Shortest Paths in Stochastic Time-dependent Networks

with Link Travel Time Correlation

Wei Dong*
Faculty of Information & Communication Technologies
Swinburne University of Technology, VIC 3122, Australia
Email: wdong@swin.edu.au
Fax: +61 3 9819 0823

Hai L. Vu
Centre for Advanced Internet Architectures, Swinburne University of Technology
Swinburne Intelligent Transport Systems Lab
Email: hvu@swin.edu.au

Yoni Nazarathy
School of Mathematics and Physics
University of Queensland, Australia
Email: y.nazarathy@uq.edu.au

Bao Quoc Vo
Faculty of Information & Communication Technologies
Swinburne University of Technology, Australia
Email: bvo@swin.edu.au

Minyi Li
Faculty of Information & Communication Technologies
Swinburne University of Technology, Australia
Email: myli@swin.edu.au

Serge Paul Hoogendoorn
Faculty of Civil Engineering and Geosciences
Delft University of Technology, Netherlands
Email: s.p.hoogendoorn@tudelft.nl
*Corresponding Author

Word count:

Text: 6198

Figures & Tables: 5×250=1250

Total: 7448

Submission date: August 1, 2012
Abstract

This paper develops a simple robust framework for the problem of finding the least expected travel time route from any node to any given destination in a stochastic and time-dependent network. We consider both spatial and temporal link travel time correlations in the proposed solution based on a dynamic programming approach. In particular, the spatial correlation is represented by a Markovian property of the link states where each link is assumed to experience congested or uncongested conditions. The temporal correlation is manifested through the time-dependent expected link travel time given the condition of the link traversed. The framework enables a route guidance system where at any decision node within a network, one can make a decision based on current traffic information about which node to take next to achieve the shortest expected travel time to the destination. Numerical examples are presented to illustrate the computational steps involved in the framework of making route choice decisions and to demonstrate the effectiveness of the proposed solution.
1. INTRODUCTION

It is envisaged that soon many drivers will be equipped with real-time in-vehicle navigation systems. Such navigation aids are designed to support drivers in finding the fastest path to destination and provide alternative route choices based on current traffic congestion information. In this context, finding a shortest path given the real-time information in a time-dependent stochastic network is a problem of both theoretical and practical nature. It is because in such a network the link cost (e.g. travel time) is continuously changing and influenced by many factors such as bad weather, traffic congestion, construction work, and day-to-day fluctuations on traffic demands, etc. Moreover, travel times on different links are strongly correlated (i.e. spatial correlation) as a road incident in one link can cause congestion on other links in the surrounding locations. It is also well known that traffic demands changed during the day (e.g. peak versus off-peak period) where travel time on the same link can be significantly different (i.e. temporal correlation). We develop in this paper a simple yet accurate framework for the problem of finding the least expected travel time route using real-time information from any node to any given destination in a stochastic and time-dependent network.

Over the past decade, the shortest path problem in transportation networks has been investigated extensively in the literature with many studies focusing on the least expected travel time under various assumptions. In the literature, there are four common models about link travel: constant link travel time, time-dependent link travel time, stochastic link travel time and stochastic time-dependent link travel time. The most basic models assume the link travel time is either constant deterministic as in Dreyfus (1), Hart et al. (2) or time-dependent deterministic as in Chabini (3).

In reality, link travel times are highly stochastic and display time-dependent characteristics. To deal with uncertainty and time-varying aspect of travel time, a number of approaches have been proposed. For instance, the shortest path issue in a network where link travel times are stochastic and time-dependent is investigated in Hall (4). In particular, the optimal route selection from any node to the destination depends on the arrival time at that node, which is not a fixed route but adaptively changes based on the least expected travel time to the destination. Later, Daniele (5) devises an algorithm for the same shortest path problem with link travel times being both random and time-dependent. Also extending from Hall's work (4), an adaptive routing algorithm for in-vehicle navigation system is studied in Fu (6) assuming better real-time information is available about the characteristic of the actual link travel times. The a-priori shortest path problem in a stochastic and time-dependent network is proposed in Nie and Wu (7). Here the full probability density function of the link travel times is assumed to be known and the calculation can be extended for the time-dependent case with increase complexity. Gao et al. (8, 9) propose adaptive route choice models in a stochastic time-dependent network with and without perfect online information about the state of the transportation network. They assume full knowledge of the joint distribution of travel times on every link of the network, and develop a routing policy based on decision rules specifying the next node based on realized travel times at the time of making the decision. Despite very good results that can be achieved using this framework, one of the main drawbacks of this approach is that it assumes knowledge of the full joint distribution of network (i.e. consisting of many links) travel times. To address this problem, we will in this paper develop an alternative approach that requires a much smaller set of parameters without significantly compromising on the accuracy of the results. Note that variants of this framework such as Nie and Wu (10), Dong et al. (11) have been developed.
using the conditional probability between adjacent links that are still being devised from the full joint probability distributions of network travel time.

Furthermore, travel times on different links are generally correlated. It is easy to observe that in a typical traffic network, congestion in a link generally influences the condition of adjacent links in terms of travel time. Evidences of this link travel times correlation have been recorded in the literature. For example, it is shown in Gajewski and Rilett (12) that the link covariance is non-zero and the distribution of the correlation coefficient can be used as a performance metric of the network. Similarly, various network dependencies are investigated in Bernard et al. (13) by calculating the correlations for travel speeds measured on paths with given numbers of intersections and distances. Their results show that the correlation is high for adjacent links, and this correlation decreases as the distance of the links increases. These motivate us to also consider the correlation between links when analysing shortest path problems with real-time information in this paper.

In most traditional approaches for finding shortest path, link costs are assumed to be independent (e.g., see Hart (2), Dijkstra (14)). These approaches are applicable in some situations, but as long as the link costs are inter-dependent, they may produce undesirable results. There has been only limited work in the literature that addresses this spatial dependence (i.e. correlation) between travel time realizations on consecutive links that make up a route. In particular, Nie and Wu (10) extend their previous work in Nie and Wu (7) to investigate the reliable shortest path problem where the known probability density function of the link travel times is now conditional on the state of the previous traversed link. In Ji et al. (15) a simulation-based method is proposed where the spatial correlations are formulated as variance-covariance matrices. Neither approach is scalable for a network with many links and suitable for on-line calculation due to their complexity. Waller and Ziliaskopoulos (16) investigate the shortest path problem in a time-invariant stochastic network with limited link correlation. Fan et al. (17) propose to use a simple one-step conditional probability to represent the correlation between the states of links and shows that the spatial dependence can be effectively treated in finding the shortest path. In this paper, we will similarly assume the spatial correlations follow a Markovian property where the probability of a link that experiences congested or uncongested conditions (or states) only depends on the condition of its immediate preceded link. However, we will further incorporate the link travel time dependencies (i.e. temporal correlation) described earlier into our framework while preserving the simplicity of the Markovian structure.

The main contributions of this paper are summarized as follows:

- We develop a general framework for shortest-time route guidance with real-time information, which incorporates both types of widely known link travel time dependencies, namely \textit{temporal dependence} and \textit{spatial dependence}. By temporal dependence, we refer to the time varying aspect of link travel time, and by spatial dependence, we refer to the correlation between travel time realizations on consecutive links that make up a route.
- We devise an algorithm based on the basic principle of dynamic programming to solve the shortest path problem in stochastic time-dependent networks with link travel time correlation.
- The proposed framework requires only just a few parameters, such as conditional average link travel time (or average flow speed) values in a given link state and time interval of the day, and the conditional transition probabilities between states (congested vs. uncongested) of the two consecutive links in the network. We show via
examples that the framework can achieve similar accuracy with a much smaller set of parameters compared to the case when the full joint distribution of network travel times is required.

- The temporal correlation is treated in this framework without significant increases in computational time compared to networks without temporal correlation. We show via examples that the proposed framework achieves the same results as a previous approach when tackling the network spatial dependencies and it improves the route choice decisions whenever there is a temporal correlation.

This paper is organized as follows. In the next section, we describe the proposed least expected travel time (LET) route selection process, which includes a formulation of the problem and the solution for shortest path selection with link travel time correlation calculation. In Section 3, an illustrative example is presented to describe the calculation process in details and to allow us to compare our approach with methods for LET route selection proposed by others. In Section 4, the performance of the proposed model is evaluated through numerical experiments on a larger network. Section 5 summarizes the paper with several concluding remarks and points to directions for future work.

2. OUR PROPOSED FRAMEWORK

In this section, we first formulate the problem and then present an on-line approach for LET route selection based on the time of arrival at the decision node and the state (or congestion level) of the incoming link.

2.1. Problem formulation

Let us consider a time-dependent network $Z = (A, N, \Gamma)$ where $A$ is a set of links, $N$ is a set of nodes and $\Gamma$ is a set of time intervals spanning the considered routing period. Let $[\tau_h, \tau_{h+1})$, $h = 0, 1, ..., |\Gamma| - 1$ be the time intervals in $\Gamma$ separated by the time points $\tau_0 < \tau_1 < ... < \tau_{|\Gamma| - 1}$ where $|\Gamma|$ is the size of the set $\Gamma$. Given a destination $D \in N$ in the above network, our aim is to find the optimal route from any node in $N$ (referred to as a decision node) to the destination $D$ with the least expected travel time. The decision of route choice is made at the time the traveler arrives at the decision node.

In this framework the spatial correlation between link travel times is taken into account via a Markovian assumption on the link states where each link is assumed to experience either congested or uncongested conditions. Although the framework can be easily extended to more than two possible levels of link congestion as briefly shown at the end of Section 2.2, we will only consider two states (congested vs. uncongested) in the numerical examples of this paper for clarity. Similar to the approach in Fan et al. (17), we also define the conditional transition probabilities between states of the two consecutive links but for a particular time interval (or time zone). The latter incorporates the temporal correlation into the proposed framework. Specifically, at any given time interval $[\tau_h, \tau_{h+1})$, for any consecutive link pair $(k, i)$ and $(i, j)$ $\in A$, $\forall k, i, j \in N$, let $\alpha_{kij}$ be the conditional probability that if link $(k, i)$ is uncongested then link $(i, j)$ is also uncongested; and, let $\lambda_{kij}$ be the conditional probability that if link $(k, i)$ is congested then link $(i, j)$ is uncongested.

The values of $\alpha_{kij}$ and $\lambda_{kij}$ can be determined based on the estimation of the link states (or a single commodity macroscopic flow) using standard techniques such as state-space cell transmission model proposed in Tampere and Immers (18). Alternatively, these values can also be calculated based on the conditional probability density function of the link travel
times given the state (congested/uncongested) of the incoming link as done in Fan et al. (17). The latter approach requires the knowledge of the link travel times distribution and some predefined travel time thresholds where the link is considered uncongested if the time required to traverse that link is less than the threshold, and considered congested otherwise. To this end, the link travel times and its distribution can be estimated directly from floating vehicles in terms of aggregation of individual measurements in a traffic network over a certain period as in Kaparias (19), or indirectly using the standard techniques such as those listed in Liu et al. (20). Note that due to the time-varying nature of the network, \( \alpha_{kij} \) and \( \lambda_{kij} \) values are time-dependent and can be different for different time intervals. For simplicity, however, in this framework we consider them to be time-invariant that will be applied across the different time zones. The rationale behind this simplification is that the threshold is a perceived value and can be different for the same link in different time intervals resulting in similar \( \alpha_{kij} \) and \( \lambda_{kij} \) values across many time zones. It is because the travel time threshold, at which the link is considered to be congested, is likely to be higher in a peak period compare to that of the off-peak period in the same day. Nevertheless, the framework can be extended to include time-dependent \( \alpha_{kij} \) and \( \lambda_{kij} \) parameters with increasing complexity.

2.2. Route choice heuristic

Our proposed approach is motivated by Bellman's principle of optimality introduced in Bellman (21). Formally, assume that the traveler arrives at node \( i \) on an incoming link \((k, i)\) at time \( t \in [\tau_h, \tau_{h+1}) \) for some \( h \in \{0,1,2,\ldots,|\Gamma| - 1\} \). Knowing the condition of link \((k, i)\) and the start time at node \( i \), the traveler wishes to arrive at the destination node \( D \) as early as possible. In such a situation, there might be multiple choices for the next node to take from the current node \( i \), which may result in different route selections (and thus different length of travel time). Moreover, once the traveller chooses a next node (e.g., \( j \)) to visit, they will further need to make a decision regarding the optimal option of the next node to visit from node \( j \). Overall, the objective here is to choose a path to travel such that the expected travel time to the destination is minimized.

Upon arrival to node \( i \) at time \( t \in [\tau_h, \tau_{h+1}) \), given the fixed values of \( \alpha_{kij} \) and \( \lambda_{kij} \), the LET route selection problem can be formulated as solving the following recursion for \( u_{ij}(\cdot) \) and \( c_{ij}(\cdot) \) below.

\[
u_{ki}(t) = \min_{j \neq i} \left\{ \alpha_{kij} \left( U_{ij}(t) + u_{ij} \left( t + U_{ij}(t) \right) \right) + \left( 1 - \alpha_{kij} \right) \left( C_{ij}(t) + c_{ij} \left( t + C_{ij}(t) \right) \right) \right\} \quad (1)
\]

\[
c_{ki}(t) = \min_{j \neq i} \left\{ \lambda_{kij} \left( U_{ij}(t) + u_{ij} \left( t + U_{ij}(t) \right) \right) + \left( 1 - \lambda_{kij} \right) \left( C_{ij}(t) + c_{ij} \left( t + C_{ij}(t) \right) \right) \right\} \quad (2)
\]

\[
u_{kD}(t) = 0, c_{kD}(t) = 0 \quad (3)
\]

where:

\( u_{ki}(t) \) - an estimate of the least expected travel time between node \( i \) and the destination node \( D \) at time \( t \) if the incoming link \((k, i)\) is uncongested;

\( c_{ki}(t) \) - an estimate of the least expected travel time between node \( i \) and the destination node \( D \) at time \( t \) if the incoming link \((k, i)\) is congested;

\( U_{ij}(t) \) - the expected travel time between node \( i \) and node \( j \) at time \( t \in [\tau_h, \tau_{h+1}) \) under uncongested conditions;
the expected travel time between node $i$ and node $j$ at time $t \in [\tau_h, \tau_{h+1})$ under congested conditions.

Furthermore, the $U_{ij}(t)$ and $C_{ij}(t)$ values in the above equations are time dependent and are calculated based on the flow speed model proposed in (19) as follows.

$$U_{ij}(t) = \frac{\overline{l}_{ij}}{v_{h(i,j)}}$$

if $\frac{l_{ij}}{v_{h(i,j)}} < \tau_{h+1} - t$ \hspace{1cm} (4)

else

$$U_{ij}(t) = \tau_{h+1} + \frac{l_{ij} - l_{ij}^0}{v_{h+1(i,j)}} - t,$$  \hspace{1cm} if $\frac{l_{ij} - l_{ij}^0}{v_{h+1(i,j)}} < \tau_{h+2} - \tau_{h+1}$, $l_{ij}^0 = v_{h(i,j)}(\tau_{h+1} - t)$ \hspace{1cm} (5)

else

$$U_{ij}(t) = \tau_{h+2} + \frac{l_{ij} - l_{ij}^3}{v_{h+2(i,j)}} - t,$$  \hspace{1cm} if $\frac{l_{ij} - l_{ij}^3}{v_{h+2(i,j)}} < \tau_{h+3} - \tau_{h+2}$,

$$l_{ij}^3 = l_{ij}^0 + v_{h+1(i,j)}(\tau_{h+2} - \tau_{h+1})$$ \hspace{1cm} (6)

... 

$$U_{ij}(t) = \tau_{|\Gamma|-1} + \frac{l_{ij} - l_{ij}^{|\Gamma|-2}}{v_{|\Gamma|-1(i,j)}} - t,$$  \hspace{1cm} if $\frac{l_{ij} - l_{ij}^{|\Gamma|-2}}{v_{|\Gamma|-1(i,j)}} < \tau_{|\Gamma|} - \tau_{|\Gamma|-1}$,

$$l_{ij}^{|\Gamma|-2} = l_{ij}^{|\Gamma|-3} + v_{|\Gamma|-2(i,j)}(\tau_{|\Gamma|-1} - \tau_{|\Gamma|-2})$$ \hspace{1cm} (7)

where $v_{h(i,j)}$ is the average flow-speed on link $(i,j)$ of length $l_{ij}$ in the time interval $[\tau_h, \tau_{h+1})$ conditioned on that it is uncongested. Similarly, $C_{ij}(t)$ can be obtained based on the corresponding $v_{h(i,j)}$ average flow-speed on link $(i,j)$ in the time interval $[\tau_h, \tau_{h+1})$ conditioned on that it is congested. Note that the $v_{h(i,j)}^u$ and $v_{h(i,j)}^c$ values can be deducted from the estimation of the link travel times distribution described earlier in Sec 2.1 for a given congestion level of the link.

More generally, for more than two possible link states (i.e. $M > 2$ possible congestion levels on the link), the proposed framework can be rewritten as follows.

$$u_{k,i}^s(t) = \min_{j \neq i} \left\{ \sum_{r=1}^{M} p_{k,i}^{sr} \left( U_{ij}^r(t) + u_{ij}^r \left( t + U_{ij}^r(t) \right) \right) \right\}, s = 1,2, ..., M$$ \hspace{1cm} (8)

$$u_{k,D}^s(t) = 0$$ \hspace{1cm} (9)

where:

- $u_{k,i}^s(t)$ - an estimate of the least expected travel time between node $i$ and the destination node $D$ at time $t$ if the incoming link $(k,i)$ is in state $s$;

- $U_{ij}^r(t)$ - the expected travel time between node $i$ and node $j$ at time $t \in [\tau_h, \tau_{h+1})$ under link state $r$, $U_{ij}^r(t) = \tau_{h+y} + \frac{l_{ij} - l_{ij}^{y-1}}{v_{h+y(i,j)}} - t,$  \hspace{1cm} if $\frac{l_{ij} - l_{ij}^{y-1}}{v_{h+y(i,j)}} < \tau_{h+y+1} - \tau_{h+y}$,
where $\gamma \geq 1$ is the number of time zones a vehicle crosses while traveling link $(i,j)$ starting at time $t \in [\tau_h, \tau_{h+1})$ and denote $l_{ij}^{\gamma-1} = 0$;

- $p_{kij}^{ST}$ - the probability that link $(i,j)$ is in state $r$ if the incoming link $(k,i)$ is in state $s$,

$$p_{kij}^{ST}(t) = \int_{t_{h(i,j)r-1}}^{t_{h(i,j)r}} P_{ij}(t,\xi)d\xi \approx p_{kij}^{ST}, \Sigma_{r=1}^{M} p_{kij}^{ST} = 1;$$

where:

$t_{h(i,j)r}, r = 1,2, ..., M$ is the travel time threshold for link $(i,j)$ to be in state $r$ and let $t_{h(i,j)0} = 0$;

$P_{ij}(t,\xi)d\xi$ is the probability traveling from node $i$ to $j$ requires time between $\xi$ and $\xi + d\xi$, and $t \in [\tau_h, \tau_{h+1})$ given that the incoming link $(k,i)$ is in state $s$.

The probability $p_{kij}^{ST}$ is approximated as time independent value across different time intervals given the threshold $t_{h(i,j)r}$. Note that in this paper the time-dependency (or temporal correlation) is mainly taken into account by the former ($U_{ij}^{T}(t)$), while the latter simplification (time-invariant $p_{kij}^{ST}$) has only secondary order impact.

3. AN ILLUSTRATIVE EXAMPLE

In this section, we demonstrate the execution of our proposed approach through a simple illustrative example. We would also compare the result of our approach with other two existing methods: i) an optimal routing policy framework introduced by Gao and Chabini (8); and ii) a dynamic programming approach proposed by Fan et al. (17). In order to compare the performance of different methods, we assume that a set of support points (i.e., the joint discrete probability distribution (or joint probability mass function, pmf) of all link travel times in the network) are available and we pick up one of the support points as the actual travel time realization in this example. Nonetheless, we emphasize here that our proposed method do not rely on support points, i.e., all the parameters required can be calculated as described in Sec 2.1 based on a set of historical data.

3. 1. The simple network

We consider a small traffic network depicted in Figure 1. There are nine links ($|A| = 9$) and seven nodes ($|\mathcal{N}| = 7$). Assume that node $i$ is the starting point and $D$ the destination point. There are two possible links from the beginning, link $(i,1)$ and link $(i,2)$, respectively. Moreover, there are four possible routes to travel to the destination.

For the sake of simplicity, assume there are only three possible support points $W = \{w_1, w_2, ..., w_L\}$ (see Table 1), where $w_i, i \in 1,2, ..., L$ is an $a \times |\Gamma|$ vector with probability $p_i$ representing one support point, and $L$ is the total number of support points. Since the set $W$ covers all the possible realizations of link travel time in the network, and each happens with probability $p_i$, so we have $\Sigma_{i=1}^{L} p_i = 1$. Here, we assign the probability to the support point $w_1, w_2$ and $w_3$ as 0.5, 0.3, 0.2, respectively. There are two time intervals 0 and 1 ($|\Gamma| = 2$) and each time interval represents three minutes length ($\tau_1 - \tau_0 = \tau_2 - \tau_1 = 3$). Travel time data beyond time 1 are assumed identical to those at time 1 for all the support points. The set of support point data is shown in Table 1, where each column vector $w_i, i \in L = \{1,2,3\}$ represents a support point. The common mean travel time for time interval 0 is 5 and for time
interval 1 is 15. The common link travel time standard deviation and the common correlation coefficient of link travel time is 1 and 0.6, respectively. In this example, we assume support point \( w_2 \) is the travel time realization (i.e., the actual travel time experienced by the driver).

### 3.2 Parameters calculation

Based on the set of support points, the parameters of \( \alpha_{kij} \) and \( \lambda_{kij} \) and the selection of travel time thresholds are calculated as follows. Let \( X_{ij}(t) \) be a random variable (RV) represent the travel time on link \((i,j)\) arriving at node \(i\) from ingoing link \((k,i)\) in time \(t\). Denote the travel time on \((k,i)\) of a particular support point \(w_n \in W\) by \(x_{ki}\). We define the follow sets

\[
C = \{w_n | P(X_{ki}(t) = x_{ki} \leq \gamma_{ki}(t)) \}\}
\]

and

\[
\bar{C} = \{w_n | P(X_{ki}(t) = x_{ki} > \gamma_{ki}(t)) \}
\]

\[
C' = \{w_n | P(X_{ki}(t) = x_{ki} \leq \gamma_{ki}(t)) \cap P(X_{ij}(t) = x_{ij} \leq \gamma_{ij}(t)) \} \subseteq C
\]

and

\[
C^* = \{w_n | P(X_{ki}(t) = x_{ki} > \gamma_{ki}(t)) \cap P(X_{ij}(t) = x_{ij} \leq \gamma_{ij}(t)) \} \subseteq \bar{C}
\]

By choosing appropriated thresholds \( \gamma_{ki}(t) \) and \( \gamma_{ij}(t) \) in each time interval, we can have

\[
\alpha_{kij} = \frac{\sum_{w_n \in C'} p_n}{\sum_{w_n \in C} p_n} \quad \text{and} \quad \lambda_{kij} = \frac{\sum_{w_n \in C^*} p_n}{\sum_{w_n \in \bar{C}} p_n}
\]

which is time invariant.

The conditional average flow speeds are calculated as

\[
v^U_{h(k,i)} = \frac{i_{kl} \sum_{w_n \in C} p_n}{\sum_{w_n \in C} p_n x_{kl}} \quad \text{and} \quad v^C_{h(k,i)} = \frac{i_{kl} \sum_{w_n \in \bar{C}} p_n}{\sum_{w_n \in \bar{C}} p_n x_{kl}}
\]

### 3.3 Calculating the shortest path

Now, assume that we are arriving at node \(i\) at time 0 and considering what route to take in order to travel to the destination \(D\). Moreover, assume that the previous link to node \(i\) is uncongested. In this example, we choose mean values as travel time thresholds for link \((1,1)\), link \((i,2)\), link \((1,D)\) and link \((2,D)\). That means, the travel time which is no more than mean value is uncongested, and congested otherwise. Consequently:

\[
\gamma_{i1}(t) = \sum_{n=1}^{3} p_n x_{i1} = 0.5 \times 4 + 0.3 \times 4 + 0.2 \times 5 = 4.2, t \in [\tau_0, \tau_1)
\]

\[
\gamma_{i2}(t) = 5.2, \gamma_{1D}(t) = 4.8, \gamma_{2D}(t) = 4.2, t \in [\tau_0, \tau_1)
\]

The average flow speeds under uncongested situation for link \((i,1)\) and link \((i,2)\) is:
The average flow speeds under congested situation for link \((i, 1)\) and link \((i, 2)\) is:

\[ v_{0(i,1)} = l_{i1} \times \left( \sum_{w_n \in E} p_n x_{i1} / \sum_{w_n \in E} p_n \right)^{-1} = (4 \times (0.5 + 0.3) / (0.5 + 0.3))^{-1} = \frac{1}{4} \]

\[ v_{0(i,2)} = l_{i2} \times \left( \sum_{w_n \in E} p_n x_{i2} / \sum_{w_n \in E} p_n \right)^{-1} = (5 \times (0.5 + 0.3) / (0.5 + 0.3))^{-1} = \frac{1}{5} \]

The average flow speeds under congested situation for link \((i, 1)\) and link \((i, 2)\) is:

\[ v_{0(i,1)} = l_{i1} \times \left( \sum_{w_n \in E} p_n x_{i1} / \sum_{w_n \in E} p_n \right)^{-1} = (5 \times 0.2 / 0.2)^{-1} = \frac{1}{5} \]

\[ v_{0(i,2)} = l_{i2} \times \left( \sum_{w_n \in E} p_n x_{i2} / \sum_{w_n \in E} p_n \right)^{-1} = (6 \times 0.2 / 0.2)^{-1} = \frac{1}{6} \]

Accordingly, for the second time interval:

\[ y_{i1} (t) = 15.2, y_{i2} (t) = 14.6, y_{1D} (t) = 15.4, y_{2D} (t) = 15, t \in [\tau_1, \tau_2] \]

In the mean time, the average flow speeds under uncongested situation for link \((i, 1)\), link \((i, 2)\), link \((1, D)\) and link \((2, D)\) is:

\[ v_{1(i,1)} = \frac{1}{15}, v_{1(i,2)} = \frac{1}{14}, v_{1(1,D)} = \frac{1}{15}, v_{1(2,D)} = \frac{1}{14} \]

Accordingly, the average flow speeds under congested situation for link \((i, 1)\), link \((i, 2)\) is:

\[ v_{c(i,1)} = \frac{1}{16}, v_{c(i,2)} = \frac{1}{16} \]

Similarly, under congested situation,

\[ v_{c(1,D)} = \frac{1}{17}, v_{c(2,D)} = \frac{1}{16} \]

In uncongested situation, the average expected travel time on link \((i, 1)\) and link \((i, 2)\) can be calculated as follows:

\[ U_{i1} (t) = \tau_1 + \frac{\left( l_{i1} - v_{0(i,1)} \right) \left( \tau_1 - \tau_0 \right)}{v_{1(i,1)} - \tau_0} = 3 + \frac{\left( 1 - \frac{1}{4} \times 3 \right)}{15} - \frac{0}{15} = 6.75, t \in [\tau_0, \tau_1] \]

\[ U_{i2} (t) = \tau_1 + \frac{\left( l_{i2} - v_{0(i,2)} \right) \left( \tau_1 - \tau_0 \right)}{v_{1(i,2)} - \tau_0} = 3 + \frac{\left( 1 - \frac{1}{5} \times 3 \right)}{14} - 0 = 8.6, t \in [\tau_0, \tau_1] \]
Similarly, the average expected travel time on link \((i, 1)\) and link \((i, 2)\) in congested situation can be calculated as follows:

\[
C_{i1}(t) = \tau_1 + \left( l_{i1} - v_{0(i,1)}^c(\tau_1 - \tau_0) \right) / v_{1(i,1)}^c - \tau_0
\]

\[
= 3 + \left( 1 - \frac{1}{5} \times 3 \right) / 16 - 0 = 9.4, t \in [\tau_0, \tau_1)
\]

\[
C_{i2}(t) = \tau_1 + \left( l_{i2} - v_{0(i,2)}^c(\tau_1 - \tau_0) \right) / v_{1(i,2)}^c - \tau_0
\]

\[
= 3 + \left( 1 - \frac{1}{6} \times 3 \right) / 16 - 0 = 11, t \in [\tau_0, \tau_1)
\]

Then, by using the same method, at next time interval, \(U_{10}(t) = 15, U_{20}(t) = 14, C_{10}(t) = 17, C_{20}(t) = 16, t \in [\tau_1, \tau_2)\) and for the link pair: link \((i, 1)\) and link \((1, D)\), link \((i, 2)\) and link \((2, D)\),

\[
\alpha_{i1D} = \frac{\sum_{w_n \in C'} p_n}{\sum_{w_n \in C} p_n} = \frac{p_1 + p_2}{p_1 + p_2} = 1
\]

Similarly, we have \(\alpha_{i2D} = 5/7, \lambda_{i1D} = 0, \lambda_{i2D} = 0\).

We can then calculate the following:

\[
u_{i1}(t) = \alpha_{i1D} \times U_{10}(t) + (1 - \alpha_{i1D}) \times C_{10}(t) = 15, t \in [\tau_1, \tau_2)
\]

\[
u_{i2}(t) = \alpha_{i2D} \times U_{20}(t) + (1 - \alpha_{i2D}) \times C_{20}(t) = 14.6, t \in [\tau_1, \tau_2)
\]

\[
c_{i1}(t) = \lambda_{i1D} \times U_{10}(t) + (1 - \lambda_{i1D}) \times C_{10}(t) = 17, t \in [\tau_1, \tau_2)
\]

\[
c_{i2}(t) = \lambda_{i2D} \times U_{20}(t) + (1 - \lambda_{i2D}) \times C_{20}(t) = 16, t \in [\tau_1, \tau_2)
\]

Finally, the minimum expected travel time conditional on previous link is uncongested in node \(i\) is decided by \(u_{ki}(t = 0) = \min_{j=1,2} \left\{ \alpha_{kij} \left( U_{ij}(t) + u_{ij} \left( t + U_{ij}(t) \right) \right) + (1 - \alpha_{kij}) \left( C_{ij}(t) + c_{ij} \left( t + C_{ij}(t) \right) \right) \right\} = 22.68\), and the next node to take is node 1.

As we assume that \(w_2\) is the actual travel time that the driver will experience, after travelling on link \((i, 1)\) (in 4 minutes), we can decide on what next node to take to the destination \(D\) via a similar process of calculation. In a summary, the sequence of nodes chosen by the proposed approach is \(i - 1 - D\), which has the actual travel time of 19 minutes.

In the following, we compare the results of our approach to an optimal routing policy framework proposed by Gao and Chabini (8) and the approach proposed by Fan et al. (17) to compute the shortest paths in stochastic networks with travel time correlation between adjacent links. In Gao and Chabini’s framework (8), perfect online information is required\(^1\).

---

\(^1\) Note that in this paper, we compare our method with the exact algorithm DOT-SPI for the perfect online information variant presented in Gao and Chabini (8). However, Gao and
That is, the travel time realization for all links in the network at the current time is assumed to be known to the user. On the other hand, Fan et al.’s approach (17) and our proposed method do not require perfect online information. Similar to the information requirement in Fan et al.’s approach, our proposed method requires only that the state (i.e., congested or uncongested) of the link on which the user traverses to reach the decision node is known. Nevertheless, Fan et al. (17) assume that the expected link travel times only depend on the link state and are time-independent. By taking into account the time-varying aspect of link travel times in addition to the dependency on the link state, we can achieve a more accurate selection of the shortest path, as illustrated in this example. In this example, to generate the expected link travel time used in Fan et al.’s approach from the support points data, we compute the average of the link travel times in all time intervals under each support point. To simplify the presentation, we also assume that the link traversed by the user to arrive at node $i$ is uncongested.

We compare the route choice of different methods, their respective actual route travel time, and the percentage difference from the minimum route travel time. In particular, percentage difference from the travel time along the shortest route is calculated as follows:

$$\frac{\text{actual route travel time} - \text{minimum route travel time}}{\text{minimum route travel time}} \times 100\%$$

In this example, based on the assumed travel time realization (i.e., support point $w_2$), the shortest route is also $i - 1 - D$. Based on the assumption that perfect online information is available, Gao and Chabini's method chooses the shortest route. Although we do not rely on perfect online information, our method also selects this shortest route. On the other hand, Fan et al.’s (17) method selects route $(i - 2 - D)$ instead of the shortest path resulting in a 11% longer travel time compared to the minimum travel time of 19 minutes. It is because the expected link travel times depend on the arrival time in this scenario.

From this small example, we observe that in Gao and Chabini’s method, the route selection process relies on perfect online information and a large data set in the form of support points. Once the online information is given, it likely finds the most accurate shortest route among all three methods. Our proposed method instead only requires the knowledge of the state of the link traversed by the user to arrive to the decision node, but it still produces a reasonably good approximation for the optimal route choice. In particular, in this small example, the selected route is also the optimal route.

Furthermore, via this small example, we can also conclude that it is crucial to take into account the time-varying aspect of link costs. As Fan et al.’s method ignores the fact that link costs vary over time, their method fails to detect the potential changes in the state of the network.

**4. AN EXTENSIVE NUMERICAL COMPARISON**

In this section, a computational test on a larger traffic network is designed to study the effectiveness of the shortest route selection strategies. A symmetric grid traffic network with 16 nodes and 24 links is adopted (see Figure 2). Assume that node 1 is the start node and node 16 is the destination. The given network can be conceptually viewed as having three different groups of links, with group A consisting of the 12 outermost links, group C

Chabini (8) also present a number of approximation methods for the cases of no-online-information and partial-online-information settings.
consisting of the 4 innermost links, and group B representing the links connecting between the nodes in group A and those in group C and having 8 links. Specifically, \( A = \{1 \rightarrow 2, 2 \rightarrow 3, 3 \rightarrow 4, 4 \rightarrow 8, 8 \rightarrow 12, 12 \rightarrow 16, 1 \rightarrow 5, 5 \rightarrow 9, 9 \rightarrow 13, 13 \rightarrow 14, 14 \rightarrow 15, 15 \rightarrow 16\} \), \( B = \{2 \rightarrow 6, 3 \rightarrow 7, 5 \rightarrow 6, 9 \rightarrow 10, 7 \rightarrow 8, 11 \rightarrow 12, 10 \rightarrow 14, 11 \rightarrow 15\} \) and \( C = \{6 \rightarrow 7, 7 \rightarrow 11, 6 \rightarrow 10, 10 \rightarrow 11\} \).

4.1. Travel time data generation

The joint distribution of all link travel time random variables is assumed to follow a multivariate normal distribution. To generate the data, the following inputs are required: 1) the number of time periods; 2) the number of support points; 3) the common mean link travel time; 4) the common link travel time standard deviation; and 5) the common correlation coefficient of link travel time. Herein the same standard deviation (1.44) of link travel time and correlation coefficient (0.5) for each pair of links are used. Moreover, each link travel time value is rounded to the nearest integer. Whilst the overall mean travel time over the entire considered period is the same, the actual average link travel time varies in each time interval for each individual group of links \((A, B, C)\) as depicted in Figure 3. Each time interval is 15 minutes long, and we experiment with various starting times in different time intervals departing from node 1. In this example, we consider thirty 15-minute time intervals covering possible departing times in a typical weekday morning.

Similar to the previous example, the conditional probabilities \( \alpha_{kij} \) (i.e., the probability that link \((i, j)\) is uncongested given the previous link \((k, i)\) being uncongested) and \( \lambda_{kij} \) (i.e., the probability that link \((i, j)\) is uncongested given the previous link \((k, i)\) being congested) can be calculated from the above joint distribution of all link travel times.

Observe that if there is no support point data available, our proposed method still works well. For example, given historical data for link travel times, we can still approximate the expected travel times for each link at different time intervals under the uncongested and congested conditions, and also the conditional probabilities for each pair of adjacent links. Thus, our method does not critically depend on the support point data being input of the algorithm.

4.2. Experiment results

In this example, we compare between our proposed method, Gao and Chabini (8) and Fan et al.'s (17) shortest path methods. Figure 3 demonstrates the mean values for different groups in all time intervals. We carry out three separate experiments for three different settings of travel times: time interval 1 to 10, 11 to 20 and 21 to 30. In each experiment, we choose the start times to be inside the first three time intervals considered in that experiment to avoid the total travel time from exceeding beyond the considered time window. To evaluate the quality of the solutions produced by different methods, we compare the route choices computed by the three methods against the optimal route according to the selected support point. The final results are then expressed as the mean of the differences for all support points.

The comparison of the travel times for the route choices produced by different methods (our proposed method, Gao and Chabini's (8) method, and Fan et al.'s (17) method, respectively) against the minimal travel time for the optimal route from node 1 to node 16 are shown in Figures 4 under different start times. The figures plot the mean differences (log scale) for each method over all support points. From these experimental results, it is clear that, due to the requirements for extensive and accurate inputs in the form of support points
data and perfect online information, Gao and Chabini's method performs very well in these experiments, with the smallest mean differences. On the other hand, our proposed method performs quite well in comparison to the method introduced by Fan et al.'s (17). In particular, since their method uses the same expected link travel times (under a given condition) across all time intervals, their method always selects a route going through the outermost links (i.e., route 1 → 2 → 3 → 4 → 8 → 12 → 16) regardless of the start times. On the other hand, our method adapts to the changing states of the network. Subsequently, while in the earlier time intervals, our method may choose a route going through the innermost links, during the peak hours (i.e., middle time intervals), our method tends to avoid those routes going through the innermost links and select routes that mostly go through the outermost links or through the connecting links.

To summarize, the results in these experiments are fairly consistent with the intuition about the advantages of our proposed method. In most cases, our method shows smaller mean differences than Fan et al.'s method indicating that considering time-dependent travel time does yield significant advantages over methods that ignore this aspect of travel time (e.g., Fan et al.'s method). It is worth noting also that, in general, most of the mean difference values for our proposed method are relatively small in most time intervals (mostly less than 1% with the exception of one time interval that reaches almost 4%). The results show that the proposed approach is reasonably effective and can provide a good approximation method for computing shortest path for stochastic time-dependent networks taking advantage of the information on link travel time correlation.

5. CONCLUSIONS AND FUTURE WORK

We have developed a simple approximate framework for finding the least expected travel time route from any node to any given destination in a stochastic and time-dependent network. Both spatial and temporal link travel time correlations are considered in the process of selecting the optimal route taking into account the information (i.e. link congestion level) available at the time of making the routing decision. A solution based on the principal of dynamic programming is derived where the optimal path is gradually formed following the real-time routing decisions at each node along that path. We showed via numerical examples that the optimal path depends strongly on both spatial and temporal correlations of travel times. Furthermore, comparing with existing work assuming full knowledge of network travel times distribution, we showed that the proposed framework can provide good results despite its simplicity and utilizing a very small set of network parameters.

For future work, it is interesting to test the proposed framework using real traffic data to see if the effectiveness of our method is preserved, and to study the impact of estimation errors on a decision regarding route choice and its resulting travel time.

ACKNOWLEDGEMENTS

This work was supported by the Australian Research Council (ARC) Future Fellowships grant FT120100723.

REFERENCES


TABLE 1 Joint travel time realizations of all links

FIGURE 1 Simple network representation

FIGURE 2 Larger network representation (16 nodes, 24 links)

FIGURE 3 Average link travel times of different link groups at different departing (start) times

FIGURE 4 Average difference in travel time using various methods (a) first 10 time intervals, (b) middle 10 time intervals and (c) last 10 time intervals
<table>
<thead>
<tr>
<th>Time interval</th>
<th>Link</th>
<th>$w_1$</th>
<th>$w_2$</th>
<th>$w_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>(i, 1)</td>
<td>4</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>(i, 2)</td>
<td>5</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>(1, D)</td>
<td>5</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>(1, 3)</td>
<td>4</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>(3, D)</td>
<td>3</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>(2, D)</td>
<td>4</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>(2, 4)</td>
<td>4</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>(4, D)</td>
<td>4</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>(i, 1)</td>
<td>15</td>
<td>15</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>(i, 2)</td>
<td>14</td>
<td>16</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>(1, D)</td>
<td>15</td>
<td>15</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>(1, 3)</td>
<td>16</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>(3, D)</td>
<td>15</td>
<td>16</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>(2, D)</td>
<td>14</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>(2, 4)</td>
<td>15</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>(4, D)</td>
<td>14</td>
<td>16</td>
<td>15</td>
</tr>
</tbody>
</table>
FIGURE 1 Simple network representation
FIGURE 2 Larger network representation (16 nodes, 24 links)
FIGURE 3 Average link travel times of different link groups at different departing (start) times
FIGURE 4 Average difference in travel time using various methods (a) first 10 time intervals, (b) middle 10 time intervals and (c) last 10 time intervals