Toll Pricing and Heterogeneous Users

Approximation Algorithms for Finding Bicriterion Time-Dependent Efficient Paths in Large-Scale Traffic Networks

Hani S. Mahmassani, Xuesong Zhou, and Chung-Cheng Lu

This paper presents both exact and approximation algorithms for finding extreme efficient time-dependent shortest paths for use with dynamic traffic assignment applications to networks with variable toll pricing and heterogeneous users (with different value of time preferences). A parametric least-generalized cost path algorithm is presented to determine a complete set of extreme efficient time-dependent paths that simultaneously consider travel time and cost criteria. However, exact procedures may not be practical for large networks. For this reason, approximation schemes are devised and tested. Based on the concept of ϵ-efficiency in multiobjective shortest path problems, a binary search framework is developed to find a set of extreme efficient paths that minimize expected approximation error, with the use of the underlying value of time distribution. Both exact and approximation schemes (along with variants) are tested on three actual traffic networks. The experimental results indicate that the computation time and the size of the solution set are jointly determined by several key parameters such as the number of time intervals and the number of nodes in the network. The results also suggest that the proposed approximation scheme is computationally efficient for large-scale bi-objective time-dependent shortest path applications while maintaining satisfactory solution quality.

Road pricing is increasingly considered as an effective demand management strategy to reduce traffic congestion and improve system performance during peak periods in many metropolitan areas. In capacity-limited transportation networks, the planning and operation of various road pricing strategies, such as road tolls, cordon (area) tolls, and high-occupancy toll lanes, call for path choice models that take into account two essential decision attributes: travel time and out-of-pocket cost. In a utility maximization framework, each trip maker can be assumed to select a path that minimizes a generalized cost function where travel time is weighted by the trip maker's value of time (VOT). The VOT relative to each trip represents how much money the trip maker is willing to trade off for unit time saving.

Various empirical studies (*1, 2*) have suggested that the VOT varies significantly across individuals because of different socioeconomic characteristics, trip purposes, attitudes, and inherent preferences. To capture the heterogeneous preferences underlying trip makers' route choice behavior, several bi-objective traffic assignment models (*3, 4*) (S. C. Dafermos, unpublished manuscript, 1981) have been proposed to generalize the classic single-objective traffic assignment by relaxing the VOT from a constant to a continuously distributed random variable. Solution procedures for the bi-objective traffic assignment problem typically involve a direction-finding step, which needs to determine bi-objective minimum paths that simultaneously seek to minimize conflicting path attributes. In addition, many applications in transportation planning and operations require efficient solution algorithms for optimum path problems with multiple objectives, such as travel time, reliability, and population exposure in the problem of shipping hazardous materials (*5*). All these applications underscore the need for computationally efficient and tractable solution procedures for multiobjective shortest path problems (MOSPs) in transportation networks.

Because no unique optimal path usually exists in terms of all the objectives in a general network, the MOSP aims to find a set of nondominated (i.e., Pareto optimal or efficient) paths. The weighting method combines different attributes into a single utility function and systematically varies the weights (e.g., VOT) to generate nondominated solution sets. This method has been widely applied to solve the MOSP, because it can utilize efficient solution algorithms for the resulting single-objective shortest path problem. In the recursive weighting algorithm presented by Dial (*6*), a weighting parameter is iteratively generated based on the slope of the line connecting the two extremes (i.e., the trade-off between two solutions), and the search process is repeated recursively until all (or a given number of) the efficient paths are found. Another common weighting method is to calculate shortest paths with randomly generated VOT from a given distribution function (*7*). Henig (*8*) introduced the concept of extreme (or supported) efficient paths, which correspond to extreme points in the boundary of the convex hull containing all the efficient points in the criterion space. He further pointed out that the weighting method can enumerate only the extreme efficient paths because nonextreme efficient solutions are dominated by a convex combination of extreme efficient solutions.

The parametric shortest path method was also applied by several researchers, such as Henig (*8*), Mote et al. (*9*), and Dial (*10*), to efficiently identify extreme efficient shortest paths. By computing the sensitivity range of the weighting parameter, the parametric weighting method is able to move directly from one extreme efficient solution to the next one without redundant calculations. As a result, the feasible range of VOT can be partitioned into a number of intervals, each corresponding to a shortest path tree. Another method to identify the entire nondominated solution set is the multilabeling approach pro-

Maryland Transportation Initiative, Department of Civil and Environmental Engineering, University of Maryland, 1173 Glenn L. Martin Hall, College Park, MD 20742.

Transportation Research Record: Journal of the Transportation Research Board, No. 1923, Transportation Research Board of the National Academies, Washington, D.C., 2005, pp. 28–36.

posed by Hansen (*11*), in which the scalar label associated with each node in the single-objective shortest path algorithm is extended to a label vector to keep track of nondominated solutions, and the label minimization step is replaced by the dominance checking step. This approach was extended by Miller-Hooks and Mahmassani (*12*) to the routing of hazardous materials in stochastic time-varying networks.

Because the MOSP is known to be NP-hard (*13*), the computational time and storage requirements for finding a complete set of efficient solutions increase exponentially with the size of the problem. Alternatively, many approximation algorithms have been developed to find a subset of efficient paths within limited computational resources. Climaco and Martin (*14*) applied the *K*-shortest path algorithm to find efficient solutions. Warburton (*15*) proposed an ϵ-approximate algorithm to quantify the degree of accuracy in approximating trade-off curves and surfaces in a multiple criterion space. With rounding and scaling techniques, he also developed a fully polynomial approximate algorithm subject to a desired degree of accuracy. In Hassin's study (*16*), the ϵ-approximate concept is combined into a binary search scheme, which iteratively adds the new weighting breakpoints, reducing the approximation error at intervals. Nielsen (*17*) applied a similar approximation scheme for solving the bicriterion shortest hyperpath problem in random time-dependent networks under a priori and adaptive route choice strategies.

Recent advances in the development of dynamic traffic assignment (DTA) models (*18*) have facilitated the design and evaluation of various road pricing scenarios that vary with location, time, and prevailing network states. DTA models are intended to describe the congestion (or queue) build-up/dissipation in response to the aforementioned demand management schemes in conjunction with intelligent transportation system measures and information supply strategies. However, existing DTA models have assumed constant VOT in evaluating pricing schemes. Consideration of heterogeneous users with a range of VOT requires computationally efficient bicriterion time-dependent shortest path (BTDSP) algorithms, for use in conjunction with route choice decisions of trip makers in the context of DTA models.

This paper focuses on the development of BTDSP algorithms for DTA applications to large networks with heterogeneous trip makers (in terms of VOT) and toll pricing on selected links. As mentioned by Dial (*3*), in the static bicriterion traffic assignment model based on utility maximization, all trips are distributed only among the set of extreme efficient paths. Therefore, this study focuses on search algorithms for extreme efficient solutions. To this end, an exact BTDSP algorithm is first developed as an extension of the static parametric approach (*8–10*). Although it is desirable to identify a complete set of extreme efficient solutions, introducing the time dimension can dramatically increase the number of extreme efficient paths, leading to considerable computational time and path storage space requirements, especially for large-scale DTA planning applications. The real-time nature of online operational DTA systems also imposes additional computational constraints. Recognizing that existing generic approximation methods (*15, 17*) lack a mechanism to effectively utilize the VOT distribution information, this paper presents a novel approximation scheme to intelligently allocate computational resources to VOT segments with different selection frequencies in the context of large-scale DTA applications. As pointed out by Marcotte (*19*), solving the parametric shortest path problem arising from network equilibrium problems approximately, by selecting in a suitable manner values of parameter, could allow bicriterion assignment problems to be solved as efficiently as single criterion problems.

The paper is structured as follows. The next section provides the problem statement, followed by a discussion of an exact solution algorithm for finding the complete set of extreme efficient paths. On the basis of the concept of ϵ-efficiency, a new measure of solution quality is then proposed to take into account the probability distribution of VOT. After presenting a binary search framework that minimizes expected approximation errors, this paper concludes with comprehensive numerical experiments on three real traffic street networks.

PROBLEM STATEMENT

Without loss of generality, the bicriterion time-dependent shortest path problem considered in this paper seeks to minimize two common path attributes simultaneously, namely, travel time and travel cost. Given a time-dependent network G (*N*, *A*), where *N* is the set of nodes and *A* is the set of directed links (i, j), the time period of interest is discretized into a set of small time intervals, $S = \{t_0, t_0 + \sigma, t_0 + 2\sigma, \ldots, t_0 + M\sigma\}$, where t_0 is the earliest possible departure time from any origin node, σ is a small time interval during which no perceptible changes in traffic conditions or travel cost occur, and M is a large number so that the intervals from t_0 to $t_0 + M\sigma$ cover the planning horizon. Each arc (i, j) is associated with two time-dependent arc attributes: $d_{ij}(t)$ and $c_{ij}(t)$, where $d_{ij}(t)$ is the time required to travel from node i to node j when departing at time t from node i, and $c_{ij}(t)$ denotes the corresponding cost of traveling arc (i, j) when departing at time t. Accordingly, $\delta_i(t)$, $\gamma_i(t)$, $\eta_i(t)$ can be denoted as the travel time, travel cost, and generalized disutility, respectively, of a path starting from origin r, at time t, to node i. The time-dependent generalized disutility function of a given arc (i, j) is assumed, given by the following linear form:

$$g_{ij}(t) = c_{ij}(t) + \alpha \times d_{ij}(t) \tag{1}$$

where α is interpreted as the value of time (i.e., the monetary cost per unit time).

In this study, heterogeneity of the user population is reflected in the value of time treated as a continuous random variable distributed across the population of travelers.

Let $p(r, s, t)$ be a feasible path starting from a given origin r, at time t, to a destination s, and let $\text{TT}(p)$ and $\text{TC}(p)$ be the travel time and travel cost, respectively, associated with path p. Let $P(r, s, t)$ be the set of all feasible paths $p(r, s, t)$ for a given (r, s, t). For simplicity and clarity, $P = P(r, s, t)$.

Definition 1. A path $p \in P$ is efficient (Pareto optimal or nondominated) if and only if it is not possible to find a different path $q \in P$ such that $\text{TT}(q) \le \text{TT}(p)$ and $\text{TC}(q) \le \text{TC}(p)$ with at least one strict inequality.

Let P^e be the set of efficient paths. An efficient path $p \in P^e$ in the solution space corresponds to an efficient point (or vector) $\text{Z}(p) = [\text{TT}(p), \text{TC}(p)]$ in the criterion space. Accordingly, the set of efficient points is denoted as Ω^e.

Definition 2. If a point $\text{Z}(p)$ lies on the boundary of the convex hull of Ω^e, then $\text{Z}(p)$ is an extreme efficient point and p is an extreme efficient path; otherwise $\text{Z}(p)$ is a nonextreme efficient point and p is a nonextreme efficient path.

As discussed previously, in the traffic assignment model based on utility maximization, all trips are distributed only among the set of extreme efficient paths. Thus, at each direction-finding step of the bi-objective DTA solution algorithm, a set of extreme efficient paths,

$P^{ex}(r, s, t)$, needs to be found for origin–destination (O-D) pair (r, s) and departure time interval t. The pth path $p \in P^{ex}(r, s, t)$ corresponds to a segment $(\alpha^{p-1}, \alpha^{p})$ in the feasible range of VOT. In the subsequent network-loading step (*6*), the probability of an O-D trip (r, s, t) using path p is

$$\text{prob}(p) = \int_{\alpha^{p-1}}^{\alpha^{p}} f(\alpha)d\alpha = \Phi(\alpha^{p}) - \Phi(\alpha^{p-1}) \tag{2}$$

where $f(\alpha)$ and $\Phi(\alpha)$ are the probability density function and cumulative density function, respectively, of random variable α.

EXACT ALGORITHM FOR FINDING COMPLETE SET OF EXTREME EFFICIENT PATHS

This section proposes an algorithm for finding the complete set of extreme efficient paths (i.e., the complete set of extreme efficient points in the criterion space) for the bi-objective time-dependent shortest path problem. The parametric shortest path approach proposed by Henig (*8*) and Mote et al. (*9*) for the static case is extended to time-dependent shortest paths.

Consider a given value of time α and the corresponding least-generalized cost path tree T, consisting of the least-generalized cost paths from origin r to each node i, for each departure time interval t. If an arc–departure time combination $[(i, j), t]$ remains out-of-tree (i.e., non-tree arc), the corresponding reduced cost should be nonnegative, leading to the following inequality:

$$\eta_i(t) + g_{ij}(t) - \eta_j[t + d_{ij}(t)] \geq 0 \tag{3}$$

For path $p(r, i, t)$, which starts from origin r, at time t, to node i, the node label with respect to generalized disutility can be expressed as the sum of the node labels in terms of travel time and travel cost.

$$\eta_i(t) = \sum_{(k,l,\tau)\in p(r,i,t)} [c_{kl}(\tau) + \alpha \times d_{kl}(\tau)] = \gamma_i(t) + \alpha \times \delta_i(t) \tag{4}$$

Let $t' = t + d_{ij}(t)$; the generalized disutility for path $p(r, j, t')$ from origin r, at time t', to node j can similarly be represented as

$$\eta_j(t') = \gamma_j(t') + \alpha \times \delta_j(t') \tag{5}$$

Substituting Equations 1, 4, and 5 back into Inequality 3 yields

$$[\gamma_i(t) + c_{ij}(t) - \gamma_j(t')] + \alpha[\delta_i(t) + d_{ij}(t) - \delta_j(t')] \geq 0 \tag{6}$$

Let $\text{RT}_{ij}(t) = \delta_i(t) + d_{ij}(t) - \delta_j(t')$ and $\text{RC}_{ij}(t) = \gamma_i(t) + c_{ij}(t) - \gamma_j(t')$ denote the reduced travel time and the reduced travel cost, respectively, of arc–departure time combination $[(i, j), t]$. On the basis of Inequality 6, the dependence of the least-generalized cost path tree on the single scalar VOT can be examined. For any out-of-tree arc for which $\text{RT}_{ij}(t) \neq 0$, the following two cases determine the sensitivity range of VOT that does not violate the reduced-cost optimality conditions.

If $\text{RT}_{ij}(t) > 0$

$$\alpha > -\frac{\text{RC}_{ij}(t)}{\text{RT}_{ij}(t)} \tag{7}$$

If $\text{RT}_{ij}(t) < 0$

$$\alpha < -\frac{\text{RC}_{ij}(t)}{\text{RT}_{ij}(t)} \tag{8}$$

Collectively, the lower and upper bounds of VOT can be calculated by scanning each out-of-tree arc–departure time combination $[(i, j), t]$

$$\alpha^{lb} = \max_{[(i,j),t]\in A}\left[-\frac{\text{RC}_{ij}(t)}{\text{RT}_{ij}(t)} : \text{RT}_{ij}(t) > 0\right] \tag{9}$$

$$\alpha^{ub} = \min_{[(i,j),t]\in A}\left[-\frac{\text{RC}_{ij}(t)}{\text{RT}_{ij}(t)} : \text{RT}_{ij}(t) < 0\right] \tag{10}$$

The least-generalized cost path tree T remains unchanged as long as $\alpha^{lb} \leq \alpha \leq \alpha^{ub}$. In other words, the closed interval $[\alpha^{lb}, \alpha^{ub}]$ defines the sensitivity range of VOT for keeping tree T optimal.

A computationally efficient time-dependent least-cost path algorithm, developed by Ziliaskopoulos and Mahmassani (*20*), is adopted in this study to solve the time-dependent least-generalized cost path (TDGCP) problem from origin r to each node i for each time interval (i.e., to obtain the TDGCP tree T for a given α). Each node $i \in N$ is associated with three vectors: $\delta_i = \{\delta_i(t)\}$, $\gamma_i = \{\gamma_i(t)\}$, and $\eta_i = \{\eta_i(t)\}$ $\forall t \in S$, corresponding to travel time, travel cost, and generalized disutility, respectively, of paths from origin r to node i for all possible departure time intervals. The algorithm is based on Bellman's general principle of optimality, and the least-generalized cost paths are calculated forward, starting from the origin node (in this implementation, and with no loss of generality). In each iteration, the algorithm selects and deletes the first node i, or "current node," from the scan eligible (SE) list. Then the current node i is scanned according to the following equation:

$$\eta_i[t + d_{ij}(t)] = \min\{\eta_j[t + d_{ij}(t)], g_{ij}(t) + \eta_i(t)\} \quad \forall j \in \Gamma\{i\} \tag{11}$$

for every time $t \in S$, where $\Gamma\{i\}$ is the set of nodes that can be directly reached from i (forward star). If at least one of the components of η_j is modified, node j is inserted in the SE list, and the other two label vectors (i.e., δ_j and γ_j) are updated accordingly. The algorithm repeats this process and terminates when the SE list is empty. The output of the algorithm includes the TDGCP tree T as well as the three node vectors δ_i, γ_i, and η_i associated with each node i. In particular, vectors γ_i and δ_i are used to calculate reduced travel time and reduced travel cost, respectively, of arc–departure time combinations, which provide essential input for Algorithm 1, described next.

With the sensitivity analysis results as a main building block, the following exact solution algorithm for finding the complete set of extreme efficient paths is presented.

Algorithm 1. Finding the complete set of extreme efficient paths

Initialize the current value of VOT α as α^{max}.
while $\alpha > \alpha^{min}$ do
Solve the TDGCP problem from origin r to each node i for each time t with α
Initialize $\alpha^{lb} = \alpha^{min}$
for each out-of-tree arc–departure time combination $[(i, j), t]$ do
Calculate $\alpha[(i, j), t] = -\dfrac{\text{RC}_{ij}(t)}{\text{RT}_{ij}(t)}$
if $\alpha[(i, j), t] > \alpha^{lb}$ and $\alpha[(i, j), t] < \alpha$, then $\alpha^{lb} = \alpha[(i, j), t]$
end for
Set $\alpha = \alpha^{lb} - \Delta$, and output α
end while.

Starting from the maximal (minimal) feasible value of VOT, the algorithm continues finding the lower (upper) bound α^{lb} of VOT for which the current shortest path tree T remains unchanged, until the minimal (maximal) feasible value of VOT is reached. It is able to directly move from one extreme efficient solution to the next one without redundant calculations on the nonextreme efficient solutions. To move to the next VOT segment and obtain a different tree T', a small value Δ needs to be subtracted from the α^{lb} found. This implies that travelers cannot distinguish differences in VOT below Δ per minute. The value of Δ also implicitly sets an upper bound for the number of breakpoints generated using Algorithm 1, with a value of $(\alpha^{max} - \alpha^{min})/\Delta$. A by-product of the preceding exact algorithm is a partition of VOT, through a series of breakpoints in the feasible range of VOT, each corresponding to an extreme efficient path tree.

APPROXIMATION METHOD FOR FINDING EXTREME EFFICIENT PATHS

When the problem (network) size is large, identifying a complete set of extreme efficient solutions may be computationally intractable. The approximation approach presented in this section aims at finding a subset of extreme efficient paths, where each path approximates (possibly) a subset of extreme efficient paths in a VOT segment, so as to minimize expected approximation errors [E(ϵ)] for a given VOT distribution—that is,

$$E(\epsilon) = \int_{a^{p-1}}^{a^{p}} \epsilon(\alpha) f(\alpha) d\alpha \tag{12}$$

where $\epsilon(\alpha)$ is the approximation error for a particular value of α.

This section first discusses possible measures for evaluating the quality of approximate solutions in a VOT interval and then develops a greedy search framework for solving a bi-objective time-dependent shortest path problem.

With the application of the TDGCP algorithm and the parametric sensitivity analysis presented in the last section with two feasible VOT α_k and α_{k+1}, where $\alpha_k < \alpha_{k+1}$, the shortest path trees T_k and T_{k+1} are obtained, together with the corresponding sensitivity intervals, denoted as $[\alpha_k^{lb}, \alpha_k^{ub}]$ and $[\alpha_{k+1}^{lb}, \alpha_{k+1}^{ub}]$. If $\alpha_k^{ub} \neq \alpha_{k+1}^{lb}$ (Figure 1) then there exists at least one extreme efficient point for VOT $\alpha \subset (\alpha_k^{ub}, \alpha_{k+1}^{lb})$. Interval $(\alpha_k^{ub}, \alpha_{k+1}^{lb})$ can be viewed as an active segment, as no extreme efficient point in this segment has been found. Let $p_1(r, i, t)$ and $p_2(r, i, t)$ denote two distinct paths from origin r to node i at time t in trees T_1 and T_2, respectively; then any extreme efficient path $p_3(r, i, t)$ in the active segment must satisfy that

$$\text{TC}[p_1(r,i,t)] < \text{TC}[p_3(r,i,t)] < \text{TC}[p_2(r,i,t)] \tag{13}$$

and

$$\text{TT}[p_2(r,i,t)] < \text{TT}[p_3(r,i,t)] < \text{TT}[p_1(r,i,t)] \tag{14}$$

Figure 2 presents the extreme efficient points Z_1 and Z_2 corresponding to paths $p_1(r, i, t)$ and $p_2(r, i, t)$ in the criterion space. Any possible extreme efficient point in the VOT interval of interest should be located inside the triangular area formed by Z_1, Z_2, and Z_4, where $Z_4 = \{\text{TT}[p_2(r, i, t)], \text{TC}[p_1(r, i, t)]\}$. With $p_1(r, i, t)$ and $p_2(r, i, t)$ available at hand, it is natural to select one of these two adjacent extreme efficient paths as the approximate solution for $\alpha \subset (\alpha_k^{ub}, \alpha_{k+1}^{lb})$, if the corresponding approximation error is acceptable. To evaluate the resulting solution quality in this bi-objective decision context, this study adapts the notion of ϵ-efficiency introduced by Warburton (*15*).

Definition 3. A path p ϵ-dominates path q if

$$\frac{\text{TT}(p)}{\text{TT}(q)} \leq 1+\epsilon \quad \text{and} \quad \frac{\text{TC}(p)}{\text{TC}(q)} \leq 1+\epsilon$$

In this case, path p is considered ϵ-efficient vis-à-vis path q. When $\epsilon = 0$, this definition reduces to the common notation of vector dominance.

It can be readily established that path $p_1(r, i, t)$ is $\epsilon_1(r, i, t)$-efficient vis-à-vis all possible extreme efficient solutions, where

$$\epsilon_1(r,i,t) = \frac{\text{TT}[p_1(r,i,t)]}{\text{TT}[p_2(r,i,t)]} - 1$$

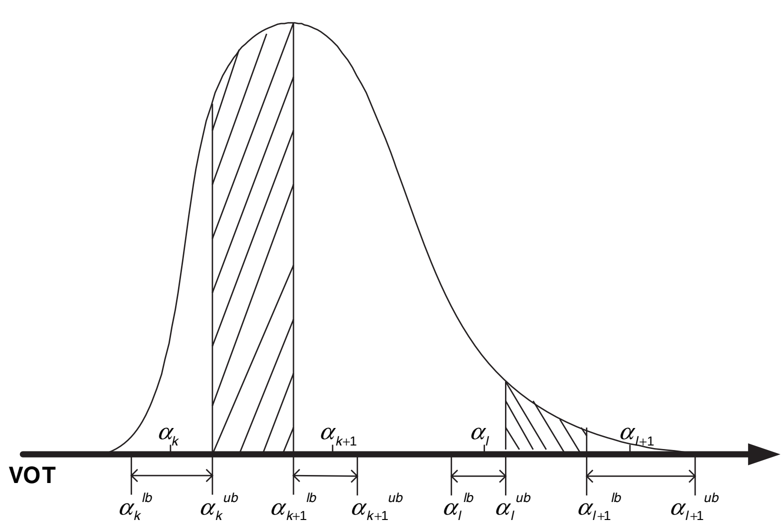

FIGURE 1 Probabilistic distribution of VOT.

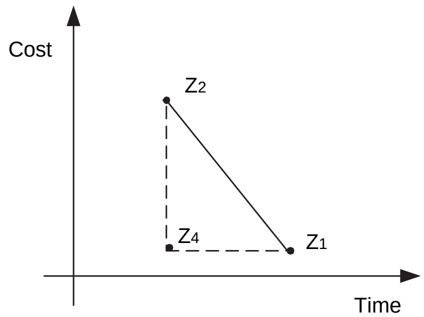

FIGURE 2 Criterion space for bi-objective shortest path problem.

This would require that

$$\frac{\mathrm{TT}[p_1(r, i, t)]}{\mathrm{TT}[p_3(r, i, t)]} \leq 1 + \epsilon_1(r, i, t) \tag{15}$$

and

$$\frac{\mathrm{TC}[p_1(r, i, t)]}{\mathrm{TC}[p_3(r, i, t)]} \leq 1 + \epsilon_1(r, i, t) \tag{16}$$

for each possible extreme efficient solution $p_3(r, i, t)$. Because $\mathrm{TC}[p_1(r, i, t)] < \mathrm{TC}[p_3(r, i, t)]$, only Inequality 15 needs to be considered. Because $\mathrm{TT}[p_3(r, i, t)] > \mathrm{TT}[p_2(r, i, t)]$,

$$\mathrm{TT}[p_1(r, i, t)]/\mathrm{TT}[p_3(r, i, t)] < \mathrm{TT}[p_1(r, i, t)]/\mathrm{TT}[p_2(r, i, t)]$$

Thus, the point (in objective space) corresponding to path $p_1(r, i, t)$ is $\epsilon_1(r, i, t)$-efficient for any point in the triangular area formed by Z_1, Z_2, and Z_4.

Similarly, it can be shown that path $p_2(r, i, t)$ is $\epsilon_2(r, i, t)$-efficient vis-à-vis all possible extreme efficient solutions, where

$$\epsilon_2(r, i, t) = \frac{\mathrm{TC}[p_2(r, i, t)]}{\mathrm{TC}[p_1(r, i, t)]} - 1$$

If

$$\mathrm{TT}[p_1(r, i, t)]/\mathrm{TT}[p_2(r, i, t)] < \mathrm{TC}[p_2(r, i, t)]/\mathrm{TC}[p_1(r, i, t)]$$

it would be advantageous to select path $p_1(r, i, t)$ because the corresponding solution efficiency $\epsilon_1(r, i, t)$ is less than $\epsilon_2(r, i, t)$; otherwise, path $p_2(r, i, t)$ would be selected. The solution quality for the VOT segment between α_k^{ub} and $\alpha_{k+1}^{\mathrm{lb}}$ would then be

$$\epsilon_k(r, i, t) = \min\left\{\frac{\mathrm{TT}[p_1(r, i, t)]}{\mathrm{TT}[p_2(r, i, t)]}, \frac{\mathrm{TC}[p_2(r, i, t)]}{\mathrm{TC}[p_1(r, i, t)]}\right\} - 1 \tag{17}$$

For the VOT distribution presented in Figure 1, the likelihood of randomly sampling an instance of VOT from active segment (α_k^{ub}, $\alpha_{k+1}^{\mathrm{lb}}$) is greater than from active segment (α_l^{ub}, $\alpha_{l+1}^{\mathrm{lb}}$). Thus, it would be desirable for the former segment, with higher choice probability, to have smaller errors to minimize the expected solution approximation error for the given VOT distribution. This consideration leads to the following statistical expectation criterion for characterizing solution quality. Note that $\epsilon_k(r, i, t) = 0$ for interval $[\alpha_k^{\mathrm{lb}}, \alpha_k^{\mathrm{ub}}]$.

$$\begin{aligned} E[\epsilon(r,i,t)] &= \sum_k \epsilon_k(r,i,t) \int_{\alpha_k^{\mathrm{lb}}}^{\alpha_k^{\mathrm{ub}}} f(\alpha) d\alpha \\ &+ \sum_k \epsilon_k(r,i,t) \int_{\alpha_k^{\mathrm{ub}}}^{\alpha_{k+1}^{\mathrm{lb}}} f(\alpha) d\alpha \\ &= \sum_k \left\{\epsilon_k(r,i,t) \times \left[\Phi(\alpha_{k+1}^{\mathrm{lb}}) - \Phi(\alpha_k^{\mathrm{ub}})\right]\right\} \end{aligned} \tag{18}$$

The preceding weighted approximation error formulation systematically captures the underlying heterogeneous characteristics of decision makers and realistically estimates actual approximation errors in the subsequent network loading process for traffic assignment. Because $\epsilon(r, i, t)$ is a nonnegative random variable, for a threshold $a > 0$, Markov's inequality (*21*) produces one possible bound on the deviation of a random variable from its mean:

$$\mathrm{prob}[\epsilon(r, i, t) > a] \leq \frac{E[\epsilon(r, i, t)]}{a} \tag{19}$$

This inequality provides a useful probabilistic performance measure of the approximation algorithms. Tighter bounds like Chebychev's bound and Chernoff's bound (*20*) could be calculated, but they require higher-order statistics of the distribution of ϵ.

Because a single run of the TDGCP algorithm generates a shortest path tree with origin r and changes the approximation errors for paths (r, i, t) $\forall\, t \in S$, $i \in N$, the measure of solution quality $Y(k, r)$ for an active VOT interval k should consider all paths in the time-dependent shortest path tree, that is,

$$Y(k,r) = \frac{1}{|N| \times |T|} \sum_{i,t} \epsilon_k(r,i,t) \times \left[\Phi(\alpha_{k+1}^{\mathrm{lb}}) - \Phi(\alpha_k^{\mathrm{ub}})\right] \tag{20}$$

In addition, one could reflect the relative importance of different O-D trip rates as follows:

$$\begin{aligned} Y(k,r) &= \frac{1}{|N| \times |T| \times \sum_{i,t} d(r,i,t)} \sum_{i,t} \left[\epsilon_k(r,i,t) \times d(r,i,t)\right] \\ &\times \left[\Phi(\alpha_{k+1}^{\mathrm{lb}}) - \Phi(\alpha_k^{\mathrm{ub}})\right] \end{aligned} \tag{21}$$

where $d(r, i, t)$ is the O-D trip rate from origin r to node i at time t.

On the basis of the preceding solution accuracy measures, the following binary search procedure is designed for intelligently allocating limited computational resources. Its basic idea is to start with a VOT interval containing the worst solution quality and iteratively shrink the intervals by adding new α values in the interval (α_k^{ub}, $\alpha_{k+1}^{\mathrm{lb}}$).

Algorithm 2. Binary search procedure

Step 1. Initialization. Solve the TDGCP problem with $\alpha^{\min}$ and $\alpha^{\max}$; add two boundary breakpoints $\alpha^{\min}$ and $\alpha^{\max}$, respectively.

Step 2. Termination criterion. If no interval has nonzero approximation errors, then terminate and output extreme efficient solutions. If the number of shortest path trees is less than a prespecified number K, continue the search process; otherwise stop.

Step 3. Interval selection. Select an interval (α_k^{ub}, $\alpha_{k+1}^{\mathrm{lb}}$) with the worst solution quality $Y(k, s)$.

Step 4. Shortest path calculation. Generate $\alpha_{new} = (\alpha_k^{ub} + \alpha_{k+1}^{lb})/2$ and solve the TDGCP problem with α_{new}.

Step 5. Evaluation of solution accuracy. Insert a new element that corresponds to α_{new} into the ordered set. Calculate solution quality measurements for two newly generated intervals (α_k^{ub}, α_{new}^{lb}) and (α_{new}^{ub}, α_{k+1}^{lb}), and go back to Step 2.

The preceding binary search procedure can be viewed as a greedy heuristic that aims to maximize the average solution quality (locally) at each iteration. By incorporating information from the parametric shortest path method, this procedure can avoid calculating the VOT within a sensitivity range $[\alpha_k^{lb}, \alpha_k^{ub}]$. Consequently, if the prespecified number K is not less than the number of breakpoints in the complete set of extreme efficient solutions, then this algorithm is able to identify all the breakpoints of VOT and corresponding extreme shortest path trees.

NUMERICAL EXPERIMENTS

Two sets of numerical experiments were conducted to examine the proposed BTDSP algorithms: one for the exact solution algorithm and the other for the approximation scheme. For convenience, and with no loss of generality, these algorithms are implemented as a backward procedure (i.e., rooted at the destination node, from all nodes to one node), because optimum path procedures in the context of DTA network models compute trees rooted at the destination. They are coded in Visual C++ 6.0 on a Windows XP platform and evaluated on a machine with a PentiumN III 2.0-GHz central processing unit and 2 GB of memory. The sizes (in terms of number of nodes and number of links) of the three real networks used in the experiments are as follows:

- Fort Worth (FW), Texas, 180 nodes and 445 links;
- Irvine, California, 326 nodes and 626 links; and
- Knoxville, Tennessee, 1,347 nodes and 3,004 links.

The planning horizon is set to 120 minutes, and the time-dependent travel times are obtained from the 2-hour simulation results output from the DYNASMART-P DTA model (*22*) for each of the three networks.

The purpose of the first set of experiments is to validate and to test the computational performance of the parametric BTDSP algorithm with respect to several problem size attributes. The first experiment is to explore the impact of introducing the time dimension to the static bicriterion shortest path (BSP) problem on the size of the solution set and associated computational effort; these issues are critical for developing online and off-line DTA models. Of particular interest is the relationship between the number of breakpoints (in the VOT range over which the BTDSP tree remains Pareto optimal) and the number of time intervals into which the planning horizon is discretized. The number of breakpoints is selected as a figure of merit because it can serve as a surrogate not only for computational time but also for size of the solution set. The length of a time interval is varied from 1 to 120. Time-dependent travel costs are randomly generated between 0.01 and 2 for every 30 minutes. The feasible range of VOT is set between \$0.0 and \$10.0 per minute. It is also assumed that travelers do not perceive differences in VOT below \$0.01 per minute, implicitly setting the maximal number of breakpoints to 1,000.

To study the impact of using different root nodes for the constructed trees, 10 different destination nodes were randomly selected from the Knoxville network. The results indicate that the number of breakpoints varies only slightly (less than 5%) for different destination nodes. Therefore, each data point in the following experiments reports the average value of 10 realizations, and for each realization a BTDSP tree is solved for a randomly selected destination.

For the three networks, Figure 3 presents the relationship between number of time intervals and number of breakpoints, and Figure 4 presents the relationship between number of time intervals and execution time. The experimental results indicate that the number of breakpoints is monotonically nondecreasing as the length of the aggregation time interval decreases. For example, the average number of breakpoints in the Knoxville network increases from 159.2 to 928.3 when the length of an aggregation time interval decreases from 120 minutes to 1 minute. As expected, for the same size of time intervals, a larger network has more breakpoints than a smaller network. For example, with a 1-minute time interval, the average number of breakpoints in the Knoxville network is 1.91 times that in the FW network.

As indicated in Figure 4, the computational times for the three networks, and especially for Knoxville, increase with the number of

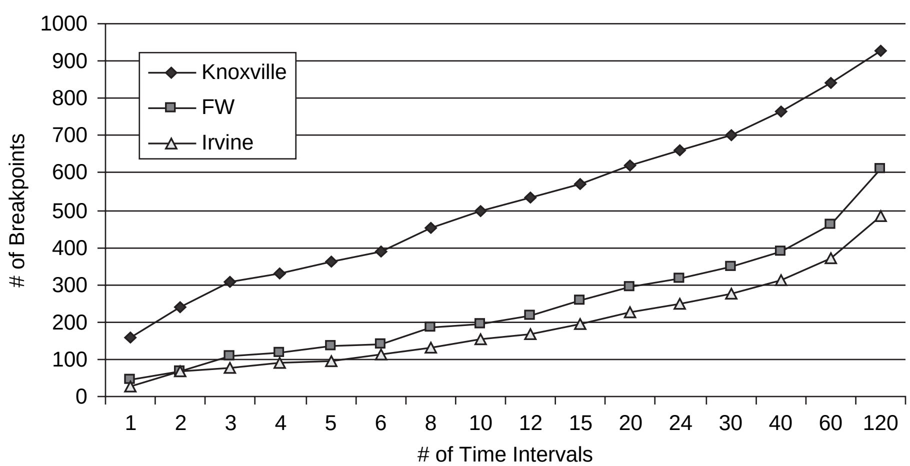

FIGURE 3 Relationship between number of time intervals and number of breakpoints.

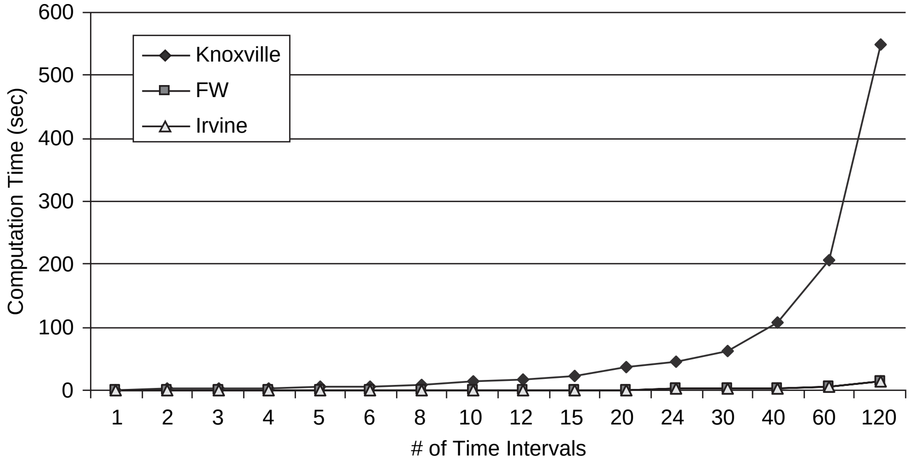

FIGURE 4 Relationship between number of time intervals and computation time.

discrete time steps. For example, computing one BTDSP tree takes an average of 9.15 minutes with a 1-minute aggregation time interval (120 time steps), compared with 1.02 seconds with a single 2-hour time interval.

The last group of experiments with the parametric BTDSP seeks to represent variable congestion pricing schemes more realistically by applying travel costs (road tolls) on only a given percent β of freeway and highway links, instead of imposing costs on all the links. The Knoxville network is used for these experiments, with an aggregation time interval of 5 minutes (24 time steps) and different travel costs generated for every 30-minute period. As indicated in Figure 5, the higher the toll link coverage, the more breakpoints in the complete solution set. In addition, even with only 10% of the freeway and highway links (around 82 links) selected as toll links (i.e., with nonzero travel cost), the corresponding solution set is still considerably large (494.2 breakpoints, equivalent to 73% of the solution set size for 100% toll link coverage).

In the second set of experiments, the Knoxville network was selected to compare the performance of alternative approximation algorithms. The length of a time interval was set to 5 minutes, and travel costs were randomly generated between 0.01 and 2 for every 30 minutes on each link. The probabilistic distribution function of VOT applied in this set of experiments is $f(\alpha) = 100\alpha \times e^{-1}$, previously used by Dial (*10*). Three alternative approximation search schemes were considered: (*a*) random weighting method, (*b*) greedy search that minimizes the maximum approximation errors among all VOT intervals, and (*c*) greedy search that minimizes the expected approximation errors. The third scheme was implemented according to the proposed binary search framework, and the second and third schemes retrieved the sensitivity analysis results by postprocessing the current least-generalized cost path tree.

Figure 6 presents the resulting expected approximation error, averaged over 10 realizations, based on Equation 18. As expected, the errors monotonically decrease with increasing number of iterations for the three approximation schemes. The decrease is more dramatic in the first 50 iterations. For the two greedy search schemes, the marginal solution improvement stabilizes after about 100 iterations, and the expected approximation errors decrease gradually toward zero. Because the random weighting method does not use the solution quality measure and the sensitivity information for each VOT interval, it exhibits a slower convergence rate. By taking into account the probabilistic distribution information, the third approximation scheme

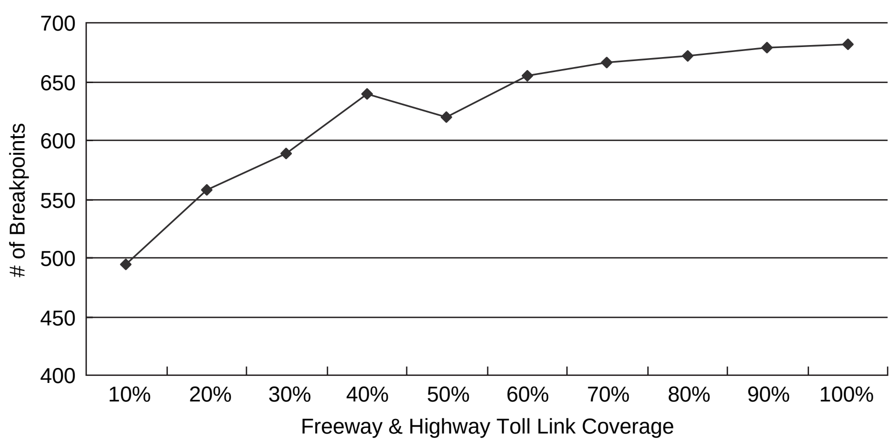

FIGURE 5 Impact of freeway and highway toll link coverage.

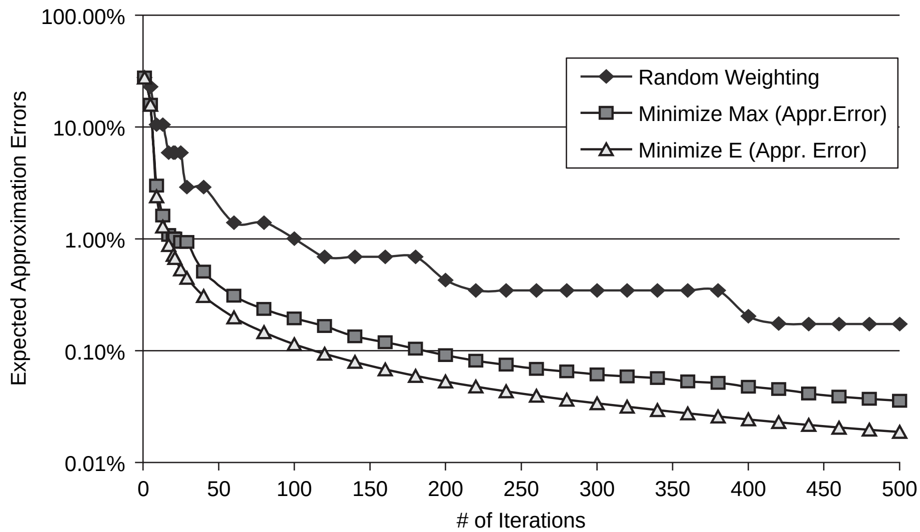

FIGURE 6 Solution quality of alternative approximation search schemes.

improves considerably the expected quality of solution compared with the first two schemes. After the first 24 iterations, on average, the expected approximation errors of the third scheme are only 53.3% and 10.2% of those of the first two methods, respectively. The three schemes take around 100, 21, and 17 iterations, respectively, to reduce the expected approximation errors to 1.0%. Furthermore, 120 iterations are required by the third scheme to reach a desired degree of accuracy of 0.1%, compared with 180 iterations for the second scheme. In summary, the proposed binary search approach successfully balances the trade-off between computational tractability and solution quality.

CONCLUDING REMARKS

With a particular emphasis on computational efficiency issues, this paper is motivated by the need to solve dynamic network assignment problems when prices are imposed on certain links of the network, and users with various preferences (value of time) react by selecting paths that minimize generalized cost of travel. This paper proposes both exact and heuristic approximation methods to find extreme efficient time-dependent shortest paths, which is a critical component for solving the motivating bi-objective DTA problem. The exact procedure is an extension and adaptation to the time-dependent case of a known parametric programming approach for solving the static bi-objective shortest path problem. It finds a complete set of extreme efficient paths from each node to a given destination at each departure time interval. However, exact procedures in this context may well explode in terms of combinatorial complexity in large networks. By using the parametric sensitivity analysis and underlying VOT distribution information, the proposed binary search framework determines a set of approximate solutions that minimize the expected approximation error over a feasible range of VOT. Both exact and approximation schemes (along with variants) were tested using three actual traffic networks. The experimental results indicate that the computation time and the size of the solution set are jointly determined by several key parameters such as the number of time intervals and the number of nodes in the network. The results also suggest that the proposed approximation scheme is computationally efficient for large-scale bi-objective time-dependent shortest path applications while maintaining satisfactory solution quality.

It has been evident over the past decade that continuing advances in dynamic network assignment methods intrinsically depend on advances in search procedures to find optimal paths that satisfy certain conditions that correspond to and are derived from the nature of the problem of interest. As the range of applicability of dynamic network assignment expands to include greater behavioral and policy realism, the complexity of the corresponding path-finding requirements increases. Road pricing is one such important policy variable that planners and policy makers are keenly interested in analyzing in a network setting. As dynamic assignment models are extended to address pricing and its implications on spatial network dynamics, the development of corresponding algorithms requires development of corresponding multiobjective path search procedures. Recognizing the heterogeneity of the network users considerably complicates the path search aspect compared with the assumption of identical users. The work presented in this paper is an example of the kind of advances to known path search algorithms that are necessary to meet the challenges of realistic dynamic network analysis tools. In addition to adaptations of exact procedures, it is often necessary to develop efficient and effective approximation schemes because the increase in computational complexity introduced by virtually any novel problem dimension tends to be quite steep. Continuing and future work in integrating the procedures described in this paper in both off-line and online DTA models, and in expanding the realm of application of these models, will undoubtedly give rise to additional challenging path-finding problems. This paper is offered as a contribution to the growing body of work on practically useful optimum path algorithms needed in the collective toolkit to advance the state of the art of dynamic network analysis.

REFERENCES

1. Small, K. A., and J. Yan. The Value of "Value Pricing" of Roads: Second-Best Pricing and Product Differentiation. *Journal of Urban Economics,* Vol. 49, 2001, pp. 310–336.

2. Lam, T. C., and K. A. Small. The Value of Time and Reliability: Measurement from a Value Pricing Experiment. *Transportation Research Part E: Logistics and Transportation Review,* Vol. 37, 2001, pp. 231–251.
3. Dial, R. B. Bicriterion Traffic Assignment: Basic Theory and Elementary Algorithms. *Transportation Science*, Vol. 30, No. 2, 1996, pp. 93–110.
4. Leurent, F. The Theory and Practice of a Dual Criteria Assignment Model with Continuously Distributed Values-of-Times. In *Transportation and Traffic Theory* (J. B. Lesort, ed.), Pergamon, Exeter, United Kingdom, 1996, pp. 455–477.
5. Elimam, A., and D. Kohler. Two Engineering Applications of a Constrained Shortest-Path Model. *European Journal of Operation Research,* Vol. 103, 1997, pp. 426–428.
6. Dial, R. B. A Model and Algorithms for Multicriteria Route-Mode Choice. *Transportation Research Part B: Methodological,* Vol. 13, 1979, pp. 311–316.
7. Leurent, F. Cost Versus Time Equilibrium over a Network. *European Journal of Operation Research*, Vol. 71, 1993, pp. 205–221.
8. Henig, M. I. The Shortest Path Problem with Two Objectives. *European Journal of Operation Research,* Vol. 25, 1985, pp. 281–291.
9. Mote, J., I. Murthy, and D. Olson. A Parametric Approach to Solving Bicriterion Shortest Path Problems. *European Journal of Operation Research,* Vol. 53, 1991, pp. 81–92.
10. Dial, R. B. Bicriterion Traffic Assignment: Efficient Algorithms Plus Examples. *Transportation Research Part B: Methodological,* Vol. 31, 1997, pp. 357–379.
11. Hansen, P. Bicriterion Path Problems. In *Multiple Criteria Decision Making: Theory and Applications* (G. Fandel and T. Gal, eds.), Springer, Heidelberg, Germany, 1980, pp. 109–127.
12. Miller-Hooks, E., and H. S. Mahmassani. Optimal Routing of Hazardous Materials in Stochastic, Time-Varying Transportation Networks. In *Transportation Research Record 1645,* TRB, National Research Council, Washington, D.C., 1998, pp. 143–151.
13. Garey, M., and D. Johnson. *Computers and Intractability: A Guide to the Theory of NP Completeness.* W. H. Freeman, San Francisco, 1979.
14. Climaco, J. C. N., and E. Q. V. Martin. A Bicriterion Shortest Path Algorithm. *European Journal of Operation Research,* Vol. 11, 1982, pp. 399–404.
15. Warburton, A. Approximation of Pareto Optima in Multiple-Objective Shortest Path Problems. *Operations Research,* Vol. 35, 1987, pp. 70–79.
16. Hassin, R. Approximation Schemes for the Restricted Shortest Path Problem. *Mathematics of Operation Research,* Vol. 17, 1992, pp. 36–42.
17. Nielsen, L. R. *Route Choice in Stochastic Time-Dependent Networks: Theory, Algorithms, and Computational Experience.* Ph.D. thesis. Department of Operations Research, University of Aarhus, Denmark, 2003.
18. Jayakrishnan, R., H. S. Mahmassani, and T.-Y. Hu. An Evaluation Tool for Advanced Traffic Information and Management Systems in Urban Network. *Transportation Research C,* Vol. 2, 1994, pp. 129–147.
19. Marcotte, P. Reformulations of a Bi-criterion Equilibrium Model. In *Reformulation—Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods* (M. Fukushima and L. Qi, eds.), Kluwer Academic, Dordrecht, Netherlands, 1999, pp. 269–292.
20. Ziliaskopoulos, A. K., and H. S. Mahmassani. Time-Dependent Shortest-Path Algorithm for Real-Time Intelligent Vehicle Highway System Applications. In *Transportation Research Record 1408,* TRB, National Research Council, Washington, D.C., 1993, pp. 94–100.
21. Ross, S. *A First Course in Probability.* Prentice-Hall, Inc., New York, 2002.
22. Mahmassani, H. S. Dynamic Network Traffic Assignment and Simulation Methodology for Advanced System Management Applications. *Networks and Spatial Economics,* Vol. 12, 2001, pp. 267–292.

The Transportation Network Modeling Committee sponsored publication of this paper.