International Journal of Geographical Information Science

ISSN: 1365-8816 (Print) 1362-3087 (Online) Journal homepage: https://www.tandfonline.com/loi/tgis20

Reliable shortest path finding in stochastic networks with spatial correlated link travel times

Bi Yu Chen , William H.K. Lam , Agachai Sumalee & Zhi-lin Li

International Journal of Geographical Information Science
Vol. 26, No. 2, February 2012, 365–386

Taylor & Francis
Taylor & Francis Group

Reliable shortest path finding in stochastic networks with spatial correlated link travel times

Bi Yu Chen[a,b]*, William H.K. Lam[a], Agachai Sumalee[a] and Zhi-lin Li[c]

[a]*Department of Civil and Structural Engineering, The Hong Kong Polytechnic University, Hong Kong, PR China;* [b]*State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, PR China;* [c]*Department of Land Surveying & Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong, PR China*

(*Received 9 February 2011; final version received 15 June 2011*)

This article proposes an efficient solution algorithm to aid travelers' route choice decisions in road network with travel time uncertainty, in the context of advanced traveler information systems (ATIS). In this article, the travel time of a link is assumed to be spatially correlated only to the neighboring links within a local 'impact area.' Based on this assumption, the spatially dependent reliable shortest path problem (SD-RSPP) is formulated as a multicriteria shortest path-finding problem. The dominant conditions for the SD-RSPP are established in this article. A new multicriteria A* algorithm is proposed to solve the SD-RSPP in an equivalent two-level hierarchical network. A case study using real-world data shows that link travel times are, indeed, only strongly correlated within the local impact areas; and the proposed limited spatial dependence assumption can well approximate path travel time variance when the size of the impact area is sufficiently large. Computational results demonstrate that the size of the impact area would have a significant impact on both accuracy and computational performance of the proposed solution algorithm.

Keywords: reliable shortest path problem; travel time uncertainty; spatial correlation; travel time reliability

1. Introduction

Shortest path problems have been intensively studied owing to their broad applications in various science and engineering disciplines. In the field of geographical information science (GIS), substantial attention has been given to the development of efficient shortest path algorithms for route guidance systems (RGS) (Huang *et al.* 2007, Zeng and Church 2009). The value of RGS is most evident when real-time traffic information, generated by advanced traveler information systems (ATIS), is incorporated. It has been recognized that ATIS-based RGS can not only help travelers to make better route choice decisions in congested road networks, but also improve overall network traffic conditions (Li *et al.* in press).

Most existing RGS assume that link travel times are deterministic. However, link travel times in urban road networks are highly stochastic, due to random traffic demand fluctuations and capacity degradations (Li *et al.* in press). In addition, link travel times are

*Corresponding author. Email: chen.biyu@gmail.com

ISSN 1365-8816 print/ISSN 1362-3087 online

http://dx.doi.org/10.1080/13658816.2011.598133
http://www.tandfonline.com

spatially correlated (Chan *et al.* 2009). For example, a traffic accident on a major urban road may also cause significant travel delays on that road's upstream links. The travel time variations and spatial correlations have been measured (in terms of variance–covariance matrices) in ATIS as an important data source for travel time estimation and prediction (Tam and Lam 2008, Chan *et al.* 2009, El Esawey and Sayed 2011). Therefore, travelers' path travel time is not deterministic but a random variable, the distribution of which is a joint distribution of all link travel times along that path.

Many empirical studies have found that travel time uncertainties have a significant impact on travelers' route choice behavior (Lam and Small 2001, Tam *et al.* 2008). These studies revealed that travelers indeed consider travel time uncertainties as a risk, when planning for important events. Clearly large travel time variations may cause late arrivals and the subsequent imposition of high penalties for travelers (e.g., missed flights). As a result, travelers tend to depart from the journey origin early, so that they can arrive at the destination with a given on-time arrival probability, termed 'travel time reliability' in the literature. Hence travel time reliability concerns are necessary inclusions in sophisticated ATIS-based RGS applications.

In order to solve the reliable shortest path problem (RSPP), Frank (1969) introduced the concept of the most reliable path, in other words, a path that maximizes on-time arrival probability for a given travel time budget. As an alternative, Chen and Ji (2005) proposed the concept of the alpha-shortest path, the aim of which is to minimize the travel time budget required to ensure a given on-time arrival probability threshold.

Based on the above two definitions, many solution algorithms have been developed for solving the RSPP in road networks with stochastic and independent link travel times. Shao *et al*. (2004) proposed a heuristic method for finding the most reliable path, based on the relationship between link travel time mean and standard deviation. Chen and Ji (2005) presented a simulation-based genetic algorithm for finding the most reliable path and the alpha-shortest path. Lim (2008) and Nikolova (2009) developed a parametric approach to determine the most reliable path for risk-averse travelers. However, this method cannot provide a solution for risk-seeking travelers, whose travel time budget is less than the least expected travel time between origin–destination (O–D) nodes. Nie and Wu (2009a) proposed a label-correcting algorithm (dynamic programming approach) to find the most reliable path by generating all nondominated paths under the first-order stochastic dominant (FSD) condition. However, spatial correlations of link travel times were not considered in these studies.

To writers' knowledge, the spatially dependent RSPP (SD-RSPP) has not received much attention in the literature. Nie and Wu (2009b) studied the SD-RSPP by assuming that the probability density function (PDF) of the travel time of a link is conditional on the state of travelers arriving at the tail node of the link. The FSD condition and the dynamic programming approach can be employed, based on this assumption, to solve the SD-RSPP. However, it is a formidable task for ATIS to generate such probabilities for a large number of links in real road networks. As an alternative approach, Ji *et al.* (2011) formulated the spatial correlations as variance–covariance matrices which can be directly obtained from ATIS. Based on this formulation, a simulation-based method was proposed to solve the SD-RSPP. Nevertheless, the simulation-based method is computationally expensive and the precision of results is dependent on the maximum number of simulations. Using the same formulation of variance–covariance matrix, Nikolova (2009) proposed a network transformation technique to solve the SD-RSPP using the parametric approach. However, this parametric method cannot provide a solution for risk-seeking travelers; and such method may encounter infinite negative cycles in the transformed network due to the negative travel time covariance.

The aim of this article is to investigate the SD-RSPP in the context of ATIS-based RGS applications, so as to aid various travelers (including risk-averse, risk-neutral, and risk-seeking travelers) make their route choice decisions under travel time uncertainties. Thus, the link travel time correlations are represented by variance–covariance matrices. In this article, the travel time of a link is assumed to be spatially correlated only with its neighboring links within a local 'impact area', in which the topological distance (measured by number of links) between any two links is less than or equal to k. The consideration of this k-limited spatial dependence in the SD-RSPP extends the work of Nie and Wu (2009b) and Nikolova (2009), which considers only travel time correlations on the adjacent links ($k = 1$). To some extent, such k-limited spatial dependence can be interpreted as Tobler's First Law of Geography that 'all things are related, but nearby things are more related than distant things' (Tobler 1970). Empirical studies based on field observations also found travel times, among adjacent links, to be strongly correlated (Gajewski and Rilett 2003). The correlation is usually very low for links that are spatially distant, even on the same street (El Esawey and Sayed 2011). Based on this assumption, an efficient solution algorithm is proposed in this article to exactly solve the SD-RSPP. The main contributions of this article are summarized as follows.

(1) Several dominant conditions for the SD-RSPP are established, in which travel time correlations among neighboring links are considered and represented by variance–covariance matrices. These established dominant conditions enable the use of efficient dynamic programming approaches to solve the SD-RSPP. Thus, the established dominant conditions have important implications regarding the algorithm design.
(2) A new multicriteria A* algorithm is proposed to solve the SD-RSPP in an equivalent two-level hierarchical network. The proposed hierarchical network can well represent travel time correlations among neighboring links and facilitate reliable shortest path findings based on established dominant conditions. Based on this hierarchical network, an efficient multicriteria A* algorithm is proposed to solve exactly the SD-RSPP. The optimality of the proposed algorithm can be formally proved. Therefore, the proposed algorithm has significant advantages over previous simulation-based method.
(3) A comprehensive case study using data from a real-world ATIS in Hong Kong is carried out. The numerical results show that link travel times are strongly correlated only within local impact areas. The proposed limited spatial dependence assumption can well approximate path travel time standard deviation when the size of impact area is sufficiently large. Computational results demonstrate that the size of the impact area has significant impacts on both the accuracy and computational performance of the proposed solution algorithm. These key findings from the case study have useful implications in the development of RGS applications.

The rest of this article is organized as follows. The definition of the SD-RSPP used in this research is presented in Section 2. The dominant conditions and properties of the SD-RSPP are introduced in Section 3. The proposed two-level hierarchical network is presented in Section 4. The solution algorithm for solving the SD-RSPP is given in Section 5. The numerical examples using data from a real-world ATIS in Hong Kong are reported in Section 6. Finally, conclusions and recommendations for further study are given in Section 7.

2. Problem statement

Let $G = (N, A, \Psi)$ be a directed network consisting of a set of nodes N, a set of links A, and a set of movements Ψ. 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}. The mean and standard deviation (SD) of link travel time are denoted by t_{ij} and σ_{ij}, respectively. Each node i has a set of successor nodes $SCS(i) = \{j : a_{ij} \in A\}$ and a set of predecessor nodes $PDS(i) = \{k : a_{ki} \in A\}$. The movement $\psi_{ijk} = (a_{ij}, a_{jk}) \in \Psi$ represents an allowed movement (e.g., through-movement or right-turn) at node j. A movement $\psi_{ijk} \notin \Psi$ means that the movement is restricted at node j (e.g., no left-turn or no U-turn).

Suppose that the nodes $r \in N$ and $s \in N$ represent the O–D nodes. Let $p_u^{rs} = \{a^1, \ldots, a^m, \ldots, a^\lambda\}$ be a path from the origin r to the destination s, consisting of λ consecutive links. The path travel time, denoted by T_u^{rs}, is the sum of the related link travel times along the path as

$$T_u^{rs} = \sum_{m=1}^{\lambda} T_{ij}^m \tag{1}$$

where T_{ij}^m is the travel time distribution of a^m (the mth link along path p_u^{rs}).

As previously indicated, the travel time of link a_{ij} in this article is assumed to be spatially correlated only within a local impact area, denoted by $G_{ij}^k = \left(N_{ij}^k, A_{ij}^k, \Psi_{ij}^k\right)$. Let d_{ij}^{qw} be the topological distance (measured by number of links) between links a_{ij} and a_{qw}. A link a_{qw} is said to be a k-neighboring link of link a_{ij} if and only if $d_{ij}^{qw} = k$. With this concept, the impact area G_{ij}^k can be formally defined as a sub-network of G, satisfying $d_{ij}^{qw} \le k$, $\forall a_{qw} \in A_{ij}^k$.

In this article, link travel times are also assumed to follow normal distributions. The normality assumption of link travel times is common in the studies of stochastic shortest path problems (Chang *et al.* 2005). Recent empirical studies based on field observations found that the use of normal distributions appears to reflect observed travel time distributions (Rakha *et al.* 2006), and the normality assumption could be sufficient from a practical standpoint, given its computational simplicity (Lim 2008). Under these two assumptions, the path travel time T_u^{rs} follows a multivariate normal distribution. Its mean and SD, respectively, denoted by t_u^{rs} and σ_u^{rs}, can be calculated as

$$t_u^{rs} = \sum_{m=1}^{\lambda} t_{ij}^m \tag{2}$$

$$\sigma_u^{rs} = \sqrt{\sum_{m=1}^{\lambda} (\sigma^m)^2 + \sum_{n=1}^{k} \sum_{m=1}^{\lambda-n} 2\mathrm{cov}(a^m, a^{m+n})} \tag{3}$$

where $\mathrm{cov}(a^m, a^{m+n})$ is the travel time covariance between links a^m and a^{m+n}.

Let $\Phi_{rs,u}^{-1}(\alpha)$ be the inverse of the cumulative distribution function (CDF) of path travel time T_u^{rs} at α confidence level. It can be expressed as

$$\Phi_{rs,u}^{-1}(\alpha) = t_u^{rs} + z_\alpha \sigma_u^{rs} \tag{4}$$

where z_α is the inverse CDF of standard normal distribution at α confidence level. The confidence level $\alpha \in (0, 1)$ is the probability of travelers arriving at the destination within the travel time budget $\Phi_{rs,u}^{-1}(\alpha)$. The on-time arrival probability α represents travelers' attitudes toward risks of being late ($\alpha > 0.5$, $\alpha = 0.5$, and $\alpha < 0.5$ for risk-averse, risk-neutral, and risk-seeking attitudes, respectively). The value of α can be predetermined based on travelers' trip purposes.

According to Chen and Ji (2005), the problem of finding the reliable shortest path which minimizes the travel time budget $\Phi_{rs,u}^{-1}(\alpha)$, required to ensure a given on-time arrival probability α, can be formally expressed as the following optimization problem

$$\underset{x_{ij}^{rs}}{\text{Min}} \quad \Phi_{rs,u}^{-1}(\alpha) \tag{5}$$

subject to

$$T_u^{rs} = \sum_{a_{ij} \in A} T_{ij} x_{ij}^{rs} \tag{6}$$

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

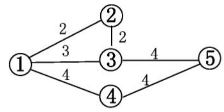
$$x_{ij}^{rs} \in \{0,\ 1\}, \quad \forall a_{ij} \in A \tag{8}$$

$$\psi_{ijk} \in \Psi, \quad \forall \psi_{ijk} \in p_u^{rs} \tag{9}$$

where x_{ij}^{rs} is the decision variable regarding the link–path incidence relationship; $x_{ij}^{rs} = 1$ means that the link a_{ij} is on the path p_u^{rs}, and otherwise $x_{ij}^{rs} = 0$. Equation (5) represents the travel time budget which travelers want to minimize. Equation (6) defines the path travel time as mentioned in Equations (1)–(3). Equation (7) ensures that the reliable shortest path is feasible. Equation (8) is concerned with the link–path incidence variables which should be binary in nature. Equation (9) ensures that all movements along the reliable shortest path are feasible.

The above formulation of SD-RSPP can be regarded as a generalization of the traditional shortest path problem. The optimal solution of the SD-RSPP depends on predetermined on-time arrival probability α. For risk-neutral scenarios ($\alpha = 0.5$), $Z_\alpha = 0$ and thus the SD-RSPP becomes the traditional shortest path problem, which is, namely, to find the path with least mean travel time t_u^{rs}. For risk-averse scenarios ($\alpha > 0.5$), travelers make route choice decisions by simultaneously minimizing t_u^{rs} and σ_u^{rs}. For risk-seeking scenarios ($\alpha < 0.5$), travelers tend to choose the optimal path with smaller t_u^{rs} but larger σ_u^{rs}.

Figure 1 illustrates the above concept, by means of a small network. In this figure, all link travel times follow normal distributions. The mean link travel times are shown on the links while the link travel time variance and covariance are given in the matrix. In the variance–covariance matrix, elements along the diagonal are the variance of link travel times and off-diagonal elements are the travel time covariance between two links. As the matrix is symmetric, only a lower triangular matrix is shown in the figure.



α	$\Phi^{-1}_{15,1}(\alpha)$	$\Phi^{-1}_{15,2}(\alpha)$	$\Phi^{-1}_{15,3}(\alpha)$
0.1	5.78	2.95	**2.57**
0.5	8	**7**	8
0.9	**10.22**	11.05	13.43

$$\begin{array}{c|cccccc|} & a_{12} & a_{13} & a_{14} & a_{23} & a_{35} & a_{45} \\ a_{12} & 2 & & & & & \\ a_{13} & -1 & 1 & & & & \\ a_{14} & -1 & -0.5 & 1 & & & \\ a_{23} & 2 & -1 & -0.3 & 2 & & \\ a_{35} & 0.3 & 1.5 & -0.4 & 2 & 6 & \\ a_{45} & -0.2 & -0.6 & 0.5 & -0.4 & -1.5 & 1 \end{array}$$

$$\Phi^{-1}_{15,1}(\alpha) = \Phi^{-1}_{a_{14}\oplus a_{45}}(\alpha) = 8 + Z_\alpha\sqrt{1+2*0.5+1}$$

$$\Phi^{-1}_{15,2}(\alpha) = \Phi^{-1}_{a_{13}\oplus a_{35}}(\alpha) = 7 + Z_\alpha\sqrt{1+2*1.5+6}$$

$$\Phi^{-1}_{15,3}(\alpha) = \Phi^{-1}_{a_{12}\oplus a_{23}\oplus a_{35}}(\alpha) = 8 + Z_\alpha\sqrt{2+2*2+2+2*2+6}$$

Figure 1. An illustrative example.

As shown in Figure 1, three paths, $p_1^{15} = a_{14} \oplus a_{45}$, $p_2^{15} = a_{13} \oplus a_{35}$, and $p_3^{15} = a_{12} \oplus a_{23} \oplus a_{35}$, are available from Node 1 to Node 5, where $\oplus$ is a path concatenation operator ($p_1^{15} = a_{14} \oplus a_{45}$ means that p_1^{15} goes through a_{14} and a_{45}). When $\alpha = 0.1$ ($Z_\alpha = -1.28$), travelers prefer to take a risk by using path p_3^{15} which has a large travel time variation so as to assign a small travel time budget $\Phi^{-1}_{15,3}(0.1) = 2.57$. When $\alpha = 0.5$ ($Z_\alpha = 0$), risk-neutral travelers prefer to choose path p_2^{15} with the least mean travel time $t_2^{15} = 7$ and set this mean travel time as the time budget for their travel. When $\alpha = 0.9$ ($Z_\alpha = 1.28$), travelers become risk-averse. They tend to use the more reliable path p_1^{15} with a small travel time standard deviation and assign a larger travel time budget $\Phi^{-1}_{15,1}(0.9) = 10.22$. Therefore, the optimal solution of the SD-RSPP depends on predetermined travelers' attitudes toward risk of being late (α).

It can also be observed from Figure 1 that the SD-RSPP is nonadditive, as the travel time budget cannot be calculated by the sum of the related link costs. For example, when $\alpha = 0.9$, the cost of path $p_1^{15} = a_{14} \oplus a_{45}$ is not equal to the aggregate cost of links a_{14} and a_{45}, $\left(\Phi^{-1}_{15,1}(0.9) = 10.22 < \Phi^{-1}_{a_{14}}(0.9) + \Phi^{-1}_{a_{45}}(0.9) = 10.56\right)$. In this case, the SD-RSPP cannot be solved easily by traditional shortest path algorithms (e.g., Dijkstra's algorithm) which build on the additive assumption. In following sections, the SD-RSPP is solved using multicriteria shortest path approach.

3. Multicriteria shortest path approach for solving SD-RSPP

The SD-RSPP can be formulated and solved as a multicriteria shortest path problem where the additive mean and variance of link travel times are considered as two independent variables. The solution algorithms, for solving the multicriteria shortest path problem, typically rely on the dominant conditions to determine a set of Pareto-optimal paths (or nondominated paths) instead of single optimal path. For the SD-RSPP, the nondominated paths can be defined as follows.

Definition 3.1 Given an on-time arrival probability α and two paths $p_u^{rj} \neq p_v^{rj} \in P^{rj}$, p_u^{rj} dominates p_v^{rj} (denoted by $p_u^{rj} \succ p_v^{rj}$), if and only if $\Phi^{-1}_{p_u^{rj}\oplus p^{j\ell}}(\alpha) < \Phi^{-1}_{p_v^{rj}\oplus p^{j\ell}}(\alpha), \forall p^{j\ell} \in P^{j\ell}, \forall \ell \in N$.

Definition 3.2 A path $p_u^{rj} \in P^{rj}$ is a nondominated path, if and only if p_u^{rj} is not dominated by $\forall p_v^{rj} \in P^{rj}$.

Based on the above two definitions, it is obvious that the reliable shortest path is one of the nondominated paths between O–D nodes. The nondominated paths can be found by recursive path extensions from the origin to the destination using generalized dynamic programming approach (Carraway *et al.* 1990). During this path search process, all dominated paths can be eliminated without further consideration, since they cannot be parts of the reliable shortest path between the O–D nodes. In the following, dominant conditions to identify dominated paths are introduced, when limited travel time correlations are considered.

Let $p^{ij,k} = \{a^1, \ldots, a^k\}$ be a path from node i to node j consisting of k consecutive links, and $p_u^{rj,k+\lambda} = p_u^{ri,\lambda} \oplus p^{ij,k}$ be a path from origin r to node j going through the sub-path $p^{ij,k}$. The FSD condition for the SD-RSPP can be formally defined as below.

Proposition 3.1 (FSD condition). Given two paths $p_u^{rj,\lambda+k} \neq p_v^{rj,\eta+k} \in P^{rj}$, $p_u^{rj,\lambda+k} \succ p_v^{rj,\eta+k}$ if $p_u^{rj,k+\lambda}$ and $p_u^{rj,k+\eta}$ satisfy $\Phi^{-1}_{p_u^{rj}}(y) < \Phi^{-1}_{p_v^{rj}}(y)\ \forall y \in (0, 1)$.

Proof See Appendix. □

In addition to the FSD condition, following mean–variance (M–V) dominant condition exists:

Proposition 3.2 (M–V dominant condition) Given an on-time arrival probability α and two paths $p_u^{rj,\lambda+k} \neq p_v^{rj,\eta+k} \in P^{rj}$, $p_u^{rj,\lambda+k} \succ p_v^{rj,\eta+k}$ if $p_u^{rj,k+\lambda}$ and $p_u^{rj,k+\eta}$ satisfy either

(1) $t_u^{rj} \leq t_v^{rj}$ and $Z_\alpha \sigma_u^{rj} < Z_\alpha \sigma_v^{rj}$ or
(2) $t_u^{rj} < t_v^{rj}$ and $Z_\alpha \sigma_u^{rj} \leq Z_\alpha \sigma_v^{rj}$

Proof See Appendix. □

The FSD condition and the M–V dominant condition are interrelated. The FSD can identify dominated paths satisfying $\Phi^{-1}_{p_u^{rj}}(y) < \Phi^{-1}_{p_v^{rj}}(y)\ \forall y \in (0, 1)$. These FSD-dominated paths can be discarded without further consideration in the path search process regardless of travelers' desired on-time arrival probability (α). In contrast to the FSD condition, the M–V dominant condition can be adopted only when the on-time arrival probability is pre-given. For risk-averse travelers ($\alpha > 0.5$), M–V-dominated paths, satisfying $\Phi^{-1}_{p_u^{rj}}(y) < \Phi^{-1}_{p_v^{rj}}(y)\ \forall y \in [0.5, 1)$, can be identified. For risk-seeking travelers ($\alpha < 0.5$), M–V-dominated paths can be determined when $\Phi^{-1}_{p_u^{rj}}(y) < \Phi^{-1}_{p_v^{rj}}(y)\ \forall y \in (0, 0.5]$ is satisfied.

Figure 2 illustrates the above two established dominant conditions in a simple network when $k = 3$ is considered. As shown in this figure, four paths from Node 1 to Node 8 go through the same sub-path $p^{58,3} = a_{56} \oplus a_{67} \oplus a_{78}$ with three links. The mean and travel time standard deviation of these four paths are given in Figure 2a, and the CDF of four path travel time distributions are illustrated in Figure 2b. It can be observed from Figure 2b that the path $p_4^{18,5}$ is FSD dominated by the path $p_1^{18,5}$ since $\Phi^{-1}_{p_1^{18}}(y) < \Phi^{-1}_{p_4^{18}}(y)\ \forall y \in (0, 1)$. Thus, for all travelers with different on-time arrival probabilities, $p_4^{18,5}$ can be discarded in the path search process. The other three paths ($p_1^{18,5}$, $p_2^{18,5}$, and $p_3^{18,6}$) should be maintained as FSD nondominated paths. If a traveler is risk-averse, $p_3^{18,6}$ can be further eliminated as an M–V-dominated path since $\Phi^{-1}_{p_1^{18}}(y) < \Phi^{-1}_{p_3^{18}}(y)\ \forall y \in [0.31, 1)$. Similarly, according to the M–V dominant condition, the path $p_2^{18,5}$ can be also determined as the M–V-dominated path for a risk-seeking traveler, since $\Phi^{-1}_{p_1^{18}}(y) < \Phi^{-1}_{p_2^{18}}(y)\ \forall y \in (0, 0.68]$. Therefore, with the given on-time arrival probability, the M–V dominant condition can help determine potential dominated paths which may not be identified under the FSD condition.

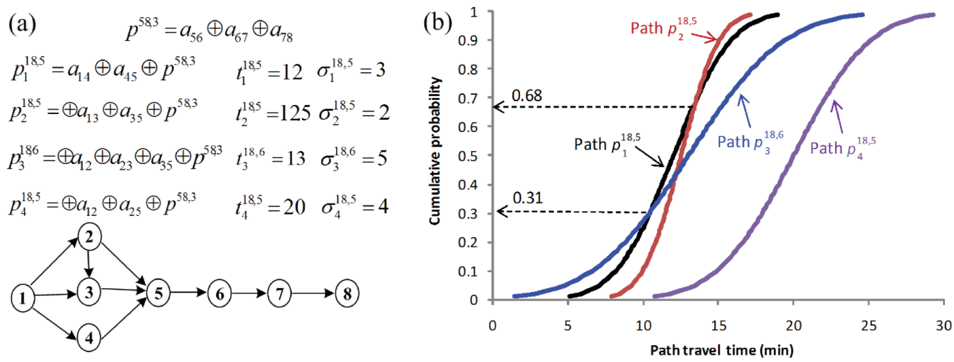


Figure 2. An illustration of dominant conditions (a) a simple network (b) path travel time distributions.

Another important SD-RSPP property is that the values of objective function ($\Phi_{rj,u}^{-1}(\alpha)$) are monotonically increasing during the path search process as follows:

Proposition 3.3 Given two paths p_u^{rj} and $p_u^{r\ell} = p_u^{rj} \oplus a_{j\ell}$, the relationship $\Phi_{r\ell,u}^{-1}(\alpha) > \Phi_{rj,u}^{-1}(\alpha)\ \forall\alpha$ always holds.

Proof $\Phi_{r\ell,u}^{-1}(\alpha) - \Phi_{rj,u}^{-1}(\alpha) = t_{j\ell} + Z_\alpha\left(\sqrt{(\sigma_u^{rj})^2 + 2\mathrm{cov}(p_u^{rj}, a_{j\ell}) + \sigma_{j\ell}^2} - \sigma_u^{rj}\right)$. When $\alpha \geq 0.5$, we have $\Phi_{r\ell,u}^{-1}(\alpha) - \Phi_{rj,u}^{-1}(\alpha) \geq t_{j\ell} + Z_\alpha\left(\sqrt{(\sigma_u^{rj} - \sigma_{j\ell})^2} - \sigma_u^{rj}\right) = t_{j\ell} - Z_\alpha\sigma_{j\ell}$. Due to the nonnegative property of travel time of link $a_{j\ell}$, we have $t_{j\ell} - Z_\alpha\sigma_{j\ell} > 0$ and thus $\Phi_{r\ell,u}^{-1}(\alpha) - \Phi_{rj,u}^{-1}(\alpha) > 0$. Similarly, we can prove $\Phi_{r\ell,u}^{-1}(\alpha) - \Phi_{rj,u}^{-1}(\alpha) > 0$ when $\alpha < 0.5$. □

It is well known that the optimal path in traditional shortest path problems is acyclic without passing the same node more than once. In the SD-RSPP, the following acyclic property holds. For convenience, a path without passing the same link twice is hereafter referred to as the acyclic path in the SD-RSPP.

Proposition 3.4 The reliable shortest path must not pass the same link more than once.

Proof Suppose $p_u^{rs} = p_u^{ri} \oplus a_{ij} \oplus \cdots \oplus a_{ij} \oplus p_u^{js}$ is the reliable shortest path passing the link a_{ij} twice. There exists a path $p_v^{rs} = p_u^{ri} \oplus a_{ij} \oplus p_u^{js}$ passing the link a_{ij} only once. Since p_u^{rs} passes more links in sub-path $a_{ij} \oplus \cdots \oplus a_{ij}$, we have $\Phi_{rs,u}^{-1}(\alpha) > \Phi_{rs,v}^{-1}(\alpha)\ \forall\alpha$ according to Proposition 3.3. Therefore, p_u^{rs} is not the reliable shortest path according to Definition 3.1, contradicting the assumption that p_u^{rs} is the reliable shortest path. □

4. A two-level hierarchical network

In this section, a two-level hierarchical network is proposed to represent road networks with spatial correlated link travel times. For clarity, the network G presented in Section 2 is hereafter referred to as the primal network.

As previously indicated, the SD-RSPP can be formulated using a generalized dynamic programming approach, and a set of nondominated paths is maintained and evaluated at the same sub-path with k consecutive links in the primal network. This dynamic programming approach, however, may not be easily implemented in the primal network, because the sub-paths with k links are not explicitly represented in this network.

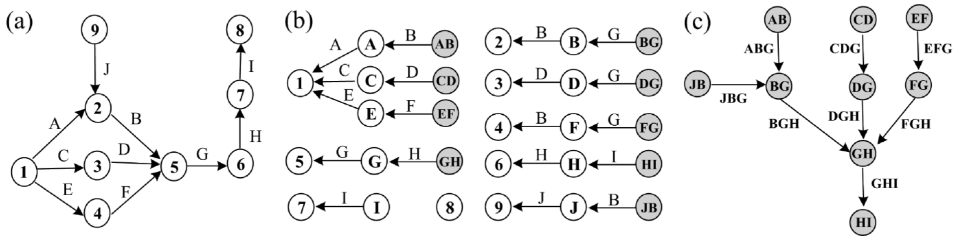


Figure 3. An illustration of hierarchical network: (a) primal network G, (b) ground hierarchy H^g and (c) top hierarchy H^t.

To facilitate such a path search approach, a two-level hierarchical network, denoted by $HG = \left(H^g, H^t\right)$, is proposed using k consecutive primal links as basis network elements. The proposed hierarchical network has two hierarchies. The ground hierarchy H^g consists of |N| directed-in-trees, where |N| is the number of nodes in the primal network G. For each primal node $i \in G$, a directed-in-tree $\bar{G}_i^g = (\bar{N}_i, \bar{A}_i)$is constructed rooting at this primal node. In each directed-in-tree $\bar{G}_i^g$, a ground hierarchical node $\bar{n}_u^{ij,\lambda} \in \bar{G}_i^g$ represents a primal path $p_u^{ij,\lambda} \in G$ with λ consecutive primal links ($\lambda \leq k-1$ always holds). In this way, all primal paths $\forall p_u^{ij,\lambda} \in G$ with λ links ($\lambda \leq k-1$) can be represented in the ground hierarchy.

Figure 3b illustrates the construction of the ground hierarchy, when $k = 3$, from the primal network, shown in Figure 3a. As illustrated in Figure 3b, there are nine directed-in-trees constructed for all nodes in the primal network. In Figure 3b, a node in these trees corresponds to a path in the primal network. For example, node $\bar{n}_{AB}$ in Figure 3b represents the path $p_1^{15,2} = a_A \oplus a_B$ in the primal network.

The nodes in the ground hierarchy can be classified into two categories: border nodes and local nodes. The border nodes for a tree $\bar{G}_i^g$, denoted by $BORDER(\bar{G}_i^g) = \{\forall \bar{n}_u^{ij,\lambda}, \lambda = k-1\}$, are defined as those nodes responding to $k-1$ consecutive primal links. The other nodes in the tree $\bar{G}_i^g$ can be defined as local nodes denoted by $LOCAL(\bar{G}_i^g) = \bar{N}_i / BORDER(\bar{G}_i^g)$. In the example shown in Figure 3b, nine nodes (denoted in gray) are classified as border nodes, since they correspond to primal paths in Figure 3a with two links ($\lambda = k-1 = 2$). These nine border nodes can be pushed into the top hierarchy as nodes in Figure 3c. The other nodes ($\lambda = 0, 1$) in Figure 3b are classified as local nodes without consideration in the top hierarchy.

The top hierarchy $H^t \in HG$ has only one network, denoted by $\hat{G}^t = \left(\hat{N}, \hat{A}, \hat{\Psi}\right)$. As mentioned, all top hierarchical nodes are the border nodes from the ground hierarchy, denoted as $\hat{N} = \{BORDER(\bar{G}_i^g), \forall \bar{G}_i^g\}$. Each top hierarchical link $\hat{a}_{iw,u} \in \hat{A}$ represents a primal path $p_u^{iw,k} = \{a_{ij}^1, \ldots, a_{\ell w}^k\}$ with k consecutive links. For the top hierarchical link $\hat{a}_{iw,u}$, its tail node $\bar{n}_u^{ij,k-1}$ and head node $\bar{n}_u^{iw,k-1}$, respectively, represent the first and last $k-1$ consecutive primal links of $p_u^{iw,k}$ (i.e., $p_u^{ij,k-1}$ and $p_u^{jw,k-1}$). Each top hierarchical movement $\hat{\psi}_{iq,u} = \{\hat{a}_{iw,u}, \hat{a}_{jq,u}\} \in \hat{\Psi}$ corresponds to a primal path $p_u^{iq,k+1} = \{a_{ij}^1, \ldots, a_{\ell w}^k, a_{wq}^{k+1}\}$ passing $k+1$ consecutive links. Similarly, this top hierarchical movement's tail link $\hat{a}_{iw,u}$ and head link $\hat{a}_{jq,u}$, respectively, represent the first and last k consecutive primal links of $p_u^{iq,k+1}$ (i.e., $p_u^{iw,k}$ and $p_u^{jq,k}$).

Figure 3c depicts the construction of the top hierarchy, where $k = 3$, from the same primal network in Figure 3a. As shown in Figure 3c, all nodes of the top hierarchical network are from the border nodes in the ground hierarchy. It can also be found from Figure 3 that the top hierarchical network consists of links and movements which, respectively, represent the paths in Figure 3a with three and four primal links. For instance, a top hierarchical movement $\hat{\psi}_{ABGH} = \{\hat{a}_{ABG}, \hat{a}_{BGH}\}$ in Figure 3c, corresponds to the primal path $p_1^{17,4} = a_A \oplus a_B \oplus a_G \oplus a_H$ in Figure 3a. This movement's tail link $\hat{a}_{ABG}$ represents the primal path $p_1^{16,3} = a_A \oplus a_B \oplus a_G$ in Figure 3a. It can be observed from Figure 3 that all acyclic primal paths ($k \leq 4$) in Figure 3a are represented in the hierarchical network (Figure 3b and c). Accordingly, this hierarchical network can be said to be a $k = 3$ complete dual network of the primal network in Figure 3a. The concept of the k complete dual network can be formally defined as follows.

Definition 4.1 A two-level hierarchical network $HG = \left(H^g, H^t\right)$ is a k complete dual network of the primal network G if and only if (i) any acyclic path $p_u^{ij,\lambda} \in G$ with $\lambda \leq k - 1$ links has a corresponding node $\bar{n}_u^{ij,\lambda} \in H^g$; (ii) any acyclic path $p_u^{iw,k} \in G$ with k links has a corresponding link $\hat{a}_{iw,u} \in H^t$; and (iii) any acyclic path $p_u^{iq,k+1} \in G$ with $k + 1$ links has a corresponding movement $\hat{\psi}_{iq,u} \in H^t$.

In the proposed hierarchical network, travel times and their correlations are stored in both ground and top hierarchies. In the ground hierarchy $H^g \in HG$, each node $\bar{n}_u^{ij,\lambda}$ maintains a travel time distribution $T_u^{ij,\lambda}$ for its corresponding primal path $p_u^{ij,\lambda} \in G$. In the top hierarchy $H^t \in HG$, each link $\hat{a}_{iw,u} \in H^t$ has a link travel time distribution $T_{iw,u}^{\hat{a}}$. Let $t_{iw,u}^{\hat{a}}$ and $\left(\sigma_{iw,u}^{\hat{a}}\right)^2$ be mean and variance of $T_{iw,u}^{\hat{a}}$, respectively. They can be calculated from the primal network as

$$t_{iw,u}^{\hat{a}} = t^k \tag{10}$$

$$\left(\sigma_{iw,u}^{\hat{a}}\right)^2 = \left(\sigma^k\right)^2 + \sum_{n=1}^{k-1} 2\text{cov}\left(a^n, a^k\right) \tag{11}$$

where a^n and a^k are nth and kth links in the corresponding primal path $p_u^{iw,k} \in G$, respectively. The travel time covariance between two adjacent top hierarchical links $\hat{a}_{iw,u}$ and $\hat{a}_{jq,u}$ is stored as an attribute of the corresponding top hierarchical movement $\hat{\psi}_{iq,u} = \{\hat{a}_{iw,u}, \hat{a}_{jq,u}\}$. This travel time covariance $\text{cov}(\hat{a}_{iw,u}, \hat{a}_{jq,u})$ can be expressed as

$$\text{cov}\left(\hat{a}_{iw,u}, \hat{a}_{jq,u}\right) = \text{cov}\left(a^1, a^{k+1}\right) \tag{12}$$

where a^1 and a^{k+1} are the first and last links of the movement's corresponding primal path $p_u^{iq,k+1} \in G$, respectively.

Let $p_u^{rs,\lambda+k-1} = \{a_{ri}^1, \ldots, a_{qs}^{\lambda+k-1}\} \in G$ be a primal path passing $\lambda + k - 1$ consecutive links. Obviously, the primal path $p_u^{rs,\lambda+k-1}$ contains λ primal sub-paths $\{p_u^{ri,k}, \ldots, p_u^{jw,k}, p_u^{\ell q,k}, \ldots, p_u^{\theta s,k}\}$with consecutive k links. These λ primal sub-paths correspond to λ top hierarchical links $\{\hat{a}_{ri,u}^1, \ldots, \hat{a}_{jw,u}^{m-1}, \hat{a}_{\ell q,u}^m, \ldots, \hat{a}_{\theta s,u}^\lambda\}$ in the top hierarchy. Let $\hat{p}_u^{rs,\lambda}$ be the top hierarchical path consisting of these λ top hierarchical links. Its mean and travel time variance, denoted by $t_{rs,u}^{\hat{p}}$ and $\left(\sigma_{rs,u}^{\hat{p}}\right)^2$, respectively are defined in this article as

$$t_{rs,u}^{\hat{p}} = t_u^{ri,k-1} + \sum_{m=1}^{\lambda} t_{\ell q,u}^{\hat{a},m} \tag{13}$$

$$\left(\sigma_{rs,u}^{\hat{p}}\right)^2 = \left(\sigma_u^{ri,k-1}\right)^2 + \sum_{m=1}^{\lambda}\left(\sigma_{\ell q,u}^{\hat{a},m}\right)^2 + \sum_{m=2}^{\lambda} 2\mathrm{cov}\left(\hat{a}_{jw,u}^{m-1}, \hat{a}_{\ell q,u}^{m}\right) \tag{14}$$

where $t_u^{ri,k-1}$ and $\left(\sigma_u^{ri,k-1}\right)^2$, respectively, are mean and travel time variance stored in the first node $\bar{n}_u^{ri,\lambda}$ of the top hierarchical path $\hat{p}_u^{rs,\lambda}$; and $\hat{a}_{jw,u}^{m-1}$ and $\hat{a}_{\ell q,u}^{m}$ are two adjacent links along the path $\hat{p}_u^{rs,\lambda}$. Using this setting, it can be proved that the top hierarchical path $\hat{p}_u^{rs,\lambda}$ and the primal path $p_u^{rs,\lambda+k-1}$ have an identical travel time distribution as below.

Proposition 4.1 Given a top hierarchical path $\forall \hat{p}_u^{rs,\lambda} \in H^t$, its travel time distribution is equivalent to that of the corresponding primal path $p_u^{rs,\lambda+k-1} \in G$.

Proof See Appendix. □

A simple path in Figure 3 can be used to illustrate the Proposition 4.1. A top hierarchical path $\hat{p}_1^{18,3} = \hat{a}_{ABG} \oplus \hat{a}_{BGH} \oplus \hat{a}_{GHI} \in H^t$ in Figure 3c corresponds to a primal path $p_1^{18,6} = a_A \oplus a_B \oplus a_G \oplus a_H \oplus a_I \in G$ in Figure 3a. The mean travel time of $\hat{p}_1^{18,3}$ is equivalent to that of $p_1^{18,6}$ as $t_{18,1}^{\hat{p}} = t_1^{15,2} + (t_{16,1}^{\hat{a},1} + t_{27,1}^{\hat{a},2} + t_{58,1}^{\hat{a},3}) = t_A + t_B + (t_G + t_H + t_I) = t_1^{18,5}$. The paths $\hat{p}_1^{18,3}$ and $p_1^{18,6}$ also have the same travel time variance as follows:

$$\begin{aligned}
\left(\sigma_{18,1}^{\hat{p}}\right)^2 &= \left(\sigma_1^{15,2}\right)^2 + \left(\sigma_{16,1}^{\hat{a},1}\right)^2 + \left(\sigma_{27,1}^{\hat{a},2}\right)^2 + \left(\sigma_{58,1}^{\hat{a},3}\right)^2 + 2\mathrm{cov}\left(\hat{a}_{16,1}, \hat{a}_{27,1}\right) \\
&\quad + 2\mathrm{cov}\left(\hat{a}_{17,1}, \hat{a}_{58,1}\right) \\
&= \left(\sigma_A^2 + \sigma_B^2 + 2\mathrm{cov}(a_A, a_B)\right) + \left(\sigma_G^2 + 2\mathrm{cov}(a_A, a_G) + 2\mathrm{cov}(a_B, a_G)\right) \\
&\quad + \left(\sigma_H^2 + 2\mathrm{cov}(a_B, a_H) + 2\mathrm{cov}(a_G, a_H)\right) \\
&\quad + \left(\sigma_I^2 + 2\mathrm{cov}(a_G, a_I) + 2\mathrm{cov}(a_H, a_I)\right) + (2\mathrm{cov}(a_A, a_H)) \\
&\quad + (2\mathrm{cov}(a_B, a_I)) = \left(\sigma_1^{18,5}\right)^2
\end{aligned}$$

With Proposition 4.1, following two important lemmas hold:

Lemma 4.1 A path extension $\hat{p}_u^{rq,m} = \hat{p}_u^{rw,m-1} \oplus \hat{a}_{\ell q,u}$ in top hierarchy is equivalent to the path extension $p_u^{rq,m+k-1} = p_u^{rw,m+k-2} \oplus a_{wq}$ in the primal network, where top hierarchical path $\hat{p}_u^{rw,m-1} \in H^t$ corresponds to the primal path $p_u^{rw,m+k-2} \in G$; top hierarchical link $\hat{a}_{\ell q,u} \in H^t$ corresponds to the primal path $p_u^{\ell q,k} \in G$; and primal link $a_{wq} \in G$ is the last link of the primal path $p_u^{\ell q,k}$.

Proof It can be easily followed by Proposition 4.1.

Lemma 4.2 Given a path $\forall p_u^{rs,\lambda} \in G$, it can be determined either a node $\bar{n}_u^{rs,\lambda} \in H^g$ or a path $\hat{p}_u^{rs,\lambda-k+1} \in H^t$ with the same travel time distribution as $p_u^{rs,\lambda} \in G$, if the hierarchical network $HG = \left(H^g, H^t\right)$ is a k complete dual network of the primal network G.

Proof When $\lambda \leq k - 1$, according to Definition 4.1, there exists a node $\bar{n}_u^{rs,\lambda} \in H^g$ representing the path $p_u^{rs,\lambda}$. According to the definition of the hierarchical network, the attribute stored at the node $\bar{n}_u^{rs,\lambda}$ is equivalent to travel time of $p_u^{rs,\lambda}$.

When $\lambda > k - 1$, $p_u^{rs,\lambda}$ contains $\lambda - k + 1$ sub-paths with consecutive k links. According to Definition 4.1, all these sub-paths, with consecutive k links, in the primal network are represented as links in the top hierarchy. Such top hierarchical links form a path $\hat{p}_u^{rs,\lambda-k+1}$. According to Proposition 4.1, $\hat{p}_u^{rs,\lambda-k+1}$ and $p_u^{rs,\lambda}$ have an identical path travel time distribution. □

With the above two lemmas, the SD-RSPP in the primal network can be equally solved in the proposed two-level hierarchical network using the generalized dynamic programming approach. As each top hierarchical link represents a primal path with k consecutive links, the nondominated paths passing the same k consecutive links (according to Propositions 3.2 and 3.3) can be directly maintained and evaluated at each top hierarchical link. In addition, as the path extension in the top hierarchy is equivalent to the path extension in the primal network, the monotonic increasing property (Proposition 3.3) and the acyclic property (Proposition 3.4) are also satisfied in the top hierarchical network. These properties contribute to the development of efficient solution algorithms given in the next section.

5. Solution algorithm

This section presents a multicriteria hierarchical A* algorithm, named *SDRSP-HA**, to solve the SD-RSPP in the proposed hierarchical network. Unless otherwise stated, the hierarchical network $HG = \left(H^g, H^t\right)$ used hereafter is a k complete dual network of the primal network. Similar to traditional A* algorithm (Zeng and Church 2009), the *SDRSP-HA** algorithm uses a heuristic valuation function $F\left(\hat{p}_u^{rj}\right) = \Phi_{rj,u}^{-1}(\alpha) + h(j)$ as a label for the top hierarchical path $\hat{p}_u^{rj} \in H^t$, where $h(j)$ is a travel time budget estimate from node $j \in G$ to destination $s \in G$, and $h(s) = 0$ at the destination. By using this heuristic valuation function, a higher priority can be assigned to the nodes closer to the destination, so as to reduce the number of examined nodes and speed up the search process.

As indicted above, the SD-RSPP in the primal network can equally be solved in the proposed two-level hierarchical network. Each top hierarchical link $\hat{a}_{ij} \in H^t$ represents a primal path $p^{ij,k} \in G$ with k consecutive links. Thus, M–V nondominated paths $p_u^{rj} = p_u^{ri} \oplus p_u^{ij,k} \in G$ passing the same $p_u^{ij,k}$ can be represented as corresponding top hierarchical paths $\hat{p}_u^{rj} = \hat{p}_u^{ri} \oplus \hat{a}_{ij} \in H^t$. Let $\widehat{P}^{rj} = \{\hat{p}_u^{rj}, \ldots, \hat{p}_v^{rj}\}$ be a set of nondominated paths maintained at the top hierarchical link $\hat{a}^{ij}$. The nondominated paths in $\hat{P}^{rj}$ are sorted in ascending order by mean travel time $t_{rj,u}^{\hat{p}}$. Nondominated paths from all top hierarchical links are maintained in a scan eligible set, denoted by $SE = \{\hat{p}_u^{rj}, \ldots, \hat{p}_v^{rw}\}$. The nondominated paths in SE are ordered by increasing value of the heuristic function, $F\left(\hat{p}_u^{rj}\right)$.

At each iteration, nondominated path $\hat{p}_u^{rj}$ at the top of SE (with minimum $F\left(\hat{p}_u^{rj}\right)$) is selected from SE for path extensions. A temporary acyclic path is constructed by extending the selected path $\hat{p}_u^{rj}$ to its successor link $\hat{a}_{qw} \in H^t$, denoted by $\hat{p}_u^{rw} := \hat{p}_u^{rj} \oplus \hat{a}_{qw}$. The dominant relationship between the newly generated path $\hat{p}_u^{rw}$ and the set of nondominated paths $\hat{P}^{rw}$ at link $\hat{a}_{qw}$ is determined using the M–V dominant condition (Proposition 3.2). If $\hat{p}_u^{rw}$ is a M–V nondominated path at link $\hat{a}_{qw}$, it is then inserted into $\hat{P}^{rw}$ and SE. The newly generated path $\hat{p}_u^{rw}$ may also dominate a set of paths in $\hat{P}^{rw}$, denoted by $\hat{P}_D^{rw}$. These dominated paths in $\hat{P}_D^{rw}$ can be eliminated from $\hat{P}^{rw}$ and SE. The algorithm continues this path search process until the destination is reached or SE becomes empty. The steps of *SDRSP-HA** algorithm are given below.

Algorithm *SDRSP-HA**

Inputs: O–D nodes (r, s) and on-time arrival probability α

Returns: the reliable shortest path

Step 1. Initialization:
For each border node $\bar{n}_u^{ri,k-1} \in \bar{G}_r$ ($\bar{G}_r \in H^g$ denotes the tree rooted at origin r)
For each top hierarchical link $\hat{a}_{ij}$ emanating from node $\bar{n}^{ri,k-1}$
Generate a new path $\hat{p}_u^{rj} := \bar{n}_u^{ri,k-1} \oplus \hat{a}_{ij}$ and calculate $h(j)$ and $F\left(\hat{p}_u^{rj}\right)$.
Set $\hat{P}^{rj} := \{\hat{p}_u^{rj}\}$ and $SE := SE \cup \{\hat{p}_u^{rj}\}$.
End for
End for
If destination $s \in \bar{G}_r$, then insert all paths $p_u^{rs} \in \bar{G}_r$ into SE.

Step 2. Path selection:
If $SE = \phi$, then Stop; otherwise, continue.
Select $\hat{p}_u^{rj}$ at the top of SE and set $SE := SE \backslash \{\hat{p}_u^{rj}\}$.
If $j = s$, then Stop; otherwise continue.

Step 3. Path extension:
For every movement $\hat{\psi}_{ijw} = \{\hat{a}_{ij,u}, \hat{a}_{qw}\}$ ($\hat{a}_{ij,u}$ denotes the last link of $\hat{p}_u^{rj}$)
Generate a new path $\hat{p}_u^{rw} := \hat{p}_u^{rj} \oplus \hat{a}_{qw}$ and calculate $h(w)$ and $F(\hat{p}_u^{rw})$.
If p_u^{rw} is acyclic, then continue; otherwise, scan next movement.
Call procedure $\hat{P}_D^{rw} := \text{CheckDominance}\left(\hat{p}_u^{rw}, \hat{P}^{rw}\right)$.
If $\hat{p}_u^{rw}$ is a nondominated path, then set $SE := SE \cup \{\hat{p}_u^{rw}\}$ and $SE := SE \backslash \hat{P}_D^{rw}$.
End for
Go to Step 2.

Procedure: CheckDomin ance

Inputs: A path $\hat{p}_u^{rj}$ and a set of nondominated path $\hat{P}^{rj}$

Returns: $\hat{P}_D^{rj}$ storing the set of paths dominated by $\hat{p}_u^{rj}$, and updated $\hat{P}^{rj}$

Step 1: Initialization
If $\alpha > 0.5$, then $\beta := 0.999$.
If $\alpha = 0.5$, then $\beta := 0.5$.
If $\alpha < 0.5$, then $\beta := 0.001$.
Set $\hat{P}_D^{rj} := \phi$ and $v := 1$.

Step 2: Dominant relationship determination
While $v \leq |\hat{P}_D^{rj}|$ and $t_u^{rj} > t_v^{rj}$ ($|\hat{P}_D^{rj}|$ is the number of paths in $\hat{P}_D^{rj}$)
If $\Phi_{rj,u}^{-1}(\beta) > \Phi_{rj,v}^{-1}(\beta)$, then return $\hat{P}_D^{rj}$.
Set $v := v + 1$.
End while
If $t_u^{rj} = t_v^{rj}$ and $\Phi_{rj,u}^{-1}(\beta) > \Phi_{rj,v}^{-1}(\beta)$, then return $\hat{P}_D^{rj}$.
Insert $\hat{p}_u^{rj}$ into $\hat{P}^{rj}$ at v^{th} position and set $v := v + 1$ (by default $|\hat{P}_D^{rj}| := |\hat{P}_D^{rj}| + 1$).
While $v \leq |\hat{P}_D^{rj}|$ and $\Phi_{rj,u}^{-1}(\beta) \leq \Phi_{rj,v}^{-1}(\beta)$
Set $\hat{P}^{rj} := \hat{P}^{rj} \backslash \{\hat{p}_v^{rj}\}$ and $\hat{P}_D^{rj} := \hat{P}_D^{rj} \cup \{\hat{p}_v^{rj}\}$.
Set $v := v + 1$.
End while
Return $\hat{P}_D^{rj}$.

The heuristic function $F(\hat{p}_u^{rj})$ is admissible if the following inequality is satisfied

$$F(\hat{p}_u^{rj} \oplus \hat{a}_{qw}) = \Phi_{rw,u}^{-1}(\alpha) + h(w) \geq F(\hat{p}_u^{rj}) = \Phi_{rj,u}^{-1}(\alpha) + h(j) \tag{15}$$

Equation (15) indicates that the heuristic function value of $F(\hat{p}_u^{rj})$ should monotonically increase with path extensions. A common admissible $h(j)$ is the Euclidean distance function $h(j) = e_{js}/V_{\max}$, where e_{js} is the Euclidean distance from node j to destination s and $V_{\max}$ is maximum travel speed (or design speed) of the network. If the heuristic function is admissible, it can be proved that the *SDRSP-HA** algorithm can obtain the optimal solution for the SD-RSPP as follows.

Proposition 5.1 If the heuristic function used is admissible, the *SDRSP-HA** algorithm can determine the reliable shortest path when the destination node is reached.

Proof Let $P^{rs} \in G$ be the set of paths containing all nondominated paths between O–D nodes. When destination node was reached, the path $\hat{p}_*^{rs}$ is selected from SE. The selected path $\hat{p}_*^{rs}$ can be either a node in the ground hierarchy or a path $\hat{p}_u^{rs} \in H^t$ in the hierarchical network. As at each iteration the path with minimum $F(\hat{p}_u^{rj})$ is selected from SE, the heuristic function value of $\hat{p}_*^{rs}$ $(F(\hat{p}_*^{rs}))$ is the minimum heuristic function value among all paths in SE. Since all paths in $\tilde{P}^{rs} = P^{rs} \backslash \{\hat{p}_*^{rs}\}$ are extended from SE and the heuristic function value monotonically increase with path extensions, the heuristic function value of $\hat{p}_*^{rs}$ is less than that of any path in $\tilde{P}^{rs}$. As $h(s) = 0$, $F(\hat{p}_*^{rs}) = \Phi_{rs,u}^{-1}(\alpha)$ is the minimum travel time budget in P^{rs} and thus $\hat{p}_*^{rs}$ is the reliable shortest path between O–D nodes. □

The performance of *SDRSP-HA** algorithm depends on the quality of $h(j)$ used. The better the travel time budget $h(j)$ estimates, the better the computational performance of the *SDRSP-HA** algorithm. When $h(j) = 0$, the *SDRSP-HA** algorithm reduces to the label-setting algorithm which uses travel time budget $\Phi_{rj,u}^{-1}(\alpha)$ as the heuristic function value for path $\hat{p}_u^{rj}$ instead of $F(\hat{p}_u^{rj})$.

With the implementation of *SE* using F-heap data structure (Fredman and Tarjan 1987), in worse case the label-setting algorithm requires $O\left(|\hat{\Psi}||\hat{P}|^2 + |\hat{A}||\hat{P}|Log(|\hat{A}||\hat{P}|)\right)$, where $|\hat{A}|$ and $|\hat{\Psi}|$ are the number of links and movements in the top hierarchical network, and $|\hat{P}|$ is maximum number of nondominated paths associated with a top hierarchical link. It is clear that this computational complexity depends on the size of $|\hat{P}|$. Theoretically, $|\hat{P}|$ grows exponentially with the network size. In practice, Nie and Wu (2009b) found that the number of nondominated paths is much smaller than the maximum possible size, especially for sparse transportation networks.

6. A case study in Hong Kong

A real-world case study, to demonstrate the applicability of the proposed solution algorithm, is described in this section. In Hong Kong, real-time traffic information on major urban roads is provided by a Real-time Travel Information System (RTIS) (http://tis.td.gov.hk/rtis/ttis/index/main_partial.jsp) (Tam and Lam 2008). In the RTIS, radio-frequency identification technology is adopted to collect real-time traffic data. Offline link travel times and variance–covariance matrices, generated by traffic flow simulators (Lam *et al.* 2002), are also adopted for the RTIS. With the use of real-time and offline traffic data, the RTIS can provide travel time estimates every 5 minutes for both links, either with or without real-time data.

As shown in Figure 4, the RTIS network consists of 1367 nodes, 3655 links, and 11,849 movements at road intersections. In this article, the RTIS data (including mean and variance–covariance matrix of link travel times) were collected at a morning peak hour (08:00–09:00) on 23 September 2010 (Thursday).

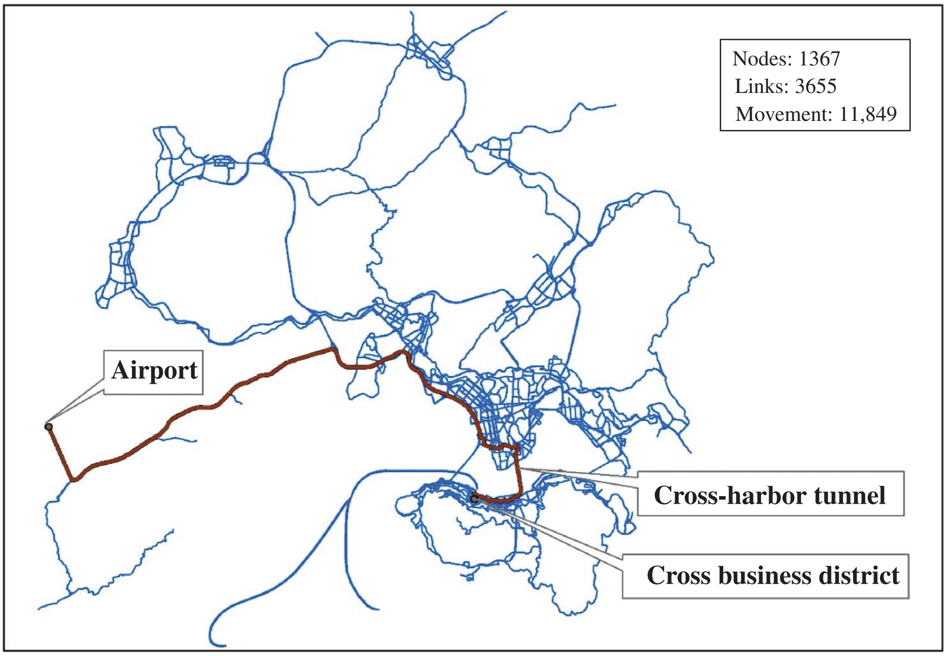


Figure 4. Hong Kong real-time travel information system (RTIS) network.

To normalize the link travel time covariance in the whole network, a correlation coefficient for every two links, a_{ij} and a_{qw}, was calculated as

$$\rho_{ij}^{qw} = \frac{\mathrm{cov}(a_{ij}, a_{qw})}{\sigma_{ij}\sigma_{qw}} \tag{16}$$

The value of ρ_{ij}^{qw} is between -1 and $+1$; $\rho_{ij}^{qw} = +1$ is the case of perfect positive correlation and $\rho_{ij}^{qw} = -1$ is the case of perfect negative correlation. Let NL_{ij}^{k} be the set of k-neighboring links for link a_{ij}. The mean absolute value (MAV) of correlation coefficients for all k-neighboring links, denoted by $E(\rho^k)$, can be calculated as

$$E(\rho^k) = \left(\frac{\sum\limits_{a_{ij}\in A} abs(p_{ij}^{qw})}{\sum\limits_{a_{ij}\in A} |NL_{ij}^{k}|} \right), \forall a_{qw} \in NL_{ij}^{k} \tag{17}$$

where $abs(p_{ij}^{qw})$ is the absolute value of ρ_{ij}^{qw} for links a_{ij} and a_{qw}; and $|NL_{ij}^{k}|$ is the number of k-neighboring links for link a_{ij}. The value of $E(\rho^k)$ is between 0 and $+1$. The larger the $E(\rho^k)$ value, the stronger the travel time correlations among k-neighboring links. The $E(\rho^k)$ value can be adopted as an indicator to measure travel time correlations among k-neighboring links.

Figure 5 shows such $E(\rho^k)$ values for the RTIS network during the peak hour. In this figure, the x-axis refers to k values and the y-axis at the left hand side represents $E(\rho^k)$ values. It can be seen from Figure 5 that with the increase of the k value the travel time

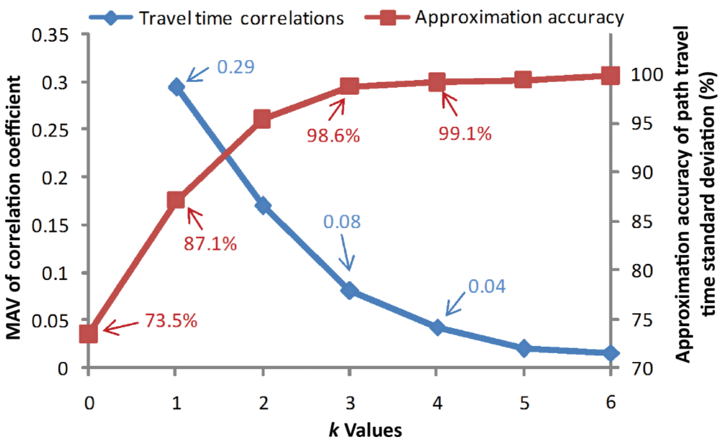


Figure 5. Travel time correlations among k-neighboring links.

correlations significantly decrease. For instance, when the k value increases from 1 to 4, $E(\rho^k)$ decreases from 0.29 to 0.04, a reduction of about 86.2%. This observation is consistent with the empirical findings of previous studies (Gajewski and Rilett 2003): the travel time correlation is usually very low for links that are spatially distant.

From the above observation, the k-limited spatial dependence assumption, that the travel time of a link correlates only with its neighboring links within a local impact area, seems valid. To justify this assumption, the approximation accuracy of path travel time standard deviation (SD) was examined using a typical cross-harbor journey from the central business district (CBD) to Hong Kong International Airport (HKIA). As shown in Figure 4, the route passing the cross-harbor tunnel (CHT) was selected for this case study, because the CHT was proved to be the most frequently used tunnel in Hong Kong and also, not surprisingly, had a large travel time variation.

Figure 5 illustrates this SD approximation accuracy under different k values during the peak hour (refers to the y-axis at right-hand side). The actual path travel time SD was 11.64 minutes. It can be found from Figure 5 that the path travel time SD can be underestimated by 26.5% when link travel time correlations are ignored ($k = 0$). This SD approximation accuracy can be improved by increasing the k value. For example, the approximation accuracy of path travel time SD can be significantly improved to 87.1%, when travel time correlations among adjacent links ($k = 1$) are considered. This SD approximation accuracy can be further improved to 99.1% when the k value increases to 4. Therefore, for this case study, the path travel time SD can be well approximated by using the k-limited spatial dependence assumption with a sufficiently large k value (e.g., $k = 4$).

Table 1 gives the sizes of the impact area under different k values. The impact area size is measured by the average number of links, denoted by $E(|A_{ij}^k|)$. It can be calculated as

$$E(|A_{ij}^k|) = \frac{\sum_{a_{ij} \in A} |A_{ij}^k|}{|A|} \tag{18}$$

where $|A_{ij}^k|$ is the number of links of the impact area G_{ij}^k for a primal link $a_{ij} \in G$ and $|A|$ is the number of links in the primal network G. It can be observed from Table 1 that the size of the impact area exponentially increases with k values. For instance, when k increases

Table 1. The sizes of impact area and hierarchical network under different k values.

k Value	Impact area (links)	Ground hierarchy (nodes)	Top hierarchy		
			Nodes	Links	Movements
1	**3.24**	–	1,367	**3,655**	11,849
2	13.59	5,022	3,655	11,849	34,555
3	38.16	16,871	11,849	34,555	106,942
4	***87.53***	51,426	34,555	***106,942***	315,172
5	174.63	157,001	106,942	315,172	958,541

from 1 to 4, the $|A_{ij}^k|$ value grows by 27 times, from 3.24 to 87.53. This result can be used to interpret the validation of the k-limited spatial dependence assumption. When k is large enough, the impact area can maintain a considerable number of links with correlated travel times; and thus the majority of link travel time correlations, along the path, can be captured within that impact area.

The proposed two-level hierarchical network was constructed using different k values (see Table 1). It can be seen from Table 1 that the size of constructed hierarchical network also exponentially increases with k values, similar to the size of impact area. For instance, when k increases from 1 to 4, the number of links in the top hierarchical network grows by 29 times from 3655 to 106,942. This growth rate is close to that of the impact area size (about 27 times). Therefore, the increase in the k value can improve the approximation accuracy of path travel time distributions, but at the cost of increasing the SD-RSPP problem size. It should be noted that when $k = 1$ (only correlations among adjacent links are considered), the primal network G can be used directly for solving the SD-RSPP. The travel time correlations among adjacent links can be maintained as an attribute of each movement in the primal network.

The computational performance of the proposed *SDRSP-HA** algorithm using different k values was tested on the RTIS network. The *SDRSP-HA** algorithm was coded in Visual C# programming language. The scan eligible (SE) was implemented using the F-heap data structure (Fredman and Tarjan 1987). The $h(j)$ used was the Euclidean distance function. All experiments were conducted in the computer with a four-core Intel Xeon 3.2 GHz CPU (only one core was used) and a Windows Server 2003 operation system. Table 2 reports the computational performance of the proposed *SDRSP-HA** algorithm using different k values. Three risk-taking scenarios, including risk-averse ($\alpha = 0.9$), risk-seeking ($\alpha = 0.1$), and risk-neutral ($\alpha = 0.5$), were tested for each k value. All reported results were the average of 100 computer runs, using different O–D nodes in each run. The 100 O–D nodes were randomly selected and the same O–D node set was used for every test performed on all hierarchical networks.

Table 2 gives the computational performance of the *SDRSP-HA** algorithm under different k values. As expected, the computational time of the *SDRSP-HA** algorithm exponentially increases with respect to the k values. It can be also observed from Table 2 that the *SDRSP-HA** algorithm runs much faster at the risk-neutral scenario than at the other two risk-taking scenarios. For example, when $k = 4$, the computational time required by a risk-averse scenario is about 58 times (79.085/1.365) larger than that required by the risk-neutral scenario. This is because when travelers are risk-neutral, the *SDRSP-HA** algorithm becomes, essentially, a traditional A* algorithm. In this case, the least number of nondominated paths is generated in the search process, since only a single path is kept at

Table 2. Computation performance of SDRSP-HA* algorithm under different k values.

	Risk-averse ($\alpha = 0.9$)		Risk-seeking ($\alpha = 0.1$)		Risk-neutral ($\alpha = 0.5$)	
k Value	$\tilde{t}$	$\lvert\hat{p}\rvert$	$\tilde{t}$	$\lvert\hat{p}\rvert$	$\tilde{t}$	$\lvert\hat{p}\rvert$
1	0.017	2.327	0.020	2.458	0.007	1.325
2	0.391	13.286	0.409	12.928	0.044	3.810
3	2.403	35.760	4.324	40.211	0.213	10.467
4	**79.085**	103.514	126.235	109.865	**1.365**	29.169

Notes: $\tilde{t}$: Average computational time (seconds).
$\lvert\hat{p}\rvert$: Average number of nondominated paths in the top hierarchical network (10^3).

each top hierarchical link. For the risk-averse and risk-seeking scenarios, additional computational effort is required to generate a considerable amount of M–V nondominated paths during the search process. It can also be found from Table 2 that the algorithm runs slightly faster (1.39 times faster when $k = 4$) in the risk-averse scenario than the risk-seeking scenario. This is positive information for the majority of travelers whose major concern is late arrival.

7. Conclusions

This article has presented a new efficient solution algorithm for solving the spatially dependent reliable shortest path-finding problem (SD-RSPP) in the context of RGS. In RGS applications, such reliable shortest paths can help travelers plan for their travel even in road networks with travel time uncertainties and achieve a given on-time arrival certainty (in terms of probability).

In this article, the travel time of a link is assumed to be spatially correlated only with the neighboring links within a local impact area. The travel time correlations with neighboring links are considered and represented by variance–covariance matrices. Based on this assumption, the SD-RSPP was formulated and solved as a multicriteria shortest path-finding problem. The two dominant conditions (first-order stochastic dominant and M–V dominant conditions) were established to solve the SD-RSPP. Based on the established dominant conditions, a new multicriteria A* algorithm was proposed to exactly solve the SD-RSPP in an equivalent two-level hierarchical network. The proposed hierarchical network can well represent travel time correlations among neighboring links. The optimality of the proposed solution algorithm has been proved theoretically and the complexity of the proposed algorithm has also been analyzed.

A case study using data from the RTIS in Hong Kong was carried out for demonstration in this article. Travel time data collected from the RTIS demonstrated that link travel times are, indeed, strongly correlated within local impact areas. This finding indicated that the common assumption that link travel times are independently distributed was erroneous. Numerical results also showed that the proposed limited spatial dependence assumption can well approximate path travel time standard deviation, if the size of the impact area is sufficiently large. Computational results indicated that the increase of the impact area size can significantly enhance the accuracy of reliable shortest path findings, but at the cost of additional computational time. Therefore, it is necessary to consider the trade-off between the accuracy and computational performance of the solution algorithm in development of the RGS applications.

Due to space limitations, the computational experiments presented in this article are by no means comprehensive. In this article, the size of a two-level hierarchical network can become huge if a large impact area is adopted for large-scale road networks. In this case, the computational time of the proposed solution algorithm may not be fast enough for real-time RGS operations. Many efficient solution algorithms (Jing *et al.* 1998) based on multi-hierarchical network approach has been proposed in the literature to solve the traditional shortest path problems with huge network size. How to integrate this multi-hierarchical shortest path approach into the proposed solution algorithm is a significant extension of the study presented in this article. It is well known that the shortest path problem is an essential sub-problem of the traffic assignment problems used to model travel choice behavior of large numbers of travelers in congested road networks. The application of the proposed algorithm in traffic assignment problems (Lam *et al.* 2008, Chen *et al.* 2011) is another possible future research direction.

Acknowledgments

The authors are thankful to the anonymous referees for their comments and suggestions that improved this article. The work described in this article was jointly supported by a research grant from the Research Grant Council of the Hong Kong Special Administration Region to the Hong Kong Polytechnic University (Project No. PolyU 5195/07E), an internal research grant J-BB7Q from the Research Committee of the Hong Kong Polytechnic University and two grants from the National Science Foundation of China (Grants #40830530 and #60872132).

References

Carraway, R.L., Morin, T.L., and Moskowitz, H., 1990. Generalized dynamic programming for multicriteria optimization. *European Journal of Operational Research*, 44, 95–104.

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

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

Chen, A. and Ji, Z.W., 2005. Path finding under uncertainty. *Journal of Advanced Transportation*, 39, 19–37.

Chen, B.Y., *et al*., 2011. An efficient solution algorithm for solving multi-class reliability-based traffic assignment problem. *Mathematical and Computer Modelling*, 54, 1428–1439.

El Esawey, M. and Sayed, T., 2011. Travel time estimation in urban networks using limited probes data. *Canadian Journal of Civil Engineering*, 38, 305–318.

Frank, H., 1969. Shortest paths in probabilistic graphs. *Operations Research*, 17, 583–599.

Fredman, M.L. and Tarjan, R.E., 1987. Fibonacci heaps and their uses in improved network optimization algorithms. *Journal of the Association of Computing Machinery*, 34, 596–615.

Gajewski, B.J. and Rilett, L.R., 2003. Estimating link travel time correlation: an application of Bayesian smoothing splines. *Journal of Transportation and Statistics*, 7, 53–70.

Huang, B., Wu, Q., and Zhan, F.B., 2007. A shortest path algorithm with novel heuristics for dynamic transportation networks. *International Journal of Geographical Information Science*, 21, 625–644.

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

Jing, N., Huang, Y.W., and Rundensteiner, E.A., 1998. Hierarchical encoded path views for path query processing: an optimal model and its performance evaluation. *IEEE Transactions on Knowledge and Data Engineering*, 10, 409–432.

Lam, T.C. and Small, K.A., 2001. The value of time and reliability: measurement from a value pricing experiment. *Transportation Research Part E-Logistics and Transportation Review*, 37, 231–251.
Lam, W.H.K., Chan, K.S., and Shi, J.W.Z., 2002. A traffic flow simulator for short-term travel time forecasting. *Journal of Advanced Transportation*, 36, 265–291.
Lam, W.H.K., Shao, H., and Sumalee, A., 2008. Modeling impacts of adverse weather conditions on a road network with uncertainties in demand and supply. *Transportation Research Part B-Methodological*, 42, 890–910.
Li, Z.C., Huang, H.J., and Lam, W.H.K., in press. Modelling heterogeneous drivers' response to route guidance and parking information system in stochastic and time-dependent networks. *Transportmetrica*.
Lim, S., 2008. *Traffic prediction and navigation using historical and current information*. Thesis (Master). Massachusetts Institute of Technology.
Nie, Y. and Wu, X., 2009a. Shortest path problem considering on-time arrival probability. *Transportation Research Part B-Methodological*, 43, 597–613.
Nie, Y. and Wu, X., 2009b. Reliable a priori shortest path problem with limited spatial and temporal dependencies. *In*: W.H.K. Lam, S.C. Wong, and H.K. Lo, eds. *Transportation and traffic theory 2009: golden jubilee*. New York: Springer, 169–195.
Nikolova, E., 2009. *Strategic algorithms*. Thesis (PhD). Massachusetts Institute of Technology.
Rakha, H., *et al*., 2006. Estimating path travel-time reliability. *In*: *Proceedings of 2006 IEEE intelligent transportation systems conference,* 17–20 September 2006, Toronto, Canada. New Jersey: IEEE, 236–241.
Shao, H., Lam, W.H.K., and Chan, K.S., 2004. The problem of searching the reliable path for transportation networks with uncertainty. *In*: *Proceeding of 9th conference of the Hong Kong society for transportation studies*, 11–14 December 2004, Hong Kong, China. Hong Kong: Hong Kong University Science & Technology, 226–234.
Tam, M.L. and Lam, W.H.K., 2008. Using automatic vehicle identification data for travel time estimation in Hong Kong. *Transportmetrica*, 4, 179–194.
Tam, M.L., Lam, W.H.K., and Lo, H.P., 2008. Modeling air passenger travel behavior on airport ground access mode choices. *Transportmetrica*, 4, 135–153.
Tobler, W.R., 1970. A computer movie simulating urban growth in the Detroit Region. *Economic Geography*, 46, 234–240.
Zeng, W. and Church, R.L., 2009. Finding shortest paths on real road networks: the case for A*. *International Journal of Geographical Information Science*, 23, 531–543.

Appendix

Let $p^{ij,k} = \{a^1, \ldots, a^k\}$ be a path from node i to node j consisting of k consecutive links, and $p_u^{rj,k+\lambda} = p_u^{ri,\lambda} \oplus p^{ij,k}$ be a path from origin r to node j going through the sub-path $p^{ij,k}$. Let $f_{p_u^{rj,k}}(\cdot)$, $\Phi_{p_u^{rj,k}}(\cdot)$, and $\Phi^{-1}_{p_u^{rj,k}}(\cdot)$ be the PDF, the cumulative probability function (CDF), and the inverse of CDF of travel time $T_u^{rj,k}$, respectively. The FSD condition and M–V dominant condition are proved as follows.

Proposition A1 (FSD condition): Given two paths $p_u^{rj,\lambda+k} \neq p_v^{rj,\eta+k} \in P^{rj}$, $p_u^{rj,\lambda+k} \succ p_v^{rj,\eta+k}$ if $p_u^{rj,k+\lambda}$ and $p_u^{rj,k+\eta}$ satisfy $\Phi^{-1}_{p_u^{rj}}(y) < \Phi^{-1}_{p_v^{rj}}(y)$, $\forall y \in (0,1)$.

Proof Given a path $\forall p^{j\ell} \in P^{j\ell}, \forall \ell \in N$ and an on-time arrival probability $\forall y \in (0,1)$, $\Phi^{-1}_{p_u^{rj} \oplus p^{j\ell}}(y) = t_u^{rj} + t^{jl} + Z_y\sqrt{\left(\sigma_u^{rj}\right)^2 + 2\mathrm{cov}\left(p^{ij,k}, p^{jl}\right) + \left(\sigma^{jl}\right)^2}$ holds under the k-limited spatial dependence assumption. When $(\tilde{\sigma}^{j\ell})^2 = 2\mathrm{cov}(p^{ij,k}, p^{j\ell}) + (\sigma^{j\ell})^2$ is used, this limited spatial dependence formula is equivalent to the following spatial independence formula $\Phi^{-1}_{p_u^{rj} \oplus \tilde{p}^{j\ell}}(y) = t_u^{rj} + t^{j\ell} + Z_y\sqrt{\left(\sigma_u^{rj}\right)^2 + \left(\tilde{\sigma}^{j\ell}\right)^2}$. As the same $(\tilde{\sigma}^{j\ell})^2 = 2\mathrm{cov}(p^{ij,k}, p^{j\ell}) + (\sigma^{j\ell})^2$ can be adopted for $\Phi_{p_v^{rj} \oplus p^{j\ell}}(y)$, $p_u^{rj,\lambda+k} > p_v^{rj,\eta+k}$ is equivalent to prove $\Phi^{-1}_{p_u^{rj} \oplus \tilde{p}^{j\ell}}(y) <$

$\Phi_{p_v^{rj} \oplus \tilde{p}^{j\ell}}(y)$ using the spatial independence formula as below. Let $f_{\tilde{p}^{j\ell}}$ be the probability density function (PDF) of modified

$$
\begin{aligned}
&\Phi^{-1}_{p_u^{rj}}(y) < \Phi^{-1}_{p_v^{rj}}(y), \forall y \in (0,1)\\
&\quad \Rightarrow \Phi_{p_u^{rj}}(b) > \Phi_{p_v^{rj}}(b), \forall b \in R^+\\
&\quad \Rightarrow \int_0^b (\Phi_{p_u^{rj}}(b-t) - \Phi_{p_v^{rj}}(b-t)) f_{\tilde{p}^{j\ell}} dt > 0, \forall b \in R^+\\
&\quad \Rightarrow \int_0^b \Phi_{p_u^{rj}}(b-t) f_{\tilde{p}^{j\ell}} dt - \int_0^b \Phi_{p_v^{rj}}(b-t) f_{\tilde{p}^{j\ell}} dt > 0, \forall b \in R^+\\
&\quad \Rightarrow \Phi_{p_u^{rj} \oplus \tilde{p}^{j\ell}}(b) - \Phi_{p_v^{rj} \oplus \tilde{p}^{j\ell}}(b) > 0, \forall b \in R^+ \text{(according to the convolution integral)}\\
&\quad \Rightarrow \Phi^{-1}_{p_u^{rj} \oplus \tilde{p}^{j\ell}}(y) < \Phi_{p_v^{rj} \oplus \tilde{p}^{j\ell}}(y), \forall y \in (0,1)
\end{aligned}
$$

Proposition A2 (M–V dominant condition) Given an on-time arrival probability α and two paths $p_u^{rj,\lambda+k} \neq p_v^{rj,\eta+k} \in P^{rj}$, $p_u^{rj,\lambda+k} > p_v^{rj,\eta+k}$ if $p_u^{rj,k+\lambda}$ and $p_u^{rj,k+\eta}$ satisfy either

(1) $t_u^{rj} \leq t_v^{rj}$ and $Z_\alpha \sigma_u^{rj} < Z_\alpha \sigma_v^{rj}$ or
(2) $t_u^{rj} < t_v^{rj}$ and $Z_\alpha \sigma_u^{rj} \leq Z_\alpha \sigma_v^{rj}$

Proof (1) When $\alpha \geq 0.5$ and $Z_\alpha \sigma_u^{rj} \leq Z_\alpha \sigma_v^{rj}$, we have $Z_\alpha \geq 0$, $\sigma_u^{rj} \leq \sigma_v^{rj}$ and $\sqrt{(\sigma_u^{rj})^2 + 2\mathrm{cov}(p^{ij,k}, p^{j\ell}) + (\sigma^{j\ell})^2} \leq \sqrt{(\sigma_v^{rj})^2 + 2\mathrm{cov}(p^{ij,k}, p^{j\ell}) + (\sigma^{j\ell})^2}$. When $\alpha < 0.5$ and $Z_\alpha \sigma_u^{rj} \leq Z_\alpha \sigma_v^{rj}$, we have $Z_\alpha < 0$, $\sigma_u^{rj} \geq \sigma_v^{rj}$ and $\sqrt{(\sigma_u^{rj})^2 + 2\mathrm{cov}(p^{ij,k}, p^{j\ell}) + (\sigma^{j\ell})^2} \geq \sqrt{(\sigma_v^{rj})^2 + 2\mathrm{cov}(p^{ij,k}, p^{j\ell}) + (\sigma^{j\ell})^2}$. Therefore, when $Z_\alpha \sigma_u^{rj} \leq Z_\alpha \sigma_v^{rj}$, for any path $p^{j\ell} \in P^{j\ell}$ we have $Z_\alpha \sqrt{(\sigma_u^{rj})^2 + 2\mathrm{cov}(p^{ij,k}, p^{j\ell}) + (\sigma^{j\ell})^2} \leq Z_\alpha \sqrt{(\sigma_v^{rj})^2 + 2\mathrm{cov}(p^{ij,k}, p^{j\ell}) + (\sigma^{j\ell})^2}$. Thus, if $t_u^{rj} < t_v^{rj}$ holds, $\Phi^{-1}_{p_u^{rj} \oplus p^{j\ell}}(\alpha) < \Phi^{-1}_{p_v^{rj} \oplus p^{j\ell}}(\alpha)$ is satisfied for $\forall p^{j\ell} \in P^{j\ell}, \forall \ell \in N$ (2) The proof is similar to (1). $\square$.

Let $T^{\hat{p}}_{rs,u}$ be the travel time distribution of path $\hat{p}_u^{rs,\lambda-k+1} \in H^t$; and $t^{\hat{p}}_{rs,u}$ and $(\sigma^{\hat{p}}_{rs,u})^2$ be the mean and variance of $T^{\hat{p}}_{rs,u}$, respectively. Proposition 4.1 can be proved as below.

Proposition A3 Given a top hierarchical path $\forall \hat{p}_u^{rs,\lambda} \in H^t$, its travel time distribution is equivalent to that of the corresponding primal path $p_u^{rs,\lambda+k-1} \in G$.

Proof Since $t^{\hat{p}}_{rs,u} = t_u^{ij,k-1} + \sum_{m=1}^{\lambda} t^{\hat{a},m}_{\ell q,u} = \sum_{m=1}^{k-1} t^m + \sum_{m=1}^{\lambda} t^{m+k-1} = \sum_{m=1}^{\lambda+k-1} t^m = t_u^{rs,\lambda+k-1}$ holds, $\hat{p}_u^{rs,\lambda}$ and $p_u^{rs,\lambda+k-1}$ have a same mean travel time. It can also be proved that these two paths also have a same travel time variance as follows.

$$
\begin{aligned}
\left(\sigma^{\hat{p}}_{rs,u}\right)^2 &= \left(\sigma_u^{ri,k-1}\right)^2 + \sum_{m=1}^{\lambda} \left(\sigma^{\hat{a},m}_{\ell q,u}\right)^2 + \sum_{m=2}^{\lambda} 2\mathrm{cov}\left(\hat{a}^{m-1}_{jw,u}, \hat{a}^m_{\ell q,u}\right)\\
&= \left(\sigma_u^{ri,k-1}\right)^2 + \left(\sigma^k\right)^2 + \sum_{n=1}^{k-1} 2\mathrm{cov}\left(a^n, a^k\right) + \sum_{m=2}^{\lambda} \left(\sigma^{\hat{a},m}_{\ell q,u}\right)^2 + \sum_{m=2}^{\lambda} 2\mathrm{cov}\left(\hat{a}^{m-1}_{jw,u}, \hat{a}^m_{\ell q,u}\right)\\
&= \left(\sigma_u^{rj,k}\right)^2 + \sum_{m=2}^{\lambda} \left(\sigma^{\hat{a},m}_{\ell q,u}\right)^2 + \sum_{m=2}^{\lambda} 2\mathrm{cov}\left(\hat{a}^{m-1}_{jw,u}, \hat{a}^m_{\ell q,u}\right)
\end{aligned}
$$

$$
\begin{aligned}
&= \left(\sigma_u^{rj,k}\right)^2 + \left(\sigma^{k+1}\right)^2 + \sum_{n=1}^{k-1} 2\mathrm{cov}\left(a^{n+1}, a^{k+1}\right) + 2\mathrm{cov}\left(a^1, a^{k+1}\right) \\
&\quad + \sum_{m=3}^{\lambda} \left(\sigma_{\ell q,u}^{\hat{a},m}\right)^2 + \sum_{m=3}^{\lambda} 2\mathrm{cov}\left(\hat{a}_{jw,u}^{m-1}, \hat{a}_{\ell q,u}^{m}\right) \\
&= \left(\sigma_u^{rw,k+1}\right)^2 + \sum_{m=3}^{\lambda} \left(\sigma_{\ell q,u}^{\hat{a},m}\right)^2 + \sum_{m=3}^{\lambda} 2\mathrm{cov}\left(\hat{a}_{jw,u}^{m-1}, \hat{a}_{\ell q,u}^{m}\right) \\
&= \ldots \\
&= \left(\sigma_u^{r\ell,\lambda+k-2}\right)^2 + \sum_{m=\lambda}^{\lambda} \left(\sigma_{\ell q,u}^{\hat{a},m}\right)^2 + \sum_{m=\lambda}^{\lambda} 2\mathrm{cov}\left(\hat{a}_{jw,u}^{m-1}, \hat{a}_{\ell q,u}^{m}\right) \\
&= \left(\sigma_u^{r\ell,\lambda+k-2}\right)^2 + \left(\sigma^{\lambda+k-1}\right)^2 + \sum_{n=1}^{k-1} 2\mathrm{cov}\left(a^{\lambda+n-1}, a^{\lambda+k-1}\right) + 2\mathrm{cov}\left(a^{\lambda-1}, a^{\lambda+k-1}\right) \\
&= (\sigma_u^{rs,\lambda+k-1})^2 \;\square
\end{aligned}
$$