use of time-dependent link-path incidence variables $\delta_{ijk}^{\tau ta}$ provides a capability for computing the actual travel time of vehicles. This circumvents the use of analytical link performance functions and/or simplistic queueing models to compute link travel times to determine the path travel times, a key shortcoming of existing path-based models. Second, they illustrate the difference between formulations based on instantaneous (or current) travel time paths and experienced travel time paths. Third, and of fundamental significance to the characterization of dynamic assignment problems, they recognize the dependence of current assignment decisions on future traffic conditions. The current literature mainly characterizes the assignment patterns as being based on current traffic conditions in the network. Use of correctly defined incidence variables allows the determination of path assignment decisions consistently with experienced travel times, so that the current assignment decisions explicitly depend on future traffic conditions.

Perhaps the most critical and difficult to characterize constraints in dynamic traffic assignment problems are (3g) that represent physical traffic flow. They state that the time-dependent link-path incidence variables are a function $F(\cdot)$ of all the assignment decisions r_{ijk}^{τ} made over the planning horizon. Implicit in these constraints is the satisfaction of FIFO, preclusion of holding of vehicles, and the representation of link interactions and other dynamic traffic phenomena. However, there are no known analytical functions that can adequately represent $F(\cdot)$. The simulation-based solution algorithms discussed in the next section use a traffic simulator to evaluate $F(r_{ijk}^{\tau})$. Given a set of assignment decisions r_{ijk}^{τ} for the entire planning horizon $\tau = 1,\dots,T$, the simulator replicates the dynamic traffic interactions in the network while satisfying the FIFO traffic property and precluding holding of vehicle at nodes. While the properties of $F(\cdot)$ are not well understood, simulation obviates the need for unrealistic link performance and/or link exit functions. The complexity of the traffic flow interactions represented by these constraints precludes the guarantee of well-behaved properties, and in turn for constraints (3e) and (3f). In addition, the dependence of current decisions on future travel conditions leads to a complicated fixed-point problem:

$$\delta_{ijk}^{\tau ta} = F\left[r_{ijk}^{\tau}(\delta),\ \forall i,j,k,\tau\right],\quad \forall i,j,k,\tau,t,a, \tag{4}$$

where δ represents the set of $\delta_{ijk}^{\tau ta}$ values for all i, j, k, τ, t and a. The equations (4) express the dependence of assignment decisions themselves on the link-path incidence variables.

Constraints (3h) and (3i) are, respectively, the definitional constraints for the number of vehicles entering and exiting links in the various time intervals. Constraints (3j) and (3k) are, respectively, the definitional constraints for the number of vehicles entering and exiting the network at node n in the time interval t. Constraints (3l) are the temporal correctness constraints that restrict the start (or departure) time interval τ of assigned vehicles to be at most the current time interval t. Constraints (3m) restrict the time-dependent incidence variables to take values of 0 or 1. Constraints (3n) represent the non-negativity constraints.

3.2. USER EQUILIBRIUM TIME-DEPENDENT TRAFFIC ASSIGNMENT (UETDTA)

The Wardrop condition for the static UE problem can be generalized to the time-dependent case, by requiring an equilibration of the experienced path travel times of users. Many existing UETDTA formulations are based on optimal control theory, and assume the equilibration of instantaneous (or current) travel times. However, an "equilibrium" based on time-dependent instantaneous travel times is conceptually meaningless, since actual travel times experienced by users can be significantly different from the current travel times, especially during peak travel periods. Users may gain travel time savings by unilaterally switching at a later time from the equilibrating instantaneous travel time paths taken when they depart.

While the objective in the SOTDTA problem is the minimization of the total system travel time over the horizon of interest, the objective here is to satisfy time-dependent UE conditions. The UE conditions are obtained through a time-dependent extension of Wardrop's first principle, namely, no user gains travel time savings by unilaterally switching routes, and are stated as:

(1) All paths $k^* \in K_{ij}$, connecting any O-D pair $(i-j)$ that are assigned vehicles in any time interval τ have the same experienced path travel times $T_{ijk^*}^{\tau}$ (equal

to $\theta_{ij}^{*\tau}$, the minimum experienced path travel time from origin i to destination j in the interval τ).

(2) All paths connecting any O-D pair that are not assigned vehicles in any time interval have experienced travel times greater than or equal to $\theta_{ij}^{*\tau}$. These conditions can be expressed as:

$$r_{ijk}^{\tau}\left(T_{ijk}^{\tau} - \theta_{ij}^{*\tau}\right) = 0, \qquad \forall i, j, k, \tau, \tag{5a}$$

$$\left(T_{ijk}^{\tau} - \theta_{ij}^{*\tau}\right) \geq 0, \qquad \forall i, j, k, \tau. \tag{5b}$$

Hence, for any used path, the path flow $r_{ijk}^{\tau} > 0$ and $(T_{ijk}^{\tau} - \theta_{ij}^{*\tau}) = 0$, and for an unused path, $r_{ijk}^{\tau} = 0$, and $(T_{ijk}^{\tau} - \theta_{ij}^{*\tau}) \geq 0$.

The constraint set is identical to the SO formulation constraint set, given by (3b)–(3n).

4. Solution algorithm

This section discusses the algorithms to solve the time-dependent traffic assignment problems discussed in the previous section. Section 4.1 discusses the significance of time-dependent marginal travel times to the SO solution. Section 4.2 discusses the SOTDTA solution algorithm conceptually and mathematically. Section 4.3 discusses the modifications to the SO solution to obtain the UE solution.

4.1. SIGNIFICANCE OF MARGINALS TO THE SO SOLUTION

The time-dependent path marginal travel time $\mathcal{T}_{ijk}^{t}$ is the effect of an additional vehicle on path k (from i to j) at time t on the system travel time. A path-based definition of the global path marginal $\mathcal{T}_{ijk}^{t}$ for a particular $i = m, j = w$, $k = p$ and $t = \kappa$, is given by

$$\mathcal{T}_{mwp}^{\kappa} = T_{mwp}^{\kappa} + \sum_{\tau} \sum_{i} \sum_{j} \sum_{k} r_{ijk}^{\tau} \frac{\partial}{\partial r_{mwp}^{\kappa}}\left(T_{ijk}^{\tau}\right), \tag{6}$$

where T_{mwp}^{κ} refers to the travel time on path p for vehicles assigned at time κ from origin m to destination w. The first term denotes the travel time experienced by an additional vehicle on path p. The second term is the aggregate influence of this vehicle on the system travel time through its effect on all other vehicles that are present in the network at any time after $\tau \geq \kappa$.

The first-order conditions for the mathematical program for the static SO assignment problem state that, at the optimum solution, the marginal travel times on all used paths connecting a given O-D pair are equal, and less than or equal to the marginal travel times on any unused routes. We similarly extend this result to the time-dependent case, as follows:

Consider the formulation (3a)–(3n). Let $\mathcal{L}(\boldsymbol{r}, \boldsymbol{u})$ denote the Lagrangian of this time-dependent SO formulation, where $\boldsymbol{u}$ denotes the vector of Lagrangian multipliers (or dual variables). The necessary conditions for a minimum of this program are given by the first-order conditions for a stationary point of the Lagrangian program $\mathcal{L}(\boldsymbol{r}, \boldsymbol{u})$, which includes the constraints (3b). The constraints (3c) and (3d) are internally conserved, (3e) through (3k) are definitional constraints. Therefore,

$$\mathcal{L}(\boldsymbol{r},\boldsymbol{u}) = z(\boldsymbol{r}) + \sum_{\tau} \sum_{i} \sum_{j} u_{ij}^{\tau}\left(r_{ij}^{\tau} - \sum_{k} r_{ijk}^{\tau} \right) \tag{7}$$

subject to

$$r_{ijk}^{\tau} \geq 0, \quad \forall i, j, k, \tau. \tag{8}$$

The first-order conditions for a stationary point of (7) and (8) are

$$r_{ijk}^{\tau} \frac{\partial \mathcal{L}(\boldsymbol{r},\boldsymbol{u})}{\partial r_{ijk}^{\tau}} = 0 \quad \text{and} \quad \frac{\partial \mathcal{L}(\boldsymbol{r},\boldsymbol{u})}{\partial r_{ijk}^{\tau}} \geq 0, \quad \forall i, j, k, \tau, \tag{9a}$$

as well as

$$\frac{\partial \mathcal{L}(\boldsymbol{r},\boldsymbol{u})}{\partial u_{ij}^{\tau}} = 0, \quad \forall i, j, \tau, \tag{9b}$$

and

$$r_{ijk}^{\tau} \geq 0, \qquad \forall i, j, k, \tau. \tag{9c}$$

The conditions (9b) and (9c) restate the conservation of O-D desires and non-negativity constraints, respectively. The conditions (9a) can be represented by obtaining the partial derivatives of the Lagrangian with respect to the path vehicle assignments. If $i = m$, $j = w$, $k = p$ and $\tau = \kappa$ represent a particular r_{mwp}^{κ}, these derivatives are given by

$$\frac{\partial \mathcal{L}(\boldsymbol{r},\boldsymbol{u})}{\partial r_{mwp}^{\kappa}} = \frac{\partial z(\boldsymbol{r})}{\partial r_{mwp}^{\kappa}} + \frac{\partial}{\partial r_{mwp}^{\kappa}}\left[\sum_{\tau} \sum_{i} \sum_{j} u_{ij}^{\tau}\left(r_{ij}^{\tau} - \sum_{k} r_{ijk}^{\tau} \right)\right], \quad \forall m, w, p, \kappa. \tag{10}$$

The two terms of (10) are evaluated separately. The second term contains r_{ij}^{τ} which is a constant and u_{ij}^{τ} which is not a function of r_{mwp}^{κ}. Also,

$$\frac{\partial r_{ijk}^{\tau}}{\partial r_{mwp}^{\kappa}} = \begin{cases} 1 & \text{if } i = m,\ j = w,\ k = p \text{ and } \tau = \kappa, \\ 0 & \text{otherwise.} \end{cases}$$

Hence,

$$\frac{\partial}{\partial r_{mwp}^{\kappa}}\left[\sum_{\tau} \sum_{i} \sum_{j} u_{ij}^{\tau}\left(r_{ij}^{\tau} - \sum_{k} r_{ijk}^{\tau} \right)\right] = -\, u_{mw}^{\kappa}. \tag{11}$$

The first term on the right hand side of (10) denotes the partial derivative of the SO objective function with respect to r^{κ}_{mwp}. From (3a), it can be represented as

$$\frac{\partial z(\boldsymbol{r})}{\partial r^{\kappa}_{mwp}} = \frac{\partial}{\partial r^{\kappa}_{mwp}}\left[\sum_{\tau}\sum_{i}\sum_{j}\sum_{k}\left(r^{\tau}_{ijk}, T^{\tau}_{ijk}\right)\right]. \tag{12}$$

If the nth order temporal interactions are included, the path assignment r^{τ}_{mwp} at time k may influence the path travel times for any i, j and k at any time $\tau > \kappa$. Expanding the right-hand side of (12),

$$\frac{\partial z(\boldsymbol{r})}{\partial r^{\kappa}_{mwp}} = T^{\kappa}_{mwp} + \sum_{\tau}\sum_{i}\sum_{j}\sum_{k} r^{\tau}_{ijk}\frac{\partial}{\partial r^{\kappa}_{mwp}}\left(T^{\tau}_{ijk}\right), \tag{13}$$

which, by definition is the path marginal travel time $\mathcal{T}^{\kappa}_{mwp}$ defined in (6). Using (11) and (13), the expression (10) can be re-written as

$$\frac{\partial \mathcal{L}(\boldsymbol{r},\boldsymbol{u})}{\partial r^{\kappa}_{mwp}} = \mathcal{T}^{\kappa}_{mwp} - u^{\kappa}_{mw}. \tag{14}$$

The first-order conditions can now be written as

$$r^{\kappa}_{mwp}\left(\mathcal{T}^{\kappa}_{mwp} - u^{\kappa}_{mw}\right) = 0, \qquad \forall m, w, p, \kappa, \tag{15a}$$

$$\left(\mathcal{T}^{\kappa}_{mwp} - u^{\kappa}_{mw}\right) \geq 0, \qquad \forall m, w, p, \kappa, \tag{15b}$$

$$r^{\kappa}_{mw} = \sum_{p} r^{\kappa}_{mwp}, \qquad \forall i, j, \tau, \tag{15c}$$

$$r^{\kappa}_{mwp} \geq 0, \qquad \forall m, w, p, \kappa. \tag{15d}$$

Equations (15a) and (15b) state that at the optimal SO assignment solution, the time-dependent marginal travel times on all used paths connecting a given O-D pair are equal, and less than or equal to the time-dependent marginal travel times on any unused routes. The dual variable u^{κ}_{mw} represents the minimum path marginal travel time value from origin m to destination w in the time interval κ, and is the marginal travel time on all used paths. Hence, the SO assignment solution is obtained by assigning vehicles to the minimum time-dependent marginal travel time paths that are consistent with (15a)–(15d). This principle provides the basis for devising the SOTDTA solution algorithm.

4.2. SOLUTION ALGORITHM FOR THE SOTDTA PROBLEM

The solution algorithm is an extension of well-known methods for the static assignment problem, with key differences in each component and significant

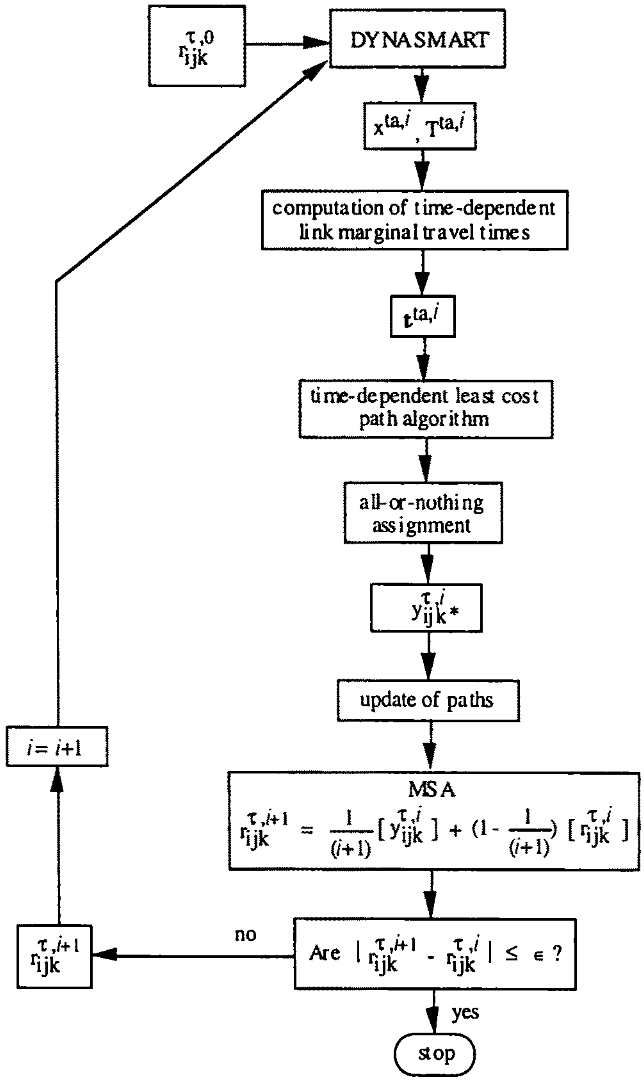

Figure 1. Algorithmic framework for the SOTDTA solution algorithm.

additional implementation challenges. Figure 1 illustrates the solution framework. It consists of a heuristic iterative procedure in which a special-purpose traffic simulation model, DYNASMART (Jayakrishnan et al. [8]), is used to represent the traffic interactions in the network, and evaluate system performance under a given assignment. As discussed earlier, the use of a traffic simulator circumvents the need for link performance functions and link exit functions, ensures FIFO, captures link interactions and precludes unintended holding of traffic. Section 4.2.2 briefly discusses the role of DYNASMART in the solution algorithm.

The vehicular and marginal trip times, estimated from the current simulation, form the basis for the direction-finding mechanism in the search process. As noted earlier, the complex interactions captured by the simulator generally preclude the kind of well-behaved properties required to guarantee a descent direction in each iteration and/or convergence of the search process to a global minimum. However, convergence of the algorithm was achieved in all the experiments reported in this paper, and many other test networks solved to date. Also, path marginals in these experiments are not necessarily global as they are based on local link level marginal travel times. However, in general, the effects of nth order temporal and spatial interactions are significantly lower than that of the direct spatial interaction represented by the local link level marginals. For example, Ghali and Smith [6] found the solution obtained using global marginals to be rather close to those obtained using three levels of approximate marginals (Peeta [13]). The determination of global marginals entails a brute force approach which is computationally inefficient even on existing highly powerful supercomputers. The time-dependent least marginal travel time paths are obtained using the time-dependent least cost path algorithm described in Ziliaskopoulos and Mahmassani [19]. The solution methodology avoids complete path enumeration between O-D pairs as illustrated in step 6 of the algorithm.

4.2.1. Description of the solution algorithm

The steps of the algorithm, shown in figure 1, are described in further detail hereafter.

Step 1. Set the iteration counter $\iota = 0$. Assign the given O-D desires r_{ij}^{τ}, $\forall i, j$ and $\tau = 1, \ldots, T$, to a time-dependent initial set of feasible paths $K(I)$. Hence, at the start of the iterative search process, the initial solution is given by the assignment $r_{ijk}^{\tau,0}$, $\forall i, j, \tau = 1, \ldots, T$, and $k \in K(I)$.

Step 2. Perform simulation of the traffic network under the set of path assignments $r_{ijk}^{\tau,\iota}$ for the entire duration of interest. Obtain time-dependent link level performance measures including link travel times T^{ta}, and the number of vehicles on links x^{ta}, $\forall t, a$.

Step 3. Compute link marginal travel times t'^{a}, $\forall t, a$. In the present implementation, approximate marginal travel times are calculated using the link travel times T^{ta} and the numbers of vehicles on links x^{ta} (Mahmassani et al. [9]).

Step 4. Compute the time-dependent least marginal travel time paths k^*, $\forall i, j$ and τ, such that $\mathcal{T}_{ijk^*}^{\tau} \leq \mathcal{T}_{ijk}^{\tau}$, $\forall k \in K_{ij}$, where K_{ij} represents the set of feasible paths from i to j.

Step 5. Perform all-or-nothing assignment of all O-D desires r_{ij}^{τ} for a given i, j and τ to the corresponding least cost marginal path k^*. This gives the auxiliary number of vehicles on paths, $y_{ijk^*}^{\tau,\iota}$, $\forall i, j$ and τ.

Step 6. Path update is done by checking if $k^* \in K_{ij}^{\tau}$, and including it if it does not, $\forall i, j$ and τ. Thereby, the solution methodology circumvents a priori complete path enumeration by checking if the auxiliary paths generated are pre-existing paths, and including them if they are not. The path assignments for the next iteration $r_{ijk}^{\tau,\iota+1}$ are obtained through a convex combination of the current path assignments $r_{ijk}^{\tau,\iota}$ and the auxiliary path assignments $y_{ijk}^{\tau,\iota}$ using the Method of Successive Averages (MSA), $\forall i, j, k$ and τ:

$$r_{ijk}^{\tau,\iota+1} = \frac{1}{(\iota+1)}\left[y_{ijk}^{\tau,\iota}\right] + \left(1 - \frac{1}{(\iota+1)}\right)\left[r_{ijk}^{\tau,\iota}\right]. \tag{16}$$

Step 7. The convergence criterion is based on the difference in the number of vehicles assigned to various paths over successive iterations. The path assignments for the next iteration $r_{ijk}^{\tau,\iota+1}$ are compared with the current path assignments $r_{ijk}^{\tau,\iota}$, $\forall i, j, k$ and τ:

$$\left| r_{ijk}^{\tau,\iota+1} - r_{ijk}^{\tau,\iota} \right| \le \varepsilon. \tag{17}$$

The number of cases, $N(\varepsilon)$, in which their absolute difference is greater than a value ε is recorded.

Step 8. (i) If $N(\varepsilon) \le \Omega$, where Ω is a pre-set upper bound on the number of violations of (17), convergence is assumed. Terminate the algorithm and output the path assignments $r_{ijk}^{\tau,\iota+1}$ as the solution to the SOTDTA problem.

(ii) If $N(\varepsilon) > \Omega$, the convergence criterion is not satisfied. Update $\iota = \iota + 1$. Go to step 2 with the new current path assignments $r_{ijk}^{\tau,\iota+1}$.

4.2.2. Role of DYNASMART in the solution algorithm

DYNASMART (DYnamic Network Assignment-Simulation Model for Advanced Road Telematics) is a fixed time step mesoscopic simulation model for ITS applications developed at the University of Texas at Austin. It is designed to model traffic patterns and evaluate overall network performance under real-time information systems, for a given network configuration (including traffic control system) and given time-dependent OD demand pattern. The modeling approach integrates a traffic flow simulator, a network path processing component, user behavior rules and information supply strategies. A comprehensive discussion on DYNASMART is presented in Jayakrishnan et al. [8], and the latest capabilities of DYNASMART are discussed in Mahmassani et al. [9]. In this study, DYNASMART is used only as a simulator given user's selected paths in the single class SOTDTA and UETDTA solution procedures.

Figure 2 illustrates the overall structure of DYNASMART from the perspective of its role in the single user class SO and UE algorithms. The double-lined

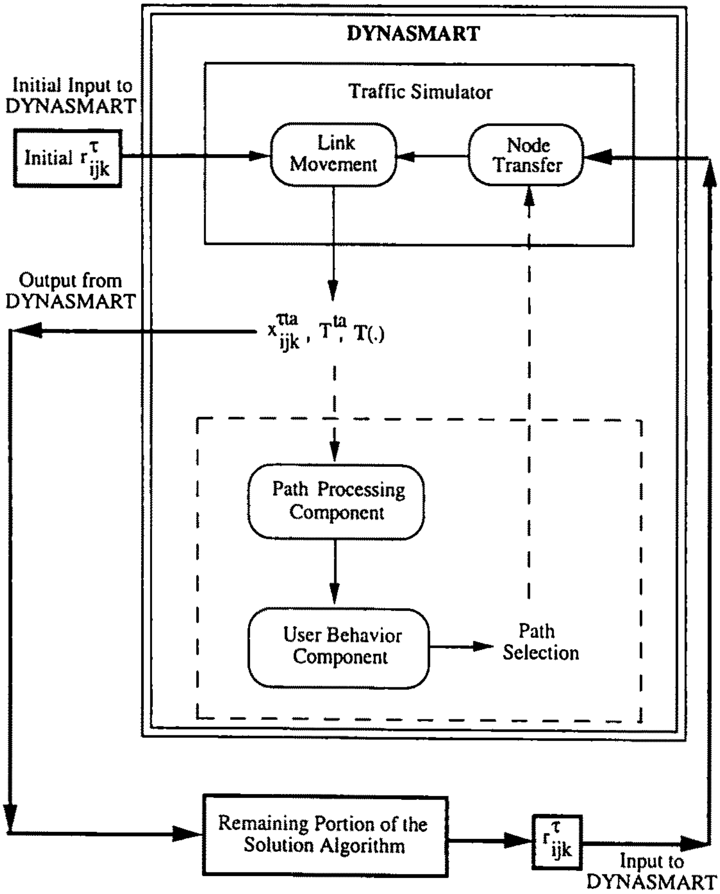

Figure 2. DYNASMART structure as part of the solution algorithm.

box in the figure encloses the various components of DYNASMART. It depicts plain-line and dashed-line arrows as part of the control logic in DYNASMART. The plain-line arrows represent that part of the logic used in the single class SOTDTA and UETDTA algorithms, in which only the traffic flow simulator is invoked. The simulator evaluates system performance and generates time-dependent aggregate and link level performance measures under a prescribed assignment strategy, which are then utilized in a direction-finding logic embedded in the iterative solution algorithm described earlier, represented by the bold-line arrows. The dashed-line arrows and box indicate the part of DYNASMART not used in the TDTA algorithms discussed here; those include the user behavior and path processing components. These two components provide the ability to analyze user response, user compliance and test alternative information supply strategies, capabilities that are lacking in existing traffic simulation models. As indicated in figure 2, vehicular paths are assigned exogenously to DYNASMART in the single user class SOTDTA and UETDTA problems, as determined by the steps of the SO or UE solution algorithms. Thus, DYNASMART is used primarily as a simulator to replicate the dynamics of traffic for a given assignment of vehicles to paths.

By simulating the set of path assignments $r_{ijk}^{\tau,t}$ for the entire duration of interest (step 2 of the algorithm), DYNASMART determines the objective function (3a′) for that iteration, represented by the total system travel time. The link movement component of DYNASMART moves vehicles consistently with constraints (3d) by solving the finite difference form of the continuity equation and using a traffic-theoretic speed-concentration relationship to model the flow (Peeta [13]). The node transfer component of DYNASMART addresses constraints (3b) and (3c) that denote the conservation of vehicles at nodes.

4.2.3. Computation of time-dependent link marginal travel times

This section discusses the methodology for computation of approximate marginals for the SOTDTA problem. Step 3 of the solution algorithm consists of determining the time-dependent link marginals t^{ta}, $\forall t, a$, and is given by

$$t^{ta} = T^{ta}(x) + x^{ta} \frac{\partial T^{ta}(x)}{\partial x^{ta}}, \tag{18}$$

where $T^{ta}(x)$ represents the travel time experienced by an additional vehicle entering link a at time t, and x is the vector of time-dependent number of vehicles on all links (x^{ta}, $\forall t, a$). Hence, spatial interactions and nth order temporal interactions are ignored in this computation. As discussed earlier, these effects may not be significant compared to the direct effect on link a at time t, $\partial T^{ta}(x)/\partial x^{ta}$. Under such conditions, the solutions obtained using the global marginals and the local marginals (18) will be relatively close. However, if the interactions are significant, the solution obtained using (18) may deviate from that obtained using the global marginals.

While $T^{ta}(x)$ and x^{ta} in (18) are obtained directly from step 2, a methodology is needed to compute $\partial T^{ta}(x)/\partial x^{ta}$. Figure 3 illustrates the approach used for the computation of the derivative $\partial T^{ta}(x)/\partial x^{ta}$. The approach used here assumes that the time-dependence of the derivative is due to "time-varying" link performance functions. This means the performance curve in figure 3 for link a at time t depends on the traffic flow conditions on the link at that time. This time-dependence is very significant; the travel time on a link can be significantly different for the same number of vehicles at two different times depending on the fraction of vehicles queued. The link performance curve for a link changes gradually over time. If the time interval between successive evaluations of marginals is small, it appears reasonable to assume that three consecutive points in time are on the same link performance curve, as illustrated in figure 3. A quadratic fit using the three points results in the time-varying link performance curve at time t and the slope of

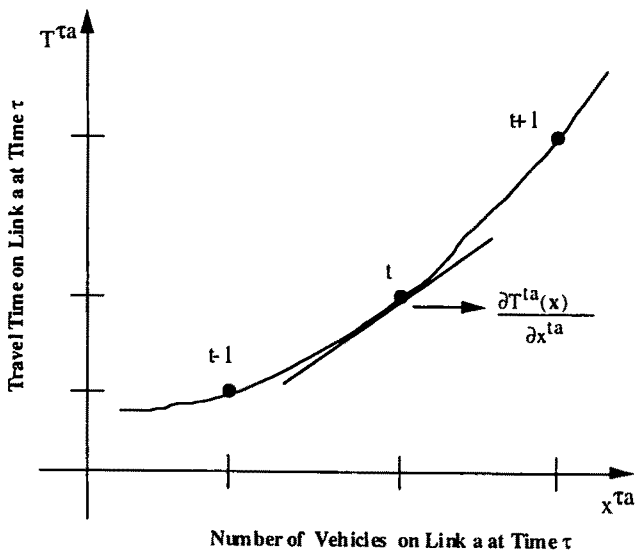

Figure 3. Computation of derivative for link marginal.

this curve at time t gives $\partial T^{ta}(x)/\partial x^{ta}$ as indicated in the figure. However, the consideration of small time intervals (in the order of a few seconds) may cause some instability in the curves because the values of travel times and the number of vehicles in successive intervals may show "jumps" at times. Hence, the length of the time interval between successive data points involves trade-offs between accuracy and robustness of the curves. We use averaging techniques to achieve stability in the curves (Peeta [13]). The simulation time interval used in DYNASMART of six seconds is too small to update the paths for a given O-D pair as no appreciable change takes place in the system in such a short duration. In the implementation of the solution algorithm, paths are updated every assignment interval (three minutes for the experiments). If Δ denotes the length of a simulation interval, then $M\Delta$ denotes the length of an assignment interval that includes M simulation intervals. The marginal values are computed for assignment intervals only and not for simulation intervals, thereby reducing the computational burden of the path-processing step.

4.2.4 Computation of time-dependent least marginal travel time paths

The least marginal travel time paths are computed using a special-purpose time-dependent least cost path algorithm (Ziliaskopoulos and Mahmassani [19]). The algorithm, which is based on the general Bellman's principle of optimality, is very efficient and is customized for use with DYNASMART in the solution methodology. It requires both average and link marginal travel times as input data.

The algorithm uses link marginal travel times as link penalties and link average travel times as link movement costs. Unlike in the static case, this seemingly unimportant detail reflects a subtle but conceptually important point for correct calculation of time-dependent marginal shortest paths. The least marginal path algorithm is used to determine the set of paths k^* with the least marginal travel times, $\mathcal{T}^{\tau}_{ijk^*}$ in step 4 of the solution algorithm.

4.3. THE USER EQUILIBRIUM SOLUTION

As in the static case, SO and UE solutions involve similar algorithmic steps, differing primarily in the specification of path travel costs that form the basis of the corresponding assignments. The UE solution can be obtained by directing O-D flows onto least travel time paths instead of least marginal cost paths in the SOTDTA solution procedure. Figure 4 illustrates the UETDTA solution framework. It differs

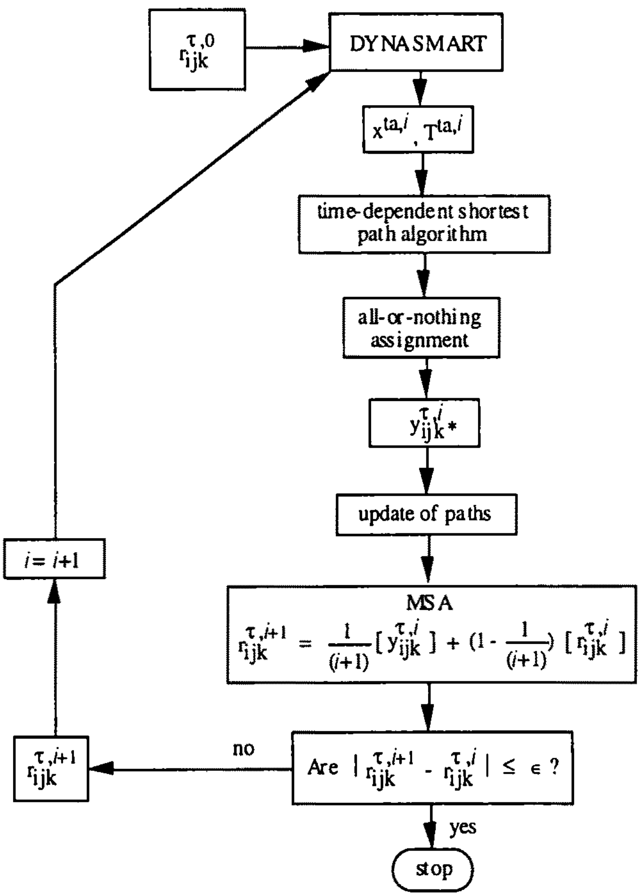

Figure 4. Algorthmic framework for the UETDTA solution algorithm.

from the SO framework in the following three elements. Step 3 of the SOTDTA algorithm, which deals with the computation of the marginal travel times is not necessary, and is excluded. Step 4 is modified as follows to obtain step 3 of the UE algorithm:

> Compute the time-dependent shortest average travel time paths k^*, $\forall i, j$ and τ, such that $T^{\tau}_{ijk^*} \le T^{\tau}_{ijk}$, $\forall k \in K_{ij}$, where K_{ij} represents the set of feasible paths from i to j.

Similarly, step 5 of the SO algorithm is modified as follows to obtain step 4 of the UE algorithm:

> Perform all-or-nothing assignment of O-D desires r^{τ}_{ij} for a given i, j and τ to the corresponding shortest average travel time path k^*. This gives the auxiliary number of vehicles on paths, $y^{\tau,l}_{ijk^*}$, $\forall i, j$ and τ.

The shortest average travel time paths are computed using a time-dependent shortest path algorithm (Ziliaskopoulos and Mahmassani [18, 19]), given the time-dependent link average travel times $T^{ta}(x)$.

5. Experimental analysis

Experiments have been conducted to derive insights on time-dependent system performance under alternative assignment strategies, namely SO and UE, and under different network loading intensities. The results suggest directions for the development of information supply strategies.

It is known from static network equilibrium theory that SO and UE lead to identical solutions only for situations where the shortest paths taken by users are simultaneously the best paths from a system viewpoint. Such situations are observed when networks are relatively uncongested so that link operating speeds are unaffected by the flows on the links (limited vehicle interactions). At the other extreme, under very highly congested conditions, system performance is not likely to be markedly different under the two assignment schemes because the opportunities for SO to sufficiently ameliorate the traffic situation would probably be limited. Between those two extremes, the extent of the difference between SO and UE solutions, particularly in terms of overall system cost, is not known. This question, for time-dependent traffic systems, is of fundamental importance to ATIS operations, with regard to the relative benefits of normative versus descriptive information supply strategies.

5.1. EXPERIMENTAL DESIGN AND SET-UP

This section first details the network configuration and traffic characteristics, followed by a description of the experimental set-up.

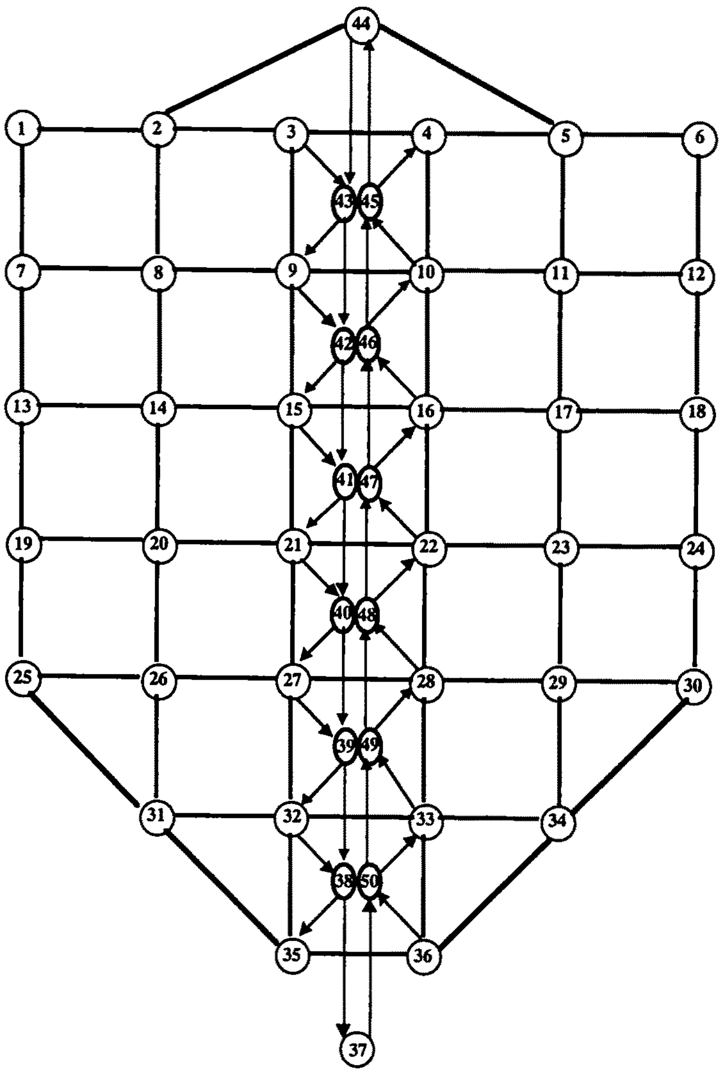

Figure 5. Network structure.

5.1.1. Network configuration and traffic characteristics

The test network consists of a freeway with a grid street network on both sides as shown in figure 5. The network consists of 50 nodes and 168 links, with all nodes except the freeway nodes serving as both origins and destinations. Consequently, the network has 38 origin nodes and 38 destination nodes (nodes 1 through 37 and node 44). All arcs are two-directional (and are actually modeled as two directed arcs) and have two lanes in each direction, except the entrance and exit ramps that connect the street network to the freeway; these are directed arcs with one lane as shown in the figure. Each link is 0.83 km (0.5 miles) long. The freeway links have a mean free speed of 91.67 kmph (55 mph) and all other links have a 50 kmph (30 mph) mean free speed. The maximum bumper-to-bumper and jam densities are assumed to be 156 vehicles/kilometer (260 vehicles/mile) and 96 vehicles/km (160 veh./mile) respectively for all links of the network. With regard to the intersection signal control, 26 nodes have pre-timed signalization, 8 have actuated signal control and the rest have no signal control. The pre-timed signals have a 60 second cycle length with two phases, each with 26 seconds of green time and 4 seconds of amber time. The actuated signals have 10 seconds of minimum green time and 26 seconds of maximum green time.

5.1.2. Experimental factors

The experimental factors considered in this study can be separated into three categories:

- *Demand levels*: The two assignment rules are compared under different network congestion levels, achieved through different network loading levels. The network loading factor (*LF*) is defined as the ratio of the total number of vehicles generated in the network during the assignment period compared to a given reference number (19220 vehicles over a 35-minute period, which represents a loading factor of 1.0). Five different loading factors are considered in the experiments, namely, 1.0, 1.4, 1.8, 2.0, and 2.2. Table 1 shows the number of vehicles generated for each loading factor for the duration of interest (35 minutes in all cases). In addition, it

Table 1

Loading factors and the corresponding number of generated vehicles for the numerical experiments.

Loading factor	Number of generated vehicles	Tagged vehicles
1.0	19220	17434
1.4	26936	24456
1.8	34656	31415
2.0	38506	34905
2.2	42371	38382

shows the corresponding number of "tagged" vehicles (vehicles generated for the 30 minute duration after the 5-minute start-up time) for which relevant performance statistics are accumulated. These represent various levels of network congestion ranging from low (for $LF = 1.0$) to moderately high (for an LF of 2.2). Two temporal loading profiles are discussed next.

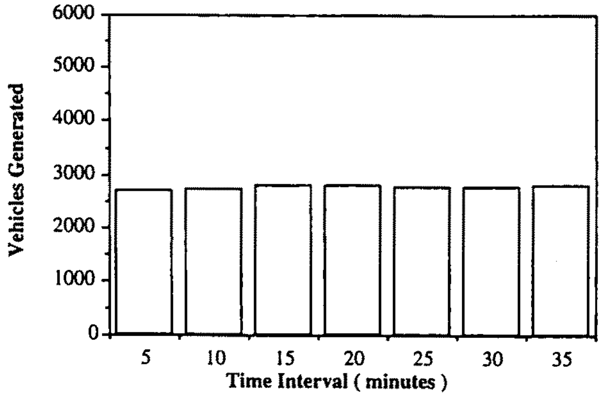

Figure 6. Loading profile I, the base case ($LF = 1.0$); specified as the number of vehicles generated (aggregated over 5 minute intervals).

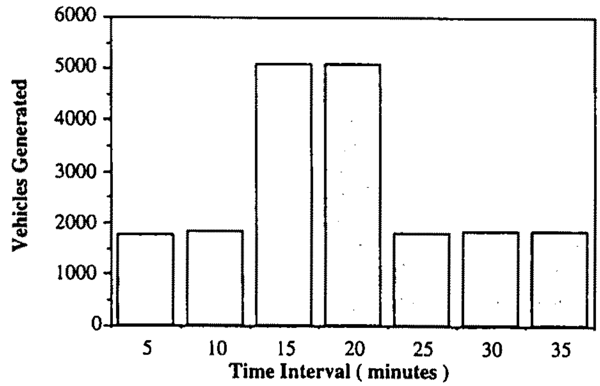

Figure 7. Loading profile II, the base case ($LF = 1.0$); specified as the number of vehicles generated (aggregated over 5 minute intervals).

• *Loading patterns*: Two temporal loading patterns are considered in the experiments, and are referred to as loading profile I and loading profile II. Loading profile I generates vehicles uniformly over the assignment duration, as illustrated in figure 6 for a 1.0 loading factor. Loading profile II, shown in figure 7, impacts the network with relatively large number of vehicles over a ten minute period which is preceded and succeeded by low levels of uniform loading for the rest of the assignment duration. While the loading patterns are time-dependent, they are shown as 5-minute aggregates in figures 6 and 7 for visual convenience. The two profiles are designed so as to represent extremes in loading conditions for the way in which they influence the system performance. A typical peak hour loading pattern would most likely lie between these two benchmarks. In both cases, vehicles are generated over a 35-minute period which includes a 5-minute start-up generation time in order for the network to be reasonably occupied, followed by a 30-minute generation of vehicles for which statistics are accumulated. No vehicles are generated afterwards. With regard to the spatial distribution of the O-D trip desires under the two loading patterns, vehicles are generated about evenly in space, both in terms of their origins and destinations, except for nodes 37 and 44 which incur only about 25% the volume (both as origins or destinations) compared to a typical node (nodes 1–36).

- *Information type*: Two information types are considered:
  (i) Vehicles are supplied time-dependent SO routes by a central controller and follow the prescribed routes to their destinations.
  (ii) Vehicles are supplied time-dependent UE routes by a central controller and follow the prescribed routes to their destinations. In a given simulation run, all vehicles are provided only one type of information. Vehicles under both the SO and UE strategies are provided the current best paths as the initial paths for their solution algorithms.

5.2. ANALYSIS OF RESULTS

Results from the various experiments form the basis for comparison of system performance, particularly user costs, under time-dependent UE and SO assignment schemes. They manifest a clear qualitative and quantitative differentiation between the SO and UE solutions, and illustrate the effectiveness of coordinated strategies vis-a-vis descriptive strategies. Figure 8 shows a typical convergence pattern of the

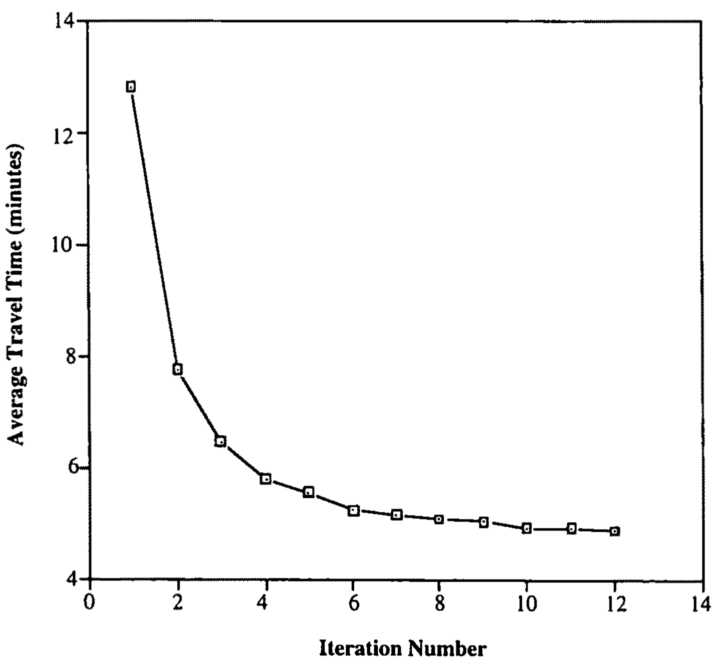

Figure 8. Typical convergence pattern for the SOTDTA solution algorithm.

solution algorithms for SOTDTA and UETDTA problems. The figure shows the progress of SO solution algorithm for a loading factor 1.4, for which the initial assignment strategy (current best paths) results in an average trip time per vehicle of about 13 minutes. At convergence, the average trip time per vehicle is slightly less than 5 minutes.

The summary statistics for the system performance under the time-dependent SO and UE assignment rules for loading profiles I and II are reported in tables 2 and 3, respectively. As expected, at low levels of network loading, when the network is relatively uncongested, the average travel times of vehicles in the network are relatively close across the different loading factors. As the load is increased, the effects of congestion become more prominent and the average travel times in the network increase at an increasing rate with the loading factor. At very high loading levels, the marginal effect of additional demand on system performance is very high. The results also indicate that there is only limited variation in the average distance traveled by vehicles under the various network loading levels, implying that greater congestion and not longer travel routes is the primary cause of the higher system trip times (the objective function seeks to minimize total system travel time only). Nevertheless, the average travel distance does increase with the loading level, reflecting an increasing percentage (though small in magnitude) of drivers assigned to longer travel routes. The average travel distances under UE for various network loading levels are smaller than the corresponding distances for SO, indicating a smaller percentage of long travel routes under UE. This may be explained by some users being assigned to longer routes under SO in order to reduce congestion elsewhere so as to reduce systemwide travel times.

Figures 9 and 10 compare the average trip times under SO and UE assignment rules for loading profiles I and II, respectively. As expected, the system performs better under the uniform loading profile I than under the more peaked profile II. As discussed above, both curves illustrate the increasing marginal effects of additional demand on system trip times.

Figure 11 highlights the difference in the quality of the solutions provided by the two assignment rules by depicting the percentage improvement in average travel time of SO over UE (as a fraction of the UE travel time) for the various average network concentrations corresponding to the various levels of network loading. At low loading levels ($LF = 1.0$), SO and UE provide essentially identical solutions. At such low concentration levels, average link speeds remain relatively unchanged due to limited interactions among vehicles, and the marginal travel time on the link is very close to the average travel time, leading to almost identical solutions under the two assignment schemes. However, as network congestion increases, SO shows substantial savings over UE. For example, SO gains 13.5% improvement over UE for moderately high network congestion ($LF = 2.2$) when the average network concentration for the entire duration of interest is about 32 vehicles per lane-mile.

Table 2

Statistics of SO and UE assignments for loading profile I.

Loading factor	Av. trip time (minutes)	Total trip time (hours)	Average trip distance (kilometers)	Total trip distance (kilometers)	Average speed (kmph)
			SO results		
1.00	3.72	1023.83	2.88	47521.25	46.42
1.40	3.94	1522.34	2.93	67984.17	44.95
1.80	4.58	2268.42	3.02	89315.00	39.57
2.00	5.55	3057.80	3.08	102155.00	33.40
2.20	7.06	4273.56	3.15	114710.42	26.85
			UE results		
1.00	3.72	1023.17	2.80	46252.08	45.20
1.40	3.96	1527.56	2.82	65410.83	42.82
1.80	4.85	2402.89	2.88	85620.00	35.63
2.00	6.04	3331.67	3.00	98475.00	29.75
2.20	8.14	4934.18	3.08	112240.42	22.73

Table 3

Statistics of SO and UE assignments for loading profile II.

Loading factor	Av. trip time (minutes)	Total trip time (hours)	Average trip distance (kilometers)	Total trip distance (kilometers)	Average speed (kmph)
			SO results		
1.00	3.93	1140.85	2.92	50773.33	44.50
1.40	4.89	1993.28	3.02	73621.67	36.93
1.80	6.84	3580.65	3.10	97358.75	27.18
2.00	8.64	5026.45	3.18	110994.58	22.08
			UE results		
1.00	3.97	1153.46	2.82	49129.17	42.60
1.40	5.10	2077.40	2.90	71120.83	34.23
1.80	7.31	3826.16	2.98	93703.75	24.48
2.00	9.28	5394.81	3.05	106547.92	19.75

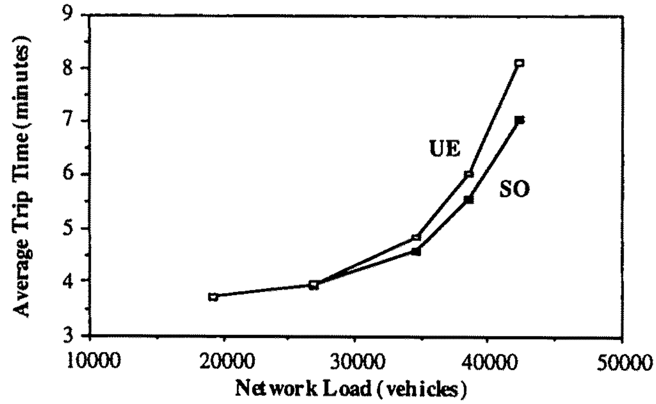

Figure 9. Comparison of average trip times (minutes) of SO and UE assignments for network loads under loading profile I.

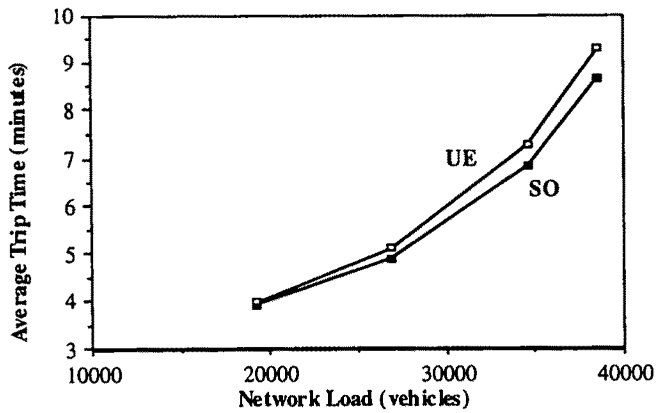

Figure 10. Comparison of average trip times (minutes) of SO and UE assignments for network loads under loading profile II.

Further differentiation between the SO and UE time-dependent assignments is illustrated in figure 12, which compares the SO and UE solutions for a more detailed set of experiments (Mahmassani and Peeta [11]) conducted with essentially the same network structure, and a loading profile that emulates peak-period network loading. A loading factor 1.0 in those experiments corresponds to 19403 vehicles generated over a 35 minute period, and congestion levels vary from the uncongested ($LF = 0.6$) to the very highly congested ($LF = 2.4$). The results essentially mirror those of the previous case, with even greater improvement levels attained by SO over UE. The main new element illustrated here is that, as network loading levels are increased further, the system reaches very high congestion levels that near gridlock, and overall network throughput drops, making it increasingly difficult to discharge all vehicles from the system in a reasonable amount of time. Under these conditions, the ability to improve overall conditions by re-routing certain vehicles

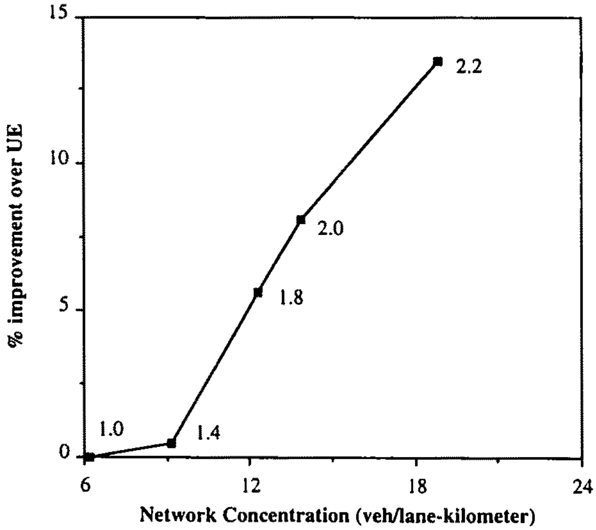

Figure 11. Percentage total trip time savings of SO over UE (as a fraction of total UE trip time) versus average network concentration for loading profile I. The number by each plotted point is the corresponding loading factor.

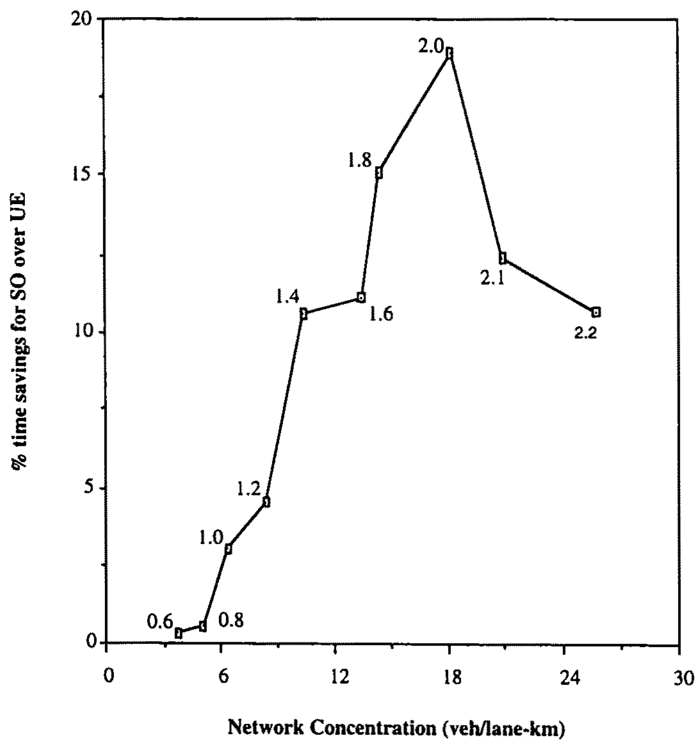

Figure 12. Percentage total trip time savings of SO over UE obtained as a fraction of total UE trip time for different loading factors (Mahmassani and Peeta [11]).

to paths with lower marginal costs diminishes, as all links become highly congested. Thus, the advantage of an SO assignment relative to UE begins decreasing, as reflected by reduced improvements of 12.4% and 10.7% for loading factors 2.1 and 2.2, respectively. The gains begin dropping rapidly beyond this point, with higher loading levels eventually yielding negligible differences in the quality of the solution provided by the two schemes.

Mahmassani and Peeta [11] also illustrate the time-dependent savings of SO over UE. Those experiments suggest that the savings of SO over UE (for a given loading factor) are not accumulated uniformly over time, but are dependent on the time-dependent network congestion profile generated, with most savings being obtained when the network is well-congested. Since the evolution of congestion is dependent on the loading profile, the shape of the network loading curve significantly affects the nature of these savings.

6. Concluding comments

The state of the art of dynamic traffic assignment covers a spectrum of problems with varying objectives and underlying assumptions. The single user class TDTA problems represent a basic class of dynamic traffic assignment problems. While realistic scenarios for ATIS operations may lack complete a priori information on time-dependent O-D desires for the planning horizon, and consist of several user classes in terms of information availability, information supply strategy and user behavior, the time-dependent traffic assignment problems serve two critical purposes. First, they are building blocks for the more realistic multiple user classes time-dependent traffic assignment (MUCTDTA) problems (Peeta [13], Mahmassani et al. [9]), and the considerably more complicated real-time or quasi-real time traffic assignment problems (Peeta [13], Peeta and Mahmassani [14]). Second, they provide benchmarks for the evaluation of alternative information supply and control strategies, thereby addressing the off-line needs of ATIS.

The solution algorithms for the SOTDTA and UETDTA problems circumvent the limitations that preclude existing analytical models from solving realistic formulations of the problem. They can be applied for general networks, and provide the foundation for developing algorithms for the MUCTDTA and the real-time assignment problems. The experiments performed using the SO and UE solution algorithms have provided insights of critical importance to the design of ATIS information supply strategies. A comparative analysis of the SO and UE strategies has illustrated the effectiveness of coordinated cooperative information supply strategies over descriptive strategies to enhance the overall system performance. If the UE solution can be considered a proxy for the system conditions when drivers have access to descriptive information through ATIS, the experiments suggest significant benefits through system optimal, coordinated route guidance, especially in heavily congested networks.

Acknowledgements

This work is based on funding provided by the Federal Highway Administration under the contract DTFH61-90-R-00074 titled "Traffic Modeling to Support Advanced Driver Information Systems (ADIS)". The authors are grateful to Ta-Yin Hu and Athanasios Ziliaskopoulos for their contributions to the implementation of the solution algorithm. Of course, the authors are solely responsible for the contents of this paper.

References

[1] M. Carey, Optimal time varying flows on congested networks, Operations Research 35(1987) 58–69.
[2] M. Carey, Nonconvexity of the dynamic assignment problem, Transportation Research 26B(1992) 127–133.
[3] C.S. Fisk and D.E. Boyce, Alternative variational inequality formulations of the network equilibrium-travel choice problem, Transportation Science 17(1983)454–463.
[4] 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(1993) 179–191.
[5] M.O. Ghali and M.J. Smith, Optimal dynamic traffic assignment of a congested city network, *Proc. of the 2nd Int. Seminar on Urban Traffic Networks*, Capri, Italy (July 1992).
[6] M.O. Ghali and M.J. Smith, Traffic assignment, traffic control and road pricing, presented at the *12th Int. Symp. on Transportation and Traffic Theory*, Berkeley, CA (July 1993).
[7] B.N. Janson, Dynamic traffic assignment for urban road networks, Transportation Research 25B(1991)143–161.
[8] R. Jayakrishnan, H.S. Mahmassani and T. Hu, An evaluation tool for advanced traffic information and management systems in urban networks, Transportation Research 2C(1994)129–147.
[9] H.S. Mahmassani, T. Hu, S. Peeta and A. Ziliaskopoulos, Development and testing of dynamic traffic assignment and simulation procedures for ATIS/ATMS applications, Technical Report DTFH61-90-C-00074-FG, Center for Transportation Research, The University of Texas at Austin (1994).
[10] H.S. Mahmassani and S. Peeta, System optimal dynamic assignment for electronic route guidance in a congested traffic network, *Proc. of the 2nd Int. Seminar on Urban Traffic Networks*, Capri, Italy (1992).
[11] H.S. Mahmassani and S. Peeta, Network performance under system optimal and user equilibrium dynamic assignments: Implications for ATIS, Transportation Research Record 1408 (1993) 83-93.
[12] H.S. Mahmassani, S. Peeta, T. Hu and A. Ziliaskopoulos, Dynamic traffic assignment with multiple user classes for real-time ATIS/ATMS applications, *Proc. of the ATMS Conf. on Management of Large Urban Traffic Systems*, St. Petersburg, Florida (1993).
[13] S. Peeta, System optimal dynamic traffic assignment in congested networks with advanced information systems, Ph.D. Dissertation, The University of Texas at Austin (1994).
[14] S. Peeta and H.S. Mahmassani, Multiple user classes real-time traffic assignment for on-line operations: A rolling horizon solution framework, Transportation Research 3C(1995)83–98.
[15] B. Ran, D.E. Boyce and L.J. LeBlanc, A new class of instantaneous dynamic user-optimal traffic assignment models, Operations Research 41(1993).

[16] M.J. Smith, A new dynamic traffic model and the existence and calculation of dynamic user equilibria on congested capacity-constrained road networks, presented at the *71st Annual Meeting of TRB*, Washington DC (1991).

[17] J.G. Wardrop, Some theoretical aspects of road traffic research, Proceedings Institution of Civil Engineers II(1) (1952) 325–378.

[18] A. Ziliaskopoulos and H.S. Mahmassani, Design and implementation of a shortest path algorithm with time-dependent arc costs, *Proc. of the 5th Advanced Technology Conference*, Washington DC (1992) pp. 1072–1093.

[19] A. Ziliaskopoulos and H.S. Mahmassani, A time-dependent shortest path algorithm for real-time intelligent vehicle/highway systems applications", Transportation Research Record 1408 (1993) 94–100.