Optimal Routing with Multiple Objectives: Efficient Algorithm and Application to the Hazardous Materials Transportation Problem

Chi Xie & S. Travis Waller*

Center for Transportation Research, School of Civil, Architectural and Environmental Engineering, The University of Texas, Austin, TX, USA

Abstract: *This article presents an efficient parametric optimization method for the biobjective optimal routing problem. The core process is a bounded greedy single-objective shortest path approximation algorithm. This method avoids the computationally intensive dominance check with labeling methods and overcomes the deficiency with existing parametric methods that can only find extreme nondominated paths. Moreover, we propose a decomposition scheme to convert a multiobjective routing problem into a number of biobjective problems. We then compare its computational performance against the classic label-correcting method over a set of synthetically generated random networks and illustrate its algorithmic advances and solution behaviors by an example application of routing hazardous materials in a U.S. northeastern highway network.*

1 INTRODUCTION

Optimal routing with respect to multiple performance measures poses important theoretical interests and practical applications. Such problems are ubiquitous across transportation, distribution, and communication networks (Current and Marsh, 1993; Boffey, 1995). They not only arise as stand-alone applications in their own right but also are often embedded as subroutines into more complex network optimization problems. The key algorithmic concern regarding a multiobjective routing problem is to identify a set of optimal routes that cover all possible or representative tradeoffs between conflicting objectives, so that a subsequent or interactive comparative assessment of cost-effective scenarios can be made.

*To whom correspondence should be addressed. E-mail: *stw@mail.utexas.edu*.

An important application area of multiobjective optimal routing is the transportation of hazardous materials. Shipments of hazardous materials, including flammable, explosive, poisonous, corrosive, and radioactive substances, liquids, or gases, as well as miscellaneous goods that are harmful to humans and the environment, constitute a significant part of the freight transportation commodities in the United States, and their volume is continuously increasing in recent decades. It was estimated that in the early 1980s, approximately 250,000 hazardous materials shipments were carried daily through the U.S. transportation system; this number has reached a level of over 800,000 per day in the year of 1998, which is by weight approximately 8.8 million tons (Rothberg and Hassan, 2000; Rothberg, 2001). Along with the increasing number of hazardous material shipments, the public is increasingly concerned about the potential risk to the population and environment caused by traffic accidents involving hazardous materials vehicles. The recently issued Hazardous Material Transportation Safety Reauthorization Act of 2001 emphasizes the importance of improving the routing regulation and operations for hazardous materials transportation.

The multiobjective setting and requirement for the hazardous materials routing problem reflect the problem's multifaceted nature. In general, a number of decision factors, such as efficiency, safety, and environment concern, must be taken into account in making routing decisions, which bear impact on different parties or stakeholders (e.g., the shipper, carrier, and affected populations). As a synthetic consideration, for example, we may incorporate the following objectives into the formulation of a hazardous materials routing problem, such as minimization of travel distance (or travel time or transportation cost), minimization of accident rate,

DOI: 10.1111/j.1467-8667.2011.00720.x

minimization of the affected population size (Turnquist, 1987; Nozick et al., 1997; Chang et al., 2005; Souleyrette and Sathisan, 1994). In this example, although it is reasonable to combine the second and third objectives to generate a new performance measure named risk exposure (e.g., Zografos and Davis, 1989), which is the product of population size and accident rate, further combination does not seem to be appropriate. This incommensurability between objectives necessitates the explicit inclusion of multiple objectives into a single optimal routing problem.

Nevertheless, in either the triobjective or the biobjective hazardous materials routing case, major changes with modeling complexity and solution algorithms occur as the number of objectives increases from one to two. Suppose a multiattribute directed network $G = (V, E, N)$, where V is the vertex or node set, E is the edge or arc set, and N is the attribute set of edges or arcs. Every arc $(i, j) \in E$ is associated with a set of attributes specified by N and its attribute vector is $c_{ij} = (c_{1,ij}, c_{2,ij}, \ldots, c_{n,ij})$, where $|N| = n$. Given origin node r and destination node s, an optimal routing problem with n objectives is to find all *efficient* or *nondominated* paths between r and s in terms of the n path attributes, where it is assumed that all arc attributes are *nonnegative* and *additive* along paths, or in other words, this optimal routing problem is of the *min-sum* type or is simply called a shortest path problem. We say that a path is efficient or nondominated if there does not exist any feasible path that is superior to this path in terms of each of the attributes/objectives. An n-objective shortest path problem (P1) can be in general written into the following vector form:

$$\min_{\mathbf{x}} \mathbf{z}(\mathbf{x}) = [z_k(\mathbf{x})]_{k=1,\ldots,n} = \left[\sum_{(i,j)\in E} c_{k,ij} x_{ij} \right]_{k=1,\ldots,n} \tag{1}$$

subject to

$$\sum_{\{j:(i,j)\in E\}} x_{ij} - \sum_{\{j:(j,i)\in E\}} x_{ji} = \begin{cases} 1 & i = r \\ 0 & \forall i \in V - \{r, s\} \\ -1 & i = s \end{cases} \tag{2}$$

$$x_{ij} = 0 \text{ or } 1 \quad \forall (i, j) \in E \tag{3}$$

where z_k is the objective function value or the sum of arc attributes along a feasible path in terms of attribute k, x_{ij} is the 0–1 decision variable with arc (i, j), indicating whether arc (i, j) is included in a path, and $\mathbf{z} = [z_k]_{1\times n}$ and $\mathbf{x} = [x_{ij}]_{1\times|E|}$ are the objective function vector and decision variable vector, respectively.

The functional form of the multiobjective shortest path problem is simply an extension from its single-objective counterpart. However, different from the single-objective problem, the integrality constraint (3) is required in the multiobjective model; otherwise, optimal solutions may contain nonintegral numbers, which fail to represent individual paths. Due to the imposition of such an integrality requirement, a multiobjective shortest path problem is known as an NP-complete problem in its worst case (see Garey and Johnson, 1979; Hansen, 1979). A numerical example shown in Figure 1 illustrates the requirement of the integrality constraint to the biobjective shortest path problem.

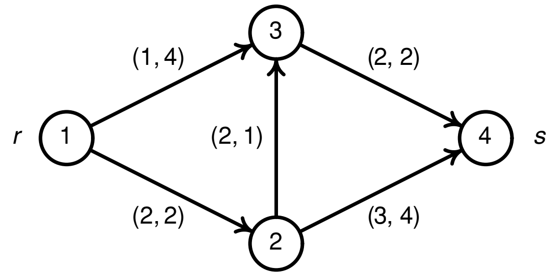


Fig. 1. An illustrative example of the biobjective shortest path problem.

For illustration, we are only concerned about finding shortest paths from node 1 to node 4 in the example network. If we consider the shortest path problem with respect to the first objective only, the resulting shortest path is 1-3-4, which corresponds to the optimal values of the decision variables: $x_{12} = 0$, $x_{13} = 1$, $x_{23} = 0$, $x_{24} = 0$, and $x_{34} = 1$; if we consider the shortest path problem with respect to the second objective, the shortest path is 1-2-3-4, where the corresponding optimal values of the decision variables are: $x_{12} = 1$, $x_{13} = 0$, $x_{23} = 1$, $x_{24} = 0$, and $x_{34} = 1$. Without the integrality requirement, the optimal solution of a single-objective shortest path problem automatically implies a single path. Then, we consider the shortest path problem with both objectives. If the integrality constraint is still not imposed, we obtain the nondominated solution set $\{x_{12} = \alpha,\ x_{13} = 1 - \alpha,\ x_{23} = \alpha,\ x_{24} = 0, \text{ and } x_{34} = 1\}$ and the nondominated objective vector set $\{(3+3\alpha, 6-\alpha)\}$, where $0 \le \alpha \le 1$. It is readily known that no solution from the nondominated solution set indeed represents an individual path unless $\alpha = 0$ or $\alpha = 1$.

The focus of this article is the deterministic multiobjective shortest path problem presented. This class of shortest path problems has been tackled by various solution strategies, including labeling methods (Hansen, 1979; Daellenbach and De Kluyver, 1980; Martins, 1984; Cox, 1984; Tung and Chew, 1988, 1992; Vincke, 1974; Loui, 1983; Warburton, 1987; Brumbaugh-Smith and Shier, 1989; Skriver and Anderson, 2000), ranking methods (Clímaco and Martins, 1982), constraint

methods (Lawler, 1976), and parametric methods (Robbins, 1983; Henig, 1985; Current et al., 1990; Coutinho-Rodrigues et al., 1999; White, 1982; Mote et al., 1991). A multiobjective shortest path problem could be also treated by heuristic or metaheuristic methods such as compromise programming (e.g., Lounis and Vanier, 2000), genetic algorithm (e.g., Liu et al., 1997), simulated annealing (e.g., Paya et al., 2008), ant system (e.g., Vitins and Axhausen, 2009), hybrid metaheuristic (e.g., Rama Mohan Rao and Shyju, 2010). For extensive surveys of general optimal routing problems, interested readers are referred to Current and Marsh (1993), Boffey (1995), and Raith and Ehrgott (2009).

The labeling methods constitute the major part of existing multiobjective shortest path solution techniques. The labeling methods include two types: label-setting methods (Hansen, 1979; Daellenbach and De Kluyver, 1980; Martins, 1984; Cox, 1984; Tung and Chew, 1988, 1992) and label-correcting methods (Vincke, 1974; Loui, 1983; Warburton, 1987; Brumbaugh-Smith and Shier, 1989; Skriver and Anderson, 2000), in terms of how label sets are updated at nodes and how shortest path labels "converge" to the optimal set. Either type of labeling methods is a multidimensional extension of its single-objective counterpart. The key algorithmic feature distinguishing a multiobjective labeling procedure from its single-objective version is merely the storage of multilabel vectors and the use of the vector dominance rule at each node in the dynamic updating process. Because of this relatively simple algorithmic logic and ease of implementation, labeling methods have become the dominating solution algorithm for the hazardous materials routing problem (see Turnquist, 1987; Nozick et al., 1997; Chang et al., 2005; Cox, 1984; Miller-Hooks and Mahmassani, 1998; Erkut, 1995, 2007). However, given the nondomination nature of solutions, the number of the label sets at a node could be as high as $\sum_{k=0}^{|V|-2} (|V|-2)!/(|V|-2-k)!$ in the worst case, where $|V|$ is the total number of nodes in the network (Tung and Chew, 1992), which makes it difficult for this type of solution method to be applied to large networks.

The major contribution of this article is the development of a parametric algorithm that can efficiently approximate the nondominated path set of a multiobjective shortest path problem. In particular, we present a polynomial-time solution procedure that can potentially identify both extreme and nonextreme nondominated paths through an iterative constrained single-objective shortest path search. Despite that it does not guarantee the completeness of the nondominated solution set, the algorithm's performance is positively justified by promising results in our tests; it also identifies "representative" solution points that can sketch the nondominated solution profile in just a few initial iterations. This latter feature makes it particularly attractive in the circumstance that we need to make a routing decision in a prompt manner.

We structure this article as follows. We first discuss the deficiencies of existing parametric methods and present our algorithmic considerations that overcome these deficiencies, which result in a constrained parametric algorithm and a problem decomposition scheme for problems with three or more objectives. Then, the characteristic and advantage of these algorithmic advances are illustrated in detail through a hazardous materials transportation case study in a U.S. northeastern highway network. Finally, the research findings are summarized and further research directions suggested in the last section.

2 PARAMETRIC ALGORITHM

2.1 Existing parametric methods

Parametric methods for multiobjective shortest path solutions appear in the literature in two types. The first type uses a parameterized utility function combining all the objectives so as to convert a multiobjective problem into a series of single-objective shortest path problems with a range of parameter values (Robbins, 1983; Henig, 1985; Current et al., 1990; Coutinho-Rodrigues et al., 1999). Each parameter in the utility function serves as a weight for its corresponding objective; nondominated solutions are obtained sequentially through exhausting the parameter range and solving the corresponding parameterized problem. This solution strategy originates from Cohon's (1978) noninferior set estimation (NISE) method for multiobjective linear programming (MOLP) problems. The second type resorts to a linear relaxation of the original integer-restricting functional form of the multiobjective problem (White, 1982; Mote et al., 1991). Extreme nondominated solutions, which all belong to the extreme solution set of the relaxed linear program, can be readily identified by the pivoting operation of a multiobjective simplex method. In the simplex solution framework, the pivoting operation iteratively moves the search from one extreme solution to another adjacent extreme one along the boundary of the feasible solution region of the relaxed linear program. Although both types of parametric methods rely on a linear relaxation to the original problem, the difference between them is that the former type of methods solves a set of single-objective LP problems while the latter directly deals with a multiobjective LP problem.

A common algorithmic feature pertaining to the parametric methods is that nondominated solutions are generated individually as a sequence, each of which can be solved in general by some polynomial-time algorithm. Those "independent" individual solution searches may be conducted in a parallel manner, which further reduces the required computational time through implementing parallel processing techniques. Moreover, as needed, the order of generating individual solutions may be customized to give priority to those most "representative" solution points. This characteristic is very useful as in many cases it is not necessary to present all the nondominated solutions to the decision maker but a subset of key solutions supporting the nondominated profile. In an actual computational process, we may suspend the solution search when the number and range of collected nondominated solutions have satisfied a certain threshold. Because of this flexibility, a parametric method can be used in a variety of different objective preference information environments, including *a priori*, *a posteriori*, or interactive preference articulation cases.

However, the deficiency of existing parametric methods is also apparent: they can only find a subset of nondominated solutions, that is, extreme nondominated solutions, which are those minimizing convex combinations of individual objectives. Unless all the objectives are commensurate with each other, the extreme solution set cannot in general represent the whole nondominated solution set. Here the commensurability implies that different objectives can be quantitatively measured by some common standard. For example, given two routing objectives, travel cost and travel time, we say that two objectives are commensurable if there exists a value of time and hence travel time can be converted to and evaluated in terms of monetary cost. The reason for this partial solution problem is that use of a utility function or linear relaxation yields a more aggressive vector dominance condition, which in principle omits all nonextreme solutions. Due to this reason, parametric methods at best can only be believed as heuristic algorithms, or they must be jointly used with other methods that are capable of finding nonextreme nondominated solutions. In fact, such a mixed solution strategy has been suggested in a few parametric algorithm implementations. For example, Current et al. (1990) used their NISE-like parametric algorithm with an auxiliary constrained shortest path algorithm; Coutinho-Rodrigues et al. (1999) suggested a combination of an NISE-like algorithm and a k-shortest path algorithm; Mote et al. (1991) developed a two-phase procedure that uses a multicriterion simplex method in the first phase and then a label correcting method in the second phase.

In addition, it must be noted that though these parametric methods have been successfully developed for biobjective shortest path problems, it does not seem straightforward to extend the algorithmic procedure to deal with problems with three or more objectives. Some researchers, for example, Boffey (1995), suggested treating extra objectives more than two as side constraints or subsidiary objectives. Such a treatment, however, might ignore some attractive nondominated solutions to the original problem. Because of this multidimensional difficulty, it is not surprising that all the parametric methods listed above have only been applied to biobjective cases.

2.2 A generic biobjective solution framework

Prior to discussing our parametric method, it is necessary to interpret a couple of fundamental terms we will use throughout this text, namely, extreme and nonextreme nondominated solutions, and full and partial nondominated objective sets.

The notion of extreme solutions plays an important role in understanding the algorithmic behavior of a parametric method. A solution $\mathbf{x}$ to a multiobjective shortest path problem formulated as (1) is an extreme solution, if it cannot be represented as a strict convex combination of two distinct feasible solutions in the feasible region F' of the integer-relaxing problem, where F' is defined by constraints (2). In other words, an extreme solution $\mathbf{x}$ is not equal to $\lambda_1\mathbf{x}_1 + \lambda_2\mathbf{x}_2$, given any $\mathbf{x}_1, \mathbf{x}_2 \in F'$, and $\lambda_1 + \lambda_2 = 1, 0 \leq \lambda_1, \lambda_2 \leq 1$, if $\mathbf{x} \neq \mathbf{x}_1$ and $\mathbf{x} \neq \mathbf{x}_2$. Extreme and nonextreme nondominated solutions are those nondominated solutions satisfying and dissatisfying this extreme condition, respectively.

A full nondominated objective set to an n-objective shortest path problem is an n-dimensional objective vector that satisfies the nondominated conditions. A partial nondominated objective set to the same problem is an m-dimensional objective vector that satisfies the nondominated condition in terms of its m objectives, where $2 \leq m \leq n$. Suppose an arbitrary feasible solution to an n-objective shortest path problem. As we will discuss later, if this solution is nondominated in terms of a partial objective set, it is nondominated in terms of the full objective set; if the solution is nondominated in terms of the full objective set, it is nondominated in terms of at least one of its partial objective set.

The algorithmic motivation of a parametric algorithm resembles the use of Lagrangian multipliers in solving an optimization problem with side constraints. In a Lagrangian relaxation method, some "hard" side constraints are relaxed and supplemented into the objective function as Lagrangian terms; then, the original problem can be tackled by repeatedly solving the

relaxed problem with updating Lagrangian multipliers. In a similar spirit, nondominated solutions of a multiobjective optimization problem may be obtained through repeatedly solving a "relaxed" single-objective problem whose objective function is a combination of all original objectives with varying weighting parameters.

In its most frequently used form, a biobjective shortest path problem may be parameterized to be a single-objective problem (P2) via a convex combination.

$$\min z_{\mathbf{w}}(\mathbf{x}) = \mathbf{w} \cdot \mathbf{z}(\mathbf{x}) = \sum_{l=1,2} w_l z_l(\mathbf{x}) = \sum_{(i,j)\in E} \sum_{l=1,2} w_l c_{l,ij} x_{ij} \tag{4}$$

subject to constraints (2)

$$x_{ij} \geq 0 \ \ \forall (i, j) \in E \tag{5}$$

where $\mathbf{w} = (w_1, w_2)$ is the parameter set and $\sum_{l=1,2} w_l = 1$ and $w_l \geq 0$, $\forall l$, hold. It can be seen that the parameterized problem is a standard shortest path problem, where the attribute of arc (i, j) is also a convex combination of the original arc attribute set, that is, $c_{\mathbf{w},ij} = \sum_{l=1,2} w_l c_{l,ij}$. Note that after this transformation, the integrality constraint (3) is no longer required for the parameterized problem; instead, a nonnegative constraint (2.5), $x_{ij} \geq 0, \forall (i, j) \in E$ is used.

A generic parametric solution procedure for generating the objective parameter values and finding nondominated solutions can be described as follows:

Step 1 (Initialization): Set $\mathbf{w}_1 = (1 - \varepsilon, \varepsilon)$ and $\mathbf{w}_2 = (\varepsilon, 1 - \varepsilon)$, where ε is a sufficiently small number, that is, $0 < \varepsilon \ll 1$. It is apparent that the values of $\mathbf{w}_1$ and $\mathbf{w}_2$ are so set as to obtain two extreme nondominated solutions that are optimal to the first and second objectives, respectively, over the whole feasible solution region. Use of a small ε value in these initial parameter sets is to avoid the risk of choosing a dominated solution when multiple optimal solutions to an initial relaxed problem are tied. The two resulting parameterized single-objective problems (P2) corresponding to $\mathbf{w}_1$ and $\mathbf{w}_2$ are then solved and the obtained optimal objective vectors, $\mathbf{z}_1 = (z_{1,1}, z_{1,2})$ and $\mathbf{z}_2 = (z_{2,1}, z_{2,2})$, are recorded into a first-in-first-out (FIFO) list as a vector pair, $(\mathbf{z}_1, \mathbf{z}_2)$.

Step 2 (Parameter generation): Select the first pair of objective vectors $(\mathbf{z}_1, \mathbf{z}_2)$ from the FIFO list and generate a new weighting parameter set $\mathbf{w} = (w_1, w_2)$ in terms of $(\mathbf{z}_1, \mathbf{z}_2)$ using the perpendicular method, which forms the following linear system:

$$\begin{bmatrix} z_{1,1} - z_{2,1} & z_{1,2} - z_{2,2} \\ 1 & 1 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} \dfrac{z_{2,2} - z_{1,2}}{(z_{2,2} - z_{1,2}) + (z_{1,1} - z_{2,1})} \\ \dfrac{z_{1,1} - z_{2,1}}{(z_{2,2} - z_{1,2}) + (z_{1,1} - z_{2,1})} \end{bmatrix} \tag{6}$$

The perpendicular method generates a vector perpendicular to the line going through the two points represented by the objective vectors $\mathbf{z}_1$ and $\mathbf{z}_2$ in the objective space. Note that either the first or second equation alone with the third equation in the linear system can lead to the above parameter result. It is readily seen that as long as $z_{2,2} - z_{1,2}$ and $z_{1,1} - z_{2,1}$ are simultaneously positive or negative, the linear system always has a feasible solution for $\mathbf{w}$. In fact, this is the sufficient and necessary condition for $\mathbf{x}_1$ and $\mathbf{x}_2$ being nondominated to each other. In other words, if $\mathbf{x}_1$ and $\mathbf{x}_2$ are nondominated solutions, there must exist a parameter set $\mathbf{w}$ satisfying $\sum_{l=1,2} w_l = 1$ and $w_{l=1,2} \geq 0$.

Step 3 (Solution generation): Solve the following parameterized, doubly constrained single-objective shortest path problem (P3) specified by $\mathbf{w} = (w_1, w_2)$ and $(\mathbf{z}_1, \mathbf{z}_2)$:

$$\min z_{\mathbf{w}}(\mathbf{x}) = \mathbf{w} \cdot \mathbf{z}(\mathbf{x}) = \sum_{l=1,2} w_l z_l(\mathbf{x}) = \sum_{(i,j)\in E} \sum_{l=1,2} w_l c_{l,ij} x_{ij} \tag{7}$$

subject to constraints (2)

$$\sum_{(i,j)\in E} c_{l,ij} x_{ij} < \max(z_{1,l}, z_{2,l}) \ \ \forall l = 1, 2 \tag{8}$$

$$x_{ij} = 0 \text{ or } 1 \ \ \forall (i, j) \in E \tag{9}$$

Note that constraint (8) can be rewritten as $\sum_{(i,j)\in E} c_{l,ij} x_{ij} \leq \max(z_{1,l}, z_{2,l}) - \varepsilon$, where ε is a very small positive number. If there exists an optimal solution $\mathbf{x}^*$ to the above single-objective shortest path problem (P3), this optimal solution is a candidate nondominated solution to the original biobjective problem (P1). We then form two new objective vector pairs, $(\mathbf{z}_1, \mathbf{z}^*)$ and $(\mathbf{z}^*, \mathbf{z}_2)$, and add them into the FIFO list. On the other hand, if no optimal solution exists, we can conclude that there does not exist any nondominated solution in the objective space confined by $\mathbf{z}_1$ and $\mathbf{z}_2$ (i.e., $[z_{2,1}, z_{1,1}]$ and $[z_{1,2}, z_{2,2}]$, if $z_{1,1} > z_{2,1}$ and $z_{1,2} < z_{2,2}$, or $[z_{1,1}, z_{2,1}]$ and $[z_{2,2}, z_{1,2}]$, if $z_{1,1} < z_{2,1}$ and $z_{1,2} > z_{2,2}$); no more search effort needs to be spent in this confined region. We then delete the used vector pair $(\mathbf{z}_1, \mathbf{z}_2)$ from the FIFO list.

Step 4 (Stopping criterion check): If the objective vector list is empty, stop the search; otherwise, go to Step 1. In case the complete set of nondominated

solution set is not required, we can use an alternative stopping criterion, such as a maximum number of iterations or a maximum number of nondominated solutions generated to terminate the search earlier.

The following proposition provides a proof for the correctness of the constrained parametric method for the biobjective shortest path algorithm and guarantees the "convergence" of the complete nondominated solution set.

Proposition 1. Given two nondominated solutions, $\mathbf{x}_1$ and $\mathbf{x}_2$, and their corresponding objective vectors, $\mathbf{z}_1$ and $\mathbf{z}_2$, we intend to construct a parameterized problem min $z_{\mathbf{w}}(\mathbf{x})$ in (4) for identifying a nondominated solution to the biobjective shortest path problem. The optimal solution to the parameterized, doubly constrained shortest path problem, if it exists, is another nondominated solution to the biobjective shortest path problem. The collection of optimal solutions of all such parameterized problems is the complete set of extreme and nonextreme nondominated solutions of the biobjective problem.

Proof. It is clear that the objective functions of the optimal solution to the parameterized, doubly constrained shortest path problem, $\mathbf{z}^* = (z_{*,1}, z_{*,2})$, satisfy the following condition:

$$z_{*,1} \in \left(\min\left\{z_{1,1}, z_{2,1}\right\}, \max\left\{z_{1,1}, z_{2,1}\right\}\right)$$
$$z_{*,2} \in \left(\min\left\{z_{1,2}, z_{2,2}\right\}, \max\left\{z_{1,2}, z_{2,2}\right\}\right)$$

Because it is the optimal solution, no other feasible solution of the parameterized problem has a smaller objective function value, and nor does other feasible solution dominate it. Therefore, the optimal solution to the parameterized problem is a nondominated solution to the biobjective shortest path problem.

Now suppose that there exists a nondominated solution $\mathbf{x}^*$ of the biobjective shortest path problem not included in the nondominated solution set identified by the constrained parametric method. We can sort all the nondominated solutions in the increasing order in terms of one of the objectives, which results in a nondominated solution sequence that has the decreasing order in terms of the other objective (see Brumbaugh-Smith and Shier, 1989). Then we choose a pair of consecutive nondominated solutions, $\mathbf{x}_1$ and $\mathbf{x}_2$, whose objective values $\mathbf{z}_1$ and $\mathbf{z}_2$ satisfy $z_{1,1} < z_{*,1} < z_{2,1}$ and $z_{1,2} > z_{*,2} > z_{2,2}$. It is apparent that such a nondominated solution $\mathbf{x}^*$ can be obtained by solving a parameterized, doubly constrained shortest path problem with its parameters determined by $\mathbf{x}_1$ and $\mathbf{x}_2$ using the perpendicular method (see (6)). This contradicts our hypothesis. Therefore, solving parameterized problems with all possible parameters will result in a complete set of nondominated solutions of the biobjective shortest path problem. ■

So far we have assumed that the parameterized problem can be optimally solved in Step 3; this optimal solution is also a nondominated solution to the biobjective problem. However, if the optimality is not guaranteed (i.e., suboptimal) in Step 3, an extra step called dominance check needs to be inserted between Step 4 and Step 5. The purpose of this step is to check whether the nondominated solution obtained in Step 3 dominates any other existing candidate nondominated solutions in the FIFO list. If any candidate nondominated solution is found to be dominated, it should be deleted from the FIFO list.

Note that for the doubly constrained shortest path problem in Step 3, an upper bound is imposed for each objective that is the larger one of the objective function values of the two parameter-generating solutions $\mathbf{x}_1$ and $\mathbf{x}_2$ (see (8)). This constraint confines the search in the restricted solution space by the two given nondominated solutions, $\mathbf{x}_1$ and $\mathbf{x}_2$, and enables nonextreme nondominated solutions, if any, to not be dominated by $\mathbf{x}_1$ and $\mathbf{x}_2$ in solving the parameterized problem. Given these added constraints, the integrality constraint (9) must be included in the constrained shortest path problem, which introduces the combinatorial complexity.

The constrained shortest path problem is an NP-hard problem due to this arising combinatorial complexity. No efficient algorithm is available to find the optimal solution of any constrained shortest path problem in polynomial time. A variety of approximate solution algorithms have been proposed for the constrained problem, including implicit enumeration method (Aneja et al., 1983), Lagrangian relaxation method (Handler and Zang, 1980; Jaffe, 1984; Carlyle et al., 2008), distributed method (Reeves and Salama, 2000; Sun and Langendorfer, 1998), ε-approximation method (Hassin, 1992; Lorenz and Raz, 2001), etc. Among these algorithm developments, Hassin's (1992), Jaffe's (1984), and Carlyle et al.'s (2008) methods can perform with a pseudo-polynomial complexity for exact solutions. A computational review of exact and approximate algorithms for the constrained shortest path problem is offered by Dumitrescu and Boland (2003). However, in general, depending on the desired solution quality, these algorithms may be still too expensive to be applied to large networks with multiple constraints (Korkmaz et al., 2002).

2.3 A constrained shortest path subroutine

To limit the computational complexity in a polynomial-time bound, we propose an approximate algorithm that

can find an optimal or near-optimal solution of the doubly constrained shortest path problem through a dynamic programming process. This procedure has an algorithmic structure analogous to a typical label-setting method for the standard shortest path problem, which starts from origin node r with setting its label equal to 0 and permanently labels nodes in the order of their distances from node r until destination node s or all nodes in the network are permanently labeled. However, this approximate algorithm maintains three labels for each node j in addition to the label recording the upstream node: one for the parameterized objective, $d(j) = w_1 d_1(j) + w_2 d_2(j)$, and two for the two single objectives, $d_1(j)$ and $d_2(j)$, respectively. The purpose of using the two single-objective labels is to make any obtained solution satisfy the objective constraint (i.e., constraint (8)). This is simply realized by a feasibility check during the dynamic scanning and updating process, which is conducted for every node j when a new temporary label set for this node is formed through an upstream node i; it simply checks whether, for each objective l, the temporary label formed by a path from origin node r to this node j via an upstream node i, $d_l(i) + c_{l,ij}$, plus the permanent label formed by the predetermined shortest path from destination node s to node j, $d_l'(j)$, satisfies the corresponding objective constraint, that is, $d_l(i) + c_{l,ij} + d_l'(j) < \max(z_{1,l}, z_{2,l})$. If the feasibility check for both objectives is passed, the temporary label for the parameterized objective, $w_1(d_1(j) + c_{1,ij}) + w_2(d_2(jt) + c_{2,ij})$, is then compared to the existing parameterized label at this node to determine if an updating for this label is required; otherwise, the temporary label set is discarded. If such an updating is warranted, the whole set of labels, including the parameterized-objective label and the two single-objective labels, are updated with the new temporary label set.

This approximate constrained shortest path solution method can be described by the following pseudo code, which synthetically serves as the major algorithmic component of Step 3 of the parametric method.

Algorithm doubly constrained single-objective shortest path subroutine
begin
  initialize $S := \emptyset$, $\bar{S} := V$;
  set $d_l(i) := \infty$, $\forall i \in V$, $\forall l = 1, 2$, $d(r) := \infty$;
  set $d_l(r) := 0$, $\forall l = 1, 2$, $d(r) := 0$;
  while $s \notin \bar{S}$ or $d_l(j) := \infty$, $\forall j \in \bar{S}$ **do**
  begin
    move node i from $\bar{S}$ to S for i satisfying $d(i) = \min\{d(j) : j \in \bar{S}\}$;
    for each $(i, j) \in A(i)$ **do**
    begin
      if $d_l(i) + c_{l,ij} + d_l'(j) < \max\left(z_{l,1}^{(k,q)}, z_{l,2}^{(k,q)}\right)$, $\forall l = 1, 2$ **then**
      begin
        if $d(i) + \sum_l w_l^q c_{l,ij} < d(j)$ **then**
        begin
          $d_l(j) := d_l(i) + c_{l,ij}$, $\forall l = 1, 2$;
          $d(j) := \sum_l w_l^q d_l(j)$;
          $pred(j) := i$;
        end
      end
    end
  end
  retrieve optimal solution **x** and its attribute set **z**;
  if $s \in S$ **then**
  begin
    conclude **x** is a nondominated solution and return it to the main routine as $\mathbf{x}^*$;
  end
  otherwise
  begin
    conclude no nondominated solution is found;
  end
end

It is readily seen that the key algorithmic steps of the constrained shortest path subroutine include a feasibility check and a dominance check. The feasibility check guarantees any chosen partial path satisfying the problem's constraint; the dominance check chooses the optimal partial path from the candidate feasible partial path set. We call the parametric method plus its subroutine—the constrained label-setting method—a constrained parametric method. As we have commented, this constrained parametric method is capable of potentially finding both extreme and nonextreme nondominated solutions, as a result of the imposition of the objective constraints as well as the use of a constrained shortest path solution procedure. The remaining concern with this algorithm is its solution efficiency and completeness.

It is observed that the constrained parametric method has two major changes compared to existing parametric methods: first, a shortest path tree from the destination node to all other nodes in the network is created at the initial stage for each objective; second, a constrained shortest path problem needs to be solved at each iteration, instead of a simple standard shortest path problem. These added algorithmic components only slightly increase the computational cost and memory requirement. The computation of a shortest path tree for each objective can be accomplished by an

ordinary labeling method, which is an efficient, one-time task and does not add a significant increase of the computing time to the whole algorithmic process. The computing time for solving the constrained problem by the approximate label-setting method is also maintained at the polynomial-time level. Compared to the label-setting method for the standard shortest path problem, the only added operations in the approximate label-setting method are two additions and two comparisons at each label updating.

Though the approximate label-setting algorithm ensures the feasibility of solutions, it does not always guarantee the solution optimality. Suboptimality may occur when the optimal path is dominated (in terms of the parameterized objective) by an infeasible path (which has at least one of the single objectives violating the corresponding constraint) at some node j when the status of the infeasible path reaching this node is feasible, that is, $d_l(i) + c_{l,ij} + d_l'(j) < \max(z_{1,l}, z_{2,l})$. Note that the left-hand side of the inequality is an upper bound of the minimum objective function value of objective l. Violation of this constraint does not necessarily mean the infeasibility of a path; some feasible path might be discarded by using this "overaggressive" check. As a result, if this discarded path happens to the real optimal path, it means that the final solution of the parameterized problem is suboptimal. However, in the parametric solution framework, the negative impact caused by this suboptimality condition could be considerably reduced by a few algorithmic settings.

There are a few algorithmic reasons contributing to this reduction. First, if the optimal solution to the parameterized, doubly constrained problem is an extreme nondominated solution to the biobjective problem, the algorithm always finds the optimal solution; in fact, in this case, removal of the objective constraints does not affect the optimization result. Second, a suboptimal solution to the parameterized, doubly constrained problem is often still a nondominated solution to the biobjective problem. So in this case, the suboptimality does not actually affect the solution quality of the biobjective problem. Third, even if a suboptimal solution is not a nondominated solution, it may be detected when a new nondominated solution that dominates it is found by the algorithm. In this case, the suboptimal solution, or the dominated solution, can be discarded from the nondominated solution set. Fourth, as we will show below, in a multiobjective shortest path problem with three or more objectives, which can be decomposed to a number of biobjective problems, any suboptimal solution obtained by solving one of the biobjective problems may be detected by solving other biobjective problems, which further reduces the possibility of suboptimal solutions appearing in the nondominated solution set.

2.4 A decomposition scheme for multiobjective problems

It seems difficult to apply a similar constrained parametric procedure to an n-objective problem when $n \geq 3$. The following proposition provides a simple strategy to decompose a multiobjective problem to a number of biobjective problems. The complete nondominated solution set of the original multiobjective problem is a union set of the nondominated solution sets of all these biobjective problems.

Proposition 2. A solution to an n-objective optimization problem $\mathbf{x}$ with its objective set $\mathbf{z}(\mathbf{x}) = (z_1(\mathbf{x}), \ldots, z_l(\mathbf{x}), \ldots, z_n(\mathbf{x}))$ is nondominated if and only if there exists at least such one m-dimensional partial objective set $\mathbf{z}'(\mathbf{x}) = (\ldots, z_l(\mathbf{x}), \ldots)$, where $2 \leq m \leq n$ and $z_l(\mathbf{x}) \in \mathbf{z}(\mathbf{x}), \forall l$, that the solution is Pareto-optimal in terms of its m objectives.

Proof. Suppose that a solution $\mathbf{x}$ to an n-objective minimization problem is nondominated in terms of the m-dimensional partial objective set $\mathbf{z}'(\mathbf{x}) = (\ldots, z_l(\mathbf{x}), \ldots)$, where $2 \leq m \leq n$. According to the definition of dominance, its nondominant condition to the optimization problem is not changed when we consider it in n objectives instead of m objectives, as $\mathbf{z}'(\mathbf{x}) \subseteq \mathbf{z}(\mathbf{x})$. The sufficiency is proved.

On the other hand, suppose that a solution $\mathbf{x}$ is nondominated in terms of the n-dimensional full objective set $\mathbf{z}(\mathbf{x})$. This means that there does not exist any other feasible solution that dominates $\mathbf{x}$, or, in other words, for any other solution $\mathbf{y}$, there exists at least one objective l so that $z_l(\mathbf{y}) > z_l(\mathbf{x})$ (in the case of a minimization problem). Thus, we can construct an m-dimensional partial objective set $\mathbf{z}'(\mathbf{x}) = (\ldots, z_l(\mathbf{x}), \ldots)$, where $2 \leq m \leq n$ and $z_l(\mathbf{x}) \in \mathbf{z}'(\mathbf{x})$. Such a condition ensures that solution $\mathbf{x}$ is nondominated in terms of these m objectives. The necessity is proved. ■

Therefore, combing the sufficient and necessary conditions, we can conclude that a solution to an n-objective minimization problem $\mathbf{x}$ with its objective set $\mathbf{z}(\mathbf{x}) = (z_1(\mathbf{x}), \ldots, z_l(\mathbf{x}), \ldots, z_n(\mathbf{x}))$ is nondominated if and only if there exists at least such an m-dimensional partial objective set $\mathbf{z}'(\mathbf{x}) = (\ldots, z_l(\mathbf{x}), \ldots)$, where $2 \leq m \leq n$ and $z_l(\mathbf{x}) \in \mathbf{z}(\mathbf{x}), \forall l$, that the solution is nondominated in terms of its m objectives.

Furthermore, it is readily known that the nondominated solution set of an n-objective problem in terms of its n objectives is a union set of all the nondominated solution sets of the same problem in terms of its m objectives, where $2 \leq m \leq n$, given that each m-objective set is a subset of the n-objective set and each objective in the n-objective set exists in at least one of the m-objective sets.

It is clear that the above conclusion provides a problem decomposition and solution combination method to dealing with a higher dimensional multiobjective optimization problem by solving a number of lower dimensional problems. All solution points of each lower dimensional problem in its objective space are a projection of those of the higher dimensional problem in the higher dimensional objective space. Under this projection scheme, we know that if we decompose an n-objective optimization problem into a set of m-objective problems, where $m \leq n$, the complete set contains $n!/m!(n-m)!$ m-objective problems. In this case, as we have developed an efficient algorithm for biobjective shortest path problems, any multiobjective shortest path problem can be indirectly tackled by solving a number of decomposed biobjective problems. For example, we can decompose a triobjective problem into three biobjective problems.

3 NUMERICAL EVALUATION AND EXAMPLE APPLICATION

In this section, we present a comparative evaluation for the proposed constrained parametric method against a labeling method—the classic multiobjective label-correcting method—and an illustrative case study of a multiobjective hazardous materials routing problem in the U.S. northeastern highway network. In particular, the main purpose of the following analysis is twofold: (1) numerically evaluate the solution quality and efficiency of the proposed algorithm over a set of synthetic networks of different problem sizes and attribute sets; (2) empirically analyze the algorithm's solution-generating behavior as well as its practical implications through a real-world routing example.

3.1 Numerical evaluation

3.1.1 Experiment design. The benchmark algorithm we choose for the numerical evaluation is the Brumbaugh-Smith and Shier's (1989) version of the classic biobjective label-correcting method. This is an exact solution method that can generate the complete nondominated solution set, and in the authors' opinion, Brumbaugh-Smith and Shier's implementation scheme is the most efficient one among all the available versions of this type of algorithm in the literature. Both the label-correcting algorithm and the constrained parametric algorithm are coded in C++. In our experiments, all the tests discussed are run on a computer equipped with a Dual-Core 3.40GHz CPU and 2GB RAM.

A set of 10 random networks of grid-like topology are generated, whose size ranges from 78 to 3,920 nodes and from 240 to 12,480 arcs (see Table 1). As we will see below, this range is sufficiently wide to exhibit the variation of the relative computation performance between the labeling method and the parametric method. The ratio of the number of arcs over the number of nodes in these networks is approximately equal to 3, which is the typical arc/node ratio of most surface transportation networks (see, e.g., Van Vliet (1978)). We also assign 10 random arc attribute sets to each of these networks, where each arc attribute set consists of a set of two-attribute vectors for all arcs. All these arc attribute values are uniformly distributed between 0 and 100 in an independent and identical manner. Following this network generation procedure, we obtain 100 topology-attribute network scenarios in total. We arbitrarily select the node at the top left corner as the origin node and the node at the bottom right corner as the destination node in each of these network-attribute scenarios.

Table 1
Performance comparison of the label-correcting method and the constrained parametric method

			Label-correcting method		*Constrained parametric method*	
Network	*Number of nodes*	*Number of arcs*	*Average number of nondominated paths*	*Average computation time (seconds)*	*Average number of nondominated paths*	*Average computation time (seconds)*
1	78	240	9.4	0.009	7.7	0.036
2	280	880	23.1	0.117	16.7	0.238
3	462	1,456	29.3	0.247	23.9	0.373
4	688	2,716	40.7	0.707	34.5	0.754
5	1,061	3,360	49.5	3.119	47.8	1.328
6	1,512	4,800	80.9	7.973	67.8	3.180
7	2,044	6,496	115.0	14.382	100.1	5.560
8	2,822	8,976	142.5	42.364	126.1	9.107
9	3,348	10,656	156.0	70.160	127.3	11.933
10	3,920	12,480	211.5	173.396	166.7	17.714

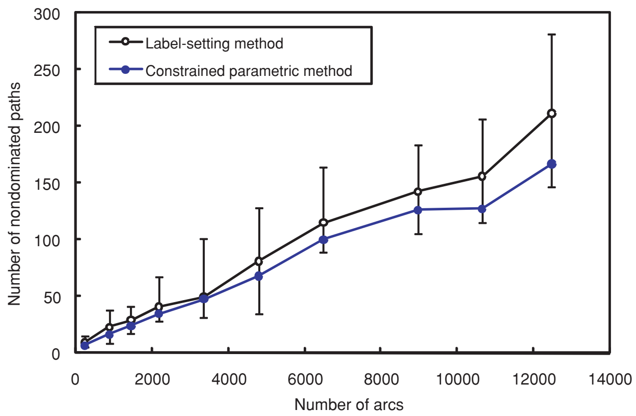
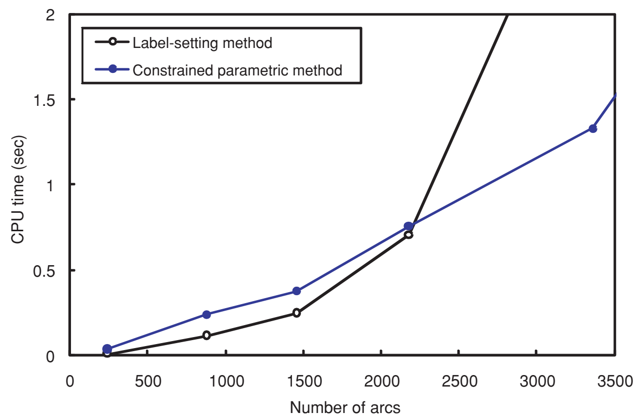


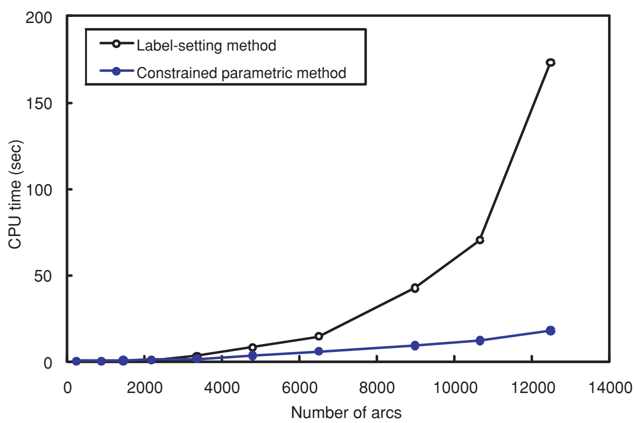
Fig. 2. Number of generated nondominated paths over network size.

3.1.2 Result analysis. Two performance measures are used to assess the computational results in the following evaluation: (1) solution quality or completeness (i.e., the number of nondominated solutions); (2) solution efficiency (i.e., the computation time). First, we evaluate the solution quality by counting the numbers of generated nondominated paths in these network scenarios with varying size. In Figure 2, we plot the average, highest and lowest numbers of nondominated paths identified by the labeling method for each network scenario over its 10 different attribute sets. It is shown that the number of nondominated paths in the complete solution set increases in an approximately linear manner with the network size increasing. In contrast, we also depict the average number of nondominated paths generated by the parametric method for each network case, as shown in Figure 2. It can be seen that the parametric method maintains a rather high solution identification rate (i.e., the ratio of the number of identified nondominated solutions over the total number of nondominated solutions) over all the network cases, in which its average solution identification rate varies from 0.70 to 0.96.

It is expected that the parametric method can perform much more efficiently than the labeling method on the price of missing a small part of the nondominated solution set. Both the average computation times consumed by the labeling method and the parametric method for each network scenario are plotted in Figure 3. Here we present this comparison in two different scales: a small scale for networks with the number of arcs less than 3,500 and a large scale for the full range of network sizes we use. It is clearly observed that, when the network size is relatively small (i.e., the number of arcs in the network is less than about 2,100), the labeling method runs faster than the parametric method (see



(a) A range of small network sizes



(b) The full range of network sizes

Fig. 3. Computation cost over network size.

Figure 3a); however, with the network size increasing, the computation time required by the labeling method increases drastically while the parametric method increases its computation time at a quite moderate speed (see Figure 3b). For the largest network (i.e., the network with 3,920 nodes and 12,480 arcs) in the set, for example, the average computation time spent by the labeling method is about 173.4 seconds; in contrast, the parametric method merely costs 17.7 seconds on average for the same network.

It is known that the computation cost of a labeling method is primarily determined by the number of dominance checks during the labeling process, while the computation cost of the proposed parametric method is a function of the number of actual nondominated paths as well as the search process of two preliminary destination-to-origin shortest paths. When the network size is small, the number of potential nondominated paths arriving at an intermediate node and hence the

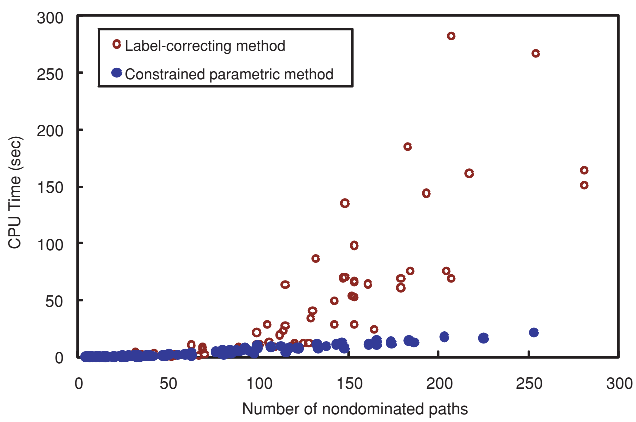


Fig. 4. Combination of solution quality and efficiency.

number of dominance checks during the search process is relatively small. Therefore, the dominance check process of a multiobjective labeling method is not significantly more costly than that of its single-objective counterpart in the same network. In contrast, the parametric method requires the nondominated paths generated one by one, each of which is a single-objective shortest path problem, plus the requirement of searching for two destination-to-origin paths. Therefore, its required computation cost is expected to be significantly higher than that of a single-objective path search in the same network.

However, when the network size becomes large, the number of potential dominance checks at any intermediate node increases in an exponential manner in its worst case, so the resulting computation cost of the labeling method. On the other hand, the computation cost of the parametric method increases with the number of actual nondominated paths, which increases only moderately in our case. Moreover, the cost for searching two destination-to-origin shortest paths, though increasing with the increasing network size in its absolute value, shrinks to a relatively small part of the whole computation cost. Apparently, it is the polynomial-time complexity feature of the parametric method that makes it outperform the most efficient labeling method in terms of solution efficiency when the size of a biobjective routing problem tends to be large.

In addition to the average computational performance, we also present the computational result of each individual topology-attribute scenario in terms of the combination of solution quality and efficiency (see Figure 4). It can be seen that across all these topology-attribute scenarios, the parametric method typically requires much less computation time in finding a comparable number of nondominated solutions than the labeling method, especially when the network size is large. This comparison of the individual results further confirms the efficiency advantage of the parametric method in searching for nondominated solutions of biobjective shortest path problems, in that it constructs a better trade-off between the solution quality and efficiency.

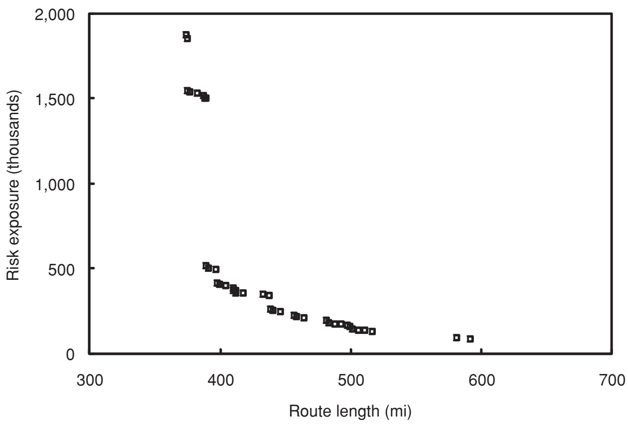
For a complete presentation, we summarize in Table 1 the problem sizes and the computation results about the solution quality and efficiency of the two biobjective optimal path algorithms from the above numerical evaluation.

These numerical results provide us with an overall understanding about the computational performance of the proposed constrained parametric method for biobjective shortest path problems; however, its algorithmic characteristics and solution behaviors and patterns may not be fully revealed by only observing the two performance measures. In fact, it is those algorithmic and solution features that make the parametric method be a more attractive tool in solving and analyzing multiobjective shortest path problems. We supplement below an example application that is of relatively small size but has realistic implications to highlight these features and also illustrate the use of the decomposition scheme.

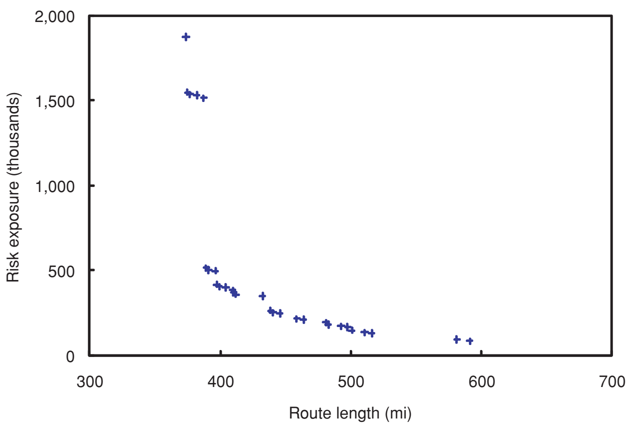
3.2 Example application

3.2.1 Problem description. The presented example problem involves shipments of a flammable liquid from an origin in Philadelphia, Pennsylvania, to a destination in Portland, Maine, through the U.S. Interstate Highway system in the northeastern states, which is created based on the data set used in Nozick et al. (1997) and Chang et al. (2005). This northeastern highway network consists of 233 nodes and 630 arcs. Three conflicting objectives are defined in this routing problem: (1) minimization of the length of the route; (2) minimization of the rate of accidents resulting in a release of hazardous material; (3) minimization of the size of populations that are potentially affected by a release of hazardous material.

In addition to this original setting of three routing objectives, we also construct a biobjective routing case that is formed with only the first objective of route length and a new objective named risk exposure. As we introduced before, this latter performance measure may be calculated on the arc level by multiplying the accident rate and the population size. This biobjective problem construction is needful and has its practical value: first, the focus of our discussion in this text is a biobjective shortest path problem; second, it is of interest to us whether such a biobjective problem could approximate the original triobjective problem well in a realistic problem setting.



(a) The label-correcting method



(b) The constrained parametric method

Fig. 5. Nondominated solutions in the objective space.

To this end, we carry out the routing analysis in two parts. First, we examine the performance of the constrained parametric algorithm in solving the biobjective routing problem; then, we discuss an analysis of the decomposition scheme applied to the triobjective routing problem and assess the similarity in the routing result between the biobjective problem and the triobjective problem.

3.2.2 Biobjective routing case. The complete solution set generated by the labeling method in 0.168 seconds contains 39 nondominated solutions. In contrast, the parametric method only finds 30 nondominated solutions with the expense of 0.262 seconds. These 30 solutions, including 17 extreme nondominated solutions and 13 nonextreme nondominated solutions, correspond to a solution identification rate of 76.9%. The generated nondominated solution points and the relationship between the number of solutions and the consumed computing time are depicted in Figures 5 and 6, respectively.

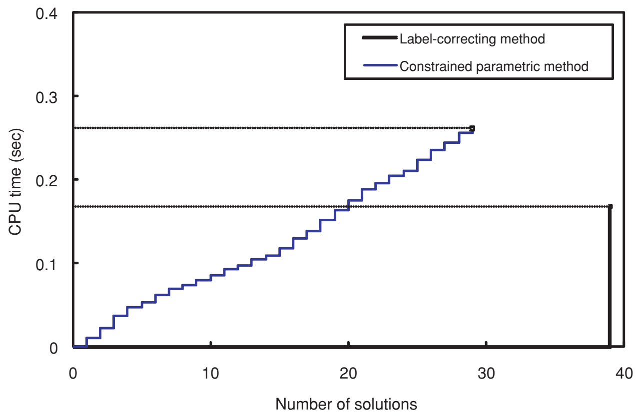


Fig. 6. Nondominated solution generation over time.

It is apparent that if a complete nondominated solution set is required, the labeling method beats the parametric method in both solution quality and efficiency. The advantage of the parametric method, however, lies in its gradual solution generation behavior. For example, if the required number of nondominated solutions is less than 20, the parametric method outperforms the labeling method (see Figure 6). More importantly, the parametric method has the priority of generating those "most representative" solution points, following the generation of the two initial extreme nondominated solutions that are exclusively optimized for each of the two objectives.

This can be seen, for instance, by labeling the first six nondominated solutions generated by the parametric search process (see Figure 7). The six solutions represent very different tradeoffs between the route length minimization and the risk exposure minimization, among which some solutions denote a dominating

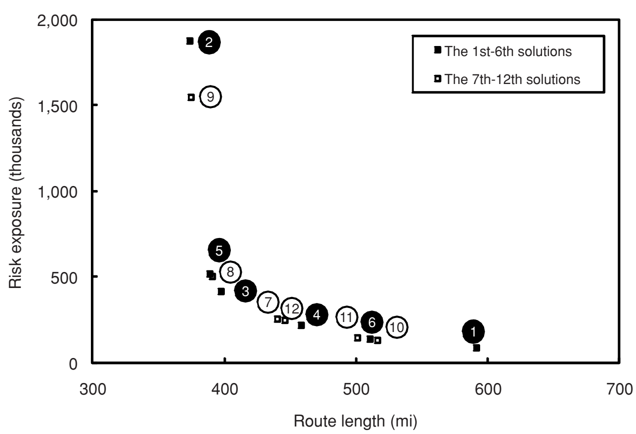


Fig. 7. Representative nondominated solutions.

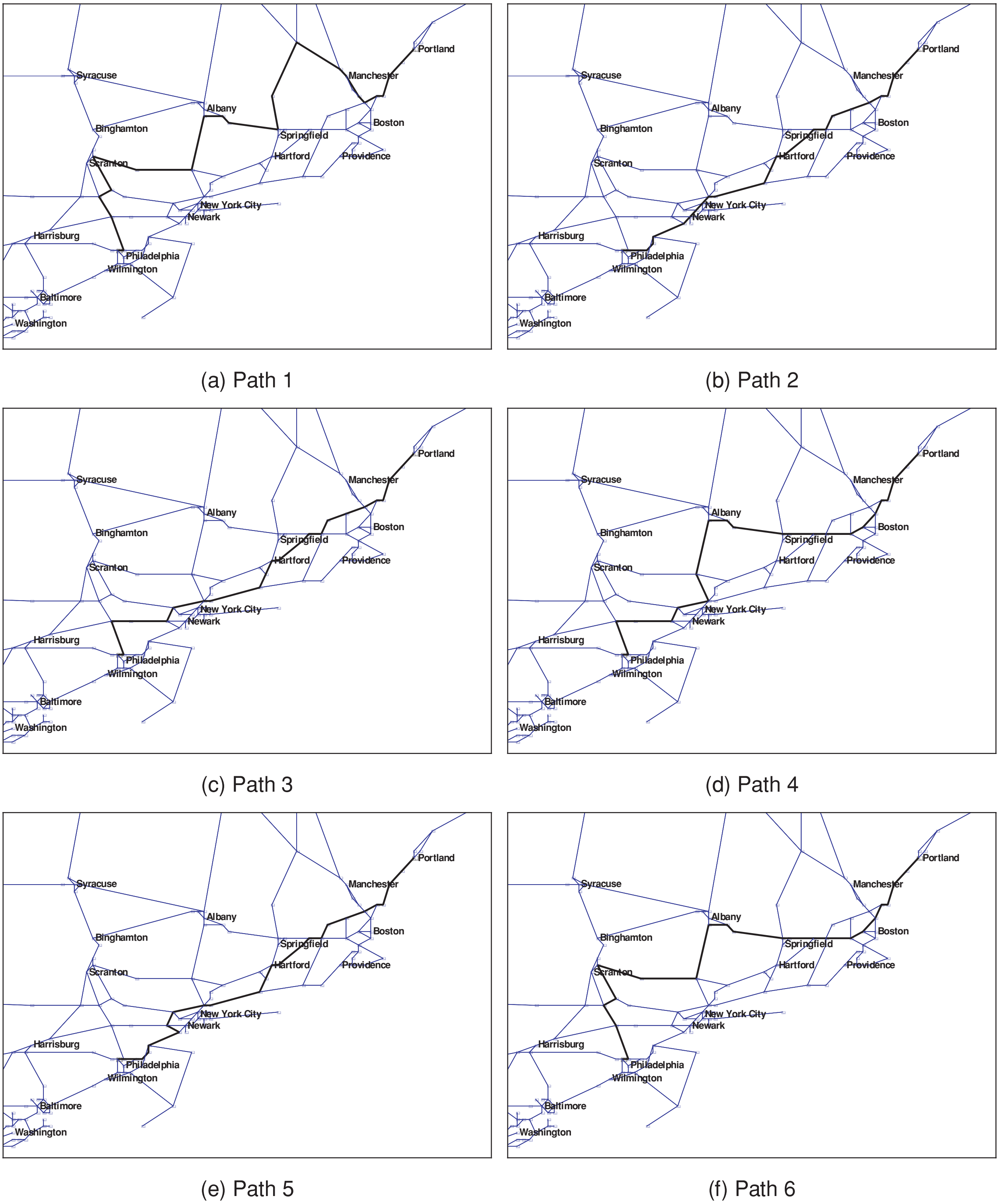
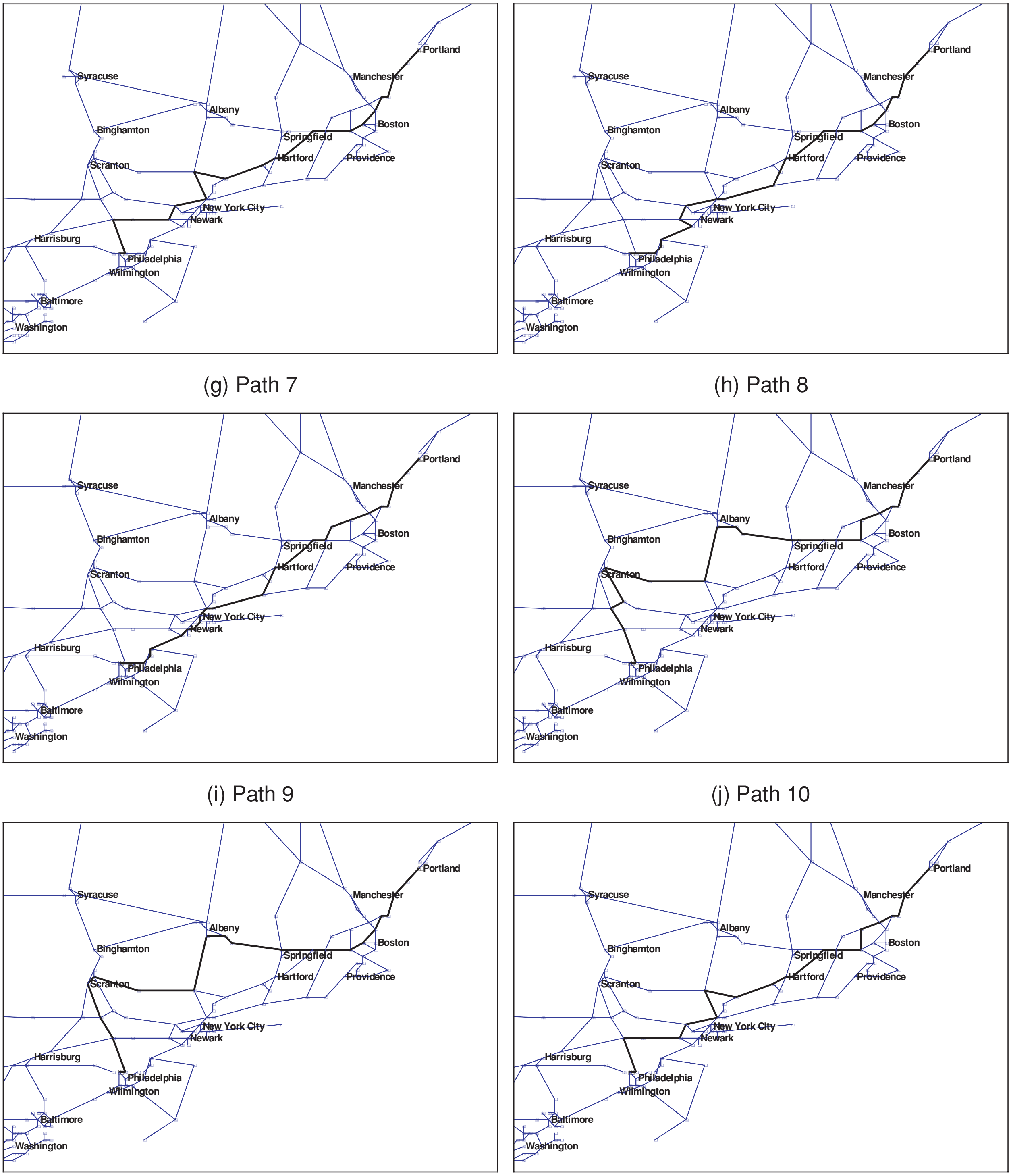


Fig. 8. Representative nondominated paths.



(g) Path 7 (h) Path 8

(i) Path 9 (j) Path 10

(k) Path 11 (l) Path 12

Fig. 8. (*Continued*)

Table 2
Nondominated solution sets of the decomposed subproblems and their synthetic result

Nondominated objective vector w.r.t route length and accident rate	*Nondominated objective vector w.r.t accident rate and population size*	*Nondominated objective vector w.r.t route length and population size*	*Nondominated objective vector w.r.t route length, accident rate, and population size*
(373.8, 143403.9)	-	(373.8, 360.8)	(373.8, 143403.9, 360.8)
(374.8, 141805.9)	-	-	(374.8, 141805.9, 360.8)
(375, 119160.7)	-	(375, 320.2)	(375, 119160.7, 320.2)
(376.8, 118376.8)	-	-	(376.8, 118376.8, 320.2)
(382, 117999.5)	-	-	(382, 117999.5, 320.2)
(387.4, 115771.3)	-	(387.4, 295.2)	(387.4, 115771.3, 295.2)
(389.2, 114987.4)	-	-	(389.2, 114987.4, 295.2)
(389.2, 42629.2)	-	-	(389.2, 42629.2, 306.7)
(391, 41845.3)	-	-	(391, 41845.3, 306.7)
(397.2, 34901.4)	-	(397.2, 258.4)	(397.2, 34901.4, 258.4)
(398.9, 34117.5)	-	-	(398.9, 34117.5, 258.4)
(409.5, 31512)	-	(409.5, 233.4)	(409.5, 31512, 233.4)
(411.3, 30728.1)	-	-	(411.3, 30728.1, 233.4)
(416.5, 30350.7)	-	-	(416.5, 30350.7, 233.4)
-	(215.8, 15350.7)	(438.5, 215.8)	(438.5, 15350.7, 215.8)
-	(182.2, 23600.7)	-	(405.5, 23600.7, 182.2)
(440.3, 23299.9)	-	-	(440.3, 23299.9, 215.8)
(445.5, 22922.5)	-	-	(445.5, 22922.5, 215.8)
-	(205.8, 20489.8)	-	(428.9, 20489.8, 205.8)
(456.8, 20301.9)	-	-	(456.8, 20301.9, 215.8)
(458.6, 19518)	-	-	(458.6, 19518, 215.8)
-	-	(480.8, 192.6)	(480.8, 19241.8, 192.6)
(482.6, 18457.9)	-	-	(482.6, 18457.9, 192.6)
(487.7, 18080.5)	-	-	(487.7, 18080.5, 192.6)
-	-	(490.6, 179)	(490.6, 18414, 179)
(497.6, 17252.7)	-	-	(497.6, 17252.7, 179)
(500.8, 14676)	-	-	(500.8, 14676, 192.6)
(506, 14298.6)	-	-	(506, 14298.6, 192.6)
-	-	(508.9, 178)	(508.9, 14632.1, 178)
(510.6, 13848.2)	-	-	(510.6, 13848.2, 179)
(515.8, 13470.8)	-	-	(515.8, 13470.8, 179)
-	(169.3, 62560.9)	(577.3, 169.3)	(577.3, 62560.9, 169.3)
(581.1, 11428.5)	-	-	(581.1, 11428.5, 155.8)
(590.9, 10600.7)	(373.8, 10600.7)	-	(590.9, 10600.7, 373.8)

preference to one of the objectives and some others may correspond to a good compromise of the two objectives. For illustrative purposes, we depict the corresponding paths of these six representative solutions in the network (see Figure 8). As we can see, Path 1 indicates a choice that predominantly minimizes the risk exposure without considering the route length, which results in a zigzag-like path itinerary; Path 2, on the other hand, indicates the shortest path decision that does not incorporate the risk exposure factor, the itinerary of which shows a route going through a few population-dense metropolitan areas; Path 4 may be regarded as a compromise choice between the route length and risk exposure, which chooses to go through some metropolitan areas and avoid some others. In many cases, this small set of nondominated solutions may be sufficient to convey to the decision maker a good evaluation. Such a result demonstrates the parametric method's capability of providing diverse choices in a very efficient manner for multiobjective decision making.

Subsequent solutions do increase our knowledge about the distribution of nondominated solutions, but most of them may merely represent a lower level of deviations from those representative solutions, because they are generated from some confined feasible region between previously generated nondominated solutions. After a certain number of iterations, newly generated nondominated solutions may be marginally different from some previous solutions. As shown in Figures 7 and 8, the next six solutions (i.e., the 7th–12th solutions)

generated by the parametric search process seem to be more or less analogous to their previous counterparts. For example, Path 9 is only slightly different from Path 2 on a few road segments between Newark and New York City, and Path 10 and Path 11 are both very close to Path 6, where the only difference between Path 10 and Path 6 is a short section from Massachusetts to New Hampshire and the road segments distinguishing Path 11 and Path 6 only appear from Allenton to Scranton, Pennsylvania. From a practical point of view, incorporating these new solutions may only add marginal value in providing more insights for final decision making.

3.2.3 Triobjective routing case. First, we compared the complete nondominated solution sets of the biobjective and triobjective problem instances (generated by the labeling method). In this particular study, the biobjective problem generates 39 nondominated solutions, while the triobjective problem generates 40 solutions. It is found that the biobjective case with the combination of accident rate and population size performs as a very good approximate to the original triobjective problem—34 of 40 nondominated solutions of the triobjective problem exist in the solution set of the biobjective problem, which cover all those example representative nondominated solutions we showed above.

Then, for the triobjective problem instance, we applied the decomposition scheme with the biobjective constrained parametric method for its nondominated solutions. Meanwhile, a triobjective labeling method extended from Brumbaugh-Smith and Shier's (18) is also implemented for the complete solution set. The nondominated solutions of the three biobjective subproblems are assembled in Table 2: the route length-population size subproblem (the 1st column), the accident rate-population size subproblem (the 2nd column), and the route length-accident rate subproblem (the 3rd column), which generate 27, 5, and 10 solutions, respectively. As an illustration, the nondominated solution profiles of the three biobjective subproblems are also depicted in Figure 9, which shows a close match of the search results between the constrained parametric method and the label-correcting method. A synthesis of these nondominated solutions results in a union set of 34 nondominated solutions for the triobjective problem (the 4th column in Table 2), which accounts for a solution identification rate of 85% (i.e., 34 out of 40). This rate is considerably higher than that from the aforementioned biobjective problem instance, 76.9% (i.e., 30 out of 39).

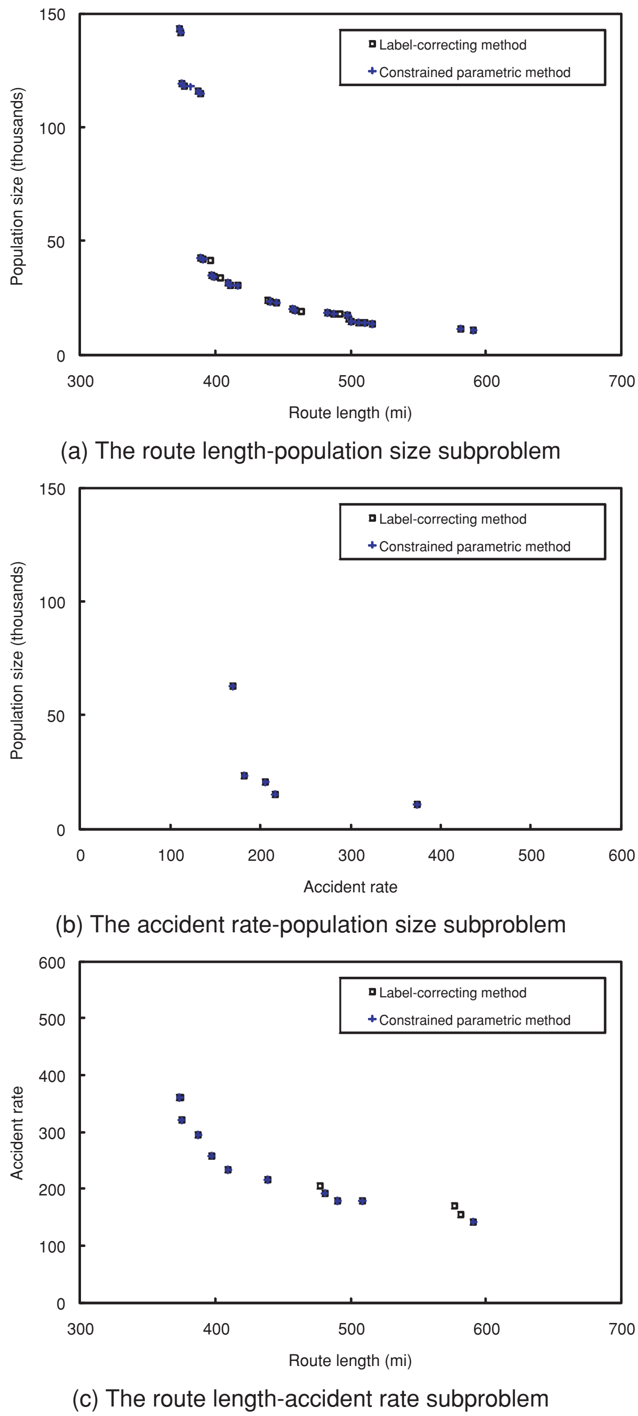


Fig. 9. Nondominated solutions of the decomposed subproblems.

4 CONCLUDING REMARKS

This article offers an alternative optimization approach for the multiobjective hazardous materials routing problem. The proposed constrained parametric method differs from existing methods in the literature: (1) compared to parametric methods, it can identify both extreme and nonextreme nondominated paths; (2) compared to labeling methods, it performs in a polynomial-time manner and prioritizes finding a set of

representative supporting solutions efficiently. This latter feature makes it especially useful in time-constrained computing environments, such as real-time, en route applications. In planning applications, it still provides attractive computational benefits, as the generation of the entire nondominated solution set is typically not required in most of real-world, large-scale problems (Current et al., 1990).

The reduced computational cost of the parametric algorithm is due to the use of an approximate label-setting procedure that solves the parameterized, doubly constrained shortest path subproblem in polynomial time. Although it may not guarantee the solution completeness, the parametric algorithm is capable of finding all extreme nondominated solutions and a large part of nonextreme nondominated solutions. Extended versions of the parametric algorithm may have greater potential in more complex multiobjective routing problems, such as problems with stochastic, time-varying attributes, alternative routing policies, and/or extra routing restrictions. In these complex cases, the combinatorial complexity may be explosively higher than their deterministic, static counterparts. The parametric optimization idea presented in this text should be further explored for efficient, practically implementable solution methods that can characterize efficient solution sets of these more challenging problems in a prompt manner.

ACKNOWLEDGMENTS

Sincere appreciation first goes to Professor Mark Turnquist of Cornell University for providing the northeastern U.S. highway network data set used in the example application. Constructive and insightful comments from seven anonymous reviewers greatly help us improve the quality and presentation of the article. Finally, the authors like to thank the editor-in-chief, Professor Hojjat Adeli, for allowing extra time in article revision.

REFERENCES

Aneja, Y. P., Aggarwal, V. & Nair, K. P. K. (1983), Shortest chain subject to side constraints, *Networks*, **13**(2), 295–302.

Boffey, B. (1995), Multi-objective routing problems, *Top*, **3**(2), 167–220.

Brumbaugh-Smith, J. & Shier, D. (1989), An empirical investigation of some bicriterion shortest path algorithms, *European Journal of Operational Research*, **43**(2), 216–24.

Carlyle, W. M., Royset, J. O. & Wood, R. K. (2008), Lagrangian relaxation and enumeration for solving constrained shortest-path problems, *Networks*, **52**(4), 256–70.

Chang, T. S., Nozick, L. K. & Turnquist, M. A. (2005), Multiobjective path finding in stochastic dynamic networks, with application to routing hazardous materials shipments, *Transportation Science*, **39**(3), 383–99.

Clímaco, J. C. N. & Martins, E. Q. V. (1982), A bicriterion shortest path algorithm, *European Journal of Operational Research*, **11**(4), 399–404.

Cohon, J. L. (1978), *Multi-Objective Programming and Planning*, Academic Press, New York, NY.

Coutinho-Rodrigues, J. M., Climaco, J. C. N. & Current, J. R. (1999), An interactive bi-objective shortest path approach: searching for unsupported nondominated solutions, *Computers and Operations Research*, **26**(6), 789–98.

Cox, R. G. (1984), Routing and scheduling of hazardous materials shipments: algorithmic approaches to managing spent nuclear fuel transport. Ph.D. thesis, School of Civil and Environmental Engineering, Cornell University, Ithaca, NY.

Current, J., & M. Marsh. (1993), Multiobjective transportation network design and routing problems: taxonomy and annotation, *European Journal of Operational Research*, **65**(1), 4–19.

Current, J. R., ReVelle, C. S. & Cohon, J. L. (1990), An interactive approach to identify the best compromise solution for two objective shortest path problems, *Computers and Operations Research*, **17**(2), 187–98.

Daellenbach, H. G. & De Kluyver, C. A. (1980), Note on multiple objective dynamic programming, *Journal of the Operational Research Society*, **31**(7), 591–94.

Dumitrescu, I. & Boland, N. (2003), Improved preprocessing, labeling and scaling algorithms for the weight-constrained shortest path problem, *Networks*, **42**(3), 135–53.

Erkut, E. (2007), Hazardous materials transportation, in C. Barnhart and G. Laporte (eds.), *Handbook in OR & MS*, vol. 14, Elsevier, Amsterdam, The Netherlands.

Erkut, E., & Verter, V. (1995), Hazardous materials logistics, in Z. Drezner (ed.), *Facility Location: A Survey of Applications and Methods*, Springer, New York, NY, pp. 467–506.

Garey, M. & Johnson, D. (1979), *Computers and Intractability: A Guide to the Theory of NP-Completeness*, Freeman, San Francisco, CA.

Handler, G. Y. & Zang, I. (1980), A dual algorithm for the constrained shortest path problem, *Networks*, **10**(4), 293–309.

Hansen, P. (1979), Bicriterion path problems, in G. Fandel and T. Gal (eds.), *Multiple Criteria Decision Making: Theory and Application, Lecture Notes in Economics and Mathematics Systems*, vol. 177, Springer-Verlag, Heidelberg, Germany, pp. 109–27.

Hassin, R. (1992), Approximation schemes for the restricted shortest path problem, *Mathematics of Operations Research*, **17**(1), 36–42.

Henig, M. I. (1985), The shortest path problem with two objective functions, *European Journal of Operational Research*, **25**(2), 281–91.

Jaffe, J. M. (1984), Algorithms for finding paths with multiple constraints, *Networks*, **14**(1), 95–116.

Korkmaz, T., Krunz, M. & Tragoudas, S. (2002), An efficient algorithm for finding a path subject to two additive constraints, *Computer Communications*, **25**(3), 225–38.

Lawler, E. L. (1976), *Combinatorial Optimization: Networks and Matroids*, Holt, Rinehart and Winston, New York, NY.

Liu, C., Hammad, A. & Itoh, Y. (1997), Multiobjective optimization of bridge deck rehabilitation using a genetic algorithm, *Computer-Aided Civil and Infrastructure Engineering*, **12**(6), 431–43.

Lorenz, D. H. & Raz, D. (2001), A simple efficient approximation scheme for the restricted shortest path problem, *Operations Research Letters*, **28**(5), 213–19.

Loui, R. (1983), Optimal paths in graphs with stochastic or multidimensional weights, *Communications of the ACM*, **26**(9), 670–76.

Lounis, Z. & Vanier, D. J. (2000), A multiobjective and stochastic system for building maintenance management, *Computer-Aided Civil and Infrastructure Engineering*, **15**(5), 320–29.

Martins, E. Q. V. (1984), On a multicriteria shortest path problem, *European Journal of Operational Research*, **16**(2), 236–45.

Miller-Hooks, E. & Mahmassani, H. S. (1998), Optimal routing of hazardous materials in stochastic, time-varying transportation networks, *Transportation Research Record*, **1645**, 143–51.

Mote, J., Murthy, I. & Olson, D. L. (1991), A parametric approach to solving bicriterion shortest path problems, *European Journal of Operational Research*, **53**(1), 81–92.

Nozick, L. K., List, G. F. & Turnquist, M. A. (1997), Integrated routing and scheduling in hazardous materials transportation, *Transportation Science*, **31**(3), 200–15.

Paya, I., Yepes, V., González-Vidosa, F. & Hospitaler, A. (2008), Multiobjective optimization of concrete frames by simulated annealing, *Computer-Aided Civil and Infrastructure Engineering*, **23**(8), 596–610.

Raith, A. & Ehrgott, M. (2009), A comparison of solution strategies for biobjective shortest path problems, *Computers and Operations Research*, **36**(4), 1299–331.

Rama Mohan Rao, A. R. & Shyju, P. P. (2010), A metaheuristic algorithm for multi-objective optimal design of hybrid laminate composite structures, *Computer-Aided Civil and Infrastructure Engineering*, **25**(3), 149–70.

Reeves, D. S. & Salama, H. F. (2000), A distributed algorithm for delay-constrained unicast routing, *IEEE/ACM Transactions on Networking*, **8**(2), 239–50.

Robbins, J. C. (1983), Routing hazardous material shipments. Ph.D. thesis, Department of Geography, Indiana University, Bloomington, IN.

Rothberg, P. F. (2001), *Hazardous Materials Transportation: Vulnerability to Terrorists, Federal Activities, and Options to Reduce Risks, CRS Report RS-21050*, National Council for Science and the Environment, Washington DC.

Rothberg, P. F., & Hassan, H. D. (2000), *Hazardous Materials Transportation Safety—Federal Program and Legislative Issues, CRS Report RS-20580*, National Council for Science and the Environment, Washington DC.

Skriver, A. J. V. & Andersen, K. A. (2000), A label correcting approach for solving bicriterion shortest-path problems, *Computers and Operations Research*, **27**(6), 507–24.

Souleyrette, R. R. & Sathisan, S. K. (1994), GIS for radioactive materials transportation, *Computer-Aided Civil and Infrastructure Engineering*, **9**(4), 294–304.

Sun, Q. & Langendörfer, H. (1998), A new distributed routing algorithm for supporting delay-sensitive applications, *Computer Communications*, **21**(6), 572–78.

Tung, T. C. & Chew, K. L. (1988), A bicriterion Pareto-optimal path algorithm, *Asia-Pacific Journal of Operational Research*, **5**(2), 166–72.

Tung, T. C. & Chew, K. L. (1992), A multicriteria Pareto-optimal path algorithm, *European Journal of Operational Research*, **6**(2), 203–09.

Turnquist, M. A. (1987), Routes, schedules and risks in transporting hazardous materials, in B. Lev *et al.* (eds.), *Strategic Planning in Energy and Natural Resources*, Elsevier Science, Amsterdam, The Netherlands, pp. 289–302.

Van Vliet, D. (1978), Improved shortest path algorithms for transport networks, *Transportation Research*, **12**(1), 7–20.

Vincke, P. (1974), Problèmes multicritères, *Cahiers du Centre D'Etudes de Recherche Opérationnelle*, **16**, 425–39.

Vitins, B. J. & Axhausen, K. W. (2009), Optimization of large transport networks using the ant colony heuristic, *Computer-Aided Civil and Infrastructure Engineering*, **24**(1), 1–14.

Warburton, A. (1987), Approximation of Pareto optima in multiple-objective, shortest-path problems, *Operations Research*, **35**(1), 70–9.

White, D. J. (1982), The set of efficient solutions for multiple objective shortest path problems, *Computers and Operations Research*, **9**(2), 101–7.

Zografos, K. G. & Davis, C. F. (1989), Multi-objective programming approach for routing hazardous materials, *Journal of Transportation Engineering*, **115**(6), 661–73.