Real-Time Route Guidance in Stochastic Time-Dependent Traffic Networks

A thesis submitted for the degree of
Doctor of Philosophy

by

Wei Dong

Faculty of Science, Engineering and Technology
Swinburne University of Technology
Australia

2014
Abstract

In-vehicle route guidance systems have become significant in alleviating congestion in urban transportation networks. Great development has taken place in this field in recent years and route guidance systems are becoming increasingly popular among vehicle drivers.

This thesis studies the problem of real-time route guidance in stochastic time-dependent traffic networks. Real-time information plays a significant role in routing decisions by online adaptation to live traffic conditions. The aim of the thesis is three-fold: 

1) the real-time determination of routes with minimum expected travel time for any connected origin-destination (OD) pair; 
2) the real-time determination of reliable routes among all feasible routes of any connected OD pair and 
3) travel time distribution analysis based on real data and evaluation of proposed methods and other related methods.

The thesis first develops a simple and robust framework to find in real-time the least expected travel time route for any OD pair in a stochastic and time-dependent network. Both spatial and temporal link travel time correlations are considered in the proposed framework. In particular, the spatial correlation is represented by a Markovian property of the link states where each link is assumed to be in either a congested or an uncongested condition. 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 policy where, at any decision
node within a network, a decision about the next link to be taken is made based on current traffic information to achieve the shortest expected travel time towards the destination. A number of tests are included to show the effectiveness of the developed methods, first based on simulated networks, and second, with respect to part of the Melbourne city network using data extracted from the urban traffic signal control system SCATS. The test results show that the proposed approach is reasonably effective in determining shortest paths in stochastic time-dependent networks, taking advantage of the information on link travel time correlations.

Secondly, this thesis is concerned with investigating reliable route guidance issues. Travellers are interested in both minimizing travel times and reducing travel time variability. In particular, travel time variability characterizes travel time reliability; the smaller the variability of the route travel time, the more reliable the route. Travel time reliability is strongly related to transport efficiency, and has a significant impact on the quality of life. Therefore, travel time minimization and travel time reliability are of equal importance to any practical route guidance system. Two approximation methods are developed in this thesis to calculate the reliability of a route travel time in a stochastic time-dependent traffic network. The proposed methods take into consideration the probabilistic nature of the travel time on individual links and their spatial correlation between adjacent links of a selected route. Reliability calculations and the accuracy of our approximation are discussed via a simple illustrative example and a larger network with synthesized data. Then travel time reliability indexes, including previously proposed models in the literature, are calculated based on the estimated travel time obtained from SCATS data and are compared and discussed. The test result shows that our proposed approximation method for travel time reliability calculation can provide good results.

Travel time is an important parameter to be considered when using a route guidance system. In this thesis, detectors data are extracted from the urban traffic signal control system SCATS. However, detectors output data do not directly provide travel time
information. Therefore, estimating travel times using these data is the first step for
the evaluation of our route guidance methods. Then link travel time distributions
are obtained by certain travel time estimation method. A detailed analysis of travel
time distribution for all the selected routes in the selected fragment of the Melbourne
network is also studied in this thesis. Here we use the goodness-of-fit test to check the
validity of the assumptions regarding the specific travel time distribution, e.g., normal,
log-normal or gamma. Based on the estimated travel time obtained from SCATS data
in the Melbourne traffic network, travel times best follow the log-normal distribution.

Thesis Supervisors: Quoc Bao Vo and Hai L. Vu
Acknowledgements

First of all, Associate Professor Quoc Bao Vo and Professor Hai L. Vu, my great supervisors, motivating my interest in learning the transport engineering field. At the beginning of the research, I did not know how to choose an interesting topic from numerous aspects and which method of research to use. I would like to take this opportunity to express my deepest gratitude to them for their valuable guidance during my PhD study. They have spent countless hours going through every detail of my research work, giving me feedback on my research, and teaching me to do the research independently. This made my journey through the PhD program much easier.

I thank Swinburne University of Technology and the Faculty of Science, Engineering and Technology for offering me a full research scholarship throughout my doctoral program. I also thank the Research Committee of the Faculty of Science, Engineering and Technology for research publication funding support and for providing me with financial support to attend academic conferences.

I would also like to thank to the Victorian Road Authority, VicRoads for providing the real-time traffic data in the Melbourne area, together with Judith Grundy, a professional editor, for correcting my English language usage and grammatical errors in my thesis.

Back at Swinburne, my thanks also go to staff members, research students and research assistants at previous CITR/CS3 and current SUCCESS for their help, suggestions, friendship and encouragement. In particular, I would like to acknowledge Professor Yun Yang, Professor Chengfei Liu, Professor TY Chen, Dr Wei Lai, Gillian Foster, Sharon Raj, Vicki Redrup, Xiao Liu, Miao Du, Gaofeng Zhang, Qiang He, Feifei Chen, Jing Gao, Xiaoyuan Xie, Daihai Cao and Wenhao Li. Thanks for all of your
encouragements, nice words, jokes, discussions, and more. I appreciated the weekly game of badminton. During my studies in Australia, I have acquainted myself with a lot of new friends in and out of Swinburne. I would like to express my thanks to them for making my life wonderful and memorable. They are Wei Xie, Tian Guo, Zhiyuan Liu, Wenhui Duan, Zhu Pan, Nanfei Cai, Ran Gu, Yuli Wei and Zhi Li. I cannot imagine what the life in a foreign country would be like for me without my friends. I would like to thank them for making my life enjoyable and unforgettable. I would also like to thank Associate Professor Majid Sarvi and Dr Andreas Hegyi, for being my thesis viewers and providing me with helpful and valuable comments on this thesis.

In addition, I wish to thank Dr Dong Yuan, Dr Jianxin Li, Dr Huai Liu and Dr Rui Zhou for their critical comments and valuable suggestions during my PhD study. They helped me a lot and always encouraged me to persist when coming across difficulties. Most importantly, Dr Minyi Li has been endlessly motivating and encouraging. Without her consistent support and encouragement, I would not have been able to complete this manuscript. In fact, to me, Minyi is not only a colleague but also an important friend of mine.

Finally, my deepest gratitude goes to my family, my parents, my aunts and my uncles for their love, understanding, constant support, help and encouragement. They have always believed in me and encouraged me to achieve with great love and encouragement. This work is dedicated to them.

Wei Dong
Melbourne, Australia
Declaration

This is to certify that this thesis contains no material which has been accepted for the award of any other degree or diploma and that to the best of my knowledge this thesis contains no material previously published or written by another person except where due reference is made in the text of the thesis.

Wei Dong
Contents

Abstract 1

Acknowledgements 4

Contents 7

List of Figures 11

List of Tables 13

1 Introduction 15

1.1 Research Questions 22

1.2 Aims of Research 25

1.3 Contributions of the Thesis 26

1.4 Outline of the Thesis 28

2 State-of-the-Art Optimal Route Guidance 30

2.1 Introduction 30

2.2 Route Guidance Issues 31

2.2.1 Static shortest path algorithms 32

2.2.2 Shortest path algorithms in stochastic networks 34

2.2.3 Real-time route guidance 38

2.3 Dynamic Traffic Assignment and Stochastic Traffic Assignment 39
3 Shortest Paths Finding in Stochastic Time-dependent Networks Considering Link Travel Time Correlation

3.1 Introduction

3.2 Methodological Framework

3.2.1 Problem formulation

3.2.2 Route choice heuristics

3.3 An Illustrative Example

3.3.1 The simple network

3.3.2 Parameters calculation

3.3.3 Calculating the shortest path

3.4 A Numerical Evaluation

3.4.1 Travel time data generation

3.4.2 Experiment results

3.5 Summary of this Chapter

4 Reliability in Stochastic Time-dependent Traffic Networks with Correlated Link Travel Times

4.1 Introduction

4.2 Reliable Route Guidance Problem Description

4.2.1 Network representation

4.2.2 Reliability calculation using the conditional expectation (CE) approximation
4.2.3 Reliability calculation using travel time distribution of the sum of two adjacent (STA) link travel times . . . . . . . . . . . . . . 92

4.3 Experiment Result . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 95
  4.3.1 An illustrative example . . . . . . . . . . . . . . . . . . . . . . . 96
  4.3.2 Calculation from the exact distribution . . . . . . . . . . . . . . 97
  4.3.3 Calculation using CE approximation method . . . . . . . . . . . 98
  4.3.4 Calculation using STA approximation method . . . . . . . . . . 102
  4.3.5 Experiment in a larger network . . . . . . . . . . . . . . . . . . . 105

4.4 Summary of this Chapter . . . . . . . . . . . . . . . . . . . . . . . . . 110

5 Validation and Discussion Using Real Traffic Data 111
  5.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 112
  5.2 Test Network . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 114
  5.3 Traffic Data . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 114
  5.4 Analysis of Travel Time Distribution based on Real Traffic Data . . . 122
  5.5 Numerical Tests . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 130
    5.5.1 Test results for the shortest path . . . . . . . . . . . . . . . . . 131
    5.5.2 Test results for reliable route guidance . . . . . . . . . . . . . 133
  5.6 Summary of this Chapter . . . . . . . . . . . . . . . . . . . . . . . . . 140

6 Conclusions and Future Work 141
  6.1 Conclusions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 141
  6.2 Future Work . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 143

Publications of the Author 146
List of Figures

1.1 ITS architecture (source: MIRA Technology) ........................................... 21
2.1 Schematic diagram of a network with two intersections (source: Cheu et al. [25]) .......................................................... 46
2.2 OD pair and related paths ................................................................. 52
3.1 LETT route selection framework ....................................................... 62
3.2 Simple network representation .......................................................... 65
3.3 Larger network representation (16 nodes, 24 links) ......................... 74
3.4 Average link travel times of different link groups at different departing (start) times ......................................................... 75
3.5 Average difference in travel time using various methods (first 10 time intervals) .............................................................. 76
3.6 Average difference in travel time using various methods (middle 10 time intervals) ............................................................ 77
3.7 Average difference in travel time using various methods (last 10 time intervals) ................................................................. 77
4.1 A small network ............................................................................. 97
4.2 Sioux Falls network ....................................................................... 106
4.3 Mean travel times of three random selected links at different time interval

5.1 Test case scenario routes of the Melbourne traffic network

5.2 Position of SCATS loop detectors (e.g., the detectors of the west-east approach are coded as 8, 9, 10)

5.3 A sample extract of SCATS data

5.4 Average difference in travel time using various methods in a real test

5.5 Average difference in travel time in a real test

5.6 Comparison between various travel time reliability measures at 6am (Kaparias et al.’s (62) method has zero total difference)

5.7 Comparison between various travel time reliability measures at 7am (Kaparias et al.’s (62) method has zero total difference)

5.8 Comparison between various travel time reliability measures at 8am (Kaparias et al.’s (62) method has zero total difference)

5.9 Comparison between various travel time reliability measures at 9am (Kaparias et al.’s (62) method has zero total difference)

5.10 Average results for comparison between various travel time reliability measures (Kaparias et al.’s (62) method has zero average difference)
## List of Tables

1.1 ITS services (source: Wong [123]) ............................................ 21

3.1 Joint travel time realisation of all links ................................. 66

3.2 Comparison among different route selection methods .............. 72

4.1 Joint travel time realization of all links ................................. 98

4.2 Evidence of correlation exists for all possible routes .............. 103

4.3 Analytical results for all possible routes ............................... 104

4.4 Data for Sioux Falls network ............................................... 108

4.5 Analytical results for a larger network for the first random selected start time ............................................ 108

4.6 Analytical results for a larger network for the second random selected start time ............................................ 108

4.7 Analytical results for a larger network for ten random selected start times ............................................ 109

5.1 Travel time comparison between ATTM model and Google Traffic ... 121

5.2 Contingency table for observed values and theoretical values ....... 124

5.3 Expected event numbers for observed values and theoretical values .... 124

5.4 The Kullback-Leibler Divergence test for the actual and transformed travel time distribution (start time 6am) ................... 125

5.5 The Kullback-Leibler Divergence test for the actual and transformed travel time distribution (start time 8:30am) .............. 126

5.6 The Chi-square test for the actual and transformed travel time distribution (start time 6am) ................................. 127
5.7 The Chi-square test for the actual and transformed travel time distribution (start time 8:30am) ................................. 128

5.8 The Percent Relative Difference test for the actual and transformed travel time distribution (start time 6am) ......................... 129

5.9 The Percent Relative Difference test for the actual and transformed travel time distribution (start time 8:30am) ................................. 130
Chapter 1

Introduction

With the rapid development of social economics, transport and traffic issues are an increasing concern for our society. Many countries are suffering from serious traffic congestion due to the growth of vehicle ownership and urbanization. Traffic congestion poses a number of social, environmental and economic challenges for many major cities around the world. Congestion largely increases traffic delays, fuel consumption, greenhouse gas emissions, air pollution and traffic accident rates. Congestion also lengthens journey times and decreases travel reliability. According to recent statistics, motor vehicle travel in Australian cities has grown significantly over the last decade, and is expected to continue to grow appreciably over the next 15 years [17]. This will impose considerable pressure on traffic networks.

Usually, traffic congestion can be classified into two categories: recurrent congestion and non-recurrent congestion. Recurrent congestion occurs nearly everyday, and the typical one is traffic congestion. During the congestion situation, the total number of drivers who need to travel in the transportation network increase and the capacity of a certain link or route deteriorates. Non-recurrent incidents include, as examples, traffic incidents, vehicle breakdowns, poor weather conditions, natural disasters and road section maintenance work. For example, vehicular crashes, breakdowns and debris
in travel lanes are the usual situations of incidents. As to special events, the travel
demand will change suddenly and make the network situation pressing, which usually
exceed the usual volume. All these lead to fluctuations in road capacity, randomly
influence the performance of a road network and increase the network instability.

In order to attenuate congestion effects, there are two main approaches. The first is
to construct new infrastructure or widen existing roads. This requires significantly high
costs, however, and there is also less and less space for new construction or expansion
due to the population growth and urbanization. The second is to make the best use of
existing infrastructure with the aid of an intelligent transportation system (ITS). ITS
integrates techniques of detection, signal processing, computation, automatic control,
and communications into the world of road infrastructures and vehicles. This is to
reduce congestion, improve safety, increase traffic productivity and decrease traffic
emissions. Some schematic diagrams of ITS are displayed in Figures 1.1. ITS provides
a wide range of services for travellers and opens up new ways of improving air quality
and energy conservation in our society [112]. One of the aspects of ITS is to assist
the users to make informed decisions by enhancing their spatial knowledge of the
traffic system. Some of the famous ITS services include Advanced Traffic Management
Systems (ATMS), Advanced Traveller Information Systems (ATIS), and Advanced
Public Transport Systems (APTS) [123]. Table 1.1 shows its widely used applications.
For instance, ATIS aims to provide travellers with up-to-date information regarding
network conditions to help that traveller avoid traffic congestion by utilizing better
information. It will be beneficial to travellers who plan to take a route where an
incident has occurred when a timely message can be provided by ATIS.

As an important part of ATIS, in-vehicle route guidance systems are designed to
improve travel flows and ease traffic congestion by informing travellers of real-time
conditions (e.g., congestion and accidents). Generally, the use of navigation systems
leads to user equilibrium. All used routes with the same OD are with the same travel
time, in which no user can lower his travel time by unilaterally shifting to another route.
In terms of the whole network, it is more likely that traffic congestion can be alleviated. In-vehicle navigation systems were initially used by a small proportion of people because these systems were expensive and mainly installed in high-priced cars. However, with the increasing demand for these goods and the cheaper cost of this system, the situation has changed and navigational systems have now become daily necessities to make travel easier [60]. Traditional route guidance systems were developed to typically find the shortest paths in terms of distance. With emerging ITS technologies, route guidance systems [31, 46, 126] can communicate with traffic control centers to obtain information of real-time traffic conditions and hence dynamically advise drivers of alternative routes [56]. The main functional components of an in-vehicle route guidance system include: the digital map database module, the positioning module, the map-matching module, the route planning module, the route guidance module, the wireless communications module and the human-machine interface module [60]. In particular, traffic conditions on most arterial roads in urban areas are well monitored in reality through frequently collected traffic information using inductive loop detectors and traffic cameras. Then in-vehicle route guidance system can receive this information via wireless communication (e.g., 3G, 4G and V2I Communications) and suggest the best route to the travellers. This thesis focuses on route guidance algorithms that are the essential part of the route planning module.

Traditional route guidance systems were developed to find the shortest path in terms of distance. Travellers take a fixed set of links along the selected route, regardless of the network conditions. In other words, most previous work focused on static situations without considering the changing traffic conditions in the road network. However, as influenced by a large number of uncertain and changing factors such as bad weather, traffic congestion, construction work, special events and day-to-day fluctuations on traffic demands, traffic conditions are time-varying and stochastic. As a consequence, travel times are stochastic time-dependent. For instance, the travel time from work to home on a particular day could be different from that on another day or on the same day of the next week. Therefore, it is a very interesting research question how an individual
traveller makes routing decisions in a stochastic and time-dependent network.

Travellers make route and departure time decisions based on their information about the current traffic network. This information can be got via a number of ways: travellers’ own experience, radio broadcast, in-vehicle route guidance system, and so on. The information can be classified as a priori and real-time [45]. A priori information is general picture of traffic status based on our knowledge of the truth. For instance, the travel time from home to city center is 15 minutes, which may vary during rush hours. Real-time information is about the traffic condition on a particular day and time, e.g. a map shows current traffic conditions in the city area using colored lines: red for heavy, yellow for medium, green for light congestion. Radio broadcast and in-vehicle route guidance system can provide real-time information. For instance, radio system can broadcast the traffic information to travellers anywhere in the radio coverage. Advanced in-vehicle route guidance systems contain travel time information under both normal and congested conditions and past incident records. Finally, the processing results are displayed as requested by travellers. Due to the stochastic characteristics of traffic network, it is significant to include real-time information when making route decisions. ATIS can provide both a priori and real-time information. As to travellers, the priori information they have is mainly related the route that they select. ATIS provides travellers a full traffic network information in the past. Nevertheless, the advantage of ATIS is mainly related to the broadcasting of real-time information, such as traffic incidents, travel time information, bad weather and work zones. It is generally believed that routing with real-time information will save travel time and enhance travel time reliability. This enables travellers to avoid being stuck in the incident link for a very long time by making alternative route choices towards their destinations.

Routing decisions with real-time information is different from those in a static network. The adaptiveness will save travel time and enhance travel time reliability. Given a start node and time, with real-time information, a decision is made on what next node to take. Under the assumption of stochastic and time-dependent traffic
network, travellers can make dynamic route decisions based on different start times rather than a constant path with a fixed of links. Then travellers may take various sets of links, depending on the network status that have been revealed during their trip. Moreover, travel times over different road sections are typically related. As to bad weather, link travel times of the whole traffic network over a certain time period are correlated. In the incident situation, link travel times are correlated near the incident location and around the incident duration. As such, how to determine the shortest paths in any stochastic time-dependent network based on travel time prediction with travel time correlation taken into account is another research question.

After understanding the primary concern of travellers on how to make adaptive routing decisions and decide the shortest travel time route from any node to any destination in a stochastic time-dependent network, another research question would be: when is travel time considered to be reliable? Travellers are also concerned about the reliability of their travel times when faced with uncertainty. Understanding travel time reliability is very important in many aspects. From the travellers point of view, travel time variability reduction means more predictable travel times. As to freight transportation, reduction of travel time variability can improve on-time deliver management. From the traffic managers point of view, less travel time variability means better stability of the network status. This will lead to less fuel emission since there are fewer vehicles undergoing acceleration/deceleration cycles [104].

The major factors affecting travel time reliability include: traffic incidents, work zones, adverse weather conditions, traffic control devices, demand fluctuation, and special events (Chen and Zhou [19]). For instance, traffic incidents are the matters that cause a part or complete block in the road section, which influence travel time variability. Work zones are connected with the construction or mending work on the roadway which could cause occasional or full road close. As to the weather conditions, the traffic network can be influenced by particular environmental situations, such as hurricane, rain and snow. Special events include sports activity in the stadium,
Christmas day and unexpected fire, which will cause extra travel time around the activity area. Some other factors such as tolls in the freeway section, parked vehicles, peak hours during the morning and evening period can also cause congestion, which lead to further influence on road network reliability. Due to the stochastic nature of traffic variables, travellers’ routing decisions will be affected and there is a need for reliable route guidance. The question is then how to establish a reliable route guidance model in a stochastic time-dependent traffic network where travellers make adaptive routing choice. Therefore, a travel time reliability definition is needed to perform the reliable route guidance. Based on the specific travel time reliability definition, this model is able to allow travellers to make adaptive reliable route selections and decide what next node to take towards the destination at the most reliable travel time.

Travellers benefit a lot from the optimal routes suggested by the route guidance system. However, if many travellers follow the same route suggested by the route guidance system, it could cause traffic congestion again. Therefore, it is interesting to evaluate the percentage of people who utilize this navigation service in the real network. The market penetration of route guidance systems, defined as the proportion of vehicles equipped with this device, has widely been recognized as an important factor to determine the actual advantage of guidance implementation. A market penetration model can be developed to study the benefits of in-vehicle route guidance systems. However, this thesis focuses on route guidance systems that determine optimal routes for travellers towards their destinations. Therefore, only route guidance algorithms are discussed in the reminder of the thesis. It will be a future work to study the benefit of route guidance systems under a certain level of market penetration and is out the scope of this thesis.

The remainder of this chapter is organized as follows. The research questions and objectives are specified in Section 1.1 and Section 1.2. Then, the main contributions to the existing knowledge and practical relevance are described in Section 1.3. Finally, the outline of this thesis is given in Section 1.4.
Chapter 1

Figure 1.1: ITS architecture (source: MIRA Technology)

Table 1.1: ITS services (source: Wong [123])

<table>
<thead>
<tr>
<th>ITS Service</th>
<th>Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advanced Traffic Management Systems</td>
<td>Traffic control</td>
</tr>
<tr>
<td></td>
<td>Incident management</td>
</tr>
<tr>
<td></td>
<td>Infrastructure maintenance management</td>
</tr>
<tr>
<td></td>
<td>Demand management</td>
</tr>
<tr>
<td></td>
<td>Policing/Enforcing traffic regulations</td>
</tr>
<tr>
<td></td>
<td>Transportation planning support</td>
</tr>
<tr>
<td>Advanced Traveller Information Systems</td>
<td>Pre-trip information</td>
</tr>
<tr>
<td></td>
<td>Route guidance and navigation</td>
</tr>
<tr>
<td></td>
<td>On-trip driver information</td>
</tr>
<tr>
<td></td>
<td>Personal information services</td>
</tr>
<tr>
<td></td>
<td>On-trip public transport information</td>
</tr>
<tr>
<td>Advanced Public Transport Systems</td>
<td>Public transport management</td>
</tr>
<tr>
<td></td>
<td>Demand responsive transport</td>
</tr>
</tbody>
</table>
1.1 Research Questions

The central problem addressed in this thesis is design of real-time route guidance system in stochastic time-dependent traffic network. The research work of this thesis is conducted according to the following research questions:

- How to describe a stochastic time-dependent network? In reality, each link travel time is shown as a random variable with time dependency characteristic.

- How to define real-time information? The real-time information can take many forms, such as link travel time, route travel time, delay time, traffic incident duration, and so on. Which information we take?

- Link travel times are not only random, but usually correlated. Then how to define correlations of link travel time?

- How to determine the shortest travel time paths in a stochastic time-dependent network based on real-time information?

- How to decide the most reliable route to take in a stochastic time-dependent traffic network?

- How to evaluate the proposed approaches with real traffic data to demonstrate the performance of the proposed frameworks and to judge its advantages and disadvantages?

A stochastic time-dependent traffic network is defined as a network where link travel times are random variables with time-dependent distributions. In this thesis, joint link travel time distributions in the network are used and assumed to be represented as a set of support points. One support point consists of the discrete values of travel time on all the links of network. Travellers are assumed to have this joint information from a set of historical data.
In the stochastic time-dependent traffic network, real-time information is assumed to be in the form of a set of link travel times at a particular time interval. Travel time can be obtained in a number of ways. For instance, it can be measured directly using probe vehicles or advanced detection technologies, such as automatic vehicle identification, automatic vehicle location, video image processing, and so on. It can also be estimated indirectly from traffic data provided by inductive loop detectors. Usually, measuring travel time directly is costly and often requires some new types of sensors. A more cost-effective way to obtain travel time is to estimate it through traffic data, particularly those provided by loop detectors that are already in place in most signalized arterial links and freeways. In particular, real data in this thesis are obtained through the loop detectors in SCATS adaptive traffic signal control system in the Melbourne network and travel time is estimated using these data.

Given an origin and destination (OD), the main concern of travellers is to determine a route with the least expected travel time for that OD pair. In the literature, some studies propose the least travel time path finding problems in a general stochastic time-dependent traffic network under the assumption of knowledge of the full joint distribution of travel times across the entire network. However, it is difficult to obtain such travel time data in reality. Therefore, a general path finding model in stochastic time-dependent networks will be developed that requires much smaller set of parameters (e.g., the conditional average link travel time (or the average flow speed) in a given link state and time interval, and the conditional transition probabilities among states of two consecutive links in the network) than the previous method, without significantly compromising on the accuracy of the results. The framework will deal with congested or uncongested conditions of link congestion for clarity and can be easily extended to more than two possible levels with the increase of computational burden.

In a real traffic network, travel times of neighboring links are usually highly correlated. Therefore, the correlation between links when analyzing route guidance
problems with real-time information should be considered. In particular, in this thesis, two types of travel time correlation (temporal and spatial correlation) are studied in depth. The spatial correlation is represented by a Markovian property of the link states where each link is assumed to experience multiple possible conditions. The temporal correlation is manifested through the time-dependent expected link travel time, given the condition of the link traversed.

Furthermore, the shortest paths determined are essentially the predictive shortest paths and may be or may not be achievable in reality. For this reason, travellers do not only wish to identify the shortest paths but also their travel time reliability. The travel time reliability definition, coefficient of variation [8] is utilized in this thesis, where average and standard deviation values are combined together in a ratio to produce a value as a travel time reliability indicator. Coefficient of variation provides a clearer picture of the trends and performance characteristics by taking into account both trip lengths and standard deviation.

Two approximation methods are developed in this thesis to obtain the reliability of a route travel time in a stochastic time-dependent traffic network. To be in detail, path travel time variance value is calculated using individual link travel time variances and covariance values between adjacent links. Then the reliability of a route travel time can be obtained and the most reliable route is calculated by minimizing coefficient of variation values in a stochastic time-dependent traffic network.

The evaluation of the proposed approach with real traffic data is a very important task as it is an effective method to demonstrate the performance of the proposed framework and to judge its advantages and disadvantages. Therefore, the proposed frameworks are evaluated with respect to part of the Melbourne city network using a large amount of data collected from the loop detectors. The performance of our proposed methods are compared with a number of existing methods in terms of route selection.
1.2 Aims of Research

The objective of this research is to propose a real-time route guidance system in stochastic time-dependent traffic network. This system provides accurate suggestions in a timely manner to help motorists navigate to their destinations by applying advanced surveillance, telecommunication and computer technologies. Here, a stochastic time-dependent traffic network is defined as a network where link travel times are random variables with time-dependent distributions. Real-time information is assumed to be in the form of the travel times of a set of links during a set of time periods. These real-time link travel times can be obtained either directly through probe vehicles and cameras or indirectly through loop detectors. The final goal is to minimize the actual travel time or to maximize the reliability of the route choice practice. This includes the evaluation of the proposed system in terms of suitably defined performance metrics and comparison with other existing route guidance methods. The evaluation of the proposed system is conducted with respect to a number of tests based on simulated networks and a real traffic network using real data.

More specifically, the research work of this thesis aims to:

- find a suitable approach to determining routes of the least expected travel times in consideration of temporal and spatial travel time correlations
- analyze travel time distribution according to real traffic data
- propose a reliable route guidance framework and develop solution algorithms
- evaluate the proposed frameworks with respect to part of the Melbourne city network using a large amount of real data collected from the network
1.3 Contributions of the Thesis

The research work in this thesis can be divided into three parts addressing real-time route guidance systems in stochastic time-dependent traffic network. The first part of this thesis focuses on finding the route with the least expected travel time from any node to any destination in a stochastic time-dependent network with dynamic programming approach. Then, in the second part of this thesis, two approximation methods to obtain the reliability of a route travel time in a stochastic time-dependent traffic network are proposed considering both spatial and temporal link travel time correlations. Finally, in the last part of this thesis, the evaluation of the proposed approaches using SCATS data in part of city network of the Melbourne in Australia is conducted to test its performance. The major contributions of this work are summarized as follows:

**Shortest route finding**

- A general framework is developed for shortest-time route guidance with real-time information. Both types of widely known link travel time dependencies, namely temporal dependence and spatial dependence, are incorporated. Temporal dependence refers to link travel times variation over time; spatial dependence refers to the correlation between travel time realizations on consecutive links that make up a route. An algorithm is devised 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 a few parameters, such as the conditional average link travel time (or average flow speed) values in a given link state and time interval, and the conditional transition probabilities among states of the two consecutive links in the network. It will be shown, with examples, that the framework can achieve similar accuracy with a much smaller set of parameters compared with when a full joint distribution of network travel times is required.
• The temporal correlation is treated in this framework without a significant increase in the 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 network spatial dependencies. It also improves the route choices whenever there is a temporal correlation.

• The real traffic network evaluation results show that the proposed approach is effective and can provide a good approximation method for computing shortest path in stochastic time-dependent networks taking advantage of the information on link travel time correlations.

Reliable route finding

• A route finding model is developed with travel time reliability incorporated as an additional route searching criterion.

• To speed up the calculation, two approximation methods are proposed to determine reliable route choice in a stochastic time-dependent traffic network.

• The evaluation results of our proposed reliable route guidance framework with the estimated travel time obtained from SCATS data in the Melbourne network show that our approximation method is reasonably effective and accurate.

Analysis of travel time distribution

• SCATS data in the Melbourne traffic network is analyzed to study travel time distribution in a typical urban traffic network.

• Kullback-Leibler Divergence, Chi-Square and Percent Relative Difference tests are selected as a test of goodness of fit. Based on the estimated travel time obtained from SCATS data, travel time distribution in the Melbourne network best fits with log-normal distribution.
1.4 Outline of the Thesis

This thesis consists of six chapters. In Chapter 2 we provide a literature review of optimal route guidance problems and discuss various shortest path algorithms, travel time distributions and urban travel time estimation methods, common travel time reliability measures. The existing models or proposed approaches are discussed in detail with detail insights and limitations analysis. In Chapter 3 we discuss a simple, robust framework for the shortest path finding problem in a stochastic and time-dependent environment. Both spatial and temporal correlations in link travel times are considered and a dynamic programming approach is employed to solve the problem. In particular, the spatial correlation is represented by a Markovian property of the link states where each link is assumed to experience multiple possible conditions. The temporal correlation is manifested through the time-dependent expected link travel time, given the condition of the link traversed. 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 compared with some other methods. Furthermore, the shortest paths determined are essentially the predictive shortest paths and may be or may not be achievable in reality. For this reason, travellers do not only wish to identify the shortest paths but also their travel time reliability. Two approximation methods of a reliability-based route selection framework are proposed in Chapter 4 to study travel time reliability for any route in a stochastic time-dependent traffic network. The proposed methods take into consideration the probabilistic nature of the travel time on individual links and their spatial correlation between adjacent links of a selected route. Finally, the proposed methods are compared with the exact values calculated from the entire joint distribution of travel times through an illustrative example. A case study of evaluating the proposed frameworks with SCATS data in the Melbourne traffic network is conducted in Chapter 5. Based on SCATS data in Melbourne, Australia, we test our proposed shortest path and reliable route guidance frameworks and compare them with several other methods. A detailed analysis of
travel time distribution, using Kullback-Leibler Divergence, Chi-Square and Percent Relative Difference tests based on the estimated travel time obtained from SCATS data, is also studied in this chapter. It concludes that travel time distribution in the Melbourne network best fits with log-normal distribution. This thesis is concluded in Chapter 6 with a number of closing remarks and some future works outlined.
Chapter 2

State-of-the-Art Optimal Route Guidance

In this chapter, literature describing optimal route guidance problems is reviewed. Results most relevant to the work in this thesis are presented. It aims to provide the necessary background knowledge that will serve us throughout this thesis. Moreover, this chapter intends to provide an overview of the methodological approaches to route guidance problems.

2.1 Introduction

As an important part of ITS, in-vehicle route guidance systems are widely used by travellers. The main function of a route guidance system is to find an optimal path between any origin and any destination at any time in a road network. The path that is considered as optimal is based on a selected objective (e.g., shortest travel distance, shortest travel time, maximum probability of arriving on time, or most reliable travel time). In essence, a route guidance problem is a shortest path problem that can be solved by various shortest path algorithms. Due to the dynamic and stochastic nature
of traffic networks, the reliability of identified shortest or efficient paths is of much concern, hence, what a traveller prefers is reliable route guidance advice.

As such, this dissertation focuses on in-car navigation type of route guidance, studies and evaluates the performance of real-time route guidance systems. The main objectives of this chapter are to review the frameworks currently available for real-time route guidance, to identify the main limitations of previous methods and to suggest approaches for improvement.

This chapter is organized as follows: Section 2.2 describes some background knowledge of shortest path algorithms. Then dynamic traffic assignment and stochastic traffic assignment models are described in Section 2.3. Section 2.4 introduces some related work about travel time distribution and discusses methods to estimate urban link travel time values from the loop detector data. The chapter is closed with a detailed description of travel time reliability.

2.2 Route Guidance Issues

With the rapid development of ITS, real-time route guidance is gaining increasing interest in providing drivers with turn-by-turn guidance recommendation and advising them how to reach their destinations with the least cost. Theoretically, any route guidance problem can eventually be reduced to a shortest path problem. The algorithms that can be employed to the solution of the shortest path problem are referred as the shortest path algorithms. According to the performance indices, the shortest path algorithm can be classified into two categories: shortest-distance-based (static type) and shortest-time-based (dynamic type). The shortest distance path algorithms focus on route length parameters and find the shortest distance route between each origin and destination pair, while the shortest travel time path algorithm determines the path with the minimum travel time.
This section is organized as follows: We first study the shortest path algorithms in deterministic networks which is useful to the introduction of routing problems in stochastic networks. Then we proceed to shortest path algorithms in stochastic networks. After this the real-time route guidance task is discussed in some detail.

2.2.1 Static shortest path algorithms

In the static network, each link has a constant cost (distance or travel time) and we term a path with minimum cost to be the shortest path. Then the static shortest path algorithm is to find a shortest path for each origin-destination pair. The shortest-distance-based algorithms can be further divided into four classes, the one-to-one, one-to-all, all-to-one and all-to-all. The one-to-one algorithm determines the shortest path between a single origin and a single destination in a network. The all-to-one refers to finding all the shortest routes from all nodes to a single destination in a traffic network. The one-to-all shortest path problem is about searching the shortest routes from one node to all other nodes in a network. The all-to-all shortest path problem consists of finding all shortest routes between all origins and all destinations in a traffic network. Dijkstra’s algorithm [29] is the most remarkable shortest path algorithm of the one-to-all type. At each step, the algorithm chooses the node with the lowest cost from the origin and finds the shortest route from the origin node to every other node in the network. Although Dijkstra’s algorithm is popular in operations research and network optimization when calculation efficiency is not a major concern, the algorithm suffers very much from its overly-high computational cost. Therefore, it is not suitable for use in real-time in-vehicle route guidance. A predecessor of Dijkstra’s algorithm, and thus a more generalized version of it, is the Bellman-Ford algorithm [105]. This algorithm computes shortest paths from a single source vertex to all of the other vertices. Later, the Floyd-Warshall algorithm [38] is designed to solve all pairs shortest paths problems by comparing all possible paths through the graph between each pair of vertices. Dial et al. [103] examine various algorithms for calculating the shortest path from one node
to all other nodes in a network. However, a major limitation of this study is that all of the test problems are randomly generated. Similarly, Glover et al. [51] propose a partitioning algorithm for finding the shortest path from one node to the other nodes in the network. Frieze and Rudolph [39] consider the problem of finding the shortest distance paths between all pairs of vertices in a random graph. All pairs shortest paths problems can also be seen in [53, 58, 93, 107].

Due to the fact that the size of actual traffic networks is usually large, it is of vital importance to have an efficient shortest path algorithm meeting these requirements. Therefore, a considerable amount of research has been focused on various acceleration methods for standard shortest path algorithms. Normally, heuristics are an effective way of reducing computation time and a common way of achieving this is by reducing the search area. The heuristic shortest path algorithm is applicable for transportation situations that requires either a quick response or repeated calculations. Many studies aim to design different kinds of heuristic shortest path algorithms. The most popular heuristic search algorithm in the literature is the A* algorithm [23] which solves one-to-one shortest path problems. The procedure of this algorithm is very similar to Dijkstra’s algorithm, the only difference being that A* uses a heuristic function to estimate the distance between any point and the destination point in the network. It does this to determine how likely it is that any node lies on the best route. Some other approaches aiming to efficiently compute the shortest route are branch-pruning, the hierarchical approach and the bi-directional approach, among others. For instance, Ahuja et al. [4] speed up the implementations of Dijkstra’s shortest path algorithm using a new data structure. Qu and Yi [98] present a novel, modified model of a pulse coupled neural network to find the shortest path with less computational costs under parallel computation. The shortest paths from a single origin node to multiple destination nodes can be found at the same time. Later, Wang et al. [120] introduce a modified pulse coupled neural network based algorithm for the finding of the shortest path in large scale systems. Simulation results indicate that the proposed method is more efficient than Dijkstra’s algorithm in the larger scale systems. A comprehensive
summary of various heuristic shortest path algorithms that have been developed can be seen in [43].

2.2.2 Shortest path algorithms in stochastic networks

Generally, in the static network, link travel times are constant values. However, in a stochastic network, link travel time values are random variables with some distributions. If the underlying network is assumed to be static (time-independent), the link travel times remain the same after they are revealed to the travelers [45]. For instance, Mirchandani [82] presents an algorithm to compute the expected shortest travel time path between two nodes in the network given each link travel time is independent discrete probability distribution. Andreatta and Romeo [49] study the shortest path problems in a static network where the topology is stochastic. A stochastic topology is defined by a deterministic set of nodes and a random set of links. A random link can be either active or not. When it is active, it is included in the network; when it is not active, it is deleted from the network [45]. In a later study by Polychronopoulos and Tsitsiklis [95], the work of Andreatta and Romeo [49] is extended by considering link travel time correlation. It assumes that the value of the link costs is random but does not change with time. Nie and Wu [87] propose a priori shortest path problem to guarantee a given likelihood of arriving on-time in a stochastic network. The proposed method determines the latest possible departure time, and the associated route, to attain a given probability of arriving at the destination at a specified arrival time or earlier.

Unlike the shortest path algorithms that address route guidance for stochastic static transportation networks, the shortest path algorithms in stochastic time-dependent networks focus on realistic situations in which travel demand, traffic conditions and various random disturbances (i.e., bad weather, incidents, special events, vehicle breakdown, work zones) are changing. Hence path travel times are not constant. In a time-dependent network, each link travel time at every time period is an individual
random variable, so travel times at different time intervals can vary. In this case, an optimal route is defined to be the one with the least expected travel time \([27, 28, 79, 81]\).

To deal with uncertainty and the time-varying aspect of travel times, a number of approaches have been proposed. Hall [52] studied for the first time the time-dependent version of the routing problem. An adaptive decision to be made at each decision node is based on the arrival time. An optimal dynamic programming based algorithm is proposed to find the shortest paths. This is shown on a small transit network example. It concludes that adaptive route choices are more effective than simple paths in a stochastic time-dependent network. However, it only considers the situation that link travel times are modeled as discrete-time stochastic processes and the proposed algorithm is applicable only for solving small networks problems because of computational constraints [42]. Pretolani [96] devises an algorithm for the shortest path problem with link travel times being both random and time-dependent. Here, directed hypergraphs are used to represent discrete, stochastic, time-dependent networks. Miller-Hooks and Mahmassani [80] present two efficient procedures for determining the least possible time paths in stochastic time-dependent networks. Here, the link travel times are discrete random variables with time-dependent probability density function. Later, Miller-Hooks and Mahmassani [81] propose an efficient means for determining lower bounds on the expected times of a priori least expected travel time path in stochastic time-dependent networks. Under this assumption, routing is based only on arrival times at decision nodes. A specialized label-setting-based algorithm is presented for determining the adaptive least expected travel time hyperpaths in stochastic time-dependent networks by Miller-Hooks [79]. It is shown that adaptive strategies can lead to improved routing decisions over a priori path selection. Sung et al. [114] propose a flow speed model which is an efficient method of ensuring consistency in a time-dependent forecasting model and finding the shortest route in any time-dependent network. Here, different vehicles experience different travel times in the same time interval on the link with changing speed. Fu and Rilett [42] implement finding the expected shortest paths in a traffic network where the link travel times are
modeled as a continuous-time stochastic process. A general probability-based formula for calculating the travel time mean and variance for a given route, based on the mean and variance of the link travel times, is developed. However, it cannot be solved exactly using standard shortest path algorithms. Instead, a heuristic algorithm is proposed to find the minimum expected paths. Gao et al. [47] propose adaptive route choice models according to realized arrival times at intermediate nodes in a stochastic time-dependent network. This model specifies the next node to take according to realized travel times at the time of making the decision, instead of a path with a fixed set of links.

In some traditional approaches for finding the shortest path, link costs (e.g., travel times) are assumed to be uncorrelated (e.g., see [87]). This is not practical in all the situations. In a real traffic network, travel times of neighboring links are usually highly correlated. For instance, it is easy to observe that in a typical traffic network, congestion in a link generally influences the condition of adjacent links. Moreover, accidents or bad weather can cause nearby link travel times to be highly correlated in the affected areas. Evidence of link travel time correlations have been recorded in the literature. For example, it is shown in [44] that the link covariance is non-zero and that the distribution of the correlation coefficient can be used as a performance metric of the network. In Xing and Zhou [125], inherent correlation among link travel times has been automatically represented by the sample multi-day historical traffic data using the Monte Carlo method. Similarly, various dependencies are investigated in [13] by calculating the correlations for travel speeds measured on paths with given numbers of intersections and distances using GPS measurements of floating cars. Their results show that the correlation is high for speed pairs of adjacent links and that this correlation decreases as the distance of the links increases. Therefore, these results motivate the consideration in this thesis of the correlation between links when analyzing shortest path problems with real-time information.

There are two types of travel time correlation defined in the literature [16,118], these being temporal and spatial correlation. Temporal correlation is defined as correlation
of travel times in the same link, at adjacent time intervals during the day, while spatial correlation refers to the travel time correlation between consecutive links that constitute a route. Only limited work was reported in the literature to address both temporal and spatial dependency. Nie and Wu [86] investigate the reliable shortest path problem, incorporating spatial correlation in a stochastic and time-dependent network, where the known probability density function of the link travel times is conditional on the state of the previous traversed link. A general dynamic programming is used to find an optimal solution to the proposed problem. Chan et al. [18] introduce methods for estimating arterial link travel time using both real-time automatic vehicle identification data and the historical (off-line) data. The spatial covariance relationships of link travel times on Hong Kong arterial roads are also calibrated by these data. A more generic algorithm to determine the route to destination with shortest expected travel time, taking into account link spatial correlation, is presented in Fan et al. [36] without time dependency consideration. Here, link travel time is defined as a random variable with probability distribution corresponding to two possible states of the link, e.g., congested or uncongested. An algorithmic approach based on dynamic programming is then proposed to find the shortest route conditioned on the status of the traversed links. Waller and Ziliaskopoulos [118] investigate the shortest path problem with limited forms of spatial and temporal dependency. For spatial dependency, each link is dependent on one predecessor link with the conditional probability matrices. The temporal dependency assumes that link cost is known when its tail node is reached. Gao and Chabini [46] present an optimal routing policy in a stochastic time-dependent network given time-dependent link travel times. They consider perfect online information to identify possible support point values for future travel time that take into account both link-wise and time-wise stochastic correlations of link travel times. Similarly, in Gao and Huang [48], a heuristic algorithm is designed for the optimal adaptive routing problem. Link travel times are correlated and a joint distribution of link travel time random variables is applied to represent the link correlation. Chen et al. [20] solve the problem of finding the reliable shortest
path by minimizing the travel time budget required to ensure a given on-time arrival probability. Here, the travel time of a link is assumed to be spatially correlated only with the neighboring links within a local impact area. The link travel time correlations are represented by variance-covariance matrices, generated by traffic flow simulators. This work considers k-limited neighboring links spatial dependence and extends the work of Nie and Wu [86] which considers only travel time correlations on the adjacent links. Ji et al. [57] formulate the spatial correlation as variance-covariance matrices which can be directly got from ATIS. Then a simulation-based method is proposed to solve the reliable shortest path problem. Nevertheless, the simulation-based method is computationally expensive and the precision of results is dependent on the maximum number of simulations [20]. Also, based on the support points, a shortest path selection method in a stochastic time-dependent network is proposed in our previous work in Dong et al. [32]. This work takes into account link travel time indeterminacy and the travel time correlation between adjacent links with real-time information.

2.2.3 Real-time route guidance

Extensive research focus has been put on real-time in-vehicle navigation systems [32, 40, 46]. In a stochastic time-dependent traffic network with real-time information, travellers can adapt to dynamic routing decisions. This decision is not a fixed route suggestion but may change over time. The best route from any given node to the final destination depends both on the node and the arrival time at that node. For example, in case of an incident in a road network, a traveller may be stuck in the traffic jam caused by the incident if they ignore real-time traffic information and take routes based on their daily experience. However, if adequate online information of the incident as well as the corresponding routing advice is available to the traveller, they can avoid being trapped in the incident link by following the route suggestion. Fu [40] studies an adaptive shortest path routing algorithm for in-vehicle navigation systems, specifying the next immediate link to take instead of a whole constant path. Real-time
information is assumed to be available about the characteristic of the actual link travel times. An efficient approximation is utilized to work out the problem and the advantage of adaptive routing systems is shown. Gao and Chabini [46] investigate the optimal routing policy problems in stochastic time-dependent networks. At each decision node, based on the real-time information and the current time, the minimum expected travel time route can be decided using the proposed algorithm. Four approximations are presented because of the complexity of the proposed algorithm. Wang et al. [121] investigate real-time route guidance in large-scale express ring roads by focusing on feedback routing performance in case of incidents. The investigating results indicate that real-time route guidance can help to ease traffic congestion. Ding et al. [30] introduce the main component of the real-time route guidance system and two different approaches to collect real-time traffic information: vehicle-to-roadside communication and vehicle-to-vehicle communication, respectively. Nadi and Delavar [83] propose a model using real-time traffic information for in-vehicle route planning. In their model, real-time traffic information in each link is detected by certain sensors and depends on the time when the vehicle arrives at that link. Recently, Dong et al. [32] discussed spatial correlations between adjacent link costs with real-time information, where the correlations between adjacent link pairs are built using conditional probability.

2.3 Dynamic Traffic Assignment and Stochastic Traffic Assignment

After understanding how an individual traveller makes route decision, it is interesting to consider the network level impact when many travellers make route decisions. It is very important to use dynamic traffic assignment (DTA) model to represent the time-dependent nature of traffic demands in congested transportation networks. DTA methods play a significant role in the ITS, which provide support to the design and operation of ATIS and ATMS [45]. DTA is a valuable analytical
technique that captures the interaction between dynamic operations and long-term adaptation by travellers, which is increasingly popular for analyzing combined planning and operational applications. DTA models long-term traveller adaptation to real-time congestion conditions in transportation systems. It also allows for consideration of a range of operational conditions, such as various weather conditions, traffic incident and travel demand variation.

There are two main approaches to tackle the DTA problem: dynamic user equilibrium and dynamic system optimum. From the travellers point of view, each traveller non-cooperatively seeks to minimize his travel time. At dynamic user equilibrium, any individual leaving his origin at any instant chooses a route that minimises his instantaneous travel time along the route to his destination. From the traffic managers point of view, the dynamic system optimum traffic assignment problem is the problem of determining time-varying link flows in a congested road network where drivers are assumed to be cooperative in minimizing total travel time.

Over the years, there have been various approaches to introducing dynamic traffic assignment models in the literature. Merchant and Nemhauser are among the first researchers to study the dynamic traffic assignment problem [77]. However, the proposed formulation is limited to deterministic, fixed-demand, single destination and system optimal case. Sadek et al. [106] develop a genetic algorithm to address dynamic traffic assignment, where link flow limitations imposed by capacity constraints are explicitly considered. The objective function is to minimize the total time that travellers spend on route to their destinations.

In a later paper by Varia and Dhingra [117], a dynamic system-optimal traffic assignment model is formulated for a congested urban road network with a number of signalized intersections. In particular, a genetic algorithm is used to minimize the overall travel cost in the network with fixed signal timings and optimization of signal timings. A policy-based stochastic dynamic traffic assignment model is developed in Gao [45] to study the network-level impact of online information provision and adaptive
routing. A generalization of Waldrop’s First Principle is utilized as the equilibrium condition. At equilibrium, each traveller follows a routing policy with the minimum perceived disutility based on his/her departure time and no one can change routing policy individually to improve the perceived disutility. A systematic comparison of static traffic assignment with a simulation-based dynamic traffic assignment approach (VISTA) is performed in [15]. The test results indicate that DTA model can account for variable demand and traffic dynamics, where traditional static model has the potential to significantly underestimate network congestion levels in traffic networks. In a recent paper, Ben-Akiva et al. [12] describe the calibration in implementing the mesoscopic dynamic traffic assignment models DynaMIT in a highly congested subarea of the city of Beijing, China. This model can replicate real traffic situations with long queues and spill backs. Comprehensive reviews of this topic can be seen in [45] and [92].

Similarly, stochastic traffic assignment models are needed to evaluate the effectiveness of various traffic management plans in case of stochastic capacity variations and travel demand fluctuations in urban traffic networks. Capacity variations are mainly resulted from traffic incidents on the road sections and travel demand varies in different times of a day, different days of a year. As such, the flow pattern which satisfies the equilibrium conditions for this problem is termed stochastic user equilibrium model.

Dial [102] proposes one of the most popular algorithm for the calculations of a logit based stochastic traffic assignment model. It can be easily applied to a large scale network, since it does not require the path enumeration over a network. However, the Dial’s algorithm sometimes generates an unrealistic flow pattern in that no flow is loaded on some paths where many vehicles are running in reality [5]. Later, a stochastic assignment model that overcomes restriction in Dial’s algorithm is introduced in [5]. In particular, Markov Chain is used as the calculation method of the proposed assignment model. Two algorithms are proposed for solving the logit based stochastic traffic assignment problems in Chen and Alfa [22]. The optimal or suboptimal step length in the search process is utilized instead of the fixed step length as in the method of
successive averages method. Nielsen et al. [89] describe a model where turn delays have been included in the solution algorithm of stochastic user equilibrium traffic assignment. A heuristic modification of stochastic user equilibrium is presented in order to consider delays at intersections. Finally, the new version of stochastic user equilibrium which includes intersection modeling is tested to see whether it provides a better description of the traffic flows than the link-based stochastic user equilibrium.

Shao et al. [109] propose a reliability-based stochastic user equilibrium traffic assignment model in view of the day-to-day demand fluctuations for multiclass transportation networks. In the model, each class of travellers has a different safety margin for on-time arrival in response to the stochastic travel times raised from demand variations. Finally, the corresponding traffic assignment problem is formulated as a variational inequality problem and a heuristic solution algorithm is developed to solve the problem. Meng et al. [76] propose a linearly constrained minimization model for stochastic user equilibrium traffic assignment problem with link capacity constraints. Later, a modified origin-based partial linearization method for solving the logit-based stochastic user equilibrium traffic assignment problem is proposed in [64]. The efficiency of the proposed solution method in computational time used and accuracy is tested with the conventional origin-based partial linearization method and the method of successive averages.

This thesis focuses on route guidance systems that determine optimal routes for travellers towards their destinations. Therefore, the route planning is of major interest and thus only route guidance algorithms are discussed in the reminder of the thesis. It will be a future work to study dynamic traffic assignment and is not in the very scope of this thesis.
2.4 Travel Time Distribution Analysis

The distribution of travel time is a significant issue when considering travel time variability and reliability. Many studies have focused on this topic in recent years. Generally we can divide them into two parts: discrete and continuous travel time distributions.

In discrete travel time distribution, travel time is a random variable with either a finite number of possible values or a countable number of possible values and corresponding probabilities. Gao and Chabini [46] solve the optimal routing policy problem in a stochastic time-dependent network, where link travel time at each time interval is modelled as random variables, with a finite number of discrete, positive and integral support points. However, when deploying the algorithm proposed for optimal route discovery, full knowledge of this probable distribution is an unrealistic assumption and the computation time is generally exponential in the number of arcs and support points. Accordingly, approximation methods are proposed in Gao [45] and a comparison among these is undertaken, both theoretically and computationally. Miller-Hooks and Mahmassani [81] showed algorithms for determining the least expected time path in stochastic time-dependent transportation networks. Link travel times are represented by discrete random variables with distribution functions that are time varying. Specific computational steps are designed to find the a priori least expected time paths from all origins to a single destination for each departure time with the associated expected times and lower bounds on the expected times of these a priori least expected time paths. A label-correcting algorithm is developed to solve this problem. Similarly, link travel times are also represented by discrete, nonnegative random variables with time varying distribution functions in Miller-Hooks [79]. An adaptive least expected time path is determined in a stochastic time varying network. This shows improved routing decisions over a priori path selection.

Compared to discrete distribution, continuous travel time distribution gains more
focus in the literature. A number of studies attempt to fit the distribution of travel times to the existing continuous probability distributions, such as normal, log-normal or gamma distribution. For instance, Wardrop [122] first points out that travel time follows a skewed distribution. Polus [94] concludes that travel times best fit a gamma distribution in a small empirical study where arterial travel time distributions are skewed to the right side. The Kolmogorov-Smirnov test is performed to find out whether the gamma distribution fits the observed data. Srinivasan and Jovanis [111] develop a heuristic algorithm to determine the number of probe vehicles required for the estimation link travel times during a peak period. A simulation model is developed to test the algorithm for the Sacramento network in the morning peak period by simplifying assumptions of normal travel time distribution. Lomax et al. [70] assume normal distribution for travel times because of simplicity. Lo et al. [69] propose an approach to relate travel time variability from stochastic network link degradations to travelers corresponding route choice behavior. The route travel time is assumed to be normal distribution according to the Central Limit Theorem. In a study on estimating path travel time reliability by Rakha et al. [99], travel times are assumed to follow normal distribution. Five methods are proposed to estimate path travel time variance from its composing segment travel time variance values. However, real traffic data are analyzed in this study and validation results show that log-normal distribution is a more appropriate presentation of roadway travel times than normal distribution. Kaparias [60] proposes two travel time reliability indices as lateness and earliness for examining reliable route finding methods under the assumption of log-normal travel time distribution. A confidence interval of the link travel time is used to indicate the minimum and maximum travel time on the link. As such, a higher indices value shows that the link is reliable. Emam and Al-Deek [33] develop a new methodology to calculate travel time reliability in a freeway corridor in Orlando, Florida. Four different travel time distributions are tested: Weibull, exponential, log-normal and normal. Through comparing the predicted and the actual travel time distribution, log-normal distribution shows the best fit with real loop detector data at the highest accuracy.
Some widely used travel time reliability measures, such as standard deviation, buffer index, and coefficient of variation, are mathematically examined for their analytical relationships in Pu [97] with the assumption of log-normal distributed travel times. Therefore, it is evident that travel time distribution format is an important factor to be considered when analyzing travel time variability and reliability.

Our proposed route guidance frameworks need urban travel time distribution. In order to establish this travel time distribution, we need to collect data for travel time. However, travel time data can hardly be measured in a real traffic network. In practice, we can use some travel time estimation models to obtain travel time data. Usually, there are two parts in urban travel time estimation (Cheu et al. [25], Ran et al. [100], Xie et al. [124], Nguyen and Gaffney [84]): cruise time and intersection delay (see Figure 2.1).

\[ \text{Arterial Link Travel Time} = \text{Cruise Time} + \text{Delay Time} \] \hspace{1cm} (2.1)

Vehicles may experience delays at a signalized intersections due to factors such as interruptions caused by traffic signal controls, traffic demand, road layout, driver characteristics, signal waiting time, acceleration and deceleration [59,110]. There are two components in the travel time estimation model. One addresses cruise time, the other is used to estimate delay time.

2.4.1 Delay estimation methods

Some models have been proposed to determine the delay value in the traffic intersection. Michalopoulos et al. [78] derive an analytical model using shock wave theory to estimate delays at signalized intersections. Akcelik and Roupail [7] propose a delay model for signalized intersections that is suitable for variable demand conditions. The major contribution of this paper is the development of an integrated framework
for an estimation process that incorporates the peaking characteristics in the demand flow pattern. Fu and Hellinga [41] build the delay patterns under two extreme traffic conditions: highly undersaturated and highly oversaturated conditions, together with an approximate model for predicting the variance of overall delay. The developed models are demonstrated through its use in a sensitivity analysis and in determining reliability-oriented optimal cycle times and levels of service. Ehsan et al. [75] use the SCATS system loop detector data to predict bus travel time. A method to estimate bus travel time is provided and a comparison with real-time GPS data of bus travel times is made. In the ARRB travel time delay model (ATTM) using SCATS data [71], travel time delay at a signalized road is made up of the following two components: uniform delay and overflow delay. Uniform delay is defined as the delay due to traffic arriving at a traffic signal with a uniform distribution. Overflow delay is an extra delay because of arrival traffic flow exceeding capacity. In particular, as a good method to estimate arterial link travel times, ATTM model is utilized in this thesis for travel time estimation. The ATTM model is introduced in detail in Chapter 5.
Chapter 2

2.4.2 Cruise time estimation methods

Several models have been proposed for arterial link cruise time or speed estimation in spite of the complexity caused by traffic signals at intersections. Normally, the cruise time can be calculated by cruise speed. One method is to use the free-flow speed recommended by the 1994 Highway Capacity Manual (HCM) [101]. Cheu et al. [25] consider the link travel time as the sum of the cruising time and delay time. They introduce a model to estimate average link travel time of signalized arterial links using data obtained from SCATS detectors. Here the link travel time is the average travel time of a vehicle after it has left the upstream stop-line detector, until the time it leaves the downstream stop-line detector. Another work by Cheu et al. [24] states that the arterial link cruise travel time is derived by dividing the segment link length by mean speed which is obtained at the detector station. ATTM [71] is developed to estimate travel time on arterial road and has been implemented on servers at VicRoads, Australia. The ATTM model relies on loop detector data (capacity, arrival flow), link geometry (link length and free flow speed) and signal timing parameters (signal cycle time and effective green time) as input. Therefore, no parameters are needed for calibration. The loop detector data and traffic signal timing parameters are readily available in a traffic control center (SCATS system server).

2.5 Travel Time Reliability

Travel time reliability is of great importance. In practice, travellers are not only concerned about the average amount of time they spend on routes towards their destinations, but also the likelihood that they can reach the destinations within an acceptable window of time of the expected arrival time. A too early or too late arrival time can create a negative travel experience and many sometimes have serious consequences. Therefore, travel time reliability is an important aspect to be considered in route guidance. For example, drivers who choose the route with the minimum
expected travel time might risk some unexpected delays caused by a significant variation in link travel times on that route. Measures of travel time reliability are used to quantify both the variability in travel times across different days and months and the variability in different times of day [66, 72]. Lyman and Bertini [73] argue for the use of travel time reliability indices as an important congestion measure in addition to existing measures of network performance. Different travel time reliability measures are analyzed, based on freeway data, for improving real-time transportation management. Chen et al. [21] state that travel time reliability is an important measure of service quality for travellers. Chen et al. [23] consider travel time reliability as an important criterion in route planning given that faster and more reliable paths are key requirements for drivers. Recker et al. [104] investigate the function of travel time reliability and variability in different risk-taking behaviour. Computation procedures are developed to calculate freeway travel time variability for both section level and route level. Shao et al. [109] find that travellers consider both average travel time and the reliability of a given route when making decisions. Travellers’ path choice behaviour under uncertainty is characterized with travel time reliability based user equilibrium principle. Gao [45] analyzes three travel time reliability measures such as travel time variance, expected early schedule delay and expected late schedule delay. A linear combination of expected travel time and travel time variance algorithms are designed to make sure the selected routes, based on perfect online information, have a higher reliability. Nie et al. [88] demonstrate the benefits and applicability of reliable route guidance using a case study in Chicago. The most reliable route can be found by maximizing the probability of arriving on time.

The main purpose of this section is to review the related work of travel time reliability. We review the main definitions of travel time reliability, followed by the literature review of relevant research work. A large variety of definitions of path travel time reliability (PTTR) can be found in the literature. This section introduces nine commonly used definitions.
Chapter 2

The first definition [33, 104] is defined as the probability of a given path travel time being less than a predefined threshold value, i.e., the probability of arriving on time. The threshold value is set to be the expected travel time plus a certain acceptable additional time [91, 104]. From a mathematical point of view, this definition can be described as the following equation:

\[
PTTR_1 = Pr\{\text{Path Travel Time} \leq \text{Threshold}\}
\]  \hspace{1cm} (2.2)

The Florida reliability definition considers the percentage of trips that are completed within a designated travel time threshold as the expected travel time plus a certain acceptable additional time (PTTR\(_1\)) [33, 91]. Travel times longer than the expected time will be deemed as unreliable. Fan et al. [35] consider the maximization of the travel time reliability for on-time arrival (PTTR\(_1\)). This identifies the next node to visit with maximum probability of arriving at the destination by a certain time. An adaptive optimal path algorithm is proposed based on the Bellman principle of optimality and the Picard method of successive approximation. Fan and Nie [34] formulate the PTTR\(_1\)-based travel time reliability problem as a stochastic on time arrival problem. Successive approximation is introduced to study the convergence properties of the algorithm. However, it is a question of whether this algorithm converges within a finite number of steps. Furthermore, they propose an approximate discrete speedy algorithm to solve the problem in a finite number of steps. This thus improves the computational efficiency significantly [85]. Seshadri and Srinivasan [108] propose another efficient algorithm to address the PTTR\(_1\)-based travel time reliability problem with normally distributed and correlated link travel times.

The second definition of PTTR is given on the basis of the 95th percentile of travel times. More precisely,

\[
PTTR_2 = \frac{95\text{th Percentile Travel Time for a Path}}{\text{Free Flow Travel Time}}
\]  \hspace{1cm} (2.3)
As shown, $PTTR_2$ is determined by the ratio of 95th percentile travel time to the free flow travel time. It evaluates how much total travel time travellers need to ensure on-time arrival. This definition is proposed by the Federal Highway Administration [37]. In Lyman and Bertini [73], the planning time index as $PTTR_2$ is tested in Portland for improving real-time transportation management.

The third definition of PTTR is based on buffer time index [37,70]. Mathematical presentation regarding buffer time and buffer time index ($PTTR_3$) can be expressed as follows:

$$PTTR_3 = \frac{\text{Buffer Time}}{\text{Average Travel Time}} \times 100\%$$

(2.4)

$Buffer Time = 95th \text{ Percentile Travel Time for a Trip} – \text{Average Travel Time}$ (2.5)

As shown in Equation 2.5, buffer time is defined as the extra time that travellers must add to their average travel time when planning trips to ensure on-time arrival. For example, a trip mean travel time is 10 minutes, a traveller should budget an extra 4 minutes to ensure on-time arrival 95% of the time in terms of 40% $PTTR_3$. The smaller the $PTTR_3$, the more reliable are selected routes.

The fourth and fifth definitions of PTTR are proposed by Van Lint and Van Zuylen [66] as skewness and width. These reliability definitions utilize the percentile of travel times to indicate reliability instead of travel time mean and variance values. Generally, the larger the travel time skewness, the higher the probability that extreme travel time (compared to mean travel time) will occur.

$$PTTR_4 = \frac{P90-P50}{P50-P10}$$

(2.6)
\[ PTTR_5 = \frac{P90-P10}{P50} \] (2.7)

where P90, P50 and P10 denote the 90th, 50th and 10th percentile travel times.

The sixth and seventh definitions are given with a mean and variance of travel times in Emmerik et al. [8] and Lomax et al. [70].

\[ PTTR_6 = \frac{\text{Standard Deviation}}{\text{Average Travel Time}} \] (2.8)

\[ PTTR_7 = (\text{Mean Time} - \text{Standard Deviation}, \text{Mean Time} + \text{Standard Deviation}) \] (2.9)

In \( PTTR_6 \), named as a coefficient of variation [8], average and standard deviation travel time values are combined together in a ratio to produce a value as a travel time reliability indicator. Lomax et al. [70] consider the travel time window as the measure of travel time reliability according to \( PTTR_7 \).

The last two commonly used definitions of PTTR, earliness and lateness, are developed by Kaparias et al. [62].

\[ PTTR_8 = \frac{\text{Minimum Travel Time}}{\text{Average Travel Time}} \] (2.10)

\[ PTTR_9 = \frac{\text{Average Travel Time}}{\text{Maximum Travel Time}} \] (2.11)

In (2.10) and (2.11), minimum travel time and maximum travel time represent the shortest and longest travel times that may be experienced on a link or route. These
two reliability indices correspond to the extreme values of the travel time distribution. A newly developed software tool, the Imperial College Navigation Software (ICNavS), is presented in Kaparias et al. [61] to implement and demonstrate the reliable route guidance framework based on $PTTR_8$ and $PTTR_9$ on a real road network.

PTTR along with these nine definitions introduced above, addresses the travel time reliability of a single path. Besides path travel time reliability, the reliability of OD pairs is also of interest in practice.

**OD travel time reliability** [104] is formulated as the probability that the weighted average travel time of a given OD pair (i.e., OD travel time) is within an acceptable domain (Figure 2.2).

\[
OD \text{ Travel Time Reliability} = Pr\{OD \text{ Travel Time} \leq \text{Threshold}\} \quad (2.12)
\]
\[
OD \ Travel\ Time = \sum_{i=1}^{N} a_i \ast \text{Travel\ Time\ of\ Path\ } i \tag{2.13}
\]

\[
\sum_{i=1}^{N} a_i = 1 \tag{2.14}
\]

The path and OD reliability suit different groups of people. Travellers are more concerned about path reliability, while OD reliability is preferred for use by traffic operators as it addresses the performance of a whole transport network.

This thesis focuses on route guidance systems that determine optimal routes for travellers towards their destinations. Therefore, the path travel time reliability is of major interest and thus only path travel time reliability is discussed in the reminder of the thesis. In particular, as the main criteria to evaluate travel time reliability, \( PTTR_6 \) is employed in this thesis. There is more research work focusing on modeling travel time reliability and reliable route guidance. Comprehensive reviews of this topic can be seen in Bates et al. [10] and Noland and Polak [90].
Chapter 3

Shortest Paths Finding in Stochastic Time-dependent Networks Considering Link Travel Time Correlation

In this chapter we develop a simple, robust framework to address the problem of finding the route with the least expected travel time from any node to any destination in a stochastic time-dependent network. A dynamic programming based approach is presented to study the shortest path problem, in which we consider both spatial and temporal link travel time correlations. In particular, the spatial correlation is represented by a Markovian property of the link states where each link is assumed to experience multiple possible 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.

3.1 Introduction

Daily traffic congestion has become a major societal and economic problem over the past decades in and around most, if not all, major cities in the world. The estimated societal costs of congestion add up to billions of dollars. These costs include not only those caused by queuing delay, but also costs due to travel time uncertainty.

Amongst the different options to relieve traffic congestion (e.g., building new infrastructure), managing traffic by means of traffic information, guidance and control turns out to be the most cost-effective one. It can strongly improve the utilization of available road infrastructure, increase the reliability and robustness of transportation networks and contribute to liveability in our densely populated metropolitan areas.

In-vehicle navigation systems are designed to support drivers by finding the fastest path to their destination and by providing alternative route choices based on information about current traffic conditions. In this context, finding the shortest path given real-time information in a time-dependent stochastic network is a problem of both theoretical and practical significance. It is problematic because in such networks, 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 of traffic demands. Moreover, travel times on different links that are strongly correlated in space as a road incident in one link can cause congestion in other links. It is also well known that traffic demands change 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).

Different variants of the shortest paths finding problem have been studied in the
literature. For instance, Gao and Chabini [46] study the least travel time path finding problems in a general stochastic time-dependent traffic network. One of the main drawbacks of this approach, however, is that it assumes knowledge of the full joint distribution of travel times across the entire network. It is difficult to obtain such travel time data in reality. Fan et al. [36] propose a multi-stage adaptive feedback control process to address least expected time path problems with correlated link costs. However, their method ignores the fact that link costs vary over time and fails to detect the potential changes. For instance, as time varies, the state of different segments of the network can change accordingly. While Miller-Hooks and Mahmassani [81] determine an optimal least-expected travel time path with respect to the arrival time at a node given time-varying discrete travel time, it assumes independent link travel times. Nie and Wu [87] study a priori shortest path problem to guarantee a given likelihood of arriving on-time in a stochastic time-dependent network. However, their proposed model targets a priori path generation instead of adaptive routing. It will increase the complexity of the problems when applying adaptive routing. Therefore, in this thesis, we will develop a general path finding model in stochastic time-dependent networks that requires much smaller set of parameters (e.g., the conditional average link travel time (or the average flow speed) in a given link state and time interval, and the conditional transition probabilities among states of two consecutive links in the network) than Gao and Chabini’s [46] method, without significantly compromising on the accuracy of the results.

In addition, few studies in the literature have simultaneously considered both the temporal and spatial correlations when searching for the shortest path. Waller and Ziliaskopoulos [118] are concerned with the shortest path problem that accounts for limited forms of spatial and temporal dependency. For spatial dependency, each link is dependent on one predecessor link with the conditional probability matrices. The temporal dependency assumes that link cost is known when its tail node is reached. However, this model is applied to time-invariant traffic network that may not produce a good result in the real situation. The same definitions of limited temporal and
spatial correlations are considered in Boyles and Waller [16]. In Nie and Wu [86], the spatial correlation is modelled by assuming the probability density functions of link traversal times to be conditional on the state of travellers arriving at the tail node of the link, and the temporal dependency is defined as the traversal time distribution being conditional on arrival time. Nevertheless, it is not scalable for a network with many links or suitable for online calculation because of its complexity. In this thesis, we will develop a comprehensive yet simple least expected travel time path finding framework with a combination of temporal and spatial correlations. The temporal correlation is calculated based on the flow speed model proposed in [114] for time-dependent link travel time. The spatial correlation is assumed to follow a Markovian property in which the probability that a link experiences multiple possible conditions (or states) only depends on the condition of its immediately preceding link. The framework deals with congested or uncongested conditions of link congestion for clarity and can be easily extended to more than two possible levels with the increase of computational burden. Compared with the requirement of probability density functions of link travel times in Nie and Wu [86], our proposed method just utilises a small set of parameters as described in the later section.

This chapter is structured as follows: First, Section 3.2 presents our proposed minimum expected travel time routing framework. Some experimental results are discussed in Section 3.3 through a simple network example. Further testing is reported with regards to an extensive network in Section 3.4. Finally, Section 3.5 gives some summaries about this chapter.

3.2 Methodological Framework

In this section, we first formulate the problem and then present an on-line approach for least expected travel time route selection based on the time of arrival at the decision node and the state (or congestion level) of the incoming link.
3.2.1 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. The number of links is denoted as $|A|$. Let $[\tau_h, \tau_{h+1})$, $h = 0, 1, \ldots, |\Gamma| - 1$ be the time intervals in $\Gamma$ separated by the time points $\tau_0 < \tau_1 < \ldots < \tau_{|\Gamma|}$ 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 traveller arrives at the decision node.

In our least expected travel time path finding framework, compared with the existing methods in the literature requiring either conditional probability density function of link travel time or variance-covariance matrices, 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 multiple possible conditions. It is the same as Fan et al. [36]'s method and seems to be the simplest method to represent spatial correlation requiring only the conditional transition probabilities between states of two consecutive links. Although the framework can be easily extended to more than two possible levels of link congestion, as briefly shown at the end of the section on route choice heuristics, for clarity only two states (congested and uncongested) will be considered in the numerical examples in this thesis. Similar to the approach in Fan et al. [36], 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. Note that $\lambda_{kij}$ captures the conditional probability that the next link is becoming uncongested.
even though the previous link is congested.

3.2.2 Route choice heuristics

Our proposed approach is motivated by Bellman’s principle of optimality, introduced in [11]. Formally, assume that the traveller arrives at node \( i \) on an incoming link \((k, i)\) at time \( t \in [\tau_h, \tau_{h+1})\) for some \( h \in \{0, 1, 2, \ldots, \Gamma - 1\} \). Beyond time period \( \Gamma - 1 \), travel times are static and deterministic. Therefore, it can be solved using a shortest path algorithm (e.g. Dijkstra’s) on the deterministic and static network. Our proposed framework is a one-to-one type, which calculates the shortest path from a single source node to a single destination node in the traffic network. Knowing the condition of link \((k, i)\) and the start time at node \( i \), the traveller 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 lengths 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 least expected travel time (LETT) route selection problem can be formulated as solving the following recursion for \( u_{ij}(\cdot) \) and \( c_{ij}(\cdot) \) below.

\[
    u_{ki}(t) = \min_{j \neq i} \{ \alpha_{kij}(U_{ij}(t) + u_{ij}(t + U_{ij}(t))) + (1 - \alpha_{kij})(C_{ij}(t) + c_{ij}(t + C_{ij}(t))) \}, \quad (3.1)
\]
Chapter 3

\[ c_{ki}(t) = \min_{j \neq i} \{ \lambda_{kij}(U_{ij}(t) + u_{ij}(t + U_{ij}(t))) + (1 - \lambda_{kij})(C_{ij}(t) + c_{ij}(t + C_{ij}(t))) \} \], \quad (3.2) 

\[ u_{kD}(t) = 0, c_{kD}(t) = 0, \quad (3.3) \]

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

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 the state-space cell transmission model proposed in Tampere and Immers [115]. 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 [36]. 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. Among various estimation methods, we adopt the method proposed in [36]. We use the mean travel time value as the threshold to classify the various link states. If the link travel time is greater than
the mean travel time, then the link is classified as congested. If it is less than the mean value, it is uncongested. 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 [60], or indirectly using the standard techniques such as those listed in [67, 74]. Note that due to the time-varying nature of the network, $\alpha_{kij}$ and $\lambda_{kij}$ values are time-dependent and can vary 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 differ for the same link in different time intervals. This results in similar $\alpha_{kij}$ and $\lambda_{kij}$ values across many time zones. This is because the travel time threshold, at which the link is considered to be congested, is likely to be higher in a peak period compared 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. The statement of LETT route selection framework is shown in Figure 3.1.

Furthermore, it is recognized that travel times on a link are temporally correlated with its travel times of previous time periods. It is worth noting that the temporal correlation in Waller and Ziliaskopoulos [118] refers to recourse, i.e., travel time can be learned before traversing a link. However, it only applies to time-independent traffic networks. In Nie and Wu [86], the temporal correlation defined as the traversal time distribution is conditional on time. This causes further complexity due to the requirement of travel time probability density function. In this thesis, however, the $U_{ij}(t)$ and $C_{ij}(t)$ values in the above equations are time-dependent and calculated based on the flow speed model proposed in [114] as follows:

$$ U_{ij}(t) = \frac{l_{ij}}{v_{h(t,i,j)}}, \text{ if } \frac{l_{ij}}{v_{h(t,i,j)}} < \tau_{h+1} - t $$  \hspace{1cm} (3.4)
LETT Route Selection Framework

Step 0: (Initialization)

0.1 compute $u_{ki}(t)$ and $c_{ki}(t)$, $\forall i \in N$, $\forall t \in [\tau|\Gamma|_{-1}, \tau|\Gamma|)$

0.2 $u_{ki}(t) = \infty$, $c_{ki}(t) = \infty$, $\forall i \in N - \{D\}$, $\forall t < \tau|\Gamma|_{-1}$

Step 1: (Main Loop)

for $\forall t \in [\tau|\Gamma|_{-2}, \tau|\Gamma|_{-1})$ to $\forall t \in [\tau_0, \tau_1)$

for $(i, j) \in A$

$$\text{temp} = \min_{j \neq i} \{a_{kij}(U_{ij}(t) + u_{ij}(t + U_{ij}(t))) + (1 - a_{kij})(C_{ij}(t) + c_{ij}(t + C_{ij}(t)))\}$$

or $$\text{temp} = \min_{j \neq i} \{\lambda_{kij}(U_{ij}(t) + u_{ij}(t + U_{ij}(t))) + (1 - \lambda_{kij})(C_{ij}(t) + c_{ij}(t + C_{ij}(t)))\}$$

if $\text{temp} < u_{ki}(t)$ or $\text{temp} < c_{ki}(t)$

$$u_{ki}(t) = \text{temp} \text{ or } c_{ki}(t) = \text{temp}$$

$$\text{arg } u_{ki}(t) = j \text{ or arg } c_{ki}(t) = j$$

Figure 3.1: LETT route selection framework

else

$$U_{ij}(t) = \tau_{h+1} + \frac{l_{ij} - l^0_{ij}}{v^u_{h+1(i,j)}} - t, \text{ if } \frac{l_{ij} - l^0_{ij}}{v^u_{h+1(i,j)}} < \tau_{h+2} - \tau_{h+1}, l^0_{ij} = v^u_{h(i,j)}(\tau_{h+1} - t) \quad (3.5)$$

else

$$U_{ij}(t) = \tau_{h+2} + \frac{l_{ij} - l^1_{ij}}{v^u_{h+2(i,j)}} - t, \text{ if } \frac{l_{ij} - l^1_{ij}}{v^u_{h+2(i,j)}} < \tau_{h+3} - \tau_{h+2},$$

$$l^1_{ij} = l^0_{ij} + v^u_{h+1(i,j)}(\tau_{h+2} - \tau_{h+1}) \quad (3.6)$$
\[ U_{ij}(t) = \tau_{[\Gamma|-1} + \frac{l_{ij} - l_{ij}^{[\Gamma|-h-2}}}{v_{[\Gamma|-1(i,j)}} - t, \quad \text{if} \quad \frac{l_{ij} - l_{ij}^{[\Gamma|-h-2}}}{v_{[\Gamma|-1(i,j)}} < \tau_{[\Gamma]-[\Gamma|-1}, \]

\[ l_{ij}^{[\Gamma|-h-2} = l_{ij}^{[\Gamma|-h-3} + v_{[\Gamma|-2(i,j)}}(\tau_{[\Gamma]-1} - \tau_{[\Gamma]-2}, \quad (3.7) \]

where \( v_{h(i,j)}^u \) is the average flow-speed on link \((i, j)\) of length \(l_{ij}\) in the time interval \([\tau_h, \tau_{h+1})\) conditional on it being uncongested. Similarly, \( C_{ij}(t) \) can be obtained based on the corresponding \( v_{h(i,j)}^c \) average flow-speed on link \((i, j)\) in the time interval \([\tau_h, \tau_{h+1})\) conditional on it being 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 Section 3.2.1 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:

\[ u_{ki}^s(t) = \min_{j \neq i} \{ \sum_{r=1}^{M} p_{ki,j}^{sr}(U_{ij}^r(t) + u_{ij}^r(t + U_{ij}^r(t))) \}, s = 1, 2, \ldots, M \quad \text{(3.8)} \]

\[ u_{kD}^s(t) = 0, \quad \text{(3.9)} \]

where:

- \( u_{ki}^s(t) \) – 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+\gamma} + \frac{l_{ij} - l_{ij}^{[\tau_{h+\gamma}(-1)}}}{v_{h+\gamma(i,j)}} - t, \quad \text{if} \quad \frac{l_{ij} - l_{ij}^{[\tau_{h+\gamma}(-1)}}}{v_{h+\gamma(i,j)}} < \tau_{h+\gamma+1} - \tau_{h+\gamma}, \]

\[ l_{ij}^{[\tau_{h+\gamma}(-1)}} = l_{ij}^{[\tau_{h+\gamma}(-2)}} + v_{h+\gamma-1(i,j)}(\tau_{h+\gamma} - \tau_{h+\gamma-1}), \quad \text{where} \quad \gamma \geq 1 \quad \text{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}^{[\tau_{h-1}]} = 0; \]
Chapter 3

- $p_{kij}^{sr}$ – the probability that link $(i, j)$ is in state $r$ if the incoming link $(k, i)$ is in state $s$, $p_{kij}^{sr}(t) = \int_{t(h(i,j))^{r-1}}^{t(h(i,j))^r} f_{ij}(t, \xi)d\xi \approx p_{kij}^{sr}$, $\sum_{r=1}^{M} p_{kij}^{sr} = 1$,

where $t(h(i,j))^{r}$, $r = 1, 2, \ldots, M$ is the travel time threshold for link $(i, j)$ to be in state $r$ and let $t(h(i,j))^0 = 0$; $f_{ij}(t, \xi)d\xi$ 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 $s$.

Similar to the previous case of two link states, $U_{ij}^r(t)$ is calculated based on the flow speed model as in Equs (3.4)-(3.7) for a given link state $r$ possibly crossing multiple time intervals, and $p_{kij}^{sr}$ is approximated as a time independent value across different time intervals given the threshold $t(h(i,j))^r$. Note that in this thesis the time-dependency (or temporal correlation) is mainly taken into account by the former ($U_{ij}^r(t)$) while the latter simplification (time-invariant $p_{kij}^{sr}$) has only secondary order impact.

It should be emphasized that only a smaller set of input parameters (e.g., the conditional average link travel time in a given link state and time interval, and the conditional transition probabilities between states (congested versus uncongested) of two consecutive links in the network) are utilised in our proposed algorithm than in Gao and Chabini’s [46] method which requires knowledge of the full joint distribution of travel times across the entire network. As such, the input data size reduces from $|A||\Gamma|L$ in Gao and Chabini’s [46] method to $M(|A||\Gamma| + M)$ in our proposed approach, where $L$ is the maximum number of support points for a single link travel time discrete distribution, $|A|$ is the number of links, $|\Gamma|$ is the number of time periods and $M$ is number of possible link states.

3.3 An Illustrative Example

In this section, we demonstrate the execution of our proposed approach through a simple illustrative example. We will also compare the result of our approach with other three existing methods: i) a optimal routing policy framework introduced by Gao
and Chabini [46]; ii) a least expected time path selection method proposed by Miller-Hooks and Mahmassani [81]; and iii) a dynamic programming approach proposed by Fan et al. [36]. 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 does not rely on support points, i.e., all the parameters required can be calculated as described in Section 3.3.1 based on a set of historical data.

### 3.3.1 The simple network

We consider a small traffic network depicted in Figure 3.2. There are nine links ($|A| = 9$) and seven nodes ($|N| = 7$). 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.
Table 3.1: Joint travel time realisation of all links

<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>

For the sake of simplicity, assume there are only three possible support points $W = \{w_1, w_2, \ldots, w_L\}$ (see Table 3.1), where $w_i, i \in 1, 2, \ldots, L$ 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 $p_i$, we have $\sum_{i=1}^{L} p_i = 1$. 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 ($\tau_1 - \tau_0 = \tau_2 - \tau_1 = 3$). Travel time data beyond time 1 are assumed to be 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).
3.3.2 Parameters calculation

Based on the assumption of the availability of a set of support points representing all link travel times in the network, in this simple illustrative example, 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 incoming link $(k, i)$ in time $t$. Denote the travel time on $(k, i)$ of a particular support point $w_n \in W$ by $x_{ki}$. The following sets are defined:

\[ C = \{ w_n | P(X_{ki}(t) = x_{ki} \leq \gamma_{ki}(t)) \}, \quad (3.10) \]

and

\[ \overline{C} = \{ w_n | P(X_{ki}(t) = x_{ki} > \gamma_{ki}(t)) \}. \quad (3.11) \]

\[ C' = \{ w_n | P(X_{ki}(t) = x_{ki} \leq \gamma_{ki}(t)) \cap P(X_{ij}(t) = x_{ij} \leq \gamma_{ij}(t)) \} \subseteq C; \quad (3.12) \]

and

\[ C^* = \{ w_n | P(X_{ki}(t) = x_{ki} > \gamma_{ki}(t)) \cap P(X_{ij}(t) = x_{ij} \leq \gamma_{ij}(t)) \} \subseteq \overline{C}. \quad (3.13) \]

By choosing appropriate thresholds $\gamma_{ki}(t)$ and $\gamma_{ij}(t)$ in each time interval, we can have
\[ \alpha_{kij} = \frac{\sum_{w_n \in C^r} p_n}{\sum_{w_n \in C} p_n}, \quad \text{and} \quad \lambda_{kij} = \frac{\sum_{w_n \in C^r} p_n}{\sum_{w_n \in C} p_n} \] (3.14)

which is time invariant.

The conditional average flow speeds are calculated as

\[ v^u_{h(k,i)} = \frac{l_{ki} \cdot \sum_{w_n \in C} p_n}{\sum_{w_n \in C} p_n x_{ki}}, \quad \text{and} \quad v^c_{h(k,i)} = \frac{l_{ki} \cdot \sum_{w_n \in C} p_n}{\sum_{w_n \in C} p_n x_{ki}} \] (3.15)

### 3.3.3 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 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 \((i,1)\), link \((i,2)\), link \((1,D)\) and link \((2,D)\). This means the travel time that is no more than mean value is uncongested, and congested otherwise. Consequently:

\[ \gamma_{i1}(t) = \sum_{n=1}^{3} p_n x_{i1} = 0.5 \times 4 + 0.3 \times 4 + 0.2 \times 5 = 4.2, \quad t \in [\tau_0, \tau_1) \]
\[ \gamma_{i2}(t) = 5.2, \quad \gamma_{1D}(t) = 4.8, \quad \gamma_{2D}(t) = 4.2, \quad t \in [\tau_0, \tau_1) \]

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

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

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

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

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

\[ v^c_{0(i,2)} = l_{i2} \times \left( \sum_{w_n \in C} p_n x_{i2} / \sum_{w_n \in C} p_n \right)^{-1} = (5 \times (0.5 + 0.3)/(0.5 + 0.3))^{-1} = \frac{1}{5} \]
is:
\[
v^c_{0(i,1)} = l_{i1} \times \left( \sum_{w_n \in C} p_n x_{i1} / \sum_{w_n \in C} p_n \right)^{-1} = (5 \times 0.2/0.2)^{-1} = \frac{1}{5}
\]
\[
v^c_{0(i,2)} = l_{i2} \times \left( \sum_{w_n \in C} p_n x_{i2} / \sum_{w_n \in C} p_n \right)^{-1} = (6 \times 0.2/0.2)^{-1} = \frac{1}{6}
\]

Accordingly, for the second time interval:
\[
\gamma_{i1}(t) = 15.2, \gamma_{i2}(t) = 14.6, \gamma_{1D}(t) = 15.4, \gamma_{2D}(t) = 15, t \in [\tau_1, \tau_2]
\]

In the meantime, the average flow speeds under the uncongested situation for link \((i, 1)\), link \((i, 2)\), link \((1, D)\) and link \((2, D)\) is:
\[
v^u_{1(i,1)} = \frac{1}{15}, v^u_{1(i,2)} = \frac{1}{14}, v^u_{1(1,D)} = \frac{1}{15}, v^u_{1(2,D)} = \frac{1}{14}
\]

Accordingly, the average flow speeds under the congested situation for link \((i, 1)\) and link \((i, 2)\) is:
\[
v^c_{1(i,1)} = \frac{1}{16}, v^c_{1(i,2)} = \frac{1}{16}
\]

Similarly, under the congested situation,
\[
v^c_{1(1,D)} = \frac{1}{17}, v^c_{1(2,D)} = \frac{1}{16}
\]

In the uncongested situation, the average expected travel time on link \((i, 1)\) and link \((i, 2)\) can be calculated as follows:
\[
U_{i1}(t) = \tau_1 + (l_{i1} - v^u_{0(i,1)}(\tau_1 - \tau_0)) / v^u_{1(i,1)} - \tau_0 = 3 + (1 - \frac{1}{2} \times 3) / \frac{1}{15} - 0 = 6.75, t \in [\tau_0, \tau_1]
\]
\[
U_{i2}(t) = \tau_1 + (l_{i2} - v^u_{0(i,2)}(\tau_1 - \tau_0)) / v^u_{1(i,2)} - \tau_0 = 3 + (1 - \frac{1}{5} \times 3) / \frac{1}{14} - 0 = 8.6, t \in [\tau_0, \tau_1]
\]

Similarly, the average expected travel time on link \((i, 1)\) and link \((i, 2)\) in the congested situation can be calculated as follows:
\[
C_{i1}(t) = \tau_1 + (l_{i1} - v^c_{0(i,1)}(\tau_1 - \tau_0)) / v^c_{1(i,1)} - \tau_0 = 3 + (1 - \frac{1}{5} \times 3) / \frac{1}{16} - 0 = 9.4, t \in [\tau_0, \tau_1]
\]
\[
C_{i2}(t) = \tau_1 + (l_{i2} - v^c_{0(i,2)}(\tau_1 - \tau_0)) / v^c_{1(i,2)} - \tau_0 = 3 + (1 - \frac{1}{6} \times 3) / \frac{1}{16} - 0 = 11, t \in [\tau_0, \tau_1]
\]
Then, by using the same method, at the next time interval, 

\[ U_{1D}(t) = 15, U_{2D}(t) = 14, C_{1D}(t) = 17, C_{2D}(t) = 16, t \in [\tau_1, \tau_2] \]

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:

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

\[ u_{i2}(t) = \alpha_{i2D} \times U_{2D}(t) + (1 - \alpha_{i2D}) \times C_{2D}(t) = 14.6, t \in [\tau_1, \tau_2] \]

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

\[ c_{i2}(t) = \lambda_{i2D} \times U_{2D}(t) + (1 - \lambda_{i2D}) \times C_{2D}(t) = 16, t \in [\tau_1, \tau_2] \]

Finally, the minimum expected travel time conditional on the previous link uncongested in node \(i\) is decided by 

\[ u_{ki}(t = 0) = \min_{j=1,2} \{\alpha_{kij}(U_{ij}(t) + u_{ij}(t + U_{ij}(t))) + (1 - \alpha_{kij})(C_{ij}(t) + c_{ij}(t + C_{ij}(t)))\} = 22.68, \text{ and the next node to take is node 1.} \]

As we assume that \(w_2\) is the actual travel time that the driver will experience after traveling on link \((i, 1)\) for 4 minutes, we can decide on what next node to take to destination \(D\) via a similar process of calculation. In 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 the shortest path finding framework proposed by Gao and Chabini [46], Miller-Hooks and Mahmassani’s [81] least expected time path method and the approach proposed by Fan et al. [36] to compute the shortest paths in stochastic networks with travel time correlation between
adjacent links. In Gao and Chabini’s framework [46], perfect online information is required.* That is, the travel time realization for all links in the network at the current time is assumed to be known to the user. Given the comprehensive full joint travel time data, Gao and Chabini’s method is supposed to perform the best among all these methods. This can be represented as a benchmark when comparing other methods. Miller-Hooks and Mahmassani’s [81] method requires the knowledge of time-dependent link travel time distributions. Nevertheless, Miller-Hooks and Mahmassani’s [81] method assumes there is no spatial correlation between adjacent links. We can show the performance of incorporating both temporal and spatial correlations in our proposed method by comparing it to Miller-Hooks and Mahmassani’s [81] method.

On the other hand, Fan et al.’s approach in [36] 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 travels to reach the decision node is known. Fan et al.’s [36] approach uses the same Markovian assumption to represent spatial correlation as our framework though without the time dependency consideration. Therefore, in their approach, differences in link travel times in different time periods are not taken into consideration. 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. Here, our proposed framework is not compared with Nie and Wu’s [86] method as introduced before, because the travel time probability density function is unavailable. In this example, to generate the expected link travel times used in Fan et al.’s approach from the support points data, we compute the average of the link travel times at all time intervals under each support point. To simplify the presentation, we also assume that the link travelled by the user to arrive at node \( i \) is uncongested.

*Note that in this research, we compare our method with the exact algorithm DOT-SPI for the perfect online information variant presented in [46]. However, [46] also present a number of approximation methods for the cases of no-online-information and partial-online-information settings.
Table 3.2: Comparison among different route selection methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Route</th>
<th>Actual time ($w_2$)</th>
<th>Percentage difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our method</td>
<td>Node i-1-D</td>
<td>19</td>
<td>0</td>
</tr>
<tr>
<td>Gao and Chabini’s [46] method</td>
<td>Node i-1-D</td>
<td>19</td>
<td>0</td>
</tr>
<tr>
<td>Miller’s [81] method</td>
<td>Node i-1-D</td>
<td>19</td>
<td>0</td>
</tr>
<tr>
<td>Fan et al.’s [36] method</td>
<td>Node i-2-D</td>
<td>21</td>
<td>0.11</td>
</tr>
</tbody>
</table>

The overall comparison is depicted in Table 3.2. 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 $i - 1 - D$. From Table 3.2 we can see that, based on the assumption that perfect online information is available, Gao and Chabini’s and Miller-Hooks and Mahmassani’s methods choose 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 method fails to choose the shortest path. Instead, their method selects route $(i-2-D)$ resulting in a 11% difference from the minimum travel time.

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. Miller-Hooks and Mahmassani’s method works well when time-dependent link travel time distributions are available. On the other hand, our proposed method only requires knowledge of the state of the link traversed.
by the user to arrive at the decision node. This method 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. For instance, as time varies, the state of different segments of the network can change in varied ways.

3.4 A Numerical Evaluation

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 3.3). 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 consisting of the 4 innermost links, group B representing the links connecting the nodes in group A and group C and having 8 links. Specifically, 

\[ A = \{1-2, 2-3, 3-4, 4-8, 8-12, 12-16, 1-5, 5-9, 9-13, 13-14, 14-15, 15-16\}, \]
\[ B = \{2-6, 3-7, 5-6, 9-10, 7-8, 11-12, 10-14, 11-15\} \] and 
\[ C = \{6-7, 7-11, 6-10, 10-11\}. \]

3.4.1 Travel time data generation

Similar to the data generation process in Gao and Chanibi [46], a multivariate normal distribution is assumed for the joint distribution of all link travel time random variables. 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; 5) the common correlation coefficient.
Chapter 3

Figure 3.3: Larger network representation (16 nodes, 24 links)

of link travel time. The total number of rows in the support point table is the number of links times the number of time intervals. The same standard deviation of link travel time and correlation coefficient for each pair of links are used. Moreover, each link travel time value is rounded to the nearest integer. The probability of each support point is uniformly distributed. While 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 3.3. 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 fifteen-minute time intervals covering possible departing times in a typical weekday morning. A common standard deviation (1.44) of link travel time, and a unique correlation coefficient (0.5)
Figure 3.4: Average link travel times of different link groups at different departing (start) times

between that on consecutive links, are used to generate the data in this numerical example.

In this experiment, 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 are 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.
3.4.2 Experiment results

In this example, we compare our proposed method to Gao and Chabini’s [46], Miller-Hooks and Mahmassani’s [81] and Fan et al.’s [36] shortest path methods. Figure 3.3 demonstrates the mean values for different groups at all time intervals. We carry out three separate experiments for three different settings of travel times: the first ten time intervals (from time interval 1 to 10), the middle ten time intervals (from time interval 11 to 20), and the last ten time intervals (from time interval 21 to 30). In each experiment, we choose the start times to be inside the first three time intervals considered in our experiment to avoid the total travel time from going out of the considered time window. To evaluate the quality of the solutions produced by different methods, we compare the route choices computed by the four 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.

![Figure 3.5: Average difference in travel time using various methods (first 10 time intervals)](image)

The comparison of the travel times for the route choices produced by different methods (our proposed method, Gao and Chabini’s method, Miller-Hooks and
Chapter 3

Figure 3.6: Average difference in travel time using various methods (middle 10 time intervals)

Figure 3.7: Average difference in travel time using various methods (last 10 time intervals)
Mahmassani’s method and Fan et al.’s method respectively) against the minimal travel
time for the optimal route from node 1 to node 16 are shown in Figures 3.5-3.7 under
different starting times. The figures plot the mean differences (in log scale) for each
method over all support points. From these experimental results, it’s quite 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 methods introduced by Miller-Hooks
and Mahmassani [81] and Fan et al. [36]. In particular, since Fan et al.’s 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 in Figure 3.3) regardless of the start times. On the other hand, our method
adapts to the changing states of the network. During earlier, off-peak time intervals,
our method may choose a route going through the innermost links but during peak
hours (i.e., middle time intervals), our method tends to avoid those routes and selects
routes that mostly go through the outermost and connecting links.

To summarize, the results in these experiments are consistent with the expected
advantages of our proposed method. In most cases, our method shows smaller mean
differences than Miller-Hooks and Mahmassani’s method suggesting the advantage of
considering spatial correlation for adjacent links. Our method also shows smaller mean
difference values 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’s 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 the shortest path for
stochastic time-dependent networks by taking advantage of the information on link
time travel time correlations. However, we need to do more experiments in order to draw
more concrete conclusions. For this purpose, our proposed approach was field tested in the Melbourne traffic network with real data. The results are reported in a later chapter. Moreover, since the proposed algorithm runs in real-time, it is important to consider computation time issue in reality. However, the computation test requires a lot of time to setup and investigate, we leave it for future work.

3.5 Summary of this Chapter

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 through taking into account the information (i.e., link congestion level) available at the time of making the routing decision. A solution based on the principle of dynamic programming is derived where the optimal route is gradually formed following the real-time routing decisions at each node along that route. We showed, via numerical examples that the optimal route depends strongly on both spatial and temporal correlations of travel times. Furthermore, comparing with existing work assuming full knowledge of network travel time distribution, we showed that the proposed framework can provide good results while utilizing a very small set of network parameters.
Chapter 4

Reliability in Stochastic
Time-dependent Traffic Networks
with Correlated Link Travel Times

Optimal route selection with reliable expected travel time has been a focus of research in transportation networks where the reliability is subject to many uncertain factors such as traffic incidents or recurring traffic congestions. We develop two approximation methods to obtain the reliability of a route travel time in a stochastic time-dependent traffic network. The proposed method takes into consideration the probability nature of the travel time on individual links and their spatial correlation between adjacent links of a selected route. Reliability calculations and the accuracy of our approximation are discussed via a simple illustrative example and Sioux Falls network.
4.1 Introduction

Generally speaking, travellers have two concerns regarding their travel times. First, given an origin and destination (OD), they wish to determine a route with the least expected travel time (LETT) for that OD pair. Second, they are concerned about the reliability of their OD travel times, in hope to take some routes with reliable travel times [70]. The first concern has been addressed in the previous chapter. This chapter focuses on the latter. In particular, we consider reliable route guidance in stochastic time-dependent networks.

Unreliable travel times are usually caused by recurring congestion (bottlenecks and poor traffic signal timing) and nonrecurring congestion (traffic incidents, adverse weather conditions, special events and work zones) [3,19]. For instance, traffic incidents are matters that cause partial or complete reduction in road capacity. As to work zones, the traffic network can be influenced by construction or mending work on the roadway causing occasional or total road closure. Special events include sports activities in stadiums, Christmas day and unexpected fire. Travellers are normally unable to accurately predict travel times in congestion and may encounter unexpected congestion by entering routes with travel times much larger than the LETT suggested by the in-vehicle route guidance system. In some cases, a longer route with smaller travel time variation could be preferred to the route with smaller expected travel time but larger travel time variation [54,68]. People would like to choose a reliable route because arriving too early or too late causes serious consequences. Recker et al. [104] observe on the freeway system in Orange County, California that both travel time and travel time variability are higher during peak hours than normal times. According to these observations, they suggest that commuters prefer departing earlier to avoid the possible delays caused by travel time variability. Travellers are interested in both travel time saving and travel time variability reduction in order to minimize the risk [19]. Travel time reliability is related to quality of life, traffic safety and transport efficiencies. This motivates us to investigate the reliable route guidance problem in this chapter.
In a real traffic network, travel times are stochastic and time-dependent. Moreover, travel times of adjacent links are usually correlated. That is, if one link is congested, it is likely that after a short time period, upstream neighboring links also become congested. Therefore, travel time correlation should be taken into account in the reliable route selection.

Based on the existing work and widely adopted definitions, travel time reliability is essentially a notion of probability. Thus travel time distributions are indispensable for the evaluation of travel time reliability. As previously stated, travel time correlation must be considered in reliable route seeking. What is really important to the calculation of the travel time reliability is the joint distribution of travel times of all links within any route.

A wide range of definitions have been proposed to measure travel time reliability in the literature. For example, Van Lint and Van Zuylen [66] introduce two travel time reliability definitions (skewness and width) utilizing the 10th, 50th and 90th percentile of travel time distribution (e.g. 50th percentile travel time is equal to median travel time value). It shows how much delay there will be on the heaviest travel days. The wider the travel time distribution, the more unreliable the route. Buffer time is defined as the extra time that travellers must add to their average travel time when planning trips to ensure on-time arrival most of the time [70]. These measures are all related to the properties of the travel time distribution. On the other hand, the Florida reliability method [91] utilizes a threshold value to distinguish between reliable and unreliable travel times. The frequency of congestion is shown when travel time exceeds some expected threshold. Nie et al. [88] propose travel time reliability by maximizing the probability of arriving at the destination through any path departing from the start node with a time budget. The input of this model is time-dependent link travel time distribution obtaining through speed values from loop detectors. Lyman and Bertini [73] use coefficient of variation (the ratio of the standard deviation to the mean travel time) as the measure of travel time reliability. Standard deviation is one of
the classical statistical range measurements. This presents an estimate of the range of transportation conditions that might be experienced by travellers [70]. A further consideration in using standard deviation as a reliability indicator is when reliability is incorporated into a cost benefit assessment [14]. Kaparias [60] presents two travel time reliability indices to reflect the shortest and longest total travel time needed under a given confidence interval that may be experienced on the link or route using link speed data.

In this thesis, we are particularly interested in coefficient of variation (CV) [8], where average and standard deviation values are combined together in a ratio to produce a value as a travel time reliability indicator. CV provides a clearer picture of the trends and performance characteristics by taking into account both trip lengths and standard deviation [70].

\[
\text{Coefficient of Variation} = \frac{\text{Standard Deviation}}{\text{Mean Travel Time}}
\]

(4.1)

Then, the most reliable route is the one which has the smallest value of CV.

This CV measure requires two parameters (mean and variance) about path travel time. One common method is to utilize path travel time distribution directly. However, it is difficult to determine such joint distribution in reality. In practice, the joint travel time distribution can be substituted by individual link parameters. Here the calculation of mean path travel time is simple, as shown in a later section of this chapter based on either the law of total expectation [1] or the definition of the mean of a sum of random variables. However, the variance value is difficult to obtain as the existence of correlation between adjacent links. Rakha et al. [99] have proposed an estimation method to calculate path travel time variance value over all link travel time distributions that make up a path. In their model, the path’s CV value is obtained through the mean coefficient of variation over all individual links. However, the underlying assumption is that the travel times of different links are statistically
independent from each other [119]. In this chapter, we first propose the conditional expectation approximation method for reliability calculation. Path travel time mean and variance value are calculated using individual link travel time means and variances, and the covariance between each link pairs. However, this method works under the assumption that individual link travel time is only related to its own state, then the conditional expected travel time value of the product of two consecutive link travel times is approximated to be the product of the conditional expected travel time values of individual links. It could underestimate the calculated route travel time variance value in reality. Therefore, the second approximation method is presented to calculate the path travel time mean and variance value using individual link travel time means and variances, and variances of the sum of two adjacent link travel times along a route. In particular, variances of the sum of two adjacent link travel times are obtained through travel time distribution of the sum of two adjacent link travel times, which is built according to the new travel time value of the sum of two adjacent link travel times. Compared to the first approximation method, the second approximation method requires some extra effort to generate new travel time distribution. As such, both methods have pros and cons, and we want to compare the performance of the proposed approximation methods and select a better one recommended for use. This aims to obtain the reliability of a route travel time and find the most reliable route by minimizing coefficient of variation values in a stochastic time-dependent traffic network. The performance of the proposed approximation methods for the path travel time variance calculation will be shown through numerical examples with synthesized data.

The remainder of this chapter is organized as follows: In Section 4.2, we describe our network settings and present the detailed calculation of reliability using our proposed methods. In Section 4.3, we provide an illustrative example to demonstrate the process of calculation and compare the results of our approximation methods with the exact values calculated from the entire joint distribution of travel times, together with an experiment in Sioux Falls network. Finally, we conclude the proposed work in Section 4.4.
4.2 Reliable Route Guidance Problem Description

In this section, we first give a representation of a stochastic time-dependent traffic network. Then we show two new approximation methods of the route mean travel time and variance calculation in our network setting.

4.2.1 Network representation

Consider a stochastic traffic network $Z = (A, N, \Gamma)$, where $A$ is the set of links, $N$ is set of nodes and $\Gamma$ is the set of time intervals from 0 to $|\Gamma| - 1$ ($|\Gamma|$ is the total number of time intervals during the period of consideration, e.g. a day period). We denote the number of links in the network as $a$ ($a = |A|$). For each link $(i,j)$ ($i,j \in A, i,j \in N$), the link travel time on $(i,j)$ is a time-dependent random variable, denoted as $T_{ij}(t)$.

Similar to the previous chapter, the spatial correlation between link travel times is taken into account via Markovian property of the link states, where each link is assumed to experience either congested (written as $C$) or uncongested (written as $U$) conditions. In particular, for any consecutive link pair $(i,j), (j,k) \in A, \forall i,j,k \in N$:

- Let $\alpha_{ijk}$ be the conditional probability that if link $(i,j)$ is uncongested then link $(j,k)$ is also uncongested; and
- Let $\lambda_{ijk}$ be the conditional probability that if link $(i,j)$ is congested then link $(j,k)$ is uncongested.

4.2.2 Reliability calculation using the conditional expectation (CE) approximation

In the previous work, Rakha et al. [99] have proposed an estimation method to calculate path travel time variance value from individual links travel time distributions
that make up a path. In the following section, we present the first new approximation method to calculate the route mean travel time and variance in detail. In this framework, we believe there is significant spatial correlation between adjacent links, then we make the assumption that the travel time on two links that are not connected to each other are independent. We calculate the path travel time mean and variance value using individual link travel time means and variances, and the covariance between each link pairs. Given the time varying link travel time distributions and a start time \( t \), we obtain the following parameter for each link:

- **Mean travel time**
  - \( E[T^U_{ij}(t)] \): mean travel time of link \((i, j)\) at time \( t \) when link \((i, j)\) is not congested;
  - \( E[T^C_{ij}(t)] \): mean travel time of link \((i, j)\) at time \( t \) when link \((i, j)\) is in congested situation.

- **Variance of link travel time**
  - \( Var[T^U_{ij}(t)] \): travel time variance value at time \( t \) when link \((i, j)\) is not congested;
  - \( Var[T^C_{ij}(t)] \): travel time variance value at time \( t \) when link \((i, j)\) is in congested situation.

- **The corresponding probability value**
  - \( P^U_{ij}(t) \): the probability that link \((i, j)\) is not congested at time \( t \);
  - \( P^C_{ij}(t) \): the probability that link \((i, j)\) is congested at time \( t \).

We now investigate the method to approximate the mean travel time value of a route. For simplicity of explanation, we first present our proposed method on a simple route that consists of only two links, e.g., link \((i, j)\) and link \((j, k)\). We will later extend our method to a more general setting that consists of more than two links.
For a route of two adjacent link \((i,j)\) and link \((j,k)\)

\[
T_R(t) = T_{ij}(t) + T_{jk}(t')
\]

(4.2)

where \(T_{ij}(t)\) is the travel time on link \((i,j)\) at time \(t\); \(T_{jk}(t')\) is the travel time on link \((j,k)\) at time \(t'\); \(t\) is the starting time at node \(i\) and \(t'\) is the arrival time at node \(j\); \(T_R(t)\) is the route travel time at time \(t\). Here, we approximate the arrival time \(t'\) at node \(j\) as the current time \(t\) plus the expected travel time on link \((i,j)\) at time \(t\) as:

\[
t' = t + E[T_{ij}(t)].
\]

The first step is to calculate the route mean travel time.

\[
E[T_R(t)] = E[T_{ij}(t)] + E[T_{jk}(t')]
\]

(4.3)

According to the law of total expectation [1], the (unconditional) expectation of a random variable \(X\), is equal to expectation of the conditional expectation of \(X\).

\[
E[X] = E[E[X|Y]]
\]

where \(E[X|Y]\) is the conditional expectation of \(X\) given the condition \(Y\). At time \(t\), link \((i,j)\) has two possible states \(U\) (uncongested) and \(C\) (congested). We denote \(S\) as the possible link states. Therefore,

\[
E[T_{ij}(t)] = E[E[T_{ij}(t)|S]] = E[T_{ij}^U(t)] \cdot P_{ij}^U(t) + E[T_{ij}^C(t)] \cdot P_{ij}^C(t)
\]

(4.4)
Similarly, we can calculate $E[T_{jk}(t')]$ with the approximation that $t' = t + E[T_{ij}(t)]$. Then the route mean travel time can be calculated as $E[T_R(t)] = E[T_{ij}(t)] + E[T_{jk}(t + E[T_{ij}(t)])]$

Then, in the second step, we calculate the route travel time variance value. According to the definition of the variance of a sum of random variables, we have

$$Var[T_R(t)] = Var[T_{ij}(t) + T_{jk}(t')]$$
$$= Var[T_{ij}(t)] + Var[T_{jk}(t')]$$
$$+ 2Cov(T_{ij}(t), T_{jk}(t')) \quad (4.5)$$

where $Cov(T_{ij}(t), T_{jk}(t'))$ is the covariance value of each connected link pair. Here the link travel time correlation has been incorporated through the calculation of covariance value.

Following the law of total variance [2], for the variance of the travel time on link $(i, j)$ at time $t$, we have

$$Var[T_{ij}(t)] = E[Var[T_{ij}(t)|S]] + Var[E[T_{ij}(t)|S]] \quad (4.6)$$

We calculate $E[Var[T_{ij}(t)|S]]$ and $Var[E[T_{ij}(t)|S]]$ as follows:

$$E[Var[T_{ij}(t)|S]] = Var[T_{ij}^U(t)] \cdot P_{ij}^U(t)$$
$$+ Var[T_{ij}^C(t)] \cdot P_{ij}^C(t) \quad (4.7)$$
\[ \text{Var}[E[T_{ij}(t)|S]] = [E[T^U_{ij}(t)] - E[T_{ij}(t)]]^2 \cdot P^U_{ij}(t) + [E[T^C_{ij}(t)] - E[T_{ij}(t)]]^2 \cdot P^C_{ij}(t) \]

(4.8)

Similarly we can calculate \( \text{Var}[T_{jk}(t')] \). In order to calculate the variance of the entire route and take into account the correlation between adjacent links, we need to calculate the covariance of each connected link pair. As mentioned earlier, the travel times on two links that are not connected to each other are independent. We can observe that

\[ \text{Cov}(T_{ij}(t), T_{jk}(t')) = E[T_{ij}(t) \cdot T_{jk}(t')] - E[T_{ij}(t)] \cdot E[T_{jk}(t')] \]

(4.9)

The expected value of the product of the travel times on two adjacent link pairs can be calculated as follows:

\[ E[T_{ij}(t) \cdot T_{jk}(t')] = E[E[T_{ij}(t) \cdot T_{jk}(t')|S_{ij}(t), S_{jk}(t')]] \]
\[ = \sum_{S_{ij}(t)=U,C} \sum_{S_{jk}(t')=U,C} P(S_{ij}(t), S_{jk}(t')) \cdot E[T_{ij}(t) \cdot T_{jk}(t')|S_{ij}(t), S_{jk}(t')] \]

(4.10)

As \( \alpha_{ijk} \) is the conditional probability that if link \((i, j)\) is uncongested then link \((j, k)\) is also uncongested:

\[ \alpha_{ijk} = \frac{P(S_{ij}(t) = U, S_{jk}(t') = U)}{P(S_{ij}(t) = U)} \]
Similarly,

\[ \lambda_{ijk} = \frac{P(S_{ij}(t) = C, S_{jk}(t') = U)}{P(S_{ij}(t) = C)} \]

Consequently, we have the following conditional probabilities:

\[
P(S_{ij}(t) = U, S_{jk}(t') = U) = \alpha_{ijk} \cdot P(S_{ij}(t) = U)\]

\[
P(S_{ij}(t) = U, S_{jk}(t') = C) = (1 - \alpha_{ijk}) \cdot P(S_{ij}(t) = U)\]

\[
P(S_{ij}(t) = C, S_{jk}(t') = U) = \lambda_{ijk} \cdot P(S_{ij}(t) = C)\]

\[
P(S_{ij}(t) = C, S_{jk}(t') = C) = (1 - \lambda_{ijk}) \cdot P(S_{ij}(t) = C)\]

For a particular state of \(S_{ij}(t)\) of link \((i,j)\) and \(S_{jk}(t')\) of link \((j,k)\), we approximate the term \(E[T_{ij}(t) \cdot T_{jk}(t')|S_{ij}(t), S_{jk}(t')\] in Equation (4.10) as follows:

\[
E[T_{ij}(t) \cdot T_{jk}(t')|S_{ij}(t), S_{jk}(t')] \approx E[T_{ij}(t)|S_{ij}(t), S_{jk}(t')] \cdot E[T_{jk}(t')|S_{ij}(t), S_{jk}(t')] \\
\approx E[T_{ij}(t)|S_{ij}(t)] \cdot E[T_{jk}(t')|S_{jk}(t')] (4.11)
\]

Here, under the assumption that individual link travel time is only related to its own state, the conditional expected travel time value of the product of two consecutive link travel times is approximated to be the product of the conditional expected travel time values of individual links. For instance, when \(S_{ij}(t) = U\) and \(S_{jk}(t') = C\), we have
\[ E[T_{ij}(t) \cdot T_{jk}(t')|U, C] \approx E[T_{ij}(t)|U, C] \cdot E[T_{jk}(t')|U, C] \]
\[ \approx E[T_{ij}(t)|U] \cdot E[T_{jk}(t')|C] \]

In a real traffic scenario, a route from origin to the destination usually consists of more than two links. Let \( H \) be a set of links that constitutes route \( R \) between decision node \( i \) and the destination, for all composing link \((i,j)\) and \((j,k)\) \( \in A; \forall i, j, k \in N \), previous Equs (4.3) and (4.5) can be easily extended as follows:

\[ E[T_R(t)] = \sum_{(i,j) \in H} E[T_{ij}(t)] \]  
(4.12)

\[ Var[T_R(t)] = \sum_{(i,j) \in H} Var[T_{ij}(t)] + 2 \sum_{(i,j),(j,k) \in H} Cov(T_{ij}(t), T_{jk}(t')) \]  
(4.13)

After we calculate the mean and variance values by the proposed approximation method, according to Equation (4.1), we can obtain the coefficient of variation as follows:

\[ \text{Coefficient of Variation} = \frac{\sqrt{Var[T_R(t)]}}{E[T_R(t)]} \]  
(4.14)

In our proposed model, only two possible level-of-service states are considered. Nevertheless, extension to multiple link states can be done with minor modifications. For more than two possible link states (i.e., \( M > 2 \) possible congestion levels on the link), the individual link travel time mean and variance can be rewritten as follows:
\[ E[T_{ij}(t)] = \sum_{r=1}^{M} E[T_{ij}^r(t)] \cdot P_{ij}^r(t) \]  
\[ Var[T_{ij}(t)] = \sum_{r=1}^{M} ([E[T_{ij}^r(t)] - E[T_{ij}(t)])^2 \cdot P_{ij}^r(t) + Var[T_{ij}^r(t)] \cdot P_{ij}^r(t) \]  

where:

- \( E[T_{ij}^r(t)] \) – mean travel time of link \((i, j)\) at time \(t\) when link \((i, j)\) is in state \(r\);
- \( Var[T_{ij}^r(t)] \) – travel time variance value at time \(t\) when link \((i, j)\) is in state \(r\);
- \( P_{ij}^r(t) \) – the probability that link \((i, j)\) is in state \(r\) at time \(t\);

### 4.2.3 Reliability calculation using travel time distribution of the sum of two adjacent (STA) link travel times

In the previous section, CE approximation method to calculate path travel time variance value from individual links travel time distributions that make up a path and covariance between each link pairs has been proposed. In the following section, we present a second method in detail to calculate the route mean travel time and variance. In this framework, we still make the assumption that the travel time on two links that are not connected to each other are independent. We calculate the path travel time mean and variance value using individual link travel time means and variances, and variances of the sum of two adjacent link travel times along a route. Given the time varying link travel time distributions and a start time \(t\), we obtain the following parameter for each link:
• Mean link travel time
  
  \[ E[T_{ij}(t)] \]: mean travel time of link \((i, j)\) at time \(t\);

• Variance of link travel time

  \[ \text{Var}[T_{ij}(t)] \]: travel time variance value of link \((i, j)\) at time \(t\);

We first calculate the mean travel time value of a route. For simplicity of explanation, we first present our proposed method on a simple route that consists of only three links, e.g., link \((i, j)\), link \((j, k)\) and link \((k, n)\). We will later extend our method to a more general setting that consists of more than three links.

For a route of three adjacent link \((i, j)\), link \((j, k)\) and link \((k, n)\)

\[
T_R(t) = T_{ij}(t) + T_{jk}(t') + T_{kn}(t'')
\]

(4.17)

where \(T_{ij}(t)\) is the travel time on link \((i, j)\) at time \(t\), \(T_{jk}(t')\) is the travel time on link \((j, k)\) at time \(t'\), \(T_{kn}(t'')\) is the travel time on link \((k, n)\) at time \(t''\); \(t\) is the starting time at node \(i\), \(t'\) is the arrival time at node \(j\) and \(t''\) is the arrival time at node \(k\); \(T_R(t)\) is the route travel time at time \(t\). Here, we approximate the arrival time \(t'\) at node \(j\) as the current time \(t\) plus the expected travel time on link \((i, j)\) at time \(t\) as: \(t' = t + E[T_{ij}(t)]\). Similarly, the arrival time \(t''\) at node \(k\) is approximated as the arrival time \(t'\) plus the expected travel time on link \((j, k)\) at time \(t'\) as: \(t'' = t' + E[T_{jk}(t')]\).

The first step is to calculate the route mean travel time. According to the definition of the mean of a sum of random variables, the route mean travel time can be calculated based on mean travel time of each link at different start time.

\[
E[T_R(t)] = E[T_{ij}(t)] + E[T_{jk}(t')] + E[T_{kn}(t'')]
\]

(4.18)
Then, in the second step, we calculate the route travel time variance value. Based on the definition of the variance of a sum of random variables and the assumption that the travel time on two links that are not connected to each other are independent, we have

$$
Var[T_R(t)] = Var[T_{ij}(t) + T_{jk}(t') + T_{kn}(t''')]
= Var[T_{ij}(t)] + Var[T_{jk}(t')] + Var[T_{kn}(t'')]
+ 2Cov(T_{ij}(t), T_{jk}(t')) + 2Cov(T_{jk}(t'), T_{kn}(t'''))
$$

where $Cov(T_{ij}(t), T_{jk}(t'))$ and $Cov(T_{jk}(t'), T_{kn}(t'''))$ is the covariance value of each connected link pair. Here the link travel time correlation has been incorporated through the calculation of covariance value.

Following the variance of a sum of two random variables, we have

$$
2 Cov(T_{ij}(t), T_{jk}(t')) = Var[T_{ij}(t) + T_{jk}(t')] - Var[T_{ij}(t)] - Var[T_{jk}(t')]
$$

(4.20)

$$
2 Cov(T_{jk}(t'), T_{kn}(t''')) = Var[T_{jk}(t') + T_{kn}(t'')] - Var[T_{jk}(t')] - Var[T_{kn}(t'')]
$$

(4.21)

Therefore, the variance of the entire route can be calculated as follows:

$$
Var[T_R(t)] = Var[T_{ij}(t) + T_{jk}(t')] + Var[T_{jk}(t') + T_{kn}(t'')] - Var[T_{jk}(t')]
$$

(4.22)
In a real traffic scenario, a route from origin to the destination usually consists of more than three links. Let $H$ be a set of links that constitutes route $R$ between decision node $i$ and the destination, for all composing link $(i,j)$ and $(j,k) \in A, \forall i,j,k \in N$, previous Equs (4.18) and (4.22) can be easily extended as follows:

$$E[T_R(t)] = \sum_{(i,j) \in H} E[T_{ij}(t)] \quad (4.23)$$

$$Var[T_R(t)] = \sum_{(i,j) \in H} Var[T_{ij}(t)] + 2 \sum_{(i,j),(j,k) \in H} Cov(T_{ij}(t), T_{jk}(t')) \quad (4.24)$$

where $2Cov(T_{ij}(t), T_{jk}(t'))$ is calculated using Equation (4.20).

After we calculate the mean and variance values by the proposed approximation method, according to Equation (4.1), we can obtain the coefficient of variation as follows:

$$Coefficient of Variation = \frac{\sqrt{Var[T_R(t)]}}{E[T_R(t)]} \quad (4.25)$$

### 4.3 Experiment Result

In the following section, we illustrate our proposed methods via a simple example and a larger network (Sioux Falls network). We consider support points [46] as the format of input data, representing the travel time data (the joint discrete probability distribution). However, we emphasize here that the proposed methods do not rely on the assumption of support points. The proposed methods are also applicable to historical data. This will be shown in Chapter 5.

The joint discrete probability distribution is represented by a set of support points
Chapter 4

\[ W = \{w_1, w_2, \ldots w_L\}, \text{ where } w_i, i \in 1, 2, \ldots L \text{ is a } a \times |\Gamma| \text{ vector with probability } p_i \]
representing one support point that consists of the discrete values of travel time on all
the links of network, with \( L \) being 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 \( p_i \), we have

\[ \sum_{i=1}^{L} p_i = 1 \quad (4.26) \]

4.3.1 An illustrative example

We first consider an illustrative example with a small traffic network that is
composed of nine links (\(|A| = 9\)), seven nodes (\( N = 7 \)) and two different time intervals
\((b = 2)\). The topology of this network is depicted in Figure 4.1. Node \( i \) is the starting
point and \( D \) the destination node. There are two possible links from node \( i \) (link \((i, 1)\),
link \((i, 2)\)) and four possible routes to the destination. There are two time intervals,
each representing five minutes. Travel time data beyond time interval 1 are assumed
to be the same with time interval 1 for all the support points. In this example, there
are three support points \( w_1, w_2, \) and \( w_3 \). The probability of these support points are
0.5, 0.3 and 0.2, respectively. These support points data are given in Table 4.1, where
each column represents a support point.

For instance, consider a route \( i - 1 - D \) and assume that the start time is \( t = 2 \). In
the following, we first calculate the exact mean and variance values of route \( i - 1 - D \),
followed by a detailed calculation using our proposed CE approximation method. Then,
we calculate the mean and variance values of route \( i - 1 - 3 - D \) using the STA
approximation method. Finally, we give a comparison among these results.
4.3.2 Calculation from the exact distribution

From Table 4.1, we can see that there are three possible travel times (as there are three support points) for the route $i - 1 - D$: 4+3=7, 4+4=8, 5+5=10. The corresponding probability is 0.5, 0.3 and 0.2, respectively. Therefore, the mean and variance of its distribution are as follows:

$$E[T_{i-1-D}(t = 2)] = 7 \times 0.5 + 8 \times 0.3 + 10 \times 0.2 = 7.9$$

$$Var[T_{i-1-D}(t = 2)] = (7 - 7.9)^2 \times 0.5 + (8 - 7.9)^2 \times 0.3 + (10 - 7.9)^2 \times 0.2 = 1.29$$

Finally, we have the CV value of route $i - 1 - D$ as follows:
Table 4.1: Joint travel time realization of all links

<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>(i,1)</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>(i,2)</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>(1,D)</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>(1,3)</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>(3,D)</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>(2,D)</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>(2,4)</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>(4,D)</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(i,1)</td>
<td>4</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>(i,2)</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>(1,D)</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>(1,3)</td>
<td>3</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>(3,D)</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>(2,D)</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>(2,4)</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>(4,D)</td>
<td>3</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

Coefficient of Variation = $\frac{\sqrt{\text{Var}(T_{i-1-D}(t=2))}}{\text{E}(T_{i-1-D}(t=2))} = \frac{\sqrt{1.29}}{7.9} = 0.14$

4.3.3 Calculation using CE approximation method

Now we calculate the mean and variance of the travel time on route $i - 1 - D$ using our CE approximation method. In this example, at time $t$, we consider the expected travel time on a link as the threshold to decide the state of the link. This means that, if a travel time value $v$ is less than the expected travel time, then $v$ is considered to be a travel time value under uncongested conditions; otherwise, $v$ is considered as a travel time value under congested conditions.

First, we calculate the mean travel time of link $(i, 1)$. At $t = 2$, there are three possible travel times on link $(i, 1): 4, 4, 5$, with the corresponding probability 0.5, 0.3
Chapter 4

and 0.2 respectively. Therefore, the expected travel time on link \((i, 1)\) at time \(t = 2\) is:

\[
E[T_{i1}(t = 2)] = 4 \times 0.5 + 4 \times 0.3 + 5 \times 0.2 = 4.2
\]

Consequently, travel time value 4 is considered as uncongested and 5 is considered as congested. Hence,

\[
E[T_{i1}^U(t = 2)] = 4 \times 0.5 + 4 \times 0.3 = 4
\]

\[
Var[T_{i1}^U(t = 2)] = 0, P_{i1}^U(t = 2) = 0.8
\]

and

\[
E[T_{i1}^C(t = 2)] = 5, Var[T_{i1}^C(t = 2)] = 0, P_{i1}^C(t = 2) = 0.2
\]

As the mean travel time of link \((i, 1)\) is 4.2, it is expected that when the driver arrives at node 1, it will be at time \(t' = 6.2\). The possible travel times on link \((1, D)\) is 3, 4 and 5, with the corresponding probability 0.5, 0.3 and 0.2 respectively. Then \(E[T_{1D}(t' = 6.2)] = 3.7\). Accordingly, we have \(E[T_{1D}^U(t' = 6.2)] = 3, Var[T_{1D}^U(t' = 6.2)] = 0, P_{1D}^U(t' = 6.2) = 0.5\) and \(E[T_{1D}^C(t' = 6.2)] = 4.4, Var[T_{1D}^C(t' = 6.2)] = 0.12, P_{1D}^C(t' = 6.2) = 0.5\)

Then the route \(i - 1 - D\) mean travel time is

\[
E[T_{i-1-D}(t = 2)] = 4.2 + 3.7 = 7.9
\]

Now we continue to calculate the route variance value. We first calculate the
variance of link \((i, 1)\) at time \(t = 2\):

\[
E[Var[T_{i1}(t = 2)|S]] = Var[T_{i1}^U(t = 2)] \cdot P_{i1}^U(t = 2) \\
+ Var[T_{i1}^C(t = 2)] \cdot P_{i1}^C(t = 2)
\]

\[
= 0
\]

\[
Var[E[T_{i1}(t = 2)|S]] = [E[T_{i1}^U(t = 2)] - E[T_{i1}(t = 2)]]^2 \\
\cdot P_{i1}^U(t = 2) + [E[T_{i1}^C(t = 2)] \\
- E[T_{i1}(t = 2)]]^2 \cdot P_{i1}^C(t = 2)
\]

\[
= 0.16
\]

Then the variance of link \((i, 1)\) is

\[
Var[T_{i1}(t = 2)] = 0 + 0.16 = 0.16
\]

Similarly, we can calculate the variance of link \((1, D)\)

\[
Var[T_{1D}(t’ = 6.2)] = 0.55
\]

Most importantly, we need to calculate the covariance value of \(Cov(T_{i1}(t = 2), T_{1D}(t’ = 6.2))\).

\[
\alpha_{i1D} = \frac{P(S_{i1}(t = 2) = U, S_{1D}(t’ = 6.2) = U)}{P(S_{i1}(t = 2) = U)} = \frac{0.5}{0.8} = \frac{5}{8}
\]
\[
\lambda_{i1D} = \frac{P(S_{i1}(t = 2) = C, S_{1D}(t' = 6.2) = U)}{P(S_{i1}(t = 2) = C)} = \frac{0}{0.2} = 0
\]

The joint probability can be obtained as

\[
P(S_{i1}(t = 2) = U, S_{1D}(t' = 6.2) = U) = \frac{5}{8} \times 0.8 = 0.5
\]

\[
P(S_{i1}(t = 2) = U, S_{1D}(t' = 6.2) = C) = (1 - \frac{5}{8}) \times 0.8 = 0.3
\]

\[
P(S_{i1}(t = 2) = C, S_{1D}(t' = 6.2) = U) = 0
\]

\[
P(S_{i1}(t = 2) = C, S_{1D}(t' = 6.2) = C) = 1 \times 0.2 = 0.2
\]

Then, based on the Equation (4.11), the conditional expected travel time value of the product of two consecutive link travel times can be approximated as follows:

\[
E[T_{i1}(t = 2) \cdot T_{1D}(t' = 6.2)|S_{i1}(t = 2) = U, S_{1D}(t' = 6.2) = U] \approx E[T_{i1}(t = 2)|S_{i1}(t = 2) = U] \cdot E[T_{1D}(t' = 6.2)|S_{1D}(t' = 6.2) = U] = 4 \times 3 = 12
\]

Similarly,

\[
E[T_{i1}(t = 2) \cdot T_{1D}(t' = 6.2)|S_{i1}(t = 2) = U, S_{1D}(t' = 6.2) = C] = 4 \times 4.4 = 17.6
\]

\[
E[T_{i1}(t = 2) \cdot T_{1D}(t' = 6.2)|S_{i1}(t = 2) = C, S_{1D}(t' = 6.2) = U] = 5 \times 3 = 15
\]

\[
E[T_{i1}(t = 2) \cdot T_{1D}(t' = 6.2)|S_{i1}(t = 2) = C, S_{1D}(t' = 6.2) = C] = 5 \times 4.4 = 22
\]

Therefore, we have

\[
E[T_{i1}(t = 2) \cdot T_{1D}(t' = 6.2)]
\]
Finally, we get
\[\text{Cov}(T_{i1}(t = 2), T_{1D}(t' = 6.2))\]
\[= E[T_{i1}(t = 2)E[T_{1D}(t' = 6.2)] - E[T_{i1}(t = 2)]E[T_{1D}(t' = 6.2)]\]
\[= 15.68 - 4.2 \times 3.7 = 0.14\]

And the variance of route \(i - 1 - D\) is
\[\text{Var}[T_{i-1-D}(t = 2)]\]
\[= \text{Var}[T_{i1}(t = 2)] + \text{Var}[T_{1D}(t' = 6.2)] + 2\text{Cov}(T_{i1}(t = 2), T_{1D}(t' = 6.2))\]
\[= 0.16 + 0.55 + 2 \times 0.14 = 0.99\]

Finally, we have the CV value of route \(i - 1 - D\) as follows:

\[
\text{Coefficient of Variation} = \frac{\sqrt{\text{Var}[T_{i-1-D}(t = 2)]}}{E[T_{i-1-D}(t = 2)]} = \frac{\sqrt{0.99}}{7.9} = 0.13
\]

### 4.3.4 Calculation using STA approximation method

As to the two links route, our STA approximation method has the same result with the exact distribution. Now we calculate the mean and variance of the travel time on route \(i - 1 - 3 - D\) using STA approximation method. The start time is also assumed to be \(t = 2\). As the mean travel time of link \((i, 1)\) is 4.2, it is expected that when the driver arrives at node 1, it will be at time \(t' = 6.2\). From Table 4.1, we can see that there are three possible travel times (as there are three support points) for the joint travel time of links \((i, 1)\) and \((1, 3)\): 4+3=7, 4+3=7, 5+5=10. The corresponding probability is 0.5, 0.3 and 0.2, respectively. Therefore, the mean and variance of its joint distribution are as follows:

\[E[T_{i1}(t = 2) + T_{13}(t = 6.2)] = 7 \times 0.5 + 7 \times 0.3 + 10 \times 0.2 = 7.6\]
Table 4.2: Evidence of correlation exists for all possible routes

<table>
<thead>
<tr>
<th>Route</th>
<th>E diff</th>
<th>CE. diff</th>
<th>STA. diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>i-1-D</td>
<td>0.52</td>
<td>0.22</td>
<td>0.52</td>
</tr>
<tr>
<td>i-1-3-D</td>
<td>2.20</td>
<td>1.14</td>
<td>1.68</td>
</tr>
<tr>
<td>i-2-D</td>
<td>1.28</td>
<td>1.28</td>
<td>1.28</td>
</tr>
<tr>
<td>i-2-4-D</td>
<td>1.24</td>
<td>0.84</td>
<td>0.84</td>
</tr>
</tbody>
</table>

\[
Var[T_{i1}(t = 2) + T_{i3}(t = 6.2)] = (7 - 7.6)^2 \times 0.5 + (7 - 7.6)^2 \times 0.3 \\
\quad + (10 - 7.6)^2 \times 0.2 \\
= 1.44
\]

Similarly, the mean and variance travel time for link (1, 3) is 3.4 and 0.64. Then, it is expected that when the driver arrives at node 3, it will be at time \( t'' = 9.6 \). Accordingly, we have \( Var[T_{i3}(t = 6.2) + T_{3D}(t = 9.6)] = 2.29 \).

Therefore, based on Equation (4.22), the variance of the entire route \( i - 1 - 3 - D \) can be calculated as follows:

\[
Var[T_{i-1-3-D}(t = 2)] = 1.44 + 2.29 - 0.64 = 3.09
\]

Finally, we get the CV value of route \( i - 1 - 3 - D \)

\[
Coefficient of Variation = \frac{\sqrt{Var[T_{i-1-3-D}(t = 2)]}}{E[T_{i-1-3-D}(t = 2)]} = \frac{\sqrt{3.09}}{10.3} = 0.17
\]

Based on the above detailed calculation, results are further analyzed to validate that whether the consideration of link travel time correlation is reasonable in our proposed frameworks. Therefore, we first calculate the sum of each individual links travel time variance values for all possible routes. Then the difference between the path travel time...
Table 4.3: Analytical results for all possible routes

<table>
<thead>
<tr>
<th>Route</th>
<th>E mean</th>
<th>E var</th>
<th>E CV</th>
<th>CE. mean</th>
<th>CE. var</th>
<th>CE. CV</th>
<th>STA. mean</th>
<th>STA. var</th>
<th>STA. CV</th>
</tr>
</thead>
<tbody>
<tr>
<td>i-1-D</td>
<td>7.9</td>
<td>1.29</td>
<td>0.14</td>
<td>7.9</td>
<td>0.99</td>
<td>0.13</td>
<td>7.9</td>
<td>1.29</td>
<td>0.14</td>
</tr>
<tr>
<td>i-1-3-D</td>
<td>10.3</td>
<td>3.61</td>
<td>0.18</td>
<td>10.3</td>
<td>2.55</td>
<td>0.16</td>
<td>10.3</td>
<td>3.09</td>
<td>0.17</td>
</tr>
<tr>
<td>i-2-D</td>
<td>5.8</td>
<td>2.56</td>
<td>0.28</td>
<td>5.8</td>
<td>2.56</td>
<td>0.28</td>
<td>5.8</td>
<td>2.56</td>
<td>0.28</td>
</tr>
<tr>
<td>i-2-4-D</td>
<td>10.1</td>
<td>2.29</td>
<td>0.15</td>
<td>10.1</td>
<td>1.89</td>
<td>0.14</td>
<td>10.1</td>
<td>1.89</td>
<td>0.14</td>
</tr>
</tbody>
</table>

variance and the sum of related individual link travel time variances is shown. Results for all possible routes are presented in Table 4.2, where E dif, CE. dif and STA. dif are the difference values between the path travel time variance and the sum of related individual link travel time variances calculated from the full distribution of travel times, CE approximation method and STA approximation method. As is shown in the Table 4.2, all the numbers are positive values. It indicates the covariance exists across links and link travel times are typically highly correlated. These results also indicate that considering spatial correlation between adjacent links in our proposed approximation method shows advantage compared to the model that assumes independent link travel times (Rakha et al. [99]).

Finally, the whole reliability calculation results are shown in Table 4.3. According to Table 4.3, the mean and variance of the travel time on route i-1-D calculated based on the exact travel time distribution is 7.9 and 1.29, respectively. The corresponding values obtained by our two approximations are 7.9 and 0.99, and 7.9 and 1.29 indicating a reasonable estimation. Results for other routes are also presented in the same Table where E mean, E var and E CV are the mean, variance and CV values calculated from the full distribution of travel times, CE. mean, CE. var and CE. CV are corresponding values obtained by CE approximation method, while STA. mean, STA. var and STA. CV are corresponding values obtained by STA approximation method. In this scenario of a small network, it can be seen that the decision regarding the most reliable route at node $i$ using the approximated mean and variance is the same as that of using the full distribution. Moreover, our two approximation methods underestimate the
variance value of the whole route travel time. When compared with CE approximation method, STA approximation method shows a better performance. It is closer to the full distribution of travel times than CE approximation method.

### 4.3.5 Experiment in a larger network

In this section, we test our proposed methods in a larger network. The city of Sioux Falls is chosen since it is a well-known test network in traffic area [9,63,116]. The Sioux Falls network consists of 24 nodes and 76 links. In this test, we consider only one origin-destination pairs from start node 1 to destination node 24 for clear comparison. For the ease of calculation, we adapt the one way direction instead of dual directions for each link. Therefore, the original 76 links will be reduced to 38 links (See Figure 4.2). The additional information such as link length and speed is shown in Table 4.4. Then the free flow travel time for each link can be obtained accordingly. Similar to the data generation process in Chapter 3, a multivariate normal distribution is assumed for the joint distribution of all link travel time random variables. The same standard deviation of link travel time and correlation coefficient for each pair of links are used. Moreover, each link travel time value is rounded to the nearest integer, which follows the same procedure as in [45] for data generation. The actual average link travel time varies in each time interval and is generated as follows. Firstly, the free flow travel time for each link is noted as the minimum travel time in the first time interval. Then the maximum mean link travel time is generated as a few times larger than the minimum travel time value, which is a random generated number between two and four. Finally, the median mean value is randomly created between the minimum and the maximum ones, which is used as the last time interval value. An example of three random chosen links’ mean travel times in each time interval are depicted in Figure 4.3. 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 twelve fifteen-minute time intervals covering possible departing times in a typical weekday morning. A common standard
deviation (1) of link travel time, and a unique correlation coefficient (0.6) between that on consecutive links, are used to generate the data in this example.

Figure 4.2: Sioux Falls network

There are 76 routes for the selected origin and destination. Here we first select part of these routes to compare with the full distribution of travel times based on CV values. In terms of free flow travel time, link 9 and link 24 have the maximum values (7.68 minutes); link 7, link 13 and link 27 have the minimum values (0.24 minutes). Therefore, we select two routes containing these two group of links, noted as Route 1 and Route 2. In terms of the number of links composing the route between the origin and destination, another two routes (Route 3 and Route 4) are selected to represent the minimum number of links (4) and the maximum number of links (16). Specifically, Route 1={1-3, 3-4, 4-5, 5-9, 9-10, 10-15, 15-19, 19-20, 20-21, 21-24}, Route 2={1-2,
Figure 4.3: Mean travel times of three random selected links at different time interval

2-6, 6-8, 8-16, 16-17, 17-19, 19-20, 20-21, 21-24, Route 3={1-3, 3-12, 12-13, 13-24} and Route 4={1-3, 3-4, 4-5, 5-6, 6-8, 8-9, 9-10, 10-11, 11-14, 14-15, 15-19, 19-20, 20-21, 21-22, 22-23, 23-24}.

The comparison is depicted in Table 4.5 and Table 4.6 for two random selected start times. We compare the percentage difference of the CV values in different approximation methods from the full distribution of travel times. In particular, percentage difference is calculated as follows:

\[ \Delta_i = \frac{\text{approximation method } i \times \text{CV value} - \text{CV value of full travel time distribution}}{\text{CV value of full travel time distribution}} \times 100\% \]

We can see clearly from the result, in this large network, that the percentage difference using our CE approximation method are always larger than STA approximation method for all selected routes in different start times. Hence, our proposed STA approximation method shows better performance than CE method. Then we repeat the experiment with another eight random selected start times. For each start time, we
Chapter 4

Table 4.4: Data for Sioux Falls network

<table>
<thead>
<tr>
<th>Link</th>
<th>Length</th>
<th>Speed limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.4</td>
<td>25</td>
</tr>
<tr>
<td>2</td>
<td>0.3</td>
<td>50</td>
</tr>
<tr>
<td>3</td>
<td>0.2</td>
<td>25</td>
</tr>
<tr>
<td>4</td>
<td>0.2</td>
<td>25</td>
</tr>
<tr>
<td>5</td>
<td>0.4</td>
<td>25</td>
</tr>
<tr>
<td>6</td>
<td>1.7</td>
<td>25</td>
</tr>
<tr>
<td>7</td>
<td>0.1</td>
<td>25</td>
</tr>
<tr>
<td>8</td>
<td>0.2</td>
<td>25</td>
</tr>
<tr>
<td>9</td>
<td>3.2</td>
<td>25</td>
</tr>
<tr>
<td>10</td>
<td>0.5</td>
<td>25</td>
</tr>
<tr>
<td>11</td>
<td>0.9</td>
<td>25</td>
</tr>
<tr>
<td>12</td>
<td>1.1</td>
<td>25</td>
</tr>
<tr>
<td>13</td>
<td>0.2</td>
<td>50</td>
</tr>
<tr>
<td>14</td>
<td>0.5</td>
<td>25</td>
</tr>
<tr>
<td>15</td>
<td>0.3</td>
<td>25</td>
</tr>
<tr>
<td>16</td>
<td>1.3</td>
<td>25</td>
</tr>
<tr>
<td>17</td>
<td>0.6</td>
<td>25</td>
</tr>
<tr>
<td>18</td>
<td>0.2</td>
<td>25</td>
</tr>
<tr>
<td>19</td>
<td>0.6</td>
<td>25</td>
</tr>
<tr>
<td>20</td>
<td>0.3</td>
<td>25</td>
</tr>
<tr>
<td>21</td>
<td>0.3</td>
<td>50</td>
</tr>
<tr>
<td>22</td>
<td>2.9</td>
<td>25</td>
</tr>
<tr>
<td>23</td>
<td>1.2</td>
<td>25</td>
</tr>
<tr>
<td>24</td>
<td>3.2</td>
<td>25</td>
</tr>
<tr>
<td>25</td>
<td>1.3</td>
<td>25</td>
</tr>
<tr>
<td>26</td>
<td>0.5</td>
<td>25</td>
</tr>
<tr>
<td>27</td>
<td>0.2</td>
<td>50</td>
</tr>
<tr>
<td>28</td>
<td>1.5</td>
<td>50</td>
</tr>
<tr>
<td>29</td>
<td>0.9</td>
<td>25</td>
</tr>
<tr>
<td>30</td>
<td>0.3</td>
<td>25</td>
</tr>
<tr>
<td>31</td>
<td>0.5</td>
<td>25</td>
</tr>
<tr>
<td>32</td>
<td>1.6</td>
<td>25</td>
</tr>
<tr>
<td>33</td>
<td>0.5</td>
<td>25</td>
</tr>
<tr>
<td>34</td>
<td>1.2</td>
<td>25</td>
</tr>
<tr>
<td>35</td>
<td>1.5</td>
<td>25</td>
</tr>
<tr>
<td>36</td>
<td>0.9</td>
<td>50</td>
</tr>
<tr>
<td>37</td>
<td>0.9</td>
<td>25</td>
</tr>
<tr>
<td>38</td>
<td>1.6</td>
<td>25</td>
</tr>
</tbody>
</table>

Table 4.5: Analytical results for a larger network for the first random selected start time

<table>
<thead>
<tr>
<th>Route</th>
<th>$\Delta_{CE}$</th>
<th>$\Delta_{STA}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Route 1</td>
<td>0.1230</td>
<td>0.0578</td>
</tr>
<tr>
<td>Route 2</td>
<td>0.6880</td>
<td>0.1334</td>
</tr>
<tr>
<td>Route 3</td>
<td>0.1427</td>
<td>0.1087</td>
</tr>
<tr>
<td>Route 4</td>
<td>0.7177</td>
<td>0.0013</td>
</tr>
</tbody>
</table>

Table 4.6: Analytical results for a larger network for the second random selected start time

<table>
<thead>
<tr>
<th>Route</th>
<th>$\Delta_{CE}$</th>
<th>$\Delta_{STA}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Route 1</td>
<td>0.8205</td>
<td>0.1272</td>
</tr>
<tr>
<td>Route 2</td>
<td>0.8722</td>
<td>0.0521</td>
</tr>
<tr>
<td>Route 3</td>
<td>0.3451</td>
<td>0.0829</td>
</tr>
<tr>
<td>Route 4</td>
<td>0.4738</td>
<td>0.1451</td>
</tr>
</tbody>
</table>
Table 4.7: Analytical results for a larger network for ten random selected start times

<table>
<thead>
<tr>
<th>Start time</th>
<th>Numbers in CE method</th>
<th>Numbers in STA method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Start time 1</td>
<td>37</td>
<td>39</td>
</tr>
<tr>
<td>Start time 2</td>
<td>5</td>
<td>71</td>
</tr>
<tr>
<td>Start time 3</td>
<td>29</td>
<td>47</td>
</tr>
<tr>
<td>Start time 4</td>
<td>39</td>
<td>37</td>
</tr>
<tr>
<td>Start time 5</td>
<td>5</td>
<td>71</td>
</tr>
<tr>
<td>Start time 6</td>
<td>36</td>
<td>40</td>
</tr>
<tr>
<td>Start time 7</td>
<td>6</td>
<td>70</td>
</tr>
<tr>
<td>Start time 8</td>
<td>4</td>
<td>72</td>
</tr>
<tr>
<td>Start time 9</td>
<td>3</td>
<td>73</td>
</tr>
<tr>
<td>Start time 10</td>
<td>11</td>
<td>65</td>
</tr>
</tbody>
</table>

calculate the percentage difference value in all possible routes over two approximation methods and note down total route numbers that one approximation method is better than the other one.

The final comparison results produced by different methods (CE method, STA method respectively) are shown in Table 4.7 under different start times. The table shows the total better performance numbers for each method over all possible routes. From these experimental results, it’s quite clear that, in most cases, STA approximation method performs very well in these experiments with larger total numbers that shows better results.

To summarize, the results in these experiments are consistent with the expected advantages of our proposed STA method. In most cases, STA method shows smaller percentage differences than CE’s method suggesting the advantage of the proposed approximation method. In the later chapter of this thesis, our proposed STA approximation method of travel time reliability calculation is field tested in the Melbourne traffic network with real data for further validation. The performance of our proposed approximation methods for travel time reliability has been investigated in the simulation environment, it is also important to consider computation time issue in reality. However, it takes time to setup and investigate the computation time in reality, and we leave it for future work.
4.4 Summary of this Chapter

In this chapter, we apply an application of reliable route selection in a stochastic time-dependent traffic network. We propose two approximation methods for calculating travel time reliability, taking into account the uncertainty of link travel times and the spatial correlation between adjacent links. The proposed methods calculate the mean travel time and the variance of a route via individual link travel time means and variances, and variances of the sum of two adjacent link travel times along a route. An illustrative example and experiment in Sioux Falls network are provided to demonstrate the calculation process of the proposed methods. The calculation results indicate that link travel times are typically highly correlated. Through a comparison with the exact values calculated via the entire discrete joint distribution of link travel times, we observe that the proposed methods compute results that are close to the exact value and STA approximation method shows better results than CE approximation method.
Chapter 5

Validation and Discussion Using Real Traffic Data

In the previous chapters, we have introduced the shortest and reliable route guidance frameworks. Some tests have been conducted through illustrative and simulation examples. The preliminary finding is that the proposed methods can provide good results with hypothesized link travel time data. However, a better validation method is through real data in real traffic networks. The evaluation of the proposed approach with real traffic data is a very important task. This is an effective way to demonstrate the performance of the proposed framework and to judge its advantages and disadvantages. This chapter discusses the evaluation of the proposed approaches using SCATS data in a part of the road network of the city of Melbourne in Australia. The work discuss in this chapter aims to test the accuracy of the developed route guidance methods and understand some important issues relating to evaluating dynamic route guidance systems in a real traffic network.
5.1 Introduction

In the previous chapters, travellers’ two major criterions in routing optimization have been addressed. Firstly, travellers wish to determine a route with the least expected travel time for a given an origin and destination. A simple, robust framework based on dynamic programming approach is developed to study this problem. When faced with uncertainty, travellers are also concerned about the reliability of their travel times. Therefore, approximation methods to obtain the reliability of a route travel time in a stochastic time-dependent traffic network have been proposed in the previous chapter and the most reliable route can be decided accordingly. Numerical tests 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. However, a better way to evaluate the proposed frameworks is to test the proposed approaches in real world scenarios with real data. The evaluation of a theoretical approach with real data is a very significant task in practice. It is the an effective way to demonstrate the characteristics of the new proposed frameworks and to identify its advantages and disadvantages. We can then draw a conclusion on whether it can be well applied in reality or not. Therefore, in this chapter, we use the experimentation to evaluate our proposed frameworks in a real traffic network, with real data, and make a comparison with a number of existing methods in terms of route selection. The performance differences are analyzed in detail and summarized in the results.

Travel time is an important parameter to be considered when using a route guidance system. In the previous chapter, travel time was estimated by certain data generation method. However, this estimation method can cause errors in the evaluation of our proposed framework. Therefore in this chapter, actual historical travel time data in a real traffic network are utilized to evaluate our proposed frameworks. Travel time can be obtained in a number of ways. Traditionally, it was collected by two main methods: the floating car technique and license plate matching. Both methods are labor intensive, costly for large-scale collection of travel time data, and as a result, unable
to supply travel time on a continuous basis. A number of new advanced technologies have emerged in the past decade that offer promising alternatives to the previous two methods. These include Advanced Vehicle Identification (AVI) and Automatic Vehicle Location (AVL). While these new systems can provide continuous, real-time travel time information, they require considerable new infrastructure investment. It will take years for full implementation of such a system to be a viable alternative of current traffic surveillance systems. Additionally, many cities have invested considerably in traffic surveillance technology such as video cameras, inductive loop detectors. Traffic conditions on most freeways and arterial roads in urban areas are well monitored through frequently collected traffic information. This information is used for traffic management and ITS applications. It is important that accurate methods can be developed to estimate travel time based on outputs from these types of traffic sensors [127]. In this thesis, we get the inductive loop detectors data from Sydney Coordinated Adaptive Traffic Systems (SCATS). This system adjusts traffic lights by measurements of inductive loop detectors in the Melbourne traffic network. As such, the main functions of vehicle detector and SCATS adaptive control systems will be introduced.

In the SCATS system, loop detectors do not directly provide travel time data. Instead, both measured traffic flow and signal timing output data are provided [84]. As such, we need to convert them to speed or travel time data. There are various methods for estimating travel time on freeways using loop detector data. However, the estimation of travel time on arterial roads is a much more challenging task because of the dynamic characteristic of traffic flow, interrupted facilities with signals and other control devices that interrupt traffic flow on arterial routes. Therefore, we analyze various travel time estimation methods and find an effective way to estimate arterial road travel time, including the possible signal delay estimate.

The chapter is structured as follows: Section 5.2 presents the test network and Section 5.3 introduces methods employed for the acquisition and processing of the data.
required to conduct the experiments. A detailed analysis of travel time distribution based on estimated travel time obtained from SCATS data is conducted for further assessment of the travel time reliability in Section 5.4. Section 5.5 reports and discusses the results obtained from the validation. Section 5.6 summarizes this chapter.

5.2 Test Network

The test network employed in the simulation study is chosen from the inner parts of Melbourne. The network serves local and regional traffic and acts as a hub for traffic entering and exiting the city area. Diversity is one of the key criteria of the network chosen with this network including many different road types ranging from freeways to minor roads. This network also contains a wide range of land uses from residential districts to central business areas. This area starts from Maidstone (node A in Figure 5.1) and ends in the Hawthorn area (node B in Figure 5.1). There is a high level of passenger demand and traffic flow especially in peak hours. Most of the corridors are main corridors to reach the Melbourne CBD. Some of the links, such as Victoria Street and Punt Road, are the busiest routes that serve both commuter and freight traffic. The network consists of 127 nodes and 134 links. It covers an area approximately 16 km long and 6 km wide. The urban traffic signal control system, SCATS, is adopted in Melbourne. We use loop detectors data extracted from the SCATS system. More specifically, loop detectors are installed in each lane at the stop line of every major intersection (Figure 5.2).

5.3 Traffic Data

Primary traffic data used in the Melbourne case study is loop detectors data that was collected during the morning period from 6am to 10am, each weekday in 2012 (totally 167 days). Within the study area, there are a total of 134 loop detectors.
Figure 5.1: Test case scenario routes of the Melbourne traffic network

Figure 5.2: Position of SCATS loop detectors (e.g. the detectors of the west-east approach are coded as 8, 9, 10)
Each detector provides 24-hour traffic information continuously. Then for the selected period, the total size of the sample data set included 240 records (recorded every one minute).

SCATS is an area wide adaptive traffic signal control system. It controls the cycle time, green splits and offsets for signalized intersections and mid-block pedestrian crossings in real-time. Based on vehicle detectors, it can adaptively modify these values to optimize the operation to suit the prevailing traffic. Alternatively, it can manage intersections in fixed-time mode where it can change plans by time of day, and day of week. It is designed to coordinate traffic signals for networks or for arterial roads. The system has loop detectors installed at every through, left-turn and right-turn lanes of a signalized traffic intersection. It can adjust signal timings throughout the system in response to measured changes of the traffic.

SCATS can also adapt traffic light timing to manage unexpected conditions and minimize delays caused by events or incidents. For instance, SCATS can be manually changed to traffic light cycles when sporting and social occur. This allows improved access and traffic flow around major venues. When taking input from loop detectors, SCATS adaptive control system employs more than 10 loop detectors for each intersection. At a signalized intersection, loop detectors are installed at each lane of signalized intersection stop lines. These detectors provide valuable data for traffic monitoring.

The SCATS system utilizes the feedback control to continuously adjust system control parameters (signal cycle time, signal offset) to minimize traffic delay and vehicle stops [84]. A sample of the SCATS system loop detectors output data is shown in Figure 5.3. Here, NB_SCATS_SITE is the intersection number in the SCATS system, QT_DATE, DT_TIME is the date and accurate time information and NB_DETETCTOR is the loop detector number in one intersection. QT_MAX_FLOW stands for the maximum flow (maximum vehicle number per hour per lane) or capacity of the lane, QT_CYCLETIME is signal cycle time (total time per signal cycle), QT_PHASETIME
Chapter 5

is green light time in one signal cycle and QT\_VO represents arrival flow (number of car passing per minute) of lanes installed with detectors. QT\_DS (degree of saturation) and QT\_VK (equivalent count) are calculated from the signal timing data.

<table>
<thead>
<tr>
<th>Wk, SCATS Lane</th>
<th>ST_DATE</th>
<th>ST_TIME</th>
<th>NB_DETECTION</th>
<th>QT_MAX_FLOW</th>
<th>QT_CYCLE_FLOW</th>
<th>QT_PHASE_FLOW</th>
<th>QT_NO</th>
<th>QT_VK</th>
<th>QT_DS</th>
<th>QT_DS</th>
<th>QT_DS</th>
</tr>
</thead>
<tbody>
<tr>
<td>434</td>
<td>01/06/2012</td>
<td>09:25:00 AM</td>
<td>8</td>
<td>1299</td>
<td>124</td>
<td>17</td>
<td>1</td>
<td>24</td>
<td>41</td>
<td>24</td>
<td>41</td>
</tr>
<tr>
<td>434</td>
<td>01/06/2012</td>
<td>09:25:00 AM</td>
<td>9</td>
<td>1274</td>
<td>124</td>
<td>11</td>
<td>15</td>
<td>73</td>
<td>61</td>
<td>73</td>
<td>61</td>
</tr>
<tr>
<td>434</td>
<td>01/06/2012</td>
<td>09:25:00 AM</td>
<td>10</td>
<td>1284</td>
<td>124</td>
<td>71</td>
<td>12</td>
<td>158</td>
<td>103</td>
<td>12</td>
<td>158</td>
</tr>
<tr>
<td>4273</td>
<td>01/06/2012</td>
<td>09:25:00 AM</td>
<td>1</td>
<td>1250</td>
<td>100</td>
<td>19</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4274</td>
<td>01/06/2012</td>
<td>09:25:00 AM</td>
<td>2</td>
<td>1280</td>
<td>120</td>
<td>39</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4274</td>
<td>01/06/2012</td>
<td>09:25:00 AM</td>
<td>3</td>
<td>1250</td>
<td>120</td>
<td>10</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4274</td>
<td>01/06/2012</td>
<td>09:25:00 AM</td>
<td>4</td>
<td>1280</td>
<td>120</td>
<td>49</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>430</td>
<td>01/06/2012</td>
<td>09:25:00 AM</td>
<td>1</td>
<td>1289</td>
<td>121</td>
<td>39</td>
<td>6</td>
<td>7</td>
<td>32</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>430</td>
<td>01/06/2012</td>
<td>09:25:00 AM</td>
<td>2</td>
<td>1374</td>
<td>121</td>
<td>74</td>
<td>28</td>
<td>38</td>
<td>80</td>
<td>28</td>
<td>38</td>
</tr>
<tr>
<td>430</td>
<td>01/06/2012</td>
<td>09:25:00 AM</td>
<td>3</td>
<td>1289</td>
<td>121</td>
<td>34</td>
<td>17</td>
<td>48</td>
<td>48</td>
<td>17</td>
<td>48</td>
</tr>
<tr>
<td>433</td>
<td>01/06/2012</td>
<td>09:25:00 AM</td>
<td>1</td>
<td>1289</td>
<td>121</td>
<td>40</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>434</td>
<td>01/06/2012</td>
<td>09:25:00 AM</td>
<td>2</td>
<td>1289</td>
<td>121</td>
<td>40</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>434</td>
<td>01/06/2012</td>
<td>09:25:00 AM</td>
<td>3</td>
<td>1289</td>
<td>121</td>
<td>40</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>434</td>
<td>01/06/2012</td>
<td>09:25:00 AM</td>
<td>4</td>
<td>1289</td>
<td>121</td>
<td>40</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>434</td>
<td>01/06/2012</td>
<td>09:25:00 AM</td>
<td>5</td>
<td>1289</td>
<td>121</td>
<td>40</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>434</td>
<td>01/06/2012</td>
<td>09:25:00 AM</td>
<td>6</td>
<td>1289</td>
<td>121</td>
<td>40</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>434</td>
<td>01/06/2012</td>
<td>09:25:00 AM</td>
<td>7</td>
<td>1289</td>
<td>121</td>
<td>40</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>434</td>
<td>01/06/2012</td>
<td>09:25:00 AM</td>
<td>8</td>
<td>1289</td>
<td>121</td>
<td>40</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>434</td>
<td>01/06/2012</td>
<td>09:25:00 AM</td>
<td>9</td>
<td>1289</td>
<td>121</td>
<td>40</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>434</td>
<td>01/06/2012</td>
<td>09:25:00 AM</td>
<td>10</td>
<td>1289</td>
<td>121</td>
<td>40</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>434</td>
<td>01/06/2012</td>
<td>09:25:00 AM</td>
<td>11</td>
<td>1289</td>
<td>121</td>
<td>40</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>434</td>
<td>01/06/2012</td>
<td>09:25:00 AM</td>
<td>12</td>
<td>1289</td>
<td>121</td>
<td>40</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>434</td>
<td>01/06/2012</td>
<td>09:25:00 AM</td>
<td>13</td>
<td>1289</td>
<td>121</td>
<td>40</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>434</td>
<td>01/06/2012</td>
<td>09:25:00 AM</td>
<td>14</td>
<td>1289</td>
<td>121</td>
<td>40</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>434</td>
<td>01/06/2012</td>
<td>09:25:00 AM</td>
<td>15</td>
<td>1289</td>
<td>121</td>
<td>40</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>434</td>
<td>01/06/2012</td>
<td>09:25:00 AM</td>
<td>16</td>
<td>1289</td>
<td>121</td>
<td>40</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>434</td>
<td>01/06/2012</td>
<td>09:25:00 AM</td>
<td>17</td>
<td>1289</td>
<td>121</td>
<td>40</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>434</td>
<td>01/06/2012</td>
<td>09:25:00 AM</td>
<td>18</td>
<td>1289</td>
<td>121</td>
<td>40</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>434</td>
<td>01/06/2012</td>
<td>09:25:00 AM</td>
<td>19</td>
<td>1289</td>
<td>121</td>
<td>40</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>434</td>
<td>01/06/2012</td>
<td>09:25:00 AM</td>
<td>20</td>
<td>1289</td>
<td>121</td>
<td>40</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Figure 5.3: A sample extract of SCATS data

In order to find the shortest routes and reliable routes in a considered traffic network, we need the link travel time distributions. However, detector output data do not directly provide such distributions. Therefore, estimating travel times using this data is the first step for the evaluation of our route guidance methods. In the literature, there are a number of methods of estimating travel times based on various data sources such as arterial links [84], floating vehicles [62,65] and simulation software [60]. These methods work well based on their particular input data. In this thesis, the available real traffic data are from loop detectors in the SCATS system. In reality, however, it is difficult to estimate various input parameters in different travel time estimation methods that not using SCATS data directly. Therefore, in this thesis, we only focus on travel time estimation methods developed using SCATS data directly. Normally, urban travel time between any two neighboring intersections includes two parts: cruise time and intersection delay (Cheu et al. [25], Ran et al. [100], Xie et al. [124], Nguyen and Gaffney [84]).
Arterial Link Travel Time = Cruise Time + Delay Time \hspace{1cm} (5.1)

For example, Ehsan et al. [75] provide a method to estimate bus travel time using the SCATS system output data. Luk [71] proposes the ARRB travel time model (ATTM) to estimate travel time on arterial road using SCATS data at each signalized road. Nguyen and Gaffney [84] discuss the method using SCATS data in Melbourne network to identify traffic congestion and estimate travel time. Among various urban travel time estimation methods, we adopt the ATTM using SCATS data [71]. The measured traffic flow and signal timing output data from the SCATS system are utilized directly as inputs to the delay and cruise time calculation. The method used in this study is presented below. Other possible methods in the literature are detailed in Chapter 2.

Cruise time is simply the time it takes to propagate across the link i.e., from one detector site to the next. The delay, experienced on the arterial signalized lines, is mainly related to the intersection where conflicting movements are controlled by traffic signals. The delay is defined as the difference in travel time when a vehicle is not affected by the controlled intersection and when a vehicle is influenced by the controlled intersection. In order to obtain the cruise time value, we need to calculate queue length first. More specifically,

Queue length for each link is calculated as the total number of queuing vehicles at an intersection [6]. The queue length is given by:

\[
Queue \ Length = \frac{r \times q_a}{3600} + \frac{Overflow \ Delay \times Q}{3600} \hspace{1cm} (5.2)
\]

where \(r\) is the effective red time (sec), \(q_a\) is arrival flow (veh/h) and \(Q\) is lane capacity (veh/h). Here \(q_a\) and \(Q\) correspond to 60*QT_VO and QT_MAX_FLOW in the above data. And the effective red time \(r\) is the difference between signal cycle time (QT_CYCLETIME) and green light time (QT_Phasetime).
Chapter 5

The calculation of queue length is related to overflow delay. Traffic seldom arrives uniformly at a traffic signal. Therefore, we need to include an extra delay (overflow delay) due to the additional flow exceeding the service flow of the intersection at the end of the green period. The overflow delay is given by:

$$Overflow\ Delay = 0.25T_p\left[(x - 1) + (x - 1)^2 + \frac{6(x - x_0)}{Q \cdot T_p}\right]^{0.5} \quad (5.3)$$

where

- $T_p =$ duration of the analysis period
- $q_a =$ arrival flow (veh/h)
- $Q =$ lane capacity (veh/h)
- $x =$ arrival flow to capacity ratio of each lane, $x = q_a/Q$

Here $x_0$ is given by $x_0 = 0.67 + s \cdot g/600$, where $g$ is green light time (sec) and $s$ is the saturation flow (veh/sec). Saturation flow $s$ is calculated as $s = (c \cdot Q)/3600g$, where $c$ is the signal cycle time (QT_CYCLETIME) and $g$ is green light time (QT_PHASETIME). The QT_DS value in SCATS is an indication whether the green light time is utilized effectively and it is not the same with the arrival flow to capacity ratio in the Equation (5.3).

We approximate the cruise time as follows:

$$Cruise\ Time = \frac{Link\ Length - Queue\ Length}{Free\ Flow\ Speed} \quad (5.4)$$

Here, the link length is measured from Google Map and queue length can be calculated using Equation (5.2). We approximate the free flow speed in different areas of the Melbourne traffic network as follows: the Melbourne CBD area is 40 km/h, outside the Melbourne CBD area is 60 km/h, and eastern freeway is 100 km/h. Delay at an
intersection is the time delay experienced by vehicles while passing the intersection. It can be determined by the ATTM model as follows:

\[ \text{Intersection Delay} = \theta \times \text{Uniform Delay} + \text{Overflow Delay} \]  \hspace{1cm} (5.5)

where uniform delay is part of intersection delay that would occur when a traffic stream arrives uniformly at a signalized approach.

More precisely, uniform delay is given by:

\[ \text{Uniform Delay} = \frac{0.5 \times r^2}{c(1 - \frac{q}{s})} \]  \hspace{1cm} (5.6)

where \( r \) is the effective red time (sec), \( c \) is the signal cycle time (sec), \( q \) is traffic flow (veh/sec) and \( s \) is the saturation flow (veh/sec). The traffic flow is assumed to be the average of arrival flow in each 15 minutes.

In the ATTM model, it does not distinguish between the condition in which the traffic is highly congested and the road is empty. In particular, it only considers the typical traffic condition in the Melbourne traffic network in the morning by setting the parameter \( \theta \). \( \theta \) is the progression factor, which ranges between 0.0 and 2.6 depending on traffic conditions and how well adjacent signals are coordinated. If \( \theta \) is 0, there is a perfect progression. If it is 1, the signal is assumed to operate as an isolated intersection. A progression factor of 2 is used for the relevant link [71].

Finally, the link travel time values are calculated based on SCATS output data. The ATTM model is one of the ways to obtain travel time values. There are some other means to obtain travel time values, such as Google Traffic. Therefore, for all possible routes from the start point to the destination, the calculated travel time using the ATTM model is compared with the travel time that obtained from Google Traffic. We set the start time at 9am in the weekday morning and the comparison result for
route travel time in minute is shown in Table 5.1. As can be seen from the table, the travel time we calculated using the ATTM model does not consistently match the travel time obtained by Google Traffic. The ATTM model is based on SCATS data and Google Traffic analyzes travel time using mobile date. Therefore, different methods for obtaining travel time can vary in certain degree. The correlation between link travel times can be different for different signal control approach. However, the proposed method is generic in that it does not depend on SCATS, and can be applied to any urban network implemented with any traffic signal control system.
5.4 Analysis of Travel Time Distribution based on Real Traffic Data

In order to have a better view of travel time variability, this section analyzes and discusses in depth the travel time distribution for all the selected routes in the selected fragment of the Melbourne network. Some travel time reliability studies assume that the travel time distribution might follow the normal distribution [55, 69, 70, 99, 111], while others insist that the log-normal distribution and gamma distribution can provide a better fit [50, 60, 113]. Here we use the goodness-of-fit test to check the validity of the assumptions regarding the specific travel time distribution, e.g., normal, log-normal or gamma. We compare normal, log-normal and gamma distributions with the exact travel time distribution based on estimated travel time obtained from SCATS data in the Melbourne traffic network. In this thesis, the Kullback-Leibler Divergence (KLD) test and the Chi-square (CS) test are selected to do the goodness of fit test. The KLD test is a non-symmetric measure of the difference between two probability distributions and gives the values of difference between them. The smaller of its value, the better fit are these two distributions. To the best of our knowledge, no study has been conducted to compare distributions using the KLD method by looking into the whole distribution. It is part of this thesis’ contribution to study travel time distribution by using the KLD test. In addition, the CS test is widely used to estimate how closely an observed distribution matches an expected distribution. An attractive feature of the CS goodness-of-fit test is that it can be applied to any distributions including discrete distributions. On the other hand, the Percent Relative Difference (PRD) test was performed in Gao [45]. This method was utilized to compare the shortest expected travel time results between proposed algorithm and approximation results. It is an indicator of how far away the approximation is from the real value. It is also appropriate to do comparisons between two travel time distributions and show the corresponding percent relative difference values. Here we also choose the PRD test to analyze travel time distribution in the selected Melbourne network.
For two discrete probability distributions, \( P = \{P_1, P_2, \ldots, P_n\} \) and \( Q = \{Q_1, Q_2, \ldots, Q_n\} \), where \( P_i, Q_i, i \in 1, 2, \ldots, n \) is the possible discrete value in the distributions \( P \) and \( Q \), and \( n \) is the total number of discrete values. The KLD of \( Q \) from \( P \) is defined to be:

\[
KLD(P, Q) = \sum_{i=1}^{n} ln\left(\frac{P_i}{Q_i}\right) \times P_i
\]  

(5.7)

Here, the Kullback-Leibler Divergence is the expectation of the logarithmic difference between the probabilities \( P \) and \( Q \), where the expectation is taken using the probabilities \( P \). The above formula (Equation 5.7) is similar to the entropy definition. Entropy is a measure of the uncertainty in a random variable, which is widely used in various fields. However, the KLD definition has nothing to do with entropy.

Chi-square test is often used to determine the goodness-of-fit between observed data and theoretical (expected) data. Observed data are the exact route travel time distribution estimated from SCATS data in the Melbourne traffic network. The theoretical data are developed on the hypothesis of normal, log-normal and gamma distributions. When the test sample is large number, the chi-square goodness-of-fit test is applied to binned data (i.e., data put into classes). The method to choose a bin is that assigning test samples to bins that naturally appear such as groupings or important thresholds. Here, in this test, the mean value of all route possible travel time probabilities is chosen as the threshold. Then the test data are divided into 2 bins, named bin 1 and bin 2, respectively. Specifically, the probability of route travel time value which is larger than mean value is noted as bin 1, and the counterpart is noted as bin 2. A contingency table is created as follows (see Table 5.2). Here, \( a \) and \( c \) are the number of events in bin 1 for the observed data and theoretical data, respectively. \( b \) and \( d \) are number of events in bin 2 for the corresponding group. Accordingly, the expected event numbers for bin 1 and bin 2 is as follows (see Table 5.3):

For the chi-square goodness-of-fit computation, the data are divided into 2 bins and
Table 5.2: Contingency table for observed values and theoretical values

<table>
<thead>
<tr>
<th>Variable</th>
<th>Bin 1</th>
<th>Bin 2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed data</td>
<td>a</td>
<td>b</td>
<td>a+b</td>
</tr>
<tr>
<td>Theoretical data</td>
<td>c</td>
<td>d</td>
<td>c+d</td>
</tr>
<tr>
<td>Total</td>
<td>a+c</td>
<td>b+d</td>
<td>a+b+c+d</td>
</tr>
</tbody>
</table>

Table 5.3: Expected event numbers for observed values and theoretical values

<table>
<thead>
<tr>
<th>Variable</th>
<th>Bin 1</th>
<th>Bin 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed data</td>
<td>((a+b)(a+c)/(a+b+c+d))</td>
<td>((a+b)(b+d)/(a+b+c+d))</td>
</tr>
<tr>
<td></td>
<td>((c+d)(a+c)/(a+b+c+d))</td>
<td>((c+d)(b+d)/(a+b+c+d))</td>
</tr>
<tr>
<td>Theoretical data</td>
<td>(a+b+c+d)</td>
<td>(a+b+c+d)</td>
</tr>
</tbody>
</table>

The test statistic is defined as

\[
\chi^2 = \sum_{i=1}^{2} \frac{(O_i - E_i)^2}{E_i}
\]  

(5.8)

where \(O_i\) is the observed number of events for bin \(i\) and \(E_i\) is expected number of events for bin \(i\).

The chi-square test is defined for the hypothesis, the null and the alternative hypotheses are:

- \(H_0\): the data follow the specified distribution.
- \(H_a\): the data do not follow the specified distribution.

Therefore, the hypothesis that the data are from a population with the specified distribution is rejected if

\[
\chi^2 > \chi^2_{1-a,k-1}
\]  

(5.9)

where \(\chi^2_{1-a,k-1}\) is the chi-square critical value with \(k-1\) degrees of freedom and significance level \(\alpha\). \(k\) is the number of bins. Usually, a fixed value of (i.e., 0.01,
Table 5.4: The Kullback-Leibler Divergence test for the actual and transformed travel time distribution (start time 6am)

<table>
<thead>
<tr>
<th>Route</th>
<th>KLD(N, A)</th>
<th>KLD(LN, A)</th>
<th>KLD(G, A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Route 1</td>
<td>0.4397</td>
<td>0.4308</td>
<td>0.4337</td>
</tr>
<tr>
<td>Route 2</td>
<td>0.3689</td>
<td>0.3648</td>
<td>0.3660</td>
</tr>
<tr>
<td>Route 3</td>
<td>0.5332</td>
<td>0.5329</td>
<td>0.5330</td>
</tr>
<tr>
<td>Route 4</td>
<td>0.3919</td>
<td>0.3920</td>
<td>0.3918</td>
</tr>
<tr>
<td>Route 5</td>
<td>0.3691</td>
<td>0.3671</td>
<td>0.3675</td>
</tr>
<tr>
<td>Route 6</td>
<td>0.5579</td>
<td>0.5521</td>
<td>0.5539</td>
</tr>
<tr>
<td>Route 7</td>
<td>0.4760</td>
<td>0.6842</td>
<td>0.6803</td>
</tr>
<tr>
<td>Route 8</td>
<td>0.4759</td>
<td>0.4740</td>
<td>0.4745</td>
</tr>
<tr>
<td>Route 9</td>
<td>0.5199</td>
<td>0.5177</td>
<td>0.5180</td>
</tr>
<tr>
<td>Route 10</td>
<td>0.4603</td>
<td>0.4731</td>
<td>0.4686</td>
</tr>
<tr>
<td>Route 11</td>
<td>0.4515</td>
<td>0.4623</td>
<td>0.4585</td>
</tr>
<tr>
<td>Route 12</td>
<td>0.4930</td>
<td>0.5015</td>
<td>0.4988</td>
</tr>
<tr>
<td>Route 13</td>
<td>0.4135</td>
<td>0.4112</td>
<td>0.4115</td>
</tr>
<tr>
<td>Route 14</td>
<td>0.4777</td>
<td>0.4841</td>
<td>0.4818</td>
</tr>
<tr>
<td>Route 15</td>
<td>0.5787</td>
<td>0.5577</td>
<td>0.5645</td>
</tr>
<tr>
<td>Route 16</td>
<td>0.5652</td>
<td>0.5312</td>
<td>0.5427</td>
</tr>
<tr>
<td>Route 17</td>
<td>0.6470</td>
<td>0.6294</td>
<td>0.6348</td>
</tr>
<tr>
<td>Route 18</td>
<td>0.5854</td>
<td>0.5493</td>
<td>0.5614</td>
</tr>
</tbody>
</table>

0.05, 0.1) are used to evaluate the null hypothesis $H_0$ at various significance levels.

In order to show which selected distribution (normal, log-normal and gamma distributions) best fit for our actual route travel time distribution, the smallest chi-square test statistic value is selected in each possible route instead of using a fixed $\alpha$ value. If the computed test statistic is small, then the observed and expected values are close and the model is a good fit to the data. The reason is that if the computed test statistic is small, then it is a low probability that a chi-square statistic falls between zero and test statistic value. Therefore the specific distribution under this small test statistic value is accepted while other possible travel time distributions will be rejected.

The definition of the Percent Relative Difference [45] between two probabilities $P$ and $Q$ is given as follows:
Table 5.5: The Kullback-Leibler Divergence test for the actual and transformed travel time distribution (start time 8:30am)

<table>
<thead>
<tr>
<th>Route</th>
<th>KLD(N, A)</th>
<th>KLD(LN, A)</th>
<th>KLD(G, A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Route 1</td>
<td>0.7616</td>
<td>0.7140</td>
<td>0.7303</td>
</tr>
<tr>
<td>Route 2</td>
<td>0.6824</td>
<td>0.6426</td>
<td>0.6561</td>
</tr>
<tr>
<td>Route 3</td>
<td>0.4244</td>
<td>0.4202</td>
<td>0.4214</td>
</tr>
<tr>
<td>Route 4</td>
<td>0.4915</td>
<td>0.3753</td>
<td>0.3748</td>
</tr>
<tr>
<td>Route 5</td>
<td>0.3691</td>
<td>0.3678</td>
<td>0.3678</td>
</tr>
<tr>
<td>Route 6</td>
<td>0.3385</td>
<td>0.3381</td>
<td>0.3381</td>
</tr>
<tr>
<td>Route 7</td>
<td>0.3799</td>
<td>0.3782</td>
<td>0.3787</td>
</tr>
<tr>
<td>Route 8</td>
<td>0.4289</td>
<td>0.4299</td>
<td>0.4294</td>
</tr>
<tr>
<td>Route 9</td>
<td>0.6951</td>
<td>0.6660</td>
<td>0.6754</td>
</tr>
<tr>
<td>Route 10</td>
<td>0.6934</td>
<td>0.6574</td>
<td>0.6692</td>
</tr>
<tr>
<td>Route 11</td>
<td>0.6909</td>
<td>0.6599</td>
<td>0.6699</td>
</tr>
<tr>
<td>Route 12</td>
<td>0.7833</td>
<td>0.7480</td>
<td>0.7596</td>
</tr>
<tr>
<td>Route 13</td>
<td>0.6162</td>
<td>0.5915</td>
<td>0.5998</td>
</tr>
<tr>
<td>Route 14</td>
<td>0.7322</td>
<td>0.7010</td>
<td>0.7116</td>
</tr>
<tr>
<td>Route 15</td>
<td>0.3923</td>
<td>0.3923</td>
<td>0.3923</td>
</tr>
<tr>
<td>Route 16</td>
<td>0.4454</td>
<td>0.4472</td>
<td>0.4466</td>
</tr>
<tr>
<td>Route 17</td>
<td>0.4486</td>
<td>0.4487</td>
<td>0.4486</td>
</tr>
<tr>
<td>Route 18</td>
<td>0.5016</td>
<td>0.5036</td>
<td>0.5029</td>
</tr>
</tbody>
</table>

\[
PRD(P, Q) = \sqrt{\frac{\sum_{i=1}^{n} (P_i - Q_i)^2}{\sum_{i=1}^{n} P_i^2}} \tag{5.10}
\]

Accordingly, based on the definition of the percent relative difference, the smaller the relative difference value, the better the fit of these two distributions. The test network is shown in Figure 5.1. We carry out two separate experiments for two different settings of start times: 6am and 8:30am represent for morning off-peak time and peak time. There are eighteen possible routes from the start point to the destination and we select all of them to do the comparison. For each route, the discrete travel time value and corresponding probabilities can be obtained through the estimated travel time data. Three statistical distributions, normal, log-normal and gamma are
Table 5.6: The Chi-square test for the actual and transformed travel time distribution (start time 6am)

<table>
<thead>
<tr>
<th>Route</th>
<th>$\chi^2_{N,A}$</th>
<th>$\chi^2_{LN,A}$</th>
<th>$\chi^2_{G,A}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Route 1</td>
<td>7.5481</td>
<td>5.5949</td>
<td>6.2115</td>
</tr>
<tr>
<td>Route 2</td>
<td>5.0529</td>
<td>5.6430</td>
<td>5.6430</td>
</tr>
<tr>
<td>Route 3</td>
<td>3.6795</td>
<td>3.6795</td>
<td>4.1954</td>
</tr>
<tr>
<td>Route 4</td>
<td>4.4019</td>
<td>3.8911</td>
<td>4.4019</td>
</tr>
<tr>
<td>Route 5</td>
<td>3.3613</td>
<td>3.3613</td>
<td>2.9228</td>
</tr>
<tr>
<td>Route 6</td>
<td>5.7671</td>
<td>2.6568</td>
<td>4.0546</td>
</tr>
<tr>
<td>Route 7</td>
<td>1.7877</td>
<td>1.1596</td>
<td>1.1596</td>
</tr>
<tr>
<td>Route 8</td>
<td>0.7267</td>
<td>0.5326</td>
<td>0.7267</td>
</tr>
<tr>
<td>Route 9</td>
<td>1.2525</td>
<td>1.2525</td>
<td>1.2525</td>
</tr>
<tr>
<td>Route 10</td>
<td>3.9924</td>
<td>3.0436</td>
<td>3.5012</td>
</tr>
<tr>
<td>Route 11</td>
<td>3.0567</td>
<td>5.1039</td>
<td>4.0115</td>
</tr>
<tr>
<td>Route 12</td>
<td>3.9267</td>
<td>2.5741</td>
<td>2.9920</td>
</tr>
<tr>
<td>Route 13</td>
<td>6.1427</td>
<td>4.9455</td>
<td>6.1427</td>
</tr>
<tr>
<td>Route 14</td>
<td>5.0613</td>
<td>5.6505</td>
<td>5.6505</td>
</tr>
<tr>
<td>Route 15</td>
<td>2.2550</td>
<td>1.8776</td>
<td>2.2550</td>
</tr>
<tr>
<td>Route 16</td>
<td>4.1652</td>
<td>5.8702</td>
<td>5.8702</td>
</tr>
<tr>
<td>Route 17</td>
<td>1.9069</td>
<td>2.2893</td>
<td>1.9069</td>
</tr>
<tr>
<td>Route 18</td>
<td>7.3464</td>
<td>6.0404</td>
<td>6.0404</td>
</tr>
</tbody>
</table>

generated using the corresponding mean route travel time and variance values. Then the statistical distributions will be discreted to get the corresponding probability values for each possible route travel time. Finally, the probability values in estimated travel time and calculated from statistical distributions will be compared using Kullback-Leibler Divergence, Chi-square test and Percent Relative Difference test based on Equs (5.7-5.10). Kullback-Leibler Divergence test results, relative difference values, are presented in Table 5.4 and 5.5 for different start time, where KLD(N, A), KLD(LN, A) and KLD(G, A) are the differences between normal distribution and actual data in the Melbourne traffic network, log-normal distribution and actual data and gamma distribution and actual data respectively. Table 5.6 and 5.7 show Chi-square test results for two start times, where $\chi^2_{N,A}$, $\chi^2_{LN,A}$ and $\chi^2_{G,A}$ are the test statistic between normal distribution and actual data in the Melbourne traffic network, log-normal distribution and actual data, gamma distribution and actual data respectively, and Table 5.8 and
Table 5.7: The Chi-square test for the actual and transformed travel time distribution (start time 8:30am)

<table>
<thead>
<tr>
<th>Route</th>
<th>$\chi^2_{N,A}$</th>
<th>$\chi^2_{LN,A}$</th>
<th>$\chi^2_{G,A}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Route 1</td>
<td>7.6579</td>
<td>6.2727</td>
<td>6.2727</td>
</tr>
<tr>
<td>Route 2</td>
<td>6.0724</td>
<td>4.3414</td>
<td>4.3414</td>
</tr>
<tr>
<td>Route 3</td>
<td>1.8535</td>
<td>2.2105</td>
<td>2.2105</td>
</tr>
<tr>
<td>Route 4</td>
<td>0.5411</td>
<td>0.5411</td>
<td>0.5411</td>
</tr>
<tr>
<td>Route 5</td>
<td>4.7458</td>
<td>4.7458</td>
<td>5.3002</td>
</tr>
<tr>
<td>Route 6</td>
<td>2.1714</td>
<td>1.8233</td>
<td>2.1714</td>
</tr>
<tr>
<td>Route 7</td>
<td>1.4560</td>
<td>1.7878</td>
<td>1.7878</td>
</tr>
<tr>
<td>Route 8</td>
<td>2.6298</td>
<td>3.0567</td>
<td>3.0567</td>
</tr>
<tr>
<td>Route 9</td>
<td>3.5591</td>
<td>4.0659</td>
<td>3.5591</td>
</tr>
<tr>
<td>Route 10</td>
<td>6.8270</td>
<td>6.1728</td>
<td>6.1728</td>
</tr>
<tr>
<td>Route 11</td>
<td>5.7104</td>
<td>6.3529</td>
<td>5.7104</td>
</tr>
<tr>
<td>Route 12</td>
<td>6.4103</td>
<td>5.1605</td>
<td>6.4103</td>
</tr>
<tr>
<td>Route 13</td>
<td>5.9893</td>
<td>4.8214</td>
<td>4.8214</td>
</tr>
<tr>
<td>Route 14</td>
<td>7.0071</td>
<td>5.6927</td>
<td>6.3312</td>
</tr>
<tr>
<td>Route 15</td>
<td>8.6090</td>
<td>9.3537</td>
<td>9.3537</td>
</tr>
<tr>
<td>Route 16</td>
<td>1.6026</td>
<td>1.2940</td>
<td>1.6026</td>
</tr>
<tr>
<td>Route 17</td>
<td>1.8182</td>
<td>1.4829</td>
<td>1.8182</td>
</tr>
<tr>
<td>Route 18</td>
<td>3.9924</td>
<td>3.5012</td>
<td>4.5175</td>
</tr>
</tbody>
</table>

5.9 summarize the Percent Relative Difference test results respectively, where PRD(N, A), PRD(LN, A) and PRD(G, A) are the difference between normal distribution and actual data in the Melbourne traffic network, log-normal distribution and actual data, gamma distribution and actual data respectively. As the route travel time is not a constant value but a random variable, it will not be included in the above table. In the literature, congestion index is usually utilized to measure congestion levels in urban areas and estimate the excess time lost in traffic congestion during peak travel periods, such as volume per capacity or free speed over actual speed. In this experiment part, there are eighteen possible routes from the start point to the destination and we select all of them to do the goodness-of-fit test. We consider all possible route travel time and corresponding probability values, which is not related to congestion index.

Based on real data analysis in the Melbourne network, the above result tables
Table 5.8: The Percent Relative Difference test for the actual and transformed travel time distribution (start time 6am)

<table>
<thead>
<tr>
<th>Route</th>
<th>PRD(N, A)</th>
<th>PRD(LN, A)</th>
<th>PRD(G, A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Route 1</td>
<td>0.8758</td>
<td>0.8660</td>
<td>0.8691</td>
</tr>
<tr>
<td>Route 2</td>
<td>0.7805</td>
<td>0.7726</td>
<td>0.7751</td>
</tr>
<tr>
<td>Route 3</td>
<td>0.9538</td>
<td>0.9527</td>
<td>0.9533</td>
</tr>
<tr>
<td>Route 4</td>
<td>0.8278</td>
<td>0.8293</td>
<td>0.8286</td>
</tr>
<tr>
<td>Route 5</td>
<td>0.8427</td>
<td>0.8432</td>
<td>0.8427</td>
</tr>
<tr>
<td>Route 6</td>
<td>0.9638</td>
<td>0.9601</td>
<td>0.9611</td>
</tr>
<tr>
<td>Route 7</td>
<td>1.2845</td>
<td>1.3088</td>
<td>1.3012</td>
</tr>
<tr>
<td>Route 8</td>
<td>0.9137</td>
<td>0.9124</td>
<td>0.9125</td>
</tr>
<tr>
<td>Route 9</td>
<td>0.9420</td>
<td>0.9416</td>
<td>0.9424</td>
</tr>
<tr>
<td>Route 10</td>
<td>0.9076</td>
<td>0.9241</td>
<td>0.9184</td>
</tr>
<tr>
<td>Route 11</td>
<td>0.8840</td>
<td>0.8939</td>
<td>0.8904</td>
</tr>
<tr>
<td>Route 12</td>
<td>0.9776</td>
<td>0.9936</td>
<td>0.9889</td>
</tr>
<tr>
<td>Route 13</td>
<td>0.9086</td>
<td>0.9118</td>
<td>0.9101</td>
</tr>
<tr>
<td>Route 14</td>
<td>0.9334</td>
<td>0.9399</td>
<td>0.9376</td>
</tr>
<tr>
<td>Route 15</td>
<td>1.0225</td>
<td>1.0084</td>
<td>1.0125</td>
</tr>
<tr>
<td>Route 16</td>
<td>0.9928</td>
<td>0.9521</td>
<td>0.9659</td>
</tr>
<tr>
<td>Route 17</td>
<td>1.0539</td>
<td>1.0452</td>
<td>1.0473</td>
</tr>
<tr>
<td>Route 18</td>
<td>1.0760</td>
<td>1.0244</td>
<td>1.0418</td>
</tr>
</tbody>
</table>

from Kullback-Leibler Divergence, Chi-square and Percent Relative Difference tests clearly suggest that the travel time distribution follows the log-normal distribution for most routes because of a smaller difference value. For the remaining routes, the normal distribution fits the observed travel time data. Then we can conclude that travel times in the Melbourne network best follow a log-normal distribution within our selected time period. Travel time is symmetric about the mean travel time value in normal distribution. In practice, there are some extreme large travel time values due to the randomness and uncertainty of the traffic network. For instance, in the morning peak hour, the behaviour of traffic dominated by a particular direction. However, normal distribution cannot capture these extreme values and hence it is not suitable to represent travel time distribution. Nevertheless, log-normal distribution can capture this large travel time value because of the skewness of its distribution. Moreover, the log-normal distribution has a similar shape of the gamma distribution [26] with
Table 5.9: The Percent Relative Difference test for the actual and transformed travel time distribution (start time 8:30am)

<table>
<thead>
<tr>
<th>Route</th>
<th>PRD(N, A)</th>
<th>PRD(LN, A)</th>
<th>PRD(G, A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Route 1</td>
<td>1.2655</td>
<td>1.2035</td>
<td>1.2255</td>
</tr>
<tr>
<td>Route 2</td>
<td>1.1461</td>
<td>1.0950</td>
<td>1.1124</td>
</tr>
<tr>
<td>Route 3</td>
<td>0.8802</td>
<td>0.8742</td>
<td>0.8758</td>
</tr>
<tr>
<td>Route 4</td>
<td>0.8331</td>
<td>0.8389</td>
<td>0.8369</td>
</tr>
<tr>
<td>Route 5</td>
<td>0.8456</td>
<td>0.8479</td>
<td>0.8465</td>
</tr>
<tr>
<td>Route 6</td>
<td>0.7273</td>
<td>0.7277</td>
<td>0.7275</td>
</tr>
<tr>
<td>Route 7</td>
<td>0.8042</td>
<td>0.8027</td>
<td>0.8031</td>
</tr>
<tr>
<td>Route 8</td>
<td>0.8534</td>
<td>0.8549</td>
<td>0.8542</td>
</tr>
<tr>
<td>Route 9</td>
<td>1.1427</td>
<td>1.1105</td>
<td>1.1208</td>
</tr>
<tr>
<td>Route 10</td>
<td>1.1538</td>
<td>1.1128</td>
<td>1.1260</td>
</tr>
<tr>
<td>Route 11</td>
<td>1.1713</td>
<td>1.1408</td>
<td>1.1507</td>
</tr>
<tr>
<td>Route 12</td>
<td>1.2916</td>
<td>1.2439</td>
<td>1.2591</td>
</tr>
<tr>
<td>Route 13</td>
<td>1.0742</td>
<td>1.0488</td>
<td>1.0572</td>
</tr>
<tr>
<td>Route 14</td>
<td>1.1544</td>
<td>1.1207</td>
<td>1.1321</td>
</tr>
<tr>
<td>Route 15</td>
<td>0.7955</td>
<td>0.7958</td>
<td>0.7956</td>
</tr>
<tr>
<td>Route 16</td>
<td>0.8648</td>
<td>0.8657</td>
<td>0.8654</td>
</tr>
<tr>
<td>Route 17</td>
<td>0.8819</td>
<td>0.8823</td>
<td>0.8822</td>
</tr>
<tr>
<td>Route 18</td>
<td>0.9033</td>
<td>0.9046</td>
<td>0.9042</td>
</tr>
</tbody>
</table>

skewness. When compared with the gamma distribution, the log-normal travel time distribution provides a closer fit over the Melbourne traffic networks.

### 5.5 Numerical Tests

This section describes numerical experiments using Matlab to evaluate the proposed frameworks in Chapter 3 and Chapter 4, in order to demonstrate their benefits in solving routing problems in a real environment. The test network and link travel time distributions described in the previous sections are employed. Then, experimental results are discussed in detail.
5.5.1 Test results for the shortest path

In this section, a computational test on a real traffic network in Melbourne (Figure 5.1) is utilized to study the effectiveness of the shortest route selection strategies. Assume that node A is the start point and node B is the destination. The joint distribution of all link travel time random variables is obtained through the travel time estimation method introduced above. In this example, we compare our proposed method, Miller-Hooks and Mahmassani’s [81], and Fan et al.’s [36] shortest path methods. Here we do not compare with Gao and Chabini’s [46] method due to the lack of full joint travel time distribution. We carry out sixteen separate experiments for sixteen different start times. We start at 6am and progress at 15 minute intervals. In order to evaluate the quality of the solutions produced by different methods, we compare the route choices computed by the three methods mentioned above against the minimal travel time route. The final results are then expressed as the mean of the differences for all possible travel time values.

The comparison of the travel times for the route choices produced by different methods (our proposed method, Miller-Hooks and Mahmassani’s method, and Fan et al.’s method, respectively) against the minimal travel time for the optimal route from start node A to the destination node B (Figure 5.1) is shown as an error bar in Figure 5.4 under different start times. The figure shows the mean differences as a bar for each method over all possible travel time values. The corresponding maximum and minimum difference values are depicted by a line. The x-axis represents each possible start time (e.g., 615 stands for 6:15am) and the y-axis stands for mean difference calculated.

As can be seen in Figure 5.4, our proposed method performs very well in these experiments, with the smallest mean differences. Miller-Hooks and Mahmassani’s method has a higher mean difference value than ours because of an independent link travel times assumption. Our proposed method performs quite well when compared to the method introduced by Fan et al. [36]. 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 fixed route regardless of the different start times. However, our proposed method can adapt to the changing states of the network.

Here, in order to show the effectiveness of our proposed method, we also compare with the route selection in static traffic network. In this case, link travel time is a constant value. The comparison of the travel times for the different route choices against the minimal travel time is also shown as an error bar in Figure 5.5 under different start times. As it is shown in Figure 5.5, our proposed method performs very
well in these experiments, with smaller mean differences in most cases. Then, we can conclude our proposed method can adapt to the changing states of the network and show advantages compared to static routing.

In summary, the results in these experiments are fairly consistent with the simulation results in the previous chapter. In most cases, our method shows smaller mean differences than Fan et al.’s method. This indicates that considering time-dependent travel time yields significant advantages over methods that ignore this aspect. Furthermore, our method shows smaller mean differences than Miller-Hooks and Mahmassani’s method suggesting that considering spatial correlation between adjacent links yields significant advantages over methods that neglect this aspect. It’s worth noting also that, in general, most of the mean difference values for our proposed method are relatively small in most time intervals (typically less than 2%). Finally, the real traffic network evaluation results show that the proposed approach is effective and can provide a good approximation method for computing the shortest path in stochastic time-dependent networks by taking advantage of the information on link travel time correlations.

Our proposed algorithms run in real-time and suggest the best route for travellers to their destination. Therefore, computation time will affect the performance of the proposed algorithm and it is an important issue to be considered. However, it requires a lot of time to setup the experiment and investigate the performance of proposed algorithm in real application. Here we leave it for future work.

5.5.2 Test results for reliable route guidance

This section will focus on the performance measurement of selected routes in the Melbourne road networks by utilizing travel time reliability measures as reported in the previous studies. Several definitions for travel time reliability have been proposed in the literature. In this thesis, we investigate the performance of five travel time
reliability methods as traffic performance measures based on real travel time data in the Melbourne road networks.

Numerous indexes have been proposed for assessing travel time reliability, or travel time variance. Here we select five of them for comparison, including:

1. Van Lint and Van Zuylen [66] propose a reliability metric for the standardized width of the travel time distribution using percentile travel times (Method 1). Generally, the larger of travel time width, the higher probability for extreme travel time (compared to mean travel time) to occur.

\[ \text{Method 1 Reliability} = \frac{P90 - P10}{P50} \]

where P90, P50 and P10 denote the 90th, 50th and 10th percentile travel times.

2. Kaparias et al. [62] introduce lateness as a travel time reliability definition that considers the maximum travel time that may be experienced on a link or route (Method 2). In reality, people try to avoid being late, and this indices reflects the longest total travel time may be experienced by travellers. The definition of lateness is expressed as follows:

\[ \text{Method 2 Reliability} = \frac{\text{Average Travel Time}}{\text{Maximum Travel Time}} \]

3. Coefficient of variation calculations through Rakha et al.’s [99] method to get path travel time variance from individual links (Method 3). Here the travel time variance of a route is computed using the expected coefficient of variation as the conditional expectation over all realizations of the various links that make up the route. Let H be a set of links that constitutes route R between decision node i and the destination (link number is n), for all composing link \((i, j) \in A, \forall i, j \in N\).
Method 3 Reliability = \frac{\sqrt{\text{Var}[T_R]}}{E[T_R]}

\text{Var}[T_R] = \frac{(E[T_R])^2}{n^2} \left( \sum_{(i,j) \in H} \frac{\text{Var}[T_{ij}]}{E[T_{ij}]} \right)^2

(4). The Florida reliability method considers the percentage of trips that are completed within a designated travel time threshold \[91\] (Method 4). Here, the travel time threshold is calculated as the median travel time plus a percentage (5%) of median travel time. From the mathematical point of view, this definition can be described as the following equation:

Method 4 Reliability = \text{Pr}\{\text{Path Travel Time} \leq \text{Threshold}\}

(5). The buffer time index \[70\], defined as the extra time that travellers must add to their average travel time when planning trips to ensure on-time arrival (Method 5). Mathematical presentation regarding buffer time and buffer time index can be expressed as follows:

Method 5 Reliability = \frac{\text{Buffer Time}}{\text{Average Travel Time}} \times 100\%

\text{Buffer Time} = 95\text{th Percentile Travel Time for a Trip} - \text{Average Travel Time}

In the previous chapter, we have proposed STA approximation method to obtain the reliability of a route travel time (note this Method 6). Now we compare all six travel time reliability methods’ results with the reliability value calculated from the
full travel time distribution.

The test area starts in Maidstone (node A in Figure 5.1) and ends in the Hawthorn area (node B in Figure 5.1). The starting time at node A is 6am. There are eighteen routes from the start point to the destination. Firstly, we calculate the travel time reliability values of the methods mentioned above for all the routes. Then we select each possible route travel time reliability value and compare this with all the other route travel time reliability values, between all six methods and the method from full distribution of travel times. Finally, we count the total number of differences for all possible routes when the comparison for the same route between two methods shows different trend. Therefore, when the total difference value is small, the tested travel time reliability method performs well and closes to the reliable route calculation using full travel time distribution. The final result is shown in Figure 5.6. The y-axis stands for total difference numbers calculated. From Figure 5.6, it can be seen that Method 2 has the best performance with the smallest difference (the red column chart is zero). The performance of other methods can be evaluated by the total difference in numbers that is shown in the figure.

Later, we repeat the test by changing starting times to 7am, 8am and 9am. The results are shown in Figures 5.7-5.9. In order to get a clear performance comparison, we get the average difference numbers among all the six methods for four different start times and show the final results in Figure 5.10. Overall, the comparison shows that Method 2 is ranked as the best one among all the comparison methods (average difference value is zero). Our proposed Method 6 performs better than Method 1 as it has a smaller total difference in numbers. Method 4 and Method 5 are the popular travel time reliability methods in the literature and also show the same results in the Melbourne road networks. However, Method 3 ranks the lowest due to the largest number difference, indicating that the Rakha et al.’s reliability approximation method does not perform well. The performance rank from high to low is Method 2, Method 6, Method 1, Method 4, Method 5 and Method 3 respectively.
Figure 5.6: Comparison between various travel time reliability measures at 6am (Kaparias et al.’s (62) method has zero total difference)

Figure 5.7: Comparison between various travel time reliability measures at 7am (Kaparias et al.’s (62) method has zero total difference)
Chapter 5

Figure 5.8: Comparison between various travel time reliability measures at 8am (Kaparias et al.’s (62) method has zero total difference)

Figure 5.9: Comparison between various travel time reliability measures at 9am (Kaparias et al.’s (62) method has zero total difference)
In the previous chapter, our proposed travel time reliability method shows good results in the simulation network. In the real traffic network test, the performance is consistent with the expected advantages of our proposed method. In this experiment, Kaparias et al.’s method [62] performs the best given a full path travel time distribution, which has zero difference value regardless of start time. In their approach, Rakha et al.’s method [99] is utilized to get path travel time variance value from individual links if link travel time distributions are available. We can see through the evaluation results that Rakha et al.’s method [99] (Method 3) does not perform well. This is due to the underlying assumption that the travel times of different links are statistically independent from each other. Therefore, Kaparias et al.’s travel time reliability method [62] shows good results only when the available data are path travel time distribution. However, it is difficult to determine path travel time distribution directly in reality. In practice, individual link parameters are available and our proposed method works well in this case. In summary, the results show that our proposed approach is reasonably effective and can provide a good approximation for computing the reliable path for stochastic time-dependent networks by including link travel time.
correlations.

5.6 Summary of this Chapter

In order to test the accuracy of the previous proposed frameworks on the shortest route and reliable route guidance with real data in a real traffic network, in this chapter numerical experiments are carried out on a road network in part of the Melbourne area based on loop detector data and traffic signal timing parameters from the SCATS system. Some background information regarding vehicle detectors and the SCATS adaptive control system, acquisition and processing of the loop detector data are reviewed in detail. We first implement a detailed analysis of travel time distribution according to estimated travel time obtained from SCATS data. The Kullback-Leibler Divergence, Chi-square and Percent Relative Difference tests of goodness-of-fit are selected to verify travel time distribution. Based on the real data in the Melbourne traffic network, the travel time distribution best fits with log-normal distribution. Then, some experiments are carried out in Matlab in a part of the traffic network in the Melbourne area and the implementation results are presented and analyzed. For the least expected travel time route finding method, evaluation results show that our proposed approach is effective and can provide a good approximation method for computing shortest travel time path in stochastic time-dependent networks taking advantage of the information on link travel time correlations. Furthermore, five travel time reliability indexes proposed in the literature and our proposed STA approximation method for calculating travel time reliability are discussed and compared with the reliability value calculated from the full travel time distribution. The final comparison result indicates that our developed STA approximation method to obtain the reliability of a route travel time is reasonably effective and accurate.
Chapter 6

Conclusions and Future Work

This closing chapter summarizes briefly the main conclusions of the research work and findings presented in this thesis. We then conclude this chapter with suggested directions for future research.

6.1 Conclusions

The aim of this research is to propose a real-time route guidance system that provides accurate and reliable suggestions in a timely manner to help motorists navigate to their destination. It includes the evaluation of the proposed system in terms of suitable defined performance metrics and comparison with a number of existing methods in terms of route selection. Tests were conducted through illustrative and simulation examples. Further to this, the evaluation of the proposed approaches, using real data from part of the city network of Melbourne, Australia, was also performed. Some conclusions which are drawn based on the research carried in this thesis are presented in the following.

In the first part, a simple approximate framework has been developed for the
determination of the route with the least expected travel time from any node, to any
given destination, in a stochastic 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., the link congestion level) available at the time
the decision is made. A solution based on the principle of dynamic programming is
derived in which the least expected travel time path is gradually formed from the
real-time routing decisions made at each node along that path. Through numerical
examples it was shown that the optimal path depends strongly on spatial and temporal
correlations of travel times. Furthermore, through a comparison with existing work that
assumes full knowledge of the distribution of network travel times, it was shown that
the proposed framework can provide good results despite its utilization of a very small
set of network parameters.

Then we studied the issue of reliable route selection in a stochastic time-dependent
traffic network as the second part of this thesis. We proposed two approximation
methods for calculating travel time reliability, that takes account link travel time
uncertainty and spatial correlation between adjacent links. Here, the path travel time
variance value is calculated using individual link travel time variances and covariance
values between adjacent links, aiming to obtain the reliability of a route travel time
and find the corresponding most reliable route. An illustrative example and Sioux Falls
network are provided to demonstrate the calculation process of the proposed methods.
Through a comparison with the exact values calculated via the entire discrete joint
distribution of link travel times, we observe that the proposed method computes results
that are close to the exact value.

In order to demonstrate the performance of the proposed frameworks, and to judge
their advantages and disadvantages, the proposed approaches are evaluated using real
traffic data in real traffic networks in the last part of this thesis. The test network
is selected from part of the city network of Melbourne, Australia and the real data
is from loop detectors in the SCATS system. Estimated travel time obtained from
SCATS data in the Melbourne traffic network has been analyzed to study travel time distribution in a typical urban traffic network. The Kullback-Leibler Divergence, Chi-Square and Percent Relative Difference tests of goodness-of-fit are conducted to verify travel time distribution. Finally, test results show that travel time distribution best fits with log-normal distribution in the Melbourne network. The performance of our proposed methods are compared with a number of existing methods in terms of route selection. In summary, the evaluation results indicate that for the least expected travel time route finding issue, our proposed approach is effective and can provide a good approximation method for computing shortest travel time path in stochastic time-dependent networks taking advantage of the information on link travel time correlations. Furthermore, our proposed STA approximation method for calculating travel time reliability and five travel time reliability indexes proposed in the literature are compared with the reliability value calculated from the full travel time distribution. The final comparison result shows that our developed STA approximation method to obtain the reliability of a route travel time is reasonably effective and accurate. It provides a good approximation for computing the reliable path for stochastic time-dependent networks by including link travel time correlations.

6.2 Future Work

Some future work may be pursued following the line of research of this thesis.

Travel time reliability issues: Reliability is an important criterion affecting travellers’ route choice when making routing decisions in a stochastic network. The main parameter in the reliable route guidance model is the travel time reliability. In this thesis, we are particularly interested in coefficient of variation (CV), where average travel time and standard deviation values are combined together in a ratio to produce a value as a travel time reliability indicator. Two approximation methods have been proposed to calculate travel time reliability. The performance of the proposed
approximation methods have been tested through illustrative example and real traffic network evaluation. Finally, STA approximation method is the best fit based on estimated travel time obtained from SCATS data. A wide range of definitions have been proposed to measure travel time reliability in the literature, which is also important to study. For example, Shao et al. [109] define travel time reliability as a minimization problem that minimize travellers’ effective travel time (the summation of mean travel time and the extra time that a traveller allows for his (or her) trip) subject to a confidence level. Nie et al. [88] propose travel time reliability by maximizing the probability of arriving at the destination through any path departing from the start node with a time budget. Here we leave it for future work to investigate.

**New road searching criterion:** In this thesis, we have been focusing on the determination of minimum expected travel time route since expected travel time is the primary concern of travellers when making routing decisions. On the other hand, travel time reliability is found to be another important factor affecting travellers’ route choice when faced with uncertainty. It would be interesting to develop an algorithm that combining travel time and travel time reliability as a new optimization criterion for finding a route to be suggested to the traveller. Then its performance can be studied computationally by evaluation in simulations with hypothesized data and traffic networks with real data.

**Stochastic dynamic traffic assignment issues:** In this thesis, a general least expected travel time route finding framework in stochastic time-dependent traffic network has been proposed. The performance is tested through experiment evaluation. However, the same path chosen by a large number of travellers departing at the same time between the same OD pair will cause link conditions change. Therefore, as a second step, a stochastic dynamic traffic assignment model should be developed to study the network level impact of real-time information provision and routing decisions. This model captures the interaction between network supply and traffic demand, aiming to predict traffic network conditions when travellers make routing decisions.
A generalization of Waldrop’s Principle can be used as the equilibrium condition and a computational test can be carried out in a real traffic network.

**Practical implementation:** In this thesis, the proposed shortest and reliable route guidance frameworks have been tested using simulation and real traffic data in a laboratory environment. Travel time data were estimated based on SCATS data in Melbourne. It will be interesting to use real experienced travel time for further investigation. Moreover, as promising route searching methods, appropriate software should be developed to implement the proposed methods in reality. These can be run on on-board units or personal handheld devices such as Iphone, and Ipad. In particular, attention would be paid to efficiency and running time issues. It is a computational burden for on-board devices with limited computation power when repeating route searching. Some optimization methods can be developed to make the best use of previous shortest path searching results, ensuring a significant reduction of computation time. Furthermore, the level of market penetration is an important factor that governs the benefit of in-vehicle route guidance to the system, equipped and unequipped travellers. Therefore, an analytical model can be developed to evaluate the traffic system benefit achieved under the percentage of in-vehicle route guidance systems adoption.
Publications of the Author


Bibliography


[46] S. Gao and I. Chabini. Optimal routing policy problems in stochastic time-

stochastic time-dependent networks. *Transportation Research Record: Journal

213, 2011.


[50] G. Giuliano. Incident characteristics, frequency, and duration on a high volume


[52] R. W. Hall. The fastest path through a network with random time-dependent

[53] Y. Han. Improved algorithm for all pairs shortest paths. *Information Processing

[54] B. Hellinga, L. Fu, and S. Selkirk. Route selection considering travel time


[56] H. Huang. Value of traveler information for adaptive routing in stochastic time-


[74] W. K. Mak, F. Viti, S. P. Hoogendoorn, and A. Hegyi. Online travel time estimation in urban areas using the occupancy of long loop detectors. In


[121] Y. Wang, M. Papageorgiou, G. Sarros, and W. J. Knibbe. Real-time route guidance for large-scale express ring roads. In Proceedings of the 9th International...


