Transportation Research Part B 96 (2017) 68–91

Contents lists available at ScienceDirect

Transportation Research Part B

journal homepage: www.elsevier.com/locate/trb

Optimizing on-time arrival probability and percentile travel time for elementary path finding in time-dependent transportation networks: Linear mixed integer programming reformulations

Lixing Yang[a,*], Xuesong Zhou[b]

[a] State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China
[b] School of Sustainable Engineering and the Built Environment, Arizona State University, Tempe, AZ 85287, USA

ARTICLE INFO

Article history:
Received 25 October 2015
Revised 19 November 2016
Accepted 21 November 2016
Available online 1 December 2016

Keywords:
Reliable path finding
On-time arrival path
Percentile path
Lagrangian relaxation

ABSTRACT

Aiming to provide a generic modeling framework for finding reliable paths in dynamic and stochastic transportation networks, this paper addresses a class of two-stage routing models through reformulation of two commonly used travel time reliability measures, namely on-time arrival probability and percentile travel time, which are much more complex to model in comparison to expected utility criteria. A sample-based representation is adopted to allow time-dependent link travel time data to be spatially and temporally correlated. A number of novel reformulation methods are introduced to establish equivalent linear integer programming models that can be easily solved. A Lagrangian decomposition approach is further developed to dualize the non-anticipatory coupling constraints across different samples and then decompose the relaxed model into a series of computationally efficient time-dependent least cost path sub-problems. Numerical experiments are implemented to demonstrate the solution quality and computational performance of the proposed approaches.

1. Introduction

1.1. Literature review

Path finding plays a fundamentally important role in the advanced traveler route guidance system, and it has received significant attention from both researchers and practitioners, in the fields of transportation science and operations research (Baldacci et al., 2011; Xing and Zhou 2013; Gounaris et al., 2013; Toriello et al., 2014; Chen and Ji 2005; Wu and Nie 2011; Orda and Rom 1990; Qian et al., 2011; Wang et al., 2016; Zhang et al., 2016). The standard shortest path problem typically finds a path starting from the origin and ending at destination with the minimized total travel time. Due to the complexity and inherent stochasticity of real-world traffic flow evolution, the actual link and path travel times in the transportation network are essentially stochastic and difficult to model in their own right. Thus, it is practically important and theoretically

* Corresponding author.
E-mail addresses: lxyang@bjtu.edu.cn (L. Yang), zhou74@asu.edu (X. Zhou).

http://dx.doi.org/10.1016/j.trb.2016.11.012
0191-2615/

challenging to study computationally efficient path finding methods under stochastic travel times that can fully capture the travelers' route choice preference under different factors of uncertainties.

Regarding travelers' route choice decisions in stochastic transportation networks, the majority of the existing studies focus on two categories: (1) expected travel time and (2) travel time reliability. In detail, the expected utility models typically use the average path travel time as an evaluation index for generating the a priori path or hyperpath between a given OD pair (see Hall 1986; Miller-Hooks 1997; Miller-Hooks and Mahmassani 2000; Gao and Chabini 2006; Huang and Gao 2012; Fan et al., 2005b), which is a most straightforward index to evaluate the performance of paths. However, in the real-world route choice applications, travelers' preferred arrival times at a destination should be taken into account, and road users also want to choose a path that minimizes not only expected travel times but also potential risk of arriving late. These criteria motivate a systematical examination of reliable path finding under a very rich set of decision making rules. Recognizing this necessity, many researchers have begun to investigate how to generate optimal and candidate paths with a more systematical consideration of travel time reliability evaluations. For instance, Nie and Wu (2009) studied the problem of finding a priori shortest paths to guarantee a given likelihood of arriving on-time. They utilized the first-order dominance rules proposed by Miller-Hooks (1997), where the travel times are typically treated as the random variables with time-variant distributions. Additionally, by focusing on a route travel time reliability measure, Samaranayake et al., (2011) designed an adaptive routing policy to maximize the probability of arriving on time at a destination with stochastic link travel times. To help travelers make their optimal route choices under travel time uncertainty, Chen et al., (2014) considered a forward and backward time-dependent reliable shortest path problem in stochastic and time-dependent road networks, and proposed two efficient solution algorithms to solve the proposed problems. Chen et al., (2010) also investigated the problem of determining the best departure time and reliable shortest path simultaneously in stochastic and time-dependent networks. In a recent study by Pan et al., (2013), the problem of finding reliable a priori shortest path aims to maximize the probability of arriving on time in a stochastic and time-dependent network on the basis of first-order stochastic dominances. Christman and Cassamano (2013) studied the problem of maximizing the probability of arriving on time in a stochastic network, and presented a dynamic programming to solve this problem. Fan et al., (2005a); Fan and Nie (2006) proposed an effective dynamic programming based optimal path algorithm for maximizing the probability of arriving on time, and they studied the convergence properties of the algorithm by introducing two special successive approximation sequences. Khani and Boyles (2015) studied the reliable path problem in the form of minimizing the sum of mean and standard deviation of path travel time. For some other related research, one can refer to Samaranayake et al., (2012); Gao et al., (2010); Shahabi et al., (2015); Andreas and Smith (2008); Andreas et al., (2008); Chen and Zhou (2010); Chen et al., al.(2016).

For the path finding process in the stochastic environment, various mathematical models have been formulated in the literature with and without considering the dynamic characteristics over the transportation networks. If no dynamic feature is considered, the problem is typically a path finding problem in the static and stochastic network. Along this line, Fan et al. (2005a), Fan and Nie (2006) proposed procedures for identifying dynamic routing policies in stochastic transportation networks, which aim to maximize the time reliability with the stationary probability density functions of the link travel times. Li et al. (2016) incorporated reliability and sustainability into a freight shortest path problem to generate efficient and environmental friendly paths, in which travel speeds on the links of the road network are modeled as discrete random variables to address uncertainty. Zockaie et al. (2013) examined how correlations in link travel times affect reliable path finding in a stochastic network, and travel times on different links are assumed to be random variables. By considering the trade-off between the risks associated with random travel time and other travel costs, Nie et al. (2012) studied optimal path problem integrated with the concept of second order stochastic dominance, which is solved by using both off-the-shelf solvers and specialized algorithms based on dynamic programming. In addition, aiming to minimize a general expected cost and maintain a reasonable arrival probability, Chen and Nie (2015) proposed a stochastic optimal path problem with relays for travelers driving with vehicles of a limited range. In their work, the travel speed on a road segment is assumed to be a discrete random variable to characterize the randomness of the road network. Note that in some congested networks (e.g., Beijing city in peak hours), the network typically has the dynamic characteristics besides of randomness. With this concern, a variety of researches focus on the dynamic and stochastic path finding problems in the literature. To effectively represent the dynamics and stochasticity of the link travel times in a real-world traffic network, different methods have been proposed in the literature, namely, the random variable with time-variant probability distributions (e.g., Miller-Hooks 1997, 2001; Nie and Wu 2009) and the sample-based time-variant link travel times (e.g., Huang and Gao 2012; Yang and Zhou 2014). In particular, the network with sample-based dynamic travel times can be viewed as having characteristics of the stochastic and time-variant networks but with complete dependency and better considerations of spatial and temporal correlations since travel times on all links at all time periods are jointly distributed. In recent decades, a variety of researchers established comprehensive frameworks for finding optimal routing policies or a set of strategies (i.e., hyperpaths) in stochastic and time-dependent networks. Interested readers can refer to Hall (1986), Miller-Hooks (1997), Miller-Hooks (2001), Nie and Wu (2009), Gao and Chabini (2006), Yang and Zhou (2014) and Chen et al. (2014).

In the stochastic networks, the most of existing literature mainly used two types of solution methods to find the favorite paths according to Miller-Hooks (2001), i.e., adaptive selection and *a priori* optimization. In the latter, the traveler is required to select a single route before departing from the origin node, in which the evaluation indexes can be either the least expected travel time (e.g., Miller-Hooks 1997) or the maximized on-time arrival probability (e.g., Nie and Wu 2009), corresponding to a series of optimization problems within the given time horizon. It should be remarked that, this *a priori* selected path cannot be updated, even though only one outcome of stochastic travel time will be finally realized. Related

body of research along this line can be found in studies by Miller-Hooks (1997), Hickman and Bernstein (1997), Sivakumar and Batta (1994), Miller-Hooks and Mahmassani (2000), Miller-Hooks and Mahmassani (2003), Xing and Zhou (2011), Huang and Gao (2012), Yang and Zhou (2014), Yang et al. (2013). In contrast, adaptive path generation with either least expected travel time or maximum on-time arrival probability can be viewed as a multi-stage recourse problem, as shown by Miller-Hooks (2001) and Fan et al. (2005a), in which one needs to select his/her route at each node after the updated travel time information is available on the adjacent links. Consequently, an optimal hyperpath with different realized probabilities can be finally produced. For this type of adaptive path generation/optimization methods, interested readers can refer to Hall (1986), Miller-Hooks (2001), Fu (2001), Yang and Miller-Hooks (2004), Fan et al. (2005a), Gao and Chabini (2006), Gao et al. (2008), Nielsen (2003), Boyles and Waller (2011), Waller and Ziliaskopoulous (2002), Yu et al. (2015), and Yang et al. (2016).

1.2. Proposed approaches

While many important contributions have been made to the reliable path finding problem for individual utility functions with time-dependent and stochastic travel times, to our knowledge, very few published studies have examined the *a priori* reliable routing problems that take into account representative decision criteria as well as capture dynamically changing travel times involving stochastic spatial and temporal dependencies. In our research, we first aim to recognize the critical needs for tackling a very challenging problem of two-stage reliable path finding under criteria different from expected utility measures with realistic link travel time representations. Equally important, our second goal is to develop a unified modeling framework and computationally efficient algorithms for this practically important application that are compatible in a large-scale regional network setting.

In this research, we adopt the representation of sample-based time-variant link travel times to capture the correlations of dynamics and randomness in a transportation network. By doing so, the measured or predicted link travel time from real-world applications can be directly embedded in the proposed modeling framework. Furthermore, scenario-specific samples enable a flexible way to construct a two-stage stochastic optimization model for specifically capturing the adaptive risk-taking behavior for extreme and complex events, compared to probability function distributions that typically only represent an aggregated version of travel time randomness.

This study aims to make the following theoretical contributions to path optimization problems in a stochastic dynamic transportation network with commonly accepted but mathematically complex risk taking criteria.

(1) Two non-expected utility criteria are formulated to measure the path travel time reliability, namely on-time arrival probability and percentile travel time. Considering the travel time data representation methods, we rigorously propose two-stage non-linear stochastic programming models for finding the most reliable paths with these different evaluation criteria. To solve the proposed models conveniently, the non-linear models are further reformulated as their equivalent integer or mixed-integer programming models by using the big-*M* methods and probability-space-time network characteristics. It is worth mentioning that, while the existing studies on reliable path findings typically focus on algorithmic developments for different decision rules rather than model formulation for standard mathematical programming, our proposed models can be successfully transformed into their equivalent linear forms that can be potentially solved by common optimization software packages or the variants of some classical algorithms in operations research fields.
(2) The second contribution of this paper is providing a highly efficient decomposition framework to solve different models in real-life transportation networks. Based on the linearized models, the hard constraints are dualized into the objective function through introducing different Lagrangian multipliers. Then, the relaxed model is further decomposed into a series of shortest path sub-problems, a univariate ratio optimization and a binary variable optimization, leading to a better utilization of underlying problem characteristics with multiple samples. The sub-gradient algorithm and branching and bounding strategy are incorporated to tighten the solution quality gaps in the searching process. The proposed approaches provide a promising algorithmic foundation for solving other closely related optimization problems involving real-world transportation networks, e.g., dynamic traffic assignment and dynamic vehicle routing with time window, which need to recognize time-dependent and stochastic travel times.

Here, we note that our study is essentially different from other reliable path models associated with path finding in dynamic and stochastic transportation networks. In the following, we mainly compare the characteristics of the proposed method with those of closely related literatures, including Nie & Wu (2009) and Chen et al. (2014). (1) The data presentation. In the existing works, they usually use the time-dependent random variables to represent the link travel times. In our paper, we use the sample-based dynamic link travel times to capture the features of the considered network. Particularly, the measured or predicted link travel times from real-world applications are consistent with our adopted data framework, which can well characterize the practical dynamics and randomness in the transportation network. (2) Modeling methods. In the works of Nie & Wu (2009) and Chen et al. (2014), they formulate this problem as non-linear mathematical models. However, with the data structure adopted in this paper, we finally prove that these models can be equivalently transformed into the linear (mixed) integer programming models, which can be potentially solved by common optimization software packages or the variants of some classical algorithms in operations research fields. (3) Solution methods. In the works of Nie & Wu (2009) and Chen et al. (2014), they use the label correcting algorithm and A* algorithm to search for the non-dominated solutions. In this paper, we give the specific measures to evaluate the quality of the solutions, and the proposed models can be easily solved by the Lagrangian relaxation based algorithm framework. Through measuring the optimality gap

between the lower and upper bounds, we can identify the performance of the solutions. Once the lower and upper bounds overlap, the exact optimal solution is finally obtained.

The rest of this paper is organized as follows. Section 2 gives a detailed description for the problem of interest. In Section 3, the formulations of the reliable path generation are stated based on two non-expected utility criteria. To solve the proposed models efficiently, Section 4 aims to investigate the equivalent models through different reformulation techniques. Section 5 describes a detailed algorithmic framework to effectively search for the optimal solution. Finally, some numerical experiments are implemented to show the performance of the proposed approaches.

2. Problem description

Consider a directed, well-connected transportation network (N, E), where N is the set of physical nodes and E is the set of physical links. The problem involves in how to select a path in this network such that the on-time arrival probability can be maximized or at least kept to a relatively high level. In the literature, one type of existing study in dynamic and stochastic networks focuses on the shortest path choice with the least expected time (LET). It should be remarked that, the LET path finding criteria might cause high risks for a single or only a few trials. In the following cases, this study will systematically examine approaches to represent reliable paths in stochastic and dynamic networks.

2.1. Data representation of link travel times

In this study, we use sample-based time-variant link travel times to characterize the dynamics and randomness of the transportation network in the stage of data preparation. In detail, the support points are used to discretize the random data (see Huang and Gao 2012), and the probability of each support point is defined as a distinctive value according to the practical situations. Let $W=\{\omega_1, \omega_2, \ldots, \omega_K\}$ represent all the considered support points, and p_{ω_k} denote the probability of support point ω_k. We then have the following relationship if we reorder the indices of support points properly,

$$p_{\omega_1} \le p_{\omega_2} \le \cdots \le p_{\omega_K} \text{ and } \sum_{k=1}^{K} p_{\omega_k} = 1$$

Typically, this form can be regarded as a generalization of the day-specified data. For instance, when a total of D days data are available, then the probability for each day can be set as 1/D. Alternatively, if we use the notation sets $\{d_1, d_2, \ldots, d_D\}$ and $\{p_{d_1}, p_{d_2}, \ldots, p_{d_D}\}$ to represent the considered days and their probabilities, then the probability of each day-specified data set is deduced by

$$p_{d_1} = p_{d_2} = \cdots = p_{d_D}$$

2.2. Reliability evaluation indexes

Next, we introduce two non-expected utility evaluation strategies associated with the stochastic path travel times. For descriptive convenience, let X denote a selected physical path in the considered network. We use $T(X, \omega)$ to represent the least path travel time on sample ω, leading to a discrete random distribution of the least path travel time. Moreover, we use $T(X)$ to denote the stochastic least travel time along path X, which is actually a random variable with realizations $T(X, \omega)$, $\omega \in W$. To avoid confusions in the following discussion, we define the following notations to represent the random variable $T(X)$, random event $\{T(X) \in B\}$ (B is a *Borel* set) and its probability $\Pr\{T(X) \in B\}$:

$$T(X) = \begin{cases} T(X, \omega_1), \text{with probability } p_{\omega_1} \\ T(X, \omega_2), \text{with probability } p_{\omega_2} \\ \cdots\cdots \\ T(X, \omega_K), \text{with probability } p_{\omega_K} \end{cases}$$

$$\{T(X) \in B\} = \{\omega \in W | T(X, \omega) \in B\},$$

$$\Pr\{T(X) \in B\} = \Pr\{\omega \in W | T(X, \omega) \in B\}.$$

With these notations, we here give two definitions to describe the reliability in the path finding process.

(1) On-time Arrival Probability (OTAP)

In general, it is desirable for a traveler to find a reliable path with the maximized on-time arrival probability. This is the case for many realistic applications. For instance, when one goes to interview before a time threshold (e.g., 8 am), he/she needs to find a reliable path to guarantee not being late due to the importance of the activity. In this sense, the recourse in path choice can be viewed as the probability that the sample-based least path travel time is less than or equal to a favorite time $\bar{T}$, calculated according to the following equation:

$$\Phi(X, \bar{T}) = \Pr\{T(X) \le \bar{T}\}. \tag{1}$$

In particular, with the sample-based data representation method, we can calculate Eq. (1) through the following process: we first rank the sample-based least travel times in an increasing sequence, denoted by $T(X, \omega_1) \le T(X, \omega_2) \le \cdots \le T(X,$

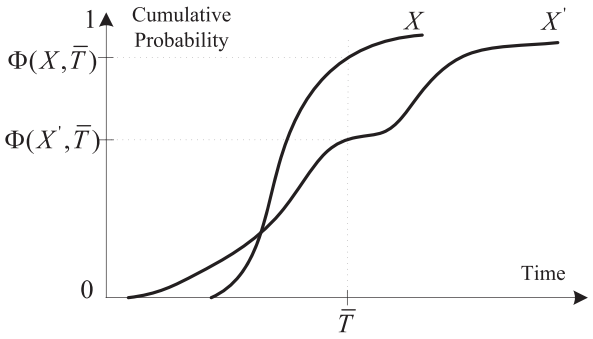

Fig. 1. An Illustration of On-time Arrival Probability.

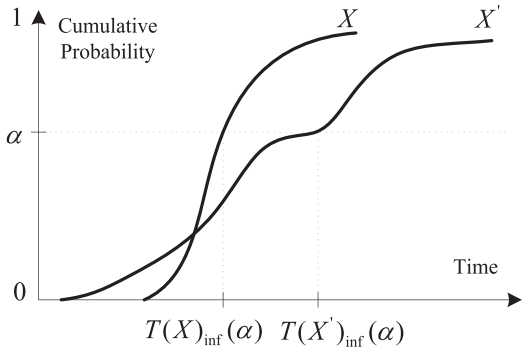

Fig. 2. An Illustration of Critical Least Travel Times.

ω_K) for simplicity; then, Eq. (1) can be calculated by

$$\Phi(X,\bar{T}) = \Pr\{T(X) \le \bar{T}\} = \Pr\{\omega \in W | T(X,\omega) \le \bar{T}\} = \sum_{k=1}^{\bar{k}} p_{\omega_k}$$

where $\bar{k}$ is the largest index of the sample whose least total travel time is not greater than $\bar{T}$, namely $\bar{k} = \max\{k | T(X,\omega_k) \le \bar{T}\}$, and p_{ω_k} is the probability of sample ω_k.

For descriptive convenience, we adopt continuous probability distribution functions to illustrate the geometric meanings of this non-expected utility index. Let X and X' be two selected paths. Fig. 1 shows probability distribution functions of their random least travel times with respected to X and X'. For the given time threshold $\bar{T}$, the corresponding probabilities are $\Phi(X,\bar{T})$ and $\Phi(X',\bar{T})$, respectively. As we have $\Phi(X,\bar{T}) \ge \Phi(X',\bar{T})$, then X corresponds to a higher chance of on-time arrival than that of X'.

(2) Percentile Travel Time (PTT)

Another concern with the reliable optimization is to choose a physical path within a given probability confidence level, resulting in the percentile reliable path choice. In this case, the recourse of the path choice will be treated as the percentile critical value of sample-based least total travel times. Typically, this method can also be regarded as a stochastic multi-objective programming, in which minimizing a critical objective function leads to the minimization of other objective functions simultaneously with a given likelihood. The percentile travel time is calculated as the critical value of the stochastic least travel time, i.e.,

$$T(X)_{\inf}(\alpha) = \inf\{r | \Pr\{T(X) \le r\} \ge \alpha\}, \tag{2}$$

where α is a pre-specified probability confidence level.

We next elaborate the calculation of the critical least travel time. Specifically, the sample-based least travel times are first ranked in an increasing sequence, denoted by

$$T(X,\omega_1) \le T(X,\omega_2) \le \cdots \le T(X,\omega_K)$$

then $T(X,\omega_{\bar{k}})$ is the α critical value to the sample-based least total travel times, where $\bar{k}$ is the smallest integer such that $\sum_{k=1}^{\bar{k}} p_{\omega_k} \ge \alpha$, i.e., $\bar{k} = \min\{k' | \sum_{k=1}^{k'} p_{\omega_k} \ge \alpha\}$

Fig. 2 illustrates the calculation of the critical least travel time. Two probability distributions of stochastic least travel time corresponding to physical paths X and X' are displayed in this figure. Typically, if we give a probability confidence level

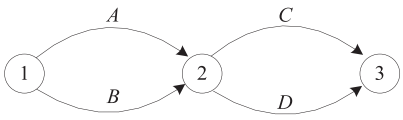

Fig. 3. A Simplified Example Network.

Table 1
Sample-based dynamic link travel times in the network (unit: min).

Timestamps	Link A				Link B				Link C				Link D			
	ω_1	ω_2	ω_3	ω_4	ω_1	ω_2	ω_3	ω_4	ω_1	ω_2	ω_3	ω_4	ω_1	ω_2	ω_3	ω_4
1	2	3	2	2	2	4	2	3	5	3	5	5	4	5	4	5
2	3	5	2	3	1	5	1	4	6	5	6	6	3	4	5	5
3	4	4	3	4	4	5	2	5	4	6	7	5	5	5	6	8
4	6	6	1	3	5	4	3	3	5	7	8	6	5	3	5	8
5	4	5	3	3	6	3	4	4	3	5	4	7	3	5	4	6
6	3	4	3	4	3	6	5	3	6	6	6	5	4	4	3	6
7	1	5	4	2	3	5	3	4	7	8	8	6	3	5	4	4

Table 2
Least path travel times (LPTT) for each sample (unit: min).

Departure Times	Path $A\rightarrow C$				Path $B\rightarrow C$				Path $A\rightarrow D$				Path $B\rightarrow D$			
	ω_1	ω_2	ω_3	ω_4	ω_1	ω_2	ω_3	ω_4	ω_1	ω_2	ω_3	ω_4	ω_1	ω_2	ω_3	ω_4
1	6	10	9	7	6	9	9	9	7	6	8	10	7	9	8	11
2	7	14	11	11	6	14	9	10	7	11	8	10	7	11	8	11
LPTT	6	10	9	7	6	9	9	9	7	6	8	10	7	9	8	11

α, the critical least total travel times have the following relationship:

$$T(X)_{\inf}(\alpha) \leq T(X')_{\inf}(\alpha)$$

If we aim to minimize the critical value, path X is obviously a better solution than X'. When the least objective is found, denoted by $T(X^*)_{\inf}(\alpha)$, we can ensure that the final outcome of the least travel time on path X^* is less than $T(X^*)_{\inf}(\alpha)$ with at least the probability confidence level α. This is the concern of the second non-expected utility criterion.

Remark 2.1. In this index, the value of parameter α reflects the risk choice preference of decision-makers. The risk seekers can set a relative small value for this parameter since it will return a reliable path with a low probability. Conversely, a risk averse decision-maker is suggested to use a larger parameter to produce the high confidence level reliable path.

2.3. An illustrative example

To illustrate the reliable path choice, a simple example will be given here to explicitly demonstrate this problem. Specifically, a network consisting of three nodes and four links is used as the experimental environment, given in Fig. 3.

Suppose that there are four sets of sample-based dynamic link travel data in this network. To characterize the temporal feature, the time horizon is discretized into six 1-min intervals, leading to 7 timestamps in total. We give the detailed time-variant link travel times in Table 1, where the probability of each sample is supposed to be 0.25.

As mentioned above, this path choice aims to produce an elementary path from the origin to destination. The selected path needs to be used as the pre-trip routing plan no matter what link travel time will finally occur. Suppose that nodes 1 and 3 are origin node and destination node, respectively, and the set of allowable departure times is assumed to be {timestamp 1, timestamp 2}. As only four paths exist on this OD pair, we can explicitly calculate their corresponding least path travel times for each sample, displayed in Table 2.

Let us take path $A\rightarrow C$ as an example. If the departure time is at timestamp 1, the total travel time should be $2 + 4 = 6$ min; otherwise, when the departure time is at timestamp 2, the total travel time should be $1 + 3 + 3 = 7$ min (waiting time at origin node, i.e., 1 min, is also included). Thus the least travel time from the origin node turns out to be 6 min. The other cases can be calculated similarly.

With the least path travel time over each sample, we then calculate each reliability index with the given parameters for these four paths (see Table 3). It is interesting to note that the most reliable path may be different for different pre-specified parameters and reliability indexes. For instance, if we set $\bar{T} = 8$ min in OTAP, the most reliable path corresponds to $A\rightarrow D$, whereas it corresponds to path $B\rightarrow C$ when we set $\alpha = 0.95$ in PTT. In general, we cannot deduce which index produces the best routing strategy in the realistic decision-making; the selection of reliability indexes is dependent on the preference of the decision-makers.

Table 3
Reliable path with different evaluation indexes and parameter settings.

Reliability index	Parameter	Path $A \rightarrow C$	Path $B \rightarrow C$	Path $A \rightarrow D$	Path $B \rightarrow D$
OTAP	$\bar{T} = 8$	0.50	0.25	0.75 (optimal)	0.50
PTT	$\alpha = 0.95$	10	9 (optimal)	10	11

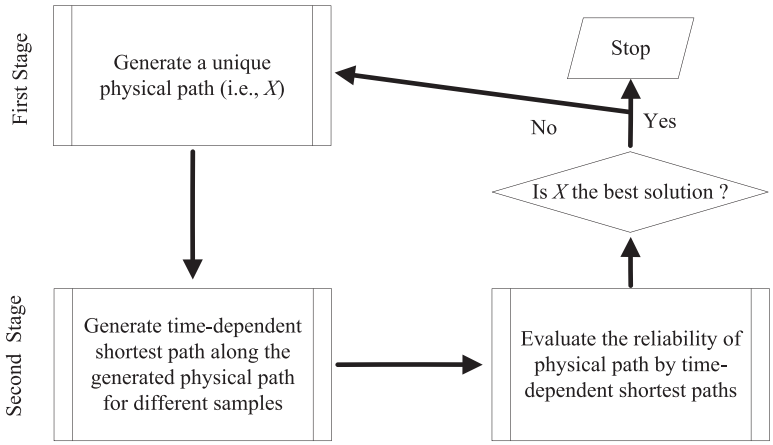

Fig. 4. Two-Stage Modeling Framework in This Study.

It is easy to see from this illustration that the path choice process involves in two specific stages. Firstly, we generate an elementary path as the potential routing plan. Then, the evaluation process will be based on the corresponding shortest path travel times over different samples (for instance, the probability that stochastic least travel time does not exceed a predetermined time threshold). For descriptive convenience, Fig. 4 gives the framework to show the main idea of modeling and solution process in this research.

As shown in Fig. 4, we formulate the problem of interest as a two-stage 0–1 stochastic integer programming problem, in which the first stage focuses on generating a unique elementary path in the network for the trip, and the second stage adaptively generates the time-dependent path choice strategies with the dynamic and stochastic link travel data. To evaluate the physical path generation, the non-expected utility reliability index of the given physical path will be calculated according to the pre-specified criteria given above. When the most reliable path is found, the search process will be terminated and it will be treated as an optimal path.

3. Mathematical formulations

In this section, two types of reliable mathematical formulations will be proposed to specify the path choice process. These models are essentially different from the traditional two-stage stochastic expected value model because of the nonlinear reliability recourse.

To describe space-time characteristics, we discretize the planning time horizon into a set of timestamps, denoted by $T=\{t_0, t_0+\delta, t_0+2\delta, \ldots, t_0+Z\delta\}$, where t_0 specifies the earliest allowable departure time from origin node R, notation δ represents the length of each time interval during which no perceptible changes of travel times take place, and Z is a sufficiently large positive integer satisfying time interval $[t_0, t_0+Z\delta]$ covers the entire planning time horizon. Associated with each physical link (i, j), a time-variant link travel time $g^{\omega}_{ijtt'} = t - t'$ is given for traveling along this link when one departs from node i at timestamp t and arrives at node j at t' over sample ω. Typically, this space-time network representation is only an approximation of the problem with the discretized time dimension, which provides an effective network framework for capturing the dynamics of the transportation network. For a given transportation network, the complexity of its space-time network is completely related to the length of discretized time intervals. A small length of each time interval necessarily corresponds to more accurate description of the network characteristics, but it potentially leads to the complexity of the problem. Thus, a trade-off between these two aspects should be taken into consideration in the stage of data preparation. (As for the space-time network representation and its various effective applications in the transportation fields, interested readers can refer to Liu and Zhou 2016; Lu et al., 2016; Mahmoudi and Zhou 2016; Yang et al., 2014, 2015, etc.)

To formulate the problem of interest, some closely related notations and parameters in the formulating process are given below for the completeness of this research:

N: set of nodes.
E: set of links.
i, j: indexes of nodes in the network, $i, j \in N$.
R: origin node.
S: destination node.
(i, j): index of link from node i to node j,$(i, j) \in E$.
ω: index of samples, $\omega \in W=\{\omega_1, \omega_2, \cdots, \omega_K\}$.
$p_{\omega:}$ probability of sample ω.
N_ω: set of time-dependent nodes on sample ω.
E_ω: set of time-dependent arcs on sample ω.
(N_ω, E_ω): space-time network on sample ω.
T: set of discrete timestamps, $T=\{t_0, t_0+\delta, \ldots, t_0+Z\delta\}$.
$\overline{T}$: pre-specified time threshold.
t, t': indexes of timestamps, $t, t' \in T$.
i_t, j_t: indexes of time-dependent nodes in the space-time network, $i_t, j_t \in N_\omega$.
$(i_t, j_{t'})$: time-dependent arc on link (i, j) from entering time t to departure time t'.
$g^{\omega}_{ijtt'}$: travel time on traffic link (i, j) from entering time t to departure time t' on sample ω.

3.1. Decision variables

As the aim of this research is to find a reliable elementary path in the transportation network, first we need to introduce the following physical link selection indicators in the process of path choice. Let x_{ij} denote the selection indicator of link (i, j)(= 1 if link(i, j) is selected,=0 otherwise). With this decision variable, it is possible to generate a physical path from the origin to destination. To evaluate the path choice, we use the corresponding time-dependent shortest path as the evaluation tool at the second stage. The time-variant arc selection indicator is denoted by $y^{\omega}_{ijtt'}$, which is the selection indicator of time-dependent arc $(i_t, j_{t'})$ over sample ω(= 1 if arc$(i_t, j_{t'})$ is selected over sample ω,=0 otherwise). Moreover, we also give the following vector representation for these two decision variables and some relevant notations. Specifically, X denotes the vector with respect to link selection indicator, which also denotes a path in the problem;Y_ω represents the vector with respect to time-dependent arc selection indicators over sample ω; Y is the vector consisting of Y_ω, $\omega \in W$; $T(X, \omega)$ is the least travel time over sample ω along physical path X; $T(X)$ is the random least travel time along physical path X.

3.2. Model constraints

In the elementary path finding model, we first give the physical flow balance constraint from origin R to destination S to guarantee the generation of a physical path, i.e.,

$$\sum_{(i,j)\in E} x_{ij} - \sum_{(j,i)\in E} x_{ji} = b_i, \forall i \in N, \tag{3}$$

where b_i denotes the flow amount at each node (i.e.,=1 for $i = R$; =− 1for $i = S$;=0 for other cases).

It is worth mentioning that a set of selected links satisfying physical flow balance constraint does not necessarily constitute an elementary path in the transportation network. Since sub-tours and loops may occur, a path can be finally generated through minimizing the total travel time, as discussed below. This constraint is rewritten as a matrix form $AX=b$ for simplicity.

According to Fig. 4, for each path generated by the physical flow balance constraint, sample-based time-dependent shortest paths will be used to evaluate each physical path choice. To demonstrate this process clearly, an illustration is given in Fig. 5 to show the time-dependent shortest path selection in the space-time network.

As displayed in Fig. 5, a simple network with three nodes and three links are considered (see left side). There are two physical paths from origin node 1 to destination node 3, i.e., $1\rightarrow 3$ and $1\rightarrow 2\rightarrow 3$. The right side of Fig. 5 gives the space-time network over one sample, where the planning time horizon is discretized into a set of timestamps by a time interval with length δ, denoted by $T=\{t_0, t_0+\delta, \ldots, t_0+7\delta\}$. In addition, the pre-specified departure time window is supposed to be $[t_0, t_0+3\delta]$. For each physical path, we find the corresponding shortest space-time path and its least travel time, which will be potentially used to evaluate the physical path, given in Table 4.

Take path $1\rightarrow 2\rightarrow 3$ as an example. Since the considered transportation network is not FIFO, the shortest space-time path corresponds to the departure time $t_0+\delta$ instead of t_0, leading to the least travel time 3δ(including the waiting time at origin node).

Table 4
Least travel time information for each path.

Path	Least travel time	Selected space–time arcs	Actual departure time
$1\rightarrow 2\rightarrow 3$	3δ	$(1_{t_0}, 1_{t_0+\delta}),(1_{t_0+\delta}, 2_{t_0+2\delta}),(2_{t_0+2\delta}, 3_{t_0+3\delta})$	$t_0+\delta$
$1\rightarrow 3$	5δ	$(1_{t_0}, 3_{t_0+5\delta})$	t_0

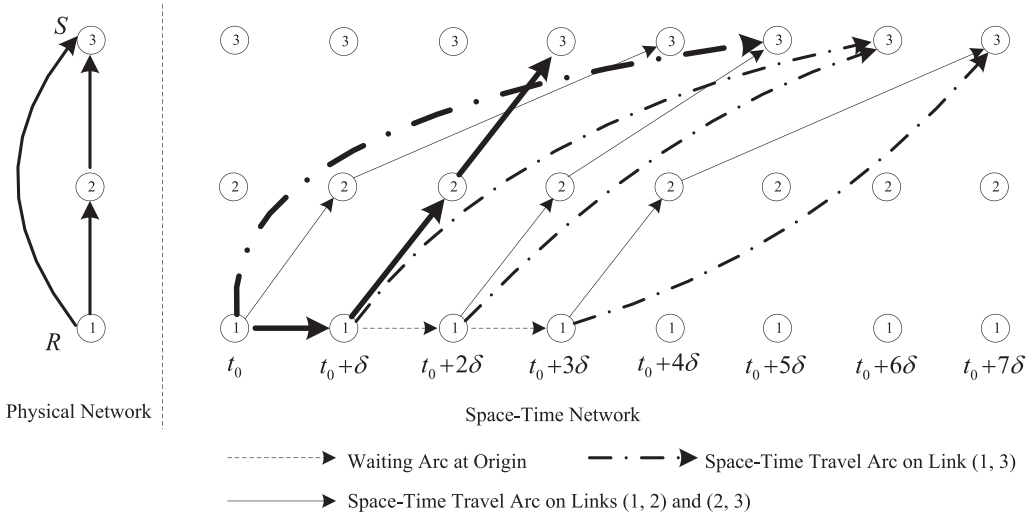

Fig. 5. An Illustration of the Shortest Space-Time Path Choice over Sample ω.

Next, we shall introduce the time-space flow balance constraint for seeking a time-dependent path over each sample ω, i.e.,

$$\sum_{(i_t, j_{t'}) \in E_\omega} y^\omega_{ijtt'} - \sum_{(j_{t'}, i_t) \in E_\omega} y^\omega_{jit't} = h^\omega_{i_t}, i_t \in N_\omega, \tag{4}$$

where $h^\omega_{i_t}$ represents the flow amount on each space-time node of (N_ω, E_ω), (i.e.,=1 for $i=R$, $t=t_0$; =− 1 for $i=S'$, $t=t_0+Z\delta$;=0 for other cases). Here, a time-dependent node $S'_{t_0+Z\delta}$ denotes the dummy destination of the trip, which only represents a trip end symbol that connects time-variant nodes S_t, $t \in \{t_0, t_0+\delta, \ldots, t_0+Z\delta\}$ without actual travel times. This constraint can be abbreviated by a matrix form $C_\omega Y_\omega = h_\omega$.

In addition, to set up the connection between the physical path and its shortest time-dependent path, we formulate the following connection constraints. That is,

$$\sum_{(i_t, j_{t'}) \in E_\omega} y^\omega_{ijtt'} = x_{ij}, (i, j) \in E, \omega \in W \tag{5}$$

In Eq. (5), the condition $(i, j) \in E$ implies all the time-dependent paths should overlap with the physical path. This constraint is simplified by $\Sigma Y_\omega = X$, $\omega \in W$ in the following discussion. In particular, if we only consider the formulation in the space-time network, this equation can also be used to substitute all the variables X in the model. In this sense, decision variables X can be regarded as the intermediate variables.

3.3. Evaluation indices and formulations

For a physical path choice X, two reliability indexes can be adopted for evaluation. We first generate a total of K time-dependent shortest paths over different samples along this physical path, leading to the following optimization problem for each sample ω at the second stage:

$$\begin{cases} \min \quad \sum_{(i_t, j_{t'}) \in E_\omega} g^\omega_{ijtt'} y^\omega_{ijtt'} \\ \text{s.t.} \quad C_\omega Y_\omega = h_\omega \\ \qquad \sum Y_\omega = X \end{cases} \tag{6}$$

Let $T(X, \omega)$, $\omega \in W$ denote the least travel times generated by model (6) over different samples. As we aim to choose the most reliable path, two non-expected utility indexes with different characteristics can be employed to evaluate the reliability of each path, given below:

(i) OTAP: $f(T(X)) = -\Pr\{T(X) \le \bar{T}\}$;
(ii) PTT: $f(T(X)) = T(X)_{\inf}(\alpha)$.

With these evaluation indexes, we formulate the model of this problem below to generate the most reliable routing plans.

$$\begin{cases} \min\ f(T(X)) \\ \text{s.t. } AX = b \\ \quad \text{where for } \forall \omega \in W, T(X,\omega) \text{ is calculated by :} \\ \quad T(X,\omega) = \min \sum\limits_{(i_t, j_{t'}) \in E_\omega} g^{\omega}_{ijtt'} y^{\omega}_{ijtt'} \\ \qquad \text{s.t. } C_\omega Y_\omega = h_\omega \\ \qquad\quad \sum Y_\omega = X \end{cases} \tag{7}$$

Remark 3.1. Typically, for each sample, if the transportation network is FIFO, there are no loops on the optimal physical paths of models with OTAP and PTT. Otherwise, loops might occur on the optimal solutions of two models. However, this is not a problem since the designed algorithm will finally return loop-free paths as the approximate optimal solution, and simultaneously as the effective upper bounds.

Remark 3.2. In the problem of interest, we use the sample-based shortest space-time paths to evaluate the generated physical path. In the second stage, we actually use the simplicity to represent the path finding process. That is, the waiting arc can be considered only at origin node, and no waiting arcs are considered at each intermediate physical node. Theoretically, this assumption is reasonable because we can aggregate the waiting time into the link travel times in the data preparation stage. For instance, if we use the reported link travel times (from a commercial traffic information service provider) to represent the sample-based dynamic link travel time data, the actual waiting time at the end of each link can be trivially added to its link travel time, which makes the problem more easy to be handled.

4. Reformulations of two-stage reliable path finding models

In this section, the proposed two-stage models will be reformulated into their equivalent linear forms for solution convenience. With the linear reformulations, it is possible to use the commercial optimization software or the variants of existing algorithms to solve approximate optimal solutions.

4.1. Equivalent nonlinear optimization models with single-stage

Following the idea of two-stage stochastic optimization models, we deduce their equivalent forms through introducing an auxiliary function, denoted by $g(X, Y_\omega) = \sum\limits_{(i_t, j_{t'}) \in E_\omega} (t' - t) y^{\omega}_{ijtt'}$, in which X represents a path selection and Y_ω represents a feasible solution of model (6) over sample ω. Then, with the similar structure of $T(X)$, we use $G(X, Y)$ to denote the random variable with a finite number of discretized values $g(X, Y_\omega)$, $\omega \in W$, i.e.,

$$G(X,Y) = \begin{cases} g(X, Y_{\omega_1}), \text{ with probability} p_{\omega_1} \\ g(X, Y_{\omega_2}), \text{ with probability} p_{\omega_2} \\ \cdots\cdots \\ g(X, Y_{\omega_K}), \text{ with probability} p_{\omega_K} \end{cases}$$

Obviously, for any given feasible solution (X, Y_ω) of model (6) over sample ω, we have the following relationship:

$$T(X,\omega) = g(X, Y'_\omega) = \min_{Y_\omega \in D(X,\omega)} \sum_{(i_t, j_{t'}) \in E_\omega} (t' - t) y^{\omega}_{ijtt'} \le g(X, Y_\omega), \omega \in W, \tag{8}$$

where $D(X, \omega)$ and Y'_ω, respectively, are the feasible domain and optimal solution of model (6) for the given X. Moreover, solution (X, Y') is typically a feasible solution to model (7). We then deduce the following propositions.

Proposition 4.1. The on-time arrival probability model (OTAP) is equivalent to the following single-stage model:

$$\begin{cases} \max\ \Pr\left\{G(X,Y) \le \bar{T}\right\} \\ \text{s.t. } AX = b \\ \quad C_\omega Y_\omega = h_\omega, \omega \in W \\ \quad \sum Y_\omega = X, \omega \in W \end{cases} \tag{9}$$

Proof. Let X^* be the optimal solution of model (OTAP). Then we have

$$\Pr\{\omega \in W | T(X^*, \omega) \le \bar{T}\} \ge \Pr\{\omega \in W | T(X, \omega) \le \bar{T}\}, \forall X. \tag{10}$$

Let Y^* be the corresponding optimal solution to model (6) over different samples along X^*. It follows that

$$\begin{aligned}\Pr\left\{G(X^*,Y^*)\le \overline{T}\right\} &= \Pr\left\{\omega\in W\ |g(X^*,Y^*_\omega)\le \overline{T}\right\}\\ &= \Pr\left\{\omega\in W\ |T(X^*,\omega)\le \overline{T}\right\}\\ &\ge \Pr\left\{\omega\in W\ |T(X,\omega)\le \overline{T}\right\}\\ &\ge \Pr\left\{\omega\in W\ |g(X,Y_\omega)\le \overline{T}\right\} \quad = \quad \Pr\left\{G(X,Y)\le \overline{T}\right\},\forall (X,Y).\end{aligned}$$

Clearly, (X^*, Y^*) is also an optimal solution to model (9). On the other hand, suppose that (X^*, Y^*) is the optimal solution to model (9), then we can also generate a feasible solution (X^*, Y') of model (OTAP) according to Eq. (8) satisfying $T(X^*,\omega)=g(X^*,Y'_\omega)\le g(X^*,Y^*_\omega),\forall\omega\in W$, which implies

$$\Pr\{\omega\in W|g(X^*,Y'_\omega)\le\bar{T}\}=\Pr\{\omega\in W|T(X^*,\omega)\le\bar{T}\}\ge\Pr\{\omega\in W|g(X^*,Y^*_\omega)\le\bar{T}\}. \tag{11}$$

Because (X^*, Y^*) is the optimal solution to model (9), we have

$$\Pr\{\omega\in W|g(X^*,Y^*_\omega)\le\bar{T}\}\ge\Pr\{\omega\in W|g(X,Y_\omega)\le\bar{T}\},\forall(X,Y). \tag{12}$$

Note that (X^*, Y') is also a feasible solution to model (9). With inequalities (11) and (12), we then draw that

$$\Pr\{\omega\in W|T(X^*,\omega)\le\bar{T}\}\}=\Pr\{\omega\in W|g(X^*,Y^*_\omega)\le\bar{T}\},$$

which implies

$$\begin{aligned}&\Pr\left\{\omega\in W|T(X^*,\omega)\le\overline{T}\right\}=Pr\left\{\omega\in W|g(X^*,Y^*_\omega)\le\overline{T}\right\}\\ &\ge Pr\left\{\omega\in W|\min_{Y_\omega\in D(X,\omega)}g(X,Y_\omega)\le\overline{T}\right\}=Pr\left\{\omega\in W|T(X,\omega)\le\overline{T}\right\},\forall X.\end{aligned}$$

Then, X^* is the optimal solution to the on-time arrival probability model (OTAP).

Proposition 4.2. Percentile travel time model (PTT) is equivalent to the following single-stage model:

$$\begin{cases}\min G(X,Y)_{\inf}(\alpha)\\ \text{s.t. } AX=b\\ \quad C_\omega Y_\omega=h_\omega,\omega\in W\\ \quad \sum Y_\omega=X,\omega\in W\end{cases} \tag{13}$$

Proof. Let X^* be the optimal solution of model (PTT). It follows that

$$T(X^*)_{\inf}(\alpha)\le T(X)_{\inf}(\alpha),\forall X. \tag{14}$$

Assume that Y^* is the corresponding optimal solution to model (6) over different samples with solution X^*. Then, by using Eq.(14), we have

$$\begin{aligned}G(X^*,Y^*)_{\inf}(\alpha) &= \inf\{r|\Pr\{\omega|g(X^*,Y^*_\omega)\le r\}\ge\alpha\}\\ &= \inf\{r|\Pr\{\omega|T(X^*,\omega)\le r\}\ge\alpha\}\\ &\le \inf\{r|\Pr\{\omega|T(X,\omega)\le r\}\ge\alpha\}\\ &\le \inf\{r|\Pr\{\omega|g(X,Y_\omega)\le r\}\ge\alpha\} \quad = \quad G(X,Y)_{\inf}(\alpha),\forall(X,Y)\end{aligned}$$

which implies that (X^*, Y^*) is the optimal solution to model (13). On the other hand, suppose (X^*, Y^*) is the optimal solution to model (13), then we can generate a feasible solution (X^*, Y') of model (PTT) according to Eq. (8) satisfying $T(X^*,\omega)=g(X^*,Y'_\omega)\le g(X^*,Y^*_\omega),\forall\omega\in W$, which implies

$$T(X^*)_{\inf}(\alpha)=G(X^*,Y')_{\inf}(\alpha)\le G(X^*,Y^*)_{\inf}(\alpha)$$

Because (X^*, Y^*) is the optimal solution to model (13), we have

$$G(X^*,Y^*)_{\inf}(\alpha)\le G(X,Y)_{\inf}(a),\forall(X,Y)$$

Note that (X^*, Y') is also a feasible solution to model (13), we then draw that

$$T(X^*)_{\inf}(\alpha)=G(X^*,Y^*)_{\inf}(\alpha)$$

Thus, we have

$$T(X^*)_{\inf}(\alpha)=G(X^*,Y^*)_{\inf}(\alpha)\le G(X,\tilde{Y})_{\inf}(\alpha)=T(X)_{\inf}(\alpha),\forall X$$

where $\tilde{Y}$ is the corresponding optimal solution to model (6) with solution X over different samples. This inequality implies that X^* is the optimal solution to the percentile travel time model (PTT).

4.2. Linearized single-stage optimization models by big-M methods

Typically, models (9) and (13) are nonlinear optimization problems because of the complexity of objective functions. To simplify the formulations, and moreover to handle the problem by commercial optimization packages or the variants of some classical algorithms, we explore the equivalent linear forms of these two formulations by introducing big-M methods in the discussion below.

Consider the following linear optimization problem, where M is a sufficiently large positive number (i.e., big-M):

$$\begin{cases} \max\ \sigma \\ \text{s.t.}\ \ g(X,Y_\omega) \le \bar{T} + z_\omega \cdot M, \omega \in W \\ \quad \sum_{\omega\in W} p_\omega z_\omega \le 1-\sigma \\ \quad \sigma \in [0,1], z_\omega \in \{0,1\}, \omega \in W \end{cases} \tag{15}$$

Proposition 4.3. The optimal objective σ^* of model (15) equals to $\Pr\{G(X,Y) \le \bar{T}\}$.

Proof: Denote $\Pr\{G(X,Y) \le \bar{T}\} = \bar{\sigma}$ and $A = \{\omega | g(X,Y_\omega) \le \bar{T}, \omega \in W\}$. Then, there exists a feasible solution $(\bar{\sigma}, z_\omega)$ for model (15) such that $z_\omega = 0$, $\omega \in A$ and $z_\omega = 1$, $\omega \in A^c$. As the objective function in model (15) is to maximize decision variable σ, it is easy to deduce $\sigma^* = \max\sigma \ge \bar{\sigma}$. Next, we prove the relationship $\sigma^* > \bar{\sigma}$ does not hold. Actually, if $\sigma^* > \bar{\sigma}$, we can necessarily find a sequence of $z^*_\omega \in \{0,1\}, \omega \in W$ such that

$$g(X,Y_\omega) \le \bar{T} + z^*_\omega \cdot M, \omega \in W,$$
$$\sum_{\omega\in W} p_\omega z^*_\omega \le 1-\sigma^*.$$

However, since $\sum_{\omega\in W} p_\omega z^*_\omega \le 1-\sigma^* < 1-\bar{\sigma} = \sum_{\omega\in A^c} p_\omega$, there exists at least one sample $\omega^* \in A^c$ satisfying $g(X,Y_{\omega^*}) \le \bar{T}$, leading to a contradiction to the definition of set A. This contradiction proves $\sigma^* = \bar{\sigma}$. The proof is thus completed.

Remark 4.1. In this linear programming model, the main focus is to find the optimal subset A of samples such that $g(X,Y_\omega) \le \bar{T}, \forall \omega \in A$, leading to the least objective value $\sum_{\omega\in A} p_\omega$. Note that if day-specific random data are used in the data representation (see Section 2), the second constraint in model (15) can be simplified as $\sum_{\omega\in W} z_\omega \le (1-\sigma)\cdot D$ since the probability for each day is 1/D.

Clearly, Proposition 4.3 shows that the objective function in model (OTAP) can be easily transformed into a linear programming-based framework. In addition, for model (PTT), let us consider the linear mixed-integer optimization problem below:

$$\begin{cases} \min\ T \\ \text{s.t.}\ g(X,Y_\omega) \le T + z_\omega \cdot M, \omega \in W \\ \quad \sum_{\omega\in W} p_\omega z_\omega \le 1-\alpha \\ \quad T \in (0,+\infty), z_\omega \in \{0,1\}, \omega \in W \end{cases} \tag{16}$$

We then have the following results.

Proposition 4.4. The optimal objective T^* of model (16) is equal to $G(X,Y)_{\inf}(\alpha)$.

Proof: Without the loss of generality, we assume $g(X,Y_{\omega_1}) \le g(X,Y_{\omega_2}) \le \cdots \le g(X,Y_{\omega_K})$. Let $\bar{k}$ be the integer satisfying $\bar{k} = \min\{k' | \sum_{k=1}^{k'} p_{\omega_k} \ge \alpha\}$. Then it is easy to deduce $G(X,Y)_{\inf}(\alpha) = g(X,Y_{\omega_{\bar{k}}})$. In this case, if we set $z_{\omega_k} = 0, \forall k \le \bar{k}$ and $z_{\omega_k} = 1, \forall k > \bar{k}$, the vector $(g(X,Y_{\omega_{\bar{k}}}), z_{\omega_1}, z_{\omega_2}, \ldots, z_{\omega_K})$ is obviously a feasible solution to model (16), which also implies $T^* \le G(X,Y)_{\inf}(\alpha) = g(X,Y_{\omega_{\bar{k}}})$. To prove this proposition, we next verify that $T^* < G(X,Y)_{\inf}(\alpha)$ does not hold through reduction to absurdity. Actually, if $T^* < G(X,Y)_{\inf}(\alpha)$, there must exist optimal solution $z^*_\omega \in \{0,1\}$, $\omega \in W$ such that

$$g(X,Y_\omega) \le T^* + z^*_\omega \cdot M, \omega \in W, \tag{17}$$

$$\sum_{\omega\in W} p_\omega z^*_\omega \le 1-\alpha. \tag{18}$$

Let $k' = \max\{k | g(X,Y_{\omega_k}) \le T^*\}$, then we have $\sum_{k\le k'} p_{\omega_k} < \alpha \le \sum_{k\le\bar{k}} p_{\omega_k}$. To keep the inequality (17), the sufficient and necessary condition turns out to be $z^*_{\omega_k} = 1, \forall k > k'$, which also leads to

$$\sum_{\omega\in W} p_\omega z^*_\omega \ge \sum_{k>k'} p_\omega z^*_\omega > 1-\alpha$$

A contradiction with Eq. (18) proves this proposition.

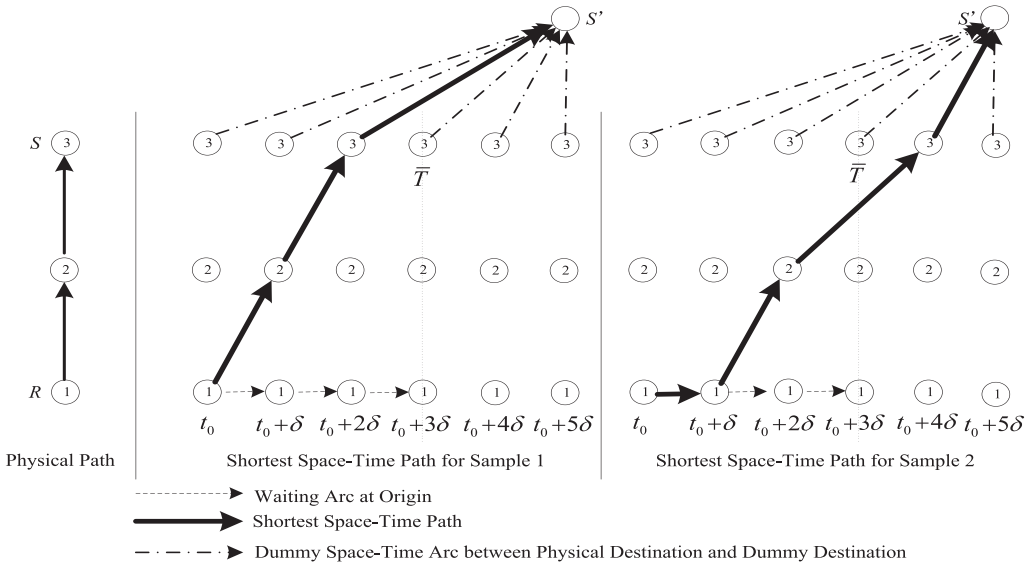

Fig. 6. Physical Path and Shortest Space-Time Paths for Different Samples.

By using propositions 4.3 and 4.4, it is possible for us to transform a non-linear programming model into a linear mixed-integer programming model. Thus, we deduce the linear single-stage equivalent models for the on-time arrival probability model (OTAP) and percentile travel time model (PTT) as follows:

$$\begin{cases} \max\ \sigma\ (\text{or}\min 1-\sigma) \\ \text{s.t.} \quad \sum\limits_{(i_t,j_{t'})\in E_\omega} (t'-t)y^{\omega}_{ijtt'} \le \bar{T} + z_\omega \cdot M, \omega \in W \\ \quad \sum\limits_{\omega\in W} p_\omega z_\omega \le 1-\sigma \\ \quad AX = b \\ \quad C_\omega Y_\omega = h_\omega, \omega \in W \\ \quad \sum Y_\omega = X, \omega \in W \\ \quad x_{ij}, y^{\omega}_{ijtt'} \in \{0,1\}, (i,j)\in E, (i_t, j_{t'}) \in E_\omega \\ \quad \sigma \in [0,1], z_\omega \in \{0,1\}, \omega \in W \end{cases} \tag{19}$$

$$\begin{cases} \min\ T \\ \text{s.t.} \quad \sum\limits_{(i_t,j_{t'})\in E_\omega} (t'-t)y^{\omega}_{ijtt'} \le T + z_\omega \cdot M, \omega \in W \\ \quad \sum\limits_{\omega\in W} p_\omega z_\omega \le 1-\alpha \\ \quad AX = b \\ \quad C_\omega Y_\omega = h_\omega, \omega \in W \\ \quad \sum Y_\omega = X, \omega \in W \\ \quad x_{ij}, y^{\omega}_{ijtt'} \in \{0,1\}, (i,j)\in E, (i_t, j_{t'}) \in E_\omega \\ \quad T \in (0,+\infty), z_\omega \in \{0,1\}, \omega \in W \end{cases} \tag{20}$$

4.3. Probability–space–time network based OTAP model

In the following discussion, we focus on deducing an equivalent formulation for the on-time arrival probability model based on the representation of the probability–space–time network on each sample (i.e., each sample corresponds to a space-time network with a pre-given probability). To this end, we first give an illustration to explain this equivalent model straightforwardly.

Take the elementary path finding as an example. Fig. 6 shows the detailed formulation process, where a path with three nodes and two links is generated. To evaluate this physical path, we adaptively select the shortest space-time paths on different probability–space–time networks over individual samples according to the aforementioned modeling methods. For

simplicity, the considered time horizon is first discretized into six timestamps with a time increment δ. A dummy destination, i.e., $S'_{t_0+5\delta}$, is used to indicate the end of the space–time trajectories in different probability–space–time networks, where a set of dummy arcs from the time-variant destination nodes to this dummy destination without actual link travel times are added in the space-time networks. Two samples are taken into account for evaluation, and the corresponding shortest space–time paths for these two samples are given with the bold arrow links. Clearly, the actual arrival time for these two samples are $t_0+2\delta$ and $t_0+4\delta$, respectively, yielding the on-time arrival probability p_{ω_1} when we consider the time threshold as $\bar{T} = t_0 + 3\delta$.

It follows from Fig. 6 that, if we minimize the number of selected dummy links whose departure times are larger than the time threshold $\bar{T}$ over all samples (here, symbol $\bar{T}$ is used to denote either a favorite travel time or a corresponding timestamp in the space-time network for notation simplicity), we can potentially obtain the corresponding maximized on-time arrival probability. Likewise, we also use $T(X)$ to represent the stochastic least path travel time along the physical path X. Considering the networks given in Fig. 6, we then have

$$Pr\left\{T(X) \le \bar{T}\right\} = \sum_{\omega\in W}\sum_{t\le\bar{T}} p_\omega y^\omega_{SS't,t_0+Z\delta}. \tag{21}$$

Alternatively, maximizing Eq. (21) is equivalent to minimizing the following equation:

$$Pr\left\{T(X) > \bar{T}\right\} = \sum_{\omega\in W}\sum_{t>\bar{T}} p_\omega y^\omega_{SS't,t_0+Z\delta}$$

We thus can transform the OTAP model into the following equivalent probability-space-time network based OTAP model (denoted by PSTOTAP) through embedding the chance constraints into the space–time networks over different samples. The detailed constraint mapping is shown below.

$$\begin{cases}\max\ \sigma \quad (\text{or } \min 1-\sigma)\\ \text{s.t.} \sum\limits_{(i_t,j_{t'})\in E_\omega}(t'-t)y^\omega_{ijtt'} \le \bar{T} + z_\omega\cdot M, \omega\in W\\ \quad \sum\limits_{\omega\in W} p_\omega z_\omega \le 1-\sigma\end{cases}$$

Chance constraints embedding into probability-space-time networks

$$\begin{cases}\min \sum\limits_{\omega\in W}\sum\limits_{t>\bar{T}} p_\omega y^\omega_{SS't,t_0+Z\delta}\\ \text{s.t. } AX = b\\ \quad C_\omega Y_\omega = h_\omega, \omega\in W\\ \quad \sum Y_\omega = X, \omega\in W\end{cases}$$

$$\begin{cases}AX = b\\ C_\omega Y_\omega = h_\omega, \omega\in W\\ \sum Y_\omega = X, \omega\in W\\ x_{ij}, y^\omega_{ijtt'}\in\{0,1\}, (i,j)\in E, (i_t,j_{t'})\in E_\omega\\ \sigma\in[0,1], z_\omega\in\{0,1\}, \omega\in W\end{cases}$$

Essentially, in the right side model, we use the dummy link selection indicators instead of total travel time to represent the on-time arrival probability. This treatment significantly simplifies the complex reliable path choice to a great extent. As for the merit of this reformulation for improving the computational efficiency, we can refer to Table 8 in Section 6.1 for the comparison of computational times with these two models.

4.4. Complexity analysis of proposed models

Next, we are particularly interested in analyzing the complexity of the proposed models. For simplicity, we use model (19) as an illustration to show the numbers of involved decision variables and system constraints. The details are shown in Table 5.

Typically, it is easy to see that the complexity of this model is closely dependent on some pre-given parameters, such as the number of samples, the number of physical links and the number of discrete timestamps over the time horizon.

Table 5
The number of involved decision variables and constraints.

Items	Total number at most
Decision variable σ	1
Decision variable z_ω	K
Decision variable x_{ij}	$\|E\|$
Decision variable $y^\omega_{ijtt'}$	$K \bullet \|E\| \bullet Z + K \bullet (Z+1)$
The first constraint in model (19)	K
The second constraint in model (19)	1
The third constraint in model (19)	$\|N\|$
The fourth constraint in model (19)	$K \bullet (\|N\| \bullet (Z+1)+1)$
The fifth constraint in model (19)	$K \bullet \|E\|$

Theoretically, in a given physical network, the large numbers of samples and discrete timestamps can well describe the inherent randomness and dynamics; however, this will necessarily increase the scale of the considered problems, potentially leading to more computational time. Thus, the compromise between these parameters and the complexity of the model should be taken into consideration in formulating the considered problem.

5. Lagrangian relaxation based approach

To design an efficient algorithmic framework for different models, we first reformulate the reliable models into separable forms, and then dualize the hard constraints into the objective function to produce the Lagrangian dual problem. Through solving Lagrangian dual problem iteratively, it is expected to generate the approximate optimal solutions to the original problem.

In the literature, Wu and Nie (2009) proved that the path X^* with the largest on-time arrival probability corresponds to the least percentile travel time when we set $\tilde{T} = T(X^*)_{\inf}(\alpha)$, and vice versa. Due to this equivalence, we only consider the reformulation of model (19) as an illustration in the following discussion, and the reformulating process of other models can be implemented with a similar method. To this end, we adopt a splitting method to decompose coupling constraints in the original model, which is similar to the approaches presented in Rockafellar and Wets (1991). Specifically, we first introduce a new physical link selection indicator variable x_{ij}^{ω} associated with different samples (i.e., $=1$ if link (i, j) is selected for sample ω; $=0$, otherwise). To keep the feasibility of the original model with the new notations, we can provide the following cycle inequality-based unique path constraint to guarantee the uniqueness of physical path choice, i.e.,

$$x_{ij}^{\omega_1} \le x_{ij}^{\omega_2}, x_{ij}^{\omega_2} \le x_{ij}^{\omega_3}, \ldots, x_{ij}^{\omega_K} \le x_{ij}^{\omega_1}, \forall (i, j) \in E.$$

It is clear that these constraints lead to a total of $|W| \bullet |E|$ nonnegative Lagrangian multipliers when we dualize this set of constraints by Lagrangian relaxation approach. In addition, we can also use the equivalent unique path constraints derived by Yang and Zhou (2014), in which some additional real-valued Lagrangian multipliers should be taken into consideration.

Next, model (19) is taken as an example to develop a generalized solution framework for the models proposed in this paper. In this model, the first two constraints and coupling constraints can be treated as the hard constraints. The first two constraints, respectively, refer to as the travel time bounding constraints and probability confidence level constraint for descriptive convenience. Three types of multipliers are introduced in dualizing these hard constraints into the objective function. That is, $\eta_\omega \ge 0$: multiplier of travel time bounding constraint, $\omega \in W$; $\rho \ge 0$: multiplier of probability confidence level constraint; $\mu_{ij}^{\omega} \ge 0$: multiplier of unique path constraint, $\omega \in W, \forall (i, j) \in E$(Here, we use cycle inequality-based unique path constraint as an illustration). Through dualizing hard constraints, we finally create a relaxed model, which can be further decomposed into different sub-problems according to its structure, given below.

Sub-problem 1: Univariate ratio optimization. The first sub-problem is associated with the optimization of variable σ, i.e.,

$$\mathrm{SP1}(\rho): \begin{cases} \min\ (1-\sigma)(1-\rho) \\ \text{s.t.} \sigma \in [0, 1] \end{cases} \tag{22}$$

This optimization model has a unique decision variable σ. For each Lagrangian multiplier ρ, we can easily produce the optimal solution and the corresponding objective by $Z^*_{\mathrm{SP1}(\rho)} = 0$ if $1-\rho \ge 0$; $Z^*_{\mathrm{SP1}(\rho)} = 1-\rho$, otherwise.

Sub-problem 2: Binary variable optimization. The second sub-problem is related to binary variable z_ω:

$$\mathrm{SP2}(\rho, \eta): \begin{cases} \min \sum_{\omega \in W} (p_\omega \rho - \eta_\omega \cdot M) z_\omega \\ \text{s.t.} z_\omega \in \{0, 1\} \end{cases} \tag{23}$$

Since only binary decision variable constraints are imposed, we can solve SP2(ρ,η) with a simple formula. That is, $z^*_\omega = 0$ if $p_\omega \rho - \eta_\omega \bullet M \ge 0$; $z^*_\omega = 1$, for other cases, where $\omega \in W$. The optimal objective of this sub-problem is calculated by:

$$Z^*_{\mathrm{SP2}(\rho,\eta)} = \sum_{\omega \in W} (p_\omega \rho - \eta_\omega \cdot M) z^*_\omega$$

Sub-problem 3: Time-dependent path optimization. Finally, note that, if a solution satisfies the space-time flow balance constraints, the mapped physical path will satisfy the physical flow balance constraints trivially. With this concern, the third sub-problem can then be derived as the following simplified form:

$$\mathrm{SP3}(\eta, \mu): \begin{cases} \min \sum_{\omega \in W} \sum_{(i_t, j_{t'}) \in E_\omega} c_{ijtt'}^{\omega} y_{ijtt'}^{\omega} \\ \text{s.t.} C_\omega Y_\omega = h_\omega, \omega \in W \end{cases} \tag{24}$$

in which we have $c_{ijtt'}^{\omega_k} = \eta_{\omega_1}(t'-t) + \mu_{ij}^{\omega_1} - \mu_{ij}^{\omega_K}$ if $k=1$, and $c_{ijtt'}^{\omega_k} = \eta_{\omega_k}(t'-t) + \mu_{ij}^{\omega_k} - \mu_{ij}^{\omega_{k-1}}$ for other cases. Then, model (24) can produce K standard time-dependent shortest path problems in their individual probability–space–time networks,

Table 6
Summary of different models.

Model index	Objective function	Model feature	Evaluation criterion	Model Type	Equivalent model
OTAP	$\Pr\{T(X) \le \bar{T}\}$	NEU model, Reliability	Probability risk	TS-SP	OS-MIP
PSTOTAP	$\Pr\{T(X) \le \bar{T}\}$	NEU model, Reliability	Probability risk	OS-IP	OS-IP
PTT	$T(X)_{\inf}(\alpha)$	NEU model, Reliability	Percentile critical value	TS-SP	OS-MIP
EVM	$E[T(X)]$	EU model	Expected value	OS-IP	OS-IP

denoted by SP3 (η, μ, ω),$\omega \in W$:

$$\mathrm{SP3}(\eta,\mu,\omega): \begin{cases} \min \sum_{(i_t,j_{t'})\in E_\omega} c^{\omega}_{ijtt'} y^{\omega}_{ijtt'} \\ \text{s.t.} C_\omega Y_\omega = h_\omega \end{cases} \tag{25}$$

For each time-dependent shortest path problem, the label correcting algorithm can be adopted to produce the optimal time-dependent path in each space-time network. For simplicity, we denote the optimal objective of SP3(η, μ, ω) as $Z^*_{\mathrm{SP3}(\eta,\mu,\omega)}$ below.

Then, the optimal objective (denoted by $L(\rho, \eta, \mu)$) of the relaxed model can be calculated by

$$L(\rho,\eta,\mu) = Z^*_{\mathrm{SP1}(\rho)} + Z^*_{\mathrm{SP2}(\rho,\eta)} + \sum_{\omega\in W} Z^*_{\mathrm{SP3}(\eta,\mu,\omega)} - \sum_{\omega\in W} \eta_\omega \bar{T}. \tag{26}$$

which is a lower bound to the optimal objective of the original problem.

The main idea behind the designed algorithm is to produce a series of candidate solutions for the original problem according to pre-specified strategies, and the best solution encountered in the searching process can be treated as the near-optimal solution for the model when the algorithm is terminated. In this study, the sub-gradient algorithm and label correcting algorithm will be integrated to produce candidate solutions for enhancing the efficiency in solving the proposed models. The algorithmic emphasis is placed on solving the following Lagrangian dual problem (27), producing the tightest lower bound (LB^*) to the primal problem simultaneously.

$$LB^* = \max_{\rho,\eta,\mu\ge 0} L(\rho,\eta,\mu). \tag{27}$$

The sub-gradient algorithm, which aims to solve the Lagrangian dual problem by iteratively updating the lower bounds, includes a point-to-point searching process. In the algorithm, the search direction at any solution is produced according to the sub-gradient directions.

Note that the Lagrangian dual problem only provides a lower bound to the primal problem due to relaxation of hard constraints. As we intend to produce an approximate optimal solution to the primal problem with the minimized duality gap, the following methods can be designed to effectively update the upper bound in the searching process. Specifically, in sub-problem SP3 (η, μ, ω), the label correcting algorithm can be adopted to find the optimal space-time path for each given set of Lagrangian multipliers and each sample ω. Thus, for every iteration of the sub-gradient algorithm, we can produce a total of K time-dependent paths from the origin to destination. Using the mapping relationship between space-time path and its physical path, we finally obtain K elementary paths in the physical networks through cancelling the potential loops, denoted by $P^n_{\omega_1}, P^n_{\omega_2}, \ldots, P^n_{\omega_K}$, where n denotes the index of the iteration in the searching process. Moreover, letting $f(P^n_\omega)$ denote the objective of the primal model, we use the following methods to update the upper bound at iteration n:

$$\min\left\{f(P^n_{\omega_1}), f(P^n_{\omega_2}), \ldots, f(P^n_{\omega_K}), UP_{n-1}\right\} \to UP_n$$

where UP_n represents the best objective encountered up to iteration n.

As discussed above, the algorithm will return a total of K physical paths from the origin to destination in each iteration. If the algorithm has $H_{\max}$ iterations at its termination, a total of $H_{\max} \bullet K$ potential solutions will be finally produced, in which the encountered best solution will be treated as the near-optimal solution of the primal problem. With these detailed techniques, the algorithmic procedure follows the standard sub-gradient algorithm. We here omit the specific description for simplicity.

Remark 5.1. This algorithm is proposed on the basis of the OTAP model. Typically, the models (PTT) and (PSTOTAP) can also be solved with a similar algorithmic framework except for the different sub-problems and sub-gradient directions in solving their corresponding Lagrangian dual problems.

Remark 5.2. In large-scale networks, the sub-gradient algorithm is probably not steady enough in searching tight lower bounds. To overcome this drawback, the branch and bound procedure can be used to distinguish different routing directions in order to generate the potential high-quality solution as soon as possible, leading to the branch and bound based sub-gradient algorithm (see Yang and Zhou 2014 for more details).

Next, to summarize the main contributions of this research, we give Table 6 to explicitly demonstrate the detailed features of different non-expected utility models and solution approaches. Yang and Zhou (2014) proposed an expected value

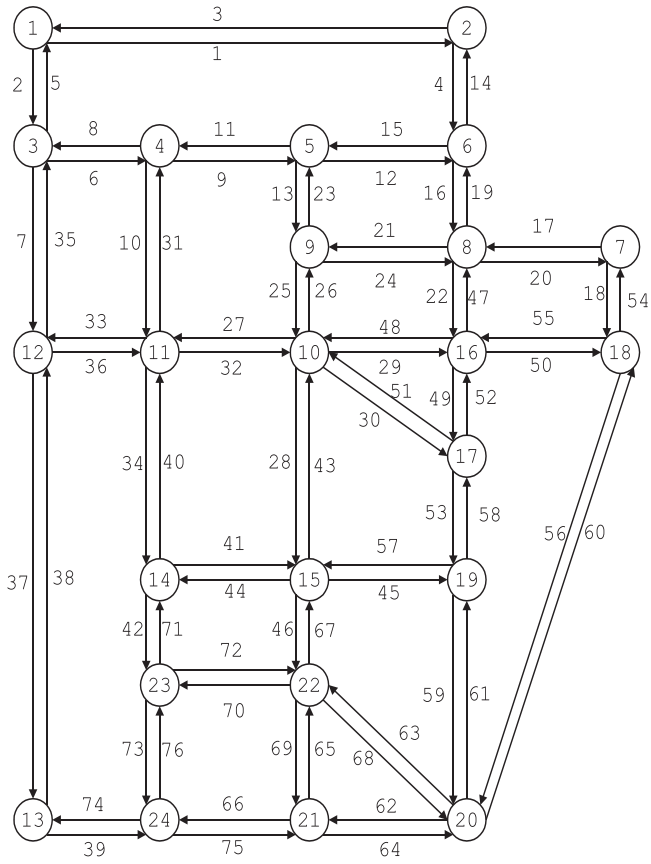

Fig. 7. Simplified Sioux Falls Network.

model (EVM) for the path generation problem in the dynamic and stochastic network. We also list the detailed characteristics of their approaches to provide an entire overview of the contributions of this research for comparison convenience.

TS-SP: two-stage stochastic programming;
OS-MIP: one-stage mixed integer programming;
OS-IP: one-stage integer programming;
EU: expected utility;
NEU: non-expected utility.

6. Numerical examples

In this section, we shall test the performance of the proposed models and algorithms through using two real-world extracted networks (i.e., Sioux Falls network and San Diego freeway network) and a grid network. All the numerical experiments are implemented on a personal computer with 2.67 GHz CPU and 4 GB memory.

6.1. Experiments in simplified sioux falls network

In this set of numerical experiments, we firstly employ the on-time arrival probability model (19) as an experimental model to solve the medium-size problem over the Sioux Falls network, given below.

As shown in Fig. 7, the experimental network consists of 24 nodes and 76 links. We use randomly generated link travel time data to present the dynamics and randomness in the real-world situations. In this set of experiments, a total of five scenarios are first taken into consideration, among which it is desirable to generate the most reliable on-time arrival path in the network with the given time threshold $\tilde{T}$. Since the model can be transformed into an equivalent linear mix-integer programming model, it can directly be solved by the commercial optimization software. With this concern, we execute this set of experiments in the GAMS optimization software with the CPLEX solver on a Windows 7.0 platform and evaluate on a personal computer with a 2.67 GHz CPU and 4 GB memory. For comparison, the proposed Lagrangian relaxation based algorithm is also encoded by Visual C++ software and implemented in the environment of Microsoft Visual Studio 2008 to show the efficiency of the proposed approaches.

Table 7
Computational times and optimal objectives of model (19) for randomly selected OD pairs.

Selected OD	Time Threshold	Opt_Obj. (CPLEX)	CTime. (CPLEX)	Opt_Obj. (LR)	CTime (s) (LR)
$1 \rightarrow 20$	28	0.4	115s	0.4	0.31
$3 \rightarrow 18$	20	0.4	175s	0.4	0.30
$3 \rightarrow 15$	15	0.4	350s	0.4	0.25
$6 \rightarrow 13$	28	0.6	79s	0.6	0.23
$7 \rightarrow 23$	27	0.8	26s	0.8	0.25
$2 \rightarrow 24$	32	0.4	102s	0.8	0.28
$8 \rightarrow 23$	21	0.6	80s	0.6	0.30
$5 \rightarrow 19$	15	0.2	262s	0.2	0.28
$4 \rightarrow 18$	19	0.6	24s	0.6	0.27
$11 \rightarrow 20$	17	0.8	14s	0.8	0.33

Table 8
Computational results for different numbers of samples.

Number of samples	Opti_Obj. in model (19) (CPLEX)	CTime. in model (19) (CPLEX)	Opti_Obj. in PSTOTAP model (CPLEX)	CTime. in PSTOTAP model (CPLEX)	Opti_Obj. in model (19) (LR)	CTime. in model (19) (LR)
8	0.63	930s	0.63	55s	0.63	0.39s
9	0.44	7409s	0.44	72s	0.44	0.45s
10	0.60	2975s	0.60	7s	0.60	0.48s
11	0.73	2954s	0.73	125s	0.73	0.59s
12	0.50	21519s	0.50	197s	0.50	0.62s

In the experiments, we randomly generate the sample-based time-dependent link travel times for each link. For simplicity, we take the equal probability for each sample in characterizing the randomness. That is, if a total of five scenarios are taken into consideration, each sample then corresponds to probability 0.2. To generate the link travel times, the set of links is first divided into two disjoint subsets E_1 and E_2 according to their physical characteristics, denoted by

$$E_1 = \{(1,2),(2,1),(12,13),(13,12),(18,20),(20,18)\}, E_2 = E/E_1$$

where E_1 and E_2 represent the sets of long links and common links, respectively. In data preparation process, a time interval with eighty minutes will be taken into consideration as the time horizon of interest. Divided by 1-min time interval, this time horizon is then discretized into a total of 80 time intervals. Then the link travel times are generated according to the following procedure:

Step 1: for each link in set E, repeat the following steps;
Step 2: for each sample ω, repeat the following steps;
Step 3: for each timestamp, if the current link belongs to E_1, randomly generate integer link travel time in interval [9, 15]; otherwise, randomly generate integer link travel time in interval [3, 7].

With these generated data, the primal model (19) has more than 30,000 binary variables, leading to a large-scale mix-integer programming model. We list the computational results for randomly selected ten OD pairs of the network in Table 7.

In the experiments, both the CPLEX solver and the Lagrangian relaxation based approach are used to solve the proposed model (19). Consequently, for each OD pair, the CPLEX solver and the proposed method can find the same optimal solutions. However, compared to the CPLEX solver, the proposed solution algorithm demonstrates it powerful efficiency in searching the optimal (or near-optimal) solutions. For instance, the computational time (CTime) for finding the optimal solutions by the Lagrangian relaxation algorithm does not exceed even one second for all the experimental cases, whereas even for the best case, the CPLEX solver still needs to consume at least 40 times more computational time than that consumed by Lagrangian relaxation approach. In the worst case, the CPLEX solver unexpectedly consumes 350 s for the OD pair $3 \rightarrow 15$ before finding the optimal solution, which illustrates the effectiveness and efficiency of the proposed Lagrangian relaxation based approaches.

In addition, to further demonstrate the computational performance of the proposed approaches, we consider five sets of link travel time data with different scales. The numbers of samples associated with these pre-specified data, respectively, are 8, 9, 10, 11 and 12. The tested OD pair is set as $4 \rightarrow 18$, and the predetermined time threshold is set as 19. In the experiments, we use both big-M based model (19) and PSTOTAP model as the experimental models to test their performances by CPLEX solver, and Lagrangian relaxation approach is still used to solve model (19). The computational results are listed in Table 8.

As expected, when more samples are considered in the numerical experiments, the computational performance of the CPLEX solver is not stable in the solution process of model (19). For instance, when we consider 8 samples in the data preparation, the computational time turns out to be 930 s. However, if the number of samples are added to 12, the computational time will be enhanced drastically to 21,519 s (above six hours), which is obviously unacceptable in some practical applications, e.g., in route guidance systems or GPS navigation. In comparison, the PSTOTAP model has the relatively better efficiency in solving the proposed experiments. As we can see, the computational time is greatly reduced in comparison to

Table 9
Computational results for randomly selected OD pairs in San Diego freeway network.

Test index	Selected OD	Time threshold	SG_LB	BBSG_LB	UB	Gap
1	1 → 84	250	2	–	2	0
2	5 → 120	250	3	–	3	0
3	8 → 48	220	1.29	2	2	0
4	20 → 107	220	6	–	6	0
5	2 → 53	300	7	–	7	0
6	44 → 97	200	2.81	3	3	0
7	88 → 32	200	1.27	1.69	2	0.31
8	124 → 41	200	8	–	8	0
9	11 → 44	190	6.79	6.83	7	0.17
10	48 → 5	210	4	–	4	0

that of model (19). All the computational times for these numerical experiments are not greater than 200 s, which shows the merits of the PSTOTAP model for improving the computational efficiency. In addition, the proposed Lagrangian relaxation-based approach demonstrates its stable performance with respect to both solution quality and computational time, in which the corresponding computational time is not greater than 1 s to generate the same optimal objectives to those found by the CPLEX solver. Typically, the CPLEX solver is less efficient in solving these reliable models because of the existence of a large number of binary variables.

6.2. Experiments with large-scale network and real-world detected data

To further test the computational efficiency of the proposed approaches, more experiments were performed on a part of the real-world freeway network near San Diego of California region with the recorded time-dependent link travel time data obtained by sensors. This network consists of 127 nodes and 284 links, in which most of links are equipped with sensors for collecting the real-time link travel times over different days.

(I) Preparation of Input Data

In the experiments, a set of ten-day time-variant link travel times obtained by detectors is used to search for the reliable a priori elementary paths. Each day corresponds to one sample with the even probability 0.1. In preparing the link travel times, with a 0.1-min travel time aggregation time interval, we take a one-hour planning time horizon from 15:00 to 16:00 into consideration, leading to a total number 600 of time intervals. All these inputted time-dependent link travel times are converted from speed readings of California's performance measurement system. Additionally, to characterize the impact of non-recurring delays, we in particular simulate the link travel times on each link over different samples with a low probability. We assume that the incident probability on each link is 0.1. Then, in preparing the input data for each day and link, we first randomly generate a number r in interval [0, 1]. If $r < 0.1$ (this case indicates the occurrence of the incident), then the time-dependent link travel times will be set as 2 or 3 times of the observed travel times.

(II) Preparation of tested model

In this set of experiments, we use the probability-space-time network based on-time arrival probability model (PSTOTAP) to test the efficiency and effectiveness of the proposed solution methods. Note that we adopt the day-specific (i.e., 10 samples) link travel times as the experimental data, then the probability of each sample will be set as $p_\omega = 0.1$ for any $\omega \in W$. With this concern, the impacts of sample probability on the objective function can be safely omitted. Then, the following objective function is used in the experimental process, i.e.,

$$\min \sum_{\omega} \sum_{t > \bar{T}} y^{\omega}_{SS't,t_0+Z\delta}$$

Clearly, this objective is in essence equivalent to the original one in the model (PSTOTAP) as the sample probabilities are supposed to be equal to each other. On the other hand, it is obvious that the optimal objective with this form should be an integer in set{0, 1, …, $|W|$}.

(III) Experimental results

Next, we test the proposed approaches for generating reliable elementary shortest paths in the considered network. A total of ten OD pairs are randomly selected from the node-link network structure. Both the standard sub-gradient (SG) and the branch and bound based sub-gradient (BBSG) are employed in the solution process. In detail, if the SG does not find a tight-enough lower bound, we shall use the branch and bound based solution framework to divide the entire search space into different disjoint solution sub-spaces, and then the SG will be used to find the lower bound on each branch. Using the bound updating strategy, the generated lower bound can be expectedly improved to a satisfactory degree. In the SG designed in searching process, the allowable maximum iteration is set as 30, and the allowable minimum gap between the upper and lower bound is set as 0.001. Thus, when either of these two conditions is violated, the searching process will be terminated. Additionally, we also avoid the potential cycle of the physical path by only searching one direction of physical links in the label correction algorithm. In the experiments, the favorite time threshold can be represented by the number of intervals. The detailed computation results are given in Table 9.

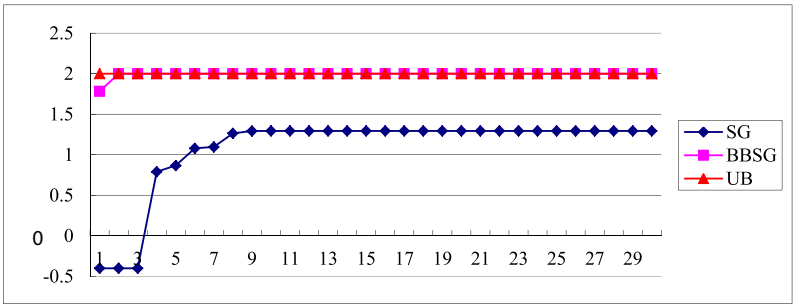

Fig. 8. Convergence Process in Two Algorithms.

As addressed above, we first adopt the standard SG to implement the numerical experiments for each selected OD pair. It is clear that, for some cases, this algorithm can produce the exact optimal solution and optimal objective value because the lower and upper bounds can finally overlap, which indicates the effectiveness of the standard sub-gradient in searching reliable shortest paths in the given network. This situation occurs respectively for experiments 1, 2, 4, 5, 8 and 10. However, for other experiments (i.e., experiments 3, 6, 7 and 9), we note that the Lagrangian relaxation based method does not directly produce a tight-enough lower bound for the original problem due to the unsteady characteristics of the sub-gradient algorithm itself. Then, we adopt branch and bound procedure to decompose the searching space in order to produce the high-quality solution and much tighter lower bounds.

For experiments 3, 6, 7 and 9, we further produce two branches according to the possibly selected links in the route generation process on the basis of the network structure. On each branch, we still use the SG to generate the lower and upper bounds. Through updating the encountered bounds by bounding policy, we find a much tighter lower bound than that generated by the SG. For instance, the SG algorithm returns a lower bound 1.29 in experiment 3, leading to a gap 0.71 from the obtained upper bound. To reduce the absolute gap, we use the branch and bound algorithmic framework to decompose the searching space into two sub-spaces, leading to two branches in the second searching level. In detail, since either link (8, 9) or link (8, 29) needs to be included in the final reliable path according to the structure of the network, two branches can be produced once we fix either of these two links in the searching process. As expected, through imposing sub-gradient algorithm on each branch, we can finally produce the most favorite lower bound that overlaps with the upper bound, producing the exact optimal solution.

To state the convergence speed of SG and BBSG, Fig. 8 shows the performance of these two algorithms in experiment 3. In this figure, the convergence process of lower bounds produced by the standard SG is fairly slow. As shown, the lower bound keeps in a relative lower level for the first several iterations. Until up to iteration 11, the encountered lower bound reaches its largest value and keeps fixed afterwards. In comparison, the BBSG demonstrates it powerful efficiency in searching the tight lower bound. Only at iteration 2, the optimal solution can be found. This is because the decomposition of solution space greatly simplifies the complexity of the original problem, which makes the algorithm find the favorite lower bound quickly.

In addition, it follows from experimental results that for all of these randomly selected OD pairs, we actually find their exact optimal objectives (i.e., the returned upper bounds) at the termination of numerical experiments, because the optimal objective function takes integer values. For instance, note that the largest absolute gap among these ten cases occurs in experiment 7 (i.e., 0.31) with the lower bound 1.69, we then deduce that the least objective should be 2.00 since it takes an integer value, which just coincides with the returned upper bound.

In this set of experiments, the computational time is closely related to the total iterations in the solution process. In each iteration, the average computational time is no more than 20 s, which consist of using the label correcting algorithm to solve the different sample-based time-variant shortest paths with the generalized link costs. Since we consider ten samples in these experiments, the sub-gradient algorithm will necessarily produce ten shortest paths in each iteration. That is to say, a single implementation of the label correcting algorithm consumes no more than two seconds in the solution process.

6.3. More experiments in a large-scale grid network

In the following, we will implement a series of numerical experiments in a randomly generated large-scale grid network to further test the performance of the proposed approaches, which consists of 900 nodes and 1205 links. In detail, we give a clear structure of the considered grid network in Fig. 9, where the bold dotted links are randomly generated with a given probability. In this network, we also considered a time horizon with one hour, which is discretized into a total of 600 timestamps by a 0.1-min time interval. The time-dependent link travel times are randomly generated in set {0.5, 0.6, 0.7, 0.8, 0.9, 1.0} (unit: min). In addition, we also consider an incident probability 0.2 on each sample in preparing the link travel times. That is, if an incident occurs, the above generated time-dependent link travel times will be multiplied by a random number in {1, 2} over the corresponding sample.

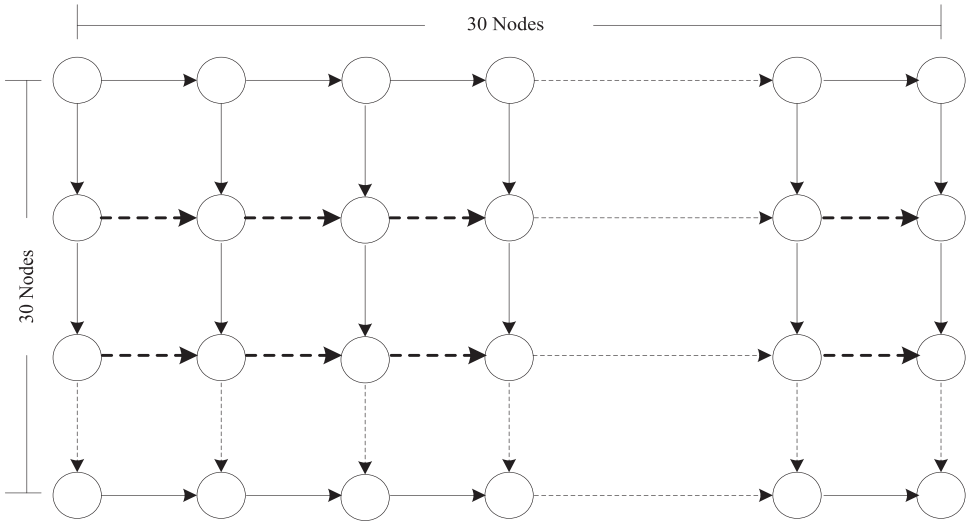

Fig. 9. Structure of the Considered Grid Network.

Table 10
Computational results for randomly selected OD pairs in grid network with 20 samples.

Test index	Lower bound	Upper bound	Returned gap	Returned relative gap	Actual relative gGap (%)
1	17.99	18	0.01	0.04%	0.00
2	15.23	16	0.77	3.87%	0.00
3	16.00	16	0.00	0.00%	0.00
4	14.70	16	1.30	6.48%	5.00
5	15.40	17	1.60	7.98%	5.00
6	11.06	12	0.94	4.68%	0.00
7	8.50	10	1.50	7.50%	5.00
8	10.00	10	0.00	0.00%	0.00
9	12.70	13	0.30	1.48%	0.00
10	18.54	20	1.46	7.29%	5.00
11	18.94	20	1.06	5.29%	5.00
12	15.64	17	1.36	6.80%	5.00
13	6.00	6	0.00	0.00%	0.00
14	17.29	18	0.71	3.54%	0.00
15	0.50	2	1.50	7.51%	5.00
16	6.45	8	1.55	7.74%	5.00
17	16.00	16	0.00	0.00%	0.00
18	11.80	12	0.20	0.98%	0.00
19	17.00	17	0.00	0.00%	0.00
20	16.32	18	1.68	8.38%	5.00

In this network, we first test the performance of the algorithm in producing the near-optimal solutions for some randomly generated OD pairs. To this end, we particularly impose the following rules to select OD pairs to be tested and the time thresholds: (1) the total number of involved links between each selected OD pair should be around or more than 30; (2) the time threshold in each implementation is randomly chosen such that the fully on-time arrival situation on the given OD pair cannot occur. In these experiments, we implement a total of 20 instances with either SG or BBSG, in which a total of 20 samples are adopted to capture of the randomness of the network. The computational results are listed in Table 10.

In these experiments, we particularly list the returned gaps, returned relative gaps and actual relative gaps for all the tests. In particular, we define these indexes according to the following equations:

Return Gap = Upper Bound – Lower Bound

$$\text{Returned Relative Gap} = \frac{\text{Upper Bound} - \text{Lower Bound}}{\text{Number of Samples}} \times 100\%$$

$$\text{Acutal Relative Gap} = \frac{\lfloor \text{Upper Bound} - \text{Lower Bound} \rfloor}{\text{Number of Samples}} \times 100\%$$

Typically, the relative gap in essence denotes the probability errors between the returned upper bound and lower bound. In fact, since we use the PSTOTAP model as an experimental model, we can easily enhance the returned lower bound by using the rounding up function, which corresponds to the actual relative gap defined above. Typically, in these tests, the

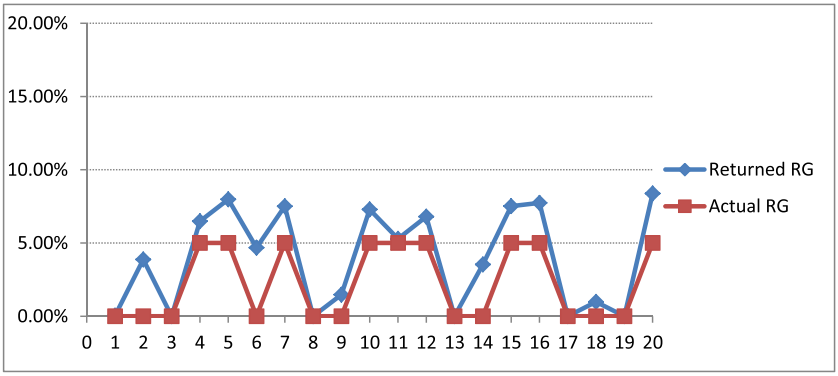

Fig. 10. Variations of Returned Relative Gaps and Actual Relative Gaps for Different Tests.

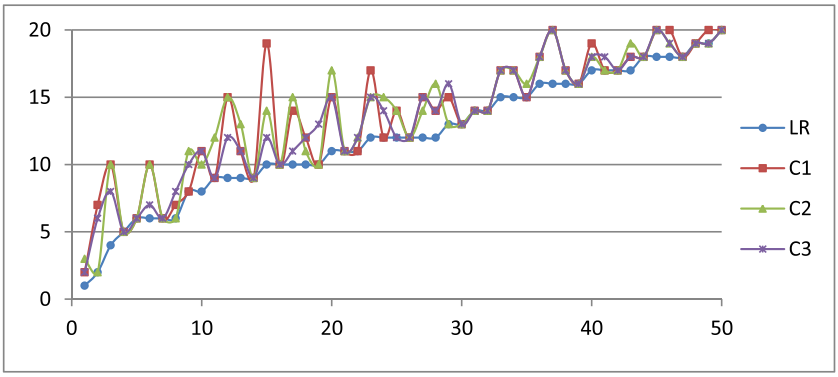

Fig. 11. Optimal Objectives of Different Instances for Different Path Finding Strategies.

returned relative gaps are not larger than 9%, and for near a half of cases, the returned relative gaps are around zero. Moreover, the actual relative gaps do not exceed 5% for all these cases, demonstrating the effectiveness and computational performance of the proposed algorithm. In addition, the average computational time of each iteration turns out to be about 120 s, which includes a total of 20 sample-based space-time path generation processes corresponding to different samples. For understanding straightforwardly, we give Fig. 10 to show the variation of the returned relative gaps (Returned RG) and actual relative gaps (Actual RG).

To further show the performance of the proposed approaches, we next implement a second set of experiments to show the effectiveness of the proposed sample-based formulation in representing spatiotemporal and stochastic correlations. For comparison, we here consider three other path finding criteria to find different paths, and the corresponding objectives in the PSTOTAP model will be used to specify the performance of the generated solutions in different path finding methods. In detail, the newly considered path finding criteria are described as follows. (C1) criterion 1: we aim to generate the shortest travel time path with the least link travel times among different samples, which is typically a risk preference decision-making process; (C2) criterion 2: we generate the shortest travel time path with the average link travel times among different samples, which is a commonly used method in the real-world applications; (C3) criterion 3: we generate the shortest travel time path with the worst link travel times among different samples, which corresponds to a risk averse decision-making process. Typically, these three criteria do not specifically take randomness into consideration. In the experiments, we would like to compare the different objective values in the PSTOTAP model for 50 randomly selected OD pairs, shown in Fig. 11, where the x-coordinate and y-coordinate respectively represent the OD pair indexes and returned objectives of the PSTOTAP model with respect to different solution criteria.

For comparative convenience in this figure, we particularly rank the indexes of these 50 experiment cases according to the increasing tendency of the near-optimal objective values returned by the Lagrangian relaxation approach. Obviously, among these solution criteria, the Lagrangian relaxation-based approach can generate the best solutions for all of these experiments, which shows the effectiveness and necessity of considering both the spatiotemporal correlation and randomness in the decision-making process. It is observed that the criteria (C1)–(C3) sometimes can generate the solutions with large gaps in comparison with Lagrangian relaxation based approach. For instance, in experiment 15, the path finding criterion (C1) returns a solution with a large gap (i.e., $19-10=9$), and the similar situations also occur for experiments 3, 12, 20, 23 and 37, respectively. In addition, we note that for some OD pairs, the solution criteria (C1)–(C3) can also find the same

optimal objectives as those obtained by the Lagrangian relaxation approach, but typically, the ratios of finding the best solutions by (C1)–(C3) do not exceed fifty percent in all the experiments.

7. Conclusions

Focusing on generating the a priori reliable path, this paper proposed different modeling methods within the stochastic optimization framework. Specifically, we adopted sample-based time-variant link travel times to represent the dynamics and randomness in the transportation network. Then two non-expected utility based reliable pathing models were formulated to characterize the two-stage elementary path generation process, i.e., the on-time arrival probability model and percentile travel time model. The single-stage equivalent models were also deduced based on the analyses of the model characteristics. A decomposition based algorithmic framework was proposed to dualize the hard constraints into the objective function for both big-*M* and probability–space–time network based reliable models. Through solving a series of sub-problems by using the label correcting algorithm, we designed the sub-gradient algorithm (or cooperated with branch and bound process) to find the tight lower bound to the original problem. The numerical examples on Sioux Falls network indicated that the proposed approaches could efficiently find the optimal (or near-optimal) solutions in comparison to commercial optimization software (e.g., GAMS with CPLEX solver). In particular, experimental results on the San Diego network and grid network shown that the algorithm can effectively find the optimal solutions with acceptable gaps between the returned lower and upper bounds, demonstrating the effectiveness of the methods proposed in this study.

Future research can focus on the following several aspects. (1) Note that the complexity of the considered problem is closely related to some pre-given parameters, such as the network scale, the number of samples and the number of involved timestamps. If these parameters turn large, the formulated models are typically large-scale problems. In this case, how to design the effective reduction methods to decrease the computational intensity is a significant topic in our further study. (2) In the numerical examples, the proposed algorithm demonstrates its effectiveness in solving the proposed models. However, in the large-scale networks (e.g., the transportation network for the Bay Area, California), the time-efficiency of the algorithm is typically not satisfactory for the real-world applications. Thus, the other efficient algorithms should also to be developed for the extremely large network-based problems. (3) The proposed modeling ideas can be expected to generalize to the other transportation fields, such as the train scheduling problems and evacuation problems.

Acknowledgments

The first author was supported by the National Natural Science Foundation of China (Nos. 71271020, 71422002, 71621001).

References

Andreas, A.K., Smith, J.C., 2008. Mathematical programming algorithms for two-path routing problems with reliability considerations. INFORMS J. Comput. 20 (4), 553–564.
Andreas, A.K., Smith, J.C., Kucukyavuz, S., 2008. Branch-and-price-and-cut algorithms for solving the reliable h-paths problem. J. Global Optim. 42, 443–466.
Boyles, S.D., Waller, S.T., 2011. Optimal information location for adaptive routing. Netw. Spat. Econ. 11 (2), 233–254.
Baldacci, R., Mingozzi, A., Roberti, R., 2011. New route relaxation and pricing strategies for the vehicle routing problem. Oper. Res. 59 (5), 1269–1283.
Chen, A., Ji, Z.W., 2005. Path finding under uncertainty. J. Adv. Transp. 39, 19–37.
Chen, A., Zhou, Z., 2010. The α-reliable mean-excess traffic equilibrium model with stochastic travel times. Transp. Res. Part B 44 (4), 493–513.
Chen, B., Lam, W.H.K., Sumalee, A., Li, Q., Tam, M., 2014. Reliable shortest path problems in stochastic time-dependent networks. J. Intell. Transp. Syst. 18 (2), 177–189.
Chen, B., Lam, W.H.K., Tam, M., 2010. Modeling departure time and route choice problems in stochastic road networks for online ATIS applications. J. Eastern Asia Soc. Transp. Stud. 8, 1796–1805.
Chen, B., Li, Q., Lam, W.H.K., 2016. Finding the K reliable shortest paths under travel time uncertainty. Transp. Res. Part B 94, 189–203.
Chen, P., Nie, Y., 2015. Stochastic optimal path problem with relays. Transp. Res. Part C 59, 48–65.
Christman, C., Cassamano, J., 2013. Maximizing the probability of arriving on time. Lect. Notes Comput. Sci. 7984, 142–157.
Fan, Y.Y., Kalaba, R.E., Moore II, J.E., 2005a. Arriving on time. J. Opt. Theory Appl. 127 (3), 497–513.
Fan, Y.Y., Kalaba, R.E., Moore II, J.E., 2005b. Shortest paths in stochastic networks with correlated link costs. Comput. Math. Appl. 49, 1549–1564.
Fan, Y.Y., Nie, Y., 2006. Optimal routing for maximizing the travel time reliability. Netw. Spat. Econ. 6 (3), 333–344.
Fu, L., 2001. An adaptive routing algorithm for in-vehicle route guidance systems with real-time information. Transp. Res. Part B 35 (8), 749–765.
Gao, S., Chabini, I., 2006. Optimal routing policy problems in stochastic time-dependent networks. Transp. Res. Part B 40 (2), 93–122.
Gao, S., Frejinger, E., Ben-Akiva, M., 2008. Adaptive route choice models in stochastic time-dependent networks. Transp. Res. Rec. 2085, 136–143.
Gao, S., Frejinger, E., Ben-Akiva, M.E., 2010. Adaptive route choices in risky traffic networks: a prospect theory approach. Transp. Res. Part C 18 (5), 727–740.
Gounaris, C.E., Wiesemann, W., Floudas, C.A., 2013. The robust capacitated vehicle routing problem under demand uncertainty. Oper. Res. 61 (3), 677–693.
Hall, R.W., 1986. The fastest path through a network with random time-dependent travel times. Transp. Sci. 20 (3), 182–188.
Huang, H., Gao, S., 2012. Optimal paths in dynamic networks with dependent random link travel times. Transp. Res. Part B 46 (5), 579–598.
Hickman, M.D., Bernstein, D.H., 1997. Transit service and path choice models in stochastic and time-dependent networks. Transp. Sci. 31 (2), 129–146.
Khani, A., Boyles, S.D., 2015. An exact algorithm for the mean–standard deviation shortest path problem. Transp. Res. Part B 81, 252–266.
Li, Q., Nie, Y., Vallamsundar, S., Lin, J., Homem-de-Mello, T., 2016. Finding efficient and environmentally friendly paths for risk-averse freight carriers. Netw. Spat. Econ. 16, 255–275.
Liu, J., Zhou, X., 2016. Capacitated transit service network design with boundedly rational agents. Transp. Res. Part B 93, 225–250.
Lu, C.C., Liu, J., Qu, Y., Peeta, S., Rouphail, N.M., Zhou, X., 2016. Eco-system optimal time-dependent flow assignment in a congested network. Transp. Res. Part B 94, 217–239.
Mahmoudi, M., Zhou, X., 2016. Finding optimal solutions for vehicle routing problem with pickup and delivery services with time windows: a dynamic programming approach based on state-space-time network representations. Transp. Res. Part B 89, 19–42.
Miller-Hooks, E.D., 1997. Optimal routing in time-varying, stochastic network: algorithms and implementations. University of Texas at Austin, USA.

Miller-Hooks, E.D., Mahmassani, H.S., 2000. Least expected time paths in stochastic, time-varying transportation networks. Transp. Sci. 34 (2), 198–215.
Miller-Hooks, E., 2001. Adaptive least-expected time paths in stochastic, time-varying transportation and data networks. Networks 37 (1), 35–52.
Miller-Hooks, E., Mahmassani, H., 2003. Path comparisons for a priori and time-adaptive decisions in stochastic, time-varying networks. Eur. J. Oper. Res. 146 (1), 67–82.
Nie, Y., Wu, X., 2009. Shortest path problem considering on-time arrival probability. Transp. Res. Part B 43 (6), 597–613.
Nie, Y., Wu, X., Homem-de-Mello, T., 2012. Optimal path problems with second-order stochastic dominance constraints. Netw. Spat. Econ. 12, 561–587.
Nielsen, L.R., 2003. Route choice in stochastic time-dependent networks. University of Aarhus, Denmark.
Orda, A., Rom, R., 1990. Shortest-path and minimum-delay algorithms in networks with time-dependent edge-length. J. ACM 37, 607–625.
Pan, Y., Sun, L., Ge, M., 2013. Finding reliable shortest path in stochastic time-dependent network. Proc. Soc. Behav. Sci. 96, 451–460.
Qian, C., Chan, C.Y., Yung, K.L., 2011. Reaching a destination earlier by starting later: Revisited. Transp. Res. Part E 47, 641–647.
Rockafellar, R.T., Wets, R.J.B, 1991. Scenarios and policy aggregation in optimization under uncertainty. Math. Oper. Res. 16 (1), 119–147.
Samaranayake, S., Blandin, S., Bayen, A., 2011. A tractable class of algorithms for reliable routing in stochastic networks. In: Proceedings of the 19th International Symposium on Transportation and Traffic Theory. Berkeley, USA.
Samaranayake, S., Blandin, S., Bayen, A, 2012. Speedup techniques for the stochastic on-time arrival problem. In: 12th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization and Systems, pp. 83–95.
Shahabi, M., Unnikrishnan, A., Boyles, S.D., 2015. Robust optimization strategy for the shortest path problem under uncertain link travel cost distribution. Comput. Aided Civil Infrastruct. Eng. 30 (6), 433–448.
Sivakumar, R.A., Batta, R., 1994. The variance-constrained shortest path problem. Transp. Sci. 28 (4), 309–316.
Toriello, A., Haskell, W.B., Poremba, M., 2014. A dynamic traveling salesman problem with stochastic arc costs. Oper. Res. 62 (5), 1107–1125.
Waller, S.T., Ziliaskopoulous, A.K., 2002. On the online shortest path problem with limited arc cost dependencies. Networks 40 (4), 216–227.
Wang, L., Gao, Z., Yang, L., 2016. A priori least expected time paths in fuzzy, time-variant transportation networks. Eng. Opt. 48 (2), 272–298.
Wu, X., Nie, Y., 2011. Modeling heterogeneous risk-taking behavior in route choice: a stochastic dominance approach. Transp. Res. Part A 45, 896–915.
Xing, T., Zhou, X., 2011. Finding the most reliable path with and without link travel time correlation: a Lagrangian substitution based approach. Transp. Res. Part B 45 (10), 1660–1679.
Xing, T., Zhou, X., 2013. Reformulation and solution algorithms for absolute and percentile robust shortest path problems. IEEE Trans. Intell. Transp. Syst. 14 (2), 943–954.
Yang, B., Miller-Hooks, E., 2004. Adaptive routing considering delays due to signal operations. Transp. Res. Part B 38 (5), 385–413.
Yang, L., Yang, X., You, C., 2013. Stochastic scenario-based time-stage optimization model for the least expected time shortest path problem. Int. J. Uncertain., Fuzziness Knowl. Based Syst. 21 (Supp01), 17–33.
Yang, L., Zhou, X., 2014. Constraint reformulation and a Lagrangian relaxation-based solution algorithm for a least expected time path problem. Transp. Res. Part B 59, 22–44.
Yang, L., Zhou, X., Gao, Z., 2014. Credibility-based rescheduling model in a double-track railway network: a fuzzy reliable optimization approach. Omega 48, 75–93.
Yang, L., Li, S., Gao, Y., Gao, Z., 2015. A coordinated routing model with optimized velocity for train scheduling on a single-track railway line. Int. J. Intell. Syst. 30, 3–22.
Yang, L., Zhang, Y., Li, S., Gao, Y., 2016. A two-stage stochastic optimization model for the transfer activity choice in metro networks. Transp. Res. Part B 83, 271–297.
Yu, X., Mai, T., Ding-Mastera, J., Gao, S., Frejinger, E., 2015. 2015. Comparison of the recursive and non-recursive models for routing policy choice in stochastic time-dependent networks. In: International Choice Modelling Conference, pp. 10–13. May.
Zhang, Y., Shen, Z.-J.M., Song, S., 2016. Parametric search for the bi-attribute concave shortest path problem. Transp. Res. Part B 94, 150–168.
Zockaie, A., Nie, Y., Wu, X., Mahmassani, H.S., 2013. Impacts of correlations on reliable shortest path finding: a simulation-based study. Transp. Res. Rec. 2334, 1–9.