Most Reliable Path-finding Algorithm for Maximizing On-Time Arrival Probability

Bi Yu Chen[a,b*], Chaoyang Shi[a,b], Junlong Zhang[b,c], William H. K. Lam[b], Qingquan Li[d,a] and Shujin Xiang[a]

[a]State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China; [b]Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hong Kong; [c]Edward P. Fitts Department of Industrial and Systems Engineering, North Carolina State University, Raleigh, NC 27695, USA; [d]Shenzhen Key Laboratory of Spatial Smart Sensing and Services, Shenzhen University, Shenzhen, China

Abstract

Finding the most reliable path that maximizes the probability of on-time arrival is commonly encountered by travelers in face of travel time uncertainties. However, few exact solution algorithms have been proposed in the literature to efficiently determine the most reliable path in large-scale road networks. In this study, a two-stage solution algorithm is proposed to exactly solve the most reliable path problem. In the first stage, the upper and lower bounds of on-time arrival probability are estimated. Dominance conditions and the monotonic property of the most reliable path problem are then established. In the second stage, the multi-criteria label-setting approach is utilized to efficiently determine the most reliable path. To illustrate the applicability of the proposed solution algorithm, a comprehensive case study is carried out using a real road network with stochastic travel times. The results of case study show that the proposed solution algorithm has a remarkable computational advantage over the existing multi-criteria label-correcting algorithm.

Key words: the most reliable path problem, travel time reliability, multi-criteria optimization

1. Introduction

The shortest path problems have been intensively studied during the past half-century, due to their broad applications in various scientific and engineering fields, such as traffic planning, intelligent transportation systems, internet routing, and etc. (Fu et al., 2014a; Polimeni and Vitetta, 2014). Following the pioneering work of Dijkstra (1959), several efficient solution algorithms and data structures have been developed for solving shortest path problems with deterministic link costs (Bast et al., 2007; Geisberger et al., 2012; Li et al., 2015).

---

[*] To cite this paper: Chen, B.Y., Shi, C., Zhang, J., Lam, W.H.K., Li, Q.Q. and Xiang, S., 2017, Most reliable path-finding algorithm for maximizing on-time arrival probability. Transportmetrica B, 5, pp. 253-269.

In recent years, there is a resurgence of interest in the studies of stochastic shortest path problems for considering travel time uncertainties in congested road networks (Nie and Wu, 2009; Chen et al., 2012; Yang and Zhou, 2014; Shahabi et al., 2015). The stochastic nature of link travel times, due to demand fluctuations and capacity degradations, has been well recognized by transportation researchers and practitioners (Mohaymany et al., 2013; Taylor, 2013; Siu and Lo, 2013; Lam et al., 2014). Many empirical studies have found that travel time uncertainty has a significant impact on travelers' route choice behaviors (Tam et al., 2008; Carrion and Levinson, 2012). This is because travel time uncertainty may result in undesirable late arrival and impose a high penalty on travelers. As a result, travelers tend to become risk-averse for their travels by departing early to ensure a high on-time arrival probability, termed as travel time reliability in the literature (Bell and Iida, 1997). Therefore, individuals' travel time reliability concerns are necessary inclusions in stochastic shortest path problems.

How to determine the optimal path considering travel time reliability has received much attention in the literature. In his seminal work, Frank (1969) introduced the concept of finding the most reliable path that maximizes travel time reliability for a given travel time budget. As an alternative definition, Chen and Ji (2005) proposed to find the α-reliable path that minimizes travel time budget required to satisfy a given travel time reliability constraint. Nie and Wu (2009) argued that the definition of the most reliable path is equivalent to that of α-reliable path, given appropriate travel time budget parameters. Chen et al. (2013a) pointed out that these two definitions have usefulness in different route searching scenarios. The α-reliable path is generally used in the pre-trip planning scenario in which travelers have freedom to choose their departure times; while the most reliable path is used in the scenario of fixed travel time budgets.

A related concept to the α-reliable path problem is the mean-standard-deviation approach that minimizes the linear combination of mean and standard deviation of path travel times (Zhou and Xing, 2011; Khani and Boyles, 2015). This mean-standard-deviation approach is equivalent to the α-reliable path problem when path travel times follow normal distributions. There are also a number of other definitions related to reliability-based shortest path problems. For example, Huang and Gao (2012) formulated traveler's risk-taking route choice behaviors based on the cumulative prospect theory (Xu et al., 2014). Chen and Zhou (2010) proposed a mean-excess travel time concept to account for the unreliability aspect of travel time variability. Watling (2006) utilized expected utility model in terms of early and late penalty to capture traveler concerns of travel time reliability. Wu and Nie (2011) summarized several existing definitions of reliability-based shortest path problems, and found that shortest path problems based on mean-excess travel time concept and expected utility model can be solved by converting them into that based on the α-reliable path concept.

Due to their non-linear objective functions, the reliability-based shortest path problems cannot be solved by classical deterministic shortest path algorithms (e.g., Dijkstra's algorithm). Mirchandani (1976) presented a recursive algorithm to determine the most reliable path. However, the algorithm requires path enumeration and can be applied to only tiny networks. To effectively determine the most reliable path and the α-reliable path, several heuristic algorithms have been developed. Chen and Ji (2005) developed a genetic algorithm for finding both the most reliable path and α-reliable path. Ji et al. (2010) extended the genetic algorithm to incorporate link travel time correlations. However, the genetic algorithms are computationally expensive and may not guarantee to obtain the optimal solution. Xing and Zhou (2013) proposed a Lagrangian substitution algorithm to find the α-reliable path with spatial correlations using sample-based formulation. Zeng et al. (2015) extended the Lagrangian substitution algorithm by representing spatial correlations based on a Cholesky decomposition. Based on K-shortest path approach, Srinivasan et al. (2014) also developed an approximation algorithm for finding the most reliable path with spatial correlations. Nevertheless, the approximation accuracy depends on the number of generated K shortest paths.

In the literature, many research efforts have been given to developing exact solution algorithms. To solve the most reliable path and α-reliable path problems, Nie and Wu (2009) proposed a multi-criteria label-correcting algorithm by generating all non-dominated paths based on the first order stochastic dominance (FSD) condition. The proposed algorithm, however, has a non-deterministic polynomial complexity, since the number of FSD non-dominated paths grows exponentially with network size. Chen et al. (2013a) investigated the dominance conditions for the α-reliable path problem. They noted that using the definition of the most reliable path, one cannot know travel time reliability until the destination has been reached, and thereby all FSD non-dominated paths have to be generated to find the most reliable path. The definition of α-reliable path with a pre-given travel time reliability preference can lead to stricter Mean-Budget (M-B) dominance condition, which can significantly reduce the number of generated FSD non-dominated paths. Based on the M-B dominance condition, efficient multi-criteria A* algorithm was developed to exactly determine the α-reliable path in large-scale road networks. Chen et al. (2012) further extended the multi-criteria A* algorithm to consider the spatial correlations in the α-reliable path problem. However, few efficient solution algorithms have been developed to exactly determine the most reliable path.

Along the line of our previous work (Chen et al., 2012, 2013a, 2013b, 2014a), this study aims to develop exact solution algorithms to efficiently determine the most reliable path in large-scale road networks. To fulfill the task, a two-stage solution algorithm is proposed. In the first stage, a deterministic shortest path search is conducted to find the least expect time path between an origin and destination (O-D) pair, so as to estimate upper and lower bounds of travel time reliability. Using estimated upper and lower bounds, the effective M-B dominance condition is investigated for solving the most reliable path problem. In the second stage, the multi-criteria label-setting technique (Chen et al., 2013a) is adopted to efficiently determine the most reliable path. To illustrate the applicability of proposed two-stage solution algorithm, a comprehensive case study is carried out using a real road network with stochastic travel times. The results of case study show that the proposed two-stage solution algorithm has a remarkable computational advantage over the existing multi-criteria label-correcting algorithm (Nie and Wu, 2009).

The rest of the paper is organized as follows. The following section presents the definition of the most reliable path problem. Section 3 presents the proposed two-stage solution algorithm for solving the most reliable path problem. Section 4 reports the case study using several large-scale networks. Section 5 presents conclusions and recommendations for further studies.

2. Problem statement

Let $G=(N,\ A)$ be a directed network consisting of a set of nodes N and a set of links A. Each node i has a set of successor nodes $SCS(i)=\{j:a_{ij}\in A\}$ and a set of predecessor nodes $PDS(i)=\{k:a_{ki}\in A\}$. Each link $a_{ij}\in A$ has a tail node $i\in N$, a head node $j\in N$ and a random travel time T_{ij}. According to previous empirical studies, link travel times could be represented by normal, log-normal, gamma or Burr distributions (Kaparias et al., 2008; Rakha et al., 2010; Susilawati et al., 2013). The mean and standard deviation of link travel time are denoted by t_{ij} and σ_{ij} respectively.

Suppose that nodes $r\in N$ and $s\in N$ represent the O-D pair. Let $P^{rs}=\{p_1^{rs},...,\ p_n^{rs}\}$ be a set of paths from origin r to destination s. Let x_{ij}^{rs} be the decision variable regarding the link-path incidence relationship, where $x_{ij}^{rs}=1$ means that the link a_{ij} is on the path p_u^{rs}, and otherwise $x_{ij}^{rs}=0$. The path travel time, denoted by T_u^{rs}, is the sum of the related link travel times along the path as

$$T_u^{rs}=\sum_{a_{ij}\in A}T_{ij}x_{ij}^{rs} \tag{1}$$

The path travel time T_u^{rs} is also a random variable. Its mean and standard deviation are denoted by t_u^{rs} and σ_u^{rs} respectively.

In this study, it is assumed that path travel times follow normal distributions and link travel times are mutually independent. These are two commonly used assumptions in the studies of stochastic shortest path problems and stochastic traffic assignment models (Chang et al., 2005; Nikolova, 2009; Chen et al., 2011, 2012). According to the central limit theorem, Lo and Tung (2003) pointed out that path travel time can be well approximated by normal distribution regardless of the types of link travel time distributions. The approximation accuracy of normal distribution is examined in Section 4.1 using a real road network. The assumption of independent link travel time distributions can be relaxed by using two-level hierarchical network model proposed by Chen et al. (2012). Under the above two assumptions, the mean and standard deviation of path travel time can be calculated as

$$t_u^{rs} = \sum_{a_{ij} \in A} t_{ij} x_{ij}^{rs} \tag{2}$$

$$\sigma_u^{rs} = \sqrt{\sum_{a_{ij} \in A} (\sigma_{ij})^2 x_{ij}^{rs}} \tag{3}$$

Let $\Phi_{T_u^{rs}}()$ be the cumulative distribution function (CDF) of path travel time distribution T_u^{rs}. Given a desirable travel time budget b, the probability that the trip can be successfully made within the travel time budget can be expressed as

$$\lambda = \Phi_{T_u^{rs}}(b) = \Phi(z_\lambda = \frac{b - t_u^{rs}}{\sigma_u^{rs}}) \tag{4}$$

where $\Phi()$ is CDF of standard normal distribution, and z_λ is the standard score of standard normal distribution at λ confidence level. This on-time arrival probability is also termed as travel time reliability (denoted by $\lambda \in (0,1)$) in the literature.

As the travel time reliability λ is monotonic increasing with the z_λ value, the most reliable path problem is equivalent to minimize the $-z_\lambda$ value. The objective function is further reformulated as 2^{-z_λ} to make objective function values positive. Then, the most reliable path problem can be formulated as the following minimization problem:

$$\underset{x_{ij}^{rs}}{Min} \quad y_\lambda^{rs} = 2^{-z_\lambda} \tag{5}$$

$$z_\lambda = \left(b - \sum_{a_{ij} \in A} t_{ij} x_{ij}^{rs} \right) \Big/ \left(\sum_{a_{ij} \in A} \sigma_{ij}^2 x_{ij}^{rs} \right) \tag{6}$$

Subject to

$$\sum_{j \in SCS(i)} x_{ij}^{rs} - \sum_{k \in PDS(i)} x_{ki}^{rs} = \begin{cases} 1 & \forall\, i = r \\ 0, & \forall\, i \neq r;\, i \neq s \\ -1 & \forall\, i = s \end{cases} \tag{7}$$

$$x_{ij}^{rs} \in \{0,1\}, \quad \forall a_{ij} \in A \tag{8}$$

Eq. (5) represents the objective function y_{λ}^{rs} which is monotonically decrease with the z_{λ} value. Eq. (6) is the z_{λ} value that travelers want to maximize. Eq. (7) ensures that the most reliable path is feasible. Eq. (8) is concerned with the link-path incidence variables which should be binary in nature.

Obviously, the most reliable path problem is non-additive and cannot be solved by existing deterministic shortest path algorithms (e.g., Dijkstra's algorithm). Nie and Wu (2009) showed that the most reliable path problem can be formulated and solved as a multi-criteria shortest path problem that relies on the first-order stochastic dominance (FSD) rule to determine a set of FSD non-dominated paths. Let p_u^{ri} be a path from origin r to node i, $\Phi_{T_u^{ri}}^{-1}()$ be the inverse of CDF of path travel time distribution T_u^{ri}, and the FSD rule is defined as follows

Proposition 1. (First order stochastic dominance) Given two paths $p_u^{ri} \neq p_v^{ri} \in P^{ri}$, p_u^{ri} FSD dominates p_v^{ri} if they satisfy $\Phi_{T_u^{ri}}^{-1}(\lambda) < \Phi_{T_v^{ri}}^{-1}(\lambda)$ for any confidence level $\forall \lambda \in (0,1)$.

Proof. See Proposition 1 in Chen et al. (2013a). □

Based on the FSD condition, the most reliable path problem can be solved by multi-criteria label-correcting algorithm (Nie and Wu, 2009) to generate all FSD non-dominated paths in the network. This solution algorithm can be computationally expensive for generating all FSD non-dominated paths in large-scale networks, because the number of FSD non-dominated paths grows exponentially with the network size. In the next section, a two-stage solution algorithm built on more effective M-B dominance condition is proposed to efficiently solve the most reliable path problem.

3. Two-stage algorithm for finding the most reliable path

This section presents the proposed two-stage algorithm, named MRP-TS. Let $p_{\alpha}^{rs} \in P^{rs}$ be the unknown most reliable path between the O-D pair, and α be its corresponding travel time reliability. In the first stage of the MRP-TS algorithm, the upper and lower bounds of α value are estimated by calculating the least expected time path (denoted by p_{θ}^{rs}). To obtain path p_{θ}^{rs}, a deterministic shortest path search is conducted from the destination to the origin. Let θ be the travel time reliability of path p_{θ}^{rs}. Its corresponding z value can be calculated by $z_{\theta} = (b - t_{\theta}^{rs}) / \sigma_{\theta}^{rs}$, where t_{θ}^{rs} and σ_{θ}^{rs} are mean and standard deviation of the path travel time respectively. According to this obtained z_{θ} value, the range of α value can be classified into following three risk-taking scenarios:

Proposition 2. Given z_θ value of the least expected time path p_θ^{rs}, the α value range of the most reliable path p_α^{rs} can be determined as

(1) If $z_\theta > 0$ (i.e., $\theta > 0.5$ risk-averse scenario), then the α value ranges within $[\theta, 1)$;

(2) If $z_\theta = 0$ (i.e., $\theta = 0.5$ risk-neutral scenario), then the α value equals to 0.5;

(3) If $z_\theta < 0$ (i.e., $\theta < 0.5$ risk-seeking scenario), then the α value ranges within $[\theta, 0.5)$.

Proof. As the least expected time path p_θ^{rs} is a feasible solution to the most reliable path problem, then its travel time reliability θ is the lower bound of the unknown α value, and thus we have $\alpha \ge \theta$. For Case (1) with $z_\theta > 0$, we have $\alpha \ge \theta > 0.5$. As the travel time reliability $\alpha \in (0,1)$, we have $\alpha \in [\theta, 1)$.

For Case (2) with $z_\theta = 0$, we have $\theta = 0.5$ and travel time budget $b = t_\theta^{rs}$. As p_θ^{rs} is the least expected time path, we have $b = t_\theta^{rs} \le t_\alpha^{rs}$ for the most reliable path p_α^{rs}. Thus, we have $z_\alpha = (b - t_\alpha^{rs}) / \sigma_\alpha^{rs} \le 0$ and $\alpha \le 0.5$. Since $\alpha \ge \theta = 0.5$ holds, we have $\alpha = 0.5$.

For Case (3) with $z_\theta < 0$, we have $\theta < 0.5$ and travel time budget $b < t_\theta^{rs}$. As p_θ^{rs} is the least expected time path, we have $b < t_\theta^{rs} \le t_\alpha^{rs}$ for the most reliable path p_α^{rs}. Therefore, $z_\alpha = (b - t_\alpha^{rs}) / \sigma_\alpha^{rs} < 0$ and $\alpha < 0.5$. Since $\alpha \ge \theta$, we have $\alpha \in [\theta, 0.5)$. □

With the estimated α value range, the effective Mean-Budget (M-B) dominance (Chen et al., 2013a) can be utilized to determine the non-dominated paths for different risk-taking scenarios.

Proposition 3. (Mean-Budget dominance) Given two paths $p_u^{ri} \ne p_v^{ri} \in P^{ri}$, p_u^{ri} M-B dominates p_v^{ri} if they satisfy $t_u^{ri} \le t_v^{ri}$ and $\Phi_{T_u^{ri}}^{-1}(\alpha) < \Phi_{T_v^{ri}}^{-1}(\alpha)$.

Proof. See Proposition 3 in Chen et al. (2013a). □

Lemma 1. If p_u^{ri} M-B dominates p_v^{ri} under λ confidence level, p_u^{ri} also M-B dominates p_v^{ri} under α confidence level if (1) $\lambda \ge \alpha > 0.5$ holds under risk-averse scenarios or (2) $0.5 > \alpha \ge \lambda$ holds under risk-seeking scenarios.

Proof. If p_u^{ri} M-B dominates p_v^{ri} under λ confidence level, we have $t_u^{ri} \le t_v^{ri}$ and $\Phi_{T_u^{ri}}^{-1}(\lambda) < \Phi_{T_v^{ri}}^{-1}(\lambda)$, according to Proposition 3. To prove p_u^{ri} M-B dominates p_v^{ri} under the α confidence level, it is equivalent to prove $\Phi_{T_u^{ri}}^{-1}(\alpha) < \Phi_{T_v^{ri}}^{-1}(\alpha)$.

We firstly prove $\lambda \geq \alpha > 0.5$ case. If $\sigma_u^{ri} - \sigma_v^{ri} < 0$, since $t_u^{ri} - t_v^{ri} \leq 0$ and $z_\alpha > 0$, we have $\Phi_{T_u^{ri}}^{-1}(\alpha) - \Phi_{T_v^{ri}}^{-1}(\alpha) = (t_u^{ri} - t_v^{ri}) + z_\alpha(\sigma_u^{ri} - \sigma_v^{ri}) < 0$. If $\sigma_u^{ri} - \sigma_v^{ri} \geq 0$, since $z_\lambda \geq z_\alpha > 0$, we have $\Phi_{T_u^{ri}}^{-1}(\alpha) - \Phi_{T_v^{ri}}^{-1}(\alpha) = (t_u^{ri} - t_v^{ri}) + z_\alpha(\sigma_u^{ri} - \sigma_v^{ri}) \leq (t_u^{ri} - t_v^{ri}) + z_\lambda(\sigma_u^{ri} - \sigma_v^{ri}) = \Phi_{T_u^{ri}}^{-1}(\lambda) - \Phi_{T_v^{ri}}^{-1}(\lambda) < 0$. Therefore, if $\lambda \geq \alpha > 0.5$ holds, then $\Phi_{T_u^{ri}}^{-1}(\alpha) < \Phi_{T_v^{ri}}^{-1}(\alpha)$ is always satisfied.

We then prove $0.5 > \alpha \geq \lambda$ case. If $\sigma_u^{ri} - \sigma_v^{ri} \leq 0$, according to $0 > z_\alpha \geq z_\lambda$, we have $\Phi_{T_u^{ri}}^{-1}(\alpha) - \Phi_{T_v^{ri}}^{-1}(\alpha) = (t_u^{ri} - t_v^{ri}) + z_\alpha(\sigma_u^{ri} - \sigma_v^{ri}) \leq (t_u^{ri} - t_v^{ri}) + z_\lambda(\sigma_u^{ri} - \sigma_v^{ri}) = \Phi_{T_u^{ri}}^{-1}(\lambda) - \Phi_{T_v^{ri}}^{-1}(\lambda) < 0$. If $\sigma_u^{ri} - \sigma_v^{ri} > 0$, according to $t_u^{ri} - t_v^{ri} < 0$ and $z_\alpha < 0$, we have $\Phi_{T_u^{ri}}^{-1}(\alpha) - \Phi_{T_v^{ri}}^{-1}(\alpha) = (t_u^{ri} - t_v^{ri}) + z_\alpha(\sigma_u^{ri} - \sigma_v^{ri}) < 0$. Therefore, if $0.5 > \alpha \geq \lambda$ holds, then $\Phi_{T_u^{ri}}^{-1}(\alpha) < \Phi_{T_v^{ri}}^{-1}(\alpha)$ is always satisfied. □

According to Lemma 1, to determine M-B non-dominated paths, the upper bound of α value can be adopted for the risk-averse scenario; while the lower bound of α value can be used for the risk-seeking scenario. The better estimation of the upper and lower bounds, the fewer non-dominated paths are generated in the path search process.

Let $\oplus$ be a path concatenation operator. $p_u^{rj} = p_u^{ri} \oplus a_{ij}$ means that the path p_u^{rj} goes through p_u^{ri} and link a_{ij}. It can prove that the objective function $y_u^{ri} = 2^{-(b-t_u^{ri})/\sigma_u^{ri}}$ monotonically increases with the path extension as below.

Lemma 2. Given any two paths p_u^{ri} and $p_u^{rj} = p_u^{ri} \oplus a_{ij}$, the relationship $y_u^{rj} > y_u^{ri}$ always holds.

Proof. According to Proposition 7 in Chen et al. (2013a), $\Phi_{T_u^{rj}}^{-1}(\lambda) \geq \Phi_{T_u^{ri}}^{-1}(\lambda)$ holds for any confidence level $\lambda \in (0,1)$. Therefore, we have $\Phi_{T_u^{rj}}(b) \leq \Phi_{T_u^{ri}}(b)$ and thus $(b - t_u^{rj}) / \sigma_u^{rj} \leq (b - t_u^{ri}) / \sigma_u^{ri}$. Therefore, $y_u^{rj} = 2^{-(b-t_u^{rj})/\sigma_u^{rj}} \geq y_u^{ri} = 2^{-(b-t_u^{ri})/\sigma_u^{ri}}$. □

Using the above propositions and lemmas, the second stage of the proposed MRP-TS algorithm is to determine the most reliable path for three different risk-taking scenarios. For the risk-neutral scenario, the most reliable path can be easily determined as the least expected time path p_θ^{rs}. For risk-seeking and risk-averse scenarios, the efficient multi-criteria label-setting technique (Chen et al., 2013a) is adopted to determine the most reliable path. The steps of the proposed MRP-TS algorithm is given as below. The detailed procedures for risk-seeking and risk-averse scenarios are respectively given in Sections 3.1 and 3.2.

Algorithm: MRP-TS

Inputs: origin r, destination s and travel time budget b

Output: the most reliable path

First stage. Risk-taking scenario determination:

Determine the least expected time path p_{θ}^{rs} using Dijkstra's algorithm.

Calculate $z_{\theta} := (b - t_{\theta}^{rs}) / \sigma_{\theta}^{rs}$.

Second Stage. The most reliable path determination:

If $z_{\theta} = 0$, then Stop and Return p_{θ}^{rs}.

If $z_{\theta} < 0$, then call MRP-RS(r, s, b, z_{θ}) procedure for risk-seeking scenario.

If $z_{\theta} > 0$, then call MRP-RA(r, s, b) procedure for risk-averse scenario.

3.1. Solution procedure for risk-seeking scenario

Based on Lemmas 1 and 2, a multi-criteria label-setting procedure (named MRP-RS) is proposed to determine the most reliable path for the risk-seeking scenario. The travel time reliability of the least expected time path (i.e., the θ value) is used for determining the M-B non-dominated paths. In the MRP-RS procedure, multiple M-B non-dominated paths under θ confidence level are maintained at each node. During the path searching process, all non-dominated paths are stored in the scan eligible (SE) and ordered by their objective function values (i.e., $y_u^{ri} := 2^{-(b-t_u^{ri})/\sigma_u^{ri}}$). At each iteration, only one non-dominated path p_u^{ri} at the top of SE (with minimum y_u^{ri}) is selected for path extensions. A temporary path is constructed by extending the selected path p_u^{ri} to its successor link a_{ij}, denoted by $p_u^{rj} := p_u^{ri} \oplus a_{ij}$. The dominance relationship between the newly generated path p_u^{rj} and existing non-dominated paths P^{rj} at node j is determined by the CheckDominance procedure given in Appendix A. If p_u^{rj} is a non-dominated path at node j, it is then inserted into P^{rj} and SE. The newly generated path p_u^{rj} may also dominate a set of paths in P^{rj}, denoted by P_D^{rj}. These dominated paths in P_D^{rj} are eliminated from P^{rj} and SE. The procedure continues this path search process until the destination is reached or SE becomes empty. When the procedure terminates, the most reliable path can be determined as the selected path. The steps of the MRP-RS procedure are given as follows.

Procedure: MRP-RS

Inputs: origin r, destination s, travel time budget b, the z_{θ} value of the least expected time path

Output: the most reliable path

Step 1. Initialization:

Create a path p^{rr} from origin r to itself and set $P^{rr} := \{p^{rr}\}$.

Set $y_u^{rr} := 0$ and $SE := \{p^{rr}\}$.

Step 2. Label selection:

If $SE = \phi$, then Stop; otherwise, continue.

Select p_u^{ri} with minimum y_u^{ri} from the SE and set $SE := SE \setminus \{p_u^{ri}\}$.

If $i = s$, then return p_u^{ri}; otherwise continue.

Step 3. Path extension:

For each successor node $j \in SCS(i)$

If $j \in p_u^{ri}$, then scan next successor node; otherwise, continue.

Generate a new path $p_u^{rj} := p_u^{ri} \oplus a_{ij}$ and calculate $y_u^{rj} := 2^{-(b-t_u^{rj})/\sigma_u^{rj}}$.

Call procedure $P_D^{rj} :=$ CheckDominance$(p_u^{rj}, P^{rj}, z_\theta)$.

If p_u^{rj} is a non-dominated path, then set $SE := SE \cup \{p_u^{rj}\}$ and $SE := SE \setminus P_D^{rj}$.

End for

Goto Step 2.

Proposition 4. The MRP-RS procedure can determine the most reliable path p_α^{rs} between origin and destination nodes when the procedure terminates at the destination node.

Proof. When the procedure terminates, the path p_u^{rs} was selected from the SE. Let $\bar{P}^{rs}$ be the path set containing all M-B non-dominated paths between the O-D pair under θ confidence level. As $0.5 > \alpha \geq \theta$ holds, according to Lemma 1, we have the most reliable path $p_\alpha^{rs} \in \bar{P}^{rs}$ and the selected path $p_u^{rs} \in \bar{P}^{rs}$. Because the label-setting technique is used in the procedure, the selected path p_u^{rs} has the minimum objective function value y_u^{rs} amongst all paths in SE. Let $\hat{P}^{rs}$ be the set of paths that have been generated and stored in SE. Thus, $y_u^{rs} \leq \hat{y}_v^{rs}$ holds for any path $\hat{p}_v^{rs} \in \hat{P}^{rs}$. Let $\tilde{P}^{rs} = \bar{P}^{rs} - \hat{P}^{rs}$ be the set of non-dominated paths that have not been generated by the procedure. Without loss of generality, let $\tilde{p}_v^{rs} = \tilde{p}_v^{ri} \oplus \tilde{p}_v^{is}$ be a path in $\tilde{P}^{rs}$ and $\tilde{p}_v^{ri}$ be its sub-path within SE. Therefore, the sub-path's objective function satisfies $\tilde{y}_v^{ri} \geq y_u^{rs}$, and thus $\tilde{y}_v^{rs} > \tilde{y}_v^{ri} \geq y_u^{rs}$ holds for any path $p_v^{rs} \in \tilde{P}^{rs}$ according to Lemma 2. Therefore, the selected path p_u^{rs} has the minimum objective value amongst all paths in $\bar{P}^{rs}$ and can be determined as the most reliable path p_{α}^{rs} . □

3.2. Solution procedure for risk-averse scenario

To determine the most reliable path for risk-averse scenario, a multi-criteria A* procedure named MRP-RA is proposed. The MRP-RA procedure has two modifications to the above MRP-RS procedure. The first modification is the way to determine the M-B dominated paths. Unlike the MRP-RS procedure using a fix θ confidence level, the MRP-RA procedure uses an adaptive confidence level λ to determine the M-B dominated paths. Initially, the confidence level is set as $\lambda = 1-\varepsilon \approx 0.999$ (i.e., using $\varepsilon = 0.001$ and $z_{\lambda} = 3.5$). During the path search process, the confidence level λ is iteratively reduced to travel time reliability of the selected path p_u^{ri} (denoted by u). According to Lemma 2, the u value is decreasing during the path search process and it is the upper bound of the optimal travel time reliability. According to Lemma 1, the use of smaller upper bound value can help identify and discard more M-B dominated paths so as to improve the efficiency of solution algorithm.

The second modification is the introduction of a heuristic function $\overline{y}_u^{ri}$ instead of objective function y_u^{ri}:

$$\overline{y}_u^{ri} = 2^{-\left(b-t_u^{ri}-h(i)\right)/\sigma_u^{ri}} \tag{9}$$

where $h(i)$ is the estimated lower bound on the mean path travel time from node i to the destination. Like A* algorithm, the use of heuristic function can speed up the most reliable path search process. When the estimated lower bound is admissible satisfying $h(j)+t_{ij} \geq h(i)$, it can prove that the heuristic function $\overline{y}_u^{ri}$ monotonically increases with the path extensions as below.

Lemma 3. Given any two paths p_u^{ri} and $p_u^{rj} = p_u^{ri} \oplus a_{ij}$, the relationship $\overline{y}_u^{rj} > \overline{y}_u^{ri}$ always holds during the path search process, if estimated lower bound satisfies $h(j)+t_{ij} \geq h(i)$.

Proof. Since $-\left(b-t_u^{rj}-h(j)\right) \leq 0$ for risk-averse scenarios, we have $\overline{y}_u^{rj} = 2^{-\left(b-t_u^{rj}-h(j)\right)/\sqrt{(\sigma_u^{ri})^2+\sigma_{ij}^2}} \geq 2^{-\left(b-t_u^{rj}-h(j)\right)/\sigma_u^{ri}}$. As $h(j)+t_{ij} \geq h(i)$, we have $\overline{y}_u^{rj} \geq 2^{-\left(b-t_u^{rj}-h(j)\right)/\sigma_u^{ri}} = 2^{-\left(b-t_u^{ri}-t_{ij}-h(j)\right)/\sigma_u^{ri}} \geq 2^{-\left(b-t_u^{ri}-h(i)\right)/\sigma_u^{ri}} = \overline{y}_u^{ri}$.□

The detailed steps of the MRP-RA procedure are given as follows.

Procedure: MRP-RA

Inputs: origin r, destination s, time budget b
Output: the most reliable path
Step 1. Initialization:
Create a path p^{rr} from origin r to itself and set $P^{rr} := \{p^{rr}\}$.
Set $z_\lambda := 3.5$, $\bar{y}_u^{rr} := 0$, and $SE := \{p^{rr}\}$.
Step 2. Label selection:
If $SE = \phi$, then Stop; otherwise, continue.
Select p_u^{ri} with minimum $\bar{y}_u^{ri}$ from the SE and set $SE := SE \setminus \{p_u^{ri}\}$.
Set $z_\lambda := (b - t_u^{ri}) / \sigma_u^{ri}$.
If $i = s$, then return p_u^{ri}; otherwise continue.
Step 3. Path extension:
For each successor node $j \in SCS(i)$
If $j \in p_u^{ri}$, then scan next successor node; otherwise, continue.
Generate a new path $p_u^{rj} := p_u^{ri} \oplus a_{ij}$ and calculate $h(j)$ and $\bar{y}_u^{rj} := 2^{-(b - t_u^{rj} - h(j))/\sigma_u^{rj}}$.
Call procedure $P_D^{rj} :=$ CheckDominance$(p_u^{rj}, P^{rj}, z_\lambda)$.
If p_u^{rj} is a non-dominated path, then set $SE := SE \cup \{p_u^{rj}\}$ and $SE := SE \setminus P_D^{rj}$.
End for
Goto Step 2.

A common admissible $h(i)$ is the Euclidean distance function $d_{is} / v_{\max}$, where d_{is} is the Euclidean distance from node i to the destination node, and $v_{\max}$ is the maximum travel speed in the network. Another admissible $h(i)$ is the least expected time from node i to the destination node. Based on Lemmas 1 and 3, the optimality of the proposed MRP-RA procedure can be proved as follows.

Proposition 5. If the estimated lower bound is admissible, the MRP-RA procedure can determine the most reliable path p_α^{rs} between origin and destination nodes when the procedure terminates at the destination node.

Proof. When the procedure terminates, the path p_u^{rs} was selected from the SE. Let $\bar{P}^{rs}$ be the path set containing all M-B non-dominated paths between the O-D pair under λ confidence level. As $\lambda \geq \alpha > 0.5$ holds, according to Lemma 1, we have the most reliable path $p_\alpha^{rs} \in \bar{P}^{rs}$ and the selected path $p_u^{rs} \in \bar{P}^{rs}$. As the label-setting technique is used in the procedure, the selected path p_u^{rs} has the minimum heuristic function value $\bar{y}_u^{rs}$ amongst all paths in SE. Let $\hat{P}^{rs}$ be the set of paths that have been generated and stored in SE. Thus, $\bar{y}_u^{rs} \le \hat{y}_v^{rs}$ holds for any path $\forall \hat{p}_v^{rs} \in \hat{P}^{rs}$. Let $\tilde{P}^{rs} = \bar{P}^{rs} - \hat{P}^{rs}$ be the set of non-dominated paths that have not been generated by the procedure. Without loss of generality, let $\tilde{p}_v^{rs} = \tilde{p}_v^{ri} \oplus \tilde{p}_v^{is}$ be a path in $\tilde{P}^{rs}$ and $\tilde{p}_v^{ri}$ be its sub-path within SE. Therefore, the sub-path's heuristic function satisfies $\tilde{y}_v^{ri} \ge \bar{y}_u^{rs}$. As the estimated lower bound is admissible, $\tilde{y}_v^{rs} > \tilde{y}_v^{ri} \ge \bar{y}_u^{rs}$ holds for any path $\forall \tilde{p}_v^{rs} \in \tilde{P}^{rs}$ according to Lemma 3. Therefore, the selected path p_u^{rs} has the minimum heuristic value amongst all paths in $\bar{P}^{rs}$ and can be determined as the most reliable path p_α^{rs}. □

3.3. Algorithm complexity analysis

The computational complexity of the proposed MRP-TS algorithm is analyzed. For the risk-neutral scenario, the MRP-TS algorithm is equivalent to the Dijkstra's algorithm with a complexity of $O\left(|A|+|N|Log(|N|)\right)$ when F-heap data structure is used (Fredman and Tarjan, 1987), where $|A|$ and $|N|$ are respectively the number of links and nodes in the network. For risk-seeking and risk-averse scenarios, the MRP-RS and MRP-RA procedures use the multi-criteria label-setting technique. In the worst case, label selection step (Step 2) requires $O\left(|N||P|Log(|N||P|)\right)$ when F-heap data structure is used, where $|P|$ is the maximum number of non-dominated paths at one network node. The path extension step (Step 3) requires $O\left(|A||P|^2\right)$. Therefore, the MRP-RS and MRP-RA procedures run in $O\left(|A||P|^2+|A|+|N|Log(|N|)\right)$. Theoretically, the MRP-RS and MRP-RA procedures have a non-polynomial complexity, because $|P|$ grows exponentially with network size. In practice, the $|P|$ value is much smaller than the maximum possible size especially for sparse transport networks (Chen et al., 2013a; Nie and Wu, 2009).

4. Case Study

4.1. Data description

This section presents a case study to demonstrate the applicability of the proposed MRP-TS algorithm. A real-world road network of Wuhan city, China was used in this case study. The Wuhan network consists of 19,306 nodes and 46,757 links. The link travel time distributions of the Wuhan network were estimated using data collected by a real floating-car system. The floating-car system in Wuhan city makes use of 11,248 taxis equipped with Global Positioning System (GPS) devices to record their traveling locations in about every forty seconds. The collected data were then matched onto the road network to re-construct taxi trajectories using multi-criteria dynamic programming map-matching (MDP-MM) algorithm (Chen et al., 2014b). Previous study has shown that the MDP-MM algorithm can achieve a very high map-matching accuracy (i.e., over 90%) for large-scale low-frequency data (Chen et al., 2014b). After the map-matching process, the taxi trajectories were separated into link segments and extracted for link travel time distribution analysis.

In this case study, the taxi trajectories in Wuhan city were collected on five weekdays (14-18 September 2009) during the morning peak period (7 a.m.–10 a.m.). Based on the previous literature, several classical distributions were considered to fit link travel time distributions, including normal, lognormal and gamma distributions (Kaparias et al., 2008; Nie and Wu, 2009; Rakha et al., 2010). The goodness-of-fit of each link travel time distribution was evaluated using the χ^2 (chi-square) test. Table 1 gives the goodness-of-fit results for these three types of travel time distributions at 5% significant level. As shown in the table, only 58.85% of link travel times can be modelled as normal distributions. This result is consistent with the previous empirical findings that the normal distribution is unable to model link travel time distributions with strong positive skew and long upper tail (Susilawati et al., 2013). The lognormal distribution was superior to the normal distribution, and it can well represent 81.64% of link travel time distributions. Nevertheless, no single type of distribution can be the best-fit distribution for all link travel times in the Wuhan network. As shown in the table, the lognormal, gamma and normal distributions respectively accounted for 50.29%, 30.25% and 19.46% of the best-fit results.

Table 1. Goodness-of-fit results for three types of link travel time distributions

χ^2 test	Normal	Lognormal	Gamma
Percentage of the significant distributions	58.85%	81.64%	67.19%
Percentage of the best-fit distributions	19.46%	50.29%	30.25%

Figure 1 shows the estimated travel time distributions in the Wuhan network. Figure 1(a) shows the mean travel speeds of all network links. In the figure, links shown in red represent congested links (< 20 km/h); yellow represents slightly congested links (20–40 km/h); and green represents uncongested links (> 40 km/h). The figure shows that 10.25% of links in the Wuhan network were congested. Figure 1(b) illustrates travel time variation in the Wuhan network. The link travel time variation was measured by coefficient of variation (CV), which is the ratio of the standard deviation to the mean. Larger values of CV indicate more uncertain in the link travel time. As illustrated in Figure 1(b), link travel times in the Wuhan network were highly stochastic, with the average CV value equal to 0.37. This high degree of travel time uncertainties was mainly due to traffic interruptions caused by traffic signals and the travel demand fluctuations in congested road networks.

[Figure 1 is about here]

As a link travel time may follow lognormal, gamma or normal distribution, the calculation of path travel times can be computational intractable. To simply the calculation, it is assumed in this study that path travel times can be approximated by normal distributions. To validate this assumption, the Monte Carlo simulation technique was adopted to calculate travel time distributions of 100 random selected paths as ground trues. Figure 2 shows the approximation accuracy of mean, 10^{th} and 90^{th} percentiles of path travel time distributions. It can be seen from the figure that the normal distribution can well approximate 10^{th} percentile path travel time with 98.3% of average approximation accuracy. The normal distribution approximation became slightly worse for 90^{th} percentile path travel time (i.e., 94.9%), due to the positive skew of path travel time distributions. Nevertheless, the approximation accuracy improved with the increase of number of links along the path. Therefore, it is reasonable to assume that the path travel time distribution follows normal distribution from a practical standpoint of route guidance applications, given its computational simplicity.

[Figure 2 is about here]

4.2. Numerical example

As shown in Figure 3, one O-D pair in the Wuhan network was extracted to demonstrate the applicability of the proposed most reliable path algorithm. Three most reliable paths were found under different input travel time budget parameters (i.e., b). Figure 4 gives the path travel time distributions of these three paths. As shown in Figure 4, when $b = 2630.9$ seconds was adopted, the least expected time path (i.e., Path 3) was determined as the most reliable path which can achieve the maximum on-time arrival probability of 50% (i.e., travel time reliability $\alpha = 0.5$). With the increase of b parameter, the travel time reliability of the expected time path was improved. When $b > 2875.5$ seconds was used, the traveler would switch to Path 2 with less travel time variation. Conversely, the decrease of b parameter leads to the degradation of travel time reliability. When $b < 2301.4$ seconds, the traveler would become risk-seeking by choosing Path 1 with larger travel time variation.

[Figure 3 is about here]

[Figure 4 is about here]

4.3. Computational performance

This section examines the computational performance of the proposed MRP-TS algorithm. In this study, the proposed MRP-TS algorithm was coded in the Visual C# programming language. The classical A* algorithm was adopted in the first stage of the proposed algorithm to find the least expected time path from destination node to origin node. The estimated lower bound $h(i)$ in the A* algorithm was the Euclidean distance function. In the second stage of the proposed algorithm, the least expected time path finding results in the first stage were utilized as the estimated lower bound $h(i)$ in the MRP-RA procedure. The priority queue was implemented using the F-heap data structure for A* algorithm, and MRP-RS and MRP-RA procedures (Fredman and Tarjan, 1987).

To comparatively evaluate and benchmark the proposed MRP-TS algorithm, an improved multi-criteria label-correcting algorithm (named BSP-LC) was implemented in this study. The BSP-LC algorithm extends the original multi-criteria label-correcting algorithm (Nie and Wu, 2009) by saving unnecessary discretization on path travel time distributions. The details of this improvement is given in Appendix B. The BSP-LC algorithm was also coded in the same Visual C# programming language. All experiments were conducted on a MacBook Air laptop with a four-core Intel i7-3667U central processing unit running at 2.0 GHz (only one core was used) and the Windows 7 operating system.

Figure 5 reports the computational performance of the proposed MRP-TS algorithm in the Wuhan network under different travel time budget parameters. All the reported computational performance was the average of 100 runs, using different O-D pairs for each run. The 100 O-D pairs were randomly selected from the road network. Given a travel time reliability threshold (i.e., the α value), the travel time budget parameter for each O-D pair was calculated by solving the α-reliable path problem based on the multi-criteria A* algorithm (Chen et al., 2013a). By using the calculated travel time budget, the proposed MRP-TS algorithm can determine the same optimal paths as the multi-criteria A* algorithm. This result also indicates the optimality of the proposed MRP-TS algorithm.

As shown in Figure 5, the pre-determined travel time budget parameter (calculated by the travel time reliability α) has a significant impact on the computational performance of the proposed MRP-TS algorithm. The proposed algorithm runs faster when travelers become more risk-neutral ($\alpha = 0.5$). For example, to determine the most reliable path in the Wuhan network, the algorithm required 275.69 milliseconds when $\alpha = 0.1$, and 162.47 milliseconds when $\alpha = 0.4$. This result is because when the α value approaches to 0.5, the M-B dominance rule (i.e., Lemma 1) can determine and discard more dominated paths, and thus reduce the number of generated non-dominated paths in the network and enhance the algorithm performance. It can be evidenced by the fact that the algorithm generated 18,115 non-dominated paths when $\alpha = 0.1$, and only 11,087 non-dominated paths when $\alpha = 0.4$.

The proposed MRP-TS algorithm performs best when $\alpha = 0.5$. In this risk-neutral scenario, only the classical A* algorithm was needed to find the least expected time path in the first stage. It also can be seen from the figure that the MRP-RA procedure for risk-averse scenarios ($\alpha > 0.5$) performs much better than the MRP-RS procedure for risk-seeking scenarios ($\alpha < 0.5$). This computational advantage is due to the effectiveness of using the least expected time path finding results of the first stage as the estimated lower bound in the MRP-RA procedure.

[Figure 5 is about here]

The performance of BSP-LC algorithm was also examined in the Wuhan network using the same set of O-D pairs and travel time budgets. The BSP-LC algorithm consumed 4743.94 milliseconds for different travel time budget parameters. Compared with the results shown in Figure 5, it can be easily observed that the proposed MRP-TS algorithm has a significant computational advantage over the BSP-LC algorithm. For example, the proposed MRP-TS algorithm runs 89.27 times faster than the BSP-LC algorithm when $\alpha = 0.9$.

It was noted that the travel time budget parameter had no impact on the performance of the BSP-LC algorithm. This is because the BSP-LC algorithm utilizes bi-criteria label-correcting approach to generate all FSD non-dominated paths in the network. Then, the most reliable path can be determined as one of generated FSD non-dominated paths, regardless of different input travel time budgets. However, this approach can generate a large amount of unnecessary M-B dominated paths. In this testing network, the BSP-LC algorithm generated 225,878 FSD non-dominated paths; but 95.09% of them can be identified as unnecessary M-B dominated paths compared with the paths generated by the MRP-TS algorithm under $\alpha = 0.1$ scenario. Conversely, the proposed MRP-TS algorithm enhances the performance of the most reliable path search by reducing the number of FSD non-dominated paths for different input travel time budgets. The proposed MRP-TS algorithm estimates the range of travel time reliability in the first stage. Based on the estimated travel time reliability range, effective M-B dominance rule can be employed to determine many dominated paths that cannot be identified by FSD rule. In addition, the proposed MRP-TS algorithm, using the multi-criteria label-setting approach, can determine the most reliable path as soon as the destination node was reached. Therefore, the proposed MRP-TS algorithm can significantly reduce the number of non-dominated paths generated, evaluated, and stored when searching for the most reliable path, so as to improve the efficiency of the path finding process.

The computational performance of the MRP-TS and BSP-LC algorithms was further examined for four networks with different sizes. These four networks were adopted from Chen et al. (2013a). For each testing network, 100 O-D pairs were randomly selected. The travel time budget for each O-D pair under $\alpha = 0.9$ scenario was calculated by solving the α-reliable path problem based on the multi-criteria A* algorithm (Chen et al., 2013a). The computational performance of both MRP-TS and BSP-LC algorithms was evaluated by the average of 100 runs, using different O-D pairs for each run. Both MRP-TS and BSP-LC algorithms can obtain the same paths as the α-reliable paths calculated by the multi-criteria A* algorithm.

Table 2 reports the computational performance of both MRP-TS and BSP-LC algorithms under $\alpha = 0.9$ scenario. As demonstrated in the Wuhan network, the proposed MRP-TS algorithm performs significantly better than the BSP-LC algorithm for all testing networks with different sizes. Especially in the Chicago regional network, the computational time of the proposed MRP-TS algorithm was about 750 times faster than that of the BSP-LC algorithm.

It also can be seen from the table that the computational performance of both algorithms degrades with the increase of the network size. For instance, the proposed MRP-TS algorithm required 3.12 milliseconds for finding the most reliable path in the G1 network with 2,000 nodes. When the Chicago regional network was analyzed, the number of nodes was increased approximately 6.5 times (12,982/2,000), and the computational time approximately 15.5 times (48.26/3.12). In contrast to the MRP-TS algorithm, the computational performance of the BSP-LC algorithm significantly degraded with the network size. For example, when the Chicago regional network was examined, the computing time required by the BSP-LC algorithm increased by about 179.8 times (36251.40/201.64). This degradation of the BSP-LC algorithm occurred because the number of generated FSD non-dominated paths exponentially increases with the network size.

Table 2. Performance of MRP-TS and BSP-LC algorithms in four networks (milliseconds)

Networks	Nodes	Links	MRP-TS	BSP-LC
Chicago regional	12,982	39,018	**48.26**	**36251.40**
G1 (40*50)	2,000	7,820	3.12	201.64
G2 (50*100)	5,000	19,700	11.31	1182.10
G3(100*100)	10,000	39,600	25.63	3208.12

5. Conclusions

This study investigated the problem of finding the most reliable path that maximizes travel time reliability in road networks with stochastic link travel times. A two stage solution algorithm (namely, MRP-TS) was proposed to exactly solve the most reliable path problem. In the first stage, the upper and lower bounds of travel time reliability were estimated by calculating the least expected time path. Based on the estimated range of travel time reliability, the effective M-B dominance rule (i.e., Lemma 1) and monotonic property of objective function (i.e., Lemma 2) was established for solving the most reliable path problem. In the second stage, a multi-criteria label-setting approach was utilized to efficiently determine the most reliable path for different risk-taking scenarios. The optimality of the proposed MRP-TS algorithm was rigidly proved.

To illustrate the applicability of the proposed MRP-TS algorithm, a comprehensive case study was carried out using the Wuhan network with real traffic data. Several typical distribution including normal, lognormal and gamma were tested. It is observed in the testing network that no single type of distribution can be the best-fit distribution for all link travel times. The lognormal, gamma and normal distributions respectively account for 50.29%, 30.25% and 19.46% of the best-fit results. The Monte Carlo simulation technique was adopted to validate the assumption of path travel times following the normal distributions. Simulation results show that the normal distribution can well approximate the path travel time distributions by achieving 98.3% and 94.9% of approximate accuracy at 10^{th} and 90^{th} percentiles. Computational performance of the proposed MRP-TS algorithm was examined using five networks with different sizes. The results of computational experiments show that the proposed MRP-TS algorithm can efficiently determine the most reliable path for all testing networks, and had potential applications in developing online route guidance systems. The proposed MRP-TS algorithm had a remarkable computational advantage over the existing multi-criteria label-correcting approach built on the FSD rule (Nie and Wu, 2009).

Several directions for future research are worth noting. Firstly, the correlations of link travel times are omitted in this study for simplicity. Empirical studies have found that link travel times in urban road networks are strongly correlated, especially among neighboring links (Chan et al., 2009). A possible way to relax this assumption would be to use a two-level hierarchical network model proposed by Chen et al. (2012). Secondly, the proposed solution algorithm only consider the road transportation model. How to extend the proposed solution algorithm to multi-model networks (Fu et al., 2014b) needs further investigations. Finally, another research direction is to extend the proposed algorithms for in-vehicle navigation applications, in which the en-route traffic information is updated continuously during journeys (Xiao et al., 2013, 2014).

Acknowledgments

The work described in this paper was jointly supported by research grants from the National Science Foundation of China (No. 41231171 and 41571149), the Research Grant Council of the Hong Kong Special Administration Region (RGC No. PolyU 152074/14E), the Research Institute of Sustainable Urban Development of the Hong Kong Polytechnic University (No. 1-ZVBY), Shenzhen Scientific Research and Development Funding Program (No. ZDSY20121019111146499), and Shenzhen Dedicated Funding of Strategic Emerging Industry Development Program (No. JCYJ20121019111128765).

References

Bast, H., S. Funke, P. Sanders, and D. Schultes. 2007. "Fast routing in road networks with transit nodes." *Science* 316 (5824):566.

Bell, M. G. H., and Y. Iida. 1997. *Transportation Network Analysis*. New York: John Wiley & Sons.

Brumbaugh-Smith, J., and D. Shier. 1989. "An empirical investigation of some bicriterion shortest path algorithms." *European Journal of Operational Research* 43 (2):216-224.

Carrion, C., and D. Levinson. 2012. "Value of travel time reliability: A review of current evidence." *Transportation Research Part A* 46 (4):720–741.

Chan, K. S., W. H. K. Lam, and M. L. Tam. 2009. "Real-time estimation of arterial travel times with spatial travel time covariance relationships." *Transportation Research Record* 2121:102-109.

Chang, T. S., L. K. Nozick, and M. A. Turnquist. 2005. "Multiobjective path finding in stochastic dynamic networks, with application to routing hazardous materials shipments." *Transportation Science* 39 (3):383-399.

Chen, A., and Z. Zhou. 2010. "The alpha-reliable mean-excess traffic equilibrium model with stochastic travel times." *Transportation Research Part B* 44 (4):493-513.

Chen, A., and Z. W. Ji. 2005. "Path finding under uncertainty." *Journal of Advanced Transportation* 39 (1):19-37.

Chen, B.Y., W.H.K. Lam, A. Sumalee, and H. Shao. 2011. "An efficient solution algorithm for solving multi-class reliability-based traffic assignment problem." *Mathematical and Computer Modelling* 54(5-6): 1428-1439.

Chen, B. Y., W. H. K. Lam, A. Sumalee, and Z. L. Li. 2012. "Reliable shortest path finding in stochastic networks with spatial correlated link travel times." *International Journal of Geographical Information Science* 26 (2):365-386.

Chen, B. Y., W. H. K. Lam, A. Sumalee, Q. Q. Li, H. Shao, and Z. X. Fang. 2013a. "Finding reliable shortest paths in road networks under uncertainty." *Networks & Spatial Economics* 13 (2):123-148.

Chen, B. Y., W. H. K. Lam, Q. Q. Li, A. Sumalee, and K. Yan. 2013b. "Shortest path finding problem in stochastic time-dependent road networks with stochastic first-in-first-out property." *IEEE Transactions on Intelligent Transportation Systems* 14 (4):1907-1917.

Chen, B. Y., W. H. K. Lam, A. Sumalee, Q. Q. Li, and M. L. Tam. 2014a. "Reliable shortest path problems in stochastic time-dependent networks." *Journal of Intelligent Transportation Systems* 18 (2):177-189.

Chen, B. Y., H. Yuan, Q. Q. Li, W. H. K. Lam, S.-L. Shaw, and K. Yan. 2014b. "Map matching algorithm for large-scale low-frequency floating car data." *International Journal of Geographical Information Science* 28 (1):22-38.

Dijkstra, E. M. 1959. "A note on two problems in connexion with graphs." *Numerische Mathematic* 1:269-271.

Frank, H. 1969. "Shortest paths in probabilistic graphs." *Operations Research* 17 (4):583-599.

Fredman, M. L., and R. E. Tarjan. 1987. "Fibonacci heaps and their uses in improved network optimization algorithms." *Journal of the ACM* 34:596-615.

Fu, X., W. H. K. Lam, and Q. Meng. 2014a. "Modelling impacts of adverse weather conditions on activity-travel pattern scheduling in multi-modal transit networks." *Transportmetrica B* 2 (2):151-167.

Fu, X., W. H. K. Lam, and B. Y. Chen. 2014b. "A reliability-based traffic assignment model for multi-modal transport network under demand uncertainty." *Journal of Advanced Transportation* 48 (1):66-85.

Geisberger, R., P. Sanders, D. Schultes, and C. Vetter. 2012. "Exact routing in large road networks using contraction hierarchies." *Transportation Science* 46 (3):388-404.

Huang, H. and S. Gao. 2012. "Optimal paths in dynamic networks with dependent random link travel times." *Transportation Research Part B* 46(5): 579-598.

Ji, Z. W., Y. S. Kim, and A. Chen. 2011. "Multi-objective alpha-reliable path finding in stochastic networks with correlated link costs: A simulation-based multi-objective genetic algorithm approach (SMOGA)." *Expert Systems with Applications* 38 (3):1515-1528.

Kaparias, I., M. G. H. Bell, and H. Belzner. 2008. "A new measure of travel time reliability for in-vehicle navigation systems." *Journal of Intelligent Transportation Systems* 12 (4):202-211.

Khani, A. and S.D. Boyles, 2015. "An exact algorithm for the mean-standard deviation shortest path problem." *Transportation Research Part B* 81(1): 252–266.

Lam, W. H. K., H. K. Lo, and S. C. Wong. 2014. "Advances in equilibrium models for analyzing transportation network reliability." *Transportation Research Part B* 66:1-3.

Li, Q. Q., B. Y. Chen, Y.F. Wang, and W. H. K. Lam. 2015. "A hybrid link-node approach for finding shortest paths in road networks with turn restrictions." *Transactions in GIS* 19(6): 915-929.

Lo, H. K., and Y. K. Tung. 2003. "Network with degradable links: capacity analysis and

design." *Transportation Research Part B* 37 (4):345-363.
Mirchandani, P. B. 1976. "Shortest distance and reliability of probabilistic networks." *Computers & Operations Research* 3 (4):347-355.
Mohaymany, A. S., M. Shahri, and B. Mirbagheri. 2013. "GIS-based method for detecting high-crash-risk road segments using network kernel density estimation." *Geo-spatial Information Science* 16 (2):113-119.
Nie, Y., and X. Wu. 2009. "Shortest path problem considering on-time arrival probability." *Transportation Research Part B* 43 (6):597-613.
Nikolova, E. 2009. "Strategic algorithms." PhD Thesis. Massachusetts Institute of Technology, USA.
Polimeni, A., and A. Vitetta. 2014. "Vehicle routing in urban areas: an optimal approach with cost function calibration." *Transportmetrica B* 2 (1):1-19.
Rakha, H., I. El-Shawarby, and M. Arafeh. 2010. "Trip travel-time reliability: issues and proposed solutions." *Journal of Intelligent Transportation Systems* 14 (4):232-250.
Shahabi, M., A. Unnikrishnan, and S.D. Boyles. 2015. "Robust optimization strategy for the shortest path problem under uncertain link travel cost distribution." *Computer-Aided Civil and Infrastructure Engineering* 30(6): 433-448.
Siu, B. W. Y., and H. K. Lo. 2013. "Punctuality-based departure time scheduling under stochastic bottleneck capacity: formulation and equilibrium." *Transportmetrica B* 1 (3):195-225.
Srinivasan, K. K., A. A. Prakash, and R. Seshadri. 2014. "Finding most reliable paths on networks with correlated and shifted log-normal travel times." *Transportation Research Part B* 66:110-128.
Susilawati, S., M. A. P. Taylor, and S. Somenahalli. 2013. "Distributions of travel time variability on urban roads." *Journal of Advanced Transportation* 47 (8):720-736.
Tam, M. L., W. H. K. Lam, and H. P. Lo. 2008. "Modeling air passenger travel behavior on airport ground access mode choices." *Transportmetrica* 4 (2):135-153.
Taylor, M. A. P. 2013. "Travel through time: the story of research on travel time reliability." *Transportmetrica B* 1 (3):174-194.
Watling, D. 2006. "User equilibrium traffic network assignment with stochastic travel times and late arrival penalty." *European Journal of Operational Research* 175 (3):1539-1556.
Wu, X. and Y. Nie. 2009. "Implementation issues for the reliable a priori shortest path problem." *Transportation Research Record* 2091:51-60.
Wu, X., and Y. Nie. 2011. "Modeling heterogeneous risk-taking behavior in route choice: A stochastic dominance approach." *Transportation Research Part A* 45 (9):896-915.
Xiao, L., and H. K. Lo. 2013. "Adaptive vehicle routing for risk-averse travelers." *Transportation Research Part C* 36:460-479.
Xiao, L., and H. K. Lo. 2014. "Adaptive vehicle navigation with en route stochastic traffic information." *IEEE Transactions on Intelligent Transportation Systems* 15

(5):1900-1912.
Xing, T., and X. Zhou. 2011. "Finding the most reliable path with and without link travel time correlation: A Lagrangian substitution based approach." *Transportation Research Part B* 45 (10):1660-1679.
Xing, T. and X. Zhou. 2013. "Reformulation and solution algorithms for absolute and percentile robust shortest path problems." *IEEE Transactions on Intelligent Transportation Systems* 14(2): 943 - 954.
Xu, H., W. H. K. Lam, and J. Zhou. 2014. "Modelling road users' behavioural change over time in stochastic road networks with guidance information." *Transportmetrica B* 2 (1):20-39.
Yang, L. and X. Zhou. 2014. "Constraint reformulation and a Lagrangian relaxation-based solution algorithm for a least expected time path problem." *Transportation Research Part B* 59: 22-44.
Zeng, W., T. Miwa, Y. Wakita, and T. Morikawa. 2015. "Application of Lagrangian relaxation approach to α-reliable path finding in stochastic networks with correlated link travel times." *Transportation Research Part C* 56: 309-334.

Appendix A.

This appendix presents the detailed steps of the *CheckDominance* Procedure used in the proposed two-stage algorithm.

Procedure: *CheckDominance*

Inputs: A newly generated path p_u^{rj}, a set of non-dominated paths P^{rj} sorting in an ascent order by mean travel times, and the z_λ value at λ confidence level

Returns: P_D^{rj} storing the set of paths dominated by p_u^{rj}

Step 1: Initialization

  Set $P_D^{rj} := \phi$ and $n := 1$.

Step 2: Dominance relationship determination

  While $n \leq |P^{rj}|$ and $t_u^{rj} \geq t_n^{rj}$ ($|P^{rj}|$ is the number of paths in P^{rj})

    If $\Phi_{p_u^{rj}}^{-1}(\lambda) > \Phi_{p_n^{rj}}^{-1}(\lambda)$, then return P_D^{rj}.

    Set $n := n+1$.

  End while

  Insert p_u^{rj} into P^{rj} at n^{th} position and set $n := n+1$ (by default $|P^{rj}| := |P^{rj}|+1$).

  While $n \leq |P^{rj}|$ and $\Phi_{p_u^{rj}}^{-1}(\lambda) < \Phi_{p_n^{rj}}^{-1}(\lambda)$

    Set $P^{rj} := P^{rj} \setminus \{p_n^{rj}\}$ and $P_D^{rj} := P_D^{rj} \cup \{p_n^{rj}\}$.

Set $n := n+1$.

End while

Return P_D^{rj}.

Appendix B.

This appendix presents an improved multi-criteria label-correcting algorithm (called BSP-LC) to solve the most reliable path problem. The multi-criteria label-correcting algorithm (Nie and Wu, 2009) utilizes the FSD condition to determine the non-dominated paths. As the FSD rule is defined with respect to the unbounded CDFs, a discretization on CDFs is employed by the MSP-LC algorithm to make path travel time distributions comparable. Let L be the number of discrete elements used to discretize CDFs. Then, each element has a probability of $\varepsilon = 1/L$. This discretization scheme essentially uses the partial CDFs (i.e., $\Phi_{T_u^{ri}}^{-1}(\lambda)$, $\forall\lambda \in [\varepsilon, 1-\varepsilon]$). The accuracy of such approximation depends on the number of used discrete elements. For example, when $L = 1000$ is adopted (i.e., $\varepsilon = 0.001$), very accurate path travel time distributions and non-dominated path set can be obtained (Wu and Nie, 2009).

It can be proved the following simplified way to determine the FSD dominated paths when path travel times follow normal distributions.

Lemma 4. Given two paths $p_u^{ri} \neq p_v^{ri} \in P^{ri}$, we have $\Phi_{T_u^{ri}}^{-1}(\lambda) < \Phi_{T_v^{ri}}^{-1}(\lambda)$ $\forall\lambda \in [\varepsilon, 1-\varepsilon]$, if they satisfy $\Phi_{T_u^{ri}}^{-1}(\varepsilon) < \Phi_{T_v^{ri}}^{-1}(\varepsilon)$ and $\Phi_{T_u^{ri}}^{-1}(1-\varepsilon) < \Phi_{T_v^{ri}}^{-1}(1-\varepsilon)$.

Proof. When $\sigma_u^{ri} - \sigma_v^{ri} \leq 0$, according to $\Phi_{T_u^{ri}}^{-1}(\varepsilon) < \Phi_{T_v^{ri}}^{-1}(\varepsilon)$, we have $(t_u^{ri} - t_v^{ri}) + z_\varepsilon(\sigma_u^{ri} - \sigma_v^{ri}) < 0$. Since $z_\lambda \geq z_\varepsilon$ holds for any $\lambda \in [\varepsilon, 1-\varepsilon]$, we have $\Phi_{T_u^{ri}}^{-1}(\lambda) - \Phi_{T_v^{ri}}^{-1}(\lambda) = (t_u^{ri} - t_v^{ri}) + z_\lambda(\sigma_u^{ri} - \sigma_v^{ri}) < (t_u^{ri} - t_v^{ri}) + z_\varepsilon(\sigma_u^{ri} - \sigma_v^{ri}) < 0$. Similarly, when $\sigma_u^{ri} - \sigma_v^{ri} \geq 0$, according to $\Phi_{T_u^{ri}}^{-1}(1-\varepsilon) < \Phi_{T_v^{ri}}^{-1}(1-\varepsilon)$, we have $(t_u^{ri} - t_v^{ri}) + z_{1-\varepsilon}(\sigma_u^{ri} - \sigma_v^{ri}) < 0$. As $z_\lambda \leq z_{1-\varepsilon}$ holds for any $\lambda \in [\varepsilon, 1-\varepsilon]$, we have $\Phi_{T_u^{ri}}^{-1}(\lambda) - \Phi_{T_v^{ri}}^{-1}(\lambda) = (t_u^{ri} - t_v^{ri}) + z_\lambda(\sigma_u^{ri} - \sigma_v^{ri}) < (t_u^{ri} - t_v^{ri}) + z_{1-\varepsilon}(\sigma_u^{ri} - \sigma_v^{ri}) < 0$. Therefore, $\Phi_{T_u^{ri}}^{-1}(\lambda) - \Phi_{T_v^{ri}}^{-1}(\lambda) < 0$ always holds for any $\lambda \in [\varepsilon, 1-\varepsilon]$. □

Using Lemma 4, the discretization on CDFs is no longer needed under the normal distributed assumption of path travel times. Then, using $\Phi_{T_u^{ri}}^{-1}(\varepsilon)$ and $\Phi_{T_u^{ri}}^{-1}(1-\varepsilon)$ as two criteria, the BSP-LC algorithm utilizes more efficient bi-criteria label-correcting algorithm

(Brumbaugh-Smith and Shier, 1989) to generate all FSD non-dominated paths in the network, so as to improve the computational performance. It should be noted that the BSP-LC algorithm only can be used for the scenario of path travel times following normal distributions. Nevertheless, the original multi-criteria label-correcting algorithm with discretization technique (Nie and Wu, 2009) can be used for any kind of travel time distribution (e.g., lognormal and gamma distributions).