USDOT Region V Regional University Transportation Center Final Report

A Decision Support Tool for Vehicle Infrastructure Integration: Understanding Information Effects and Advancing Data Fusion Algorithms for Traffic Management Applications

By

Srinivas Peeta, Principal Investigator
Professor of Civil Engineering
Purdue University
peeta@purdue.edu

and

Lili Du
Postdoctoral Research Associate
Purdue University
ldu@purdue.edu

and

Yong Hoon Kim
Graduate Research Assistant
School of Civil Engineering
Purdue University
Kim523@purdue.edu

Report Submission Date: October 15, 2009

DISCLAIMER

USDOT Region V Regional University Transportation Center Final Report

TECHNICAL SUMMARY

A Decision Support Tool for Vehicle Infrastructure Integration: Understanding Information Effects and Advancing Data Fusion Algorithms for Traffic Management Applications

Introduction

This research seeks to explore vehicle-to-vehicle information networks to understand the interplay between the information communicated and traffic conditions on the network. A longer-term goal is to develop a decision support tool for processing and storage of large amount of real-time (probe) data for advancing the state of the art in Vehicle Infrastructure Integration (VII). The fundamental concept in VII is that the (probe) vehicles serve as data collectors and anonymously transmit traffic information to transportation agencies to facilitate proactive strategies for traffic management and safety. The project develops new data fusion algorithms for travel time estimation and online stochastic routing which provides a clear representation of the benefit of information exchange between vehicles in VANETs.

A Vehicular Ad Hoc Network (VANET) is composed of smart vehicles in which advanced wireless communication equipment are installed so that inter-vehicular communications are enabled. As a newly emerging paradigm of the Advanced Traveler Information Systems (ATIS), the VANET has received significant interest recently. There are many challenging research topics in this promising area, which can be grouped into two major categories. The first category includes the issues regarding the performance of the communication networks; for example, the network connectivity, the communication capacity, the broadcasting protocol, and so on. The second category comprises issues about traffic networks such as traffic data or information fusion, and online routing. This study focuses on building the information fusion framework to predict the short-term link travel time distribution with the real-time travel-time information provided by a VANET.

Findings

In this study, we model the short-term link travel time variability as stochastic time-dependent random variables with discrete distributions. In the context of VANETs, real-time travel time information is considered as a travel time interval with information quality represented by information accuracy and time delay. Based on this information framework, we proposed an adaptive learning process to refine the short-term travel time distribution step by step. The learning process employs the knowledge of long-term historical travel time distribution as the starting point and adaptively updates it using the obtained real-time travel time information. The realized short-term travel time distribution obtained in

the last iteration will be used as the historical travel time distribution in the next iteration when new real-time information is available. The core of the proposed learning procedure is the information fusion model formulated as a non-linear program, which mathematically integrates the historical travel time information and real-time travel time information incorporating the information quality. To evaluate the efficiency of the proposed methodologies, we establish numerical experiments and also demonstrate their applications to online stochastic routing in VANETs.

The study illustrates that the long-term historical travel time distribution for a traffic link cannot describe the short-term traffic distribution robustly. Motivated by this issue, the study develops an information fusion model that uses an iterative learning process to estimate the short-term travel time distirbution dynamically. The numerical expriments conducted in this research suggest that the proposed information fusion model has good accuracy, robustness, and adaptivity. Depending on the real-time information accuracy and time delay, the robustness of the information fusion model will be observable as the procedure progress. Parametric analysis indicates that the accuracy of real-time information substantially influences the performance of the proposed information fusion model.

Recommendations

The proposed study provides a methodology to fuse information from multiple sources, where the format of information is characterized by a discrete distribution and a confidence interval. Hence, it has the potential to be applied to application domains in which dynamic changes to the distribution need to be tracked and characterized. Specifically, based on the proposed information fusion model, several potential future research directions can be envisaged. First, the proposed information fusion models can be extended to fuse two distributions with information quality still incorporated; these include the historical travel time distribution and real-time travel time distribution. Second, the current experiments were conducted based on synthetic data generated using a simulator. It would be insightful to validate the proposed information fusion model based on field data. We expect that the generated short-term travel time estimation will have better accuracy since the real-time information now provides more comprehensive insights about the actual travel time conditions on the link. Third, this study does not differentiate the information by vehicle classes. Involving this factor as part of the information fusion models is another important extension that has significant practical utility.

Contacts
For more information:

Srinivas Peeta
Principal Investigator
Civil Engineering
Purdue University
peeta@purdue.edu

NEXTRANS Center
Purdue University - Discovery Park
2700 Kent B-100
West Lafayette, IN 47906

nextrans@purdue.edu
(765) 496-9729
(765) 807-3123 Fax
www.purdue.edu/dp/nextrans

USDOT Region V Regional University Transportation Center Final Report

A Decision Support Tool for Vehicle Infrastructure Integration: Understanding Information Effects and Advancing Data Fusion Algorithms for Traffic Management Applications

By

Srinivas Peeta, Principal Investigator
Professor of Civil Engineering
Purdue University
peeta@purdue.edu

and

Lili Du
Postdoctoral Research Associate
Purdue University
ldu@purdue.edu

and

Yong Hoon Kim
Graduate Research Assistant
School of Civil Engineering
Purdue University
Kim523@purdue.edu

Report Submission Date: October 15, 2009

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1. INTRODUCTION

1.1 *Background and motivation*

A Vehicular Ad Hoc Network (VANET) is composed of smart vehicles on which advanced wireless communication equipment are installed so that inter-vehicular communications are enabled. As a newly emerging paradigm of the Advanced Traveler Information Systems (ATIS), the VANET has received significant interest recently. There are many challenging research topics in this promising area, which can be grouped into two major categories. The first category includes the issues regarding the performance of the communication networks; for example, the network connectivity, the communication capacity, the broadcasting protocol, and so on. The second category comprises issues about traffic networks such as traffic data or information fusion, and online routing. This study focuses on building the information fusion framework to predict the short-term link travel time distribution with the real-time travel-time information provided by a VANET. To avoid confusion in terminology, we hereafter differentiate between data and information by defining information as processed data. From this point of view, the proposed framework is built upon information instead of data.

Travel time is a very important input for the Dynamic Route Guidance Systems which is one of the most promising applications of VANET. It allows travelers to choose times and routes before the trip as well as en-route in an informed manner. In a VANET, real-time traffic data can be collected from multiple sources such as individual vehicles or road side sensors and then disseminated to all other vehicles in the network through wireless communication. Based on the embedded data processing models, the real-time data further generates the real-time information which represents the realized real-time

traffic conditions. However, due to the dynamic characteristic of traffic flow which significantly impacts the data dissemination, propagation, and processing in VANET, the quality of the data collected by VANET is neither constant nor uniform (Ziliaskopoulos and Zhang, 2003; Jin and Recker, 2006; Artimy, et al., 2006; Ukkusuri and Du, 2008; Du and Ukkusuri, 2008(a,b)) among roads. Thereby, the quality of data collected at some time instance or from certain road links may be inferior. Due to these issues with real-time data available in VANET, recent studies such as Tam and Lam (2007), and Kothuri, et al. (2008), showed that many existing travel time estimation or prediction methods can produce results with only a limited degree of accuracy. For example, the field experiments of Kothuri, et al. (2008) reported that 15% of the estimated travel time data in their study exhibited errors over the FHWA-suggested error threshold of 20%. Hence, the quality of data affects the accuracy of information obtained from processing the data. An additional issue with communication networks is the likelihood of time delay in transmitting data. Hence, the dissemination of the data/information in VANETs can entail a time delay, which directly affects the ability of the real-time information in predicting the traffic conditions. In summary, the ability to leverage the real-time data in VANETs necessitates the design of an information fusion method which addresses the information quality issues. To the best of our knowledge, the proposed research represents the first effort to study the information fusion while factoring the aforementioned aspects of information quality.

Another important aspect that motivates this study is the need to predict the distribution of the travel time rather than the actual value of the travel time in a short-term context. In recent years, the increased deployment of advanced intelligent transportation infrastructure has enabled the provision of real-time traffic information to travelers. In this context, the reliability of the information has been widely recognized as a critical issue for any ATIS to be successfully deployed. Since travel time distribution is a key basis to analyze the level of information reliability (Rakha, et al., 2006; Hollander and Liu, 2008; Kaparias, et al., 2008; Dong and Mahmassani, 2009), there is a need for methodologies to estimate the short-term travel time distribution. However, most previous efforts have focused on the travel time variability in a long-term context; for

example rush/non-rush hours, daily or seasonal variability (Anantharam, 1998; Hollander and Liu, 2008). In addition, general normal or log-normal distributions have proven unsuitable to represent short-term travel time variability (He, et al., 2002). For these reasons, this study seeks to address the issue of short-term travel time distributions, so as to bridge gaps both in the research literature and practical applications.

1.2   *Problem statement*

The proposed study focuses on determining the short-term travel time distribution based on the differences in the travel times on the same link obtained at different times from different travelers. The reasons for such variability in a short time period are the arrival time, driver behavior, and demand fluctuations, which are too subtle to be predicted by travelers (Hollander and Liu, 2008). This study seeks to develop a mechanism which adaptively estimates the short-term travel time distribution based on both the historical travel time distribution and the real-time travel time information. More specifically, the link travel time distribution is initially extracted from long-term historical travel time information and then updated whenever new real-time travel time information is available. Information accuracy as well as time delay are incorporated in the information fusion procedure. Also, rather than addressing the information fusion on an individual link separately, the proposed study uniformly fuses information from multiple sources and estimates the short-term travel time distribution over the entire local network. As demonstrated in Chapter 7, the proposed framework can support the more reliable stochastic on-line routing in VANETs.

1.3   *Organization of the research*

The reminder of this research is organized as follows. Chapter 2 reviews the previous efforts on data fusion techniques used for short-term travel time estimation, and the travel time distribution modeling. Chapter 3 discusses the information network in the context of the proposed research. Chapter 4 presents the mathematical information fusion model to estimate the short-term travel time distribution. Chapter 5 introduces the test-bed and the synthetic data which are used to setup the numerical experiments. Chapter 6 evaluates the proposed information fusion models through numerical experiments. The

applicability of the information fusion models is demonstrated in Chapter 7 using an online stochastic routing application. Chapter 8 summarizes the research and its contributions, and proposes future research directions.

CHAPTER 2.   LITERATURE REVEIW

2.1   *Data fusion techniques*

The proposed study is in the domain of short-term travel time estimation (prediction) problems in the literature. Numerous methods have been developed by previous studies. For a comprehensive review, readers are referred to the three survey papers: Vlahogianni, et al. (2004), Lin, et al. (2005), and Krishnan and Polak (2008).

In the context of input data sources, most previous work used either indirect travel-time data such as volume, occupancy, and speed (Roden, 1996; D'Angelo, et al., 1999; Pant, et al., 1998; Wu, et al., 2004) or direct travel-time data from fixed sensors such as loop detectors (Kwon, et al., 2000; Jiang and Zhang 2003; Stathopoulos and Karlaftis, 2003). Recent advances in wireless communication techniques provide an ability to develop reliable data fusion algorithms for short-term travel time predictions from floating vehicle data (probe vehicle data) (Stathopoulos and Karlaftis, 2003; Chen and Chien, 2001; Yang, et al., 2004; Wunderlich, et al., 2000; (Li and cDonald, 2002). Previous studies suggest that floating car data has the advantages of providing real-time data, covering a broader area, and tracking traffic conditions continuously. However, it is also characterized by limitations such as skewed distributed in space and time (that is, dense observations in some time or space but sparse in others), and data uncertainties due to missing data or errors. Therefore, it requires specialized procedures to deal with (Turner, 1998).

In the context of prediction models, the existing data fusion methods can be categorized by parametric and non-parametric models (Vlahogianni, et al., 2004). Non-parametric models, such as artificial neural networks (ANN) and k-nearest neighbors have the ability to derive meaningful insights from complicated situations or imprecise

data (Smith, et al., 2002). Therefore, many research studies have used ANN techniques to predict travel time, such as Park, et al. (1999), Dia and Berkum (2001), Clark (2003), Bajwa, et al. (2005), van Lint (2006), and van Lint (2008) proposed a k-nearest neighbor algorithm for short-term travel time prediction. However, non-parametric models usually need both historical data and online data to obtain good performance. Further, ANN suffers from large computational load, over-fitting issues and generalization problems. Hence, non-parametric models have limitations in addressing the problem of interest (Tu, 1996).

In contrast to the non-parametric prediction models, parametric models are more suited for studying functional dependencies between factors (Smith, et al., 2002). Many regression models (linear, non-linear, local weighted regression model, etc.) have been applied to short-term travel time prediction in prior studies (Zhang and Rice, 2003; Wunderlich, Kaufman and Smith, 2000; Wu, et al., 2003; Sun, et al., 2003; Zhong, et al., 2005). Among them, Zhang and Rice (2003) and Wunderlich, et al. (2000) integrated data from vehicle probes and other data sources. In addition, Wunderlich, et al. (2000) embedded their prediction models into a decentralized route guidance architecture. Time series models, such as Auto-Regressive Integrated Moving Average (ARIMA) family of models were introduced into traffic forecasting by Ahmed and Cook (1979), and further developed by Oda (1990), Van Der Voort, et al. (1996), and J. Yang (2005). Among them, Yang J. (2005) used the traffic data collected by vehicle probes equipped with Global Positioning System (GPS). Kalman filter models support estimations of past, present, and even future states. They enable the continuous updating of the state variable as new observations become available (Chien and Kuchipudi, 2003). Several approaches based on Kalman filters were developed to predict dynamic freeway short-term travel time (Chen and Chien, 2001; Chien and Kuchipudi, 2003; Nanthawicit, et al., 2003; Yang, et al., 2004; van Lint, 2008). These models applied vehicle probe data. However, parametric models are applicable only when the dependent function between the input and output variables is known. In addition, their performance is sensitive to the data quality. Outlier detection is necessary and they typically cannot deal with data sets with

missing values (Tu, 1996; Vlahogianni, et al., 2004). These disadvantages to some extent limit the applications of parametric models.

2.2   Travel time distribution estimation

The proposed study addresses the distribution of the travel time based on real-time travel time information in VANETs. Previous data fusion models for short-term travel time estimation provide valuable reference for this research to estimate the short-term travel time distribution. This is reflected when we discuss our information fusion methodology in later chapters. However, the proposed study is more generalized and falls into the category of travel time distribution estimation in the literature. In the literature, there are few efforts that focus on developing tools for estimating the travel time distribution (He, et al., 2002; Hollander and Liu, 2008). The state-of-the-art further suggests the need to develop methodologies to predict the short-term travel time distribution.

CHAPTER 3. INFORMATION NETWORK MODEL

We first model the dynamic local information map, which is the basis to dynamically guide the individual trips. In the example shown in Figure 3.1, the dynamic local information network is composed of the road network topology and the stochastic travel time (historical/real-time) information. The network topology is discussed first followed by the introduction of the formulation of the travel time information. The following notation is used hereafter: 1) a, b, c, d, e: the indices for links, 2) i, j, k, g: the indices for travel time states, 3) n, v, w: the indices for nodes, 4) r: the superscript representing real-time information, 5) h: the superscript representing historical information, and 6) f: the superscript representing information fusion. Other notation will be defined when they are first introduced.

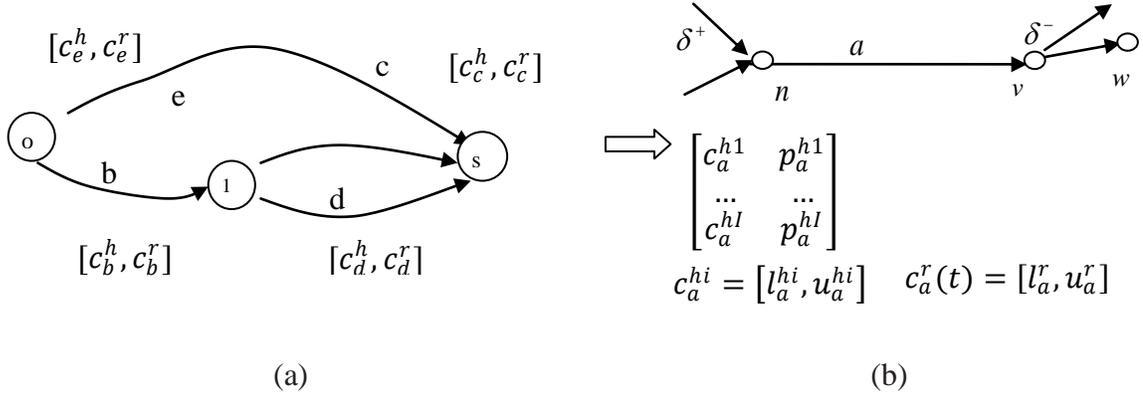(a)   (b)

Figure 3.1: An example of arc in the information network.

3.1 *Road network*

The road network is represented as a directed graph G (*N*, *A*), where the nodes correspond to intersections and the arcs correspond to road links. Let *N* be the set of all nodes and *E* be the set of all arcs (|E| = M). The direct link from node v to node w is

denoted by $[v, w]$. In the sample arc of Figure 3.1(b), the set of arcs incident from node n is denoted by $\bar{\delta}(n)$ while the set of arcs incident to node n is denoted by $\delta^{+}(n)$. We use o and s to denote the origin node and the destination node, respectively.

3.2   *Travel time distribution*

Long-term travel time distribution is generally considered to follow the normal or log-normal distribution under the conditions that travel times are independent on separate routes and travel time per unit section obeys identical distributions (Rakha, et al., 2006). However, He, et al. (2002) suggest that due to the complexity of real-world travel time distribution, the best way to describe short-time travel time distribution is by applying empirical distributions using a histogram from sample data. Using this perspective, the proposed study considers the travel time at each arc, $c_a(t)$ as a random variable whose short-term travel time variability is represented by a time-dependent discrete distribution (histogram). Thereby, the time domain is separated into small time intervals indexed by t = 0, 1, …, T. In each time interval, the random variable $c_a(t)$ has the possible states $\left\{s^i = [l_a^i(t), u_a^i(t)]\right\}_{i=1}^{I_a}$, where I_a is the total number of states of the discrete travel time distribution on Link a.

The discrete long-term historical travel time distribution, c_a^h for each arc is assumed to be known in this study. The corresponding PDF is given by Equation (1) below:

$$f(c_a^h) = \left[[l_a^{hi}, u_a^{hi}], p_a^{hi}\right]_{i=1}^{I_a}, a \epsilon E \tag{1}$$
$$s^{hi} = [l_a^{hi}, u_a^{hi}].$$

Under the proposed framework, the travel time obeys a continuous uniform distribution in each time interval. Correspondingly, the long-term historical mean and variance of the link travel time can be calculated using (2), (3), and (4):

$$E(c_a^i) = \frac{u_a^{hi} - l_a^{hi}}{2} \tag{2}$$

$$E(c_a) = \sum_{i=1}^{I_a} p_a^{hi} E(c_a^i) \tag{3}$$

$$\sigma_a = \sqrt{\sum_{i=1}^{I_a} \int_{l_a^{hi}}^{u_a^{hi}} (x - E(c_a))^2 p_a^{hi} dx}. \tag{4}$$

The long-term historical travel time distribution varies within large time intervals, such as rush hour/non-rush hour in a day, and weekday/weekend in a week. Therefore, the long-term historical information will be updated within a relatively long time interval.

3.3   Real-time travel time information

Real-time travel time is represented by a possible travel time range, $c_a^r(t) = [l_a^r, u_a^r]$ associated with a time-dependent probability $p_a^r(t)$ representing the information accuracy (He, et al., 2002), and a time delay (Δt) representing how "fresh" the information is. More specifically, the information accuracy $p_a^r(t)$ is a conditional probability that the realized real-time travel time is in state i given that the actual travel time is in state i:

$$p_a^r = p\left(c_a^r \in s_a^i(t) | s_a^i(t)\right), i = 1, \dots, I_a. \tag{5}$$

In addition, we assume that the probability that the realized real-time travel time is in state $j \neq i$, given the actual travel time is in state i, is uniformly distributed among all the possible states j, i.e.,:

$$p\left(c_a^r \in s_a^j(t) | s_a^i(t)\right) = \frac{1 - p_a^r}{I_a - 1}. \tag{6}$$

The time delay introduced here (Δt) can be explained as the gap between the time stamp that the real-time travel time information is generated and the time stamp the information arrives at the vehicle for updating the link travel time distribution[1]. Unlike the information accuracy, the time delay measures the uncertainty of the traffic condition from the time dimension. The proposed real-time travel time formulation incorporates the

---

[1] Case 1. Assume only individual vehicles' travel time data is disseminated in VANET, and the real-time traffic information is locally generated by individual vehicle itself based on the collected real-time traffic data. Each piece of real-time travel time data will arrive at vehicles in a time series order. Thus, it is unrealistic to generate link travel time estimation based on a single piece of data. Hence, this study assumes that the real-time data will be processed as a batch, and uses the mean of the arriving time across all the real-time data in a batch as the time stamp representing the arriving time of the information. Then, even if the travel time distribution is updated immediately once the real-time information is generated, a time delay still exists between the time stamps of information arrival and the distribution update.

Case 2. Assume the real-time traffic information is disseminated in VANET. There is a time delay between the time stamp that the travel time information is generated somewhere by a vehicle and the time stamp that the travel time information arrives at the other vehicle (and when it is used to update the corresponding link travel time distribution).

information errors that may be introduced into the travel time estimation during the information propagation, collection, and analysis procedure in VANET. The explicit relationship between the information quality and the network performance benchmarks can be found through extensive experiments, which is beyond the scope of this study. In summary, the information map such as the one shown on the left of Figure (1) provides the basis on which we build our information fusion model.

CHAPTER 4. INFORMATION FUSION MODEL

Previous studies (Tam and Lam, 2007; Kothuri, et al., 2008) suggest that integrating historical data and real-time data will improve the accuracy of estimating link travel time variability. Along this line, this study seeks a new methodology that mathematically integrates the historical travel time distribution and real-time travel time information so that the short-term travel time distribution can be dynamically updated. The approach uses the notion that historical information is more stable but may not represent current conditions, while real-time information is more "fresh" but with reliability issues. We first present the methodology at a conceptual level, followed by a detailed mathematical exposition.

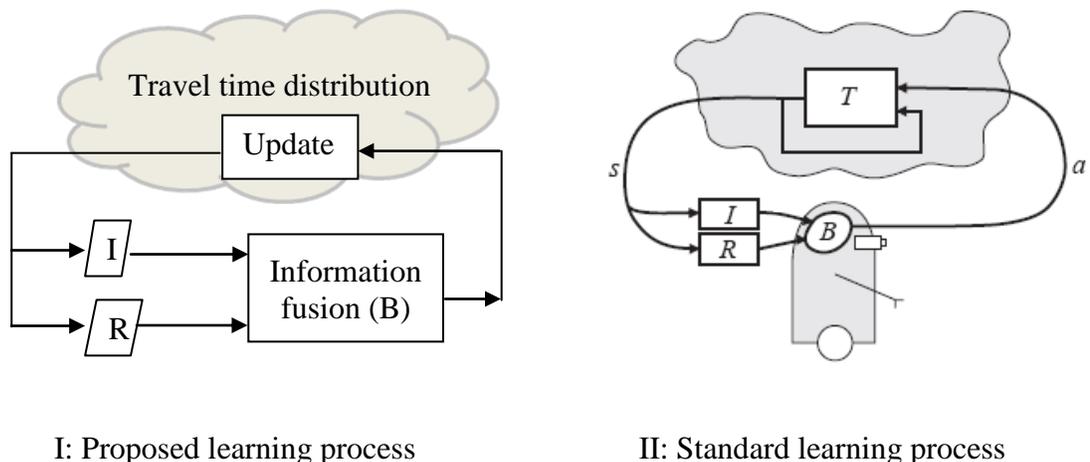I: Proposed learning process   II: Standard learning process

Figure 4.1: The proposed learning process and standard reinforcement learning

4.1 Adaptive learning process

The proposed approach aims to track the short-term travel time distribution as an adaptive learning process which is demonstrated on the left-hand side of Figure 4.1. The system iteratively refines its historical travel time knowledge (H) when new real-time information (I) is available at the vehicle. This adaptive learning process is similar to the reinforcement learning [2] approach in the literature. Using the framework of the reinforcement learning process depicted on the right-hand side of Figure 4.1, the learning process is described hereafter. The input information is the real-time traffic information (I) indicating the real-time traffic state (s). The information fusion works as the agent of action (a). The Behavior (B) of the agent is to choose the information fusion model and the reward (R) of the learning step is the reduction of uncertainty in the travel time distribution. The core technique of the learning process is to choose the information fusion model by which the historical travel time distribution and real-time traffic information are integrated to describe the latest travel time distribution. In the initial iteration, the historical travel time distribution is represented by the long-term travel time distribution estimation, and in the following iterations the realized short-term travel time distribution from the previous iteration is used as the historical travel-time distribution for the next iteration. The technical details of the information fusion model are discussed in the following sections.

4.2 Information fusion

To present the framework of the proposed information fusion, we first analyze the relationship between the two kinds of available travel time information, and then describe the key idea related to fusing them. Next, the mathematical model for the proposed information fusion problem is provided.

4.2.1 Travel time information analysis

The first difficulty to address the proposed information fusion issue comes from the frameworks of the historical and real-time traffic information. The historical travel time

---

[2] In the standard reinforcement-learning model, an agent is connected to its environment via perception and action. On each step of interaction the agent receives as input, i, some indication of the current state, s, of the environment; the agent then chooses an action, a, to generate as output (Kaelbling, Littman and Moore, 1996)

distribution of a link is available as a set of discrete states with corresponding probability. Each state is represented by an interval of the possible travel time on a given link. The real-time travel time information of a link is provided in the form of a possible travel time interval associated with the information quality. From the perspective of the information format, the proposed information fusion problem is to integrate a discrete distribution and a confidence interval. This precludes previously used data fusion methods such as artificial neural network, and linear regression, from being employed here.

We describe the relationship between the historical travel time interval and the real-time travel time interval through the six possible cases shown in Figure 4.2. For Case 1, Case 2 and Case 3, this study treats them under the following two situations:

1) Real-time information exactly covers one or more travel time intervals in the historical travel time interval sets:

$$f(c_a^r) = \left[[l_a^r(t), u_a^r(t)], p_a^r\right], l_a^r(t) = l_a^{hi}, \text{and } u_a^r(t) = u_a^{hj}, \forall i, j \epsilon I_a, a \epsilon E. \tag{7}$$

2) Real-time information interlaces two or more travel time intervals in the historical travel time distribution:

$$f(c_a^r) = \left[[l_a^r(t), u_a^r(t)], p_a^r\right], l_a^r(t) \neq l_a^{hi}, \text{or } u_a^r(t) \neq u_a^{hj}, \forall i, j \epsilon I_a, a \epsilon E. \tag{8}$$

As illustrated in Figure 4.2, theoretically, there are three other possible cases: (i) in Case 4, the lower bound of the real-time travel time interval is less than the smallest value in the historical travel time distributio; (ii) in Case 5, the upper bound of the real-time travel time interval is greater than the largest value in historical travel time distribution; and (iii) in Case 6, the real-time travel time interval completely lies outside the historical travel time data range. Since the historical travel time distribution is obtained from a large data set, it is reasonable to assume that it covers various traffic conditions and includes all the possible traffic states for each individual links over a long enough time window. Therefore, it is acceptable to ignore Cases 4 to 6 in this study. Later, we show that Case 2 and Case 3 can be converted to Case 1 using the proposed decomposition technique, and the information fusion model developed covers all three cases.

Case 1

Case 4

Case 2

Case 5

Case 3

Case 6
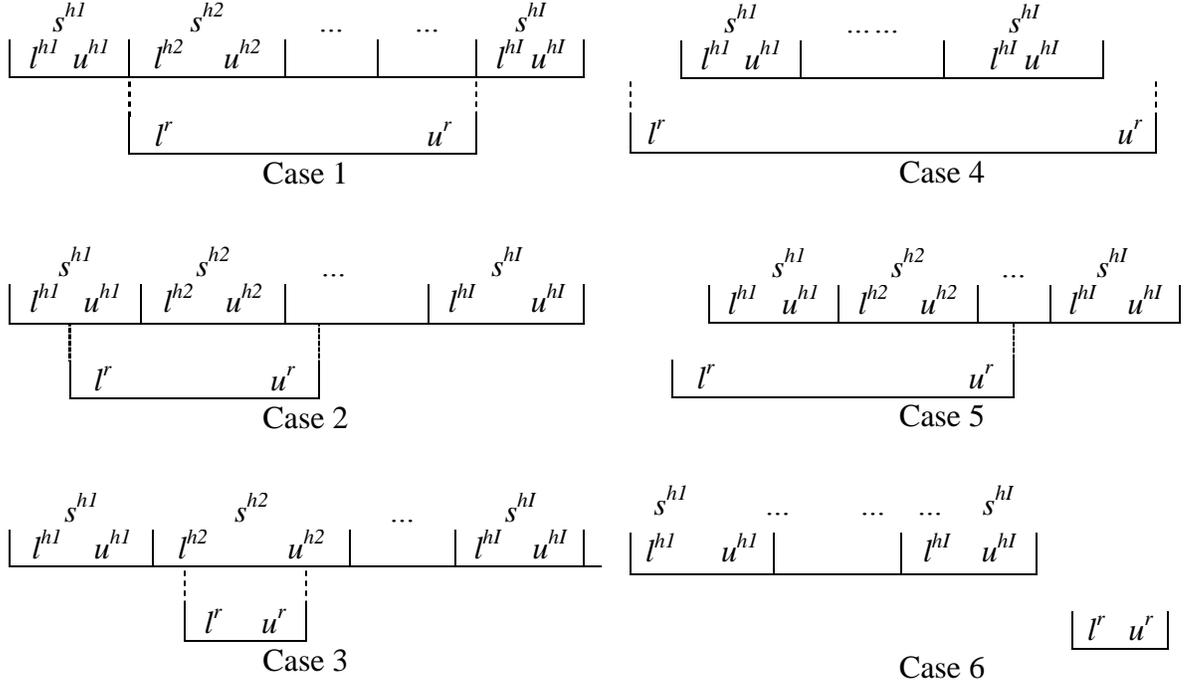
Figure 4.2: The relationship between the historical travel time distribution and the real-time travel time information

4.2.2 Main idea

Our initial idea to fuse the probability of the travel time states in a link by the historical travel time distribution and the real-time travel time information is mathematically described as follows:

$$p_a^i = x_a^i p_a^{hi} + \left(1 - x_a^i\right) q_a^i \, \forall i, j \epsilon I_a, a \epsilon E, \tag{9}$$

where,

$$q_a^i = p(s_a = i | s_a^r = i) = \frac{p(s_a = i, s_a^r = i)}{p(s_a^r = i)} \tag{10}$$

$$= \frac{p(s_a^r = i | s_a = i) p(s_a = i)}{p(s_a^r = i | s_a \neq i) p(s_a \neq i) + p(s_a^r = i | s_a = i) p(s_a = i)}$$

$$= \frac{p_a^r p_a^{hi}}{p^* + p_a^r p_a^{hi}}, a \epsilon E.$$

$$p^* = p(s_a^r = i | s_a \neq i)p(s_a \neq i) = \sum_{j \neq i} p(s_a^r = i | s_a = j)p(s_a = j)$$

$$= \sum_{j \neq i} \frac{1 - p_a^r}{I_a - 1} p(s_a = j) = \frac{1 - p_a^r}{I_a - 1} \sum_{j \neq i} p(s_a = j) = \frac{1 - p_a^r}{I_a - 1} \left(1 - p_a^{hi}\right)$$

To derive p^*, we assume the same probability for each event where the actual travel time state is i but the real-time information reports the state $j \neq i$. This concept is presented by Equation (6) in Section 2.3.
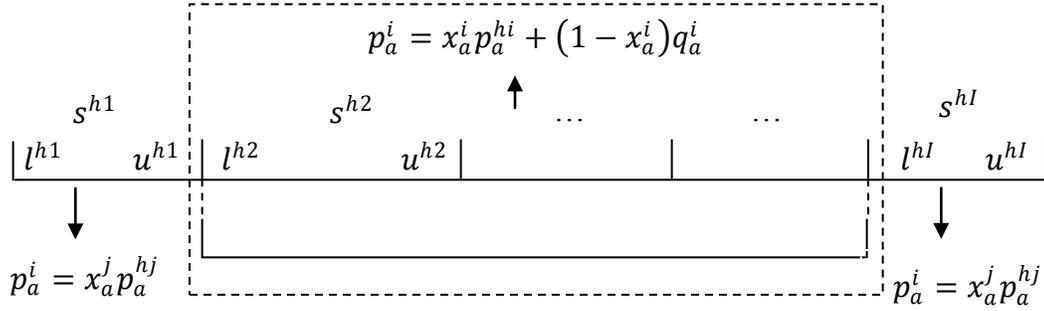

Figure 4.3: Information fusion model for Case 1

The convex combination presented in Equation (9) is a standard information fusion model used for travel time estimation in transportation research as well as practice. In this study, it is used to perform the travel time distribution estimation. As demonstrated in Figure 4.3, we apply the convex combination to each possible travel time state so that the probability of each travel time state in the historical distribution can be updated based on the real-time information. The immediate issue arising from the above model is how to decide the weight of different information sources in the information fusion model. Previous studies usually assign a fixed empirical weight without considering the dynamic traffic situation. In this study we develop a mathematical model to explore the optimal weights in considering the real-time information accuracy and time delay. There are several issues that need to be addressed before we present the mathematical model.

First, the objective of the information fusion here is to improve our knowledge about the link travel time distribution. To avoid the linguistic fuzziness, we quantitatively

evaluate the travel time distribution knowledge using Shannon's information entropy, which is a measure of the uncertainty associated with a random variable (Shannon, 1948).

Second, the weights of the real-time information are related to their corresponding information quality. To address this concern, we first discuss some characteristics of real-time information and historical information we identified in this study. When real-time information is available in state i, if the real-time information is associated with a long time delay, then historical information will gain more weight in the information fusion procedure. If the travel time of a link itself varies significantly (large value of variance, σ_a) over time and the real-time information has a high accuracy (high value of p_a^r), then the real-time information is assumed more reliable to represent the traffic condition. Correspondingly, the historical information has less weight in the information fusion procedure. Equation (11) reflects this logic: the weight of the historical information for a given link x_a^i is directly proportional to the real-time information time delay, Δt_a, but inversely proportional to the variation of travel time, σ_a, and the accuracy of real-time information, p_a^r. This relationship works for the weight of real-time information in the reverse manner.

$$x_a^i \propto \frac{\Delta t_a}{\sigma_a p_a^r} \tag{11}$$

Thereby, the information fusion among different links should be unified according to their information quality so that the link having better real-time information employs smaller weight of the historical information in the fusion model, and vice versa. This is used to introduce constraints (18) and (19) in the proposed model M1 discussed later.
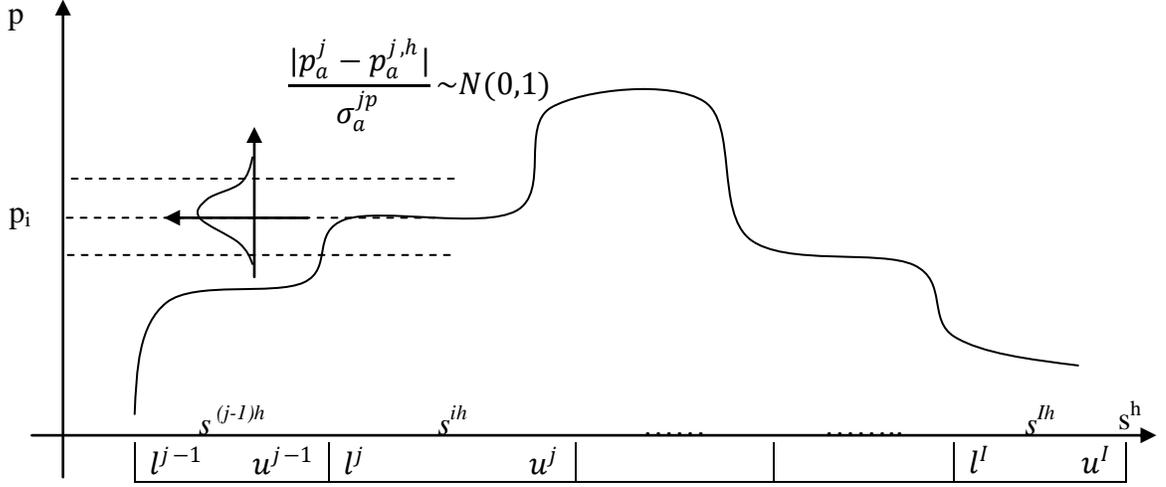
Figure 4.4: The probability that a travel time state on Link a obeys normal distribution

Third, with respect to the consistency of link travel time, the information fusion model should be designed in a way so that the travel time distribution is not changed significantly and arbitrarily by each single update action. This means that the adjustment of the travel time state probability by the new real-time information should be limited in ability. To address this constraint mathematically, we consider the probability of each travel time state on a link, p_a^i as a random variable, whose fluctuation obeys a normal distribution with the mean equal to the long-term historical probability of the travel time state, p_a^{hi}. The corresponding variance σ_a^{ip} cannot be obtained from the proposed information framework, but is achievable by collecting more historical data[3]. The range of the adjustment in the information fusion model at each iteration is designed to stay in a confidence interval with the significance level, α[4]. It is equal to $P\left(\frac{\left|p_a^j - p_a^{hj}\right|}{\sigma_a^{jp}} \leq z_{\frac{\alpha}{2}}\right) = 1 - \alpha$. This idea is illustrated in Figure 4.4. Mathematically, this constraint can be expressed as below:

$$\frac{\left|p_a^j - p_a^{hj}\right|}{\sigma_a^{jp}} \leq z_{\frac{\alpha}{2}}, \forall\, j \in I_a, a \in E. \tag{12}$$

[3] We briefly introduce the process as follows: we can collect multiple sets of historical data. From each data set obtain the probability of a travel time state, and then calculate the sample variance of the probability of the travel time state.
[4] Suppose that we have hypotheses, H_0: mean $(p_a^i) = p_a^{hi}$, H_a: mean $(p_a^i) \neq p_a^{hi}$. The significance level is the maximum probability of rejecting H_0 when H_0 is true (Bain and Engelhardt, 2000).

4.2.3 Mathematical model

Treating historical travel time distribution and real-time travel time information (c_a^h and c_a^r) as the input parameters, we develop a nonlinear programming model M1 to generate the information fusion model. In M1 we assign indices $k = \max\{i|s^i \in [l_a^r, u_a^r]\}$, $g = \min\{i|s^i \in [l_a^r, u_a^r]\}$ for the links in Case 1 as described in Equation (8).

$$\textbf{M1} \quad min \sum_{a=1}^{E} \sum_{i=1}^{I_a} \left(-p_a^j log p_a^j\right) \tag{13}$$

s.t.
$$p_a^i = x_a^i p_a^{hi} + \left(1 - x_a^i\right)q_a^i, \forall\, i \in [g_a, k_a], a \in E \tag{14}$$

$$p_a^j = x_a^j p_a^{hi}, \forall\, i \neq j, j \in I_a, a \in E \tag{15}$$

$$x_a^i \geq 0, \forall\, i \in [g_a, k_a], a \in E \tag{16}$$

$$x_a^i \leq 1, \forall\, i \in [g_a, k_a], a \in E \tag{17}$$

$$x_a^i - x_b^{i'} \geq 0, \forall\, \frac{\Delta t_a}{\sigma_a p_a^{ri}} \geq \frac{\Delta t_b}{\sigma_b p_b^{ri}}, i \in [g_a, k_a], i' \in [g_b, k_b], a \neq b \in E \tag{18}$$

$$x_a^i - x_b^{i'} \leq 0, \forall\, \frac{\Delta t_a}{\sigma_a p_a^{ri}} \leq \frac{\Delta t_b}{\sigma_b p_b^{ri}}, i \in [g_a, k_a], i' \in [g_b, k_b], a \neq b \in E \tag{19}$$

$$\left|p_a^j - p_a^{hj}\right|/\sigma_a^{jp} \leq z_{\alpha/2}, \forall\, j \in I_a, a \in E \tag{20}$$

$$\sum_{j=1}^{I_a} p_a^j = 1, \forall a \in E \tag{21}$$

$$p_a^j \geq 0, j \in I_a, \forall a \in E \tag{22}$$

The objective function (13) represents the objective of the information fusion model: based on the information we have, the proposed information fusion model aims to refine the current travel time distributions in the local map so that the information uncertainty is minimized. Shannon's Entropy is employed to evaluate the information uncertainty. Constraint (14) is a convex combination, which integrates the link travel time state probabilities of the historical and real-time information for a given link. Constraints (18) and (19) illustrate the linkage of the information fusion among different links in the local network map. That is, the link having better real-time information has smaller weight for historical information in its fusion model, and vice versa. Constraint (15)

updates the probabilities of other travel time states which are not covered by real-time information. Constraints (16) and (17) guarantee that the probabilities of all travel time states obey the basic properties of probability after they are updated by the information fusion model. Constraints (21) and (22) represent the basic properties of probability. Constraint (20) limits the difference between the updated distribution and the previous distribution.
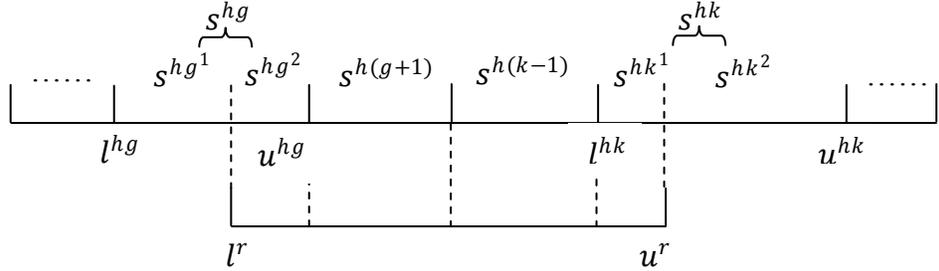

Figure 4.5: Method of converting Case 2 to Case 1

For the links in Case 2, we convert it to Case 1 by the method demonstrated in Figure 4.5. We decompose the real-time interval into several sub-intervals such that the real-time information can be represented by Equation (23).

$$f(c_a^r) = \left[[l_a^r, u_a^r], p_a^r\right] = \tag{23}$$

$$\left\{\left[[l_a^r, u_a^{hg}], p_a^r\right], \left[[l_a^{h(g+1)}, u_a^{h(g+1)}], p_a^r\right], \dots, [l_a^{hk}, u_a^r], p_a^r]\right\},$$

where $g = \{i | l_a^r \in s^i\}, k = \{i | u_a^r \in s^i\}$. Assuming c_a^h obeys uniform distribution in each state, we further decompose the corresponding historical interval $\left[[l_a^{hg}, u_a^{hg}], p_a^{hg}\right]$ into two states:

$$(g^1): \left[[l_a^{hg}, l_a^r], \ p_a^{hg} \frac{l_a^r - l_a^{hg}}{u_a^{hg} - l_a^{hg}}\right], \tag{24}$$

$$(g^2): \left[[l_a^r, u_a^{hg}], \ p_a^{hg} \frac{u_a^{hg} - l_a^r}{u_a^{hg} - l_a^{hg}}\right]. \tag{25}$$

That is, we insert a new state g^1: $\left[l_a^{hg}, l_a^r \right] = \left[l_a^{hg^1}, l_a^r \right]$. In the same way, the other

historical interval $\left[[l_a^{hk}, u_a^{hk}], p_a^{hk} \right]$ is decomposed into the two states below:

$$(\text{k}^1): \left[[l_a^{hk}, u_a^r], p_a^{hk} \frac{u_a^r - l_a^{hk}}{u_a^{hk} - l_a^{hk}} \right], \tag{26}$$

$$(\text{k}^2): \left[[u_a^r, u_a^{hk}], p_a^{hk} \frac{u_a^{hk} - u_a^r}{u_a^{hk} - l_a^{hk}} \right]. \tag{27}$$

Therefore, Case 2 is simplified to Case 1. Model **M1** becomes suitable for Case 2 with

the minor modifications that the state set I_a is changed to $I_a^{'} = \left\{ I \backslash \{ g_a, k_a \} \right\} \cup$

$\{ g_a^1, g_a^2, k_a^1, k_a^2 \}$, and the set of states i is enlarged by adding $\{ g_a^2; k_a^1 \}$. Clearly, Case 1 is a

special situation of Case 2 when the additional states $\{ g_a^2, k_a^1 \}$ are empty. To address

Case 3, we decompose the state of the historical travel time distribution, which includes

the real-time travel time information. Modifying the corresponding state I_a (the set of

states i), we get the information fusion model for Case 3.

In summary, there are three key points in above methodology to solve the

proposed information fusion problem: (i) The real-time information provides new

realizations for the travel time distribution. Therefore, it can be used to refine the known

travel time distribution to reduce its uncertainty; (ii) Since multiple sources of

information (real-time information and historical travel time distribution) are considered,

we propose to improve the discrete historical travel time distribution by the convex

combination fusion models. The weights of the information from different sources are

decided by the nonlinear programming model; (iii) The information fusion for different

links are conducted with reference to the levels of the information quality. From the

perspective of information quality, the information fusion for all the links in the local

map are developed uniformly.

CHAPTER 5. TEST-BED AND SYNTHETIC DATA

In this chapter, we first describe the Borman Expressway traffic network from which a specific sub-network is used to analyze our information fusion model. After that, we introduce the procedures used to generate historical travel time data and the real-time travel time information. Based on this experimental setup and the procedure to examine the proposed information fusion model, we perform experiments whose results are discussed in the next chapter.

5.1   *Test-bed*

Our experiments are based on traffic simulation on the Borman network shown in Figure 5.1, which consists of a 16 mile section 94 (called the Borman Expressway), I-90 toll freeway, I-65, and the surrounding arterials. The network has 197 nodes and 458 links, and is divided into 14 zones. The Borman Expressway is a highly congested freeway with 30% - 70% semi-trailer truck traffic. To manage traffic under incidents and peak period congestion, an advanced traffic management system has been installed on the Borman network to provide travelers real-time traffic information.

        In our experiments, we collect the traffic information on only a small sub-network (of the Borman network) with 19 links to test our information fusion models though the traffic simulation itself is done on the entire network. The traffic flow through the Borman network is considered as the general traffic environment. The topology of the small network is indicated by the dotted lines in Figure 5.1.
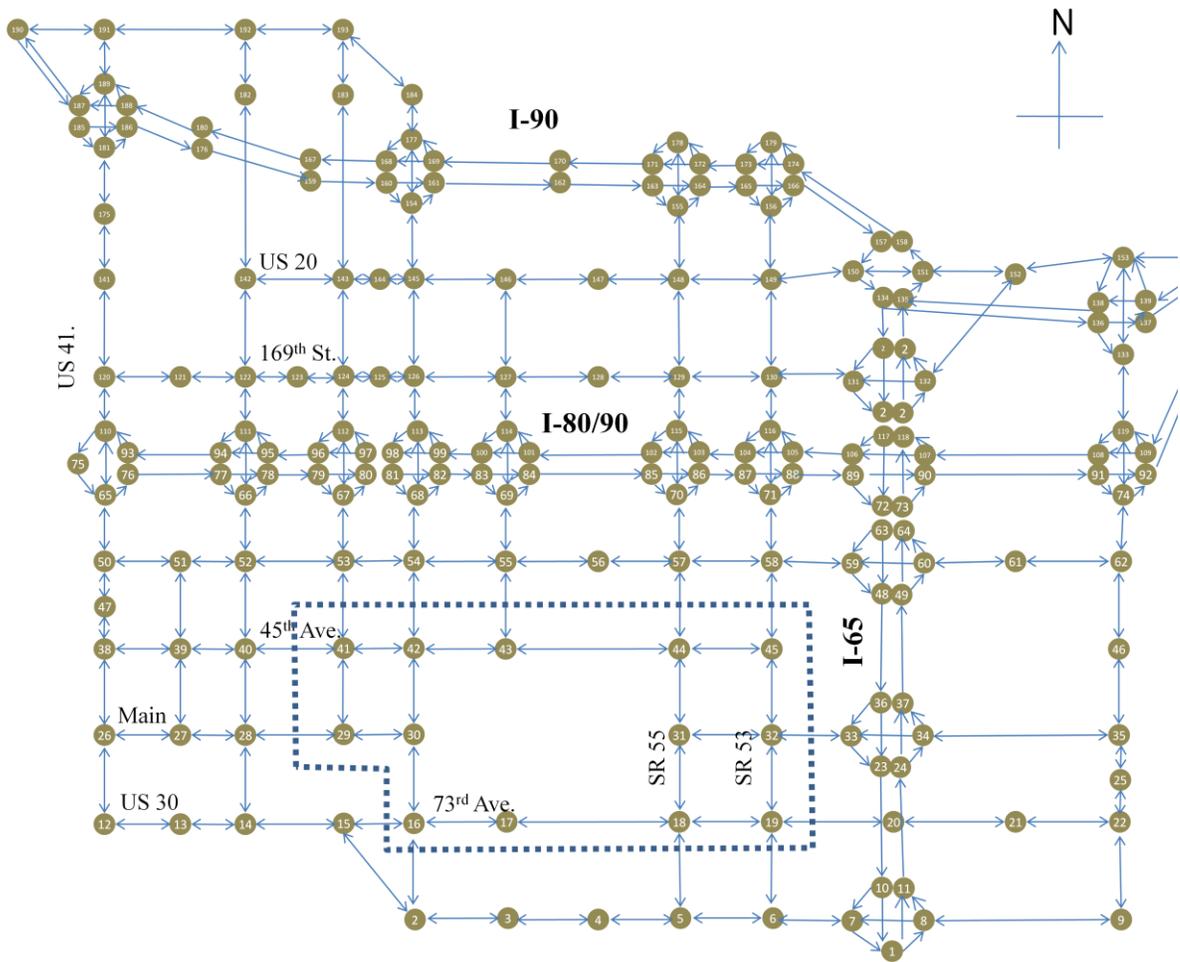
Figure 5.1: Borman network and the sub-network

5.2   *Synthetic data generation*

Due to the non-availability of field experiment data, this study develops a procedure to generate synthetic data for both long-term historical travel time distributions and the real-time travel time information. The proposed procedure and the corresponding simulations are consistent with traffic realism.

5.2.1 Generation of the long-term historical travel time distribution

The long-term historical travel time distribution is generated using the DYNASMART simulator. The experiments first simulate the traffic flow in the Borman network, and

then collect the aggregated link travel time for the links of interest for one hour each day. Next, using the collected data, we obtain the long-term historical travel time distribution, c_a^h using the framework defined in Section 2.2. More specifically, to describe the long-term historical travel time using various states, this study clusters the link travel time into different travel time states by dividing the range of travel time data (the interval between the maximum and minimum travel time) using an interval of fixed length. The probability of each state is estimated by the ratio of the travel time data falling in the corresponding travel time slot to that for all the travel time data obtained during the experiments. Travel time states vary across individual links.

In summary, a key issue in this procedure is how to group the travel time data into travel time states for each link. In addition to the approach used here, this issue can also be solved using other cluster algorithms, which has been extensively studied in literature (Jain, et al., 1999).

5.2.2 Generation of real-time information

To obtain the real-time information, we develop a semi-random procedure to mimic the real-time information collection process in field experiments. As demonstrated in Figure 5.2, the designed procedure uses the simulation data on certain days to represent the actual/true traffic condition, and then combines the long-term historical travel time distribution and the actual traffic condition to generate the real-time travel time interval associated with information accuracy and time delay.

More specifically, at time t, we first randomly generate the time delay (Δt_a) and the information accuracy (p_a^r). The time delay Δt_a implies that the real-time information is obtained at time $(t - \Delta t_a)$; the corresponding actual traffic condition is $c_a(t - \Delta t_a)$. Further, since the "collected" real-time travel time interval $[l_a^r, u_a^r]$ is based on the actual travel time and the accuracy of information collection systems at that time, we simulate the data collection procedure using the following mechanism.

Another random value α is employed to simulate the dynamic performance of traffic information collection systems. If $\alpha < p_a^r$, it indicates that "good" real-time travel time interval is obtained, that is, with the probability p_a^r. The collected real-time traffic information represents the actual traffic condition at that time stamp. Mathematically, the

"good" real-time travel time interval is described by $[l_a^r, u_a^r] \subset [l_a^{hi}, u_a^{hi}] = s_a^i(t)$ with information accuacy p_a^r. If $\alpha > p_a^r$, consider the collected real-time travel time interval has poor accuracy. It means that the real-time travel time interval randomly floats in a broader range and does not necessarily reflect the actual traffic condition well. Mathematically, the poor real-time information is described by $[l_a^r, u_a^r] \subset \cup_{j=1}^{I_a} s^j(t)$ with the information accuracy $\frac{1-p_a^r}{I_a-1}\beta$, where β is the number of states covered by real-time information.

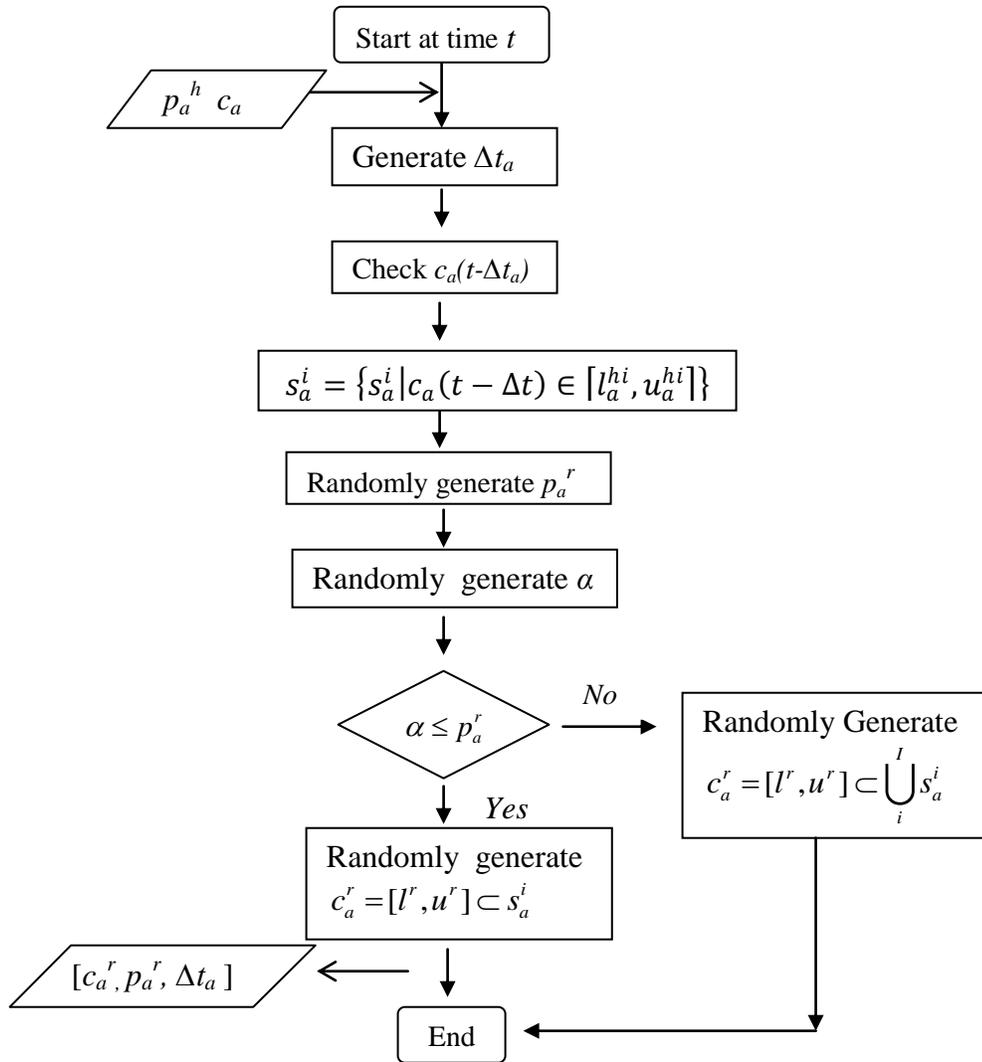

Figure 5.2: Semi-random procedure: generate real-time information

CHAPTER 6. NUMERICAL EXPERIMENTS

This chapter discusses results of the numerical experiments to evaluate the proposed information fusion model. Its performance is evaluated from the following three aspects (van Lint, 2006).

1. Accuracy; the difference between the actual travel time distribution and the estimated short-term travel time distribution should be small.

2. Robustness; the estimated short-term travel time distribution is close to the long-term historical travel time distribution if the information quality is very poor and approaches the actual travel time distribution if the information quality is good.

3. Adaptivity; the information fusion model is able to track the actual travel time changes and adapt the estimated short-term travel time distribution accordingly.

6.1   Experiment procedure

Using the test-bed network and synthetic data, the following procedure analyzes the performance of the information fusion model:

1. Run Borman network in DYNASMART for 90 minutes on 55 days based on the randomly generated O-D demand on each day.

2. Collect the travel time data for the test network with 19 links.

3. Use the data from the first 50 days to generate the long-term historical travel time distribution for each link.

4. Choose a day in the last 5 days, and track the link short-term travel time distribution by applying the proposed information fusion models using the following procedure: (i) From time stamp 30 to 70, we randomly generate real-time travel time information for each link every 2 minutes, and (ii) Using this

sequence of the real-time information generated, we update the historical link travel time distribution through the information fusion models.

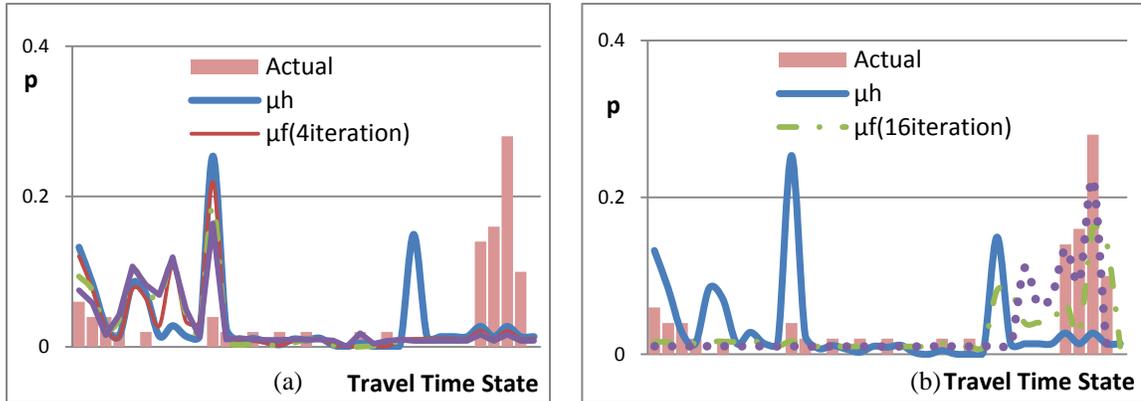5. Examine the efficiency of the information fusion model.

Figure 6.1: Travel time distributions on a link

f (*): short-term travel time distribution at iteration *;

h: historical travel time distribution; Actual: actual travel time distribution on that day

6.2   Short-term travel time distribution

We first test the efficiency of the information fusion models by tracking the short-term travel time distribution in terms of its ability to approach the actual travel time distribution. Without loss of generality, we randomly pick a link in the study network and demonstrate its short-time travel time distribution at different iterations in Figure 6.1. The results show two important observations: (1) the long-term historical travel time distribution does not represent the actual travel time distribution of this link on this day. This is consistent with the observation of He, et al. (2002); (2) Though the short-term travel time distributions estimated by the information fusion model were much different from the actual travel time distribution in the first few iterations, and stayed closer to the long-term historical travel time distribution (see f(4), f(8), and f(12) in Figure 6.1(a)), after more iterations, the information fusion model enabled the short-term travel time distribution to gradually approach the actual link travel time distribution on that day (see f(16) and f(20) in Figure 6.1(b)). Therefore, by tracking the link short-term travel time

distribution, we conclude that the proposed information fusion models demonstrate good quality in terms of accuracy and adaptivity.

6.3   *Mean of short-term travel time distribution*

To further demonstrate the efficiency of the information fusion models, we additionally evaluate the performance of the estimated short-term travel time distribution by comparing four different link travel times at time t:

i) c_a : the actual link travel time, which is obtained from the DYNASNART simulation.

ii) μ^h: the mean of the long-term historical travel time distribution, which is a fixed value.

iii) μ^f: the mean of the latest short-term travel time distribution which is iteratively updated by the information fusion model.

iv) μ^r: the mean of the real-time travel time interval.

Figure 6.2 presents the results for the four sample links, which cover the typical features of real-time information in terms of quality, such as the consistently good real-time information on Link 1 with μ^r always close to c_a; inaccurate real-time information on Links 2, 3 and 4 with μ^r sometimes far away from c_a; long time delay on Link 3 where μ^r shifts to right of c_a with similar shape. A detailed analysis of the curves in Figure 6.2 suggests the following characteristics for the information fusion models:

1. The curves for Link 1 and Link 3 show that as the actual travel time $c_a(t)$ changes (decreases on Link1 and increases on Link 3), the information fusion model can correspondingly adjust the short-term travel time distribution, and its mean μ^f approaches the actual travel time in an adaptive manner. Thus, the information fusion model demonstrates good adaptivity.
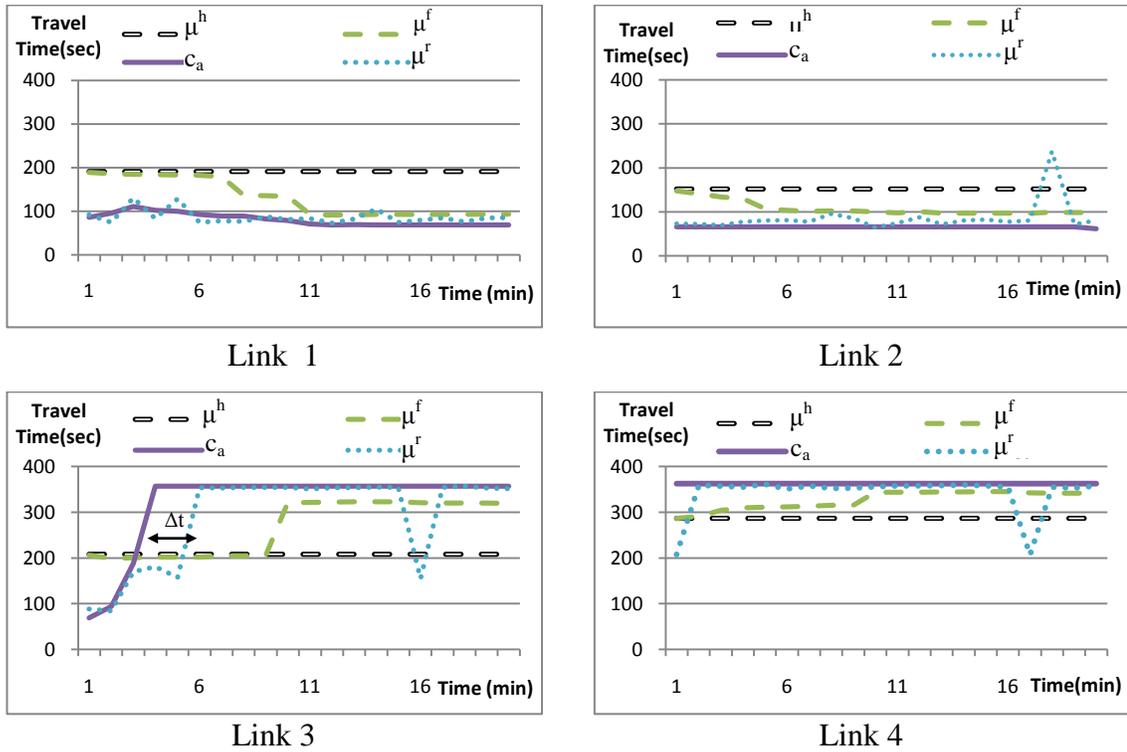
Figure 6.2: Results for information fusion models

2. The curves for Links 2, 3, and 4 demonstrate that when the real-time travel time is not accurate (at time 17 and 18 on Link 2, at time 16 on Link 4) or has long time delay (time 1 to time 6 on Link 4), the information fusion model pushes the short-term travel time distribution closer to the historical travel time distribution, but once the time delay is decreased or information accuracy is improved, the information fusion drives the short-term travel time distribution to the actually travel time distribution relatively quickly. Hence, the information fusion model does not direct the short-term travel time estimation to a worse situation due to inaccurate real-time information. In the worst case, it will ensure that the average performance is as good as the historical travel time distribution. Thus, the information fusion model is robust.

3. The curves on all links demonstrate that the mean of the short-term travel time distribution μ^f always gradually approaches the actual travel time $c_a(t)$ as our

information fusion model updates the historical travel time distribution step by step. Thus, the proposed information fusion model has good accuracy.

Overall, the proposed information fusion model performs well in estimating the short term travel time distribution, whose mean approaches the actual travel time step by step when good real-time travel time information is constantly available, but stay at the same level as the historical travel time distribution in case bad real-time travel time information is obtained. The performance for other links in our experiments is given in Appendix A.

6.4   Parametric sensitivity analysis

We are additionally interested in the significance of the key parameters in our information fusion models. Sensitivity analysis is performed for these parameters: the time delay Δt, information accuracy p_a^r , and the significance level of the state probability, α. The specific experiments are:

1) Keeping the time delay Δt and the significance level α constant, we explore the impact of information accuracy, p_a^r.

2) Keeping the information accuracy p_a^r and the significance level α constant, we explore the significance of time delay, Δt .

3) Keeping the time delay Δt and the information accuracy p_a^r constant, we explore the effect of the significance level, α.

To compare the performance of our information fusion models, we use the root mean of squared error (RMSE) calculated by Equation (28), in which M is the number of links in the study network.
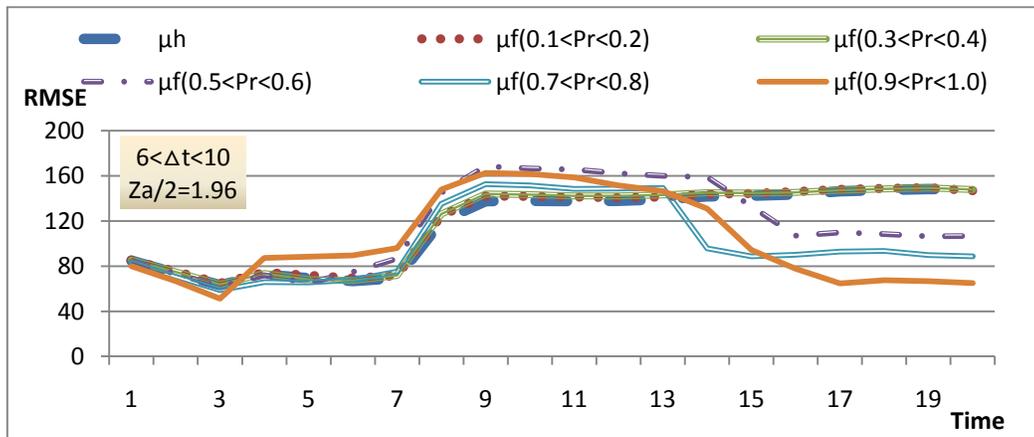
$$RMSE = \sqrt{\frac{\sum_a \left(\mu_a^f(t) - c_a(t)\right)^2}{M}} \tag{28}$$

6.4.1 Sensitivity analysis for information accuracy

We first test the significance of the information accuracy in our information fusion models. Selecting $z_{\alpha/2} = 1.96$ ($\alpha = 0.05$), the experiments for various information accuracy are separated into two groups according to the time delay levels ($6 < \Delta t < 10\ min$ representing the long time delay and $1 < \Delta t < 5\ min$ representing the short time

delay). The results in Figure 6.3 show that whether the time delay is short or long and the information accuracy is low or high, the curves of μ^f (representing the RMSE value of μ^f) are almost always lower or close to the curve of μ^h (representing the RMSE value of μ^h). In addition, in the beginning periods, the curves of μ^f and μ^h are close to each other; but later as the estimated short-term travel time distributions are updated by the information fusion model, the curves of μ^f with higher information accuracy levels, such as $0.9 < p^r < 1$, is usually lower than the curves of μ^f with lower information accuracy, such as $0.3 < p^r < 0.4$. Comparing the curves with the same information accuracy level but different time delay in Figure 6.3(a) and Figure 6.3(b), we find the curves of μ^f with shorter time delay usually are improved earlier than the one with longer time delay.

These observations indicate two insights about the impact of the information accuracy on the performance of the information fusion models. First, irrespective of whether the time delay is long or short, higher information accuracy improves the performance of the proposed information fusion models and produces a better estimation for the short-term travel time distribution. Second, the effect of the information fusion model takes some time to be observed, but shorter the time delay results in a quicker improvement.
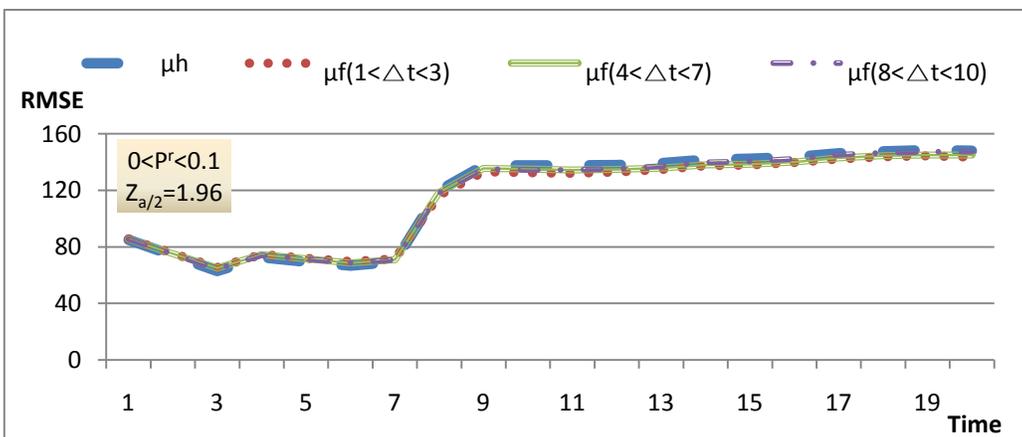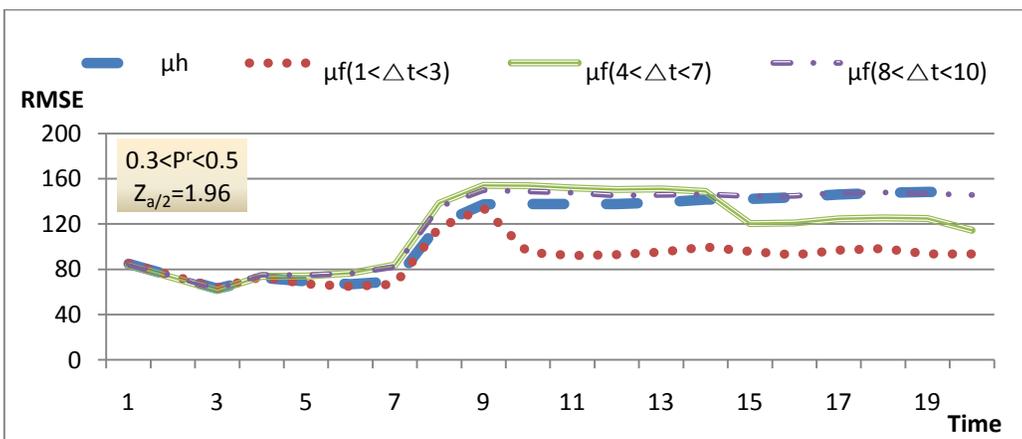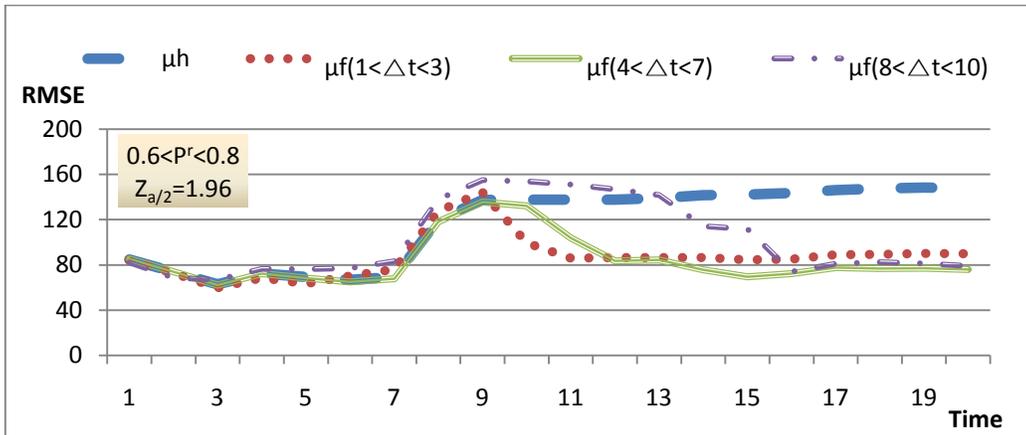

(a)
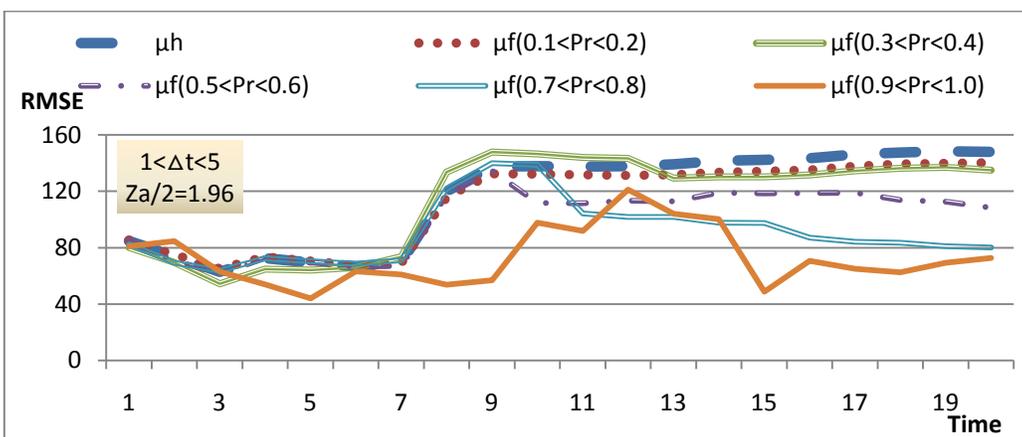
Figure 6.3: Sensitivity analysis for information accuracy

Figure 6.4: Sensitivity analysis for time delay

6.4.2 Sensitivity analysis for time delay

Here, we test the sensitivity of the time delay in the information fusion model. Assigning the significance level, $\alpha = 0.05$ (i.e. $z_{\alpha/2} = 1.96$), we test the performance of the information fusion under three different time delay levels: $1 < \Delta t < 3min$, $4 < \Delta t < 7min$, and $8 < \Delta t < 10min$. In addition, under each time delay level, the experiments are conducted for three different information accuracy levels: $0 < p^r < 0.1$, $0.3 < p^r < 0.5$, and $0.6 < p^r < 0.8$. The curves in Figure 6.4 demonstrate that under low information accuracy, the results of information fusion will always be close to the long-term historical travel time distribution as seen in Figure 6.4(a). Under moderate or high information accuracy levels (see Figure 6.4(b) and Figure 6.4(c)), the curves of μ^f move below the curve of μ^h. In addition, the curve of μ^f with shorter time delay, such as $1 < \Delta t < 3$ will go below the curve of μ^f with longer time delay, such as $4 < \Delta t < 7$, and $8 < \Delta t < 10$. These findings indicate that our information fusion model is more sensitive to time delay as the information accuracy is at a relative high level. It also implies that information accuracy is more significant than time delay to our information fusion model.

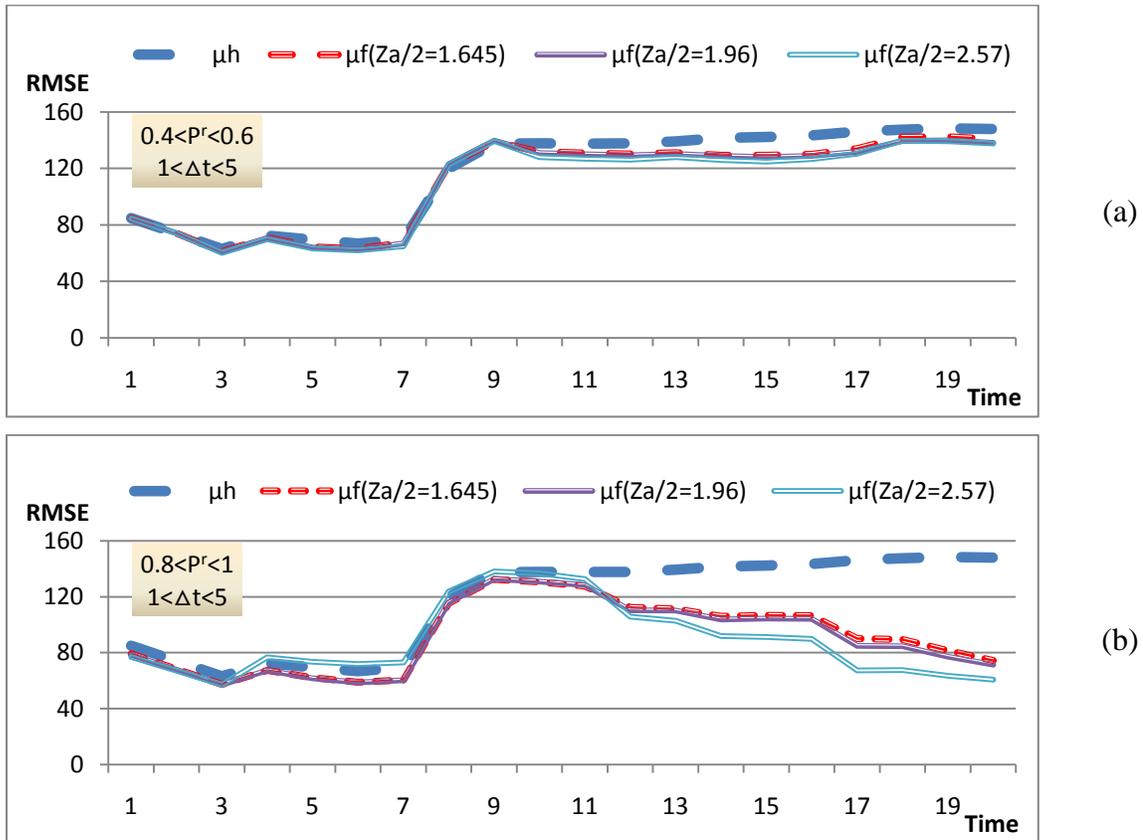*6.4.3 Sensitivity analysis for significance level*

Figure 6.5: Sensitivity analysis for significance level

We also test the sensitivity of the significance level for the state probability in the information fusion model. Keeping the time delay greater than 1 minute and less than 5 minutes, we test the sensitivity of the significance levels $\alpha = 0.01, 0.05,$ and 0.1 (i.e. $z_{\alpha/2} = 2.575, 1.96,$ and 1.645) under the conditions $0.4 < p^r < 0.6$ and $0.8 < p^r < 1$, respectively. The results show that when the information accuracy is high, $0.8 < p^r < 1$, the higher significance level will improve the performance of the proposed information fusion model (see Figure 6.5(b)). However, when the information accuracy is low or even moderate ($0.4 < p^r < 0.6$), the significance level almost has no influence on the results of the performance of information fusion (see Figure 6.5(a)). This indicates that the information accuracy dominates the impact, namely, the significance level is only important to the proposed information fusion when the information accuracy is at a relatively high level.

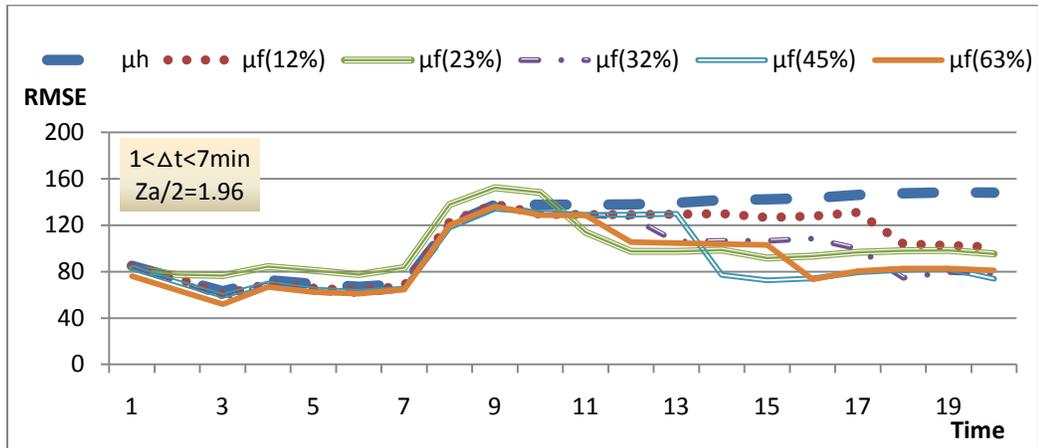*6.4.4 Sensitivity analysis for good information rate*

Figure 6.6: Sensitivity analysis for good information rate

One more sensitivity analysis tests the effect of the good real-time information rate. To facilitate the setup of this experiment, we define good real-time information as the quality that $p_a^r \geq 0.7$ and $\Delta t \leq 5$ minutes. Accordingly, without loss of generality, we maintain the time delay in 0~7 minutes, assign the significance level $\alpha = 0.05$ (i.e. $z_{\alpha/2} = 1.96$), and examine the performance of our information fusion models when the real-time information is available with different percent of "good" information.

The results shown in Figure 6.6 demonstrate that even when only a small percent of information is good, such as 12%, the curve of μ^f has smaller RMSE values than the curve of μ^h, which means across all links in the study, the mean of the short-term travel time distribution provides a more accurate estimation on actual travel time conditions than the historical travel time distribution. In addition, as the percentage of good information increases, such as 63% in the figure, μ^f has even smaller RMSE value. This indicates that the mean of the short-term travel time distribution estimated by our information fusion model has a better approximation to the actual travel time condition if more good real-time information is available. In other words, the information fusion modes are sensitive to the rate of good real-time information.

CHAPTER 7. APPLICATION IN ONLINE STOCHASTIC ROUTING

To further evaluate the value of the information fusion, this study applied the proposed information fusion model to the on-line stochastic routing problem (OLSRP) since it is the core component of the decentralized ATIS based on VANET. The following sections first introduce the routing policy and the modified label correcting algorithm, and then demonstrate the experiments of applying the proposed information fusion models to the on-line stochastic routing.

7.1   Routing policy

Previous studies have proved that the optimal solution to the OLSRP is not a single deterministic path but a routing policy (Hall, 1986). A policy is an identification of the next action of travelers given the network condition at the current time. This study seeks an on-line routing policy while focusing on the real-time information quality. Along with a closed-loop routing rule in (Fu, 2001), we develop the routing policies which identify the optimal alternative paths by employing the discrete stochastic travel time information on each link. The mathematical model is demonstrated below:

$$a^* (t) = argmin\{E[c_a(t)] + c_{is}(t), \forall\, a = [n, i] \in \delta^-(n)\}, \tag{29}$$

where, $a_1^*(t)$ represents the optimal arc that the traveler is suggested at time t so that under our routing policy he/she will experience the expected minimum travel time from current node to the destinations. $E(c_a(t))$ is the expected realization travel time of arc a at the time t. c_{is} is the expected minimum travel time from node i to the destination s given that the traveler is routed under the provided policy. It can be further expressed as below:

$$c_{is}(t) = E\big[min\{c_a(t) + c_{js}\,|\forall\, a = (i, j) \in \delta^-(i)\}\big], \tag{30}$$

$$c_{ss}(t) = 0. \tag{31}$$

Equations (29), (30), and (31) are formulated as a recurrent relation. In theory, the problem can be solved by dynamic programming (Bellman, 1965). Unlike Fu (2001), where they use a continuous travel time distribution and develop an approximate algorithm, we employ a discrete travel time distribution and further develop the exact routing algorithm to solve the problem.

7.2   Routing algorithm

To solve the shortest path under the proposed routing policy, we design the corresponding algorithm to identify the optimal arc at each decision node. The steps of the algorithm are:

1. Sort the index of nodes in the studied network by Topological Order Algorithm (Ahuja, et al., 1993).

2. Update the label of each node by the label-correcting algorithm in a reverse order of the node index. The label of each node j represents the expected minimum travel time from the node to the destination.

3. Calculate the optimal arc a^* by Equation (29).

The core of the above algorithm falls in the framework of the label-correcting algorithm, but several additional points need to be clarified:

- We apply the mean of the travel time state, which is an interval in the short-term travel time distribution, to represent the link travel time condition in the routing algorithm. This transformation makes the routing problem solvable since it is hard to compare the travel time of different links if their travel time is represented by intervals.

- The index number of each node in the network will be sorted based on Topological Ordering Algorithm (Ahuja, et al., 1993). It will make the general label-correcting algorithm more efficient.

- The proposed algorithm works on a directional acyclic graph (DAG). Section 7.2.1 below demonstrates our procedure to abstract the DAG for this study.

- The label of each node in the network is the expected minimum travel time from the current node to the destination. Section 5.2.2 below presents the designed algorithm to calculate the label for each arc in this study.

7.2.1 Generate DAG network

For the experiments, it is necessary to abstract an acyclic sub-network from the given network taking into account of the specific origin and destination. While this issue is well-known, only limited algorithms are provided. As it is not a focus of this paper, we employ the node geometrical position data, represented by coordinate (X ,Y) to develop a more straightforward method for our study. The following steps illustrate our approach:

1. Given the origin node $O(X_o, Y_o)$ and the destination node $S(X_s, Y_s)$, we identify the routing searching space composed of the egress/ingress/moving area, that is, node sets (see Figure 7.1 for an example) by the formula below:

- Nodes in egress area:

$$n_{ea} = \left\{ \forall v \in N | \sqrt{(X_v - X_o)^2 + (Y_v - Y_o)^2} < R_{ea} \right\} \tag{32}$$

- Nodes in ingress area:

$$n_{ia} = \left\{ \forall v \in N | \sqrt{(X_v - X_s)^2 + (Y_v - Y_s)^2} < R_{ia} \right\} \tag{33}$$

- Nodes in moving area:

$$n_{ma} = \left\{ \forall v \in N | \sqrt{(X_v - X_o)^2 + (Y_v - Y_o)^2} + \sqrt{(X_v - X_d)^2 + (Y_v - Y_d)^2} \right. \tag{34}$$
$$\left. < R_{ma} \right\}$$

2. With the node geometrical data, we pick the significant links $[v, w]$ (from node v to node w) in egress, ingress, and moving areas by the rules below:

- Links in egress area:

$$l_{ea} = \left\{ \forall [v, w] \in E | \sqrt{(X_w - X_o)^2 + (Y_w - Y_o)^2} \right. \tag{35}$$
$$\left. - \sqrt{(X_v - X_o)^2 + (Y_v - Y_o)^2} > 0 \right\}$$

- Links in ingress area:

$$l_{\text{ia}} = \left\{ \forall [v, w] \in E \big| \sqrt{(X_w - X_s)^2 + (Y_w - Y_s)^2} - \sqrt{(X_v - X_s)^2 + (Y_v - Y_s)^2} \right. \tag{36}$$
$$\left. < 0 \right\}$$

- Links in moving area:

$$l_{\text{ma}} = \left\{ \forall [v, w] \in E \big| \sqrt{(X_w - X_s)^2 + (Y_w - Y_s)^2} - \sqrt{(X_v - X_s)^2 + (Y_v - Y_s)^2} \right. \tag{37}$$
$$\left. < 0 \right\}$$

3. Final Directed Acyclic Graph (DAG) is $G(n_{ea} \cup n_{ia} \cup n_{ma}, l_{ea} \cup l_{ia} \cup l_{ma})$.

The produced DAG used in this study is identified by the dotted line in Figure 7.1, in which we choose $R_{ea} = R_{\text{in}} = 10$ mile for the egress/ingress area, and $R_m = 45$ mile for the moving area. The generated DAG includes enough search area for the given O-D in our experiments.
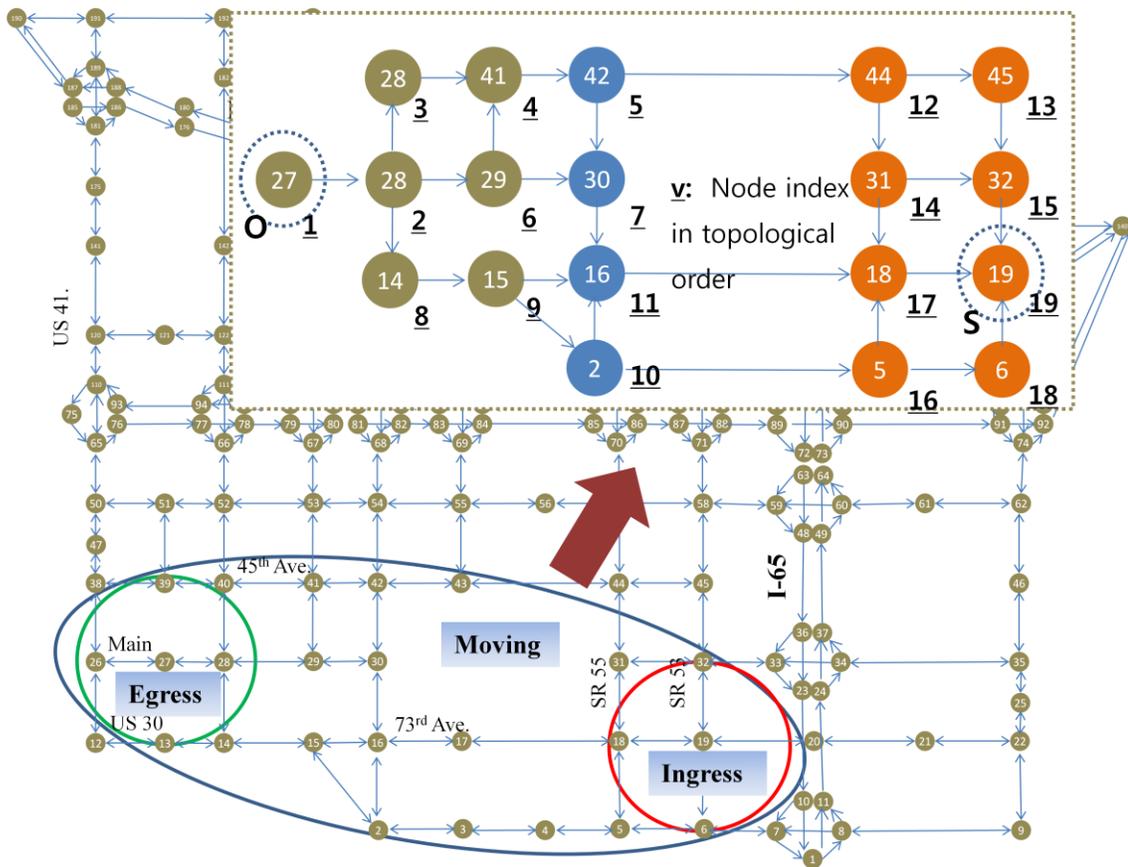

Figure 7.1: The DAG used in online stochastic routing

7.2.2 Calculate the labels for individual nodes

We develop a subroutine to calculate the label of individual node denoted by $\pi_v = E[min\{c_a(t) + \pi_w | \forall\, a = [v, w] \in \delta^-(v)\}]$ (i.e. Equation (30)), which is the expected minimum travel time from node i to the destination s under our routing policy. The idea of this subroutine is based on the following observation. Since the label-correcting algorithm goes in a backward manner, π_w is a permanent label when we calculate the label π_v. In addition, since $c_a(t)$ is a discrete stochastic variable with limited number of states, there is only a limited number of values in $\varphi = \{c_a(t) + \pi_w | \forall\, a = [v, w] \in \delta^-(v)\}$. Once we sort the values in φ and refer it to as φ', we correspondingly obtain a sequence of the states on all links $[v, w] \in \delta^-(v)$. Then we can find the case that the entire travel time states of arc $a' \in \delta^-(v)$ are listed behind the entire states of the other arc $a \in \delta^-(v)$. Using γ to indicate the position that the first state of arc a' behind the entire states of arc a, we recognize that all the values in φ' after γ are impossible to be the value in $min\{\varphi\}$, since arc a in any way will show up with a state which produces a relatively smaller value than all other values behind γ in φ'. With this idea, the proposed algorithm to calculate the label of any node v includes the following steps:

1. Sort the possible values in φ to obtain φ'.
2. Remove all the impossible values in φ' in terms of $min\{\varphi\}$. The final list is denoted by Γ.
3. Calculate the corresponding probability for each value in Γ.
4. Calculate the expected minimum travel cost from node v to node s under our routing policy.

7.3   Experiments

The aim of the application experiments is to demonstrate the value of the information fusion model when it is used to guide travelers' trips. For a fixed network and O-D pair, the driver will be routed with or without information fusion model, respectively. We compare the paths as well as the travel time (Σ) that the travelers actually experience. Table 7.1 shows the travel time of the travelers whose trip is routed by the real-time travel time information (r), the long-term historical travel time distribution (h), or the

short-term travel time distribution generated by the proposed information fusion model (f). The results demonstrate that when the information quality is relative good, such as $p^r > 0.9$, the estimated short-term travel time distribution works as well as the real-time information and is able to identify the possible best path. On the other hand, when the real-time information is very bad, such as $p^r > 0.3$, the shortest path searched by the estimated short-term travel time distribution is as good as the long-term travel time distribution, but better than the real-time information. Hence, the information fusion avoids the worst case that the traveler is misguided by the bad information. More importantly, when the real-time information quality is moderate, such as $p^r > 0.7$, the proposed information fusion efficiently obtains the shortest path, by which the traveler experiences lesser travel time than under the guidance of either real-time information or long-term historical travel time distribution. The application results indicate the importance of considering the information quality in actual applications of VANETs. In addition, it further demonstrates the value as well as the efficiency of the proposed information fusion model.

Table 7.1 Results for online stochastic routing application

	$1 < \Delta t < 5$, $z_{\alpha/2} = 2.575$		
	$p^r > 0.9$	$p^r > 0.7$	$p^r > 0.3$
Σ^f	27→28→14→15→16→1 →19 = 1631.6	27→28→14→15→16→18 →19 = 1631.6	27→28→14→15→2→16→18 →19 = 2059.8
Σ^r	27→28→14→15→16→18 →19 = 1631.6	27→28→29→30→16→17 →18→19 = 1691.6	27→28→29→41→42→44→45→ 32→19 = 2121.3
Σ^h	27→28→14→15→2→16→18→19 = 2059.8		
Σ	27→28→14→15→16→18→19= 1631.6 (possible best path)		

CHAPTER 8. CONCLUSIONS AND FUTURE RESEARCH

8.1 Summary

In this study, we model the short-term link travel time variability as stochastic time-dependent random variables with discrete distributions. In the context of VANET, real-time travel time information is considered as a travel time interval with information quality represented by information accuracy and time delay. Based on this information framework, we proposed an adaptive learning processing to refine the short-term travel time distribution step by step. The learning process employs the knowledge of long-term historical travel time distribution as the start point and adaptively updates it using the obtained real-time travel time information. The realized short-term travel time distribution obtained in the last iteration will be used as the historical travel time distribution in the next iteration when new real-time information is available. The core of the proposed learning procedure is the information fusion model developed by the proposed nonlinear formulation (M1), which mathematically integrates the historical travel time information and real-time travel time information incorporating the information quality. To evaluate the efficiency of the proposed methodologies, we establish numerical experiments and also demonstrate their applications to online stochastic routing in VANETs.

Based on the results in the numerical experiments, the proposed information fusion model demonstrates good qualities in accuracy, robustness, and adaptivity. The short-term travel time estimation always predicts the travel time better than the long-term travel time distribution. Moreover, the mean of the short-term travel time distribution will gradually approach the actual travel time when the real-time information has good

quality. Furthermore, our sensitivity analysis indicates that the proposed information fusion models have the following characteristics.

i)   The proposed information fusion models are more sensitive to the information accuracy, p^r compared to time delay Δt and significance level, α.

ii)   Shorter time delay and relative higher significance level will improve the accuracy of short-term travel time estimations as the information accuracy is at a reasonably high level, such as greater than 0.3.

iii)   The improvement effect of the information fusion models on short-term travel time distribution takes some time; a shorter time delay will make the process quicker.

iv)   Good real-time information will always benefit the performance of the proposed information fusion models.

The application to online stochastic routing shows that the proposed information fusion models enable good prediction of the short-term travel time distribution so that the travelers have robust guidance. More importantly, when the quality of the real-time information is moderate, our information fusion model will provide a more accurate estimation about the travel time distribution. This will benefit the routing algorithm to find a better path for travelers than individually using historical or real-time travel time information.

8.2   *Future research directions*

There are several potential future research directions. First, the real-time formulation can be generalized to a discrete distribution and each state with the corresponding information quality. Then, the proposed information fusion models can be extended to fuse two distributions, namely historical travel time distribution and real-time travel time distribution. Information quality is still incorporated. We expect that the generated short-term travel time estimation will have better accuracy since the real-time information now provides more comprehensive insights about the actual travel time condition on the link. In addition, this study does not differentiate the information by vehicle classes. Involving this factor as part of the information fusion models is another important extension work. Lastly, when the field experiment data is available in the

future, more comprehensive experiments can be performed to further examine the efficiency of the proposed methodology. In general, the proposed study provides a methodology to fuse the information from multiple sources, where the framework of information is featured as a discrete distribution or a confidence interval with uncertain errors.

REFERENCES

Ahmed, M., and Cook, A. (1979). Analysis of freeway traffic time-series data by using Box-Jenkins techniques. *Transportation Research Record , 922*, 1-9.

Ahuja, R. K., Magnanti, T. L., and Orlin, J. B. (1993). *Network Flows: Theory, Algorithms, and Applications.* Prentice Hall, USA.

Anantharam, V. (1998). *Estimation Of Travel Time Distribution And Detection Of Incidents Based On Automatic Vehicle Classificatin*. Retrieved August 9, 2009, from California Partners for Advanced Transit and Highways (PATH): http://repositories.cdlib.org/its/path/papers/UCB-ITS-PWP-98-12

Artimy, M., Robertson, W., and Phillips, W. (2006, September 17-20). Minimum transmission range in vehicular ad hoc networks over uninterrupted highways. *In the 9th International IEEE Conference on Intelligent Transporation Systems* .

Bain, L. J., and Engelhardt, M. (2000). *Introduction To Probability And Mathematical Statistics* (2nd ed.). Duxbury Press.

Bajwa, S., Chung, E., and Kuwahara, M. (2005). Performance evaluation of an adaptive travel time prediction model. *Proceedings of 8th International IEEE Conference on Intelligent Transportation Systems.* Vienna, Austria.

Bellman, R. (1965). *Dynamic Programming.* Princeton, NJ, USA: Princeton University Press.

Chen, M., and Chien, S. I. (2001). Dynamic freeway travel time prediction using probe vehicle data: link-based vs. path-based. *Proceedings of TRB 80th Annual Meeting (CD-ROM).* Washington DC, USA.

Chien, S. I., and Kuchipudi, C. M. (2003). Dynamic travel time predictioin with real-time and historic data. *Journal of Transportation Engineering , 129* (6), 608-616.

Clark, S. (2003). Traffic prediction using multivariate nonparametric regression. *Journal of Transportation Engineering-ASCE , 129* (2), 161-168.

D'Angelo, M., Al-Deek, H., and Wang, M. (1999). Travel time prediction for freeway corridors. *Proceedings of 78th TRB Annual Meeting (CD-ROM).* Washington, DC.

Dia, H., and Berkum, E. C. (2001). An object oriented neural network approach to short-term traffic forecasting. *European Jouranl of Operation Research , 131* (2), 253-261.

Dong, J., and Mahmassani, H. S. (2009). Flow breakdown and travel time reliability. *Transportation Research Board 88th Annual Meeting.* Washington DC.

Du, L., and Ukkusuri, S. (2008(a)). Optimization models to characterize the broadcast capacity of vehicular ad hoc networks. *Accepted for publication on Transportaion Research Part C .*

Du, L., and Ukkusuri, S. (2008(b)). The relative mobility of vehicles improves the performance of information flow in vehicle ad hoc networks. *Accepted for publication on Networks and Spatial Economics .*

Fu, L. (2001). An adaptive routing algorithm for in-vehicle route gudance systems with real-time information. *Transportation Research Part B , 32* (7), 749-765.

Hall, R. W. (1986). The fast path through a network with random time-dependent travel time. *Transportation Science , 20* (3), 182-188.

He, R. R., Liu, H. X., Kornhauser, A. L., and Ran, B. (2002). *Temporal and special variability of travel time.* Technical Report, University of California, Irvine, Institute of Transportation Studies, Irvine, USA.

Hollander, Y., and Liu, R. (2008). Estimation of the distribution of travel times by repeated simulation. *Transportation Research Part C , 16*, 212-231.

Jain, A. K., Murty, M. N., and Plynn, P. J. (1999). Data clustering: a review. *ACM Computing Surveys , 31* (3), 264-322.

Jiang, G., and Zhang, R. (2003). Travel time prediction for urban aterial road: a case on china. *Vehicle Electronics Conference, 2001. IVEC 2001. Proceedings of the IEEE International.*

Jin, W. L., and Recker, W. W. (2006). Instantaneous information propagation in a traffic stream through inter-vehicle communication. *Transportation research. Part B : methodological , 40* (3), 230-250.

Kaelbling, L. P., Littman, M. L., and Moore, A. W. (1996). Reinforcement Learning: A Survey. *Journal of Arti , 4*, 237-285.

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

Kothuri, S. M., Tufte, K. A., Fayed, E., and Bertini, R. L. (2008). Toward understanding and reducing errors in real-time estimation of travel times. *Transportation Research Record: Journal of the Transportation Research Board , 2049*, 21-28.

Krishnan, R., and Polak, J. W. (2008). Short-term travel time prediction: an overview of methods and recurring themes. *Proceedings of the Transportation Planning and Implementation Methodologies for Developing Countries Conference (TPMDC).* Mumbai, India.

Kwon, J., Coifman, B., and Bickel, P. (2000). Day-to-day travel time trends and travel time prediction from loop detector data. *Transportation Research Record: Journal of Transportation Reserch Board , 1974*, 120-129.

Li, Y., and cDonald, M. (2002). Link travel time estimation using single GPS equipped probe vehicle. *Proceedings of The IEEE 5th International Conference On Intelligent Transportation Systems, IEEE.* Singapore.

Lin, H. E., Zito, R., and Taylor, M. A. (2005). A review of travel time prediction in transportation and logistics. *Proceedings of The Eastern Asia Society for Transportation Studies*, *5.*

Nanthawicit, C., Nakatsuji, T., and Suzuki, H. (2003). Application of probe-vehicle data for real-time traffic-state estimation and short-term travel-time prediction on a freeway. *Transportation Research Record , 1855*, 49-59.

Oda, T. (1990). An algorithm for prediction of travel time using vehicle sensor data. *Third International Conference on Road Traffic Control*, (pp. 40-44). London England.

Pant, P., Polycarpou, M., Sankaranarayanan, P., Li, B., Hu, X., and Hossain, A. (1998). Travel time prediction sysytems (TIPS) for freeway work zones. *ASCE Proceedings of the ICTTS'98 onTracffic abd Transportation Studies*, (pp. 20-29). Beijing, China.

Park, D., Rilett, L., and Han, G. (1999). Special basis neural networks for real-time travel time forecasting. *Journal of Transportation Engineering , 125* (6), 515-523.

Rakha, H., El-Shawarby, I., Arafeh, M., and Dion, F. (2006). Estimating path travel-time reliability. *Proceedings of the IEEE Intelligent Transportation Systems Conference 2006.* Toronto, Canada.

Roden, D. B. (1996). Forecasting travel time. *Transportation Research Record: Journal of Transportation Research Board , 1518*, 7-12.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal , 27*, 379-423 and 623-656.

Smith, B., Williams, B., and Oswald, R. (2002). Comparison of parametric and nonparametric models for traffic flow forecasting. *Transportation Research Part C: Emerging Technologies , 10* (4).

Stathopoulos, A., and Karlaftis, M. G. (2003). A multivariate state space approach for urban traffic flow modeling and prediction. *Transportation Research Part C: Emerging Technology , 11C* (2), 121-135.

Sun, H., Liu, H. X., Xiao, H., He, R. R., and Ran, B. (2003). Short-term traffic forecasting using the local linear regression model. *Transportation Research Board Annual Meeting (CD-ROM).* Washington, DC, USA.

Tam, M. L., and Lam, W. H. (2007). Real-time travel time estimation using automatic vehicle identification data in Hong Kong. *Lecture Notes in Computer Science , 4413*, 352–361.

Tu, J. (1996). Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal of Clinical Epidemiology , 49* (11), 1225-1231.

Turner, S. M. (1998). *Travel time data collection handbook.* Technical, U. S. Federal Highway Administration, Texas Transportation Institute, Office of Highway Information Management.

Ukkusuri, S., and Du, L. (2008). Geometric connectivity of vehicular ad hoc networks: Analytical characterization. *Transportation Research Part C , 16* (5), 615-634.

Van Der Voort, M., Dougherty, M., and Watson, S. (1996). Combining Kohonen maps with ARIMA time series models to forecast traffic flow. *Transportation research Part C: Emerging technologies , 4* (5), 307-318.

van Lint, J. W. (2008). Online learning solution for feeway travel time prediction. *IEEE Transactions on Intelligent Transportation Systems , 9* (1), 38-47.

van Lint, J. W. (2006). Reliable real-time framework for short-term freeway travel time prediction. *Journal of Transportation Engineering , 132* (12), 921-932.

Vlahogianni, E. I., Golias, J. C., and Karlaftis, M. G. (2004). Short-term traffic forecasting: overview of objectives and methods. *Transport Reviews , 24* (5), 533-557.

Wu, C. H., Ho, J., and Lee, D. (2004). Travel-time prediction with support vector regression. *IEEE Transactions on Intelligent Transportation Systems , 5* (4), 276-281.

Wu, C. H., Wei, C. C., Su, D. C., Chan, M. H., and Ho, J. M. (2003). Travel time prediction with support vector regression. *Proceedings of 2003 Conference on Intelligent Transportation Systems.* Shanghai, China.

Wunderlich, K. E., Kaufman, D. E., and Smith, R. L. (2000). Travel time prediction for decentralized route guidance architectures. *IEEE Transactions on Intelligent Transportation Systems , 1* (1), 4-14.

Yang, F., Yin, Z. Z., Liu, H. X., and Ran, B. (2004). On line recursive algorithm for short-term traffic prediction. *Transportation Research Record: Journal of Transportation Research Board. , 1897*, 1-8.

Yang, J. (2005). A study of travel time modeling via time series analysis. *Proceedings of 2005 IEEE Conference on Control Applications, 2005*, (pp. 855-860). Toronto, Canada.

Zhang, X., and Rice, J. A. (2003). Short-term travel time prediction. *Transportation Research Part C: Emerging Technology , 11C* (3-4), 187-210.

Zhong, M., Sharma, S., and Lingras, P. (2005). Short-term traffic prediction on different types of roads with genetically designed regression and time delay ANN models. *Journal of Computation in Civil Engineering , 19* (1), 94-103.

Ziliaskopoulos, A. K., and Zhang, J. (2003). A zero public infrastructure vehicle based traffic information System. *Proceedings of the 82th Transportation Research Board Annual Meeting (CD-ROM).* Washington, DC, USA.

Appendix A

This appendix shows the mean of estimated short-term travel time distribution for the 19 links in the study network.