Reformulation and Solution Algorithms for Absolute and Percentile Robust Shortest Path Problems

Xuesong Zhou, Member, IEEE and Tao Xing

Abstract—To model driver route choice behavior under inherent traffic system stochasticity, and further provide better route guidance with travel time reliability guarantees, this paper examines two models to evaluate the travel time robustness: absolute and $\alpha$-percentile robust shortest path problems. A Lagrangian relaxation approach and a scenario-based representation scheme are integrated to reformulate the minimax and percentile criteria under day-dependent random travel times. The complex problem structure is decomposed into several subproblems that can be solved efficiently as the standard shortest path problems or univariate linear programming problems. Large-scale numerical experiments with real-world data are provided to demonstrate the efficiency of the proposed algorithms.

Index Terms—Route guidance, traffic information systems, algorithms, traffic planning.

I. INTRODUCTION

TRAVEL time reliability has been widely recognized as an important element of a traveler’s route and departure time scheduling, especially for risk-averse commuters. In recent years, transportation planning and management agencies have begun to shift their focus more toward monitoring and improving the reliability of transportation systems through traffic information provisions and integrated corridor management. In addition to providing predicted travel times, the new generation of personal navigation systems needs to estimate the potential uncertainty and variability of origin-to-destination (OD) trip times, and further suggest the route that can maximize the trip time reliability or on-time performance with respect to uncertainties including inherent traffic dynamics, measurement errors and possible poor forecasts.

A wide range of definitions and formulations have been proposed to quantify travel time reliability, including (1) percentile travel time and absolute robust travel time, (2) on-time arrival probability, and (3) travel time variation expressed in terms of standard deviation or coefficient of variation. From a trip planning point of view, the first class of criteria highlights travel time guarantees over uncertain traffic situations, while the second and third definitions emphasize the probability of later arrivals for a given preferred arrival time or a given buffer time index.

Fig. 1, adapted from a recent FHWA report [1], shows a distribution of travel times on eastbound State Route 520 in Seattle, based on 3096 observation samples taken on weekdays between 4:00 to 7:00 pm. In the heavy-tailed travel time distribution along this 11.5-mile corridor, the mean and standard deviation statistics (i.e. 15.9 min and 5.5 min) are insufficient to fully measure the extreme delay during the daily commutes, where contributing factors may include traffic crashes or severe weather conditions. In particular, the worst or absolute robust travel time is about 31.5 min (during the survey period of four months), while the 95% percentile travel time is around 22.5 min.

In this study, we will focus on the absolute robust shortest path (ARSP) and percentile robust shortest path (PRSP) problems, and a finite number of link-based travel time samples (e.g., from different days) are used to describe the travel time distribution. The absolute robust shortest path problem under consideration aims to find the path that minimizes the maximum path travel time over all samples. Similarly, the a-percentile robust shortest path problem is defined as the path that minimizes the travel time within which a-percentile trips in all samples are completed. ARSP tend to emphasize the extreme tail of the travel time distribution, which might be unlikely to occur. PRSP is able to systematically balance the trade-off between the overall risk and uncertainty, and also provides a better statistical measure to avoid possible outliers in the real-world data sample set. Additionally, PRSP can meet the needs for travelers with different degrees of risk-avoidance preferences.

The concept of the percentile robust path has also attracted increasing attention recently in the area of transportation network analysis. A traffic assignment model proposed by Nie [2] considers a percentile user equilibrium where travelers follow PRSPs between each OD pair, and the PRSP problem is solved by assuming independent probability distributions of link travel times as a result of stochastic service rates. Ordonez and Stier-Moses [3] examined a user equilibrium model for risk-averse users, where the robust shortest path model proposed by Bertsimas and Sim [4] is embedded to capture the trade-off between the normal and worst cases of link travel times.

Manuscript received October 3, 2011.
X. Zhou and T. Xing are with the Department of Civil & Environmental Engineering, University of Utah, Salt Lake City, UT 84112 USA (e-mail: zhou@eng.utah.edu; tao.xing@utah.edu).
The ARSP problem and its variants have been extensively studied in the last few decades. Murty and Her [5] proposed a relaxation based label-correcting procedure to provide exact solutions for the ARSP problem. Specifically, two pruning techniques, namely one-row and Lagrangian-based relaxation, were used to improve the algorithm efficiency. Their approach was later enhanced by Bruni and Guerriero [6] by using heuristic rules and evaluation functions to better guide the solution search procedure. Yu and Yang [7] studied both ARSP and robust deviation shortest path (RDSP) problems by using a set of scenarios to capture the uncertainty of travel time. They first proved that both ARSP and RDSP problems are NP-complete under limited scenarios and NP-hard for an unbounded number of scenarios, and then proposed dynamic programming algorithms with a pseudo-polynomial computational time and a few heuristic methods. Mainly focusing on the RDSP problem, Karasan et al. [8] proposed a simple ARSP approximation algorithm by setting each link travel time to its upper bound over all scenarios/samples. Montemanni and Gambardella [9] developed two algorithms for the ARSP problem, namely a Benders decomposition-based algorithm and a solution method by generating duality reformulation and solving through mixed integer linear programming techniques.

In a recent study by Fan et al. [10], a path finding algorithm was proposed to minimize the probability of arriving at the destination later than a specified arrival time. Nie and Wu [11] developed solution algorithms with first-order stochastic dominance rules for the routing problem with a given on-time arrival reliability. Many studies [12]-[14] have been devoted to calibrating the travel time reliability measure through standard deviation or variance. There are also a number of other definitions related to robust shortest paths. For example, Yu and Yang [7] considered the robust deviation shortest path problem that minimizes the maximum deviation of the path length from the optimal path length of the corresponding scenario, and Sigal et al. [15] suggested using the probability of being the shortest path as an optimality index.

Using a sampling-based representation scheme, this research utilizes historical travel time records from multiple days of traffic measurements to capture day-by-day traffic dynamics and the complex spatial network correlations. Specifically, a scenario (corresponding to travel time samples on a day) is considered as a realization of random travel time distributions. Recognizing the complexity in on-line path computation for path reliability measures, Chen et al. [16] proposed a dynamic updating method using off-line pre-calculated candidate paths based on historical travel time samples. In this research, we focus on how to efficiently find approximate solutions for the ARSP and PRSP problems, and a Lagrangian relaxation based algorithm is used to generate satisfactory feasible solutions and provide the corresponding quality evaluation on large-scale real-world networks. In particular, we adopt a variable splitting approach to reformulate the minimax objective function of the ARSP problem and the percentile definitional constraint of the PRSP problem. The variable splitting approach was proposed by Joerntsen and Naesberg [17]. To reformulate a complex objective function, auxiliary variables and additional constraints are introduced so that easy-to-solve subproblems can be constructed in a Lagrangian relaxation solution framework. This approach was adapted by Larsson et al. [18] to solve a minimum cost network flow problem with a concave objective function, and recently by Xing and Zhou [19] to handle the nonlinear and nonadditive cost function associated with the quadratic forms of the standard deviation term in a reliable path problem. To find paths that can maximize travel time reliability on a network with normally distributed and correlated link travel times, Seshadri and Srinivasan [20] proposed bounds-based optimality conditions and an efficient path generation procedure.

The remainder of this paper is structured as follows. The next two sections provide formal problem statements, theoretical derivations and algorithmic development for both absolute and percentile robust shortest path problems. Section 4 evaluates the performance of proposed algorithms through numerical experiments on a large-scale network with real-world observation data.

II. FINDING ABSOLUTE ROBUST SHORTEST PATH

A. Problem Statement and Assumptions

Consider a directed, connected transportation network \(G(N, A)\) consisting of a set of nodes \(N\) and a set of links \(A\). In this study, we assume a set of link travel time samples or measurements is available for the same time period of \(D\) days, for example, for the peak hour from 8am to 9am over \(d = 1, \ldots, D\) weekdays during a three-month period \((D = 60)\), where \(d\) is the index of random scenarios (in a stochastic optimization framework) or the index of data collection days (from a traffic data mining perspective).

With a sufficiently large sample set, the calculated path travel time measure is able to capture the inherent spatial and temporal correlation of link travel times. Interested readers are referred to discussions by Seshadri and Srinivasan [20], and Xing and Zhou [19] on different models for representing spatial correlation for link travel times.
For notational simplicity, this study considers time-invariant travel times during the analysis time period of individual days, but the presented solution framework can be extended to a space-time expanded network by adding mapping constraints between physical links and space-time arcs, as illustrated in a recent study \[21\].

We further denote the link from node \(i\) to node \(j\) as a paired index of \(ij\), and accordingly the travel time of each link \(ij\) at the sample day \(d\) is expressed as \(c_{ij,d}\). For a given OD pair \((r, s)\), a set of binary variables \(X = \{x_{ij} | ij \in A\}\) represents the selection of links on a path (i.e. a path solution). The travel time for a path \(X\) at sample day \(d\) is then written as:

\[
T_d = \sum_{ij \in A} c_{ij,d} x_{ij}
\]

subject to a flow balance constraint

\[
\sum_{j \in A} x_{ij} - \sum_{j \in A} x_{ji} = b
\]

where \(b = \begin{cases} 1 & i = r \\ 0 & i \in N - \{r, s\} \\ -1 & i = s \end{cases}\) represents the flow status

for each node \(i\) in the network.

Given a day-dependent sample set, the robust shortest path problem aims to find a single path solution \(X\) that satisfies a certain robustness criterion over all realized samples/scenarios from different days. Specifically, two types of criteria are considered for the evaluation of the path travel time from different days. Specifically, two types of criteria are considered for the evaluation of the path travel time from different days. Specifically, two types of criteria are considered for the evaluation of the path travel time from different days. Specifically, two types of criteria are considered for the evaluation of the path travel time from different days. Specifically, two types of criteria are considered for the evaluation of the path travel time from different days.

The ARSP problem is mathematically expressed as

**Problem P0:**

\[
z_0 = \min_x \max_d \sum_{ij \in A} (c_{ij,d} x_{ij})
\]

subject to the flow balance constraint (2).

**B. Variable Splitting-based Model Reformulation**

To reformulate the proposed minimax objective function (3), a variable splitting approach is adopted in this study. Expressly, we first introduce an auxiliary variable \(y\) into the objective function,

\[
y = \max_d \sum_{ij \in A} (c_{ij,d} x_{ij})
\]

so the minimax problem is transformed to a standard minimization problem format as \(z = \min y\). The auxiliary variable \(y\) is defined as the maximum path travel time for the path \(X\) from \(r\) to \(s\) over all samples, and \(y\) also corresponds to the absolute robust travel time for a path solution \(X\).

Further, the maximization sub-problem in (4) can also be equivalently expressed as a set of inequality constraints for \(y\) over different days \(d = 1, \ldots, D\):

\[
y \geq \sum_{ij \in A} (c_{ij,d} x_{ij}), \quad \forall d
\]

Consequently, the ARSP problem \(P0\) is formulated as \(P1\)

**Problem P1:**

\[
z_1 = \min y
\]

subject to constraints (2) and (5).

To further iteratively find solutions for \(P1\) efficiently, a Lagrangian relaxation based approach is implemented. That is, we introduce a set of non-negative Lagrangian multipliers \(\mu_d\) to relax and dualize the inequality constraint set (5) into the objective function (6).

\[
\min y + \sum_d \mu_d \left( \sum_{ij \in A} (c_{ij,d} x_{ij}) - y \right)
\]

By re-grouping variables in (7), we now consider a Lagrangian problem:

**Problem L1:**

\[
L(\mu_1, \mu_2, \ldots, \mu_D)
\]

\[
= \min \sum_d \mu_d \left( \sum_{ij \in A} c_{ij,d} x_{ij} \right) + \left(1 - \sum_d \mu_d\right) y
\]

subject to constraint (2).

For any feasible (non-negative) value set of the Lagrangian multipliers \(\mu_d\), the objective function value of the Lagrangian dual problem \(L(\mu_1, \mu_2, \ldots, \mu_D)\) provides a lower bound to the optimal value \(z_1^*\) of the original problem \(P1\). By iteratively adjusting the Lagrangian multiplier set for \(L1\), we want to maximize the dual objective function in (8) and therefore improve the lower bound estimate of the primal problem. Additionally, we will use paths generated through solving the dual problem to discover better solutions, which can also reduce the upper bound to the optimal value \(z_1^*\) of the primal problem.

**C. Lagrangian Decomposition**

In the dual problem \(L1\), (8) can further be decomposed into and solved by two independent sub-problems for primal variables \(x\) and auxiliary variable \(y\), respectively.

\[
L(\mu_1, \mu_2, \ldots, \mu_D) = L_x(\mu_1, \mu_2, \ldots, \mu_D) + L_y(\mu_1, \mu_2, \ldots, \mu_D)
\]

The subproblem \(L_x(\mu_1, \mu_2, \ldots, \mu_D)\) is a binary integer programming problem for the primal variable set \(x\), and it can be solved efficiently using standard label correcting or label setting algorithms \[22\] for new link cost values of \(\sum_d \mu_d c_{ij,d}\).

**Subproblem SP1:**

\[
L_x(\mu_1, \mu_2, \ldots, \mu_D) = \min \left\{ \sum_{ij \in A} \left( \sum_d \mu_d c_{ij,d} \right) x_{ij} : \sum_{j \in A} x_{ij} - \sum_{j \in A} x_{ji} = b \right\}
\]

The second part of the dual problem in (9) is a linear
minimization problem for the single variable $y$. As a linear function, the optimal value of $L_y$ is achieved at one extreme point of the feasible range of $y$.

**Subproblem SP2:**

$$L_y(\mu_1, \mu_2, \ldots, \mu_D) = \min \left(1 - \sum_d \mu_d\right)y$$  \hspace{1cm} (11)

To find the optimal solution for subproblem SP2, we need to consider a feasible range $[y^{LB}, y^{UB}]$ for $y$ and propose a solution procedure for $L_y$:

**Proposition 1:** Depending on the value of $1 - \sum_d \mu_d$, the variable $y$ in subproblem SP2 is selected at one extreme point of its feasible range for the optimal value of $L_y$, e.g.:

$$y = \begin{cases} 
    y^{LB} & 1 - \sum_d \mu_d \geq 0 \\
    y^{UB} & 1 - \sum_d \mu_d < 0 
\end{cases}$$  \hspace{1cm} (12)

Proof: The above proposition can be easily derived as subproblem SP2 is a univariate linear program with $1 - \sum_d \mu_d$ as the cost coefficient.

By referring back to the definitional equation (4) for variable $y$, for any feasible path solution $X$ of the primal problem, it corresponds to an upper bound on the optimal objective function. For example, we can find the shortest path (with a cost function as the path distance), and use the corresponding maximum day-specific travel time as the upper bound $y^{UB}$. In the iterative search process to be presented below, the upper bound $y^{UB}$ can be updated once a new path is discovered with a lower value of the maximum day-specific travel time compared to the current $y^{UB}$.

Essentially, $y^{UB}$ should provide a lower bound to the maximum day-specific travel time on the optimal path. In this study, we first compute $c^{\min}_{ij} = \min_d \left\{c_{ij,d}\right\}$ as the least travel time of link $(ij)$ across different days, and then use $\sum_{ij \in A} c^{\min}_{ij} x_{ij}$ as the objective function to find the least travel time path and the corresponding path travel time $T^{\min}$. As $c^{\min}_{ij}$ is the least possible travel time of each link, $T^{\min} \leq z^{LB}$ for sure, for both ARSP and PRSP problems.

**D. Subgradient Method**

Let us denote $L^*$ to be the maximum value of $L(\mu_1, \mu_2, \ldots, \mu_D)$ over different Lagrangian multiplier sets:

$$L^* = \max_{\mu_1, \mu_2, \ldots, \mu_D \geq 0} L(\mu_1, \mu_2, \ldots, \mu_D)$$  \hspace{1cm} (13)

In order to find a tighter lower bound for the primal problem, we adopt a subgradient approach to iteratively search the Lagrangian multiplier set and the corresponding values of $x$ and $y$.

The search directions of $\mu$ are typically calculated as the subgradient of $L$:

$$\nabla L(\mu_1, \mu_2, \ldots, \mu_D) = \left(\sum_{\forall \mu \in A} c_{\mu} x_{\mu} - y, \sum_{\forall \mu \in A} c_{\mu} x_{\mu} - y, \ldots, \sum_{\forall \mu \in A} c_{\mu} x_{\mu} - y\right)$$  \hspace{1cm} (14)

Let us use $k$ to denote the number of iterations. Starting from any feasible initial value set, we first find solutions $x^{LB}_y$ and $y^{LB}$ for subproblems SP1 and SP2, respectively. Then the values of the Lagrangian multipliers $\mu^{k+1}_d$ at iteration $k+1$ are updated using the following subgradient equation:

$$\mu^{k+1}_d = \mu^k_d + \theta^k_d \left(\sum_{\forall \mu \in A} c_{\mu} x^{LB}_y - y^{LB}\right)$$  \hspace{1cm} (15)

where the step-size set $\theta^k_d$ can be calculated by using the following heuristic algorithm:

$$\theta^k_d = \lambda^k_d \left[z^{UB}_1 - L(\mu^k_1, \mu^k_2, \ldots, \mu^k_D)\right]$$  \hspace{1cm} (16)

In (16), $z^{UB}_1$ is the current best objective function value for feasible solutions in the primal problem and can be updated when a tighter upper bound is found. A scalar $\lambda^k_d$ chosen between 0 and 2 is used in this study to adjust the step-size $\theta^k_d$ of the search process and ensure non-negativity of Lagrangian multipliers.

**E. Solution Procedure**

The overall algorithm for solving the absolute robust shortest path problem is described below.

**Algorithm 1:**

**Step 1: Initialization**
Set iteration number $k = 0$; Choose positive values to initialize the set of Lagrangian multipliers $\mu_0$;
Select initial values for $y^{UB}$ and $y^{LB}$. The upper bounds and lower bounds of the original problem $P1$ are $z^{UB}_1 = y^{UB}$ and $z^{LB}_1 = y^{LB}$.

**Step 2: Solve decomposed dual problems**
Solve subproblem SP1 using a standard shortest path algorithm and find a path solution $x$; Solve subproblem SP2 with (12) in Proposition 1 and find a value for $y$; Calculate primal, dual and gap values, and update the upper and lower bound of $y$.

**Step 3: Update Lagrangian multipliers**
If $k > K_{\text{max}}$ or the gap is smaller than a predefined tolerance gap, terminate the algorithm, otherwise go back to Step 2. $K_{\text{max}}$ is a predetermined maximum iteration value.

**F. Solution Quality Measurement**
To measure the path solution quality, we define $\varepsilon = z^{UB} - z^{LB}$ as the duality gap between the lower bound $z^{LB}$ and the upper bound $z^{UB}$ of the optimal solution. As a result, the gap between the optimal value $z$ and the objective function value of the current best solution $z^{UB}$ is no larger than the gap $\varepsilon$. To
normalize the duality gap for comparison purposes, we can define a relative optimality measure as

$$
\varepsilon^* = \frac{z^{UB} - z^{LB}}{z^{UB}} \geq \frac{z^* - L^*}{z^*}
$$

(17)

With a reasonably small relative gap, we provide a satisfied solution quality guarantee on the suggested absolute robust path. It is important to notice that, due to the approximate nature of the Lagrangian relaxation estimator, there could still be a positive gap even if the optimal solution of the primal problem has been achieved.

The above proposed algorithm has a complexity of $O(|A|K)$ where $K$ is the number of iteration (e.g., 10-20 for our experiments in Section IV), while the complexity of Yu and Yang’s heuristic solution algorithm is $O(|A|D)$ with $D$ being the number of scenarios/days. Our algorithm is comparatively more efficient when a large number of scenarios (say $D=50$) is required to achieve a low sampling error in capturing the network travel time stochasticity and dynamics.

III. FINDING $\alpha$-PERCENTILE ROBUST SHORTEST PATH

An $\alpha$-percentile robust shortest path problem aims to minimize the $\alpha$-percentile path travel time among all feasible paths. For instance, given $\alpha = 0.9$, each feasible path solution $x$ of the given OD pair has a path reliability measure $y(x)$ corresponding to the 90th-percentile travel time. This measure ensures that, over all $D$ sample days, 90% of those days have day-specific path travel times less than $y(x)$. Among all feasible paths, the path solution $x^*$ with a minimum 90th-percentile travel time is then considered as the 90th-percentile robust shortest path. As a special case, the absolute robust shortest path problem can be viewed as the 100th-percentile robust shortest path, where the path reliability measure $y(x)$ is minimized and $y(x) = \max_d \sum_{ij \in A} (c_{ij,d} x_{ij})$

A. Problem Formulation

The $\alpha$-percentile represents the value below which $\alpha$ percent of the observations (in ascending order) may be found. To represent the $\alpha$-percentile travel time over a finite number of days $d=1,\ldots,D$, let us first denote the sorted path travel times of a path on different days as $T_1 \leq T_2 \leq \ldots \leq T_n$. The percentage $\alpha$ then corresponds to the $n^{th}$ value and $T_n$, where $n = \alpha \times D$. For example, when considering $\alpha = 0.9$ and $D = 100$ sample days, $n=90$. If $\alpha \times D$ turns out to be a floating point number, especially when the sample size $D$ is small, then the rank $n$ can be obtained by rounding to the nearest integer of the value of $\alpha \times D$.

To model the $\alpha$-percentile robust shortest path problem, we introduce the following formulation:

Problem $P2$: $z_2 = \min y$

subject to

$$
\sum_{ij \in A} (c_{ij,d} x_{ij}) - y \leq Mw_d, \quad \forall d
$$

(18)

$$
\sum_{d} w_d \leq (1-\alpha)D
$$

(19)

where $M$ is a sufficiently large number and $w_d$ is a binary variable for sample $d$.

When $w_d$ is 0, (18) reduces to

$$
y \geq \sum_{ij \in A} (c_{ij,d} x_{ij})
$$

(20)

and this active inequality should be held for all day-specific path travel times $T_\alpha$ less than the $\alpha$-percentile travel time.

When $w_d=1$, (18) leads to an always-feasible and inactive constraint

$$
\sum_{ij \in A} (c_{ij,d} x_{ij}) - y \leq M
$$

(21)

To make sure (21) is valid for those sample days $d$ on which $T_\alpha$ is greater than the final $y^*$ (i.e. the optimal $\alpha$-percentile travel time), the parameter $M$ should be sufficiently large.

To ensure the robust path travel time measure $y$ is larger than for a certain percentage of days, the variable set $w$ is constrained by (19). For example, for $\alpha = 0.9$ and $D = 100$, $\sum_d w_d \leq (1-0.9) \times 100 = 10$, so there are a total of 10 inactive constraints, and 90 active constraints. Because problem $P2$ needs to minimize the variable of $y$, the optimization result needs to select the $n = \alpha \times D$ active constraints for travel time values on different days ranking from smallest to largest.

B. Illustrative Example

Consider a single origin-destination pair with two parallel paths, as shown in Fig. 2. In this illustrative network, both paths share a common link $A$.

Table 1 shows the link and path travel times (min) over $D=4$ sample days. As calculated in this table, for the ARSP problem, path AB (along links A and B) has a minimax travel time of 12 min over four sample days. On the other hand, for a 75th-percentile robust shortest path problem, each path can only have $(1-75\%)\times 4 = 1$ day of sampled travel time greater than the variable of $y$. For path AB, only day 4 has a $w = 1$, leading to its 75th-percentile travel time as 11 min. Path AC needs to set $w = 1$ on day 3, leading to its 75th-percentile travel time as 10 min and therefore the optimal solution to the PRSP problem.

![Network of the illustrative example](image)
Table I: Travel Time Calculation of the Illustrative Example (min)

<table>
<thead>
<tr>
<th>Link A</th>
<th>Link B</th>
<th>Link C</th>
<th>Path AB</th>
<th>Path AC</th>
<th>w for Path AB</th>
<th>w for Path AC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day1</td>
<td>3</td>
<td>5</td>
<td>6</td>
<td>8</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Day2</td>
<td>4</td>
<td>7</td>
<td>6</td>
<td>11</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>Day3</td>
<td>5</td>
<td>6</td>
<td>8</td>
<td>11</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>Day4</td>
<td>4</td>
<td>8</td>
<td>6</td>
<td>12</td>
<td>1</td>
<td>10</td>
</tr>
</tbody>
</table>

100th percentile: 12 (ARSP solution) 13
75th percentile: 11 10 (PRSP solution)

C. Lagrangian Relaxation and Decomposition

Following the same Lagrangian relaxation modeling framework for the ARSP problem, we introduce a set of non-negative Lagrangian multipliers \( \mu_d \) and \( v \) to relax the inequality constraint set (18) and (19) into the objective function:

\[
\min y + \sum_{d} \mu_d \left( \sum_{j=1}^{n} c_{j,d}x_{j} - y - M w_d \right) + v \left( \sum_{d} w_d - (1-\alpha)D \right)
\] (22)

By re-grouping variables in (22), a Lagrangian dual problem is constructed with three sets of independent variables:

**Problem L2:**

\[
L(\mu_1, \ldots, \mu_D, v) = \min \sum_{i=1}^{n} \left( \sum_{d} \mu_d c_{i,d} \right) x_i
\] + \left( 1 - \sum_{d} \mu_d \right) y + \sum_{d} \left( v - M \mu_d \right) w_d - v (1-\alpha) D
\] (23)

We then divide the dual function into three independent subproblems. For notational convenience, we denote a constant variable \( h = -v (1-\alpha) D \).

\[
L(\mu_1, \ldots, \mu_D, v) = L_\mu(\mu_1, \ldots, \mu_D) + L_v(\mu_1, \ldots, \mu_D, v) + h
\] (24)

The first two subproblems in the dual function are identical to subproblems SP1 and SP2 in the ARSP problem, and both can be solved efficiently. The third part of the dual problem in (24) is a number of binary integer problems, each corresponding to a day \( d \) and a single variable \( w_d \).

**Subproblem SP3:**

\[
L_w(\mu_1, \ldots, \mu_D, v) = \left( v - M \mu_d \right) w_d, \quad \forall d
\] (25)

**Proposition 2:** For each univariate linear programming problem in subproblem SP3, the optimal value of variable \( w_d \) is determined according to the given values of \( \mu_d \) and \( v \), i.e.:

\[
w_d = \begin{cases} 
0 & v - \mu_d M \geq 0 \\
1 & v - \mu_d M < 0
\end{cases}
\] (26)

Proof: The above proposition can be easily derived as subproblem SP3 is a univariate linear program with \( \left( v - M \mu_d \right) \) as the cost coefficient, and variable \( w_d \) is bounded by an interval of \([0, 1]\).

D. Subgradient Method

Similar to the ARSP problem, we need to improve the upper and lower bounds of the primal problem by iteratively maximizing the dual problem in (23). The subgradient method is implemented here with two sets of Lagrangian multipliers \( \mu \) and \( v \).

\[
\nabla L(\mu_1, \ldots, \mu_D, v) = \left( \sum_{j=1}^{n} c_{j,d} x_j - y - M w_d \right) + v \left( \sum_{d} w_d - (1-\alpha)D \right)
\] (27)

\[
\sum_{j=1}^{n} c_{j,d} x_j - y - M w_d, \sum_{d} w_d - (1-\alpha)D
\]

\[ \mu_d^{k+1} = \mu_d^{k} + \theta_d^k \left( \sum_{j=1}^{n} c_{j,d} x_j - y - M w_d \right), \quad \forall d \]

\[ v^{k+1} = v^k + \theta_v^k \left( \sum_{d} w_d - (1-\alpha)D \right) \]

A heuristic algorithm is used to update the step-size set \( \theta_d^k \) and \( \theta_v^k \):

\[ \theta_d^k = \lambda_d^k \left[ y^{UB} - L(\mu_1^k, \ldots, \mu_D^k, v^k) \right], \quad \forall d \]

\[ \theta_v^k = \lambda_v^k \left[ y^{UB} - L(\mu_1^k, \ldots, \mu_D^k, v^k) \right] \]

E. Solution Procedure

The overall algorithm for solving the \( \alpha \)-percentile robust shortest path problem is described below.

**Algorithm 2:**

**Step 1:** Initialization
Set iteration number \( k = 0 \);
Choose positive values to initialize the set of Lagrangian multipliers, \( \mu_d \) and \( v \);
Select initial values for \( M, y^{UB} \) and \( y^{LB} \).

**Step 2:** Solve decomposed dual problems
Solve Subproblem SP1 using a standard shortest path algorithm and find a solution \( x \);
Solve Subproblem SP2 with (12) in Proposition 1 and find a value for \( y \);
Solve Subproblem SP3 with (26) in Proposition 2 and find values for \( w_d \);
Calculate primal, dual and gap values, and update the upper and lower bounds of the optimization problem \( P2 \).

**Step 3:** Update Lagrangian multipliers
Update Lagrangian multipliers with (27-31)

**Step 4:** Termination condition test
If \( k > K_{\max} \) or the gaps are smaller than the predefined toleration gap, terminate the algorithm, otherwise go back to Step 2.
F. Illustrative Numerical Examples

Now we apply the proposed Lagrangian relaxation approach in Algorithms 1 and 2 to find the ARSP and 75\textsuperscript{th}-percentile PRSP in the sample network (Fig. 2). Tables II and III show some key intermediate computational results in the first few iterations of the search procedure.

<table>
<thead>
<tr>
<th>Iteration</th>
<th>$\mu_1$</th>
<th>$\mu_2$</th>
<th>$\mu_3$</th>
<th>$\mu_4$</th>
<th>$\gamma$</th>
<th>$L_{x1}$</th>
<th>$L_{x2}$</th>
<th>$L_x$</th>
<th>$L_y$</th>
<th>$L$</th>
<th>$L_B$</th>
<th>$L_U$</th>
<th>Gap</th>
<th>Relative Gap</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>8</td>
<td>10.5</td>
<td>10.5</td>
<td>10.5</td>
<td>0</td>
<td>10.5</td>
<td>10.5</td>
<td>13</td>
<td>2.5</td>
<td>19%</td>
</tr>
<tr>
<td>2</td>
<td>0.5</td>
<td>0.75</td>
<td>1.5</td>
<td>0.75</td>
<td>13</td>
<td>37.75</td>
<td>39</td>
<td>37.75</td>
<td>-32.5</td>
<td>5.25</td>
<td>10.5</td>
<td>15</td>
<td>1.5</td>
<td>12.5%</td>
</tr>
<tr>
<td>3</td>
<td>0.01</td>
<td>0.45</td>
<td>1.2</td>
<td>0.6</td>
<td>12</td>
<td>25.43</td>
<td>26.19</td>
<td>25.43</td>
<td>-15.12</td>
<td>10.31</td>
<td>10.5</td>
<td>15</td>
<td>1.5</td>
<td>12.5%</td>
</tr>
<tr>
<td>4</td>
<td>0.01</td>
<td>0.34</td>
<td>1.09</td>
<td>0.6</td>
<td>12</td>
<td>22.96</td>
<td>23.6</td>
<td>22.96</td>
<td>-12.42</td>
<td>10.54</td>
<td>10.54</td>
<td>12</td>
<td>1.465</td>
<td>12%</td>
</tr>
<tr>
<td>5</td>
<td>0.01</td>
<td>0.25</td>
<td>0.99</td>
<td>0.6</td>
<td>12</td>
<td>21.02</td>
<td>21.58</td>
<td>21.02</td>
<td>-10.31</td>
<td>10.71</td>
<td>10.71</td>
<td>12</td>
<td>1.2892</td>
<td>11%</td>
</tr>
<tr>
<td>6</td>
<td>0.01</td>
<td>0.19</td>
<td>0.94</td>
<td>0.6</td>
<td>12</td>
<td>19.6</td>
<td>20.1</td>
<td>19.6</td>
<td>8.76</td>
<td>10.84</td>
<td>10.84</td>
<td>12</td>
<td>1.16</td>
<td>10%</td>
</tr>
<tr>
<td>7</td>
<td>0.01</td>
<td>0.14</td>
<td>0.89</td>
<td>0.6</td>
<td>12</td>
<td>18.51</td>
<td>18.95</td>
<td>18.51</td>
<td>-7.57</td>
<td>10.94</td>
<td>10.94</td>
<td>12</td>
<td>1.06</td>
<td>9%</td>
</tr>
</tbody>
</table>

Additional notations in Tables II & III:
- $L_{x1}$, $L_{x2}$: objective function values of subproblem SP1 for paths AB and AC, respectively.
- $L_x$: the cost of the shortest path found in subproblem SP1.
- $L_y$: optimal value of subproblem SP2 at each iteration.
- $L$: the value of the dual problem for each iteration.
- $L_B$: lower bound of the solution, obtained from the best dual value $L$.
- $U_B$: upper bound of the solution, generated from the best primal value among the paths uncovered up to the current search iteration.
- Gap: the difference between $U_B$ and $L_B$.
- Relative Gap: as defined in (17).

In the above two examples, by iteratively configuring weights $\mu$ on different samples, the proposed approach successfully uncovered the optimal solutions (Path AB for ARSP and Path AC for PRSP). Specifically, starting with uniform distributed values ($1/4 = 0.25$), the Lagrangian multipliers are adjusted to improve the lower bound of the optimal solution even after the optimal upper bounds have been achieved.

It should be remarked that, although the optimal solution was found in both problems, a relative gap still exists due to the approximation nature of the Lagrangian relaxation method.

In the proposed subproblem SP3, the parameter $M$ is included in the dualized objective function (25). Given its corresponding negative sign, a larger value of $M$ could lead to a lower value in the final optimal function for (25) and therefore a looser lower bound estimator. Thus, we need to select a value for parameter $M$ which is not only sufficiently large enough to make inequality (21) valid, but also small enough to construct a tight lower bound for function (25).

In the illustrative example, $M$ is selected to be 4 minutes so that it is larger than the maximum value of the gap between 100\% and 75\% travel times, which is 3 minutes for Path AC.

IV. Numerical Experiments

In this section, numerical experiments are conducted on a large-scale real-world transportation network for the Bay Area, California, which is comprised of 53,124 nodes and 93,900 links. Specifically, 8,511 links (9.1\% of links) of the entire network are freeways with a total length of 1,774.8 miles (i.e., 15.8\% of the total mileage), while 85,389 links (90.9\%) are arterial roads with a total length of 9,431.8 miles (i.e., 84.2\% of the total mileage). The algorithm is implemented in C# on the Windows Vista platform and evaluated on a personal computer with an Intel Core Duo 1.8GHz CPU and 2 GB memory.

The samples of link travel time are calculated based on available historical records from the NAVTEQ traffic database. In particular, 73 days of travel time measurements between November 2009 and February 2010 are collected for the time interval of 9:00 AM to 9:15 AM of each sample day. As the observation data used in this study (mainly from freeway segments) cover about 4.1\% of the total mileage in the Bay Area, random sample travel times are generated for links without data coverage. For simplicity, this research does not remove traffic data from weekend days and holidays. Fig.
3 shows the test network and its sensor data coverage.

As short-distance OD pairs might be covered by no or inadequate raw observations, and they typically have very limited alternative routes to examine, this study imposes the following rules to select OD pairs to be tested: (1) the average path travel time is larger than 45 minutes, and (2) the measurement coverage on the least expected travel time path is larger than 30% in distance.

As a result, an OD-pair set \( U \) containing \( u = 246 \) random OD pairs is generated from the Bay Area network. The performance of proposed algorithms are assessed using the average relative gap, which is calculated as the average value of the relative gaps for all 246 OD pairs under a predefined maximum number of iterations \( K_{\text{max}} \), e.g., 
\[
\bar{e} = \frac{\sum_{(r,s) \in U} e_{(r,s),K_{\text{max}}}}{u}.
\]

Additionally, the average objective function value of primal and dual problems among all OD pairs are also used to demonstrate the improvement of the solution quality over the iterative procedure.

As shown in Fig. 4, the average gap decreases along with the increase of the predefined maximum number of iterations \( K_{\text{max}} \). Figs. 5 and 6 illustrate that, for both models, after about 5 iterations, the reduction of upper bound becomes very slow, while the lower bound keeps improving. The average gap of the 95% percentile robust path problem is larger than that of the absolute robust path problem, which can be explained by the difference in the constructed dual objective functions, and in particular the additional complexity in turning the parameter \( M \) for subproblem \( SP3 \).
Overall, our experiments indicate that 20 iterations are sufficient for both models to achieve relatively small gap values, and the solution quality improvement begins to diminish after 10 iterations. It should be mentioned that a small duality gap can still exist even when an optimal solution is found, mainly due to the inherent limitation of Lagrangian lower bound estimation techniques.

V. EXTENSIONS AND CONCLUSIONS

As an emerging research topic on modeling driver route choice behavior and providing reliability oriented route guidance, the robust shortest path problem is studied in this paper with two modeling criteria: absolute robust shortest path and percentile robust shortest path. To reformulate the complex minimax objective function in the ARSP problem, we applied a variable splitting and relaxation technique to generate a dual problem that provides tight lower bounds for the optimal solution. Furthermore, a subgradient method is adopted in the solution procedure algorithm to iteratively improve both upper and lower bounds of the original problem. Along this line, the α-percentile robust shortest path problem is reformulated as a set of easy-to-solve subproblems by introducing auxiliary variables and additional definitional constraints. The comprehensive experiment results on a large-scale network with real-world travel time measurements demonstrated that 10-20 iterations of standard shortest path algorithms for the reformulated models can offer a very small relative duality gap of about 3-6%.

The model presented in this paper assumes day-dependent but time-invariant travel times. In our future research, one challenging task is how to consider both absolute and percentile robust shortest path problems in stochastic and time-dependent networks. To capture traffic dynamics, one needs to first build space-time expanded networks, and then establish a complex multi-stage stochastic integer programming model. In this case, a large number of nonanticipativity constraints need to be incorporated to model the first-stage path choice decisions across different scenarios. Interested readers are referred to a paper by Rockafellar and Wets [23] on the use of various Lagrangian formations to handle scenario-based stochastic integer programming problems. Recently, Yang and Zhou [21] examined some alternative reformulation schemes to relax the nonanticipativity constraints as linear objective functions, so that the underlying time-dependent shortest path problem can be solved using an efficient label-correcting algorithm by Ziliaskopoulos and Mahmassani [24] on each sample day.

Our future research directions also include the following extensions: (1) incorporate ARSP and PRSP models (for individual commuters) into the route choice component for network-wide dynamic traffic assignment and flow management problems; and (2) develop effective distributed or parallel computing techniques to improve the overall computational efficiency, as the proposed subproblems can be solved independently.

ACKNOWLEDGMENT

The authors would like to thank the California Partners for Advanced Transit and Highways (PATH) in the University of California, Berkeley and NAVTEQ Corporate for providing traffic measurement and map data of the Bay Area, CA.

REFERENCES


Tao Xing received his B.Eng. degree in communication engineering from Beijing University of Posts & Telecommunications, Beijing, China, in 2005, and his M.Sc. degree in radio frequency communication systems from University of Southampton, Southampton, UK, in 2006. He is currently pursuing the Ph.D. degree in transportation engineering with the Department of Civil and Environmental Engineering at the University of Utah. His current research interests include real-time traveler information systems, multi-criteria routing, real-time traffic estimation and prediction, and heterogeneous sensor location problems.

Xuesong Zhou (M’07) received his B.S. degree in Railroad Management from Beijing Jiaotong University, Beijing, China, in 1995, and his Ph.D. degree in Civil Engineering from the University of Maryland, College Park, in 2004. He is currently an Assistant Professor in the Department of Civil and Environmental Engineering at the University of Utah. His current research interests include analytical modeling of transportation systems, real-time dynamic traffic estimation, and large-scale routing and rail scheduling. He is the Co-Chair of the IEEE ITS Society Technical Committee on Traffic and Travel Management and serves as a Committee Member for TRB Committee on Transportation Network Modeling (ADB30).